Diversity within the Adenovirus fiber knob hypervariable loops influences primary receptor interactions.

Alexander T. Baker[1], Alex Greenshields-Watson[2], Lynda Coughlan[3], James A. Davies[1], Hanni Uusi-Kerttula[1], David K. Cole[2,4], Pierre Rizkallah[2], Alan L. Parker[1]*

[1] Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom
[2] Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff CF14 4XN, United Kingdom
[3] Icahn School of Medicine at Mount Sinai, Department of Microbiology, New York, New York 10029-6574
[4] Present address: Immunocore Ltd, Abingdon, Oxon, United Kingdom

Corresponding author:

*Alan L. Parker
Division of Cancer and Genetics
Henry Wellcome Building
Cardiff University School of Medicine
Heath Park
Cardiff
CF14 4XN
Email: ParkerAL@Cardiff.ac.uk

Keywords: adenovirus, receptor, CD46, CAR, crystal structure

Word count: Abstract: 150 words

Main body (excluding methods and figure legends): 4027

Figure count: 8 + 3 Supplementary figures + 1 Supplementary Table

Abstract

Adenovirus based vectors are of increasing relevance for wide ranging therapeutic applications. As vaccines, vectors derived from species D serotypes 26 and 48 are demonstrating promising efficacy as protective platforms against infectious diseases. Significant clinical progress has been made, yet definitive studies underpinning mechanisms of entry, infection, and receptor usage are currently lacking. Here, we performed structural and biological analysis of the receptor binding fiber-knob protein of Ad26/48, reporting crystal structures, and modelling putative interactions with two previously suggested attachment receptors, CD46 and CAR. We provide clear evidence of low affinity interaction with CAR, with modelling suggesting affinity is attenuated through extended, semi-flexible loop structures, providing steric hindrance. Conversely, *in silico* and *in vitro* experiments were unable to provide evidence of interaction between Ad26/48 fiber-knob with CD46. Our findings provide new insight to the cell-virus interactions of Ad26/48, with important implications for the design and engineering of optimised Ad-based therapeutics.

40 **ARTICLE**

41 Adenoviruses (Ads) are increasingly important vectors for wide ranging therapeutic interventions,
42 from gene delivery and oncolytic agents to platforms for vaccine applications[1–3]. As vaccine vectors,
43 their use clinically has been popularised by their excellent safety profile coupled with their ability to
44 induce robust cellular and humoral immunogenicity in humans[4]. Phylogenetically, Ads are diverse
45 and subdivided across 7 species, A-G[5], based classically on serological cross-reactivity, receptor
46 usage, haemagglutination properties and, more recently, phylogenetic sequence similarity[6,7].

47 Most experimental and clinical studies have focussed on the well-studied species C adenovirus, Ad5.
48 Although potently immunogenic, the efficacy of vaccines based on Ad5 appears hampered by high
49 seroprevalence rates in humans, and enthusiasm for their use as clinical vaccine platforms has been
50 dampened by the well-publicised failure of the MERCK STEP vaccine trial. This trial, to evaluate an
51 Ad5-based HIV vaccine encoding HIV gag/pol/nef antigens, was abandoned due to apparent lack of
52 efficacy upon 1st term analysis. The study also identified a non-significant trend towards increase HIV
53 acquisition in a specific high-risk, uncircumcised subset of patients who also had high levels of
54 baseline pre-existing neutralising antibodies to Ad5[8,9]. As a result, attention has switched from Ad5
55 based vectors towards the development of alternative adenoviral serotypes with lower rates of pre-
56 existing immunity. Most notably, vectors under development include those based on species D
57 serotypes including Ad26, which has entered Phase-III clinical trials as an Ebola vaccine and recently
58 reported promising immunogenicity in an HIV trial, or chimeric vectors utilising the hexon hyper
59 variable regions (HVRs) of Ad48 which have undergone Phase-I evaluation as an HIV vaccine[3,10–12].
60 However, despite extensive clinical advances using these vaccine vectors, we possess very limited
61 knowledge of their basic biology, particularly with regards to the determinants underpinning their
62 tropism, mechanisms of cellular entry and receptor usage. Therefore, in this study, we address these
63 shortcomings through analysis of adenoviral diversity in the context of their receptor binding, fiber
64 proteins. Whilst Ad are historically divided into seven species, A-G, this may underestimate their
65 diversity[13–15]. Phylogenetic examination of 56 human Adenovirus fiber proteins from different
66 species shows deviation from the taxonomy expected based upon whole genome taxonomy, likely
67 due to recombination events as seen in other Adenoviral proteins[14]. Here, we have sought to
68 generate high resolution crystallographic structures of the cellular interacting fiber-knob domain of
69 species D Adenoviruses Ad26 and Ad48, to enable modelling of their putative interactions with
70 either human Coxsackie and Adenovirus Receptor (CAR) or CD46 (Membrane Cofactor Protein,
71 MCP), both of which have both been previously suggested as primary receptors for these clinically
72 significant viral vectors[16,17]. In this study, we employ an integrative workflow utilising X-ray
73 crystallography, *in silico* modelling, and *in vitro* assays to dissect previous findings suggestive of

74 interactions of Ad26 and Ad48 with CAR[5,18] and for CD46[16,17,19–21] . Our findings provide shed new
75 light on the cell-virus interactions of adenovirus and with potential implications for the design and
76 engineering of optimised Ad-based therapeutics, both for vaccine applications and oncolytic
77 development, allowing us to minimise off-target or undesirable interactions *in vivo*.

78 **RESULTS**

79 **Genetic variability in adenovirus fiber-knob protein**

80 We generated phylogenetic trees of human adenoviruses 1-56 which revealed broader
81 differentiation in the fiber-knob domain (Fig. 1A) than might be expected based upon taxonomy of
82 the whole virus (Fig. 1B). In both phylogenetic trees the adenoviruses divide into seven clades
83 corresponding to the seven adenoviral species, A-G. However, the tree based upon the fiber-knob
84 domain (Fig. 1A) shows the species D adenoviruses forming a greater number of sub-groups than in
85 the whole genome tree suggesting greater diversity in the receptors of this species than might be
86 expected when comparing serotypes. This may be the result of recombination events, as reported
87 previously for other adenoviral proteins[13,15,22–25]. The opposite is observed in species B adenoviruses,
88 where simpler groupings are seen when analysing fiber-knob domain alignment than by whole
89 genome analysis. When divided into sub-species based on whole genome they divide into species B1
90 and B2, but when the tree is generated based on fiber-knob alone the species B viruses do not divide
91 into similar groups. The significance of this fiber-knob diversity is unclear, but it has previously been
92 suggested that the species B1/B2 designation may more closely represent the tissue tropism, than
93 receptor usage[26,27].

94 We next calculated the amino acid variability at each position in the aligned adenoviral fiber-knob
95 sequences, which revealed regions of broad conservation corresponding to β-strands which make up
96 the main fold of the fiber-knob trimer (Fig. 1C). The positions corresponding to the β-strands of Ad5,
97 as originally reported by Xia et al[28], are shown by arrows. Comparison shows that the more N-
98 terminal (A, B, and C) and C-terminal (I and J) β-strands have greater homology across the adenoviral
99 species than the other sheets. This may relate to the intervening loops between the less tightly
100 conserved β-strands (D, E, F, G, H) being more apical, a region which is often involved in receptor
101 interactions[27,29–33].

102 **Structural analysis of Ad26 and Ad48 fiber-knob domain**

103 To investigate diversity of species D adenoviruses, recombinant, 6-His tagged fiber-knob protein
104 from Ad26 and Ad48 (hereafter referred to as Ad26K and Ad48K) were generated, purified and used
105 to determine X-ray crystallographic structures of Ad26 (PDB: 6FJN) and Ad48 (PDB: 6FJQ) at
106 resolutions of 0.97Å and 2.7Å, respectively (Fig. 2A-D). The monomers (Red, Green, Blue) form an
107 anti-parallel β-barrel, typical of Adenoviral fiber-knob protein, as described by Xia *et al* (PDB 1KNB)[28].
108 Each monomer interacts with two neighbouring copies to form homotrimer with 3-fold symmetry
109 (Fig. 2B,D) and a highly stable interface (suppl. 1A). Stability analysis using PISA (Protein Interactions,

110 Surfaces and Assemblies) software calculates the Ad26 and Ad48 trimers to have >20% lower
111 interface energy than that of Ad5, indicative of a more stable interaction (Suppl. 1A) between the
112 monomers of the Ad26 and Ad48 fiber-knobs.

113 As with the pan species analysis (Fig. 1C), variability of the aligned species D fiber-knobs (Fig. 2E)
114 confirmed that the β-strands comprising the hydrophobic core of the trimers are highly conserved,
115 with β-strands demonstrating a high degree of overlap in both spatial position and sequence
116 variability (suppl. 1B).

117 **Hypervariable loops within the fiber-knob domain are stabilised by inter-loop interactions**

118 Particularly relevant to this study are the DG, GH, HI and IJ loops, linking the indicated strands
119 corresponding to those in the originally reported Ad5 fiber-knob loops (suppl. 1B)[28]. These loops
120 have previously been shown to be critical in engagement of CD46 for Ad11, 35, and 21[34–37].
121 Alignment of these loops with the corresponding loops of Ad5 (species C, CAR interacting), Ad35
122 (species B, CD46 interacting), Ad11 (species B, CD46 interacting), and Ad37 (species D, CAR and Sialic
123 acid/GD1a glycan interacting) reveals different topologies in these critical receptor interacting
124 regions (Fig. 3A)[36–39]. The HI loops of Ad26K and Ad48K are most homologous to those of Ad35K and
125 Ad37K respectively in terms of amino acid sequence identity (Suppl.1B) and spatial alignment. The
126 Ad26K DG loop is most homologous to Ad35K but incorporates a three amino acid insertion, while
127 Ad48K DG loop displays a differing and unique topology. The GH and IJ loops of Ad5, 26, 37, and 48K
128 demonstrate similar spatial arrangements (though the similarity does not extend to the sequence
129 identity) but differ from the CD46 utilising Ad11K and Ad35K.

130 The high-resolution structures obtained allowed us to robustly characterise the loops, seen in the
131 electron density maps (Fig. 3B, C). To assess loop flexibility and mobility, we assessed the B-factors
132 (also known as temperature factor), a measure of the confidence in the position of an atom which
133 can be used to infer flexibility. By assessing the B-Factors, the relative flexibility of the moieties of
134 interest can be inferred. While the apical domains of some loops demonstrated increased B-Factors
135 relative to the rest of the molecule, the loops' B-factors are surprisingly low (Fig. 3B, C), suggesting
136 that they may exhibit limited flexibility.

137 To investigate whether the different loop conformations were the product of flexibility, or
138 stabilisation we investigated the inter-loop interactions in the Ad26 and Ad48 structures. This
139 analysis demonstrated that the GH loop of Ad26K (like those of Ad5, 37 and 48K) does not extend
140 directly away from the G and H β-sheets, but forms a β-hairpin (Fig. 3A, B) that is maintained by
141 seven polar contacts within the neighbouring IJ loop which restrict the loops' orientation (Fig. 4A,

142 Suppl. 2A). Polar contacts were also observed at the apex of several loops, most notably the GH and
143 CD loops of both fiber-knobs (Fig. 4A, B). The IJ loops form fewer intramolecular polar contacts but
144 are stabilised by interactions with the adjacent CD and DG loops (Suppl. 2C, D). These interactions
145 retain the apical residues in a stable conformation, rather than leaving the side chains fully labile.

146 The B-Factors of the Ad48K DG loop were observed to be polarised normal to the hairpin, with the
147 outer face of the loop having higher B-Factors compared to the inner face (Fig.4C). This is likely the
148 result of polar contacts formed between Ser-307, Gln-308, Ala-309, and Leu-304 with Asp-359 and
149 Gln-357 of the opposing monomer stabilising the conformation of the DG loop. The proline rich
150 nature of this loop provides further rigidity (Suppl. 3).

151 Based on this analysis of the inter and intra-loop bonds we suggest that these adenoviral loops may
152 not be fully flexible variable regions, but organised receptor engagement motifs with carefully
153 evolved structures. This has direct implications for receptor engagement of these viruses, as the
154 loops govern previously characterised interactions with CAR and CD46, and are directly involved in
155 their pathogenicity[26,34,40,41].

156 **Modelling of Ad26K and Ad48K in complex with CAR suggests a low affinity binding interaction**

157 Both CD46 and CAR have been proposed as primary attachment receptors for Ad26 and Ad48[16,17].
158 Previously generated crystal structures of Ad11K in complex with full length CD46 (PDB: 3O8E), and
159 Ad37K in complex with CAR-D1 domain (PDB: 2J12) reveal the loops to be essential to receptor
160 interactions[42]. To investigate the ability of Ad26K and Ad48K to bind these receptors we generated
161 homology models by alignment of the new Ad26K and Ad48K fiber-knob structures modelled with
162 the existing fiber-knobs in complex with the receptor of interest and performed energy minimisation
163 to optimise the conformation to achieve the lowest possible energy interface with which to analyse
164 the interaction. We performed similar experiments with the well described CAR and CD46 binding
165 fiber-knob proteins of Ad5 (PDB: 6HCN) and Ad35 (PDB: 2QLK), respectively, as controls.

