

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/281407578>

# Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker and Compact Domain Conformation

ARTICLE in **BIOCHEMISTRY** · AUGUST 2015

Impact Factor: 3.02 · DOI: 10.1021/acs.biochem.5b00610

---

READS

29

## 8 AUTHORS, INCLUDING:



[Francisco Zarate-Perez](#)  
Virginia Commonwealth University

7 PUBLICATIONS 18 CITATIONS

[SEE PROFILE](#)



[Michael Linden](#)  
King's College London

49 PUBLICATIONS 2,619 CITATIONS

[SEE PROFILE](#)



[Carlos R. Escalante](#)  
Virginia Commonwealth University

40 PUBLICATIONS 1,817 CITATIONS

[SEE PROFILE](#)

# 1 Structural Studies of AAV2 Rep68 Reveal a Partially Structured Linker 2 and Compact Domain Conformation

3 Faik N. Musayev,<sup>†</sup> Francisco Zarate-Perez,<sup>‡</sup> Martino Bardelli,<sup>§</sup> Clayton Bishop,<sup>‡</sup> Emil F. Saniev,<sup>‡</sup>  
4 R. Michael Linden,<sup>§,II</sup> Els Henckaerts,<sup>§</sup> and Carlos R. Escalante<sup>\*‡</sup>

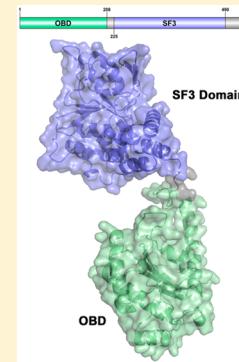
5 <sup>†</sup>Department of Medicinal Chemistry, School of Pharmacy, <sup>‡</sup>Department of Physiology and Biophysics, School of Medicine,  
6 Virginia Commonwealth University, Richmond, Virginia 23298, United States

7 <sup>§</sup>Department of Infectious Diseases, King's College London, London SE1 9RT, United Kingdom

8 <sup>II</sup>UCL Gene Therapy Consortium, UCL Cancer Institute, University College London, London WC1E 6DD, United Kingdom

9  Supporting Information

10 **ABSTRACT:** Adeno-associated virus (AAV) nonstructural proteins Rep78 and Rep68 carry out all  
11 DNA transactions that regulate the AAV life cycle. They share two multifunctional domains: an  
12 N-terminal origin binding/nicking domain (OBD) from the HUH superfamily and a SF3 helicase  
13 domain. A short linker of ~20 amino acids that is critical for oligomerization and function  
14 connects the two domains. Although X-ray structures of the AAV5 OBD and AAV2 helicase domains  
15 have been determined, information about the full-length protein and linker conformation is not  
16 known. This article presents the solution structure of AAV2 Rep68 using small-angle X-ray scattering  
17 (SAXS). We first determined the X-ray structures of the minimal AAV2 Rep68 OBD and of the  
18 OBD with the linker region. These X-ray structures reveal novel features that include a long  
19 C-terminal  $\alpha$ -helix that protrudes from the core of the protein at a 45° angle and a partially  
20 structured linker. SAXS studies corroborate that the linker is not extended, and we show that a  
21 proline residue in the linker is critical for Rep68 oligomerization and function. SAXS-based rigid-  
22 body modeling of Rep68 confirms these observations, showing a compact arrangement of the two  
23 domains in which they acquire a conformation that positions key residues in all domains on one face of the protein, poised to  
24 interact with DNA.



25 **T**he nonstructural Rep proteins from adeno-associated virus  
26 (AAV) are multifunctional proteins with specialized  
27 domains equipped to handle the complex interactions with  
28 DNA during the AAV life cycle.<sup>1,2</sup> AAV has a single-stranded  
29 DNA genome of ~4.7 kb containing two major open reading  
30 frames (ORFs) flanked by inverted terminal repeats (ITRs).  
31 The ITRs form T-shaped hairpin structures and contain the cis-  
32 regulatory functions required for replication, transcriptional  
33 regulation, and possibly site-specific integration.<sup>3–9</sup> The stem of  
34 the hairpin forms a double-stranded region containing a Rep  
35 binding site (RBS) made up of several 5'-GCTC-3' repeats.  
36 A terminal resolution site (trs) located upstream of the RBS is  
37 the site of a strand- and site-specific endonuclease reaction  
38 required to complete the replication of the AAV genome.<sup>3,11,12</sup>  
39 Despite its limited genome size, AAV generates eight different  
40 polypeptide chains. The right ORF regulated by the P<sub>40</sub>  
41 promoter produces three capsid proteins (VP1–3) and an  
42 assembly activating protein (AAP).<sup>13–18</sup> The left ORF uses two  
43 different promoters, and alternative splicing, to generate four  
44 nonstructural proteins: two large Rep proteins (Rep78 and  
45 Rep68) transcribed from the p5 promoter and two small Rep  
46 proteins (Rep52 and Rep40) regulated through the p19  
47 promoter.<sup>14,19–21</sup> Thus, Rep40 is equivalent to the helicase  
48 domain of Rep68. Most of the biochemical activities re-  
49 quired for transcriptional regulation,<sup>22,23</sup> DNA replication<sup>24,25</sup>  
50 and site-specific integration<sup>26–28</sup> are carried out by the large

