Molecular Therapy
Methods & Clinical Development
Original Article

Multiparametric domain insertional profiling of adeno-associated virus VP1

Mareike D. Hoffmann,[1] Alina C. Zdechlik,[2] Yungui He,[1] David Nedrud,[2] George Aslanidi,[3] Wendy Gordon,[2] and Daniel Schmidt[1]

[1]Department of Genetics, Cell Biology & Development, University of Minnesota, Minneapolis, MN 55455, USA; [2]Department of Biochemistry, Molecular Biology & Biophysics, University of Minnesota, Minneapolis, MN 55455, USA; [3]The Hormel Institute, Austin, MN 55912, USA

Several evolved properties of adeno-associated virus (AAV), such as broad tropism and immunogenicity in humans, are barriers to AAV-based gene therapy. Most efforts to re-engineer these properties have focused on variable regions near AAV's 3-fold protrusions and capsid protein termini. To comprehensively survey AAV capsids for engineerable hotspots, we determined multiple AAV fitness phenotypes upon insertion of six structured protein domains into the entire AAV-DJ capsid protein VP1. This is the largest and most comprehensive AAV domain insertion dataset to date. Our data revealed a surprising robustness of AAV capsids to accommodate large domain insertions. Insertion permissibility depended strongly on insertion position, domain type, and measured fitness phenotype, which clustered into contiguous structural units that we could link to distinct roles in AAV assembly, stability, and infectivity. We also identified engineerable hotspots of AAV that facilitate the covalent attachment of binding scaffolds, which may represent an alternative approach to re-direct AAV tropism.

INTRODUCTION

Recombinant adeno-associated virus (rAAV) has proved to be safe and able to drive long-term expression in dividing and non-dividing human cells. Several AAV-based therapeutics have been approved by the US Food and Drug Administration (FDA) and numerous clinical trials using AAV for the treatment of genetic diseases are underway.[1–4] Despite the exceptional clinical potential of naturally evolved AAV serotypes, they could be substantially improved with respect to production yield, DNA packaging capacity, immunogenicity, cell type specificity, and infectivity.[5]

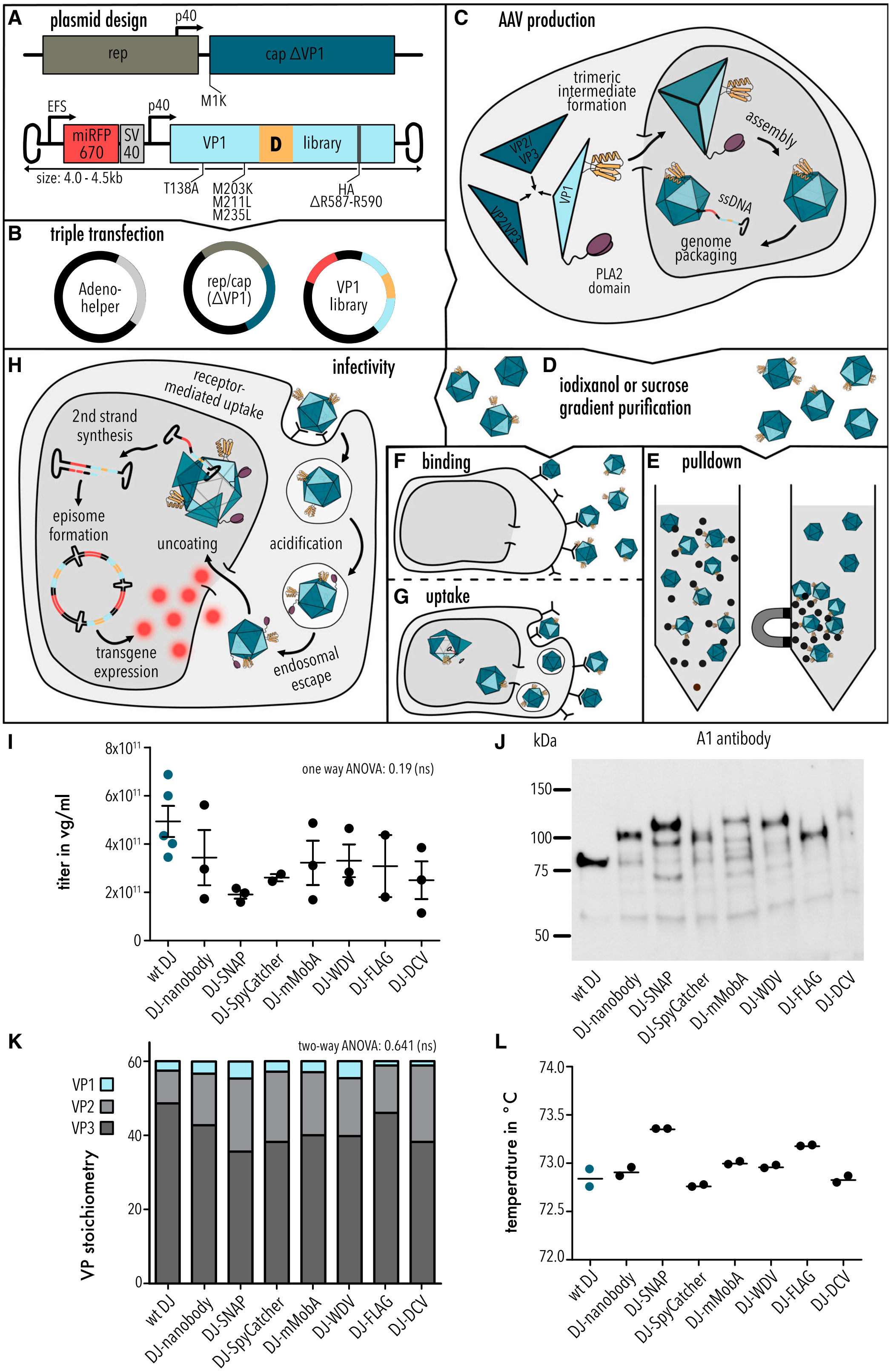
Addressing AAV's drawbacks through engineering is facilitated by its relatively simple structural and genetic organization. The ~4.7-kb single-stranded DNA (ssDNA) genome comprises two genes, *rep* and *cap*, flanked by inverted terminal repeats (ITRs). One open reading frame (ORF) of the capsid gene *cap* encodes for three viral proteins VP1 (737 amino acids [aa], 87 kDa), VP2 (600 aa, 72 kDa), and VP3 (535–503 aa, 62 kDa). It is expressed from the p40-promoter (C terminus of the *rep* gene) and translated from overlapping ORFs in a way that VP2 is lacking the N terminus of VP1 and VP3 is missing the N-terminal part of VP1/VP2.[6] Other *cap* ORFs express the assembly-activating protein (AAP)[7] and membrane-associated accessory protein (MAAP[8]). For recombinant production of AAV, the *rep* and *cap* genes between the ITRs are replaced by the transgene of interest. The capsid assembly takes place in the nucleus and entails a stochastic oligomerization process forming an empty capsid in the nucleus first,[9–11] followed by packaging of the ssDNA payload.[9–11] The capsid shell is composed of 60 VP monomers at an average ratio of 1:1:10 VP1, VP2, and VP3, respectively.[12–15] This ratio is highly divergent and assembly is stochastic, such that every capsid has a unique structural assembly.[11] The icosahedral capsid features a cylindrical pore at the 5-fold interface, depressions surrounding the 5-fold pore continuing through the 2-fold axis, as well as protrusions at the 3-fold axis.[16] Although the overall topology of AAV capsids is conserved across serotypes, Govindasamy et al. determined variable regions (VRs; VR1–9) mapping to surface loops of the capsid.[17] These VRs are commonly involved in virus binding to (co-)receptors on the surface of a target cell, such as heparan sulfate or AAVR,[18,19] and the subsequent process of endocytosis.[20–22] The unique N-terminal portion of VP1, also known as VP1u, is located inside the capsid and is indispensable for infection. Upon infection, acidification during endosomal trafficking causes unfolding of the VP1u domain so that it can be externalized through the pore.[23–27] A conserved phospholipase A2 domain (PLA2[28]) and nuclear localization signals (NLSs[29]), which are part of VP1u, can then facilitate endosomal escape and nuclear entry.[22] Upon interaction of the AAV with the nuclear pore complex, the virions enter the nucleus, where they are forwarded to the nucleus and the genetic payload is released.[22,27]

A variety of capsid engineering approaches have been applied in the past to improve the natural infection efficiency of AAVs ranging from shuffling of natural AAV serotypes, recovery of ancestral serotypes,

---

Received 17 April 2023; accepted 21 October 2023;
https://doi.org/10.1016/j.omtm.2023.101143.

Correspondence: Mareike D. Hoffmann, Department of Genetics, Cell Biology & Development, University of Minnesota, 420 Washington Ave SE, Minneapolis, MN 55455, USA.
E-mail: hoff1570@umn.edu
Correspondence: Daniel Schmidt, Department of Genetics, Cell Biology & Development, University of Minnesota, 420 Washington Ave SE, Minneapolis, MN 55455, USA.
E-mail: schmida@umn.edu



(legend on next page)

and peptide display. These methods resulted in significantly advanced capsid variants, such as AAV-DJ,[30] Anc80,[31] AAV-PHP.eB,[32] or AAV2.7m8.[33] Moreover, there have been successes with incorporating larger, structured protein domains. The first was fusing the green fluorescent protein (GFP) N terminally to VP2, which was useful for visualizing intracellular trafficking of AAV particles.[34] In the same manner, *Gaussia* luciferase[35] and an ankyrin repeat protein (DARPin[36]) were successfully incorporated into the capsid. Other studies incorporated domains for a cell type-specific targeting into VR4 of VP1 or VP2, including nanobodies,[37] HUH tags,[38] or DARPins.[39] While the aforementioned studies focus on the termini and VR4 or VR8 for domain insertions or peptide insertions, respectively, few studies attempted to comprehensively survey permissive capsid regions for peptide and domain insertions. Among them, Judd et al. constructed a random insertion library of mCherry into the VP3-encoding section of VP1 of AAV2, identifying only a single clone, in VR4, that tolerated insertion.[40] Another example is the study from Koerber et al.,[41] in which a transposon-based saturation mutagenesis approach was used to insert a hexahistidine peptide in AAV2 and AAV8, again pointing to VR8 as the prime peptide insertion spot. Taken together, there is still a paucity of large-scale domain insertional datasets that comprehensively assess (1) the effect on distinct biologically relevant AAV properties, i.e., assembly, binding, and infection, as well as (2) the effect of inserting domains with different physicochemical properties. In the absence of these data, the boundaries of AAV capsid plasticity with respect to accepting domain insertions while maintaining fitness (i.e., assembly, packaging, cell entry, etc.) are yet to be fully understood.

Our goal in this study was to survey as many proxies (AAV fitness phenotypes) for these distinct virion functions as possible. We posit that a comprehensive multiparametric AAV fitness dataset will facilitate the optimization of engineering AAV along multiple axes.[8,42,43] Furthermore, based on our previous studies in other protein models,[44,45] we hypothesized that systematic domain insertion (i.e., perturbation scanning) across different measured phenotypes may identify capsid determinants for assembly, stability, and dynamics, thereby revealing the topological organization of virus functions.

To achieve this goal, we here combined saturated programmable insertion engineering (SPINE[46]) with sequencing-based fitness assays to comprehensively determine multiple AAV fitness phenotypes upon insertion of the FLAG peptide tag as well as several large, structured protein domains into VP1 of AAV-DJ. Our data revealed a surprising robustness of AAV viral capsids to accommodate large protein domain insertions. We also found strong positional, domain-type, and fitness phenotype dependence of insertion permissibility, which we can map to contiguous structural units of the AAV capsid and link to distinct roles in AAV assembly, stability, and infectivity. We also identified additional engineerable hotspots of AAV accepting insertion of small protein tags that facilitate the covalent attachment of antibodies. These hotspots may enable alternative approaches to re-direct AAV tropism.

RESULTS

Domain insertional profiling in AAV-DJ VP1

We turned to AAV-DJ[30] as the testbed for inserting a peptide tag (FLAG) and six different protein domains (Figure S1) in between every two residues of VP1. Our rationale for focusing on VP1, as opposed to the more abundant VP3 or the non-essential VP2, was as follows: as the least abundant VP isoform, there are, on average, between one and five copies of VP1 incorporated per capsid.[12–15] Note that this is an average copy number based on bulk measurements[12–15]; because assembly is a stochastic process, there are many particles with copy numbers at the extreme tails (i.e., 0 copies or >10 copies).[11] We reasoned that keeping a low number of VPs carrying inserted domains, which are potentially very disruptive, is more likely to result in assembled and functional virions. This is akin to applying a low or intermediate amount of selection pressure in directed protein evolution experiments, which can reveal more facetted fitness phenotypes.[47] Furthermore, we had observed in a prior study that inserting large domains into VR4 (part of the VP common region) completely abolished AAV production unless it was limited to VP1 (or VP2) only.[38] Most importantly, focusing on VP1 allows us to interrogate the effect of domain insertion on AAV cell entry; VP1 is required for AAV infectivity as it contains the PLA2 domain that mediates endosomal escape.[26,28,48]

To generate separate expression constructs for the VP1 domain insertion library and *cap* expressing only VP2 and VP3, we duplicated the AAV-DJ *cap* gene and introduced mutations (M1K; T138A/M203K/M211L/M235L) to suppress expression of VP1 or VP2/VP3, respectively (Figure 1A; and Table S1). We also replaced the VP1 heparin binding domain (HBD: residues 587–590) with a hemagglutinin (HA) tag—originally intended but not used—for western blot tracking. The VP1-only *cap* gene was then subjected to SPINE.[46] In brief, the VP1 gene was divided into 14 fragments and insertion variants carrying a temporary genetic handle at each residue position within each fragment were micro-array synthesized as oligo libraries. The genetic handle was subsequently replaced by the FLAG peptide tag or one of the six domains. This cloning approach resulted in a

Figure 1. Design and analysis of AAV domain insertion libraries

(A) Schematic of library design. (B) Plasmids used for the triple transfection in AAV production. (C) Schematic of AAV capsid assembly. (D) Purification of AAV via gradient purification. (E–H) Schematic of workflows analyzing the fitness of AAV domain insertion libraries (i.e., pull-down, binding, uptake, and infectivity assays). (I) Quantification of packaging titer via qPCR. Data are means ± SEM. One-way ANOVA test p value 0.19, not significant (ns). (J) Representative western blot image of AAV domain insertion libraries stained with A1 antibody (detecting VP1 subunits). Predicted sizes for VP1 and its domain fusions are VP1, 81.9 kDa; VP1-nanobdy, 96.0 kDa; VP1-SNAP, 102.4 kDa; VP1-SpyCatcher, 92.7 kDa; VP1-mMobA, 104.0 kDa; VP1-WDV, 98.7 kDa; VP1-FLAG, 83.8 kDa; VP1-DCV, 95.5 kDa. (K) Western blot quantification of VP1, VP2, and VP3 subunits. Data are means (n = 3). Two-way ANOVA test p value 0.641, not significant (ns). (L) T_m of AAV libraries obtained by DSF assay. Data are means (n = 2).

library of VP1 variants with a peptide tag or protein domain, flanked by short linkers, inserted in between every two residues. We chose the FLAG peptide tag (DYKDDDDK) as one type of insertion because of its size similar to prior AAV peptide insertion screens.[49–51] With an eye toward redirecting AAV tropism, we focused on protein domains that themselves have retargeting abilities (nanobody binding to GFP[52]) or that enable covalent linkage of retargeting moieties (SNAP tag to link O6-benzylguanine derivatives;[53] SpyCatcher to link SpyTag fusions[54]), and three different HUH tags (WDV, DCV, mMobA) to covalently link ssDNA-conjugated molecules in a sequence specific manner.[55–58]

Since insertions can affect any of the steps in AAV packaging and infection, we attempted to independently assay different insertion variant fitness phenotypes. The relatively large insertion variant library size (744 AAV positions (including HA tag) × 7 motifs = 5,208 variants for each assay) necessitated a high-throughput format. We therefore devised assays in which the different fitness phenotypes of an insertion variant were assessed by next-generation (NextGen) sequencing (NGS) variant populations before and after a fitness test.[59] A requirement for our approach was a stringent linkage between genotype (the insertion variant) and the measured phenotype (determined by properties of the capsid into which this VP1 variant is assembled). We achieved this by flanking the VP1 variant library with ITRs such that the gene encoding a specific VP1 variant is likely to become packaged into the capsid that incorporated this variant following assembly (payload sizes for the different libraries are shown in Figure S1C). As a measure to reduce cross-packaging (a mismatch between packaged VP1 variant gene and VP1 variant protein that is incorporated into the capsid), we transfected producer cells with a significantly reduced plasmid copy number per cell. This has been demonstrated to reduce cross-packaging.[60]

We then used these VP1 variant input libraries, stratified by inserted motif, for helper-free virus production followed by gradient purification (Figures 1B–1D). Using NGS of packaged genomes from full capsids, we assayed AAV packaging fitness. Note, that the term “packaging” convolutes two processes, assembly of preformed capsid in the nucleus, which we cannot measure directly, and DNA packaging (requiring a preformed capsid) mediated by Rep52/40.[61] Other AAV properties, such as its propensity to aggregate and reducing yield from producer cells, are captured by packaging fitness as well. We calculated the packaging fitness by counting the frequency of a given VP1 variant (i) after packaging (s) relative to the frequency of that variant in the input library (u), normalized to wild-type AAV (wt):

$$W^i = \ln\left(\frac{f_s^i}{f_u^i} * \frac{f_u^{wt}}{f_s^{wt}}\right)$$

We measured absolute wild-type fitness by spiking in an AAV-DJ ITR-flanked VP1-only *cap* gene containing 10 synonymous mutations into the VP1-library mix (Table S1). Wild-type AAV-DJ and AAV-DJ with silent mutations showed no difference in production titers (Figure S2). Using a similar approach to count VP1 variants before and after selection, we established assays that determine pull-down fitness (using affinity purification material), cell binding, cell uptake, and infectivity fitness (Figures 1E–1H). Infectivity assays were based on transduction of HEK293FT cells with miRFP670-nano[62] expressed from the AAV payload (Figure 1A). This enabled flow sorting of infectious variants (enriched in miRFP670nano^{high} cells) (Figure S3).

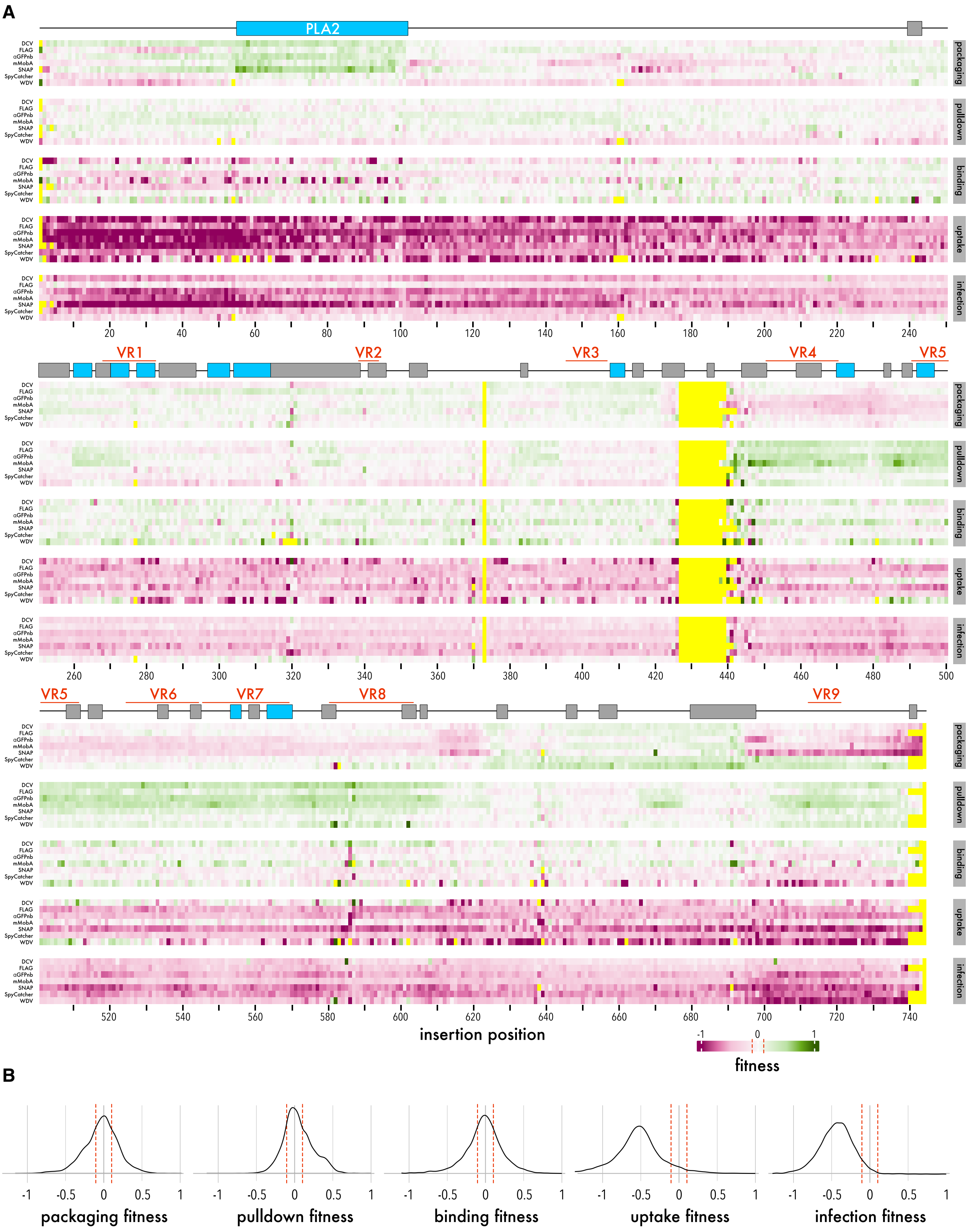
VP1 insertion library quality and completeness

Overall, coverage for all phenotype assays and inserted motifs was excellent (median dataset completeness is 98.5%; see Table S2 for sequencing statistics). Biological replicates were highly correlated for plasmid library, packaging, and pull-down assays (Pearson correlation coefficient: 0.67–0.99; Figure S4A). The infection assay was very noisy for some domains (Pearson correlation coefficient 0.38–0.88), likely related to small number of cells collected for the miRFP670nano^{high} cell pool (Figure S3). For all tested motifs, the majority of missing insertion positions (no data in either replicate; for example, positions 428–445) were missing in all phenotypes and the input library, suggesting that the dropout rate was related to library construction (Figures S4B and S4C). Since replicate 2 had a higher read depth and completeness overall, we used this dataset for further analysis. Median read depth across tested motifs and phenotype assays was 635 reads per position (Figure S4D; Table S2).