166 Modelling of Ad26K and Ad48K in complex with the CAR-D1 domain revealed a region of high
167 homology with the CAR utilising Ad5 fiber-knob, here after termed the α-interface (Fig. 5A).
168 Sequence alignment with Ad5K shows that many of the residues previously shown to be critical for
169 CAR interaction in Ad5K are conserved in Ad26K and Ad48K, including Ser-408, Pro-409, and Tyr-
170 376[38]. The same is true of residues predicted to interact with CAR directly, such as Lys-417 (number
171 is for Ad5K). The residues predicted to form direct CAR binding interactions for Ad5K, 26K, and 48K
172 are pictured in complex with the maximum spatial occupancy of the energy minimised CAR-D1 (Fig.
173 5C). The high levels of homology are seen to extend to the fold as well as the linear sequence.

174 Calculation of the binding energies for the modelled complexes predicted a stable α-interface for all
175 complexes modelled, albeit weaker for the non-CAR utilising Ad35K (Fig.5D) which has low sequence
176 conservation with Ad5K. However, the interaction is complicated by a second CAR interface, termed
177 the β-interface (Fig. 6A). The loops forming the β-interface with CAR-D1 differ between Ad5, 26, and
178 48 fiber-knob (Fig. 6B). The shorter Ad5K DG loop does not clash with the CAR-D1 surface, whereas
179 the extended Ad26K DG loop forms a partial steric clash, with surface seen to clash with the aligned
180 CAR-D1, and Ad48K DG loop is seen to form an even larger steric clash. Whilst the longer loop of
181 Ad26K is expected to be more flexible than that of Ad5K, the Ad48K DG loop is surprisingly stable
182 due to the characteristics described (Fig.4C).

183 **In vitro IC_{50} evaluation of Ad26K and Ad48K confirms diminished affinity CAR interaction**
184 **compared to Ad5K.**

185 Our modelling studies indicate that the longer, inflexible DG loop of Ad48 would be likely sterically
186 hinder the Ad48K: CAR interaction at the β-interface to a greater extent than the more modest
187 inhibition of the smaller and more labile loop of Ad26K, which in turn would exhibit more inhibition
188 of CAR binding than that of Ad5K, where no steric inhibition is observed. Competition inhibition
189 assays using recombinant fiber-knob protein to inhibit antibody binding to CAR receptor in CHO-CAR
190 cells (which express CAR, while the parental cell line is established to be non-permissive to Ad
191 infection) support our observation (Fig. 6C). The IC_{50} (the concentration of protein required to inhibit
192 50% of antibody binding) of Ad5 is 7.0 ng/10^5 cells, while Ad26K and Ad48K demonstrate IC_{50} values
193 15.7 and 480 times higher at 0.110μg/10^5 cells and 3.359μg/10^5 cells, respectively, reflecting their
194 reduced ability to engage CAR due to steric inhibition at the β-interface (Fig. 6C).

195 **Modelling of Ad26K and Ad48K in complex with CD46 indicates minimal interaction**

196 A similar approach was taken to model Ad26K and Ad48K in complex with CD46. Alignments with the
197 previously published Ad11K-CD46 complex were generated and energy minimised to obtain the
198 lowest energy state of the complex [35,43]. This interface utilises loops DG, GH, HI and IJ to form a
199 network of polar interactions with the CD46 Sc1 and Sc2 domains (Fig. 7A)[36].

200 When Ad35K was modelled in complex with CD46, a network of polar contacts between Ad35 and
201 CD46 was predicted (Fig. 7B) similar to that observed in the Ad11-CD46 complex crystal structure,
202 PDB: 3O8E (Fig. 7A)[35]. Previously Ad35K residues Phe-243 , Arg-244, Tyr-260, Arg-279, Ser-282, and
203 Glu-302 (underlined in Fig.7C) have been implicated as key contact residues for CD46 interaction,
204 and are conserved in Ad11K (highlighted in blue, Fig.7C)[37]. Our modelling suggests that conversely,
205 Ad26K and Ad48K are predicted to form very few polar contacts with CD46 with just 2 contacts

206 predicted for Ad26K (Fig. 7D) and 3 predicted for Ad48K (Fig. 7E). Furthermore, they do not share
207 any of the critical CD46 binding residues which have been reported previously (underlined, Fig. 7C)
208 for Ad11K and Ad35K, or any of the predicted interacting residues (blue highlight, Fig. 7C).

209 We again employed PISA to calculate the binding energy of the various modelled and energy
210 minimised fiber-knob CD46 complexes (fig.8A). Ad11K, the strongest known CD46 binding
211 adenovirus[34], was predicted to have the lowest binding energy reflecting its high stability interface,
212 with Ad35 demonstrating a similar but slightly reduced binding energy. Conversely, Ad26K and
213 Ad48K are predicted to have lower binding energies, similar to that which may be expected for
214 random proteins passing in solution, indicating that an interaction between either the Ad26K, or
215 Ad48K with CD46 is unlikely[44].

216 While still low compared to the known CD46 interacting Ad11K and Ad35K binding energies, that for
217 Ad5K was higher than expected for a known non-CD46 interacting adenovirus (Fig. 8A). Inspection of
218 the model shows that this is due to the close proximity of the large Ad5 HI loop to CD46 (Fig. 8B).
219 The residues involved in the predicted interaction are not conserved in any known CD46 interface
220 and suggesting these are random interaction. Furthermore, interaction between the DG loop and
221 CD46 is integral to known CD46 binding interfaces and is prevented by the Ad5K HI loop laying
222 between them (Fig. 8B).

223 **IC_{50} evaluation of Ad26K and Ad48K confirms lack of CD46 interaction**

224 Antibody competition inhibition assays in CHO-BC1 cells (CHO cells transduced to express the BC1
225 isoform of CD46) were used to test the predictions made by modelling (Fig. 8C). These data confirm
226 that recombinant Ad26K and Ad48K proteins are incapable of inhibiting antibody binding to CD46 at
227 any concentration tested (up to 2ng/cell), whilst the well-defined CD46 interacting Ad35K
228 demonstrates strong inhibition, with a calculated IC_{50} of 0.003µg/10^5 cells. Based on these findings it
229 is highly unlikely that Ad26 or Ad48 are capable of engaging CD46 as a primary receptor.

230 **DISCUSSION**

231 This study reveals the crystal structure of two adenovirus proteins critical to primary receptor
232 engagement, Ad26 and Ad48 fiber-knob, which are important viral vectors currently in human
233 clinical trials[3,11,12]. Despite their advanced development, the field lacks fundamental knowledge
234 regarding the mechanisms of infection for these viral vector platforms. The work we described here
235 provides a combined crystallographic, *in silico,* and *in vitro* approach to investigate adenovirus fiber-
236 knob: receptor interactions with CAR and CD46, two receptors previously proposed to be utilised by
237 these viruses[5,9,17].

238 Analysis of the phylogenetic relationship between 56 Adenovirus serotypes, both whole genome and
239 fiber-knob domain alignment, confirms diversification into seven adenoviral species[7]. However,
240 generating the phylogenetic tree with fiber-knob sequences, rather than whole genomes, shows
241 additional diversity, not revealed by the traditional taxonomic approach. Adenovirus species D
242 breaks up into several additional sub-clades when focused on the fiber-knob, suggesting greater
243 receptor diversity than might be expected based on the whole virus phylogeny. Similar observations
244 have previously been made in species D hexon and penton[13].

245 In contrast to species D, the phylogeny of species B adenoviruses, which are known to utilise
246 desmoglein 2 and CD46 as primary receptors, is simplified when focused upon the fiber-knob,
247 indicating less diverse receptor usage[40,45–47]. This simplification in comparison to the whole genomic
248 alignment implies that much of the species diversity must lay in other proteins. The E3 protein, for
249 example, is known to be highly diverse within species B adenoviruses, having previously been
250 exploited in the selection of the oncolytic (cancer killing) virus enadenotucirev, which is currently in
251 clinical trials[13,48,49].

252 That we see such opposing effects on the species B and D phylogenetic trees when focusing on the
253 fiber-knob, highlights the limitations of simple taxonomic approaches. The current Adenoviral
254 taxonomy is based on antibody neutralisation assays, which are limited by antibodies reliance on
255 surface accessible proteins in the capsid, and does not account for diversity in other viral proteins, as
256 the above suggests for species B. This supports a taxonomic proposal based upon viral genetics
257 rather than antibody neutralisation, as has previously been suggested[13,50].

258 Many studies on adenovirus neutralisation have focused upon neutralising antibodies (NAbs) which
259 bind to the hexon[51–53]. Following intramuscular vaccination with non-replicating adenoviral vectors,
260 most NAbs are targeted to the hexon; a reflection of its high abundance and surface availability in
261 each viral capsid[51,54]. However, during natural infection, many NAbs target the fiber protein[54],

262 presumably due to the abundance of fiber produced in the early stages of AdV replication to loosen
263 cell-cell junctions and facilitate viral spread, prior to lysis and entry of large amounts of whole virus
264 to the blood stream[55]. For individuals with pre-existing anti-adenovirus immunity derived from
265 natural infections, anti-fiber NAbs are likely to limit the use of vectors with common adenovirus fiber
266 proteins by neutralisation of the viral vector prior to its therapeutic effect. Thus, for the
267 development of vectors to circumvent pre-existing anti-adenovirus immunity for therapeutic use
268 further exploration of this fiber protein diversity may be beneficial, as well as the on-going studies
269 using hexon HVR pseudotypes to circumvent anti-hexon immunity[11,56].

270 Analysis of the adenovirus loops reveals an intricate network of polar interactions which stabilise
271 their three-dimensional structures. These bonds appear to hold the loops in a conformation which,
272 in the case of Ad11 and Ad35, facilitates receptor binding. In the Ad26 and Ad48 fiber-knob
273 structures presented in this study the loops are also held in a stable conformation, though not one
274 which enables CD46 interaction.

275 Modelling of Ad26K/Ad48K in complex with CD46 suggested few contacts, and interface energy
276 calculations using these models predict a weak binding energy. This was confirmed by *in vitro*
277 competition inhibition assays. These findings appear contradictory with previous studies which
278 suggest CD46 as the primary receptor for these viruses. However, our findings improves our
279 knowledge of the cell entry mechanisms of these viruses and the vectors derived from them, and
280 does not diminish the observed effectiveness of these vaccines. However, if CD46, a protein
281 expressed on the surface of all nucleated cells, is not the receptor for these viruses then it is as yet
282 unknown what the primary tissue tropism determinant is for these clinically significant viruses[57,58].

283 A similar methodology was applied to the interaction Ad26K/Ad48K with CAR. Inspection of the
284 modelled complexes indicated a conserved α-interface enabling CAR binding in adenovirus 5, 26, and
285 48, fiber knobs. However, the structure of the β-interface interaction appears to indicate a novel
286 mechanism modulating the fiber-knob: CAR affinity. When occupying the intermonomer cleft in the
287 conformation shown in figure 6B, the DG loops of Ad26 and Ad48 are likely to inhibit CAR binding by
288 steric hindrance, but if the loops were to shift into a conformation which relieves this clash CAR
289 binding could occur. Therefore, the ability of these vectors to interact with CAR is a function of the
290 steric hindrance provided by these loops, sterically hindering the ability of the fiber-knob domain to
291 engage CAR in a permissive conformation.

292 The inter-loop contacts described in Figure 4 A, B, and supplemental figure 2, and the normalised B-
293 factors described in figure 6B will influence the molecular dynamics of the DG-loops. Loops which
294 can occupy a CAR inhibitory conformation but have fewer stabilising contacts, such as that of Ad26K

295 (Suppl. 2A, C), should be more permissive to CAR binding. While loops which are less flexible and/or
296 stabilised in a CAR inhibitory conformation, such as Ad48K (Fig. 6B, Suppl. 2B, D) should result in a
297 fiber-knob which is less able to bind CAR. This hypothesis fits the competition inhibition studies
298 shown in Figure 6C, which demonstrate that Ad26K has and apparent ~15x lower affinity for CAR
299 than Ad5K, and Ad48K has 500x lower affinity.

300 Species D adenoviruses (Suppl. 3) have a large range of different DG loops. Most sequences have
301 lengths equal to, or greater than, that of Ad26K, making it plausible that they too could modulate
302 the fiber-knob interaction with CAR. However, the magnitude of this effect will be dependent on the
303 individual molecular dynamics of the DG-loops and its interactions with adjacent residues.