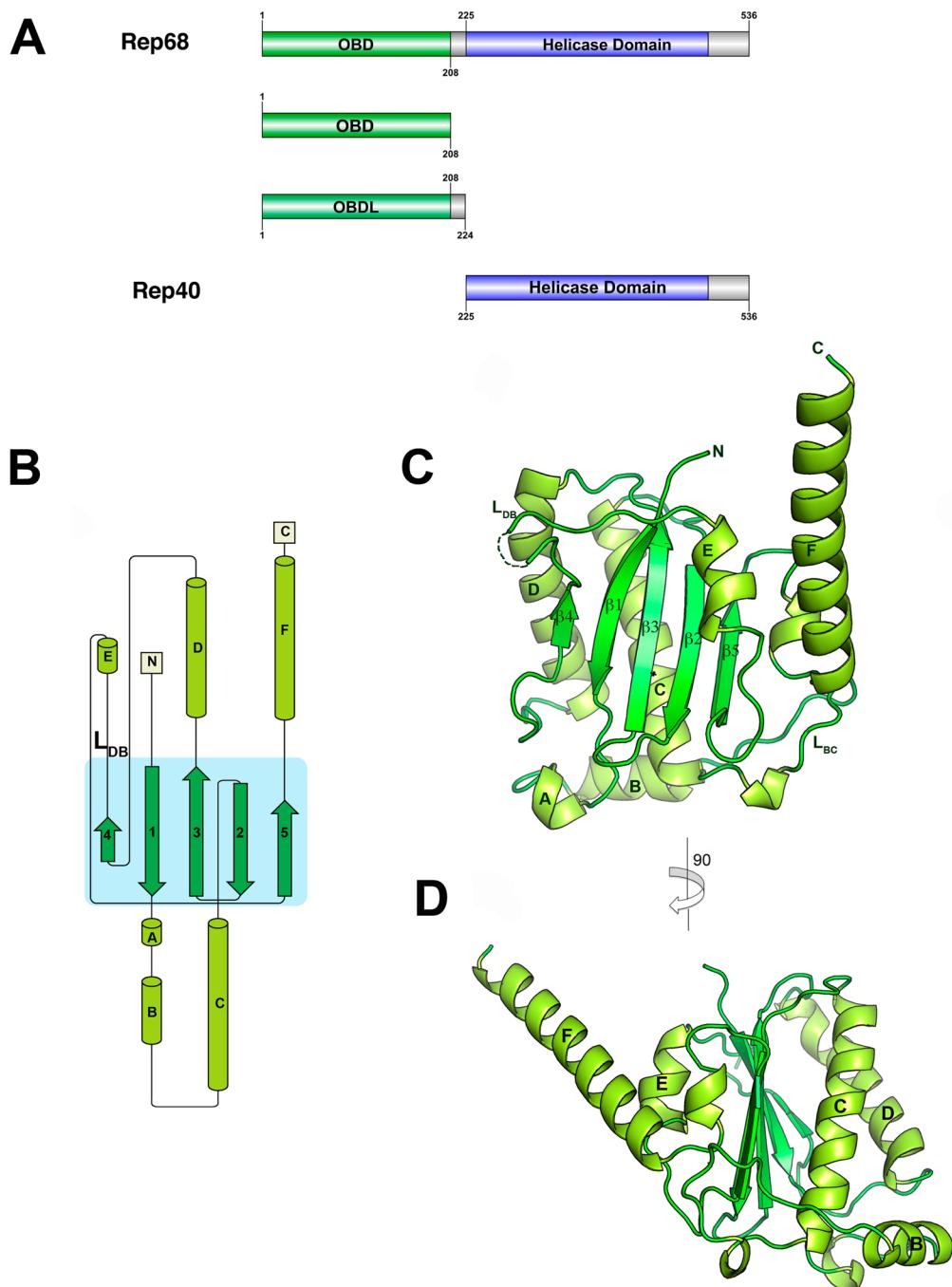
Rep78/Rep68 proteins, whereas the small Reps are thought  
51 to be important for DNA packaging into preformed  
52 capsids.<sup>29,30</sup>

The wide spectrum of functions performed by Rep78/Rep68  
53 is a direct reflection of the multifunctionality of their domains.  
54 The N-terminal origin binding domain (OBD) displays site-  
55 and strand-specific endonuclease activity and recognizes the  
56 GCTC repeats.<sup>24,31–34</sup> Structurally, the OBD is a member of  
57 the HUH endonuclease family specialized in the cleavage and  
58 rejoicing of single-stranded DNA substrates (ssDNA), which  
59 occurs during transposition, bacterial conjugation, rolling-circle  
60 replication of bacterial DNA and bacteriophages, and in the  
61 replication of small eukaryotic viruses.<sup>35</sup> This family is  
62 characterized by two signature motifs that participate in the  
63 nuclelease reaction: the HUH motif, which consists of two  
64 histidine (H) residues separated by a large hydrophobic residue  
65 (U), and the Y motif, with either one or two tyrosine (Y)  
66 residues. Structurally, the common feature of HUH endonu-  
67 cleases is a five-stranded antiparallel  $\beta$ -sheet surrounded by  
68  $\alpha$ -helices, where the HUH motif is found in one of the strands.  
69 The position of the Y motif can vary: in the relaxase subgroup,  
70

Received: June 3, 2015

Revised: August 20, 2015





**Figure 1.** Structure of AAV2 OBD. (A) Domain structure of AAV2 Rep68 protein: OBD is shown in green, Rep40 (SF3 helicase domain), in blue, and linker and C-terminal tail, in gray. The Rep40 protein spans residues 225–536 of Rep68. (B) Topology diagram of AAV2 OBD. (C) Ribbon diagram of the OBD structure. The  $\alpha$ -helices are light green, and  $\beta$ -strands are dark green. Secondary structure elements are labeled with  $\alpha$ -helices A–F, and  $\beta$ -sheets are numbered 1–5. The DNA binding loop  $L_{DB}$  with missing residues in the structure is shown as a dotted line. (D) A view of the structure rotated by 90° clockwise.  $\alpha$ -Helix F is shown protruding from the core of the structure at an angle of almost 45°.

it is located N-terminally, whereas it is found at the C-terminus in the Rep subclass.<sup>36–38</sup>

The helicase domain found in all Rep proteins is a representative of the SF3 family helicase with 3' to 5' unwinding activity.<sup>38–43</sup> The ATPase core is a modified version of the AAA<sup>+</sup> domain with additional features that include an N-terminal helical bundle and a  $\beta$ -hairpin that is involved in DNA binding during translocation.<sup>38</sup> Consequently, the large Rep proteins can interact with DNA in two different modes: one is mediated by the OBD and recognizes DNA in a

sequence-specific manner, whereas the SF3 helicase domain interacts with DNA nonspecifically.<sup>44</sup> Furthermore, the large Rep proteins show a complex and dynamic oligomerization behavior, which may add an additional level of regulation to Rep interactions with DNA.<sup>45</sup> How the two domains cooperate with each other in order to interact with DNA and how they are arranged in the context of Rep78/Rep68 is not known. Moreover, studies have established that the linker region plays a critical role in the oligomerization and function of AAV Rep proteins; therefore, it is likely that this region acquires a

particular secondary structure and may not simply be extended.<sup>46,47</sup> To answer these questions, we performed X-ray crystallography and small angle X-ray scattering (SAXS) studies on the AAV2 Rep68 functional domains and full-length protein. Our results show that the OBD has a long extended helix that includes part of the linker, which is not extended, but is tilted in a way that brings the two domains into a compact configuration. SAXS studies on a monomeric version of full-length Rep68 confirm this observation and show for the first time the overall domain architecture of a full-length AAV2 Rep protein.