Domain insertions do no alter bulk properties

After helper-free production and iodixanol gradient purification, we measured virus titers by qPCR. While titers appeared somewhat lower for insertion libraries, this difference was not significant (Figure 1I; one-way ANOVA p value 0.19). Western blot with the A1 antibody (which recognizes a VP1-unique epitope) confirmed that VP1 was incorporated into some virions for all libraries (Figure 1J). However, we observed VP1 bands at various smaller sizes, suggesting that the libraries contained some unstable VP1 domain fusions that are susceptible to degradation. We next used the B1 antibody densitometry to determine bulk VP1/VP2/VP3 ratios (Figures 1K and S5). We found that wild-type AAV-DJ virion contains on average three VP1 copies per 60-mer capsid, in line with prior studies.[11,15] There was no significant difference in VP composition for the different insertion libraries (two-way ANOVA p value 0.641), nor VP1 content (one-way ANOVA p value 0.206). To further characterize our libraries on the bulk level, we measured capsid melting temperatures using differential scanning fluorimetry. While SNAP and FLAG libraries had slightly elevated melting points (Figures 1L and S6), all libraries were overall remarkably similar and within a 1°C range of AAV-DJ, suggesting that domain insertions in VP1 did not significantly affect bulk capsid stability of the insertion library. We next used negative-stain electron microscopy to compare full/empty capsid ratios of wild-type AAV-DJ, SNAP, and nanobody libraries and found that full particles comprised between 80% and 90% of all samples (Figure S7).

Note that the western blot, capsid melting assays, and electron microcopy data merely determine average properties of our libraries; with these bulk measurements, we cannot determine VP ratio at the



(legend on next page)

individual capsid level or measure how individual VP1 variants and their copies affect capsid stability or full/empty ratios. Previous studies have shown that the actual single-capsid stoichiometry varies tremendously, even without any modification, and up to 60% of capsids lack either VP1 or VP2 altogether.[11] Taken together, our data suggest that on average (i.e., considering the entire insertion library) bulk properties are not altered by domain insertions. This suggests either that assembly of AAV capsid with VP1 variants is a rare event (such that bulk properties are mostly determined by VP2/VP3-only capsids) or that many VP1 variants are compatible with capsid assembly, suggesting a certain degree of conformational plasticity. To resolve this question, we analyzed our libraries for their insertional fitness on the single-capsid level.

High-resolution AAV fitness profiles across different phenotypes

By measuring enrichment or depletion of individual VP1 variants in different selection assays, we found that different insertion types at different sites do have divergent impacts on AAV fitness. These data are summarized in Figure 2, showing a heatmap of insertion fitness of all seven motifs in all 744 VP1 positions, segregated by measured fitness phenotypes (packaging, pull-down, binding, uptake, and infectivity). Fitness values are mapped from magenta over white to green, corresponding to lower to higher than wild-type AAV-DJ fitness (white). Poisson errors were generally low (Figure S8) and only slightly elevated at the junctions between the 14 different fragments used in the SPINE-mediated assembly of the AAV insertion library. Median error with respect to dynamic range of each fitness assay was (0.102/2) log units = 5.1%, which is typical for these types of high-throughput protein fitness assays.[63] While overall there was a strong dependence on insertion position and type of inserted domain for all assays, uptake and infection fitness were decreased for nearly every position. One notable exception was the increased fitness of the WDV domain between positions 480 and 530. We currently do not know the underlying mechanism, but domain-idiosyncratic fitness effects support the idea that AAV capsid structure and function are sensitive to the structural context of a perturbation.

Effects of domain insertions in VP1u

Focusing on packaging fitness, we saw a distinct increase in fitness when motifs (in particular, nanobody, SNAP tag, or DCV HUH tag) were inserted into the PLA2 domain of VP1u (Figure 3). For the same positions, a pull-down with the respective affinity materials (e.g., GFP-agarose beads) showed significant enrichment when the nanobody, mMobA, DCV, and SNAP tag were inserted into VP1u (residues 1-160; Figures 4 and S9). Both observations can only be true if these VP1 variants are incorporated in the capsid such that the inserted motif is accessible on the capsid exterior. While speculative, a likely explanation is that the inserted motif forces VP1u to remain external, thus removing a steric barrier to genome packaging, which would result in increased recovery of this variants after both fitness assay selections. FLAG tag insertions into VP1u, which, because of their small size, presumably did not interfere with internalization, did not increase pull-down fitness. Neither did WDV insertions, which is consistent with their deleterious impact on packaging fitness. Overall binding fitness was neutral (peaked around wild-type fitness; Figures 2 and S10), which was expected as most VP subunits (VP2 and VP3) contain the wild-type determinants of proteoglycan and AAVR binding.[21,64] However, we observed a notable drop for binding fitness for VP1u insertions of motifs that package well (e.g., nanobody; Pearson correlation coefficient -0.641), possibly due to steric hindrance of virus binding when it carries large external motifs. The only motifs that were not impaired for binding upon VP1u insertions are the FLAG tag, which can be explained by its small size, as well as SpyCatcher and WDV, which are motifs that our pull-down data suggest were not compatible with virus packaging when inserted into VP1u.

Consistent with the role and required timing of VP1u in virus trafficking upon endocytosis, uptake fitness (our assay measures presence of viral genome in any compartment inside the cell[65]) was impaired for all motifs inserted into the N terminus of VP1 (Figures 2 and S11). It has previously been demonstrated that premature exposition of VP1u decreases infectivity,[66] meaning that variants with (as we speculate) pre-externalized VP1u escape less efficiently from the endosome and are degraded in the lysosomal compartment.[22] We observed the same for infectivity fitness, with the notable exceptions of FLAG and WDV.

Effects of domain insertions near AAV's 3-fold axis

Insertion into protrusion near the 3-fold axis (residues 420–620; including VR4–8) generally impaired virus packaging (Figures 2 and 3), which is consistent with the highly interdigitated structure of this region and its role in early assembly of VP trimers that are then forwarded to the nucleus as capsid building blocks.[9,10,27] VP1 with insertions in this region may interfere with efficient trimer assembly (through a kinetic mechanism or by promoting off-pathway products), which would lower the overall capsid assembly efficiency. Nevertheless, several lines of evidence suggested that there appears to be some plasticity in trimer assembly to accommodate VP1 insertion variants so that these trimers can be incorporated into assembling capsid. For one, the overall hit to packaging fitness depended on the specific inserted motif. As expected, we saw that FLAG peptide insertions were relatively benign, but so were insertions of two HUH tags (DCV and WDV) and SpyCatcher (Figures 2 and 3). There was no clear correlation with motif size, hinting at more complex determinants for insertion fitness in this region. Second, despite impaired packaging fitness, pull-down fitness was greater than wild

Figure 2. Fitness of AAV domain insertion libraries

(A) Heatmap representing the fitness of each motif insertion in each VP1 position for AAV packaging, pull-down, binding, uptake, and infectivity assays. Green indicates higher and magenta lower fitness than AAV-DJ (white). Yellow denotes positions without data. VP1 secondary structure elements and VR1-9 are indicated on top. (B) Fitness distributions of all AAV libraries compared to AAV-DJ (fitness = 0) ± standard error (red dashed lines).

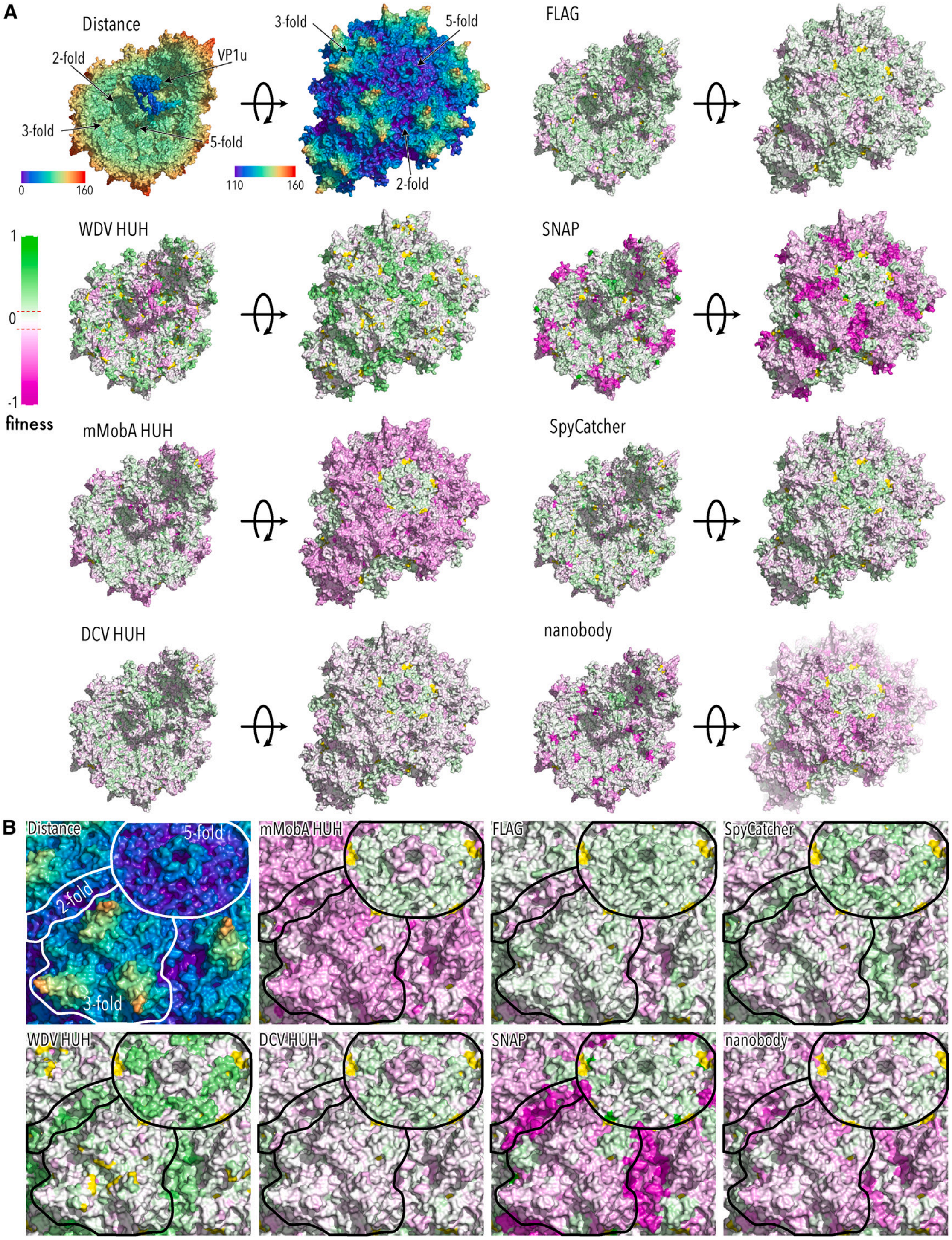


Figure 3. Packaging fitness of AAV domain insertion libraries mapped to the capsid structure

(A) Top left corner: AAV-DJ capsid structure view from the inside (left) and outside (right) radially color cued. The 2-, 3-, and 5-fold axes are indicated. VP1u domain was modeled using RoseTTAFold[98] and manually positioned. All other structures show packaging fitness heatmaps of the indicated domain insertions. Green indicates higher and magenta lower fitness than AAV-DJ (RCSB PDB: 7KFR). (B) Zoom of the outside structures from (A). The 2-, 3-, and 5-fold axes are outlined.

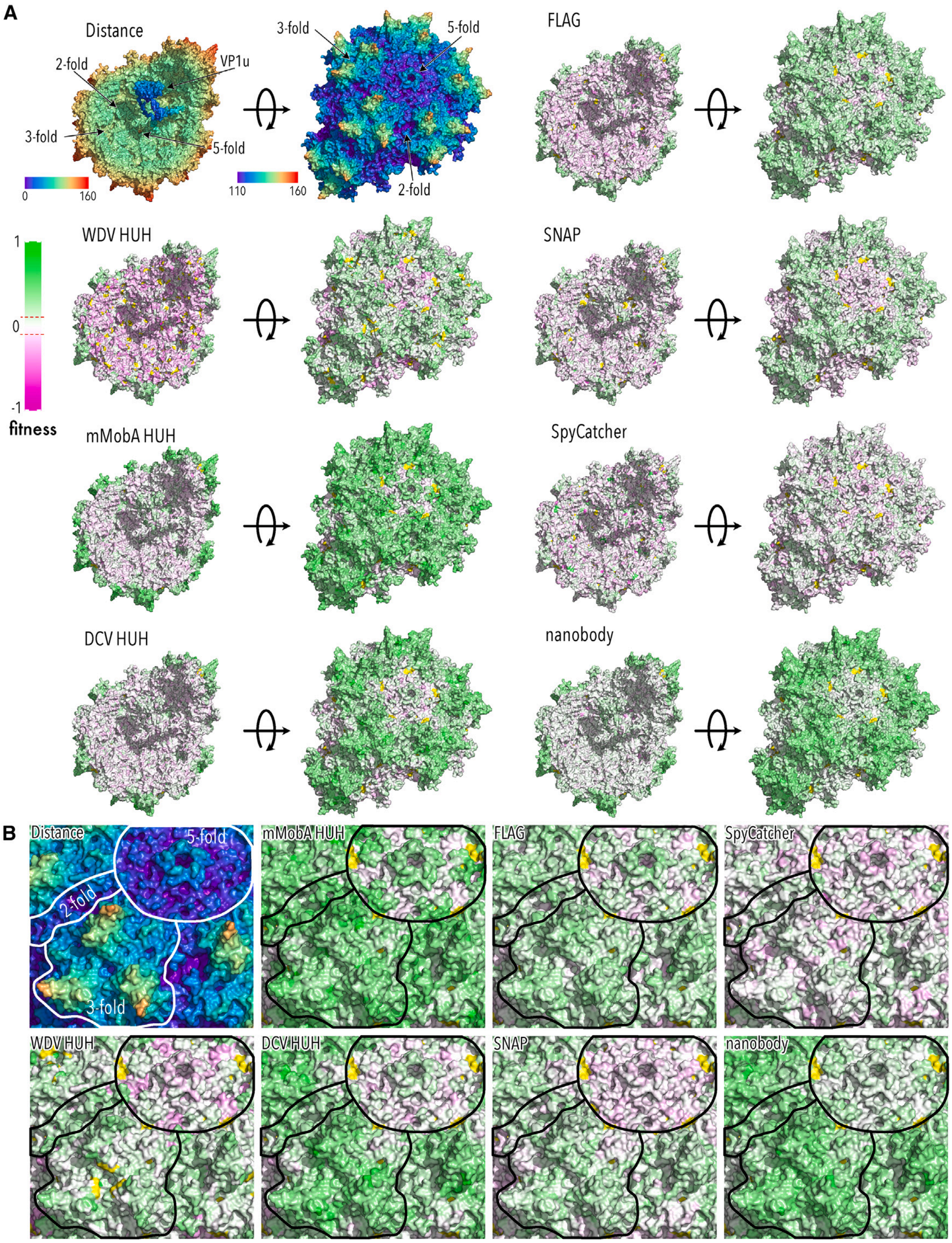


Figure 4. Pull-down fitness of AAV domain insertion libraries mapped to the capsid structure

(A) Top left corner: AAV-DJ capsid structure view from the inside (left) and outside (right) radially color cued. The 2-, 3-, and 5-fold axes are indicated. VP1u domain was modeled using RoseTTAFold[98] and manually positioned. All other structures show pull-down fitness heatmaps of the indicated domain insertions. Green indicates higher and magenta lower fitness than AAV-DJ (white) (RCSB PDB: 7KFR). (B) Zoom of the outside structures from (A). The 2-, 3-, and 5-fold axes are outlined.

type for all motifs, suggesting that they can become incorporated into purified AAV capsids, albeit at a lower overall efficiency (Figures 2 and 4).

Similar to what we observed for insertions into the N terminus of VP1, binding fitness was not impaired for insertions in the 3-fold protrusion (Figures 2 and S10). In fact, the three HUH tags showed increased binding, which could be due to their generally higher cationic surface charge (Figure S1B) aiding interaction with negatively charged components of the extracellular matrix, such as proteoglycans. Consistent with higher surface binding, we found higher uptake fitness than wild-type AAV-DJ in the case of DCV and WDV when inserted into VR5 or VR8 (Figures 2 and S11). We note that binding and uptake fitness measured in this high-throughput assay for insertion of mMobA into VR4 match our results from our previous engineering of this region.[38] Despite higher binding and uptake, none of the insertions in this region could achieve wild-type infection efficiency (Figure 2), suggesting that domain insertion affected later steps of virus trafficking to the nucleus.

Effects of domain insertions near AAV's 2-fold valleys

The neighborhood near the 2-fold symmetry center, which includes VR9, emerged as another region with distinct motif-specific phenotypes. This is consistent with earlier studies showing that dynamics of the 2-fold regions are essential for genome packaging[67] and AAV infectivity.[68] Here, we observed both strongly deleterious fitness (e.g., SNAP tag) and strongly beneficial fitness (SpyCatcher and WDV). In all cases, pull-down fitness was positive (Figure 4), suggesting some degree of VP1 variant incorporation in this region. Interestingly, binding, uptake, and infectivity were impaired for most motifs.

Unbiased clustering of insertion fitness reveals contiguous functional units in AAV capsids

Taking our fitness measurements across all motifs, all insertion positions, and all measured phenotypes in aggregate, we noticed patterns in fitness variance that appear correlated in contiguous regions of the AAV capsid. For example, packaging fitness varied predominantly in the PLA2 domain of VP1u and the 2-fold symmetry axis (Figure 5A). Focusing on residues unique to the 2-fold interface, we found that packaging fitness distributions of interface and non-interface residues were not significantly different when DCV or FLAG peptide were inserted (Figure 5C). However, fitness was significantly improved for mMobA, SpyCatcher, or WDV insertions into interface residues. SNAP tag insertions were strongly deleterious. The variance at the 3-fold symmetry axis was markedly different. As described above, all motifs except WDV lowered packaging fitness, which resulted in lower overall fitness variance at this interface (Figure 5D). Remarkably, uptake fitness variance was greatest in the protrusion around the 3-fold axis and still considerably high along the 2-fold axis (Figure 5B). For all measured phenotypes, variance was relatively low around the 5-fold symmetry axis (Figure S12).

If we think of different inserted motifs as different degrees of perturbation (e.g., weak for FLAG peptide insertion, strong for large nanobody insertion), then the positional and domain-type dependence of insertion permissibility we observed in our data suggests that we were measuring spatially resolved information of how different aspects of AAV fitness responds to these different degrees of perturbation. To link insertion permissibility phenotypes to mechanistic structure/function relationships, we used an unbiased clustering approach (Uniform Manifold Approximation and Projection [UMAP]).[69] This resulted in five robust clusters that map to regions of the capsid with distinct roles in AAV biology (Figure 6A). Importantly, these clusters mapped to structurally contiguous (not interspersed) regions of the AAV capsid (Figures 6B and 6C). Cluster 1 contains VP1u in addition to residues lining the bases of the 3-fold and 5-fold axes. Cluster 2 forms an extended network that comprises the HI loop and connects to the 3-fold axis protrusions. Cluster 3 represents protrusion at the 3-fold axis and residues on the external turns of the DE loop that line the pore at the 5-fold symmetry axis. Cluster 4 maps to the depression near the 2-fold symmetry axis and buried regions, which are part of the 3-fold axis. Cluster 5 predominantly maps to residues that line the capsid interior and the 5-fold pore, or that interdigitate HI loop, external residues of the 2-fold valley, and 3-fold protrusions (Figure 6B).

To understand the underlying mechanisms that drive clustering, we segregated fitness phenotype distributions by cluster identity (Figure 6D). Considering packaging fitness, we found that insertion into two clusters (1 and 5; containing VP1u and the network of residues that connect HI loop to the base of the 3-fold protrusion) were associated with improved packaging fitness, while clusters 3 and 4, which represent the interdigitated external region 3-fold and 2-fold axes, were associated with poor packaging fitness. Insertions into capsid-lining regions (cluster 2) were neutral with respect to packaging. With different measured phenotypes, these association patterns change: considering pull-down fitness, we found that clusters representing buried residues or those lining the capsid interior (i.e., clusters 1, 2, and 5) were associated with poor fitness compared to those in externally accessible regions (cluster 3 and 4). For uptake, cluster 1 (which contains VP1u) had the worst fitness and cluster 3 (the 3-fold protrusion) was closest to wild-type fitness.

Disulfide crosslinking to probe conformational flexibility

Conformational dynamics between different capsid regions plays a central role in AAV structure and function, including packaging and infectivity.[9,24,26,67] Thus, clusters may emerge from our data because they represent units of correlated conformational plasticity that are also correlated in the way they respond to a perturbation (i.e., domain insertion). We therefore would expect that changing inter-cluster dynamics affect packaging and/or infectivity more than intra-cluster changes. One way to test this idea is by replacing two proximal residues by cysteines, such that a cystine disulfide link is potentially formed once AAV is exposed to an oxidizing environment (i.e., after release from producer cells). If the two mutated residues that undergo a cystine disulfide link are part of the same cluster of correlated conformational plasticity, and assuming that the individual cysteine substitutions are benign, we expect a less significant effect on