304 Assuming this mechanism of CAR binding regulation is broadly applicable, it may have important
305 implications for adenoviral vector design. The presence of a high affinity receptor for the virus can
306 mask the low affinity CAR interaction, creating a "hidden tropism" only observed if the virus is forced
307 to rely upon it. Expression of CAR on human erythrocytes suggests the potential for sequestration of
308 virotherapies in the blood[59]. CAR expression in lung epithelial tissues offers another site for potential
309 off target activity[60,61]. Therefore, many virotherapies previously thought to be non-CAR binding
310 adenoviruses may in fact demonstrate weak CAR tropism, driving off target infections or resulting in
311 sequestration of the vector in organs other than that target. This may not be of grave consequence
312 for non-replicating vectors, such as viral vaccines, but in vectors which rely upon controlled
313 replication of the virotherapy in the target tissues, such as cancer targeting oncolytic vectors, this
314 could result in off-target infection, dysregulated expression of therapeutic proteins, and reduced
315 delivery to the desired point of need.

316 The work undertaken in this study presents, for the first time, the crystal structures of the fiber-knob
317 domain protein of adenovirus 26 (PDB: 6FJN), and adenovirus 48 (PDB: 6FJQ) fiber-knob protein. In
318 addition we report a new crystal structure for adenovirus serotype 5 fiber-knob protein (PDB: 6HCN)
319 with improved resolution compared to the existing structure (PDB: 1KNB)[28]. We utilise these
320 structures to investigate the ability of these proteins to interact with the putative receptors, CAR and
321 CD46, through computational modelling and *in vitro* experimentation. We demonstrate that
322 adenovirus 26 and 48 fiber-knob domains have a weak ability to bind CAR, and no demonstratable
323 CD46 interaction, suggesting that these viruses are unlikely to utilise these proteins as a primary
324 receptor *in vivo*. We suggest that CAR binding is moderated by a previously unreported mechanism
325 of steric inhibition which may apply to other adenoviruses and demonstrate an *in silico* methodology
326 capable of rapidly predicting these interactions. These findings enhance our understanding of the

327 virology of adenovirus infection, and have direct implications for virotherapy vector design, which
328 often rely upon carefully controlled receptor tropisms to achieve specificity and efficacy[9,18,29,62].

329

330 **METHODS**

331 **Analysis of adenovirus genetic diversity**

332 **Genome alignment**

333 Representative whole nucleotide genomes of adenoviral species 1-56 were selected from the
334 National Center for Biotechnology Information (NCBI), and aligned using the EMBL-EBI Clustal
335 Omega tool[63–65]. Fiber-knob domain amino acid sequences were derived from the same genome
336 sequences, defined as the translated nucleotide sequence of the fiber protein (pIV) from the
337 conserved TLW hinge motif to the protein C-terminus, and aligned in the same manner as the whole
338 genomes. Phylogenetic relationships were inferred using the maximum likelihood method based
339 upon the Jukes Cantor model for the whole genome nucleotide analysis[66], and the Poisson
340 correction model for the fiber-knob amino acid analysis[67], using the MEGA X software[68]. Confidence
341 was determined by bootstrap analysis (500 replicates)[69] and trees displayed condensed at 70%
342 confidence (percentage confidence values shown at each node).

343 **Fiber-knob amino acid variability:**

344 Amino acid sequence variability scores were calculated from the Clustal omega aligned amino acid
345 sequences of the fiber-knob domains of either Adenoviruses 1-56, or only the species D
346 adenoviruses. Analysis was performed using the protein variability server (PVS), using a consensus
347 base sequence and the Wu-Kabat method[70,71].

348 **Expression and purification of Recombinant Fiber-Knob protein:**

349 Purification of recombinant Ad26K and Ad48K proteins was performed as previously described[18,72,73].
350 Briefly, fiber-knob DNA sequences were synthesised commercially by GenScript (Nanjing, China) with
351 flanking BamHI and HindIII restriction digest sites, in frame, cloned into pQE-30 vector by BamHI,
352 HindIII double digest and ligation, and the resultant plasmid was transformed into SG13009 *E.coli* by
353 heat-shock. For expression and purification, E.coli were resuspended in lysis buffer (50mM Tris,
354 pH8.0, 300mM NaCl, 1% (v/v) NP40, 1mg/ml Lysozyme, 1mM β-mercaptoethanol) and incubated at
355 room temperature for 30mins. Lysate was clarified by centrifugation at 30,000g for 30mins and
356 filtered through a 0.22µm syringe filter (Millipore, Abingdon, UK). Clarified lysate was then loaded
357 onto a 5ml HisTrap FF nickel affinity chromatography column (GE) at 2.0ml/min and washed with 5
358 column volumes into elution buffer A (50mM Tris [pH8.0], 300mM NaCl, 1mM β-mercaptoethanol).
359 Protein was eluted by 30min gradient elution from buffer A to B (buffer A + 400mM Imidazole).
360 Fractions were analysed by reducing SDS-PAGE, and Fiber-knob containing fractions further purified
361 using a superdex 200 10/300 size exclusion chromatography column (GE) in crystallisation buffer (10

362 mM Tris [pH 8.0] and 30 mM NaCl). Fractions were analysed by SDS-PAGE and pure fractions
363 concentrated by centrifugation in Vivaspin 10,000 MWCO (Sartorius, Goettingen, Germany)
364 proceeding crystallisation.

365 **Competition Inhibition Assays:**

366 **Sample Preparation:**

367 Competition inhibition assays were performed as previously described[18]. Briefly, CHO cells
368 expressing the appropriate receptor (CAR: CHO-CAR, CD46: CHO-BC1), were pre-incubated with
369 0.0001-100µg/10^5 cells of recombinant soluble knob protein. Primary staining was performed with
370 anti-CD46 antibody (MEM-258) or anti-CAR antibody (RmcB) to complex receptors unbound by fiber-
371 knob protein, and secondary staining was performed with Alexa-647 labelled goat anti-mouse
372 F(ab')2 (ThermoFisher) [18,72,73]. All antibodies were used at a concentration of 2µg/ml.

373 **Flow cytometric analysis:**

374 Samples were run in triplicate and analysed by flow cytometry on Attune NxT (ThermoFisher), and
375 analysed using FlowJo v10 (FlowJo, LLC) by gating on singlets, cell population, and Alexa-647 positive
376 cells, respectively. Total fluorescence (TF) was used as the measure of inhibition, rather than
377 percentage of fluorescent cells in the total population, to account for the presence of multiple
378 receptor copies per cell surface which can enable partial inhibition of antibody binding on the cell
379 surface, while the cell remains fluorescently labelled. TF was defined as the percentage of Alexa-647
380 positive cells in the single cell population for each sample and multiplied by the median fluorescent
381 intensity (MFI) of the Alexa-647 positive single cell population in each sample. Data points are the
382 mean total fluorescence of three biological replicates with error given as the standard deviation
383 from the mean. IC_{50} curves were fitted by non-linear regression as previously described, and used to
384 determine the IC_{50} concentrations[18,72,73].

385 **Crystallisation and structure determination.**

386 Protein samples were purified into crystallisation buffer (10 mM TRIS [pH 8.0] and 30 mM NaCl). The
387 final protein concentration was approximately 7.5 mg/ml. Commercial crystallisation screen
388 solutions were dispensed into 96-well plates using an Art-Robbins Instruments Griffon dispensing
389 robot (Alpha Biotech, Ltd), in sitting-drop vapour-diffusion format. Drops containing 200 nl of screen
390 solution and 200 nl of protein solution were equilibrated against a reservoir of 60 µl crystallisation
391 solution. The plates were sealed and incubated at 18°C.

392 Crystals of Ad5K appeared in PACT *Premier* condition D04 (0.1 M MMT, pH 7.0, 20% PEG 1500),
393 within 1 to 7 days. Crystals of Ad26K appeared within 1 to 7 days, in PACT *Premier* (Molecular
394 Dimensions, Suffolk, UK) condition A04 (0.1 M MMT [DL-Malic acid, MES monohydrate, Tris], pH 6.0,
395 25% PEG 1500. Crystals of Ad48K appeared in PACT *Premier* condition D02 (0.1 M Bis-Tris-propane,
396 pH 6.5, 20% PEG 3350, 0.2M NaNO3), within 2 weeks. Crystals were cryoprotected with reservoir
397 solution to which ethylene glycol was added at a final concentration of 25%. Crystals were harvested
398 in thin plastic loops and stored in liquid nitrogen for transfer to the synchrotron. Data were collected
399 at Diamond Light Source beamline I04, running at a wavelength of 0.9795Å. During data collection,
400 crystals were maintained in a cold air stream at 100°K. Dectris Pilatus 6M detectors recorded the
401 diffraction patterns, which were analysed and reduced with XDS, Xia2[74], DIALS, and Autoproc[75].
402 Scaling and merging data was completed with Pointless, Aimless and Truncate from the CCP4
403 package[76]. Structures were solved with PHASER[77], and COOT[78] was used to correct the sequences
404 and adjust the models. REFMAC5[79] was used to refine the structures and calculate maps. Graphical
405 representations were prepared with PyMOL[80]. Reflection data and final models were deposited in
406 the PDB database with accession codes: Ad5K, **6HCN**; Ad26k, **6FJN**; and Ad48k, **6JFQ**.

407 **Energy Minimisation:**

408 Energy Minimisations were performed using the YASARA self-parametrising energy minimisation
409 algorithm as performed by the YASARA energy minimisation server, and results were visualised in
410 PyMoll[43,80,81].

411 **Calculation of Interface Energy:**

412 Interface energies were calculated using QT-PISA using biological protein assemblies and excluding
413 crystallographic interfaces[82]. Values are the mean of the three symmetrical interfaces in each trimer
414 and error is the standard deviation from the mean, any values above -3.0 $kcalmol^{-1}$ were considered
415 to be background as shown as a red dashed line on graphs[44].

416 Sequence visualisations:

417 Performed using the Clustal Omega multiple sequence alignment algorithm and visualised with
418 BioEdit[63–65,83].

419 **B-Factor Normalisation:**

420 Comparing order between different structures by comparing individual B-factors can be misleading.
421 Post-refinement B-factors relate to the Wilson B-factor, which can vary widely between data sets,
422 even from the same crystal preparation. A valid comparison between different structures can be

423 achieved by comparing normalised B-factors instead. Normalisation was performed by dividing
424 individual atomic B-factors by the average B-factor of the whole data set, quantifying the range of
425 internal flexibility in a structure. This normalised B- factor can then be compared between different
426 data sets.

427 **Acknowledgments:**

428 *The authors would like to thank Diamond Light Source for beamtime (proposal mx14843), and*
429 *the staff of beamline I04 for assistance with data collection. We are also grateful to Elmar Kreiger*
430 *and the team who maintain the YASARA energy minimisation server, as well as Pedro Reche and the*
431 *Immunomedicine Group for maintenance of the Protein Variability Server.*