## MATERIALS AND METHODS

**Protein Expression and Purification.** The DNA region encoding amino acids 1–208 (OBD) and 1–224 (OBD plus linker) from adeno-associated virus type 2 (AAV2) (GeneBank protein\_id = AF043303.1) was cloned into pET15b (Novagen) using restriction sites *Nde*I and *Xho*I. The residue C151 was mutated to serine, as it was found to produce disulfide bonds and inhibit crystallization, and Rep-C151S was shown to be fully functional in supporting the AAV life cycle.<sup>45</sup> The OBD constructs were overexpressed in *Escherichia coli* strain BL21 pLysS at 37 °C in Luria–Bertani (LB) broth. IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside) was added to a final concentration of 1 mM when an OD of 0.6 was reached. Cells were harvested after 5 h and stored at –80 °C. The cell pellet was resuspended in binding buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM imidazole, 10% glycerol, 1 mM TCEP, pH 7.9) and lysed by sonication. The OBD was purified with a Ni-NTA column (Qiagen) using step gradients of 10 and 30 mM imidazole to wash nonspecific proteins binding to the column and was eluted with 100 mM imidazole. Protein was loaded onto a HiLoad desalting column (GE) to change into thrombin buffer (25 mM Tris-HCl, 200 mM NaCl, 10% glycerol pH 8.0). His-tag was cut by addition of thrombin (1 unit/mg) and removed by passing through a Ni-NTA column. The untagged OBD was collected from the flow through, concentrated, and further purified by gel filtration on a HiLoad 16/60 Superdex 75 column (GE Healthcare) previously equilibrated with gel filtration (GF) buffer (25 mM Tris-HCl, 200 mM NaCl, 1 mM TCEP, pH 7.5). The protein was concentrated to ~40 mg/mL using Millipore Centricon (10 kDa cutoff). AAV2 Rep40 was purified as described elsewhere.<sup>38</sup> AAV-2 Rep68<sub>wt</sub> (1–536) and Rep68Y224Δ were expressed in *E. coli* strain BL21 pLysS at 18 °C as described in earlier reports.<sup>46</sup> In brief, histidine-tagged Rep68 was purified in a Ni-NTA column, and after PreScission protease cleavage of the His-tag, the protein was purified on a HiLoad Superdex 200 16/60 column (GE Healthcare). OBDL was purified using the same procedure. Crystallographic data was collected using our X-ray home source that consists of a Rigaku Micromax 007 X-ray generator and a Raxis IV<sup>+</sup> area detector.

**Crystallization, X-ray Data Collection, and Structure Determination.** Crystallization was carried out using the hanging-drop method with commercially available screening kits at 4 °C. Crystals grew after 2 to 3 days in 50 mM cacodylate, pH 6.5, 80 mM sodium acetate, 15 mM magnesium acetate, and 8–10% isopropanol. Crystals were cryoprotected in reservoir buffer and supplemented with 20% MPD before flash freezing them in liquid nitrogen. The crystals diffracted to 2.3 Å and belonged to space group P2<sub>1</sub>2<sub>1</sub>2 with unit cell dimensions  $a = 183.4$  Å,  $b = 154.4$  Å,  $c = 38.8$  Å. Diffraction data was collected at the National Synchrotron Light Source

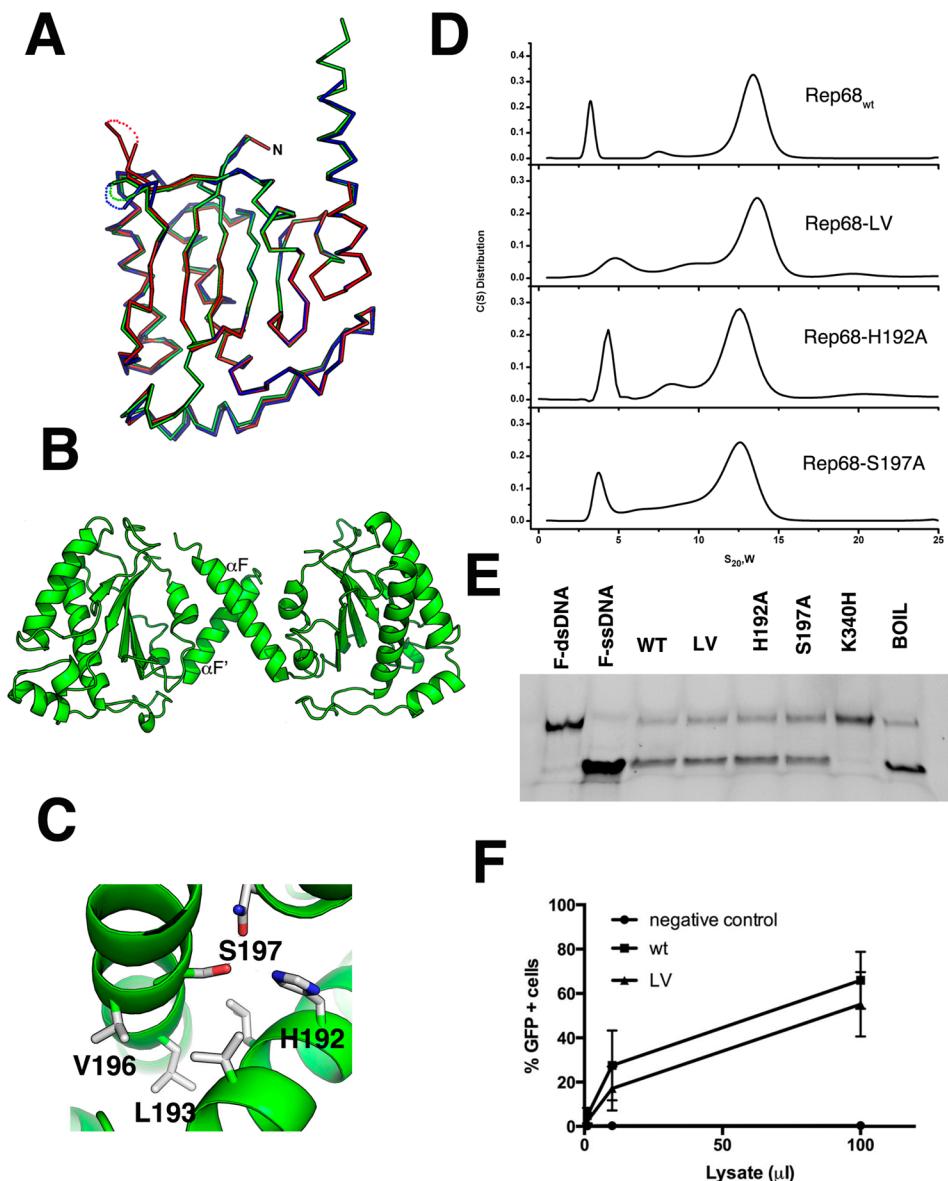
**Table 1. Data Collection and Refinement Statistics**

data collection	OBD (1–208)	OBDL (1–224)
space group	P2 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub>
Cell Dimensions		
$a, b, c$ (Å)	186.4, 154.4, 38.8	75.6, 178.7, 130.4
$\alpha, \beta, \gamma$ (deg)	90, 90, 90	90, 91.7, 90
wavelength (Å)	0.9792	1.54
resolution (Å)	30–2.30 (2.34–2.30)	30–2.6 (2.69–2.6)
no. of measured	566 473	426 030
no. of unique	50 960	105 074
data coverage (%) <sup>a</sup>	99.8 (100)	99.1 (98.9)
$R_{\text{merge}}(\%)^{a,b}$	0.099 (0.402)	0.075 (0.311)
$I/\sigma^{a,c}$	17.7 (8.6)	12.4 (4.1)
refinement statistics		
resolution range	30.0–2.3	29.8–2.6
reflections	50 960	105 074
$R_{\text{cryst}}(\%)^c$	22.5	18.1
$R_{\text{free}}(\%)^d$	24.2	24.8
non-hydrogen atoms	5024	17677
protein	4704	17493
metal	2	11
water	314	173
average B-factors (Å <sup>2</sup> )	49.8	59.1
protein	50.0	59.2
metal	49.2	88.4
water	46.4	47.3
RMS Deviations		
bonds (Å)	0.02	0.016
angles (Å)	1.4	1.53
Ramachandran Plot Quality		
most favored (%)	99.0	97.0
additional allowed (%)	1.0	3.0
generously allowed (%)	0	0.0
disallowed (%)	0	0

<sup>a</sup>Values for the outermost shells are given in parentheses. <sup>b</sup> $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$ , where  $I$  is the integrated intensity of a given reflection. <sup>c</sup> $R_{\text{cryst}} = \sum (\|F_o\| - \|F_c\|) / \sum |F_o|$ . <sup>d</sup>For  $R_{\text{free}}$  calculations, 5% of data was excluded from refinement.