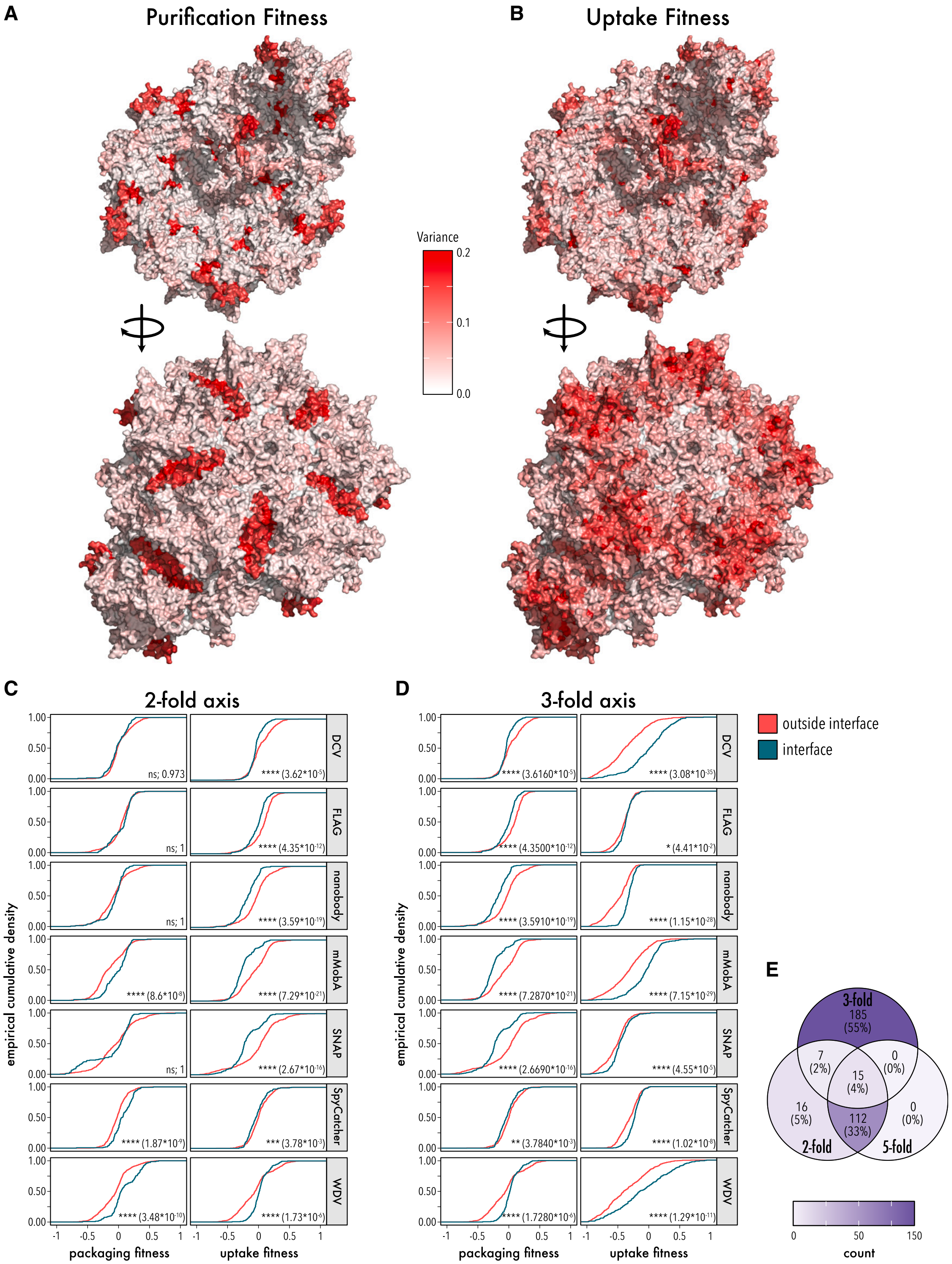
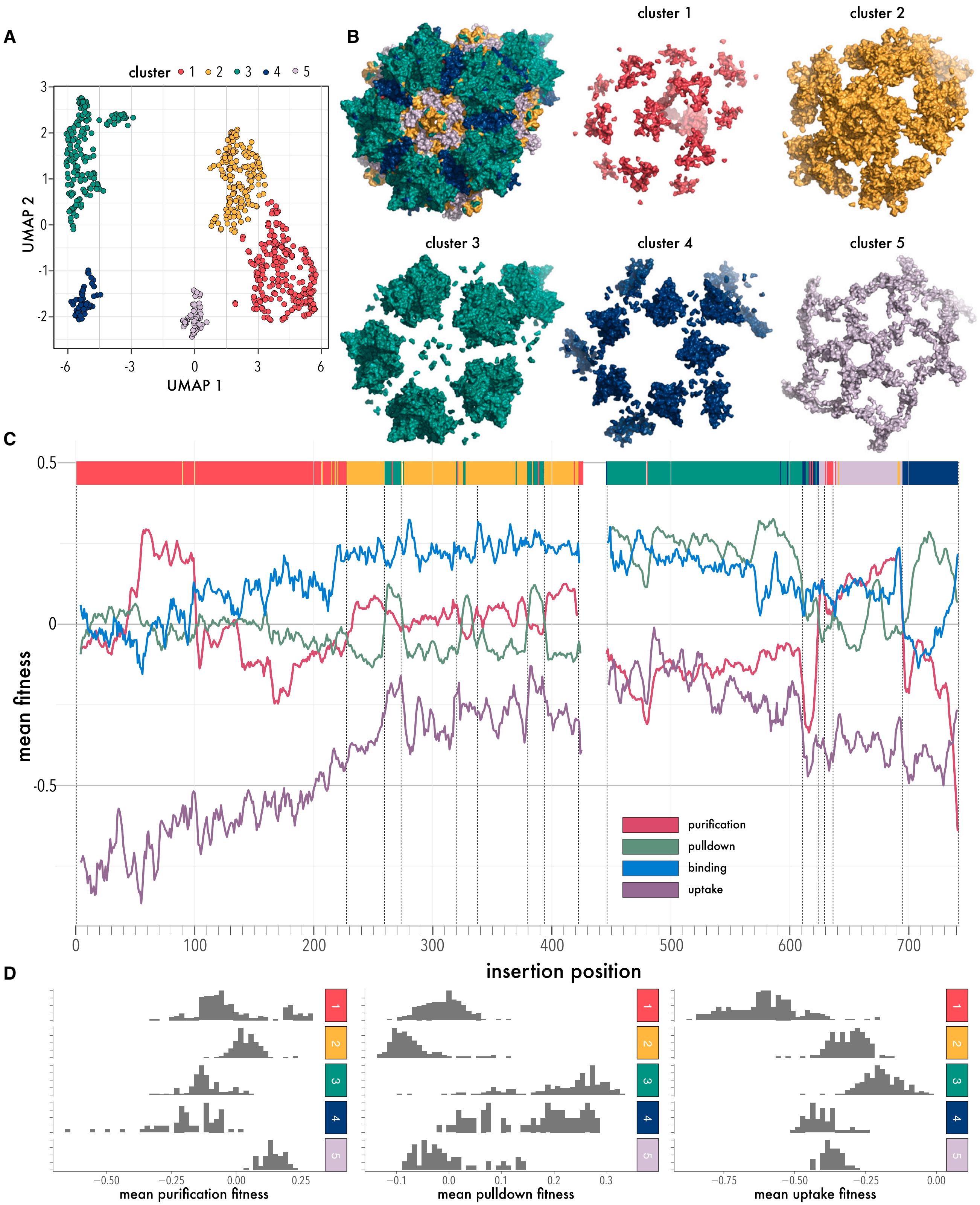


Figure 5. Variance of packaging and uptake fitness

(A and B) Variance of packaging fitness (A) and uptake fitness (B) from all domain insertion libraries mapped to the AAV capsid structure. The capsid inside (top) and the capsid outside (bottom) are shown (RCSB PDB: 7KFR). (C and D) Empirical cumulative density insertional fitness of residues within (petrol green) and outside (red) the 2-fold axis (C) and 3-fold axis (D). Significance of distribution differences was tested using a two-sided, two-sample Kolmogorov-Smirnov test. Significance level and p values are shown. (E) Venn diagram showing which of the 335 AAV interface residues are unique and shared among interfaces.



(legend on next page)

packaging and/or infectivity compared to when the two residues belong to different clusters. Our choices of residue pairs are summarized in Figure S13A. Unlike domain insertions, which were only done in VP1, cysteine substitutions were introduced into all VPs. All single and double mutants produced near to wild-type titers (Figure S13A; one-way ANOVA not significant [NS]; Dunnett's test with wild-type AAV-DJ as control, NS).

Some single cysteine mutants had dominant deleterious effects on infectivity (e.g., W608C, H292C). For double mutants in the "within cluster" set that had wild-type single-mutant infectivity, infectivity was comparable to wild-type (e.g., H625C/Y426C; Figure S13B), suggesting the minimal disruption of conformational dynamics. For pairs that belong to different clusters, all but one (H643C/Y350C) showed effects on infectivity that differed from what was predicted based on the individual single mutants (Figure S13B; see the Note S2 for a description of individual variants). While disulfide bond formation is also influenced by the protein structure and surrounding amino acids and we are not testing an exhaustive set of potential crosslinks, our results support the idea that at least the variation of infectivity fitness that contributes to cluster identity has a basis in different conformational dynamics.

Engineerable hotspots near the 2-fold axis and in the HI loop

We previously used the HUH tag mMobA in AAV-DJ VR4 to covalently link targeting scaffolds to the AAV capsid, which redirected AAV tropism *in vitro*.[38] Here, we interrogated the entire AAV capsid for suitable HUH tag insertion hotspots. Comparing fitness maps for all three HUH tags, there were many differences among tags, which is likely related to their different biophysical properties (Figure S1). We noticed that WDV was remarkably different from all other inserted domains in several regards. For one, packaging fitness was improved over wild-type when this domain was inserted into HI loops or along the 2-fold axis (Figure 3). Binding fitness was generally strong, but insertions into the 2-fold axis were deleterious (Figure S10); this was the only insertion type for which we saw a deleterious phenotype for this assay. For uptake fitness, we observed a strong segregation in fitness between 3-fold protrusion and 2-fold and 5-fold axes (Figure S11). Given that most previous studies have investigated VRs in the 3-fold protrusion for capsid engineering, we turned our attention to insertion sites near the 2-fold and 5-fold axes that had near wild-type packaging fitness in the NGS-based assay. We produced 10 VP1 WDV insertion variants individually as crude cell lysates and measured titers, which all were comparable to wild-type (Figure 7A, one-way ANOVA NS). Testing each crudely enriched variant for the ability to infect HEK293FT cells, we found that several WDV variants inserted into surface exposed sites retained infection potency (Figure 7B). Among those were insertions into the 3-fold protrusions (S268), DE loop of the 5-fold pore (T331), HI loop (N664), and two sites along the 2-fold axis (Y702, K708). All sites had positive fitness in the NGS-based pull-down assay, suggesting the WDV-VP1 does become incorporated into AAV capsid (Figures 2 and 4). For two variants at particularly non-intuitive regions, namely N664 in the HI loop and K708 at the 2-fold axis (Figure 7C), we produced iodixanol-gradient-purified virus, which both trended to produce at higher titer compared to wild-type (Figure 7D). All variants incorporated VP1 as confirmed by western blot (Figure S14), albeit at levels that varied for N664 but not K708 (one-way ANOVA for VP1 content, p value 0.00993; Dunnett's test for pairwise comparison with wild-type AAV-DJ as control, N664 p value 0.0074; K708 p value 0.4755). As we have previously shown,[38] HUH tags mediate the attachment of ssDNA antibodies to AAV, which in turn increased infectivity in cells that express, on the cell surface, the antigen recognized by the antibody. Using surface-expressed GFP (GFP-GPI) as a test case, we tested infectivity of the two purified WDV variants and wild-type AAV-DJ with and without conjugation to an ssDNA-linked anti-GFP antibody. Note that expression of GFP-GPI alone reduced cell health, likely related to endoplasmic reticulum (ER) stress. Although infectivity of WDV variants was unaffected by conjugation to anti-GFP just like for wild-type AAV-DJ (expected as it was a mock conjugation since it does not contain an HUH tag), we saw a boost to infectivity for both N664-WDV and K708-WDV upon co-expression of surface-expressed GFP (Figure 7E), suggesting that anti-GFP became conjugated to WDV and then enhanced infectivity by directing AAV toward surface-expressed GFP as a binding receptor (two-way mixed measures ANOVA p value 0.0063).

DISCUSSION

Viral vectors are an essential component for gene delivery in therapies treating inherited disorders and cancer. AAV is widely used in both approved therapies and in ongoing clinical trials because of its good safety in humans and ability to drive long-term expression in both dividing and non-dividing cells. Unfortunately, several evolved properties of AAV are mismatched to clinical needs (e.g., broad tropism, limited payload capacity, existing human serum-immunity) or they pose biomanufacturing challenges (e.g., scale-up of helper virus-free production, yield of full virions to maximize potency).

Motivated by these challenges, there have been extensive efforts to re-engineer AAV properties in the past, including directed-evolution approaches, such as repeated mutagenesis, capsid shuffling,[70] viral display of short peptides,[49–51] and adding larger, structured targeting scaffolds, such as antibodies, nanobodies, DARPins, or affibodies.[36–39,71] Most of these studies have focused on regions in the 3-fold protrusion, commonly VR8, VR4, or VP termini. Recently, deep mutagenesis of the entire capsid protein, combined with

Figure 6. Unbiased clustering of insertion fitness

(A) UMAP cluster analysis of the AAV domain insertional profiling data resulting in five distinct clusters. (B) Cluster map to distinct capsid regions: (1) N terminus of VP1 and bases of the 3-fold and 5-fold interface in red; (2) the HI loop, the pore and the inner connecting residue layer in yellow; (3) protrusions of the 3-fold interface in green; (4) 2-fold axis in blue; and (5) the surrounding of the 5-fold pore in light lilac. (C) Mean domain insertion fitness of packaging, pull-down, binding, and uptake shown by residues aligned to the UMAP clusters in (A). (D) Distribution of insertion fitness of packaging, pull-down, and uptake for each cluster.

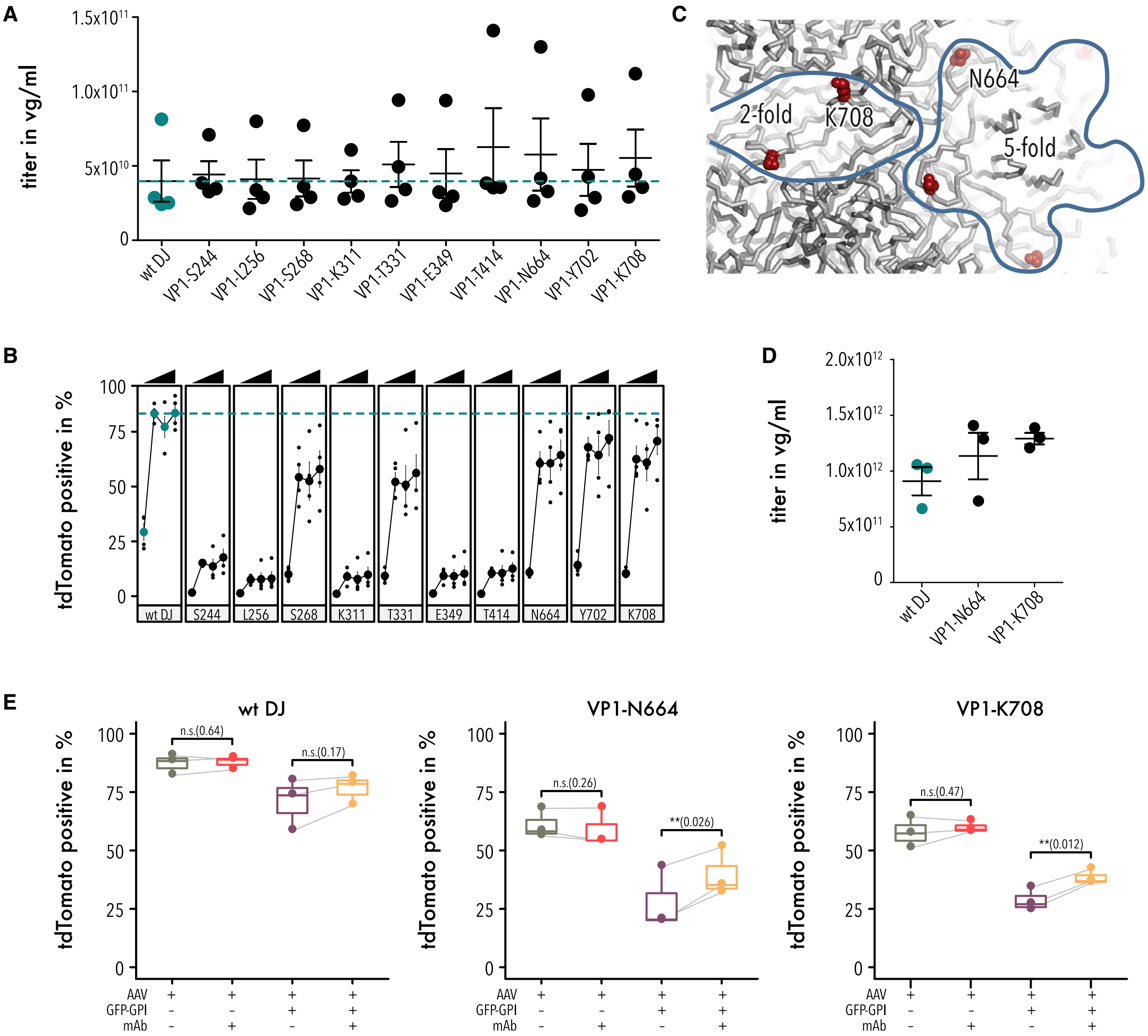


Figure 7. Fitness of WDV insertion variants

(A) Quantification of packaging titer of AAV-DJ and 10 different WDV insertion variants via qPCR. Data are means ± SEM. (B) Infection fitness of the WDV insertion variants in (A) quantified by measuring the percentage of tdTomato-positive cells 48 h post transduction at the MOIs of 1×10^2, 1×10^3, 5×10^3 and 4×10^4 vg/cell. Data are mean ± SEM. (C) Zoom to the 2-fold and 5-fold axes (outlined) of the capsid surface. Positions of N664 and K708 are shown as red spheres. (D) Quantification of packaging titers of AAV-DJ and WDV insertion variants N664 and K708. Data are mean ± SEM. (E) Infection fitness of the WDV insertion variants N644 and K708 quantified by measuring the percentage of tdTomato-positive cells 48 h post transduction at an MOI of 1×10^4 vg/cell. If indicated, cells were co-expressing GFP-GPI and/or an ssDNA-anti-GFP antibody was added. Data shown as boxplots. Lower and upper hinges of boxes indicate 25th and 75th percentile, respectively. Mean is indicated by a horizontal bar in each box. Whiskers extend 1.5 × IQR. Two-way mixed measures ANOVA was used to test the significance of variance between means (AAV-DJ, N664, K708) for conditions with and without GFP present. Presence of the monoclonal antibody (mAb) was a significant source of variation when GFP-GPI was expressed (p value 0.0063) but not without (p value 0.514). Significance levels and p values for pairwise comparisons using a Bonferroni correction are shown.

machine learning,[8,42,72] demonstrated that learned sequence and function relationships can aid the prediction of sequence variation to improve desired AAV traits. For example, Ogden et al. performed a comprehensive single-amino-acid substitution, insertion, and deletion screen.[8] They identified a higher tolerance in the VRs and particularly in the 3-fold protrusions. While comprehensive, this study focused on single-amino-acid changes. Based on prior work in other protein systems,[44,45] we hypothesized that the domain insertion permissibility differs substantially considering the larger size, structure/folding, and polarity of entire domains.

We thus combined the concept of deep mutagenesis with scaffold insertion to systematically measure the fitness of AAV containing VP1 with seven different motifs inserted in between every two residues. This comprehensive analysis quantitatively links where different structured motifs can be inserted into VP1 to retain compatibility with AAV virion packaging, cell binding, and uptake. Note that fitness measurements are derived from AAV capsids with a variable copy number of VP1 located at random faces, which could affect phenotype penetrance; e.g., a higher number of VP1 proteins might lower the packaging fitness due to steric hinderance of genome packaging but could increase the cell type-specific infection potency due to improved receptor binding mediated by the inserted domain or large number of PLA2 domains. The fitness profiles show that there is a strong dependence on insertion position and type of inserted motif, with several regions showing diverging fitness for different measured phenotypes (e.g., VP1u for packaging vs. cell uptake; Figure 2). This highlights that the outcome of sequence variation (encompassing single-residue changes to large-scale insertions) can have multi-facetted impact on AAV properties, and it calls for integration of assays across several clinically relevant AAV attributes to safeguard against inadvertent optimization for undesired traits. In recent years, we have gained broad access to precision variant library engineering,[46,73,74] NextGen sequencing (enabling counting number of sequence variants in a highly diverse population before and after applying a test for fitness[59]), unified analytical frameworks to interpret these large dataset,[75] and machine learning approaches for deciphering sequence/function relationships.[76] Taken together, generating and interrogating large AAV variant libraries has become feasible. Our focus can now shift toward what these datasets tell us about AAV biology and how they can guide and accelerate viral vector engineering.

For example, while fitness of many insertion sites is consistent with known AAV structure and function (e.g., the importance of trimer assembly along 3-fold axis in capsid assembly[9]), the high packaging fitness when motifs are inserted into the PLA2 domain of VP1u was surprising. Several studies have elucidated VP1u and VP2 dynamics as part of events after virus uptake by the cell, in which the externalization of VP1u PLA2 is a required step for endosomal escape.[28,48] VP1u internalization during maturation of AAV virions is less well understood but appears to be coordinated with genome packaging.[67] The existence of such mechanism would reconcile recently proposed models of stochastic assembly[11] (which should result in both internalized and externalized VP1u) and the well-known requirement for heating or pH lowering to induce VP1u externalization before it can be detected by VP1-specific antibodies.[24,26,77,78] It is plausible that domain insertions interfere with this process, leaving VP1u externalized, which would leave more room inside the AAV capsid for genome packaging by Rep52/40 due to lack of steric hindrance (one VP1 occupies 85×10^3 Å^3 or one-thirty-fifth of the available space inside the capsid), but at the loss of infectivity. Interestingly, VP1u and VP2 of the related parvovirus B19V have acquired a receptor-binding domain insertion just upstream of PLA2 and are always external,[79–82] supporting the idea that addition of extra domains into VP1u leaves it externalized. Given that precise timing of VP1u externalization in the correct endosomal compartment is required for maximum infectivity (premature externalization reduced infectivity[66]), this may point to an evolutionary mechanism that balances virion packaging efficiency with infectivity. If the full/empty ratio is fundamentally constrained by a packaging/infectivity balance, this would have implications for biomanufacturing of AAV in which one of the major ongoing efforts is to find ways to enrich full particles. One prediction of our hypothesis is that full particles may contain fewer VP1, on average, compared to empty particles, and this is negatively affecting infection potency. The observation that overexpression of VP1 inhibits rAAV packaging is consistent with this idea.[83] There may exist a Goldilocks regimen, just the right copy number of VP1, that maximizes both genome content and endosomal escape. Further research is required to fully test this hypothesis, including measuring VP stoichiometry and genome content at the single-capsid level.

Unbiased clustering of insertion fitness across several phenotypes also revealed a topological organization of AAV into regions that can be linked to correlates in AAV capsid assembly, genome packaging, and infectivity. Given that many of these roles have a basis in distinct regimes of capsid stability and dynamics, we hypothesize that inserting different domains (which represent different degrees of perturbation) probes conformational plasticity at or near the insertion site. Put simply, it probes whether the insertion site is conformationally rigid (allowing no insertions) or conformationally flexible (allowing some or all insertions). Clusters emerge because conformational plasticity of different capsid regions impinges differently with different measured phenotypes (e.g., capsid flexing required for efficient genome packaging[67] vs. flexing during externalization of VP1u/PLA2 after cell uptake, which only happens in full, but not empty, capsids[24]).

Similar ideas of spatially contiguous protein regions linked to specific functions have been proposed in the past, including protein “sectors” mapped through measuring amino acid co-evolution,[84,85] regional conformational flexibility mapped by circular permutation profiling[86] and domain insertion,[44,45,73] or revealing the functional architecture of an enzyme from high-throughput enzyme variant kinetics.[87] At their core, all these approaches use mutations to perturb sequence/function relationships. Similarly, by perturbing VP1 through domain insertion and measuring how it responds (in terms of packaging, infectivity, etc.), we learn how AAV structure intersects with AAV function. Going forward, domain insertional profiling in the background of different genetic backgrounds (i.e., serotypes) may further separate general principles of AAV packaging, function, and serotype-specific properties (e.g., stability, immune evasion, tropism). As we have done here, these should focus on alterations of VP1 or VP2, which have a significantly lower copy number in the assembled capsid and are more likely to be benign compared to VP3 or all VP modifications.

Our systematic domain insertion approach also revealed additional opportunities for viral engineering. Following the intuition that engineering non-conserved, surface-exposed, and tropism-determining loops is the likeliest path to change AAV properties, much of AAV engineering so far has focused on N termini of capsid proteins or variable loops of the 3-fold protrusions.[36,38,40,71,88–92] However, comprehensive studies we and others conducted[45,86] suggest that there is significantly more to explore. We found that permissibility to domain insertion is not correlated with conservation, surface exposure, or other static, structural features. Instead, dynamic features, such as regional flexibility, are predictive as to where a domain insertion is tolerated. In this study, we identified two regions near the 2-fold and 5-fold axes that can tolerate the insertion of HUH tags (15 kDa), which in turn enable the covalent linkage of antibodies (150 kDa). In a proof of principle, we found that insertions have little impact on production levels and provide a modest boost to infecting cells that express the antibody's cognate antigen. Further research and engineering are required to fully leverage the potential of these engineerable hotspots. They represent an exciting opportunity to sidestep the constraint of directed evolution of targeting the region near the 3-fold axis. As this region is important for receptor binding, it is also the most antigenic region. In fact, the binding sites for several proteoglycans, AAVR, and neutralizing antibodies overlap.[18,19,21,93] Approaches that shuffle the sequence of this region must apply selection pressure to selectively remove antibody binding while retaining the mode of cell binding and uptake. By providing an alternative site to which a targeting scaffold can be linked, it may be possible to address this challenge more effectively.