432 **REFERENCES**

433 1. Combination Adenovirus + Pembrolizumab to Trigger Immune Virus Effects - Full Text View -
434 ClinicalTrials.gov. Available at:
435 https://clinicaltrials.gov/ct2/show/NCT02798406?term=adenovirus&recrs=a&cond=Glioblastom
436 a&rank=2. (Accessed: 23rd September 2017)
437 2. A Safety and Efficacy Study of ChAdOx1 LS2 and MVA LS2 - Full Text View - ClinicalTrials.gov.
438 Available at:
439 https://clinicaltrials.gov/ct2/show/NCT03203421?term=adenovirus&recrs=a&cond=malaria&ra
440 nk=2. (Accessed: 23rd September 2017)
441 3. Long-term Safety Follow-up of Participants Exposed to the Candidate Ebola Vaccines
442 Ad26.ZEBOV and/or MVA-BN-Filo - Full Text View - ClinicalTrials.gov. Available at:
443 https://clinicaltrials.gov/ct2/show/NCT02661464?term=adenovirus&recrs=a&cond=ebola&rank
444 =1. (Accessed: 23rd September 2017)
445 4. Lee, C. S. *et al.* Adenovirus-mediated gene delivery: Potential applications for gene and cell-
446 based therapies in the new era of personalized medicine. *Genes Dis.* **4,** 43–63 (2017).
447 5. Abbink, P. *et al.* Comparative Seroprevalence and Immunogenicity of Six Rare Serotype
448 Recombinant Adenovirus Vaccine Vectors from Subgroups B and D. *J. Virol.* **81,** 4654–4663
449 (2007).
450 6. Family - Adenoviridae. in *Virus Taxonomy* (eds. King, A. M. Q., Adams, M. J., Carstens, E. B. &
451 Lefkowitz, E. J.) 125–141 (Elsevier, 2012). doi:10.1016/B978-0-12-384684-6.00009-4
452 7. International Committee on Taxonomy of Viruses. *Virus taxonomy: classification and*
453 *nomenclature of viruses: sixth report of the International Committee on Taxonomy of Viruses*.
454 (Springer-Verlag, 1995).
455 8. Mast, T. C. *et al.* International epidemiology of human pre-existing adenovirus (Ad) type-5, type-
456 6, type-26 and type-36 neutralizing antibodies: correlates of high Ad5 titers and implications for
457 potential HIV vaccine trials. *Vaccine* **28,** 950–957 (2010).
458 9. Uusi-Kerttula, H., Hulin-Curtis, S., Davies, J. & Parker, A. L. Oncolytic Adenovirus: Strategies and
459 Insights for Vector Design and Immuno-Oncolytic Applications. *Viruses* **7,** 6009–6042 (2015).
460 10. Priddy, F. H. *et al.* Safety and Immunogenicity of a Replication-Incompetent Adenovirus Type 5
461 HIV-1 Clade B *gag/pol/nef* Vaccine in Healthy Adults. *Clin. Infect. Dis.* **46,** 1769–1781 (2008).
462 11. Baden, L. R. *et al.* First-in-Human Evaluation of a Hexon Chimeric Adenovirus Vector Expressing
463 HIV-1 Env (IPCAVD 002). *J. Infect. Dis.* **210,** 1052–1061 (2014).
464 12. Barouch, D. H. *et al.* Evaluation of a mosaic HIV-1 vaccine in a multicentre, randomised, double-
465 blind, placebo-controlled, phase 1/2a clinical trial (APPROACH) and in rhesus monkeys (NHP 13-
466 19). *The Lancet* **392,** 232–243 (2018).
467 13. Robinson, C. M. *et al.* Molecular evolution of human adenoviruses. *Sci. Rep.* **3,** 1812 (2013).
468 14. Rux, J. J., Kuser, P. R. & Burnett, R. M. Structural and Phylogenetic Analysis of Adenovirus Hexons
469 by Use of High-Resolution X-Ray Crystallographic, Molecular Modeling, and Sequence-Based
470 Methods. *J. Virol.* **77,** 9553–9566 (2003).
471 15. Singh, G. *et al.* Overreliance on the Hexon Gene, Leading to Misclassification of Human
472 Adenoviruses. *J. Virol.* **86,** 4693–4695 (2012).
473 16. Teigler, J. E., Iampietro, M. J. & Barouch, D. H. Vaccination with Adenovirus Serotypes 35, 26,
474 and 48 Elicits Higher Levels of Innate Cytokine Responses than Adenovirus Serotype 5 in Rhesus
475 Monkeys. *J. Virol.* **86,** 9590–9598 (2012).
476 17. Li, H. *et al.* Adenovirus Serotype 26 Utilizes CD46 as a Primary Cellular Receptor and Only
477 Transiently Activates T Lymphocytes following Vaccination of Rhesus Monkeys. *J. Virol.* **86,**
478 10862–10865 (2012).
479 18. Uusi-Kerttula, H. *et al.* Pseudotyped αvβ6 integrin-targeted adenovirus vectors for ovarian
480 cancer therapies. *Oncotarget* **7,** 27926–27937 (2016).

481 19. Camacho, Z. T., Turner, M. A., Barry, M. A. & Weaver, E. A. CD46-Mediated Transduction of a
482 Species D Adenovirus Vaccine Improves Mucosal Vaccine Efficacy. *Hum. Gene Ther.* **25,** 364–374
483 (2014).
484 20. Weaver, E. A. & Barry, M. A. Low Seroprevalent Species D Adenovirus Vectors as Influenza
485 Vaccines. *PLoS ONE* **8,** (2013).
486 21. Geisbert, T. W. *et al.* Recombinant Adenovirus Serotype 26 (Ad26) and Ad35 Vaccine Vectors
487 Bypass Immunity to Ad5 and Protect Nonhuman Primates against Ebolavirus Challenge. *J. Virol.*
488 **85,** 4222–4233 (2011).
489 22. Wang, Y. *et al.* Phylogenetic evidence for intratypic recombinant events in a novel human
490 adenovirus C that causes severe acute respiratory infection in children. *Sci. Rep.* **6,** 23014 (2016).
491 23. Ebner, K., Pinsker, W. & Lion, T. Comparative Sequence Analysis of the Hexon Gene in the Entire
492 Spectrum of Human Adenovirus Serotypes: Phylogenetic, Taxonomic, and Clinical Implications. *J.*
493 *Virol.* **79,** 12635–12642 (2005).
494 24. Li, P. *et al.* Circulation of HAdV-41 with diverse genome types and recombination in acute
495 gastroenteritis among children in Shanghai. *Sci. Rep.* **7,** 3548 (2017).
496 25. Al Qurashi, Y. M. A., Alkhalaf, M. A., Lim, L., Guiver, M. & Cooper, R. J. Sequencing and
497 phylogenetic analysis of the hexon, fiber, and penton regions of adenoviruses isolated from AIDS
498 patients. *J. Med. Virol.* **84,** 1157–1165 (2012).
499 26. Pache, L., Venkataraman, S., Nemerow, G. R. & Reddy, V. S. Conservation of fiber structure and
500 CD46 usage by subgroup B2 adenoviruses. *Virology* **375,** 573–579 (2008).
501 27. Russell, W. C. Adenoviruses: update on structure and function. *J. Gen. Virol.* **90,** 1–20 (2009).
502 28. Xia, D., Henry, L. J., Gerard, R. D. & Deisenhofer, J. Crystal structure of the receptor-binding
503 domain of adenovirus type 5 fiberprotein at 1.7 Å resolution. *Structure* **2,** 1259–1270 (1994).
504 29. Baker, A. T., Aguirre-Hernández, C., Halldén, G. & Parker, A. L. Designer Oncolytic Adenovirus:
505 Coming of Age. *Cancers* **10,** 201 (2018).
506 30. Amoureux, M.-C. *et al.* Polysialic Acid Neural Cell Adhesion Molecule (PSA-NCAM) is an adverse
507 prognosis factor in glioblastoma, and regulates olig2 expression in glioma cell lines. *BMC Cancer*
508 **10,** 91 (2010).
509 31. Lenman, A. *et al.* Human Adenovirus 52 Uses Sialic Acid-containing Glycoproteins and the
510 Coxsackie and Adenovirus Receptor for Binding to Target Cells. *PLOS Pathog.* **11,** e1004657
511 (2015).
512 32. Arnberg, N. *et al.* Adenovirus Type 37 Binds to Cell Surface Sialic Acid Through a Charge-
513 Dependent Interaction. *Virology* **302,** 33–43 (2002).
514 33. Arnberg, N., Edlund, K., Kidd, A. H. & Wadell, G. Adenovirus Type 37 Uses Sialic Acid as a Cellular
515 Receptor. *J. Virol.* **74,** 42–48 (2000).
516 34. Cupelli, K. *et al.* Structure of Adenovirus Type 21 Knob in Complex with CD46 Reveals Key
517 Differences in Receptor Contacts among Species B Adenoviruses. *J. Virol.* **84,** 3189–3200 (2010).
518 35. Persson, B. D. *et al.* Structure of the Extracellular Portion of CD46 Provides Insights into Its
519 Interactions with Complement Proteins and Pathogens. *PLOS Pathog.* **6,** e1001122 (2010).
520 36. Persson, B. D. *et al.* Adenovirus type 11 binding alters the conformation of its receptor CD46.
521 *Nat. Struct. Mol. Biol.* **14,** 164–166 (2007).
522 37. Wang, H. *et al.* Identification of CD46 binding sites within the adenovirus serotype 35 fiber knob.
523 *J. Virol.* **81,** 12785–12792 (2007).
524 38. Kirby, I. *et al.* Identification of Contact Residues and Definition of the CAR-Binding Site of
525 Adenovirus Type 5 Fiber Protein. *J. Virol.* **74,** 2804–2813 (2000).
526 39. Nilsson, E. C. *et al.* The GD1a glycan is a cellular receptor for adenoviruses causing epidemic
527 keratoconjunctivitis. *Nat. Med.* **17,** 105–109 (2011).
528 40. Fleischli, C. *et al.* Species B adenovirus serotypes 3, 7, 11 and 35 share similar binding sites on
529 the membrane cofactor protein CD46 receptor. *J. Gen. Virol.* **88,** 2925–2934 (2007).
530 41. Ismail, A. M. *et al.* Selection Pressure in the Human Adenovirus Fiber Knob Drives Cell Specificity
531 in Epidemic Keratoconjunctivitis. *J. Virol.* **90,** 9598–9607 (2016).

532 42. Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E. J. & Cusack, S. Structural and Mutational
533 Analysis of Human Ad37 and Canine Adenovirus 2 Fiber Heads in Complex with the D1 Domain
534 of Coxsackie and Adenovirus Receptor. *J. Biol. Chem.* **281,** 33704–33716 (2006).
535 43. Krieger Elmar, Koraimann Günther & Vriend Gert. Increasing the precision of comparative
536 models with YASARA NOVA—a self-parameterizing force field. *Proteins Struct. Funct.*
537 *Bioinforma.* **47,** 393–402 (2002).
538 44. Jensen, J. H. Predicting accurate absolute binding energies in aqueous solution: thermodynamic
539 considerations for electronic structure methods. *Phys. Chem. Chem. Phys.* **17,** 12441–12451
540 (2015).
541 45. Vassal-Stermann, E. *et al.* Mapping of Adenovirus of serotype 3 fibre interaction to desmoglein 2
542 revealed a novel 'non-classical' mechanism of viral receptor engagement. *Sci. Rep.* **8,** (2018).
543 46. Wang, H. *et al.* Desmoglein 2 is a receptor for adenovirus serotypes 3, 7, 11 and 14. *Nat. Med.*
544 **17,** 96–104 (2011).
545 47. Segerman, A., Arnberg, N., Erikson, A., Lindman, K. & Wadell, G. There are two different species
546 B adenovirus receptors: sBAR, common to species B1 and B2 adenoviruses, and sB2AR,
547 exclusively used by species B2 adenoviruses. *J. Virol.* **77,** 1157–1162 (2003).
548 48. Kuhn, I. *et al.* Directed Evolution Generates a Novel Oncolytic Virus for the Treatment of Colon
549 Cancer. *PLOS ONE* **3,** e2409 (2008).
550 49. Illingworth, S. *et al.* Preclinical Safety Studies of Enadenotucirev, a Chimeric Group B Human-
551 Specific Oncolytic Adenovirus. *Mol. Ther. Oncolytics* **5,** 62–74 (2017).
552 50. HAdV Working Group, http://hadvwg.gmu.edu/.
553 51. Bradley, R. R., Lynch, D. M., Iampietro, M. J., Borducchi, E. N. & Barouch, D. H. Adenovirus
554 serotype 5 neutralizing antibodies target both hexon and fiber following vaccination and natural
555 infection. *J. Virol.* **86,** 625–629 (2012).
556 52. Bradley, R. R. *et al.* Adenovirus Serotype 5-Specific Neutralizing Antibodies Target Multiple
557 Hexon Hypervariable Regions. *J. Virol.* **86,** 1267–1272 (2012).
558 53. Ma, J. *et al.* Manipulating Adenovirus Hexon Hypervariable Loops Dictates Immune
559 Neutralisation and Coagulation Factor X-dependent Cell Interaction In Vitro and In Vivo. *PLoS*
560 *Pathog.* **11,** (2015).
561 54. Parker, A. L. *et al.* Effect of neutralizing sera on factor x-mediated adenovirus serotype 5 gene
562 transfer. *J. Virol.* **83,** 479–483 (2009).
563 55. Walters, R. W. *et al.* Adenovirus Fiber Disrupts CAR-Mediated Intercellular Adhesion Allowing
564 Virus Escape. *Cell* **110,** 789–799 (2002).
565 56. Keele, B. F. *et al.* Adenovirus prime, Env protein boost vaccine protects against neutralization-
566 resistant SIVsmE660 variants in rhesus monkeys. *Nat. Commun.* **8,** 15740 (2017).
567 57. Andrews, P. W. *et al.* A human cell-surface antigen defined by a monoclonal antibody and
568 controlled by a gene on human chromosome 1. *Ann. Hum. Genet.* **49,** 31–39 (1985).
569 58. Uhlen, M. *et al.* A pathology atlas of the human cancer transcriptome. *Science* **357,** eaan2507
570 (2017).
571 59. Seiradake, E. *et al.* The Cell Adhesion Molecule "CAR" and Sialic Acid on Human Erythrocytes
572 Influence Adenovirus In Vivo Biodistribution. *PLOS Pathog.* **5,** e1000277 (2009).
573 60. Cohen, C. J. *et al.* The coxsackievirus and adenovirus receptor is a transmembrane component of
574 the tight junction. *Proc. Natl. Acad. Sci.* **98,** 15191–15196 (2001).
575 61. Awasthi, V., Meinken, G., Springer, K., Srivastava, S. C. & Freimuth, P. Biodistribution of
576 Radioiodinated Adenovirus Fiber Protein Knob Domain after Intravenous Injection in Mice. *J.*
577 *Virol.* **78,** 6431–6438 (2004).
578 62. Uusi-Kerttula, H. *et al.* Ad5NULL-A20 - a tropism-modified, αvβ6 integrin-selective oncolytic
579 adenovirus for epithelial ovarian cancer therapies. *Clin. Cancer Res.* clincanres.1089.2018 (2018).
580 doi:10.1158/1078-0432.CCR-18-1089
581 63. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments
582 using Clustal Omega. *Mol. Syst. Biol.* **7,** 539–539 (2011).