(NSLS) at Brookhaven National Laboratory beamline X6a. The data were processed with the program HKL2000,<sup>48</sup> and the structure was solved by molecular replacement using the program PHENIX. We used the structure of the AAV5 OBD as a search model (PDB ID: 1MSS). Model building was carried out using PHENIX,<sup>49</sup> and manual building, using the program COOT.<sup>50</sup> OBDL data was collected using our X-ray home source that consists of a Rigaku Micromax 007 X-ray generator and a Raxis IV<sup>+</sup> area detector.

**Sedimentation Velocity.** Sedimentation velocity experiments were carried out using a Beckman Optima XL-I analytical ultracentrifuge (Beckman Coulter Inc.) equipped with an eight-position AN-60Ti rotor. Rep protein samples were loaded in the cells, using, in all cases, the GF buffer. Samples in double sector cells were centrifuged at 25 000 rpm. In all experiments, temperature was kept at 20 °C. Sedimentation profiles were recorded using UV absorption (280 nm) and interference scanning optics. For analysis of the data, the program Sedfit was used to calculate sedimentation coefficient distribution profiles using the Lamm equation.<sup>51</sup>



**Figure 2.** Superposition of OBD molecules in the asymmetric unit. (A) Ribbon diagram of the superposition of the three molecules in the asymmetric unit. Molecule A is shown in green; molecule B, blue; and molecule C, red. The three loops L<sub>DB</sub> have different conformations and are shown as dotted connections. (B) Ribbon diagram of the dimer in the asymmetric unit formed by molecules A and B. The small dimer interface occurs through residues in α-helix F. (C) Details of the residues involved in the oligomeric interaction with Ser197, Val196, and Leu193 from molecule A interacting with His192, Leu193, and Val196 from molecule B. (D) Sedimentation velocity profiles of Rep68<sub>wt</sub>, Rep68-L193AV196A, Rep68-H192A, and Rep68-S197A. Experiments were carried out at 20 °C and 25 000 rpm on a Beckman XL-I analytical ultracentrifuge. Scans were collected every 2 min using absorbance at 280 nm. Data was analyzed using the program Sedfit.<sup>51</sup> (E) Helicase assay of Rep68 mutants. Fluorescent-labeled DNA molecule has a 3' single-stranded tail and an 18 bp region (first lane). Upon ATP addition, helicase activity is shown as the 18 nucleotide fluorescein-labeled ssDNA is displaced (second lane). Rep68, L193AV196A, H192A, and S197 proteins all unwind DNA. K340H mutant is ATPase-negative and does not unwind DNA. Last lane shows the DNA substrate after heating at 100 °C for 5 min. (F) Comparison of the production of rAAV2-GFP infectious particles in the presence of wt or LV mutant. Various volumes of supernatant (in mL, x axis) from 293T cells producing rAAV2-GFP were added to HeLa cells, and the percentage of GFP-positive infected cells was determined by FACS analysis, as described in Materials and Methods. Data is presented as average ± standard deviation from three independent experiments.

174 **Small-Angle X-ray Scattering (SAXS).** Data were  
 175 collected at three different concentrations that produce a single  
 176 homogeneous population, as determined by sedimentation  
 177 velocity studies. Synchrotron SAXS measurements were  
 178 performed at Brookhaven National Laboratories at beamline  
 179 X9 and at the Advanced Light Source at the Lawrence Berkeley  
 180 National Laboratory at the SYBILS beamline. Three different  
 181 concentrations of each sample were prepared and measured  
 182 (1, 2, and 3 mg/mL). All data were processed with the package  
 183

ATSAS.<sup>52</sup> Buffer subtraction was carried out using beamline-specific software. Radii of gyration ( $R_g$ ) were evaluated using the Guinier approximation,  $sR_g < 1.3$ . Distance distribution functions and maximum diameters  $D_{\max}$  were calculated using the program GNOM.<sup>53</sup> SAXS molecular envelopes were calculated using the programs DAMMIN and GASBOR.<sup>54,55</sup> Conformational flexibility of the linker and C-terminal tail was analyzed with the program EOM.<sup>56</sup> We used the structures of Rep40 (PDB ID: 1s9h) and of OBDL and connect the two

domains with a flexible linker (residues 215–224) to generate an atomic model of AAV2 Rep68. Using this initial model, we performed rigid-body and molecular dynamics using BILBOMD to generate the best Rep68 model that fits the SAXS data.<sup>57</sup>

**DNA Helicase Assay.** The helicase assay was based on a modification of the strand-displacement assay described elsewhere.<sup>58</sup> The DNA substrate (28:18) consists of 3' tail of 10 nucleotides adjacent to 18 bp. The top strand has been labeled with fluorescein (F) at the 5' end. All reactions were performed in a final volume of 50 μL in a buffer containing 25 mM HEPES, 50 mM NaCl, pH 7.0. For the reaction, 1 μM protein was mixed with 0.5 μM ds F-DNA (28:18). The reaction was started by addition of 5 mM ATP-MgCl<sub>2</sub> and 2.5 μM trap ssDNA. Reaction was incubated at 25 °C for 1 h. EDTA was used to stop the reaction at a final concentration of 20 mM. Aliquots of 10 μL were loaded in a 12% bis-acrylamide gel (30%) (19:1) using 6× loading dye (0.25 xylene cyanol FF, 30% glycerol). For densitometry and analysis of the bands, a Gel Doc EZ Imager was used, using the automatic lane and band detection tool. Lane background subtraction, white illumination, and an activation time of 300 s were used for the analysis.