MATERIALS AND METHODS

Cloning and library generation

All plasmids and libraries used in this study are listed in Table S1 and were generated either by classical restriction enzyme cloning or Golden Gate Assembly.[94] Restriction enzymes were obtained from NEB; standard oligos, labeled oligos, as well as gBlocks from IDT; and oligo pools for library cloning from Agilent Technologies. For amplification of DNA sequences for cloning and NGS, the PrimeSTAR Max DNA polymerase (Takara Bio) and for colony PCRs the OneTaq Quick-Load Master Mix Polymerase was used (NEB). PCR products were analyzed on 1% TAE agarose gels, cut out and purified using the Zymoclean Gel DNA Extraction Kit (Zymo Research) by following the manufacturer's instructions. Post cloning, plasmids were transformed into NEB Stable Competent *Escherichia coli* cells and libraries into MegaX DH10B T1R Electrocomp Cells (Thermo Fisher), before plated on LB plates containing either carbenicillin (100 μg/mL) alone or a combination with chloramphenicol (25 μg/mL) depending on the selection marker(s) on the plasmids and libraries. To assess coverage of libraries, a small amount from the transformed cells was taken, and serial dilutions were prepared and plated on LB plates with the respective selection marker(s). The next day, colonies were counted to estimate coverage and colony PCRs were run to verify library diversity. Plasmids were isolated using the Zyppy Plasmid Miniprep Kit, ZymoPURE II Midiprep Kit, or the ZymoPURE II Maxiprep Kit (all Zymo Research).

For the generation of AAV-DJ insertion libraries, an altered cap-DJ gene sequence was used with mutated start sites for VP2 (T138A) and VP3 (M203K, M211L, and M235L), a T176A mutation eliminating the BsmBI cutting site, and a replacement of the HBD domain (R587-R590) by an HA tag (AYPYDVPDYAA). The libraries were created using SPINE.[46] In brief, the *cap* DJ-VP1 sequence was split up into 14 fragments, and oligos containing a genetic handle behind every amino acid position were designed. Every oligo contained barcodes for amplification, matching BsmBI restriction sites to assemble the 14 fragments, and BsaI restriction sites to swap out the handle. The genetic handle was first replaced by a chloramphenicol expression cassette flanked by BsmBI cutting sites to insert a selection marker for library presence. Next, the library was transferred into an AAV plasmid backbone encoding an elongation factor-1 alpha short (EFS)-driven miRFP670nano sequence terminated by an SV40-polyA, a p40 promoter with BsmBI sites for the library insertion and flanked by ITRs. Last, the chloramphenicol was replaced by different domains (nanobody, SpyCatcher, SNAP, mMobA, WDV, or DCV) or a FLAG tag. While the domains had 5-aa SGGGG-domain-GGGGS linkers, the FLAG tag was flanked by short SG-FLAG-GS linkers only. The DJ-VP1 silent mutation plasmid, which was used as a reference, was designed by introducing 10 silent mutations into the VP1-only DNA sequence of AAV-DJ. To this end, codons for either arginine, serine, or leucine were altered by two nucleotides each at positions that were 200–300 bp apart from each other. The silent mutation VP1 coding sequence was ordered as a gBlock and cloned into the same backbone as the plasmid libraries, i.e., miRFP670nano expression cassette and a p40 promoter to drive VP1 expression and flanked by ITRs. Cysteine point mutations were introduced by site-directed mutagenesis of the wild-type rep2-capDJ plasmid. WDV insertion variants were generated by inserting the domain into the above-mentioned p40-driven and altered AAV-DJ-VP1-only sequence by Golden Gate Cloning.

Tissue culture

HEK293FT cells (Invitrogen) and 293AAV cells (Cell Biolabs) were maintained in DMEM (Gibco) containing 4.5 g/L D-glucose, L-glutamine, 110 mg/L sodium pyruvate, and supplemented with 10% fetal bovine serum (FBS) (Gibco) and 100 U per mL penicillin/100 μg per mL streptomycin (Gibco). Cells were kept in a humidified cell culture incubator at 5% CO_2 and 37°C and passaged every 2–3 days when reaching 70%–90% confluency. For experiments with HEK293FT cells, plates were pre-coated with growth factor-reduced basement membrane matrix Matrigel (Corning) prior to seeding. If applicable, HEK293FT cells were transfected with a plasmid encoding GFP-GPI using Turbofect (Invitrogen) while seeding and according to the manufacturer's protocol. The amounts of DNA used for transfection are further specified in the sections of the different assays. For AAV productions, 293AAV cells were used only until reaching passage 10.

AAV crude lysate production

293AAV cells were seeded into six-well plates at a density of 500,000 cells per well. The next day, cells were transfected with 2.5 μg of DNA using PEI and an equimolar ratio of the plasmids necessary for the respective AAV production. Three days post transfection, cells were harvested by flushing off the cells by pipetting. Cells were washed with PBS once and then subjected to five freeze and thaw cycles by alternating between liquid nitrogen and a 37°C water bath. Cell debris was pelleted by centrifugation at 18,000 rpm at 4°C for 10 min and the supernatant containing the AAV particles was stored at −20°C until use.

Purified AAV production

Large-scale productions of AAV were either done by the University of Minnesota Viral Vector and Cloning Core using a sucrose gradient or in lab following published iodixanol gradient density protocols.[95,96] In brief, 4 million 293AAV cells were seeded into 15-cm dishes and transfected using PEI and 47 μg of total DNA per dish 48 h post seeding. For the AAV-DJ library control, an equimolar triple transfection was used composed of an Adeno-helper plasmid, a plasmid encoding the *rep2* and *capDJ* genes, and a transgene plasmid encoding an EFS promoter-driven miRFP670nano sequence flanked by ITRs. For AAV-DJ insertion library productions, a plasmid ratio of 1:0.1:0.1 of an adeno-helper plasmid, a plasmid encoding rep2 and only VP1 of cap-DJ, as well as the respective AAV-DJ insertion library in which the AAV-DJ silent mutation variant was spiked in was transfected. To top up to 47 μg of total DNA, a pUC19 stuffer plasmid was added. At 72 h post transfection, cells were detached with a cell lifter and cells pelleted by centrifugation at 400 × *g* for 15 min. The cell pellet was washed once with PBS and resuspended in a buffer containing 2 mM $MgCl_2$, 0.15 M NaCl, and 50 mM Tris-HCl at pH 8.5. Cells were cracked open using five freeze and thaw cycles. Free genomic and plasmid DNA was digested with a Benzonase Nuclease (Sigma-Aldrich). Cell debris was removed by centrifugation and, subsequently, the lysate transferred into ultracentrifugation tubes (Beckman Coulter). The iodixanol discontinuous gradient (15%, 25%, 40%, and 60% iodixanol concentration) was layered underneath the cell lysate. Density gradient centrifugation was done at 50,000 rpm for 2 h at 4°C using a 70.1 Ti rotor (Beckman Coulter). Post centrifugation, the 40% iodixanol phase containing the AAV was isolated, aliquoted, and stored at −80°C until use. For pull-down, binding, and uptake, as well as the differential scanning fluorimetry (DSF) assays, AAV samples were dialyzed to PBS supplemented with 5% glycerol using 10-kDa Amicon Ultra-15 Centrifugal Filter Units (Millipore, Sigma). Certain assays required preparation of purified virus: for the pull-down assay, we observed an increased stickiness to the beads in iodixanol preps compared to preps in PBS. For the binding and internalization assay, we decided to use PBS as buffer, because we think that the iodixanol slows down the settlement of AAV particles and consequently the attachment of the particles to the cells. For the binding and internalization assays, the settle time matters as the AAV particles are only incubated for 1 h at 4°C before unbound particles are washed away. For the DSF, a fluorescent dye is used and the iodixanol quenches the fluorescent signal.

qPCR

To determine the titer of crude lysate AAV samples, 1–5 μL of the crude lysate were mixed with PBS supplemented with 2 mM $MgCl_2$ to a final volume of 50 μL. Then, 0.1 μL of ultrapure Benzonase Nuclease (Sigma-Aldrich) was added. Samples were incubated for 30 min at 37°C to digest non-encapsidated DNA. Next, 5 μL of a 10× Proteinase K buffer (100 nM Tris-HCl, pH 8.0, 10 mM EDTA, and 10% SDS) and 1 μL of Proteinase K (20 mg/mL; Zymo Research) were added to stop the DNA digest and start the protein digest to free the ssDNA from the AAV particles. Samples were incubated for 20 min at 50°C, followed by a heat inactivation of the enzymes for 5 min at 95°C. The DNA was purified using the DNA Clean & Concentrator-5 kit (Zymo Research) according to the manufacturer's instructions for ssDNA purification. For purified AAV samples, the Benzonase digest step was skipped and only the Proteinase K and DNA purification steps were performed. All samples were diluted 1:1,000 in H_2O prior to qPCR. The viral genome (vg) quantification was done on a QuantStudio5 Real-Time PCR System (Applied Biosystems) using the PowerUp SYBR Green Master Mix (Applied Biosystems) and following the manufacturer's instructions. To calculate the viral titer (vg/mL), a plasmid standard with a known concentration of plasmid copies was used. Primer sets binding either the cytomegalovirus (CMV) enhancer or the p40 promoter of the AAV genomes, as well as in the plasmid standard, were selected (Table S3).

Pull-down assay of libraries

Between 1×10^9 and 1×10^{10} vg were used as input material to bind to different magnetic beads for pull-down assays: SNAP-Capture Magnetic Beads (NEB) for SNAP tag insertion; Pierce Anti-DYKDDDDK Magnetic Agarose (Thermo Scientific) for FLAG tag insertion; and Streptavidin Magnetic Beads (NEB) for nanobody, SpyCatcher, and HUH tag insertions. The 80-μL bead slurry for SNAP pull-downs and 50-μL bead slurry for FLAG pull-downs were washed three times with 300 μL of wash buffer I (0.15 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA). Then, AAV libraries were mixed with wash buffer I and the beads to a final volume of 300 μL before incubated on a slow shaker for 30 min at room temperature. Beads were washed again three times with 300 μL of wash buffer I for SNAP beads and with PBS (pH 7.4, Gibco) for FLAG beads to remove unbound AAV particles and finally resuspended in 50 μL of PBS. For pull-down assays with streptavidin beads, 100 μL of bead slurry was washed three times with 300 μL of wash buffer II (0.15 M NaCl, 20 mM Tris-HCl pH 7.5, 1 mM EDTA). For nanobody and SpyCatcher binding, beads were pre-incubated with either 320 pmol of biotinylated superfolder-GFP or 1,000 pmol of biotinylated SpyTag, respectively, for 30 min on a slow shaker at room temperature in a total volume of 300 μL in wash buffer II. Afterward, unbound superfolder-GFP and SpyTag were removed by washing the beads three times with 300 μL of wash buffer II. For HUH tag pull-downs, AAV libraries were first reacted with 1 nmol of biotinylated ssDNA oligos (sequences are given in Table S3) in PBS supplemented with 0.05% v/v salmon sperm DNA (Invitrogen), 1 mM $MgCl_2$, and 1 mM $MnCl_2$ for 15 min at 37°C. Next, beads were incubated with AAV libraries for 30 min on a slow shaker at room temperature, before unbound AAV particles were removed by washing three times with 300 μL wash buffer II. Beads with bound AAV particles were resuspended in 50 μL of PBS. Viral genomes of samples after the pull-down were purified using the Quick-DNA Microprep Plus Kit (Zymo research) and by following the manufacturer's protocol.

Binding and uptake assays of libraries

The binding and uptake assays were performed as previously described[65] but using HEK293FT cells. In brief, 375,000 HEK293FT cells were seeded into six-well plates using 2 mL of culturing medium per well. The next day, cells were incubated for 30 min at 4°C. Afterward, the media was aspirated and 200 μL of cold DMEM containing AAV particles at an MOI of 1e4 vg/cell added. The cells were further incubated for 1 h at 4°C. Next, cells were washed three times with ice-cold PBS to eliminate unbound AAV particles. For the binding assay, 150 μL of ice-cold PBS was added, the cells detached with a cell scraper, and the cell suspension transferred to a microcentrifuge tube. For the uptake assay, 1 mL of pre-warmed culturing medium was added immediately after the PBS wash and the cells incubated for 2 h in a cell culture incubator to allow for uptake of the AAV particles. Then, the cells were detached by trypsinization and collected in a microcentrifuge tube before being washed three times with 200 μL of PBS. The DNA from binding and uptake samples was purified using the Quick-DNA Microprep Plus Kit (Zymo research) and by following the manufacturer's protocol.

Infectivity assay of libraries

Here, 150,000 HEK293FT cells were seeded into 12-well plates using 1 mL of culturing medium per well. The next day, cells were transduced with purified AAV (in PBS supplemented with 5% glycerol) at an MOI of 2×10^5 vg/cell. To this end, the medium was aspirated and cells were washed with 500 μL of PBS once. Purified AAVs were mixed with DMEM without supplements to a final volume of 500 μL and added onto the cells. After 2 h of incubation in the cell culture incubator, 1.5 mL of culturing medium (with supplements) was added. At 24 h post transduction, the temperature was reduced to 33°C to promote protein expression rather than cell growth.[97] At 72 h post transduction, cells were prepared for cell sorting as follows. Medium was aspirated, cells were washed with 500 μL of PBS, and 500 μL of Accutase solution (Sigma-Aldrich) was added and incubated at room temperature until all cells detached. The cell suspension was transferred to a microcentrifuge tube and centrifuged for 3 min at 400 × *g* to pellet cells. Cells were washed two times with 500 μL of PBS, before being resuspended in 650 μL of cell sorting buffer (PBS supplemented with 5 mM EDTA and 2.5% FBS) and passed through a 35-μm cell strainer to avoid cell clumps. Cell sorting was performed by the University of Minnesota Flow Cytometry Resource (UFCR) on a FACS Aria II instrument (BD Biosciences) with an 85-μm nozzle by sorting miRFP670nano positive (excitation 640-nm laser, emission 670/30-nm bandpass filter). Post sorting, the DNA was extracted from the cells using the Quick-DNA Microprep Plus Kit (Zymo research) and by following the manufacturer's protocol.

NGS preparation

Purified DNA samples (library plasmid DNA, ssDNA from purified AAV, and DNA extracted after pull-down, binding, uptake, and infectivity assays) were amplified using primers binding 50 base pairs up- and downstream of the VP1 coding sequence (Table S3). For amplification, the PrimeSTAR Max DNA Polymerase (Takara Bio) was used according to the manufacturer's recommendation with a 25-μL reaction volume, an annealing temperature of 62°C, and an elongation time of 15 s. At least five reactions were pooled for each DNA sample, whereas the cycle number was kept at a minimum to obtain >50 ng per sample. PCR products were purified using the DNA Clean & Concentrator-5 kit (Zymo Research) and by following the manufacturer's instructions for dsDNA purification. The DNA was eluted in a 10 mM Tris buffer with pH 8.0. Prior to sequencing, the DNA was quantified using the Qubit 1× dsDNA HS assay kit and a Qubit 4 Fluorometer (both Invitrogen), and the DNA of the insertion libraries from the same assay were pooled in an equimolar ratio. Then, >50 ng of each DNA pool was submitted to the University of Minnesota Genomics Center, where Nextera XT libraries were created, and samples sequenced using a NovaSeq SPrime 150 paired-end run.

Sequencing data analysis and enrichment calculation

Forward and reverse reads were aligned individually using a DIP-seq pipeline,[73] slightly modified for SPINE compatibility and for updated python packages. The code for handling data from domain insertion library sequencing is available at: https://github.com/SavageLab/dipseq. This pipeline results in.csv spreadsheets (available as processed data, along scripts to reproduce manuscript figures, which are available at https://github/com/Schmidt-lab/AAV_Insertion_Profiling), indicating insertion position, direction, and whether it was in frame. Fitness was calculated from the frequency of a given VP1 variant (*i*) after packaging (*s*) relative to the frequency of that variant in the input library (*u*), normalized to wild-type AAV (*wt*):

$$W^i = \ln\left(\frac{f_s^i}{f_u^i} * \frac{f_u^{wt}}{f_s^{wt}}\right)$$

Fitness standard error for each variant was calculated assuming a Poisson distribution.

$$SE = \sqrt{\frac{1}{Count_{i,s}+0.5}+\frac{1}{Count_{i,u}+0.5}+\frac{1}{Count_{wt,s}+0.5}+\frac{1}{Count_{wt,u}+0.5}}$$

Mapping of fitness data onto the AAV capsid structure

Fitness data were mapped onto the structure of AAV-DJ (PDB: 7KFR) using Pymol. VP1u, which is not resolved in the cryoelectron microscopy (cryo-EM) structure of AAV-DJ, was modeled using RoseTTAFold.[98] The highest-scored predicted VP1u structure was manually positioned in Pymol adjusting peptide-backbone bond dihedral angles to avoid clashes with the experimentally determined capsid structure.

Western blot

For VP protein analysis 3×10^9 to 1×10^{10} vg of purified AAV were mixed with 12.5 μL of 4× Laemmli Sample Buffer (Bio-Rad, supplemented with 10% 2-mercaptoethanol) and topped up to a final volume of 50 μL with PBS. Samples were denatured for 10 min at 95°C and afterward chilled on ice. Protein samples, and 5 μL of the Precision Plus Protein Dual Color Standard (Bio-Rad), were separated by molecular weight on a 7.5% precast polyacrylamide gel (Bio-Rad) in Tris/Glycine/SDS Electrophoresis Buffer (Bio-Rad) for 85 min at 120 V. Next, proteins were transferred to a nitrocellulose membrane (pore size 0.45 μm; Thermo Scientific) in an ice-cold blotting buffer (25 mM Tris base, 96 mM glycine, 20% methanol) for 80 min at 110 V. The membrane was washed once in TBS-T (20 mM Tris base, 137 mM NaCl, pH 7.6, 0.05% Tween 20) and incubated in 5% skim milk solution in TBS-T for 2 h at room temperature on a slow shaker to block non-specific binding. A primary antibody detecting either all three VP proteins (anti-AAV VP1/VP2/VP3 mouse monoclonal, B1, supernatant, Progen) or only VP1 (anti-AAV VP1 mouse monoclonal, A, lyophilized, purified, Progen) was diluted 1:250 in 5% skim milk solution in TBS-T and incubated overnight at 4°C. The next day, the membrane was washed four times for 5 min in TBS-T on a shaker before the secondary antibody was added (anti-mouse immunoglobulin [Ig] G-peroxidase antibody produced in goat, Sigma-Aldrich). The secondary antibody was diluted 1:50,000 in 5% skim milk solution in TBS-T and incubated for 2 h at room temperature on a slow shaker. Afterward, the membrane was washed again four times for 5 min in TBS-T at room temperature to remove unbound antibodies before the SuperSignal West Dura Extended Duration Substrate kit solution (Thermo Scientific) was applied and incubated for 2 min at room temperature. The chemiluminescence signal was detected with an Amersham Imager 600 (GE Healthcare) using exposure times between 1 s and 10 min, depending on the signal intensities. Quantification of VP expression was done using ImageJ.

DSF

First, 5,000× SYPRO Orange dye (Invitrogen) was diluted 1:100 in PBS with 5% glycerol. Then, each sample was prepared by mixing 50× SYPRO Orange dye 1:10 with $>5 \times 10^9$ vg of purified AAV samples in PBS with 5% glycerol to a final volume of 25 or 50 μL. Samples were mixed by pipetting up and down and into a 0.1-mL MicroAmp Fast Optical 96-Well Reaction Plate (Applied Biosystems). The plate was sealed with an optical adhesive cover (Applied Biosystems) and spun down for 2 min at 1,000 × *g*. A melt-curve experiment was run on a QuantStudio5 Real-Time PCR instrument (Applied Biosystems) using the x1-m4 filter set (excitation filter, 470/15 nm; emission filter, 623/14nm) and the following settings: 30°C for 2 min, temperature increase from 30°C to 99°C in 0.5°C and 30-s increments, and a final incubation step of 2 min at 99°C. Lysozyme at a concentration of 0.1 mg/mL with a determined melting temperature of 70°C was used as a reference control within each run. Post-processing of the melt-curve data was done in MATLAB R2021a (The Mathworks). A smoothing spline (smoothing parameter p = 0.9) was fitted to the data before calculating the numerical gradient $\partial(\text{Fluorescence})/\partial(\text{Temperature})$.

Transmission electron microscopy

To quantify the empty to full capsid ratio, negative staining and transmission electron microscopy was performed by the Characterization Facility, University of Minnesota. At least 300 AAV particles were manually counted per sample.

Infectivity assay of cysteine mutants and WDV variants

Here, 75,000 HEK293FT cells were seeded per well of a 24-well plate using 0.5 mL of culturing medium. If applicable, cells were transfected with 100 ng of GFP-GPI plasmid while seeding. The next day, medium was exchanged and cells transduced with crude lysates or purified AAV at the indicated MOIs. At 48 h post transduction, cells were prepared for flow cytometry as follows. Medium was aspirated, cells were washed with 500 μL of PBS, 250 μL Accutase solution (Sigma-Aldrich) was added, and they were incubated at room temperature until all cells detached. Cell suspension was transferred into a microcentrifuge tube and centrifuged for 3 min at 400 × *g* to pellet cells. Cells were washed two times with 300 μL of PBS before being resuspended in 600 μL of flow cytometry buffer (PBS supplemented with 5 mM EDTA and 2.5% FBS) and passed through a 35-μm cell strainer to avoid cell clumps. Flow cytometry was performed either on a LSRFortessa X-20 or a FACSymphony A3 Cell Analyzer (both BD Biosciences) equipped with 561- and 488-nm lasers to detect tdTomato- and GFP-positive cells, respectively. A minimum of 10,000 single cell events were recorded per sample. Data analysis was performed using the FlowJo 10.8.0 software (BD Biosciences).

DATA AND CODE AVAILABILITY

Library staging and accessory plasmids (including sequence information) are available upon request. Sequencing data generated in this study have been deposited in the Sequence Raw Archive (https://www.ncbi.nlm.nih.gov/sra) under accession code PRJNA950466; refer to for corresponding metadata and read statistics. The code for handling data from domain insertion library sequencing is available at https://github.com/SavageLab/dipseq. The SPINE code is available at https://github.com/schmidt-lab/SPINE. The version used for this study is archived at https://zenodo.org/badge/latestdoi/223953195. Processed read count data, along with R scripts to reproduce manuscript figures, are available at https://github.com/Schmidt-lab/AAV_Insertion_Profiling. A Shiny app to display 3D views of AAV insertion fitness maps in a web browser is available at https://github.com/Schmidt-lab/shinyAAViewerR.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2023.101143.

ACKNOWLEDGMENTS

The project described was supported by a Career Development Award from the American Society of Gene & Cell Therapy (M.D.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the American Society of Gene & Cell Therapy. This work was supported by the National Institutes of Health (R01GM141152 to D.S.) and (P30DA048742 to D.S.). Some AAV productions used in this study were generated by the University of Minnesota Viral Vector and Cloning Core (Minneapolis, MN). We thank the University of Minnesota Genomics Center for providing NGS service and technical support. This work was also supported by the resources and staff at the University of Minnesota University Imaging Centers (UIC). The UIC RRID is SCR_020997. Transmission electron microscopy was carried out in the Characterization Facility, University of Minnesota, which receives partial support from the NSF through the MRSEC (award number DMR-2011401) and the NNCI (award number ECCS-2025124) programs.