583 64. Li, W. *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids*
584 *Res.* **43,** W580-4 (2015).
585 65. McWilliam, H. *et al.* Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **41,**
586 W597–W600 (2013).
587 66. Jukes, T. H. & Cantor, C. R. Evolution of Protein Molecules. in *Mammalian Protein Metabolism*
588 21–132 (Elsevier, 1969). doi:10.1016/B978-1-4832-3211-9.50009-7
589 67. Zuckerkandl, E. & Pauling, L. Evolutionary Divergence and Convergence in Proteins. in *Evolving*
590 *Genes and Proteins* (eds. Bryson, V. & Vogel, H. J.) 97–166 (Academic Press, 1965).
591 doi:10.1016/B978-1-4832-2734-4.50017-6
592 68. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics
593 Analysis across Computing Platforms. *Mol. Biol. Evol.* **35,** 1547–1549 (2018).
594 69. Felsenstein, J. CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP.
595 *Evol. Int. J. Org. Evol.* **39,** 783–791 (1985).
596 70. Kabat, E. A., Wu, T. T. & Bilofsky, H. Unusual distributions of amino acids in complementarity-
597 determining (hypervariable) segments of heavy and light chains of immunoglobulins and their
598 possible roles in specificity of antibody-combining sites. *J. Biol. Chem.* **252,** 6609–6616 (1977).
599 71. Garcia-Boronat, M., Diez-Rivero, C. M., Reinherz, E. L. & Reche, P. A. PVS: a web server for
600 protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic*
601 *Acids Res.* **36,** W35–W41 (2008).
602 72. Coughlan, L. *et al.* Retargeting Adenovirus Serotype 48 Fiber Knob Domain by Peptide
603 Incorporation. *Hum. Gene Ther.* **25,** 385–394 (2014).
604 73. Coughlan, L. *et al.* In Vivo Retargeting of Adenovirus Type 5 to αvβ6 Integrin Results in Reduced
605 Hepatotoxicity and Improved Tumor Uptake following Systemic Delivery. *J. Virol.* **83,** 6416–6428
606 (2009).
607 74. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. D Biol.*
608 *Crystallogr.* **69,** 1260–1273 (2013).
609 75. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D*
610 *Biol. Crystallogr.* **67,** 293–302 (2011).
611 76. Dodson, E. J., Winn, M. & Ralph, A. [32] Collaborative computational project, number 4:
612 Providing programs for protein crystallography. in *Methods in Enzymology* **277,** 620–633
613 (Academic Press, 1997).
614 77. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
615 78. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D*
616 *Biol. Crystallogr.* **60,** 2126–2132 (2004).
617 79. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta*
618 *Crystallogr. D Biol. Crystallogr.* **67,** 355–367 (2011).
619 80. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 2.0.
620 81. Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in
621 homology modeling: Four approaches that performed well in CASP8. *Proteins Struct. Funct.*
622 *Bioinforma.* **77,** 114–122 (2009).
623 82. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol.*
624 *Biol.* **372,** 774–797 (2007).
625 83. Hall, T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for
626 Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41,** 95–98 (1999).
627

628

629 **Figure legends**

630 **Figure 1: Phylogenetic analysis of adenoviruses mapped by whole genome and fiber-knob domain.**
631 Condensed maximum likelihood trees (percentage confidence shown by numbers next to nodes)
632 were generated from alignments of fiber-knob domain amino acid sequences of adenoviruses 1-56
633 (A) or whole genome NT sequences (B). Adenoviruses divide into 7 subspecies, as denoted in the
634 key, regardless of alignment used, but the species D adenoviruses divide into additional sub-species
635 when determined by fiber-knob alignment. Numbers next to nodes denote confidence. Wu-Kabat
636 variability analysis of the Clustal omega aligned fiber-knob domains amino acid sequences of
637 adenoviruses 1-56 (C) reveals regions of low amino acid variability corresponding to beta-sheets. The
638 locations of Ad5 β-strands, as described by Xia et al (1995), are aligned to the structure and shown
639 by arrows, the corresponding positions are coloured in red.

640 **Figure 2: Overview of the Ad26 and Ad48 fiber-knob protein structures (PDB 6FJN and 6FJQ,**
641 **respectively)**. The surface representation of the trimeric Ad26K biological assembly is shown side-on
642 with the cartoon representation shown for the nearest monomer (A) and the top-down view of the
643 same Ad26K trimer, as it would appear looking towards the virion, is seen as a cartoon
644 representation (B), with each monomer coloured in Red, Green, or Blue with the hypervariable loops
645 extending between the β-strands (Dark blue) coloured as follows: AB-Green, BC – Purple, CD –
646 Brown, DE – Orange, DG – Red, EG – Pink, GH – Purple, HI – Yellow, IJ – Light Blue. The Ad48K trimer
647 is shown similarly (C,D). The Wu-Kabat variability plot of the fiber-knob domains of species D
648 adenoviruses shows regions of low variability (E) with the locations of the Ad48K β-strands shown by
649 arrows above the graph, and the positions coloured blue, or red when the position is a β-strand in
650 both Ad26K and Ad48K.

651 **Figure 3: Comparison of HI, DG, GH, and IJ loops of adenoviruses used in this study**. The
652 hypervariable loops of Ad26K (Green) and Ad48K (Cyan) relevant to this study (HI, DG, GH, and IJ)
653 are shown in the context of the control virus fiber-knob domains, Ad5K (Orange), Ad11K (Yellow),
654 Ad35K (Pink), and Ad37K (Purple). The electron density achieved in the loops of Ad26K (B) and
655 Ad48K (C) are shown as mesh. The fitted residues are seen as stick representations, with oxygen and
656 nitrogen atoms coloured red and blue, respectively and other atoms coloured according to their
657 relative B-factors with warmer colours indicating higher B-factor values.

658 **Figure 4: Hypervariable loop conformations and contacts residues**. The residues comprising the
659 indicated loops of Ad26K (A) and Ad48K (B) are shown diagrammatically with numbers indicating the
660 start and end residues of each loop depicted. The network of intraloop polar interactions is shown
661 by solid lines (one polar bond), and dashed lines (two polar bonds, colour variations are only for ease

662 of viewing) between interacting residues, similar interloop bonds are also present as visualised in
663 Suppl. 2. Residues forming part of a helical motif are shaded in blue. The Ad48K DG-loop is seen to
664 form contacts to the opposing monomer across the inter-monomeric cleft (C). The labelled residues
665 forming polar contacts (shown as sticks) are coloured by relative B-factor, with warmer colours
666 indicating higher relative B-factors, as is the cartoon representation of the loop. The opposing Ad48K
667 monomer is seen as a ribbon representation of the carbon-α chain in cyan and the surface of the
668 Ad48K trimer seen as a semi-transparent grey surface.

669 **Figure 5: Modelling of the Ad26K and Ad48K interaction with CAR at the α-interface.** The α -
670 interface region is shown by the box on the structural alignment of Ad5K (Orange), 26K (Green), and
671 48K (Cyan) fiber-knob domain crystal structures in complex with CAR-D1 domain (Grey) as
672 determined by homology alignment to the previously reported Ad37K CAR-D1 structure (PDB: 2J12)
673 (A). The aligned amino acid sequence of the investigated fiber-knobs (B) and the predicted α-
674 interface forming CAR-D1 binding residues are highlighted in blue, with the underlined residues
675 representing the Ad5K amino acids shown by Kirby *et al* (2000) to be important for CAR interaction.
676 Conservation of key residues can be seen between Ad5K, Ad26K, and Ad48K fiber-knobs. This
677 conservation is visualised, with the contact residues comprising the α-interface with Ad5K, 26K, and
678 48K shown as sticks in complex with the energy minimised CAR-D1 domain (Grey), shown as the
679 surface of the maximum spatial occupancy of the aligned CAR-D1 monomers from each of the
680 energy minimised models in complex with Ad5K, 26K, and 48K fiber-knobs (C). (D) Plots the binding
681 energy of the energy minimised fiber-knob proteins to CAR-D1 complex in the α-interface. Lower
682 binding energy indicates a more stable interface with the red line depicting 3.0 $KcalMol^{-1}$, which can
683 be considered background.

684 **Figure 6: Modelling of the Ad26K and Ad48K interaction with CAR at the β-interface.** The β-
685 interface region is shown by the box on the structural alignment of Ad5K (Orange), 26K (Green), and
686 48K (Cyan) fiber-knob domain crystal structures in complex with CAR-D1 domain (Grey) as
687 determined by homology alignment to the previously reported Ad37K CAR-D1 structure (PDB: 2J12)
688 (A). A dot surface shows the surface of Ad5K, 26K, and 48K DG-loops in the inter-monomer cleft (B).
689 The boxes denote the maximum B-factor of the corresponding loops, which are shown as putty
690 representations with thicker regions indicating higher relative B-factors, from which we can infer the
691 relative stability of the loops. Antibody competition inhibition assay (C) shows the relative inhibitory
692 ability of the Ad5, 35, 26, and 48, fiber-knob domains in CAR expressing CHO-CAR cells, with the
693 calculated IC_{50} values shown in boxes. IC_{50} curves are fitted by non-linear regression. Error bars
694 represent standard deviation of 3 biological replicates.

695 **Figure 7: Modelling of the Ad26K and Ad48K with CD46.** Red dashes show contacts between the
696 energy minimised crystal structure of CD46 SC1 and SC2 domains (grey cartoon) and Ad11K in
697 complex (PDB 3O8E). The known CD46 interacting fiber-knob, Ad35K (purple), is aligned to the
698 above crystal structure and energy minimised (B). Amino acid sequence alignment of the tested
699 fiber-knob proteins (C) shows conservation of residues previously shown by Wang *et al* (2007) to be
700 key to CD46 binding (underlined) between the known CD46 binding fiber-knobs, Ad11K and Ad35K.
701 Residues highlighted in blue are predicted to form direct contacts with CD46 in the energy
702 minimised models. Similar alignments to that performed with Ad35K are shown for Ad26K (Green -
703 D) and Ad48K (Cyan - E). In all models red dashes indicate polar contacts between the residues
704 shown as stick representations.

705 **Figure 8: Binding energetics and affinities of Ad26K and Ad48K with CD46**. Calculation of the
706 predicted binding energies for the energy minimised fiber-knob: CD46 models are compared on the
707 bar chart (A), lower $Kcalmol^{-1}$ values indicate a stronger interaction, the red line at 3.0 $KcalMol^{-1}$,
708 denotes an interface energy which can be considered negligible (random proteins passing in
709 solution). The HI loop (red) of the Ad5 fiber-knob (orange) is seen to extend between CD46 (grey)
710 and the DG loop (B). The antibody competition inhibition assay (C) shows the relative inhibitory
711 ability of the Ad5, 35, 26, and 48, fiber-knob domains in CD46 expressing CHO-BC1 cells, with the
712 calculated IC_{50} values shown in boxes. IC_{50} curves are fitted by non-linear regression. Error bars
713 represent standard deviation of 3 biological replicates.