**Fluorescent Anisotropy Binding Assays.** Binding assays were performed using 5 nM fluorescein labeled 41-mer DNA containing the Rep binding site (TGGCGGCCGGTTGGG-GCTCGCGCTCGCTCGCTGGGCG). Rep68 at different concentrations was mixed with DNA at a final volume of 300 μL using the following buffer: 25 mM HEPES (pH 7.0), 100 mM NaCl, 1 mM TCEP. Fluorescence readings were taken on a PC1 fluorimeter (ISS, Inc.) with excitation and emission filters at 492 and 528 nm. Tubes were equilibrated at 20 °C for 20 min before measurement. Each anisotropy point is the average of 10 measurements. Anisotropy is calculated as the ratio of the difference between vertical and horizontal emission intensities over the total normalized intensity. The fraction of DNA bound (*B*) was calculated using eq 1

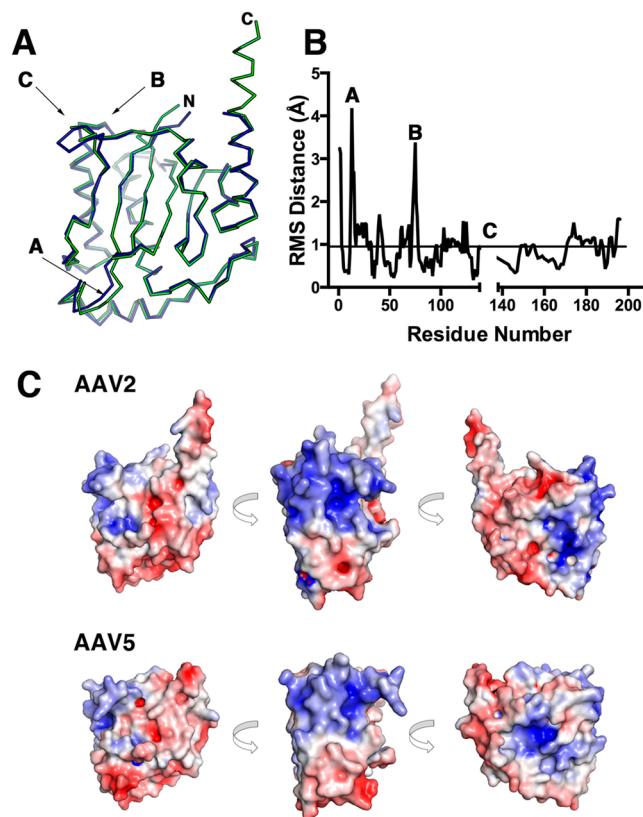
$$B = ([A]_x - [A]_{\text{DNA}})/([A]_{\text{FINAL}} - [A]_{\text{DNA}}) \quad (1)$$

where  $[A]_x$  represents the anisotropy measured at protein concentration  $x$ ,  $[A]_{\text{DNA}}$  is the anisotropy of free fluorescence DNA, and  $[A]_{\text{FINAL}}$  is the anisotropy at saturation. Data was fit to a single binding site model using the program PRISM6 (GraphPad). Each experiment was done in triplicate.

**AAV Infectious Particles Assay.** 293T cells were transfected with three plasmids using polyethylenimine (PEI): an AAV2 ITR-containing plasmid encoding a CAG-controlled GFP gene (pTRUF11), a helper plasmid expressing AAV2 Rep (wt or L193A-V196A cloned from the pHsRep68LV/15b) and Cap derived from pDG, and a third construct containing the adenovirus helper functions (HG11 plasmid).<sup>59,60</sup> The presence of the L193A–V196A double mutation was confirmed by sequencing (Eurofins). After 72 h, the cell supernatant was harvested and increasing volumes of supernatant were used to infect HeLa cells. The percentage of GFP-positive HeLa cells was determined 48 h postinfection by FACS (FACSCanto, BD Biosciences).

## RESULTS

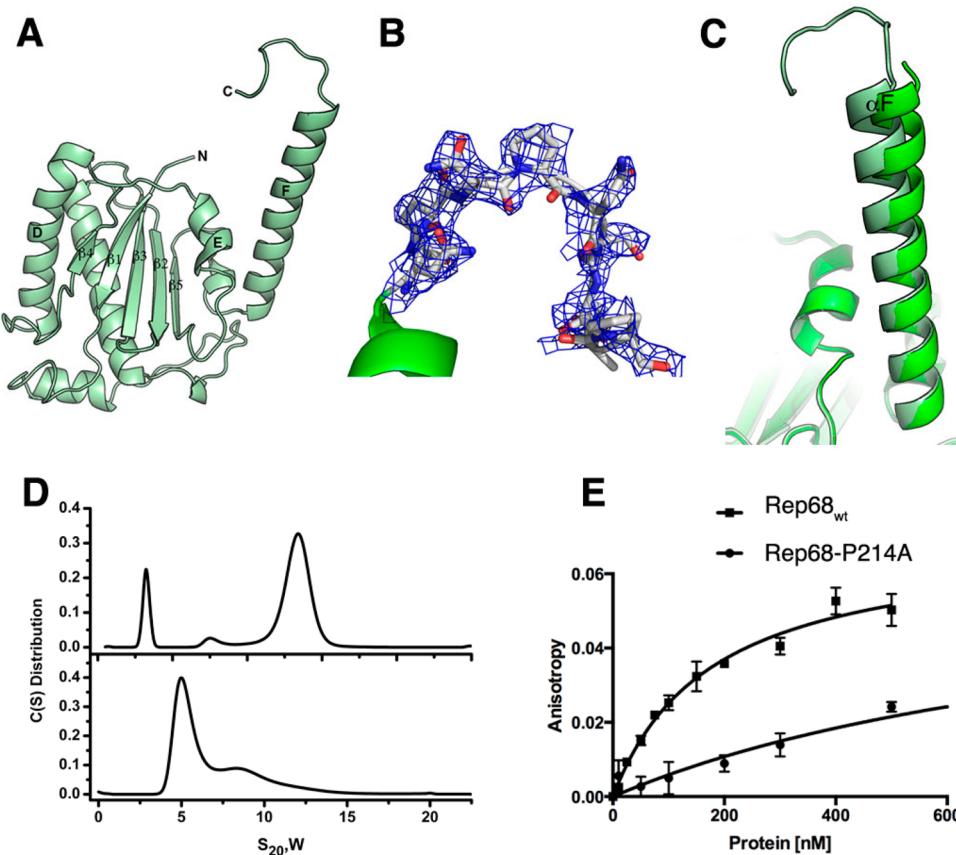
We opted to structurally characterize full-length AAV2 Rep68 using small-angle X-ray scattering and rigid-body modeling. A diagram of the different domains and protein constructs used in this study is shown in Figure 1A. We first determine the crystal



**Figure 3.** Comparison of AAV2 and AAV5 OBD structures. (A) Ribbon diagram of the superposition of the AAV2 (green) and AAV5 (blue) OBD structures. The N and C termini are marked. Regions that have the largest rmsd are indicated by arrows. (B) Plot of rmsd differences against residue number between AAV2 and AAV5 OBD. (C) Solvent-accessible electrostatic surface comparison of AAV2 and AAV5 OBDs showing three views of 90° counter-clockwise rotations. Positive regions are shown in blue, and negative, in red.

structure of its OBD including the region spanning the linker, followed by SAXS studies of Rep40 (SF3 helicase domain) to study the dynamic behavior of the C-terminal tail. Finally, to determine the relative orientation of its two functional domains, we determined the SAXS solution structure of Rep68.