AUTHOR CONTRIBUTIONS

M.D.H., A.C.Z., and D.S. designed the study with input from W.R.G. and G.A. M.D.H., A.C.Z., D.N., and D.S. conducted the experiments with assistance from Y.H. M.D.H. and D.S. analyzed the data and authored the manuscript. All authors have given approval to the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Russell, S., Bennett, J., Wellman, J.A., Chung, D.C., Yu, Z.F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., et al. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. Lancet *390*, 849–860.
2. Hoy, S.M. (2019). Onasemnogene Abeparvovec: First Global Approval. Drugs *79*, 1255–1262.
3. Ylä-Herttuala, S. (2012). Endgame: glybera finally recommended for approval as the first gene therapy drug in the European union. Mol. Ther. *20*, 1831–1832.
4. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat. Rev. Drug Discov. *18*, 358–378.
5. Li, C., and Samulski, R.J. (2020). Engineering adeno-associated virus vectors for gene therapy. Nat. Rev. Genet. *21*, 255–272.
6. Jay, F.T., Laughlin, C.A., and Carter, B.J. (1981). Eukaryotic translational control: adeno-associated virus protein synthesis is affected by a mutation in the adenovirus DNA-binding protein. Proc. Natl. Acad. Sci. USA *78*, 2927–2931.
7. Sonntag, F., Schmidt, K., and Kleinschmidt, J.A. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc. Natl. Acad. Sci. USA *107*, 10220–10225.
8. Ogden, P.J., Kelsic, E.D., Sinai, S., and Church, G.M. (2019). Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. Science *366*, 1139–1143.
9. Mateu, M.G. (2013). Assembly, stability and dynamics of virus capsids. Arch. Biochem. Biophys. *531*, 65–79.
10. Medrano, M., Fuertes, M.Á., Valbuena, A., Carrillo, P.J.P., Rodríguez-Huete, A., and Mateu, M.G. (2016). Imaging and Quantitation of a Succession of Transient Intermediates Reveal the Reversible Self-Assembly Pathway of a Simple Icosahedral Virus Capsid. J. Am. Chem. Soc. *138*, 15385–15396.
11. Wörner, T.P., Bennett, A., Habka, S., Snijder, J., Friese, O., Powers, T., Agbandje-McKenna, M., and Heck, A.J.R. (2021). Adeno-associated virus capsid assembly is divergent and stochastic. Nat. Commun. *12*, 1642.
12. Buller, R.M., and Rose, J.A. (1978). Characterization of adenovirus-associated virus-induced polypeptides in KB cells. J. Virol. *25*, 331–338.
13. Johnson, F.B., Ozer, H.L., and Hoggan, M.D. (1971). Structural proteins of adenovirus-associated virus type 3. J. Virol. *8*, 860–863.
14. Rose, J.A., Maizel, J.V., Inman, J.K., and Shatkin, A.J. (1971). Structural proteins of adenovirus-associated viruses. J. Virol. *8*, 766–770.
15. Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A.J.R. (2014). Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. J. Am. Chem. Soc. *136*, 7295–7299.
16. Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M.S. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc. Natl. Acad. Sci. USA *99*, 10405–10410.
17. Govindasamy, L., Padron, E., McKenna, R., Muzyczka, N., Kaludov, N., Chiorini, J.A., and Agbandje-McKenna, M. (2006). Structurally mapping the diverse phenotype of adeno-associated virus serotype 4. J. Virol. *80*, 11556–11570.
18. Lerch, T.F., O'Donnell, J.K., Meyer, N.L., Xie, Q., Taylor, K.A., Stagg, S.M., and Chapman, M.S. (2012). Structure of AAV-DJ, a retargeted gene therapy vector: cryo-electron microscopy at 4.5 Å resolution. Structure *20*, 1310–1320.
19. Meyer, N.L., and Chapman, M.S. (2022). Adeno-associated virus (AAV) cell entry: structural insights. Trends Microbiol. *30*, 432–451.
20. Pillay, S., Meyer, N.L., Puschnik, A.S., Davulcu, O., Diep, J., Ishikawa, Y., Jae, L.T., Wosen, J.E., Nagamine, C.M., Chapman, M.S., and Carette, J.E. (2016). An essential receptor for adeno-associated virus infection. Nature *530*, 108–112.
21. Meyer, N.L., Hu, G., Davulcu, O., Xie, Q., Noble, A.J., Yoshioka, C., Gingerich, D.S., Trzynka, A., David, L., Stagg, S.M., and Chapman, M.S. (2019). Structure of the gene therapy vector, adeno-associated virus with its cell receptor. Elife *8*, e44707.
22. Riyad, J.M., and Weber, T. (2021). Intracellular trafficking of adeno-associated virus (AAV) vectors: challenges and future directions. Gene Ther. *28*, 683–696.
23. Bleker, S., Sonntag, F., and Kleinschmidt, J.A. (2005). Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity. J. Virol. *79*, 2528–2540.
24. Kronenberg, S., Böttcher, B., von der Lieth, C.W., Bleker, S., and Kleinschmidt, J.A. (2005). A conformational change in the adeno-associated virus type 2 capsid leads to the exposure of hidden VP1 N termini. J. Virol. *79*, 5296–5303.
25. Nam, H.J., Gurda, B.L., McKenna, R., Potter, M., Byrne, B., Salganik, M., Muzyczka, N., and Agbandje-McKenna, M. (2011). Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking. J. Virol. *85*, 11791–11799.
26. Venkatakrishnan, B., Yarbrough, J., Domsic, J., Bennett, A., Bothner, B., Kozyreva, O.G., Samulski, R.J., Muzyczka, N., McKenna, R., and Agbandje-McKenna, M. (2013). Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. J. Virol. *87*, 4974–4984.
27. Cotmore, S.F., and Tattersall, P. (2014). Parvoviruses: Small Does Not Mean Simple. Annu. Rev. Virol. *1*, 517–537.
28. Zádori, Z., Szelei, J., Lacoste, M.C., Li, Y., Gariépy, S., Raymond, P., Allaire, M., Nabi, I.R., and Tijssen, P. (2001). A viral phospholipase A2 is required for parvovirus infectivity. Dev. Cell *1*, 291–302.
29. Grieger, J.C., Snowdy, S., and Samulski, R.J. (2006). Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. J. Virol. *80*, 5199–5210.
30. Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). *In vitro* and *in vivo* gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol. *82*, 5887–5911.
31. Zinn, E., Pacouret, S., Khaychuk, V., Turunen, H.T., Carvalho, L.S., Andres-Mateos, E., Shah, S., Shelke, R., Maurer, A.C., Plovie, E., et al. (2015). In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. Cell Rep. *12*, 1056–1068.
32. Chan, K.Y., Jang, M.J., Yoo, B.B., Greenbaum, A., Ravi, N., Wu, W.L., Sánchez-Guardado, L., Lois, C., Mazmanian, S.K., Deverman, B.E., and Gradinaru, V. (2017). Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. *20*, 1172–1179.
33. Dalkara, D., Byrne, L.C., Klimczak, R.R., Visel, M., Yin, L., Merigan, W.H., Flannery, J.G., and Schaffer, D.V. (2013). In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. Sci. Transl. Med. *5*, 189ra76.
34. Warrington, K.H., Gorbatyuk, O.S., Harrison, J.K., Opie, S.R., Zolotukhin, S., and Muzyczka, N. (2004). Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. J. Virol. *78*, 6595–6609.

35. Asokan, A., Johnson, J.S., Li, C., and Samulski, R.J. (2008). Bioluminescent virion shells: new tools for quantitation of AAV vector dynamics in cells and live animals. Gene Ther. *15*, 1618–1622.

36. Münch, R.C., Janicki, H., Völker, I., Rasbach, A., Hallek, M., Büning, H., and Buchholz, C.J. (2013). Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer. Mol. Ther. *21*, 109–118.

37. Eichhoff, A.M., Börner, K., Albrecht, B., Schäfer, W., Baum, N., Haag, F., Körbelin, J., Trepel, M., Braren, I., Grimm, D., et al. (2019). Nanobody-Enhanced Targeting of AAV Gene Therapy Vectors. Mol. Ther. Methods Clin. Dev. *15*, 211–220.

38. Zdechlik, A.C., He, Y., Aird, E.J., Gordon, W.R., and Schmidt, D. (2020). Programmable Assembly of Adeno-Associated Virus-Antibody Composites for Receptor-Mediated Gene Delivery. Bioconjugate Chem. *31*, 1093–1106.

39. Michels, A., Frank, A.M., Günther, D.M., Mataei, M., Börner, K., Grimm, D., Hartmann, J., and Buchholz, C.J. (2021). Lentiviral and adeno-associated vectors efficiently transduce mouse T lymphocytes when targeted to murine CD8. Mol. Ther. Methods Clin. Dev. *23*, 334–347.

40. Judd, J., Wei, F., Nguyen, P.Q., Tartaglia, L.J., Agbandje-McKenna, M., Silberg, J.J., and Suh, J. (2012). Random Insertion of mCherry Into VP3 Domain of Adeno-associated Virus Yields Fluorescent Capsids With no Loss of Infectivity. Mol. Ther. Nucleic Acids *1*, e54.

41. Koerber, J.T., Jang, J.H., Yu, J.H., Kane, R.S., and Schaffer, D.V. (2007). Engineering adeno-associated virus for one-step purification via immobilized metal affinity chromatography. Hum. Gene Ther. *18*, 367–378.

42. Bryant, D.H., Bashir, A., Sinai, S., Jain, N.K., Ogden, P.J., Riley, P.F., Church, G.M., Colwell, L.J., and Kelsic, E.D. (2021). Deep diversification of an AAV capsid protein by machine learning. Nat. Biotechnol. *39*, 691–696.

43. Ravindra Kumar, S., Miles, T.F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., et al. (2020). Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods *17*, 541–550.

44. Coyote-Maestas, W., He, Y., Myers, C.L., and Schmidt, D. (2019). Domain insertion permissibility-guided engineering of allostery in ion channels. Nat. Commun. *10*, 290.

45. Coyote-Maestas, W., Nedrud, D., Suma, A., He, Y., Matreyek, K.A., Fowler, D.M., Carnevale, V., Myers, C.L., and Schmidt, D. (2021). Probing ion channel functional architecture and domain recombination compatibility by massively parallel domain insertion profiling. Nat. Commun. *12*, 7114.

46. Coyote-Maestas, W., Nedrud, D., Okorafor, S., He, Y., and Schmidt, D. (2020). Targeted insertional mutagenesis libraries for deep domain insertion profiling. Nucleic Acids Res. *48*, 1010.

47. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., and Tawfik, D.S. (2006). Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. Nature *444*, 929–932.

48. Girod, A., Wobus, C.E., Zádori, Z., Ried, M., Leike, K., Tijssen, P., Kleinschmidt, J.A., and Hallek, M. (2002). The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J. Gen. Virol. *83*, 973–978.

49. Perabo, L., Büning, H., Kofler, D.M., Ried, M.U., Girod, A., Wendtner, C.M., Enssle, J., and Hallek, M. (2003). In vitro selection of viral vectors with modified tropism: the adeno-associated virus display. Mol. Ther. *8*, 151–157.

50. Varadi, K., Michelfelder, S., Korff, T., Hecker, M., Trepel, M., Katus, H.A., Kleinschmidt, J.A., and Müller, O.J. (2012). Novel random peptide libraries displayed on AAV serotype 9 for selection of endothelial cell-directed gene transfer vectors. Gene Ther. *19*, 800–809.

51. Waterkamp, D.A., Müller, O.J., Ying, Y., Trepel, M., and Kleinschmidt, J.A. (2006). Isolation of targeted AAV2 vectors from novel virus display libraries. J. Gene Med. *8*, 1307–1319.

52. Rothbauer, U., Zolghadr, K., Tillib, S., Nowak, D., Schermelleh, L., Gahl, A., Backmann, N., Conrath, K., Muyldermans, S., Cardoso, M.C., and Leonhardt, H. (2006). Targeting and tracing antigens in live cells with fluorescent nanobodies. Nat. Methods *3*, 887–889.

53. Juillerat, A., Gronemeyer, T., Keppler, A., Gendreizig, S., Pick, H., Vogel, H., and Johnsson, K. (2003). Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules *in vivo*. Chem. Biol. *10*, 313–317.

54. Zakeri, B., Fierer, J.O., Celik, E., Chittock, E.C., Schwarz-Linek, U., Moy, V.T., and Howarth, M. (2012). Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. Proc. Natl. Acad. Sci. USA *109*, E690–E697.

55. Everett, B.A., Litzau, L.A., Tompkins, K., Shi, K., Nelson, A., Aihara, H., Evans Iii, R.L., and Gordon, W.R. (2019). Crystal structure of the Wheat dwarf virus Rep domain. Acta Crystallogr. F Struct. Biol. Commun. *75*, 744–749.

56. Lovendahl, K.N., Hayward, A.N., and Gordon, W.R. (2017). Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. J. Am. Chem. Soc. *139*, 7030–7035.

57. Smiley, A.T., Tompkins, K.J., Pawlak, M.R., Krueger, A.J., Evans, R.L., Shi, K., Aihara, H., and Gordon, W.R. (2023). Watson-Crick Base-Pairing Requirements for ssDNA Recognition and Processing in Replication-Initiating HUH Endonucleases. mBio *14*, e0258722.

58. Tompkins, K.J., Houtti, M., Litzau, L.A., Aird, E.J., Everett, B.A., Nelson, A.T., Pornschloegl, L., Limón-Swanson, L.K., Evans, R.L., Evans, K., et al. (2021). Molecular underpinnings of ssDNA specificity by Rep HUH-endonucleases and implications for HUH-tag multiplexing and engineering. Nucleic Acids Res. *49*, 1046–1064.

59. Fowler, D.M., and Fields, S. (2014). Deep mutational scanning: a new style of protein science. Nat. Methods *11*, 801–807.

60. Schmit, P.F., Pacouret, S., Zinn, E., Telford, E., Nicolaou, F., Broucque, F., Andres-Mateos, E., Xiao, R., Penaud-Budloo, M., Bouzelha, M., et al. (2020). Cross-Packaging and Capsid Mosaic Formation in Multiplexed AAV Libraries. Mol. Ther. Methods Clin. Dev. *17*, 107–121.

61. King, J.A., Dubielzig, R., Grimm, D., and Kleinschmidt, J.A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. EMBO J. *20*, 3282–3291.

62. Oliinyk, O.S., Shemetov, A.A., Pletnev, S., Shcherbakova, D.M., and Verkhusha, V.V. (2019). Smallest near-infrared fluorescent protein evolved from cyanobacteriochrome as versatile tag for spectral multiplexing. Nat. Commun. *10*, 279.

63. Schmiedel, J.M., and Lehner, B. (2019). Determining protein structures using deep mutagenesis. Nat. Genet. *51*, 1177–1186.

64. Stagg, S.M., Yoshioka, C., Davulcu, O., and Chapman, M.S. (2022). Cryo-electron Microscopy of Adeno-associated Virus. Chem. Rev. *122*, 14018–14054.

65. Berry, G.E., and Tse, L.V. (2017). Virus Binding and Internalization Assay for Adeno-associated Virus. Bio. Protoc. *7*, e2110.

66. Sonntag, F., Bleker, S., Leuchs, B., Fischer, R., and Kleinschmidt, J.A. (2006). Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. J. Virol. *80*, 11040–11054.

67. Gerlach, B., Kleinschmidt, J.A., and Böttcher, B. (2011). Conformational changes in adeno-associated virus type 1 induced by genome packaging. J. Mol. Biol. *409*, 427–438.

68. Horowitz, E.D., Finn, M.G., and Asokan, A. (2012). Tyrosine cross-linking reveals interfacial dynamics in adeno-associated viral capsids during infection. ACS Chem. Biol. *7*, 1059–1066.

69. McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.

70. Wu, P., Xiao, W., Conlon, T., Hughes, J., Agbandje-McKenna, M., Ferkol, T., Flotte, T., and Muzyczka, N. (2000). Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. J. Virol. *74*, 8635–8647.

71. Hagen, S., Baumann, T., Wagner, H.J., Morath, V., Kaufmann, B., Fischer, A., Bergmann, S., Schindler, P., Arndt, K.M., and Müller, K.M. (2014). Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed enzyme prodrug therapy. Sci. Rep. *4*, 3759.

72. Becker, J., Fakhiri, J., and Grimm, D. (2022). Fantastic AAV Gene Therapy Vectors and How to Find Them-Random Diversification, Rational Design and Machine Learning. Pathogens *11*, 756.

73. Nadler, D.C., Morgan, S.A., Flamholz, A., Kortright, K.E., and Savage, D.F. (2016). Rapid construction of metabolite biosensors using domain-insertion profiling. Nat. Commun. *7*, 12266.

74. Plesa, C., Sidore, A.M., Lubock, N.B., Zhang, D., and Kosuri, S. (2018). Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. Science *359*, 343–347.

75. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol. *18*, 150.

76. Yang, K.K., Wu, Z., and Arnold, F.H. (2019). Machine-learning-guided directed evolution for protein engineering. Nat. Methods *16*, 687–694.

77. DiPrimio, N., Asokan, A., Govindasamy, L., Agbandje-McKenna, M., and Samulski, R.J. (2008). Surface loop dynamics in adeno-associated virus capsid assembly. J. Virol. *82*, 5178–5189.

78. Wobus, C.E., Hügle-Dörr, B., Girod, A., Petersen, G., Hallek, M., and Kleinschmidt, J.A. (2000). Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. J. Virol. *74*, 9281–9293.

79. Cotmore, S.F., McKie, V.C., Anderson, L.J., Astell, C.R., and Tattersall, P. (1986). Identification of the major structural and nonstructural proteins encoded by human parvovirus B19 and mapping of their genes by procaryotic expression of isolated genomic fragments. J. Virol. *60*, 548–557.

80. Kaufmann, B., Simpson, A.A., and Rossmann, M.G. (2004). The structure of human parvovirus B19. Proc. Natl. Acad. Sci. USA *101*, 11628–11633.

81. Kawase, M., Momoeda, M., Young, N.S., and Kajigaya, S. (1995). Most of the VP1 unique region of B19 parvovirus is on the capsid surface. Virology *211*, 359–366.

82. Ros, C., Gerber, M., and Kempf, C. (2006). Conformational changes in the VP1-unique region of native human parvovirus B19 lead to exposure of internal sequences that play a role in virus neutralization and infectivity. J. Virol. *80*, 12017–12024.

83. Wang, Q., Wu, Z., Zhang, J., Firrman, J., Wei, H., Zhuang, Z., Liu, L., Miao, L., Hu, Y., Li, D., et al. (2017). A Robust System for Production of Superabundant VP1 Recombinant AAV Vectors. Mol. Ther. Methods Clin. Dev. *7*, 146–156.

84. McLaughlin, R.N., Poelwijk, F.J., Raman, A., Gosal, W.S., and Ranganathan, R. (2012). The spatial architecture of protein function and adaptation. Nature *491*, 138–142.

85. Rivoire, O., Reynolds, K.A., and Ranganathan, R. (2016). Evolution-Based Functional Decomposition of Proteins. PLoS Comput. Biol. *12*, e1004817.

86. Atkinson, J.T., Jones, A.M., Zhou, Q., and Silberg, J.J. (2018). Circular permutation profiling by deep sequencing libraries created using transposon mutagenesis. Nucleic Acids Res. *46*, e76.

87. Markin, C.J., Mokhtari, D.A., Sunden, F., Appel, M.J., Akiva, E., Longwell, S.A., Sabatti, C., Herschlag, D., and Fordyce, P.M. (2021). Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics. Science *373*, eabf8761.

88. Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.L., Yang, B., Huber, N., Pasca, S.P., and Gradinaru, V. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. *34*, 204–209.

89. Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deléage, G., and Hallek, M. (1999). Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat. Med. *5*, 1438.

90. Michelfelder, S., Lee, M.K., deLima-Hahn, E., Wilmes, T., Kaul, F., Müller, O., Kleinschmidt, J.A., and Trepel, M. (2007). Vectors selected from adeno-associated viral display peptide libraries for leukemia cell-targeted cytotoxic gene therapy. Exp. Hematol. *35*, 1766–1776.

91. Müller, O.J., Kaul, F., Weitzman, M.D., Pasqualini, R., Arap, W., Kleinschmidt, J.A., and Trepel, M. (2003). Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. Nat. Biotechnol. *21*, 1040–1046.

92. Münch, R.C., Muth, A., Muik, A., Friedel, T., Schmatz, J., Dreier, B., Trkola, A., Plückthun, A., Büning, H., and Buchholz, C.J. (2015). Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors. Nat. Commun. *6*, 6246.

93. Large, E.E., and Chapman, M.S. (2023). Adeno-associated virus receptor complexes and implications for adeno-associated virus immune neutralization. Front. Microbiol. *14*, 1116896.

94. Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One *3*, e3647.

95. Grieger, J.C., Choi, V.W., and Samulski, R.J. (2006). Production and characterization of adeno-associated viral vectors. Nat. Protoc. *1*, 1412–1428.

96. Zolotukhin, S., Byrne, B.J., Mason, E., Zolotukhin, I., Potter, M., Chesnut, K., Summerford, C., Samulski, R.J., and Muzyczka, N. (1999). Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther. *6*, 973–985.

97. Lin, C.Y., Huang, Z., Wen, W., Wu, A., Wang, C., and Niu, L. (2015). Enhancing Protein Expression in HEK-293 Cells by Lowering Culture Temperature. PLoS One *10*, e0123562.

98. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876.

OMTM, Volume *31*

Supplemental information

Multiparametric domain insertional profiling of adeno-associated virus VP1

Mareike D. Hoffmann, Alina C. Zdechlik, Yungui He, David Nedrud, George Aslanidi, Wendy Gordon, and Daniel Schmidt

Note S1: Linking clusters to AAV function

Using existing AAV structure / function relationships knowledge, we can begin to make association between each of the five identified clusters and their specific roles in AAV packaging and infection:

1. The strong association between poor pulldown fitness, uptake fitness, and cluster 1 is obvious considering it being comprised of mostly buried or internal residues and containing VP1u, which encodes the required PLA2 domain and nuclear localization signals.[1–4] Conversely, association with higher packaging fitness would be compatible with motif insertion promoting externalization of VP1u, thus decreasing steric hindrance with the packaged genome.

2. The sensitivity of cluster 3 with respect to packaging efficiency is consistent with trimer formation and stability, which are key determinants of capsid assembly.[5, 6] Most binding sites for cellular receptors (e.g., proteoglycan, AAVR) are located near the 3-fold axis, as well.[7–13] Flexibility of the 2-fold interface has previously been linked not only to AAV infectivity[14], but also genome packaging[15], which would explain how insertions may drive poor packaging fitness.

3. The relatively neutral fitness of cluster 2 (e.g., packaging, uptake) is consistent with most insertions ending up on the capsid interior, not interfering with assembly or cellular uptake. While the 5-fold pore is part of this cluster, a numeric simulation of VP1 copy numbers ranging from 1-10 suggests that around half of the twelve 5-fold pores are assembled from non-VP1 only, and thus available as an alternative pathway for Rep-mediated genome packaging.