Figure 1

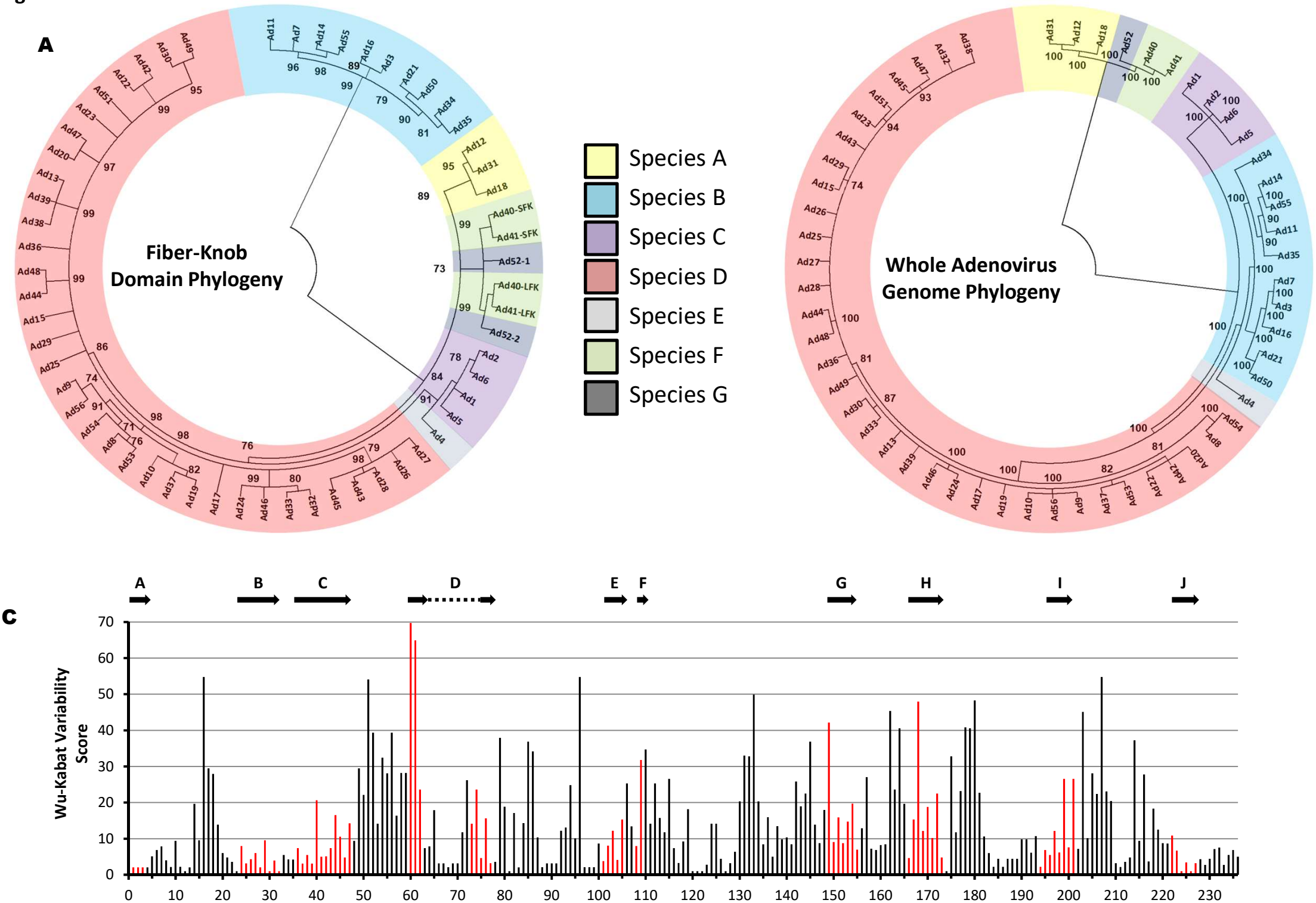

A
Fiber-Knob Domain Phylogeny
Whole Adenovirus Genome Phylogeny
Species A
Species B
Species C
Species D
Species E
Species F
Species G
C
Wu-Kabat Variability Score
A
B
C
D
E
F
G
H
I
J

Figure 2

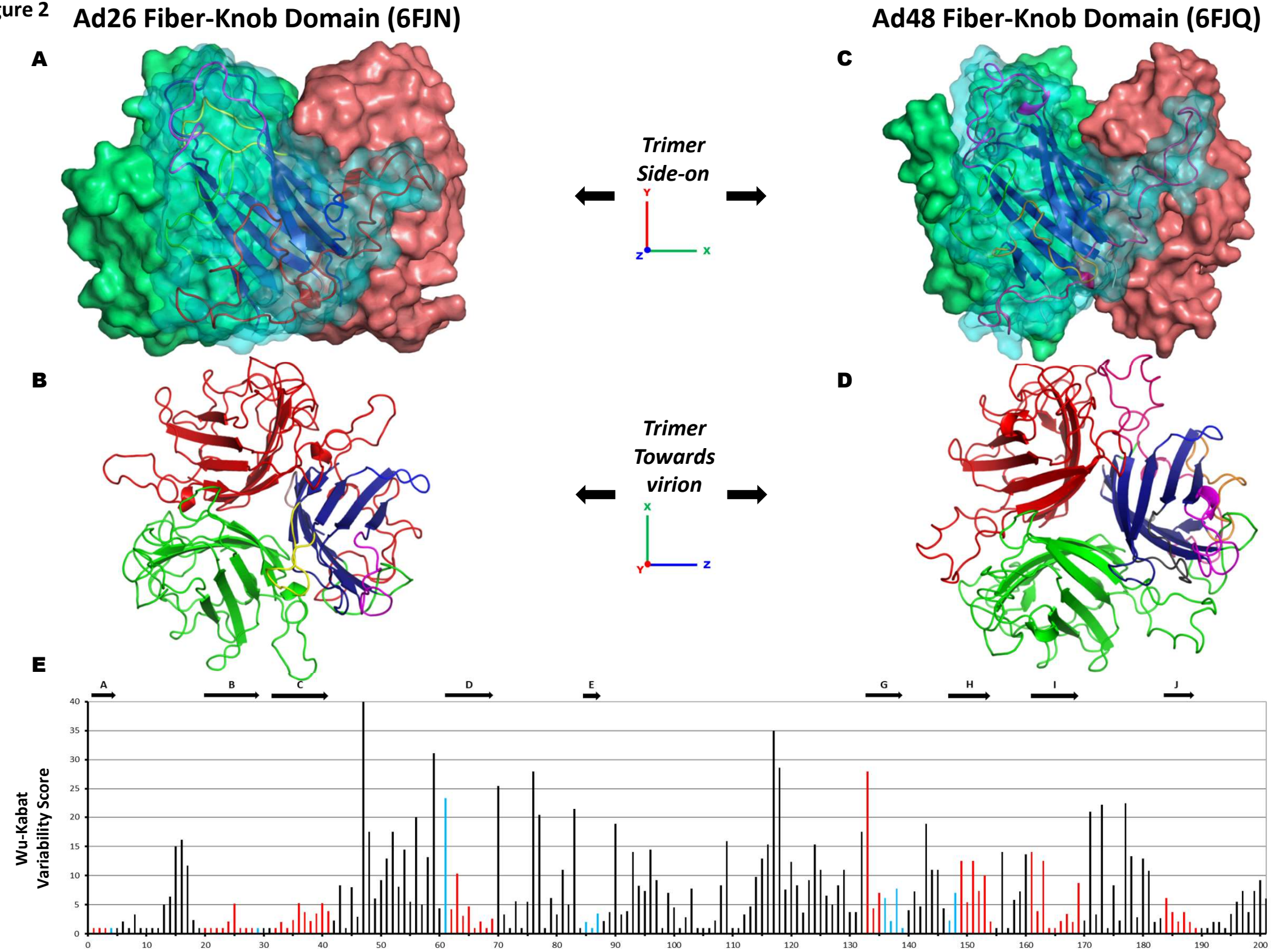

Ad26 Fiber-Knob Domain (6FJN)
Ad48 Fiber-Knob Domain (6FJQ)
A
B
C
D
E
Trimer
Side-on
Trimer
Towards
virion
Y
X
Z
Wu-Kabat
Variability Score
A
B
C
D
E
G
H
I
J

Figure 3

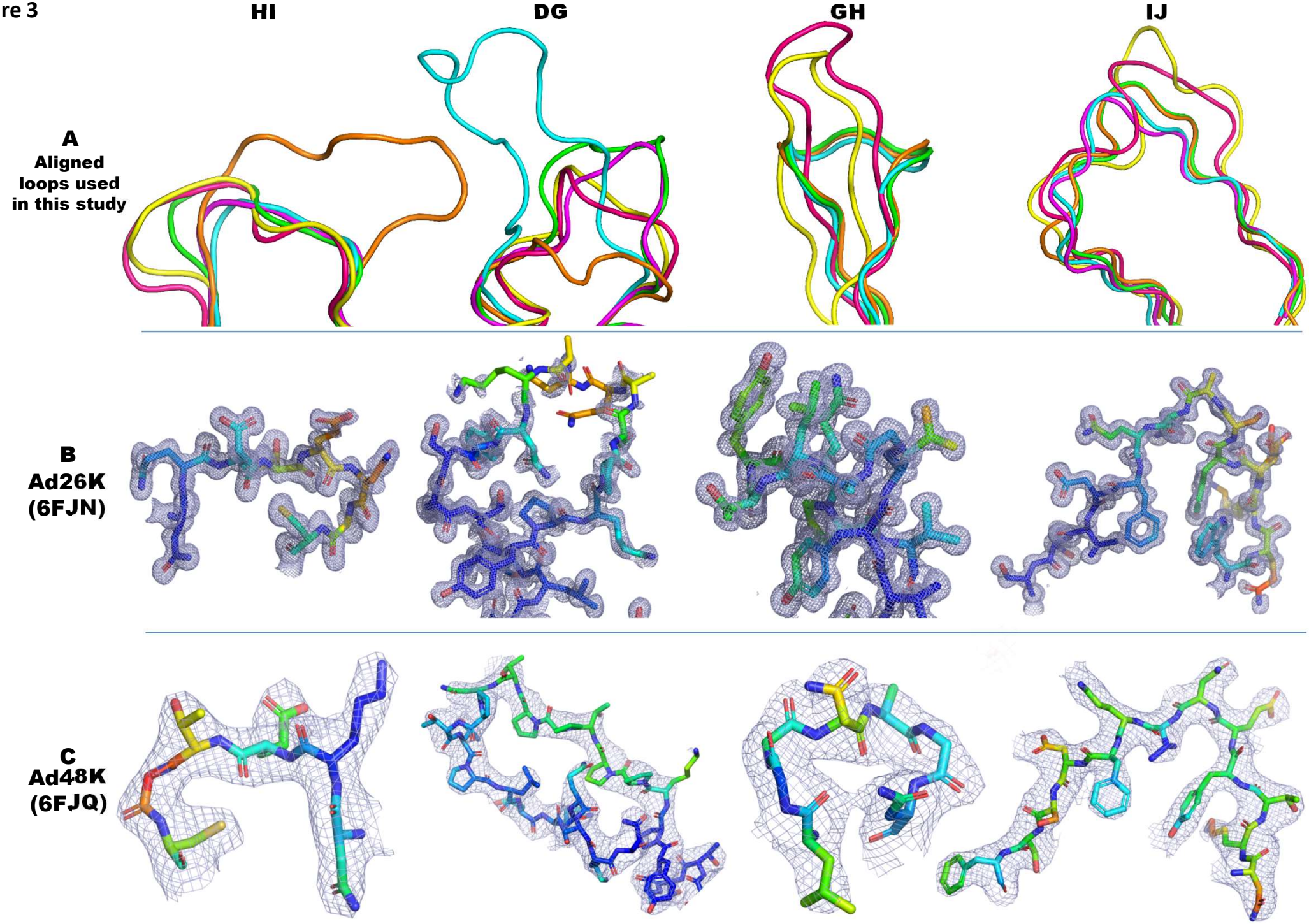

HI
DG
GH
IJ
A
Aligned loops used in this study
B
Ad26K
(6FJN)
C
Ad48K
(6FJQ)