**Structure of AAV-2 Rep OBD.** The OBD construct (residues 1–208) was initially obtained from limited proteolysis experiments.<sup>61</sup> The structure was solved by molecular replacement and refined to 2.3 Å resolution (Table 1). There are three molecules in the asymmetric unit: molecule A, spanning residues 1–138 and 143–206, molecule B, spanning residues 1–139 and 143–201, and molecule C, covering residues 1–131 and 141–191. The missing residues are all part of the DNA binding loop (L<sub>DB</sub>). Molecules A and B form a dimer interacting through their C-terminal helix F, whereas molecule C does not make any close contacts. The OBD structure is characterized by a central five-stranded antiparallel β-sheet flanked by six α-helices, three at each side of the sheet (Figure 1B). The topology of the β-sheet follows a 4–1–3–2–5 order, with helices E and F on one side of the sheet and helices B–D positioned on the opposite side (Figure 1C,D). There are two long loops: the DNA binding loop L<sub>DB</sub> connects β-strand 4 to α-helix E and loop L<sub>BC</sub> connects α-helices B and C and protrudes underneath α-helix F (Figure 1C). L<sub>DB</sub> was identified as being important in making specific contacts with the major groove of GCTC repeats and shows the highest



**Figure 4.** Structure of OBDL. (A) Ribbon diagram of the structure spanning residues 1–214. (B) Electron density of region 206–214 contoured at 0.9 $\sigma$ . The density fits most of the residues, but it suggests that the region is dynamic. (C) Superposition of OBD and OBDL structures illustrating the flexibility of helix F. (D) Comparison of the sedimentation velocity profiles between Rep68<sub>wt</sub> (top) and Rep68-P214A (bottom). (E) Comparison of binding isotherms of Rep68<sub>wt</sub> and Rep68-P214A. Fluorescent anisotropy binding assays were carried out with a 41-mer AAVS1 DNA labeled with fluorescein. Apparent dissociation constants were determined from triplicate experiments and fitted to a one binding site-specific binding model.

degree of variation among all AAV Rep isoforms.<sup>31,62</sup> This is a dynamic region, as shown by the lack of electron density in all three molecules found in the asymmetric unit. The C-terminal helix  $\alpha$ F is long, spanning residues 182–206, and protrudes at an approximately 45° angle from the plane of the  $\beta$ -sheet (Figure 1D). The residues at the end of helix F appear to be dynamic, as we could see only up to residue 201 in molecule B and to residue 191 in molecule C. Helix F is loosely packed against helix E and the loop connecting  $\beta$ S to  $\alpha$ F. The interactions that keep these three elements together are sparse, suggesting that helix F is flexible. Several dynamic regions in the AAV2 OBD can be observed by the superposition of the three molecules in the asymmetric unit (Figure 2A). The three molecules superimpose with an overall rmsd over 197  $C_{\alpha}$  of 0.62 Å. Two regions can be identified with larger than average values: region 1 includes residues 15–34 spanning helices A and B, and region 2 includes residues from L<sub>DB</sub>. Not surprisingly, based on the AAV5 OBD–RBS structure, both of these regions are involved in DNA interactions: region 1 is involved in recognition of the ITR hairpin stem 2, and L<sub>DB</sub> recognizes part of the GCTC repeat.<sup>31</sup> Thus, folding and stabilization of these elements must occur upon DNA binding. Two of the OBD molecules in the asymmetric unit form a dimer through interactions with residues present in  $\alpha$ -helix F, resembling a pseudocoiled coil (Figure 2B). These include L193, V196, H192, and S197 making hydrophobic and hydrogen-bond

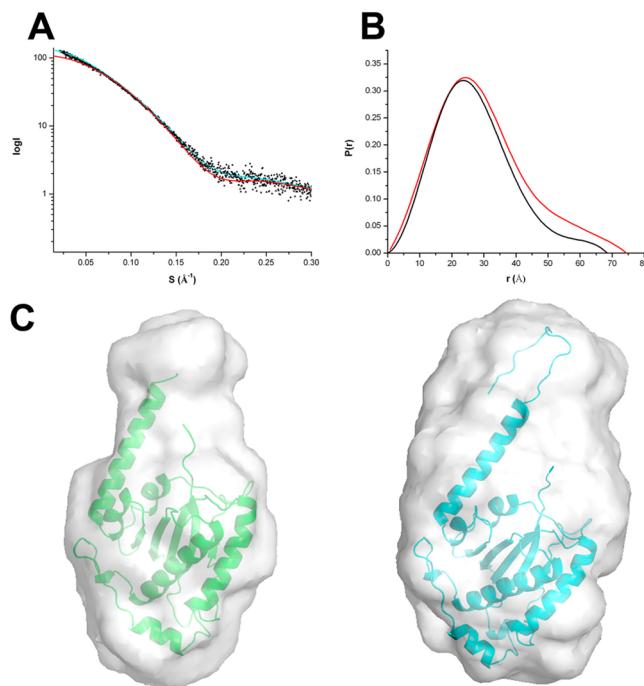
interactions (Figure 2C). This type of interface was also observed in the AAV5 OBD–RBS complex, which may suggest a functional role for this interface in the context of full-length Rep78/Rep68 proteins.<sup>31</sup> To determine if these interactions are functionally relevant in Rep68 function and their effect on the AAV life cycle, we examined the ability of several Rep68 mutants to oligomerize, to unwind dsDNA, and to support the production of infectious particles. Figure 2D shows that Rep68 mutants H192A, S197A, and L193AV196A have similar sedimentation velocity profiles as that of Rep68<sub>wt</sub>, forming the 13S peak that corresponds to heptameric/octameric ring species.<sup>45</sup> In addition, the mutants have similar unwinding activity as that of the wild-type protein (Figure 2E). Thus, the different mutants are functionally equivalent to Rep68<sub>wt</sub>. This conclusion is supported by the ability of the Rep68-L193AV196A double mutant to produce infectious viral particles as efficiently as wild-type Rep68 (Figure 2F). These results show that the dimer interactions are generated during crystal packing and are not involved in Rep68 oligomerization and function.

#### Structural comparison of the AAV2 and AAV5 OBD.

The Rep proteins from the majority of AAV serotypes show an overall homology with the AAV2 OBD of 95% (Figure S1). In contrast, AAV5 Rep has only 59% identity over the N-terminal 224 residues. However, the AAV2 and AAV5 OBD structures superimposed well, with an overall rmsd of 0.94 Å for 191