4. Cluster 5 has an intricate structure, with the HI loops that surround the 5-fold pore like the blades of an aperture and connecting to the base of the 3-fold axis. Prior studies that have suggested a link between conformational changes at the 3-fold protrusion upon binding cell surface proteoglycan are communicated to conformational change at the 5-fold pore, priming the release of VP1u.[16]

Note S2: Effects of cysteine mutants

Several of the cysteine pairs we hand-picked involve interfaces that undergo conformational dynamics during infection:

1. Mutating F671C in the HI loop was strongly deleterious to infection, but this phenotype was rescued in the background of H255C, which by itself was benign. Prior studies have shown that heparin binding near the 3-fold and 2-fold axes induces an HI loop re-arrangement and an iris-like opening of the channel located at the 5-fold axis.[16] HI loop deletions, substitutions, and insertions have shown that this loop, while flexible in amino acid composition and length, is critical for proper VP1 incorporation and infectivity [17]. The same study showed that interaction of the HI loop with the underlying EF loop is mediated by hydrophobic pi-stacking interactions (F661/P373; in AAV2 numbering) and that disrupting this interaction lowers infectivity by preventing VP1 incorporating into assembled capsid. Our results are reminiscent of this mechanism.

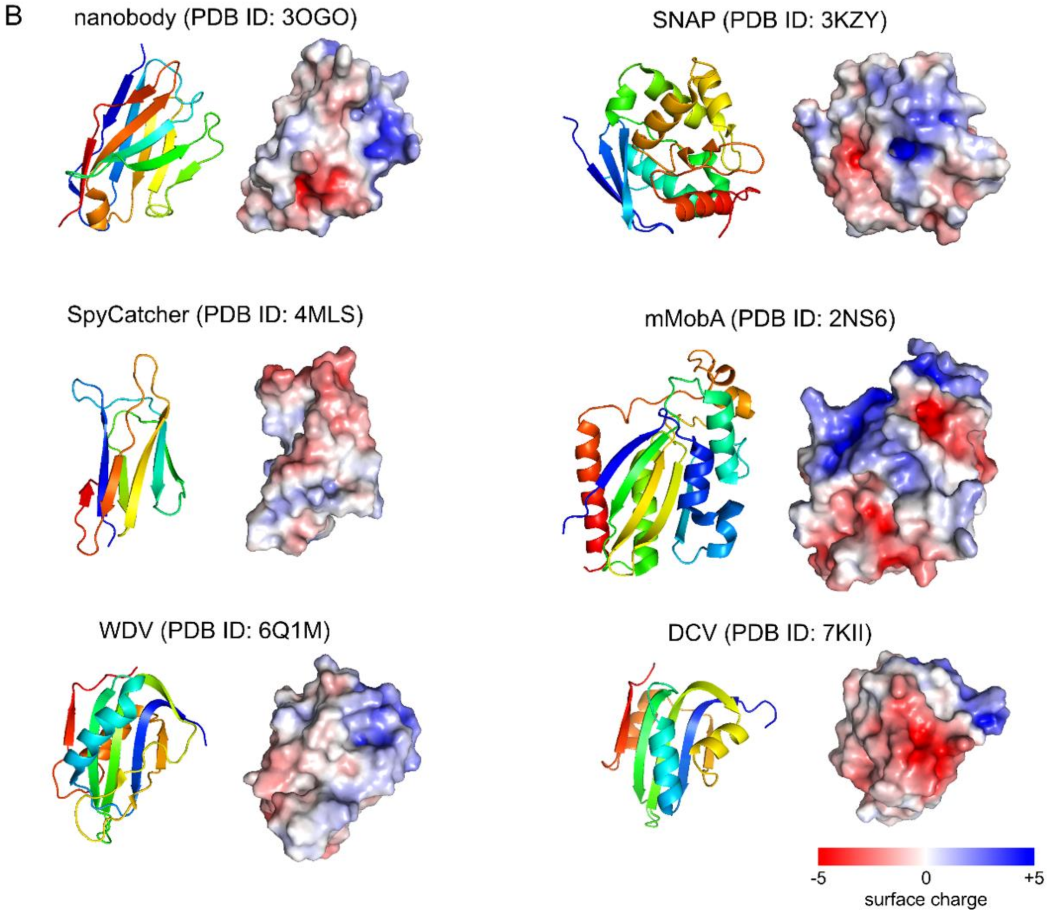
2. Positions F671/H255, which can form NH···π hydrogen bonds, are both conserved across AAV serotypes. H225C may be benign as this supports formation of an aromatic-thiol π hydrogen bond, thus allowing VP1 to incorporate or retaining structural rearrangement after receptor binding. Conversely, F671C is disruptive as it removes the aromatic component of π-stacking interactions, prevents VP1 incorporation and/or disrupts these rearrangements. The H255C/F671C double mutant may rescue infectivity by forming a disulfide bond to substitute as a stand in for π interactions.

3. H423C (at the base of the 3-fold axis) and V613C (with the 2-fold interface) individually had little effect, but together strongly impaired infectivity. This trend held true for adjacent pairs (H360C/437C; H428CL737C) that similar linked clusters comprising the 3-fold protrusion and 2-fold axis. Interestingly, several prior studies have linked AAV infectivity and conformational dynamics at the 3-fold and 2-fold axes. In addition to structural rearrangements in the HI loops, heparin binding to AAV2 causes significant rearrangement of 3-fold protrusions and the 2-fold valleys.[16] Selective oxidation of tyrosine residue at the 2-fold dimer interface lowered infectivity.[14] A mutation (R432A in AAV2) remodeled intramolecular and intermolecular hydrogen bond networks propagating from the 3-fold to both 2-fold and 5-fold axes.[18, 19]

We would like to note here that more factors than an oxidizing environment play a role in the formation of disulfide bonds, including the structure of the protein, but also the surrounding amino acids and the pH value. Furthermore, post-translational modifications other than disulfide bonds can also occur in an oxidizing environment, the effects of which cannot be distinguished from those of disulfide bonds in our experimental approach.

A

Domain	Length	Size	Termini distance	Isoelectric point
nanobody	116aa	12.85kDa	39.2Å	8.134
SNAP	181aa	19.16kDa	40.1Å	6.250
SpyCatcher	88aa	9.54kDa	21.0Å	4.989
mMobA	185aa	20.78kDa	11.2Å	6.503
WDV	136aa	15.47kDa	25.6Å	6.537
DCV	110aa	12.26kDa	27.2Å	6.211



C

Library	Domain	Linker	Total Size
FLAG	24bp	SG-domain-GS	3999bp
SpyCatcher	264bp	SGGGG-domain-GGGGS	4257bp
DCV	330bp	SGGGG-domain-GGGGS	4323bp
nanobody	348bp	SGGGG-domain-GGGGS	4341bp
WDV	408bp	SGGGG-domain-GGGGS	4401bp
SNAP	543bp	SGGGG-domain-GGGGS	4536bp
mMobA	555bp	SGGGG-domain-GGGGS	4548bp

Figure S1

Properties of inserted domains and sizes of library payloads. (A) Physical descriptions of domains used in this study. (B) Cartoon representation of domain structures (left) and surface representation with net surface charge shown as a gradient from red (negative), over white (neutral), to blue (positive). (C) Sizes of library payloads.

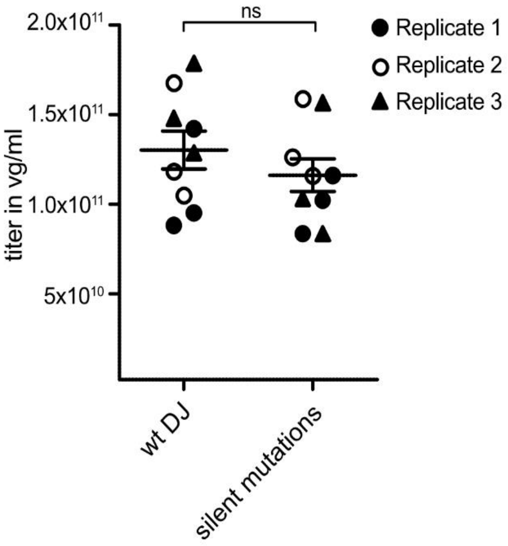
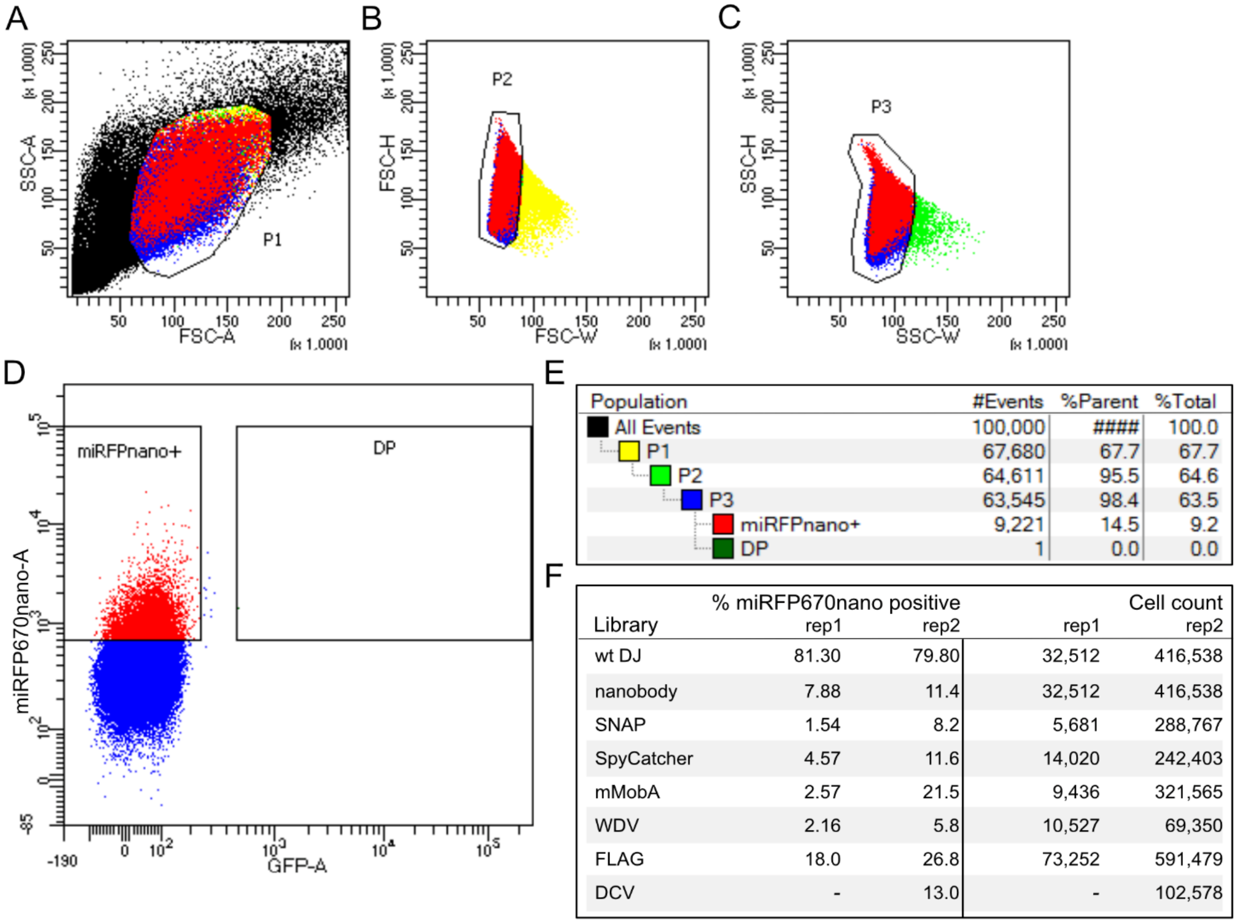


Figure S2

Packaging fitness of silent mutation variant compared to AAV-DJ. Crude lysate packaging titers quantified via qPCR. Three replicates with three technical replicates each were performed. Data are means ± SEM. No statistical significance (ns) between AAV-DJ and the silent mutations variant by an unpaired, two-sided Student's t-test (p-value: 0.3327).



E

Population	#Events	%Parent	%Total
■ All Events	100,000	####	100.0
🟨 P1	67,680	67.7	67.7
🟩 P2	64,611	95.5	64.6
🟦 P3	63,545	98.4	63.5
🟥 miRFPnano+	9,221	14.5	9.2
🟩 DP	1	0.0	0.0

F

Library	% miRFP670nano positive		Cell count	
	rep1	rep2	rep1	rep2
wt DJ	81.30	79.80	32,512	416,538
nanobody	7.88	11.4	32,512	416,538
SNAP	1.54	8.2	5,681	288,767
SpyCatcher	4.57	11.6	14,020	242,403
mMobA	2.57	21.5	9,436	321,565
WDV	2.16	5.8	10,527	69,350
FLAG	18.0	26.8	73,252	591,479
DCV	-	13.0	-	102,578

Figure S3

Infection fitness assay gating scheme. (A) Whole HEK293FT cells are gated on side (SSC-A) and forward scattering (FSC-A). (B-C) Forward scattering height (FSC-H), forward scattering width (FSC-W), and side scattering width (SSC-W) are used to gate single cells. (D) Cells are further gated using miRFP670nano as an infection marker (representative example). (E-F) Sort statistic for each gated cell population.

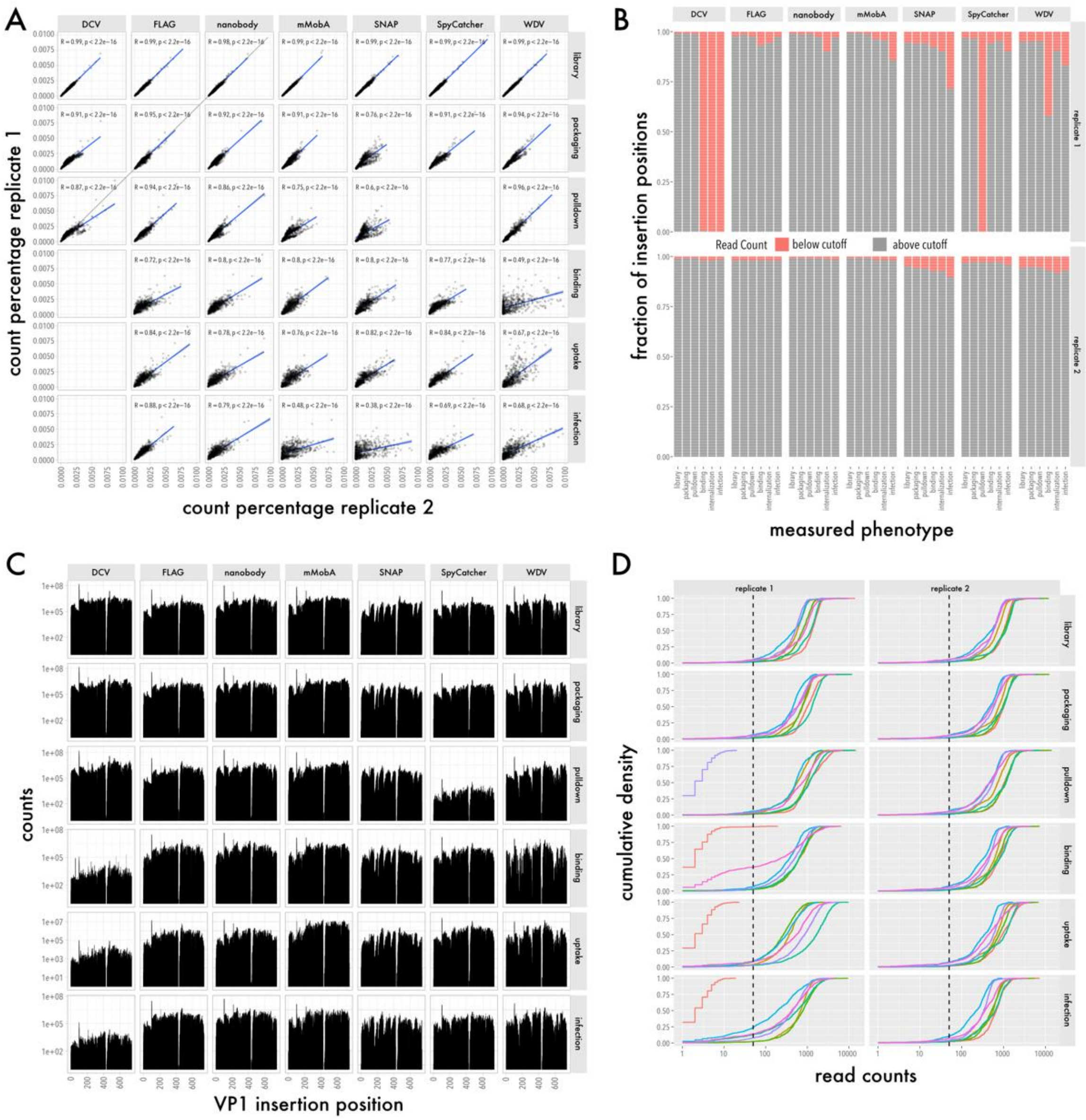


Figure S4

Fitness assay biological replicates, data completeness, and depth. (A) Read counts of replicate 1 are plotted against read counts of replicate 2 for the plasmid library, packaging, pulldown, binding, uptake, and infection assays. Data for all seven inserted domains are shown. Linear correlation was calculated (blue line). Pearson correlation coefficient and p values are shown. (B) Contingency plots showing the fraction of insertion positions below (red) and above (grey) the read count quantity cut-off of 50 reads for all seven domains and measured phenotypes. Data are shown for replicate 1 (top) and replicate 2 (bottom). (C) Total read counts of the plasmid library, packaging, pulldown, binding, uptake, and infection assays for each position are represented by insertion position. Data for all seven inserted domains are shown. (D) Cumulative density plots showing the read counts of the plasmid library, packaging, pulldown, binding, uptake, and infection assays for all seven inserted domains for replicate 1 and replicate 2. Cut off for sufficient read quantity was set to 50 reads and is represented by the black dashed lines.

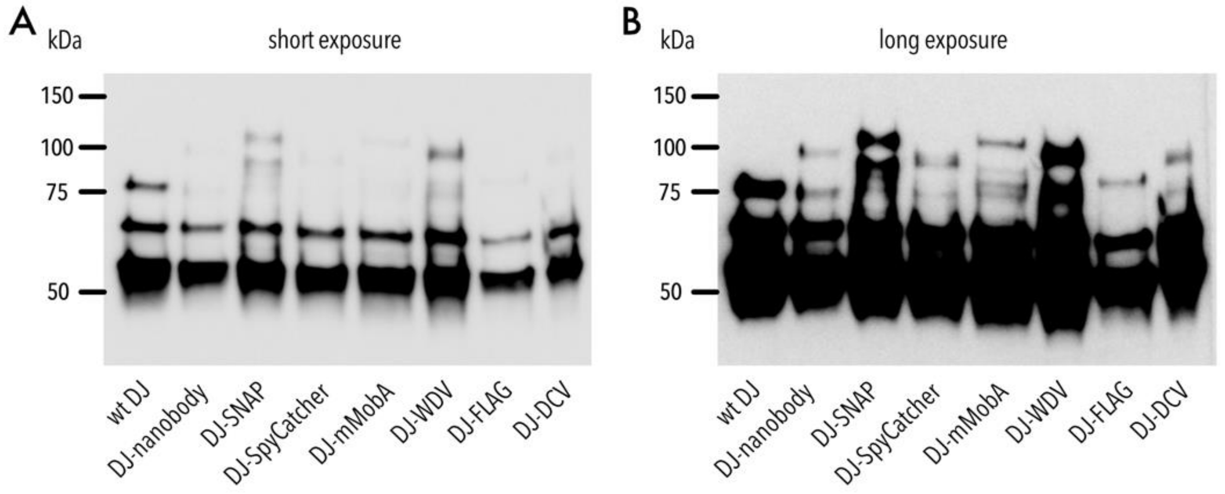


Figure S5

Western blot of AAV domain insertion libraries. (A-B) Representative Western blot image of AAV domain insertion libraries stained with B1 antibody (detecting VP1, VP2 and VP3 subunits) at a short exposure time (A) and long exposure time (B).

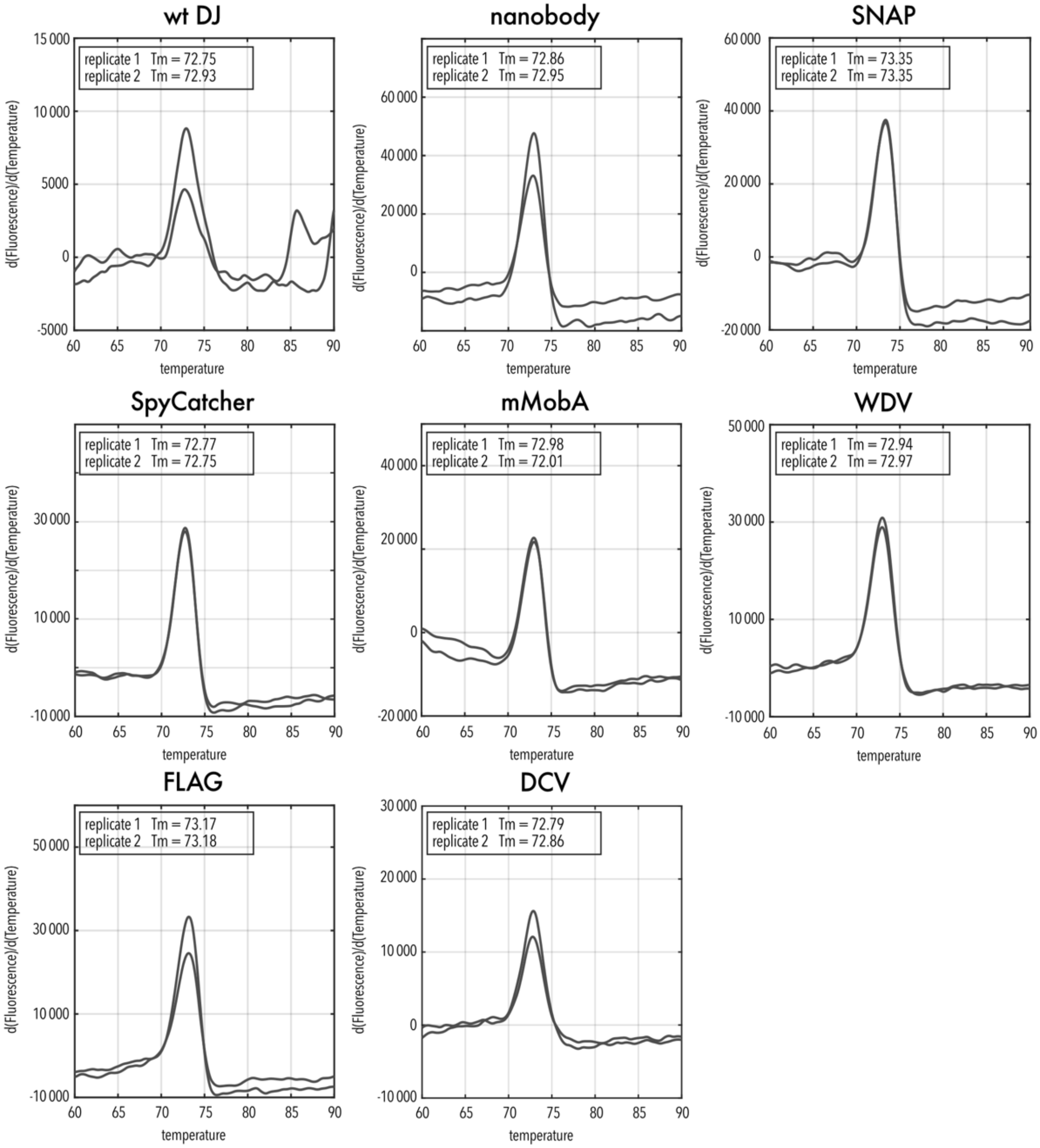


Figure S6

Thermal profiles of AAV domain insertion libraries. Data of two replicates are shown as "∂(Fluorescence)/∂(Temperature)" versus Temperature in °C.