Figure 4

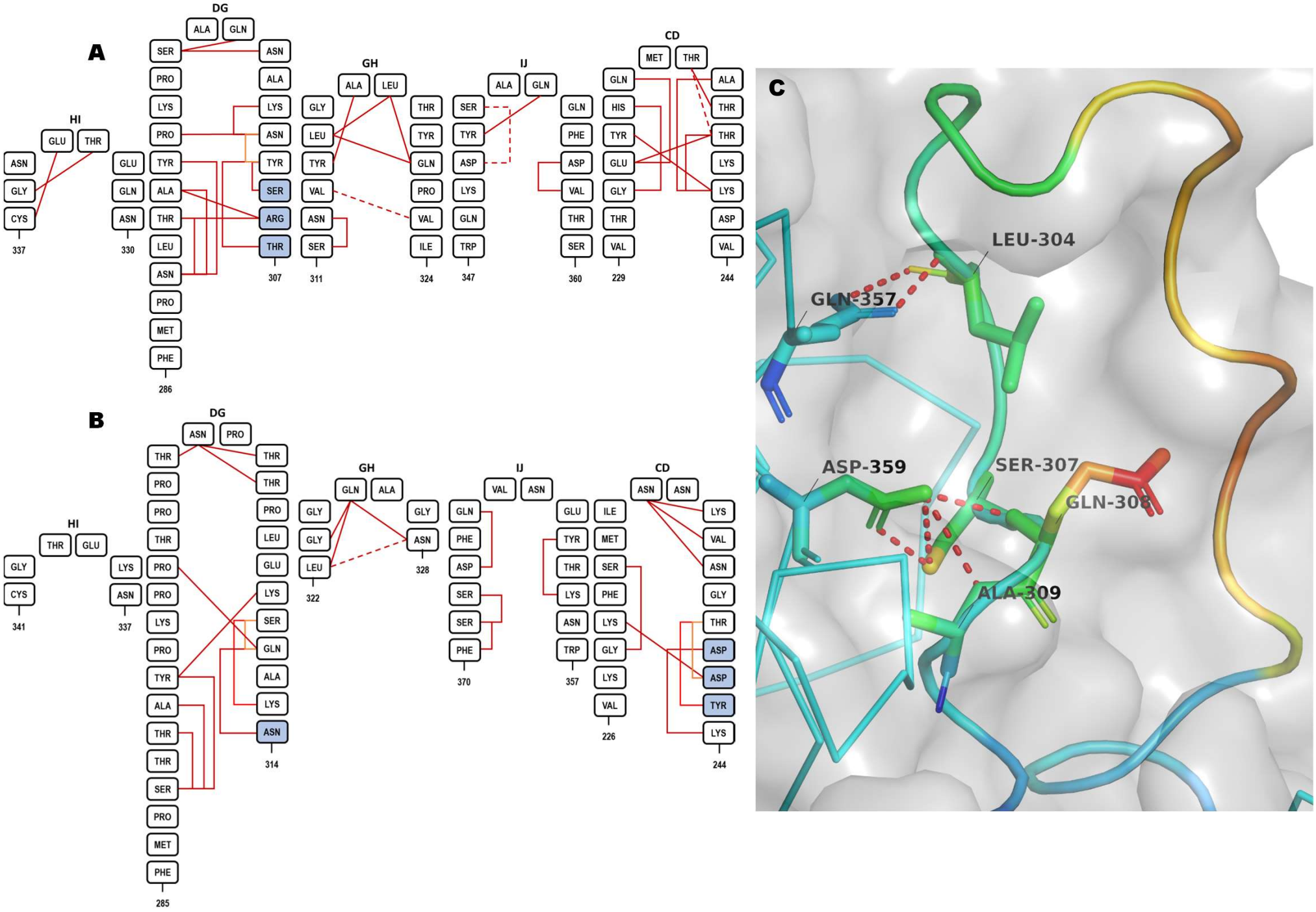

A
DG
ALA GLN
SER PRO LYS PRO TYR ALA THR LEU ASN PRO MET PHE
286
ASN ALA LYS ASN TYR SER ARG THR
307
HI
GLU THR
ASN GLY CYS
337
GLU GLN ASN
330
GH
ALA LEU
GLY LEU TYR VAL ASN SER
311
THR TYR GLN PRO VAL ILE
324
IJ
ALA GLN
SER TYR ASP LYS GLN TRP
347
GLN PHE ASP VAL THR SER
360
CD
MET THR
GLN HIS TYR GLU GLY THR VAL
229
ALA THR THR LYS LYS ASP VAL
244
B
DG
ASN PRO
THR PRO PRO THR PRO TYR ALA THR THR SER PRO MET PHE
285
THR THR PRO LEU GLU LYS SER GLN ALA LYS ASN
314
HI
THR GLU
GLY CYS
341
LYS ASN
337
GH
GLN ALA
GLY GLY LEU
322
GLY ASN
328
IJ
VAL ASN
GLN PHE ASP SER SER PHE
370
CD
ASN ASN
GLU TYR THR LYS ASN TRP
357
ILE MET SER PHE LYS GLY LYS VAL
226
LYS VAL ASN GLY THR ASP ASP TYR LYS
244
C
LEU-304
GLN-357
ASP-359
SER-307
GLN-308
ALA-309

Figure 5

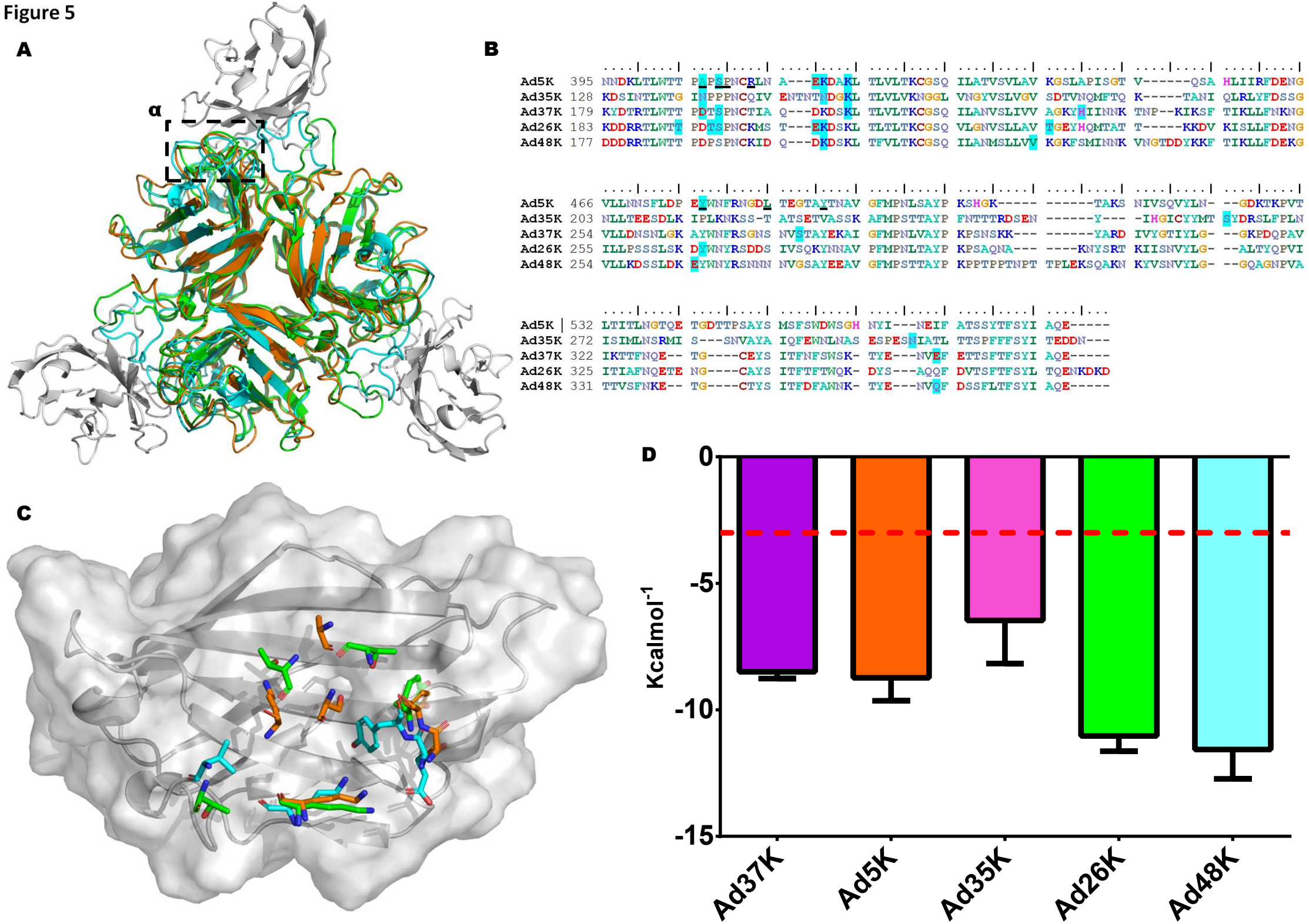

A
α
B
Ad5K 395 NNDKLTLWTT PAPSPNCRLN A---EKDAKL TLVLTKCGSQ ILATVSVLAV KGSLAPISGT V------QSA HLIIRFDENG
Ad35K 128 KDSINTLWTG INPPPNCQIV ENTNTNDGKL TLVLVKNGGL VNGYVSLVGV SDTVNQMFTQ K-----TANI QLRLYFDSSG
Ad37K 179 KYDTRTLWTT PDTSPNCTIA Q---DKDSKL TLVLTKCGSQ ILANVSLIVV AGKYHIINNK TNP--KIKSF TIKLLFNKNG
Ad26K 183 KDDRRTLWTT PDTSPNCKMS T---EKDSKL TLTLTKCGSQ VLGNVSLLAV TGEYHQMTAT T-----KKDV KISLLFDENG
Ad48K 177 DDDRRTLWTT PDPSPNCKID Q---DKDSKL TFVLTKCGSQ ILANMSLLVV KGKFSMINNK VNGTDDYKKF TIKLLFDEKG
Ad5K 466 VLLNNSFLDP EYWNFRNGDL TEGTAYTNAV GFMPNLSAYP KSHGK----- ------TAKS NIVSQVYLN- --GDKTKPVT
Ad35K 203 NLLTEESDLK IPLKNKSS-T ATSETVASSK AFMPSTTAYP FNTTTRDSEN ------Y--- -IHGICYYMT SYDRSLFPLN
Ad37K 254 VLLDNSNLGK AYWNFRSGNS NVSTAYEKAI GFMPNLVAYP KPSNSKK--- ------YARD IVYGTIYLG- --GKPDQPAV
Ad26K 255 ILLPSSSLSK DYWNYRSDDS IVSQKYNNAV PFMPNLTAYP KPSAQNA--- ----KNYSRT KIISNVYLG- --ALTYQPVI
Ad48K 254 VLLKDSSLDK EYWNYRSNNN NVGSAYEEAV GFMPSTTAYP KPPTPPTNPT TPLEKSQAKN KYVSNVYLG- --GQAGNPVA
Ad5K 532 LTITLNGTQE TGDTTPSAYS MSFSWDWSGH NYI---NEIF ATSSYTFSYI AQE-----
Ad35K 272 ISIMLNSRMI S---SNVAYA IQFEWNLNAS ESPESNIATL TTSPFFFSYI TEDDN---
Ad37K 322 IKTTFNQE-- TG----CEYS ITFNFSWSK- TYE---NVEF ETTSFTFSYI AQE-----
Ad26K 325 ITIAFNQETE NG----CAYS ITFTFTWQK- DYS---AQQF DVTSFTFSYL TQENKDKD
Ad48K 331 TTVSFNKE-- TG----CTYS ITFDFAWNK- TYE---NVQF DSSFLTFSYI AQE-----
C
D
Kcalmol-1
0
-5
-10
-15
Ad37K
Ad5K
Ad35K
Ad26K
Ad48K

Figure 6

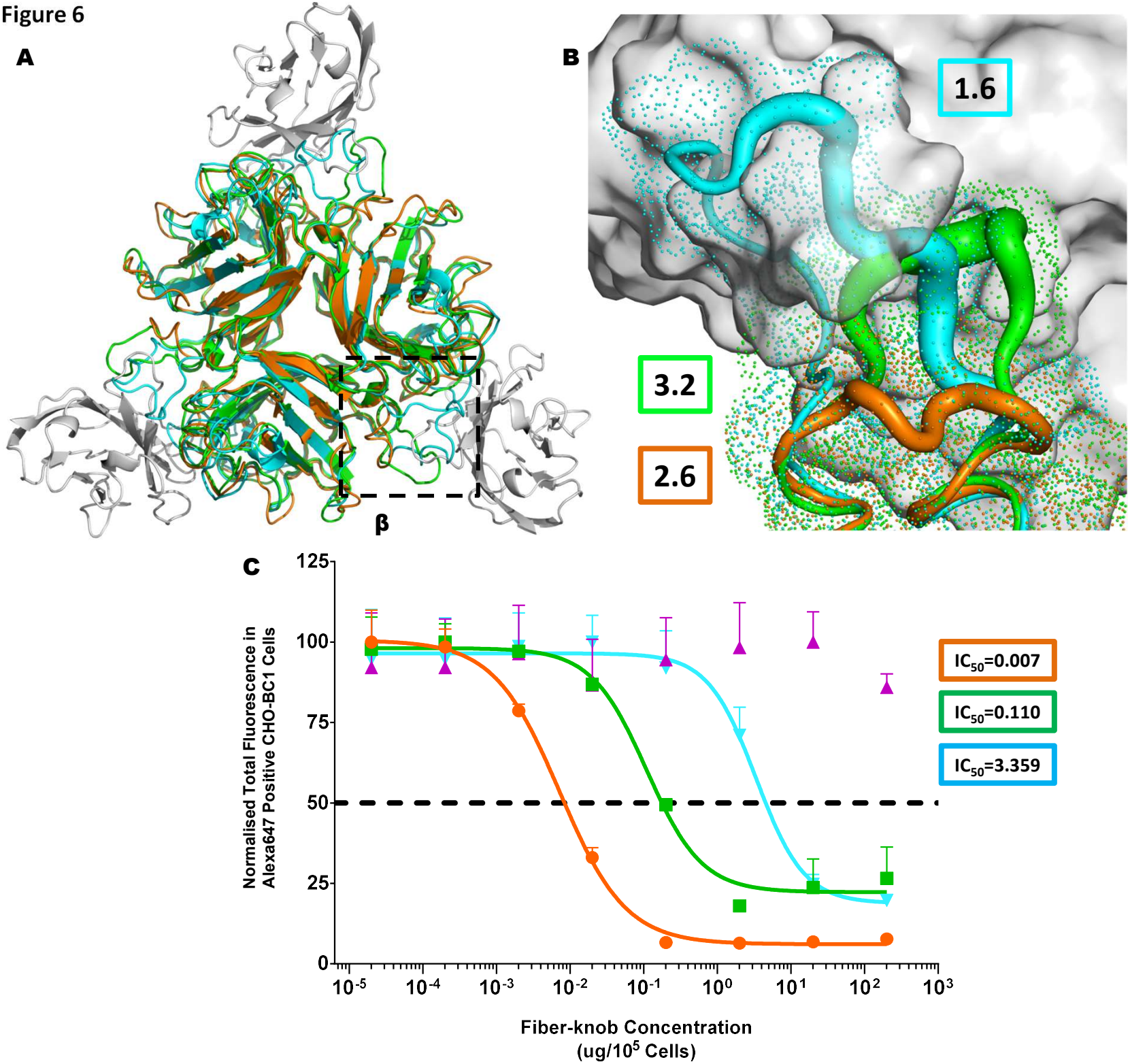

A
β
B
1.6
3.2
2.6
C
Normalised Total Fluorescence in Alexa647 Positive CHO-BC1 Cells
125
100
75
50
25
0
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
Fiber-knob Concentration
(ug/10^5 Cells)
IC_{50}=0.007
IC_{50}=0.110
IC_{50}=3.359