aligned  $C_{\alpha}$  carbons. As shown in Figure 3A,B, there are, nonetheless, three regions that show differences: region A (residues 12–15) includes the linker connecting  $\beta$ 1 to  $\alpha$ A, region B (74–77) spans the turn connecting  $\beta$ 3 to  $\alpha$ D, and region C (139–147) consists of L<sub>DB</sub>. Among all AAV serotypes, AAV5 and AAV8 have the largest sequence variations. For instance, AAV5 has a shorter L<sub>DB</sub> loop than AAV2, whereas in AAV8, the loop is longer by 2 residues. In addition, the AAV5 protein has four additional residues after position 198. Moreover, sequence conservation in the  $\alpha$ F region is low in AAV5, suggesting that this structure may indeed be shorter than in AAV2 (Figure S1). The overall electrostatic surface representations of the two proteins look similar, but subtle differences are visible. In both structures, one side of the surface is highly positively charged; however, in AAV2 OBD, the positive patch is wider and more intense. This is seen more explicitly if we compare the positive surface potential values between the two OBDs, with 2130.6 kcal/molq for AAV2 and 1720.4 kcal/molq for AAV5.<sup>63</sup> This area includes L<sub>DB</sub>, helix D, and the N-terminal half of  $\alpha$ -helix C. The former two interact with the GCTC repeats, whereas the latter is involved in the recognition of the RBE' hairpin in the ITRs.<sup>31</sup> Helix F is mostly negatively charged, but in AAV2, there is a slightly positive patch located in the first three turns that is not visible in AAV5 (Figure 3C). Because the structure of the AAV5 construct includes up to residue 197 and the sequence homology around this region is less than 80%, it is not apparent whether helix F in AAV5 will be as extended as in AAV2. Whether these structural and sequence variations found in L<sub>DB</sub> and the linker region account for the differences in nicking specificity observed for AAV2 and AAV5 Rep proteins remains to be determined and will require structures with DNA substrates along the nicking reaction pathway.<sup>64,65</sup>

**Crystal Structure of OBD-Linker.** We and others have described previously that the linker region connecting the OBD and the SF3 helicase domain of the large Rep proteins is pivotal for the assembly of multimeric Rep protein complexes.<sup>46,47</sup> To better characterize this region, we expressed, purified, and crystallized a construct containing the OBD and the linker that we termed OBDL, spanning residues 1–224. OBDL crystals diffracted to 2.6 Å and belong to space group P2<sub>1</sub>, with 11 molecules in the asymmetric unit. The electron density corresponding to the C-terminus of helix F and the linker region varies among the different molecules in the asymmetric unit; nevertheless, we were able to build a model of OBDL up to residue 214 (Figure 4). We did not detect any electron density from residues 215–224. The structure shows that helix F spans to residue 205 and is followed by a loop that makes a downward turn toward the core of the protein at proline 209 (Figure 4B). Superposition of OBD with OBDL shows that helix F is flexible and moves as a rigid body with respect to the core of the protein, pivoting around the loop that connects it with strand  $\beta$ 5. While the two OBD structures superimpose with an overall rmsd of ~0.8 Å, the two  $\alpha$ F helices superimpose with an rmsd of only 1.3 Å and appear at a different angle with respect to the core of the protein, with a difference between them of ~2.4° (Figure 4C). The OBDL structure also implies that the linker is not in an extended conformation but, instead, has a particular structure that is constrained by the two prolines in the linker. This particular conformation may bring the two domains together in the context of full-length Rep68. To test the importance of the proline residues, we mutated proline 214 to alanine and measured the effect of this on Rep68's



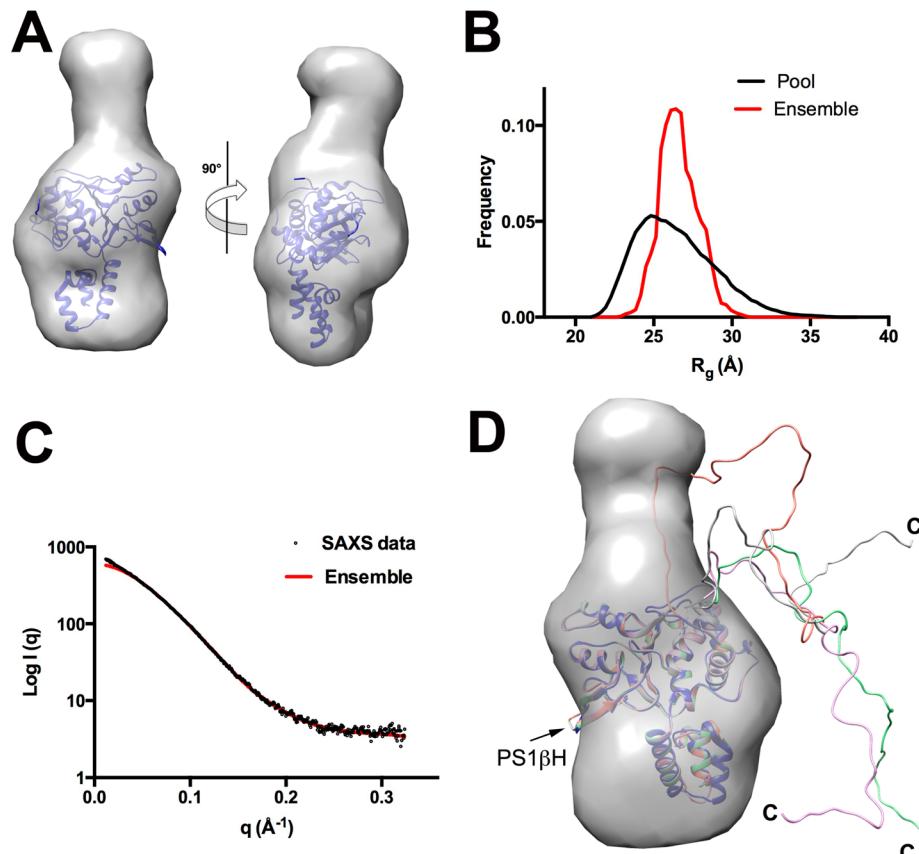
**Figure 5.** SAXS analysis for OBD and OBDL. (A) SAXS scattering data from OBD (blue) and OBDL (red) in solution. Superimposed is the scattering calculated from the atomic models for OBD (■) and OBDL (△). (B) Normalized pair distribution functions  $P(r)$  for OBD (black) and OBDL (red) calculated from the experimental scattering curves using the program PRIMUS. (C) Low-resolution envelopes of DAMMIN and GASBOR models for OBD (left) and OBDL (right) are superimposed with the X-ray structure of OBD and OBDL, respectively.

properties. Figure 4D shows that this mutation drastically alters the oligomeric behavior of Rep68 and affects its ability to form the 13S oligomer species (Figure 4D). Moreover, the mutation affects the ability of Rep68 to bind double-stranded DNA containing a Rep binding site (RBS). Figure 4E shows that the Rep68-P214A mutant binds DNA ~7 times less than wild-type protein. Taken together, our results show that the formation of the linker is critical for the function of Rep68.