A

sample	full	empty	undecided	total	% full
wt DJ	309	17	18	344	89.8
nanobody	508	48	26	582	87.3
SNAP	377	50	40	467	80.7

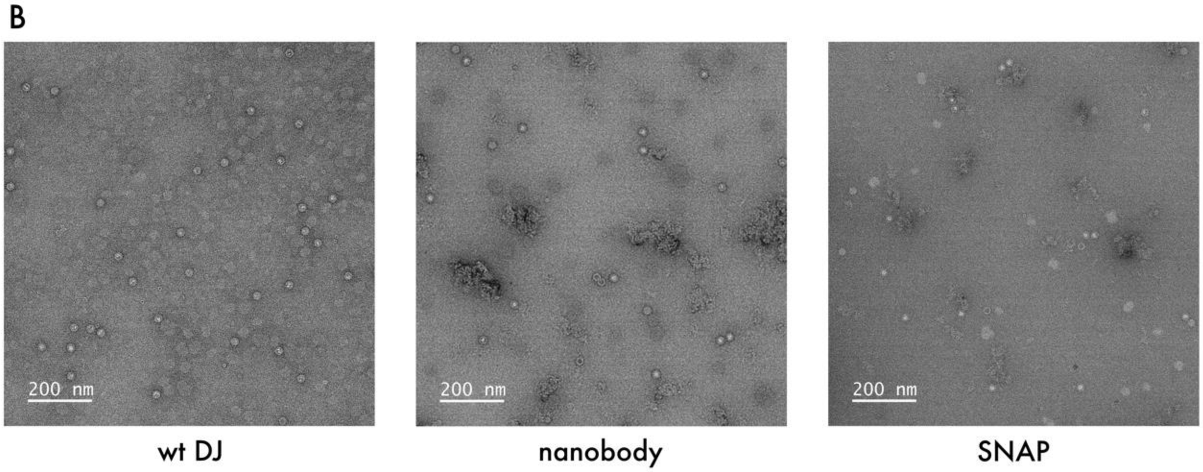


Figure S7

Quantification of empty to full capsid ratio by negative staining transmission electron microscopy. (A) More than 300 capsids were counted manually and grouped into full, empty, and undecided and the percentage of full capsids was calculated. (B) Representative transmission electron microscopy images of the indicated samples.

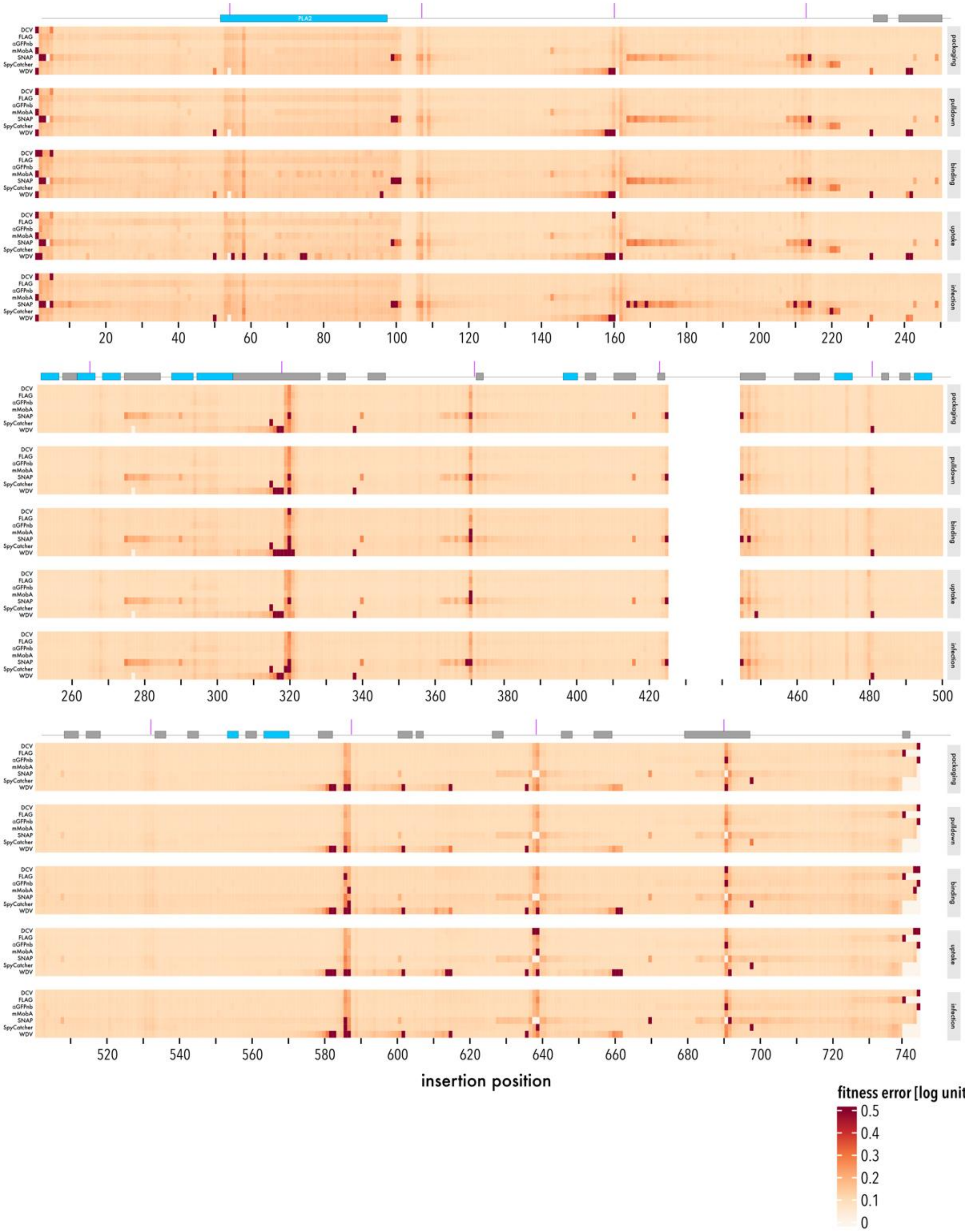


Figure S8

Accuracy of domain insertional profiling data. Standard errors of AAV fitness of all seven domain insertion libraries by insertion position and fitness assay. Secondary structure elements and VR1-9 of the capsid are indicated on top. Boundaries of oligos from VP1 assembly using SPINE are indicated by purple vertical bars.

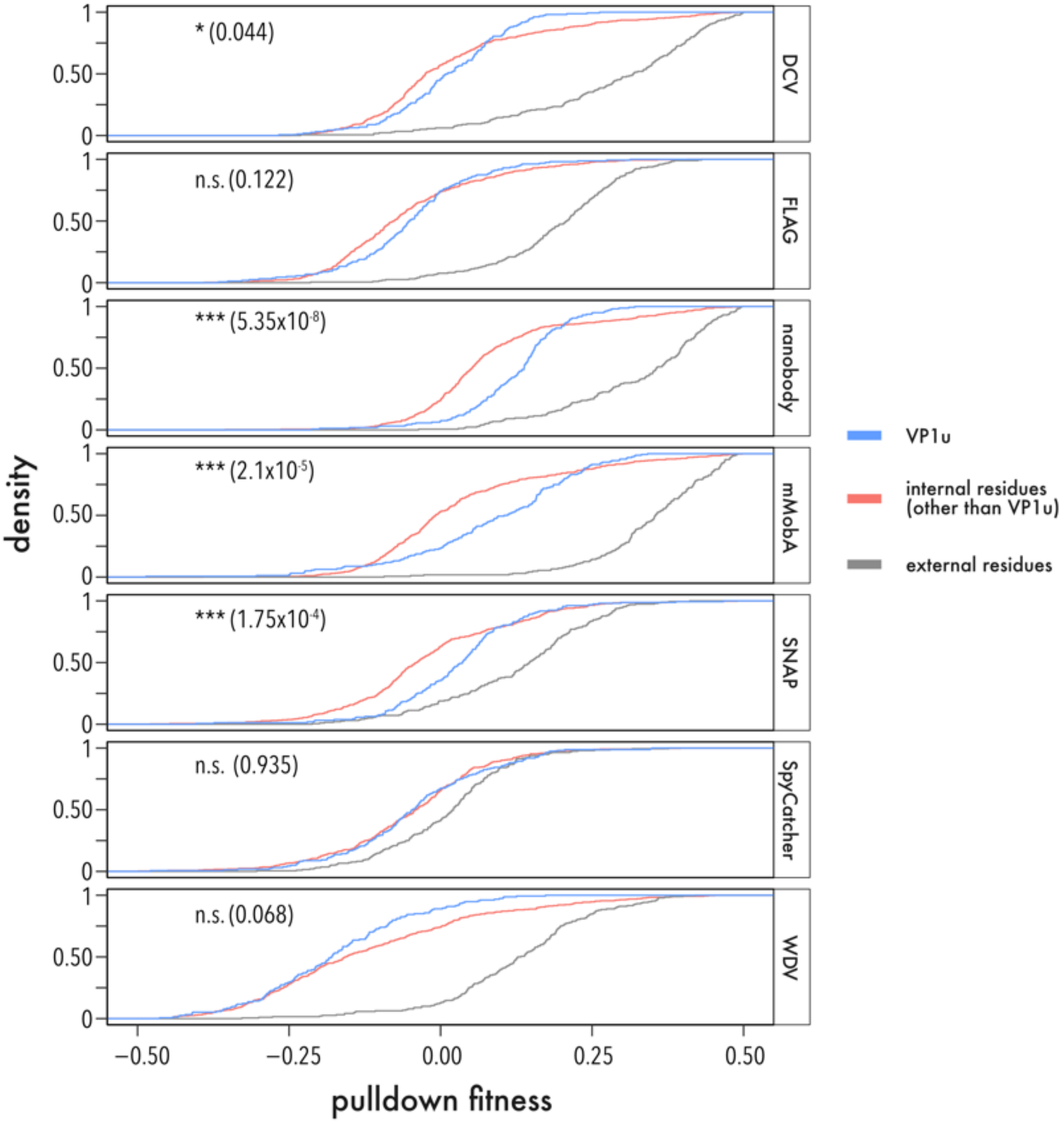


Figure S9

Distribution of pulldown fitness by residue location. Cumulative density function of pulldown fitness for each inserted motif stratified by residue location; within VP1u (blue), buried or exposed inside the capsid and not in VP1u (red), external residues (gray). Significance of distribution differences was tested using a two-sided, two-sample Kolmogorov-Smirnov test. Significance level and p values are shown.

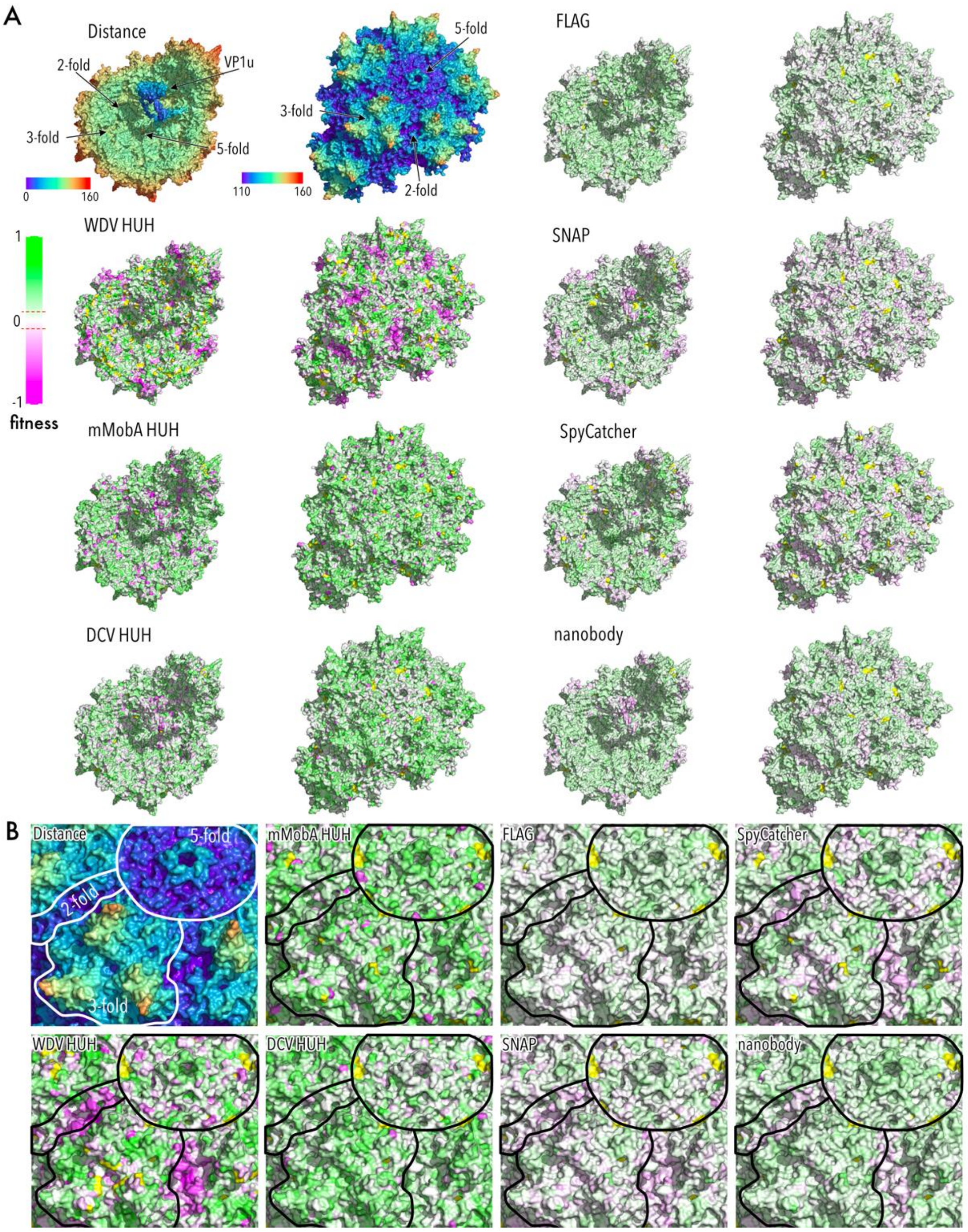


Figure S10

Binding fitness of AAV domain insertion libraries mapped to the capsid structure. (A) Top left corner: AAV-DJ capsid structure view from the inside (left) and outside (right) radially color-cued. 2-, 3-, and 5-fold axes are indicated. VP1u domain was modeled using RoseTTAFold (98) and manually positioned. All other structures show binding fitness heatmaps of the indicated domain insertions. Green indicates higher and magenta lower fitness than AAV-DJ (white) (RCSB PDB 7KFR). (B) Zoom of the outside structures from (A). 2-, 3-, and 5-fold axes are outlined.

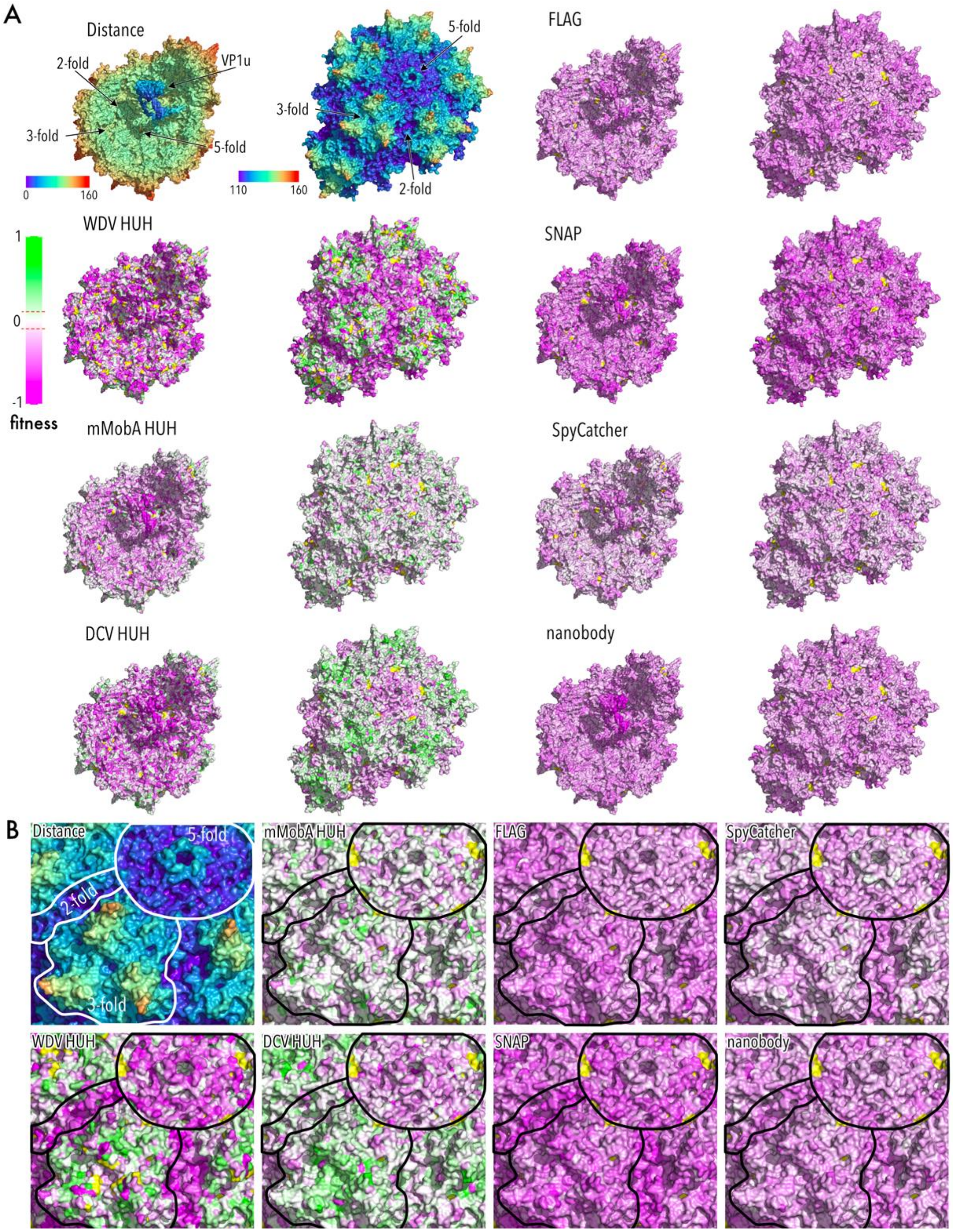


Figure S11

Uptake fitness of AAV domain insertion libraries mapped to the capsid structure. (A) Top left corner: AAV-DJ capsid structure view from the inside (left) and outside (right) radially color-cued. 2-, 3-, and 5-fold axes are indicated. VP1u domain was modeled using RoseTTAFold (98) and manually positioned. All other structures show uptake fitness heatmaps of the indicated domain insertions. Green indicates higher and magenta lower fitness than AAV-DJ (white) (RCSB PDB 7KFR). (B) Zoom of the outside structures from (A). 2-, 3-, and 5-fold axes are outlined.

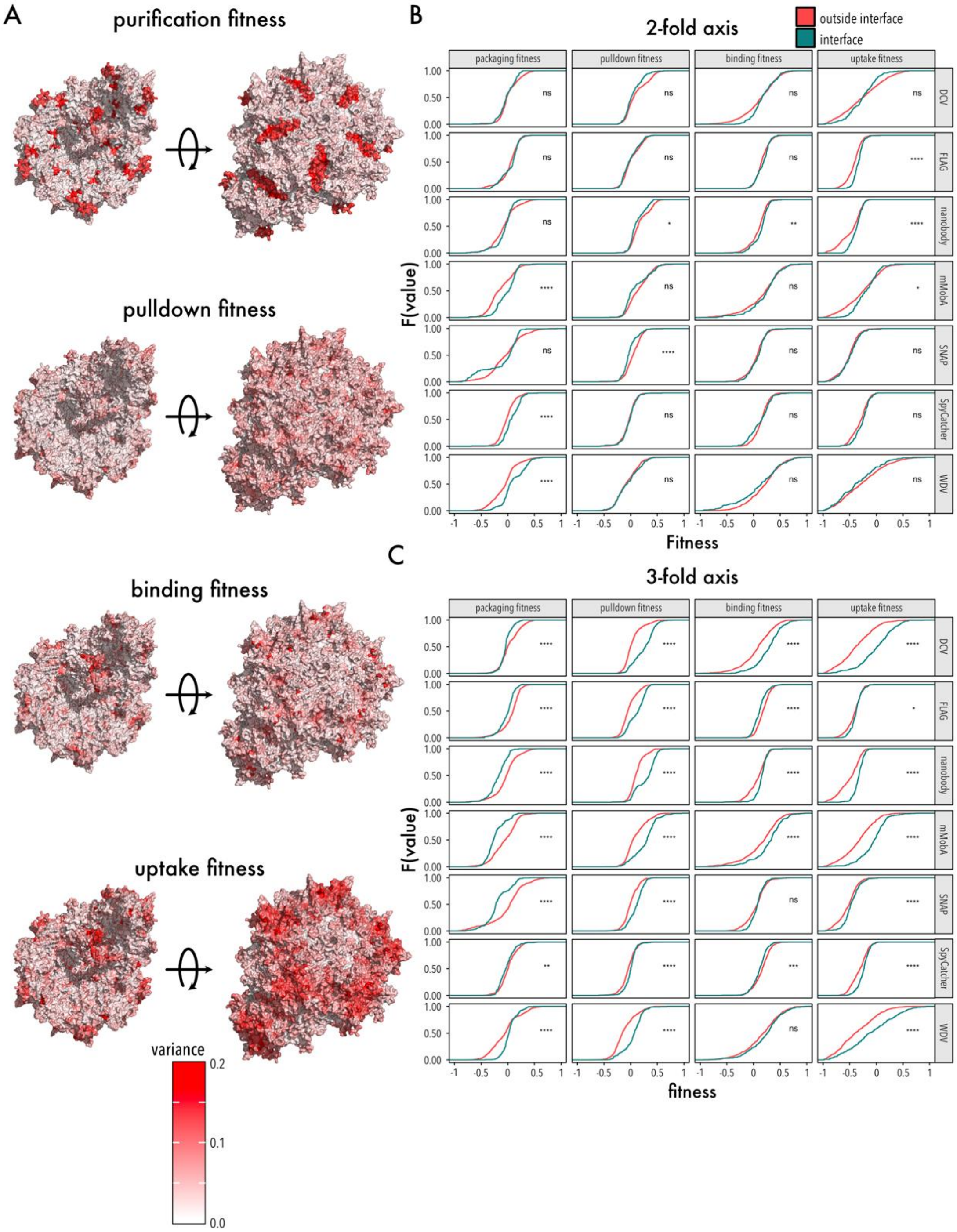


Figure S12

Variance of all measured fitness assays. (A) Variance of different fitness measures from all domain insertion libraries mapped to the AAV capsid structure. The capsid inside (left) and the capsid outside (right) are shown (RCSB PDB 7KFR). (B) Empirical cumulative density insertional fitness of residues within (petrol green) and outside (red) at the 2-fold axis (top) and 3-fold axis (bottom). Significance of distribution differences was tested using a two-sided, two-sample Kolmogorov-Smirnov test. Significance levels are shown (**** < 0.0001, *** < 0.001, ** < 0.01, * < 0.05).