Figure 7

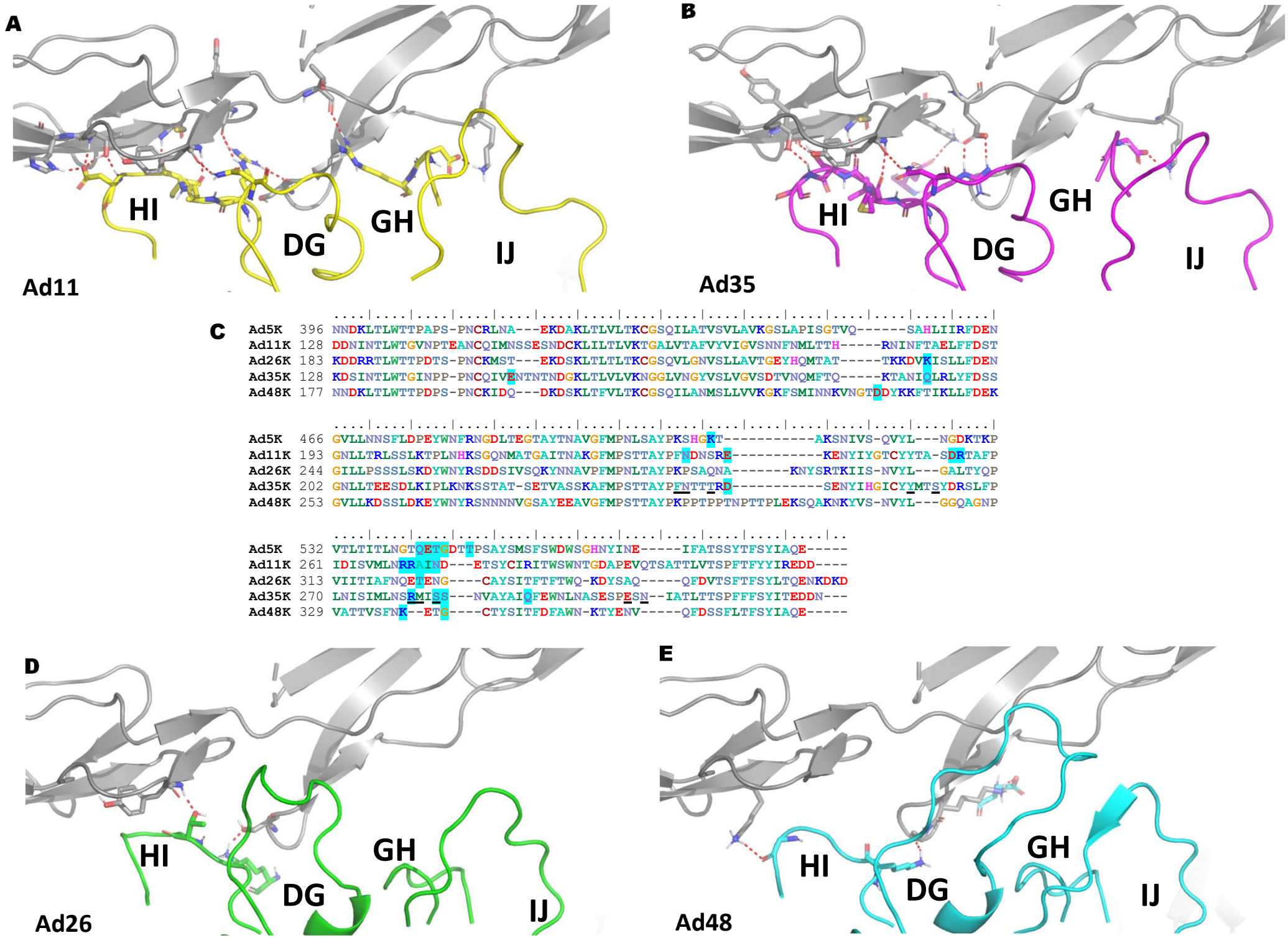
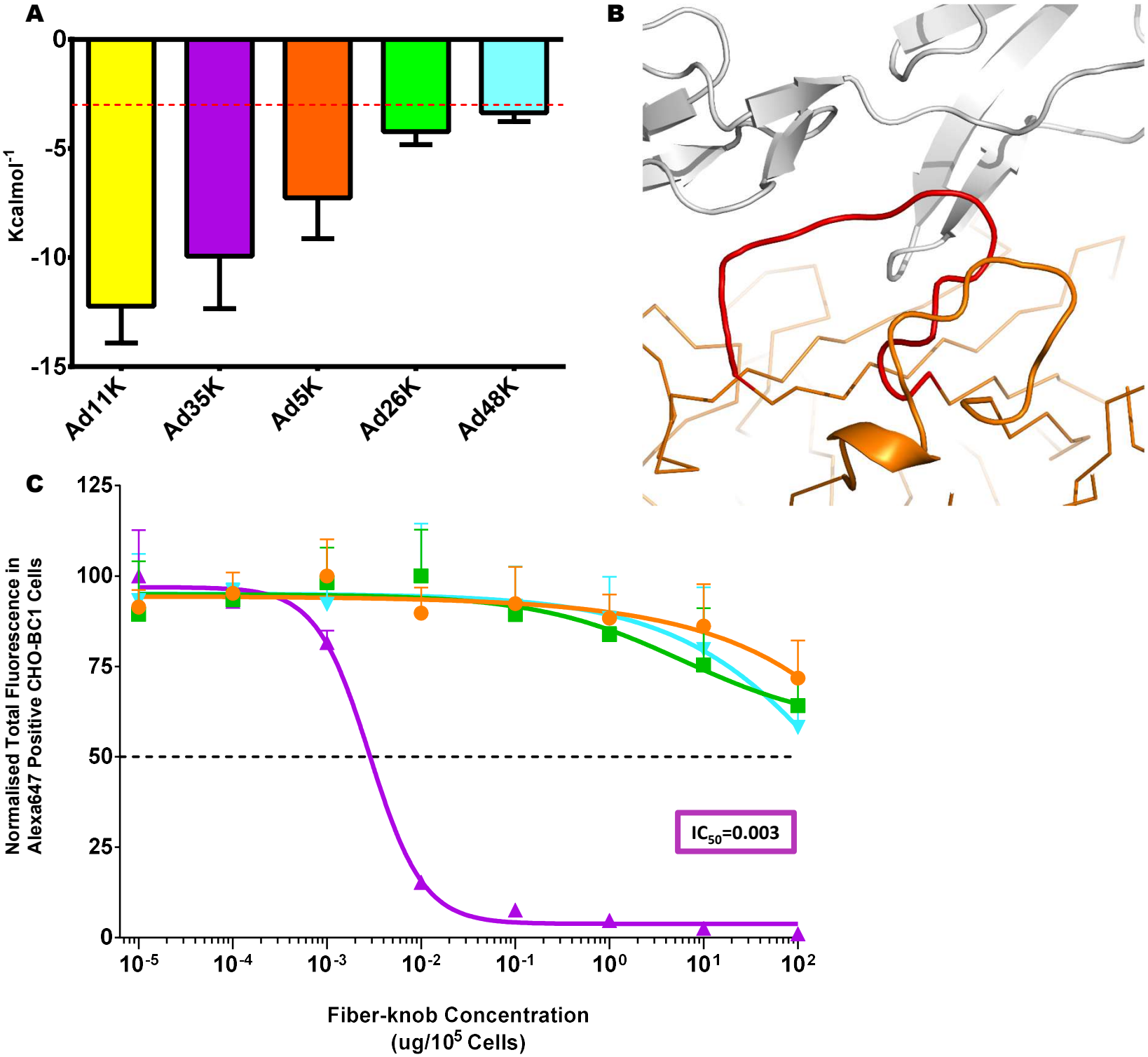

A
HI
DG
GH
IJ
Ad11
B
HI
DG
GH
IJ
Ad35
C
Ad5K 396 NNDKLTLWTTPAPS-PNCRLNA---EKDAKLTLVLTKCGSQILATVSVLAVKGSLAPISGTVQ------SAHLIIRFDEN
Ad11K 128 DDNINTLWTGVNPTEANCQIMNSSESNDCKLILTLVKTGALVTAFVYVIGVSNNFNMLTTH-----RNINFTAELFFDST
Ad26K 183 KDDRRTLWTTPDTS-PNCKMST---EKDSKLTLTLTKCGSQVLGNVSLLAVTGEYHQMTAT-----TKKDVKISLLFDEN
Ad35K 128 KDSINTLWTGINPP-PNCQIVENTNTNDGKLTLVLVKNGGLVNGYVSLVGVSDTVNQMFTQ-----KTANIQLRLYFDSS
Ad48K 177 NNDKLTLWTTPDPS-PNCKIDQ---DKDSKLTFVLTKCGSQILANMSLLVVKGKFSMINNKVNGTDDYKKFTIKLLFDEK
Ad5K 466 GVLLNNSFLDPEYWNFRNGDLTEGTAYTNAVGFMPNLSAYPKSHGKT-----------AKSNIVS-QVYL---NGDKTKP
Ad11K 193 GNLLTRLSSLKTPLNHKSGQNMATGAITNAKGFMPSTTAYPFNDNSRE-----------KENYIYGTCYYTA-SDRTAFP
Ad26K 244 GILLPSSSLSKDYWNYRSDDSIVSQKYNNAVPFMPNLTAYPKPSAQNA-------KNYSRTKIIS-NVYL---GALTYQP
Ad35K 202 GNLLTEESDLKIPLKNKSSTAT-SETVASSKAFMPSTTAYPFNTTTRD-----------SENYIHGICYYMTSYDRSLFP
Ad48K 253 GVLLKDSSLDKEYWNYRSNNNNVGSAYEEAVGFMPSTTAYPKPPTPPTNPTTPLEKSQAKNKYVS-NVYL---GGQAGNP
Ad5K 532 VTLTITLNGTQETGDTTPSAYSMSFSWDWSGHNYINE-----IFATSSYTFSYIAQE-----
Ad11K 261 IDISVMLNRRAIND---ETSYCIRITWSWNTGDAPEVQTSATTLVTSPFTFYYIREDD----
Ad26K 313 VIITIAFNQETENG----CAYSITFTFTWQ-KDYSAQ-----QFDVTSFTFSYLTQENKDKD
Ad35K 270 LNISIMLNSRMISS---NVAYAIQFEWNLNASESPESN--IATLTTSPFFFSYITEDDN---
Ad48K 329 VATTVSFNK--ETG----CTYSITFDFAWN-KTYENV-----QFDSSFLTFSYIAQE-----
D
HI
DG
GH
IJ
Ad26
E
HI
DG
GH
IJ
Ad48

Figure 8


A
0
-5
-10
-15
Kcalmol-1
Ad11K
Ad35K
Ad5K
Ad26K
Ad48K
B
C
125
100
75
50
25
0
Normalised Total Fluorescence in Alexa647 Positive CHO-BC1 Cells
IC50=0.003
10-5
10-4
10-3
10-2
10-1
100
101
102
Fiber-knob Concentration
(ug/105 Cells)