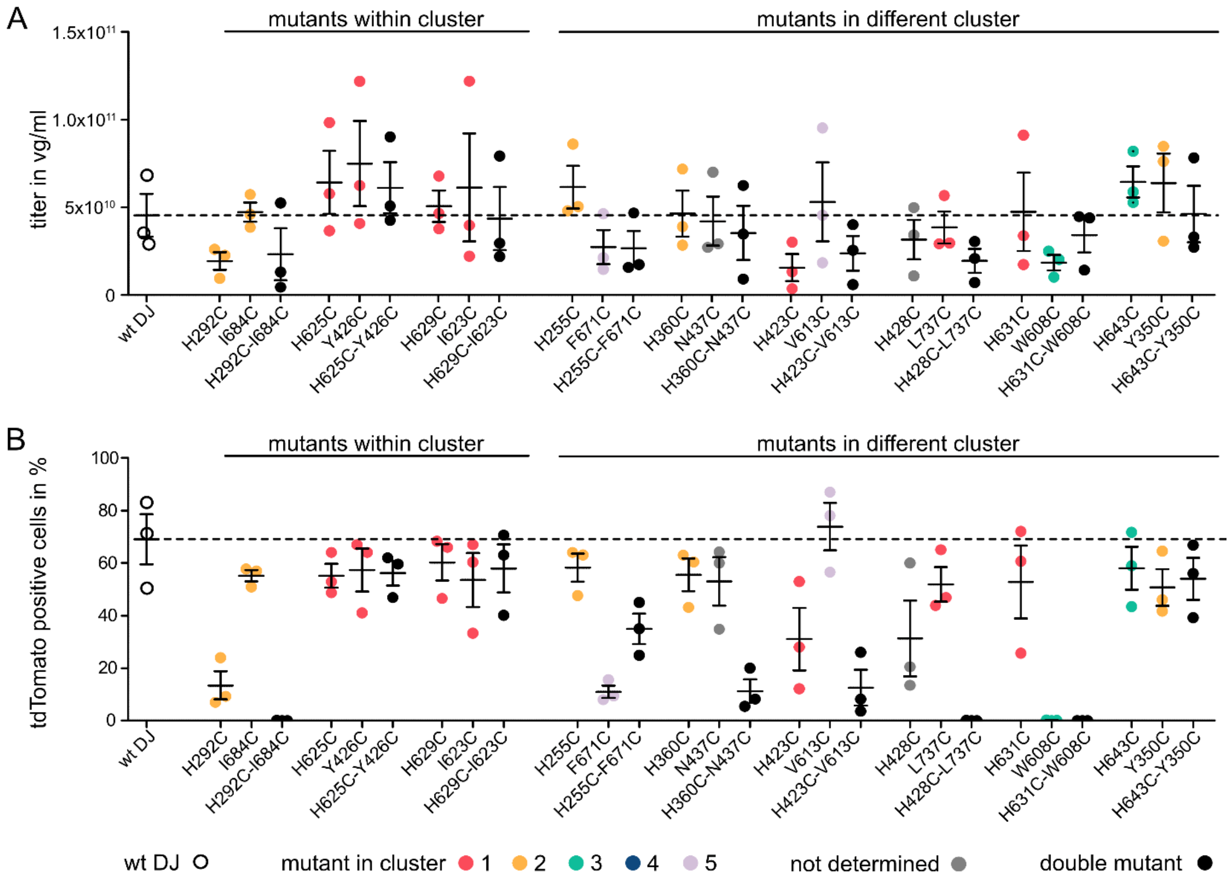


Figure S13

Packaging and infection fitness of cysteine mutants. (A) Crude lysate packaging titers quantified via qPCR. (B) Infection fitness quantified by measuring the percentage of tdTomato positive cells 48 hours post transduction with an MOI of 1×10^4 vg/cell. Data are means ± SEM. Data points of single mutants are colored by cluster membership, missing residues are gray, double mutants are black, and wildtype AAV-DJ fitness is shown as open circles and horizontal dashed line.

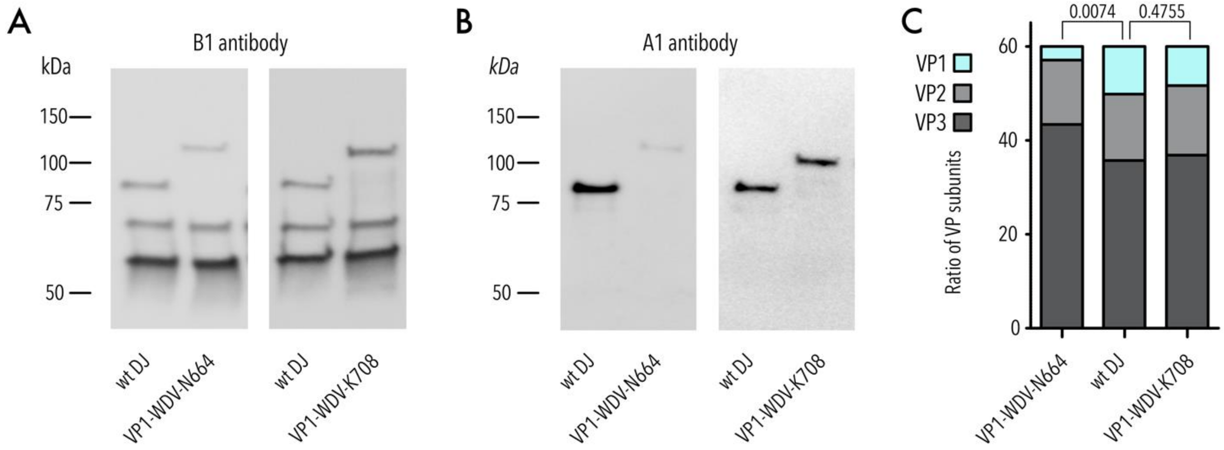


Figure S14

Quantification of VP ratios of WDV insertion variants N664 and K708. (A) Representative Western blot image of AAV domain insertion libraries stained with B1 antibody (detecting VP1, VP2 and VP3 subunits). (B) Representative Western blot image of AAV domain insertion libraries stained with A1 antibody (detecting VP1 subunits). (C) Western blot quantification of VP1, VP2, and VP3 subunits. Data are means (n=3). VP1 content varies significantly (one-way ANOVA p-value: 0.00993, Dunnett's test for pairwise comparison with wildtype AAV-DJ as control: N664 p-value 0.0074; K708 p-value 0.4755.

Table S1

Plasmids and libraries used in this study.

Plasmids not generated, but used for this study			
#	Name	Encodes/ Use	Origin
01	Adeno-helper	Adenohelper proteins for AAV productions	Cellbiolabs
02	AAV-DJ	rep2-capDJ for packaging of AAV-DJ	Cellbiolabs
03	tdTomato	ITR_CAG-tdTomato-WPRE	based on Addgene Plasmid #105554
04	DJ-M1K	rep2-capDJ mutated VP1 start codon	Schmidt Lab (Alina Zdechlik)
05	DJ-VP1	DJ VP1 only (mutated start sites for VP2 & VP3), HA tag instead of HBD; used as cloning intermediate	Schmidt Lab (Alina Zdechlik)
06	CMV-GFP-GPI	Expression of GFP at outer cell membrane	Schmidt Lab (Alina Zdechlik)
07	pUC19	Stuffer plasmid	ThermoFisher
Plasmids and libraries generated for this study			
#	Name	Use	Origin
08	library_entry	ITR_EFS-driven miRFP670nano & p40 with BsmBI sites to insert DJ-VP1 chlor; used control plasmid	Schmidt Lab (Mareike Hoffmann)
09	DJ-VP1_SilMut	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with silent mutations	Schmidt Lab (Mareike Hoffmann)
10	DJ-H255C	rep2-capDJ-H255C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
11	DJ-F671C	rep2-capDJ-F671C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
12	DJ-H255C-F671C	rep2-capDJ-H255C-F671C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
13	DJ-H292C	rep2-capDJ-H292C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
14	DJ-I684C	rep2-capDJ-I684C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
15	DJ-H292C-I684C	rep2-capDJ-H292C-I684C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
16	DJ-H360C	rep2-capDJ-H360C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
17	DJ-N437C	rep2-capDJ-N437C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
18	DJ-H360C-N437C	rep2-capDJ-H360C-N437C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
19	DJ-H423C	rep2-capDJ-H423C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
20	DJ-V613C	rep2-capDJ-V613C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
21	DJ-H423C-V613C	rep2-capDJ-H423C-V613C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
22	DJ-H428C	rep2-capDJ-H428C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
23	DJ -L737C	rep2-capDJ -L737C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
24	DJ-H428C-L737C	rep2-capDJ-H428C-L737C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
25	DJ-H625C	rep2-capDJ-H625C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
26	DJ-Y426C	rep2-capDJ-Y426C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
27	DJ-H625C-Y426C	rep2-capDJ-H625C-Y426C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)

28	DJ-H629C	rep2-capDJ-H629C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
29	DJ-I623C	rep2-capDJ-I623C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
30	DJ-H629C-I623C	rep2-capDJ-H629C-I623C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
31	DJ-H631C	rep2-capDJ-H631C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
32	DJ-W608C	rep2-capDJ-W608C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
33	DJ-H631C-W608C	rep2-capDJ-H631C-W608C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
34	DJ-H643C	rep2-capDJ-H643C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
35	DJ-Y350C	rep2-capDJ-Y350C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
36	DJ-H643C-Y350C	rep2-capDJ-H643C-Y350C; cysteine mutant	Schmidt Lab (Mareike Hoffmann)
37	DJ-VP1-L256-WDV	p40-driven DJ-VP1 with WDV insertion L256	Schmidt Lab (Mareike Hoffmann)
38	DJ-VP1-S268-WDV	p40-driven DJ-VP1 with WDV insertion S268	Schmidt Lab (Mareike Hoffmann)
39	DJ-VP1-K311-WDV	p40-driven DJ-VP1 with WDV insertion K311	Schmidt Lab (Mareike Hoffmann)
40	DJ-VP1T331-WDV	p40-driven DJ-VP1 with WDV insertion T331	Schmidt Lab (Mareike Hoffmann)
41	DJ-VP1-E349-WDV	p40-driven DJ-VP1 with WDV insertion E349	Schmidt Lab (Mareike Hoffmann)
42	DJ-VP1-T414-WDV	p40-driven DJ-VP1 with WDV insertion T414	Schmidt Lab (Mareike Hoffmann)
43	DJ-VP1-S244-WDV	p40-driven DJ-VP1 with WDV insertion S244	Schmidt Lab (Mareike Hoffmann)
44	DJ-VP1-N664-WDV	p40-driven DJ-VP1 with WDV insertion N664	Schmidt Lab (Mareike Hoffmann)
45	DJ-VP1-Y702-WDV	p40-driven DJ-VP1 with WDV insertion Y702	Schmidt Lab (Mareike Hoffmann)
46	DJ-VP1-K708-WDV	p40-driven DJ-VP1 with WDV insertion K708	Schmidt Lab (Mareike Hoffmann)
lib01	DJ-VP1 chlor	Template for domain insertion library, contains handle with chloramphenicol	Schmidt Lab (Mareike Hoffmann)
lib02	pAAV_DJ-VP1-chlor	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with chlor handle	Schmidt Lab (Mareike Hoffmann)
lib03	pAAV_DJ-VP1-nanobody	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with nanobody insertion	Schmidt Lab (Mareike Hoffmann)
lib04	pAAV_DJ-VP1-SNAP	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with SNAP insertion	Schmidt Lab (Mareike Hoffmann)
lib05	pAAV_DJ-VP1-SpyCatcher	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with SpyCatcher insertion	Schmidt Lab (Mareike Hoffmann)
lib06	pAAV_DJ-VP1-mMobA	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with mMobA insertion	Schmidt Lab (Mareike Hoffmann)
lib07	pAAV_DJ-VP1-WDV	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with WDV insertion	Schmidt Lab (Mareike Hoffmann)
lib08	pAAV_DJ-VP1-FLAG	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with FLAG insertion	Schmidt Lab (Mareike Hoffmann)
lib09	pAAV_DJ-VP1-DCV	ITR_EFS-driven miRFP670nano & p40-driven DJ-VP1 with DCV insertion	Schmidt Lab (Mareike Hoffmann)

Table S2

Sequencing statistics.

Sequencing Run	Pool Name	Assay	Repl.	Total Reads	Mean Quality R1	Mean Quality R2	Motif	Aligned Reads	Pos	Median Reads / Position	Coverage [x-fold]
Schmidt_Project_053	S1	Insertion Library	1	49,263,317	34.8	33.8	DCV	984,082	744	1295	1,323
							FLAG	440,403		564	592
							GFPnb	612,028		768	823
							mMobA	860,533		1187	1,157
							SNAP	374,315		450	503
							SpyCatcher	387,705		503	521
							WDV	663,006		849	891
	S2	Packaging Assay	1	51,060,989	34.8	33.9	DCV	866,748	744	996	1,165
							FLAG	588,126		757	790
							GFPnb	596,785		745	802
							mMobA	972,338		1280	1,307
							SNAP	355,788		448	478
							SpyCatcher	464,490		566	624
							WDV	495,047		562	665
	S3	Pulldown Assay	1	47,622,985	34.8	33.4	DCV	1,103,510	744	1058	1,483
							FLAG	479,651		557	645
							GFPnb	773,476		934	1,040
							mMobA	953,608		1099	1,282
							SNAP	431,198		515	580
							SpyCatcher	1,664		2	2
							WDV	684,376		731	920
	S4	Binding Assay	1	37,616,155	34.8	33.8	DCV	1,753	744	1	2
							FLAG	544,918		598	732
							GFPnb	562,553		611	756
							mMobA	495,428		611	666
							SNAP	311,261		355	418
							SpyCatcher	412,666		474	555
							WDV	359,617		128	483
	S5	Uptake Assay	1	47,345,415	34.7	33.4	DCV	1,781	744	2	2
							FLAG	308,480		360	415
							GFPnb	235,074		257	316
							mMobA	1,330,134		584	1,788
							SNAP	289,337		300	389
							SpyCatcher	728,389		832	979

							WDV	490,064		516	659
	S6	Infectivity Assay	1	43,873,293	34.8	33.6	DCV	1,597	744	2	2
							FLAG	530,282		680	713
							GFPnb	685,511		762	921
							mMobA	556,769		527	748
							SNAP	248,758		169	334
							SpyCatcher	383,865		424	516
							WDV	434,667		393	584
Schmidt_Project_054	S1	Insertion Library	1	51,407,231	35.1	34.7	DCV	763,808	744	997	1,027
							FLAG	540,227		694	726
							GFPnb	800,432		1018	1,076
							mMobA	798,364		1115	1,073
							SNAP	386,969		490	520
							SpyCatcher	429,433		555	577
							WDV	441,373		567	593
	S2	Packaging Assay	1	52,569,434	35.1	35.2	DCV	792,792	744	1004	1,066
							FLAG	579,228		754	779
							GFPnb	753,802		945	1,013
							mMobA	731,258		982	983
							SNAP	360,936		456	485
							SpyCatcher	439,536		580	591
							WDV	446,429		513	600
	S3	Pulldown Assay	1	55,102,309	35.2	34.8	DCV	867,430	744	1033	1,166
							FLAG	584,332		704	785
							GFPnb	878,514		1068	1,181
							mMobA	831,938		1134	1,118
							SNAP	364,218		452	490
							SpyCatcher	414,933		539	558
							WDV	421,941		450	567
	S4	Binding Assay	1	49,251,857	35.2	35	DCV	634,851	744	769	853
							FLAG	451,741		602	607
							GFPnb	561,059		700	754
							mMobA	617,489		789	830
							SNAP	256,278		315	344
							SpyCatcher	334,592		448	450
							WDV	367,036		400	493
	S5	Uptake Assay	1	54,099,860	35.2	35	DCV	815,491	744	904	1,096
							FLAG	510,011		670	685
							GFPnb	630,885		774	848

							mMobA	780,744		1049	1,049
							SNAP	284,480		336	382
							SpyCatcher	431,248		563	580
							WDV	459,667		408	618
	S6	Infectivity Assay	1	47,910,960	35.1	34.6	DCV	534,941	744	658	719
							FLAG	414,577		555	557
							GFPnb	421,517		507	567
							mMobA	466,958		622	628
							SNAP	179,921		215	242
							SpyCatcher	249,392		332	335
							WDV	318,616		350	428

Table S3

DNA oligos used in this study.

Use/ Target	Direction	Sequence	Modification
qPCR/ p40[20]	forward	TTTCCGGTGGGCAAAGG	
	reverse	GCTCACTTATATCTGCGTCACT	
qPCR/ CMV-enhancer[21]	forward	AACGCCAATAGGGACTTTCC	
	reverse	GGGCGTACTTGGCATATGAT	
NGS amplification PCR / cap	forward	CTGGTCAATGTGGATTTGGATGACTGC	
	reverse	GCAGACCAAAGTTCAACTGAAACGAATCAAC	
Pulldown/mMobA[22]	-	CCAGTTTCTCGAAGAGAAACCGGTAAGTGCACCCTCCC	3' Biotin
Pulldown/WDV[23]	-	TAATATTACCGGATGCCGCGC	3' Biotin
Pulldown/DCV[23]	-	TATTATTACCAGAAATGATGA	3' Biotin
Binding, Internalization, and Infectivity/WDV[23]	-	TAATATTACCGGATGCCGCGC	3' Amino Modifier C6 dT

Supplemental References

1. Girod, A., Wobus, C. E., Zádori, Z., Ried, M., Leike, K., Tijssen, P., Kleinschmidt, J. A., and Hallek, M. (2002). The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J Gen Virol *83*, 973-978.
2. Zádori, Z., Szelei, J., Lacoste, M. C., Li, Y., Gariépy, S., Raymond, P., Allaire, M., Nabi, I. R., and Tijssen, P. (2001). A viral phospholipase A2 is required for parvovirus infectivity. Dev Cell *1*, 291-302.
3. Venkatakrishnan, B., Yarbrough, J., Domsic, J., Bennett, A., Bothner, B., Kozyreva, O. G., Samulski, R. J., Muzyczka, N., McKenna, R., and Agbandje-McKenna, M. (2013). Structure and dynamics of adeno-associated virus serotype 1 VP1-unique N-terminal domain and its role in capsid trafficking. J Virol *87*, 4974-4984.
4. Grieger, J. C., Snowdy, S., and Samulski, R. J. (2006). Separate basic region motifs within the adeno-associated virus capsid proteins are essential for infectivity and assembly. J Virol *80*, 5199-5210.
5. Mateu, M. G. (2013). Assembly, stability and dynamics of virus capsids. Arch Biochem Biophys *531*, 65-79.
6. Medrano, M., Fuertes, M. Á., Valbuena, A., Carrillo, P. J., Rodríguez-Huete, A., and Mateu, M. G. (2016). Imaging and Quantitation of a Succession of Transient Intermediates Reveal the Reversible Self-Assembly Pathway of a Simple Icosahedral Virus Capsid. J Am Chem Soc *138*, 15385-15396.
7. Large, E. E., and Chapman, M. S. (2023). Adeno-associated virus receptor complexes and implications for adeno-associated virus immune neutralization. Front Microbiol *14*, 1116896.
8. Lerch, T. F., O'Donnell, J. K., Meyer, N. L., Xie, Q., Taylor, K. A., Stagg, S. M., and Chapman, M. S. (2012). Structure of AAV-DJ, a retargeted gene therapy vector: cryo-electron microscopy at 4.5 Å resolution. Structure *20*, 1310-1320.
9. Meyer, N. L., Hu, G., Davulcu, O., Xie, Q., Noble, A. J., Yoshioka, C., Gingerich, D. S., Trzynka, A., David, L., Stagg, S. M. et al. (2019). Structure of the gene therapy vector, adeno-associated virus with its cell receptor, AAVR. Elife *8*, e44707.
10. Pillay, S., Meyer, N. L., Puschnik, A. S., Davulcu, O., Diep, J., Ishikawa, Y., Jae, L. T., Wosen, J. E., Nagamine, C. M., Chapman, M. S. et al. (2016). An essential receptor for adeno-associated virus infection. Nature *530*, 108-112.
11. Stagg, S. M., Yoshioka, C., Davulcu, O., and Chapman, M. S. (2022). Cryo-electron Microscopy of Adeno-associated Virus. Chem Rev *122*, 14018-14054.
12. Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M. S. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A *99*, 10405-10410.
13. Meyer, N. L., and Chapman, M. S. (2022). Adeno-associated virus (AAV) cell entry: structural insights. Trends Microbiol *30*, 432-451.
14. Horowitz, E. D., Finn, M. G., and Asokan, A. (2012). Tyrosine cross-linking reveals interfacial dynamics in adeno-associated viral capsids during infection. ACS Chem Biol *7*, 1059-1066.
15. Gerlach, B., Kleinschmidt, J. A., and Böttcher, B. (2011). Conformational changes in adeno-associated virus type 1 induced by genome packaging. J Mol Biol *409*, 427-438.

16. Levy, H. C., Bowman, V. D., Govindasamy, L., McKenna, R., Nash, K., Warrington, K., Chen, W., Muzyczka, N., Yan, X., Baker, T. S. et al. (2009). Heparin binding induces conformational changes in Adeno-associated virus serotype 2. J Struct Biol *165*, 146-156.
17. DiPrimio, N., Asokan, A., Govindasamy, L., Agbandje-McKenna, M., and Samulski, R. J. (2008). Surface loop dynamics in adeno-associated virus capsid assembly. J Virol *82,* 5178-5189.
18. Bleker, S., Pawlita, M., and Kleinschmidt, J. A. (2006). Impact of capsid conformation and Rep-capsid interactions on adeno-associated virus type 2 genome packaging. J Virol *80*, 810-820.
19. Drouin, L. M., Lins, B., Janssen, M., Bennett, A., Chipman, P., McKenna, R., Chen, W., Muzyczka, N., Cardone, G., Baker, T. S. et al. (2016). Cryo-electron Microscopy Reconstruction and Stability Studies of the Wild Type and the R432A Variant of Adeno-associated Virus Type 2 Reveal that Capsid Structural Stability Is a Major Factor in Genome Packaging. J Virol *90*, 8542-8551.
20. Nonnenmacher, M., Wang, W., Child, M. A., Ren, X. Q., Huang, C., Ren, A. Z., Tocci, J., Chen, Q., Bittner, K., Tyson, K. et al. (2021). Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. Mol Ther Methods Clin Dev *20*, 366-378.
21. Negrete, A., and Kotin, R. M. (2007). Production of recombinant adeno-associated vectors using two bioreactor configurations at different scales. J Virol Methods *145*, 155-161.
22. Lovendahl, K. N., Hayward, A. N., and Gordon, W. R. (2017). Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags. J Am Chem Soc *139*, 7030-7035.
23. Tompkins, K. J., Houtti, M., Litzau, L. A., Aird, E. J., Everett, B. A., Nelson, A. T., Pornschloegl, L., Limón-Swanson, L. K., Evans, R. L., Evans, K. et al. (2021). Molecular underpinnings of ssDNA specificity by Rep HUH-endonucleases and implications for HUH-tag multiplexing and engineering. Nucleic Acids Res *49*, 1046-1064.