**SAXS Studies of OBDL Validate the Conformation of the Linker.** To gain further structural information about the linker domain and to validate the OBDL X-ray structure, we performed solution studies on OBD and OBDL. The concentration of NaCl in the buffer was kept at 0.5–1 M to prevent formation of oligomers induced by the presence of the linker.<sup>46</sup> The calculated sedimentation coefficient for both constructs is ~2.0 S (Figure S2). The small difference in S value suggests that the linker region is not in an extended conformation. However, the smaller sedimentation velocity coefficient of OBDL implies a slightly more elongated shape. We subjected both constructs to SAXS studies under conditions that generated single species, as determined by sedimentation velocity. Scattering profiles from three different concentrations were collected, and a final scattering curve was obtained by merging the best curves using the program Almerge.<sup>66</sup> The scattering profiles for OBD and OBDL are shown in Figure 5A. The merged data was used to calculate both the radius of gyration ( $R_g$ ) and the  $P(r)$  distribution function as described in the Material and Methods (Figure 5B). *Ab initio* models were produced using the programs DAMMIN and GASBOR. For each protein, 10 independent models were generated and averaged with Damaver. The final models from

Table 2. Hydrodynamic Parameters

parameter	OBD (1–208)	OBDL (1–224)	Rep40 <sub>wt</sub> (225–536)	Rep68Y224AΔ (1–490)
S (s) <sup>a</sup>	2.1	1.97	nd <sup>d</sup>	nd
R <sub>g</sub> (nm) <sup>b</sup>	21.5 ± 0.3	23.2 ± 2.3	27.6 ± 3.8	38.9 ± 0.1
D <sub>max</sub> (nm) <sup>c</sup>	68.2	74.4 (104.96)	98.8	123.8
χ	1.0 <sup>e</sup>	1.9 <sup>e</sup>	0.5 <sup>f</sup>	0.8 <sup>g</sup>

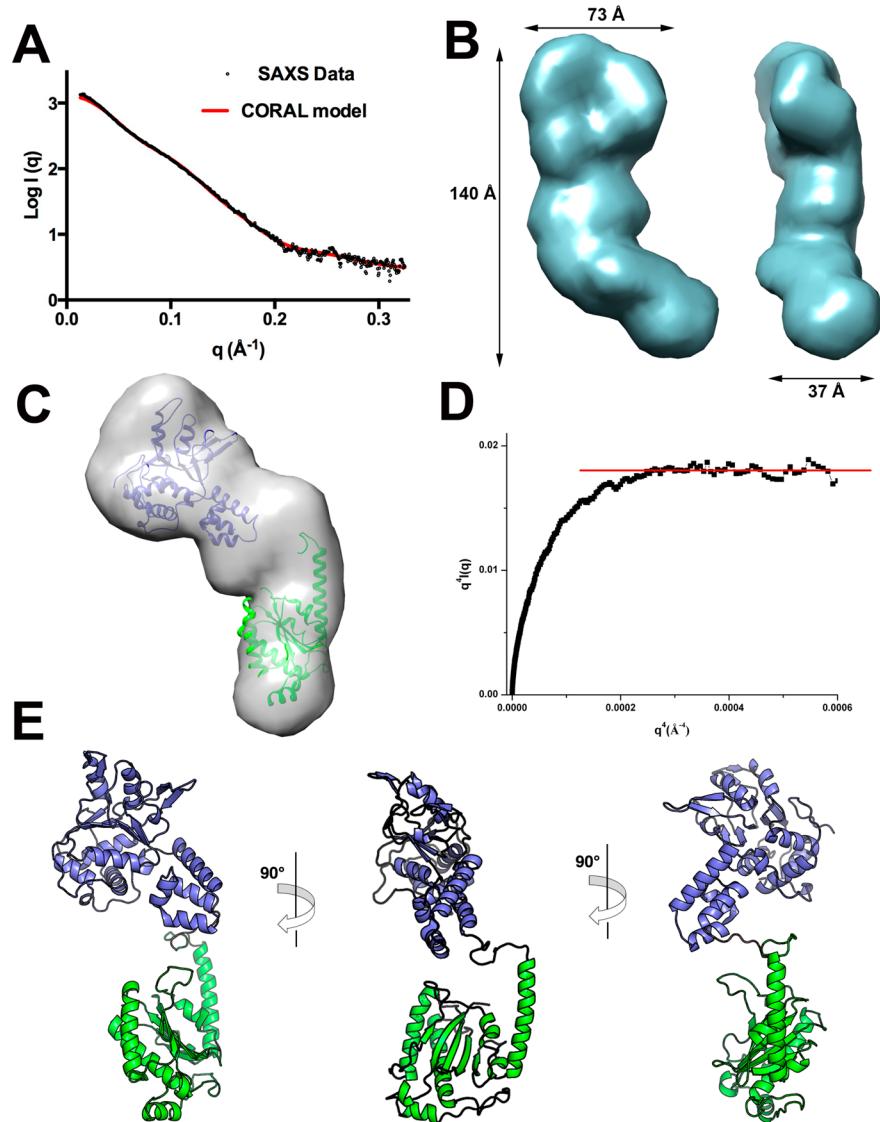
<sup>a</sup>Determined from sedimentation velocity data. <sup>b</sup>Determined from Guinier analysis. <sup>c</sup>Determined from P(r) analysis. <sup>d</sup>nd, not determined.<sup>e</sup>Experimental/model data fit using FoXS.<sup>24</sup> <sup>f</sup>Experimental/model data fit using EOM. <sup>g</sup>Experimental/model data fit using CORAL.

**Figure 6.** SAXS analysis of Rep40/Rep68 C-terminal tail. (A) Shape reconstruction of Rep40<sub>wt</sub> and the docked X-ray structure of Rep40. (B) Comparison of the R<sub>g</sub> distribution of the initial EOM generated pool (black line) and the selected model ensemble (red line). (C) Scattering curve of Rep40<sub>wt</sub> (black) and fit of the optimal EOM model ensemble (red). (D) Superposition of the ensemble generated model showing the preferred conformations of the C-terminal tails.

the two independent reconstructions were aligned using the program Supcomb, resulting in an overall normalized spatial discrepancy (NSD) value of 0.54, suggesting an excellent correlation between the two independent reconstructions.<sup>67</sup> The resulting GASBOR models for OBD and OBDL are shown in Figure 5C with the superimposed X-ray structures of the corresponding constructs. The calculated R<sub>g</sub> and D<sub>max</sub> values are similar for both models (Table 2). The OBDL GASBOR *ab initio* model shows an envelope that resembles a round cylinder, suggesting that, under our experimental conditions, the region between 208 and 224 is not extended but may resemble the overall conformation seen in the OBDL crystal structure (Figure 5C, right panel). To further corroborate the hypothesis that the linker is neither extended nor highly flexible, we carried out molecular dynamics using BILBOMD to determine if an ensemble with multiple conformations of the linker described the SAXS scattering profile of OBDL.<sup>57</sup> The results show that inclusion of more than one conformation

does not improve the fit to the experimental data significantly (data not shown).

**Modeling Flexibility of C-Terminal Tail of Rep68.** AAV Rep proteins Rep40 and Rep68 have a C-terminal tail of 46 residues that extends from the core of the helicase domain, as determined by limited proteolysis experiments.<sup>38</sup> To determine the overall flexibility of the C-terminal tail, we carried out SAXS experiments of Rep40<sub>wt</sub> at different concentrations, and data were analyzed as described previously. A GASBOR model of Rep40 clearly resembles the overall shape of the helicase domain and has a characteristic flat disk shape with an additional elongated density that is not represented in the X-ray structure, which should correspond to the C-terminal tail (Figure 6A). However, the dimensions are smaller than those expected from a fully extended C-terminal tail. To analyze the flexibility of the tail, we used EOM software to search for an optimal ensemble of conformations that will best fit the scattering data. This method generates a large pool of models

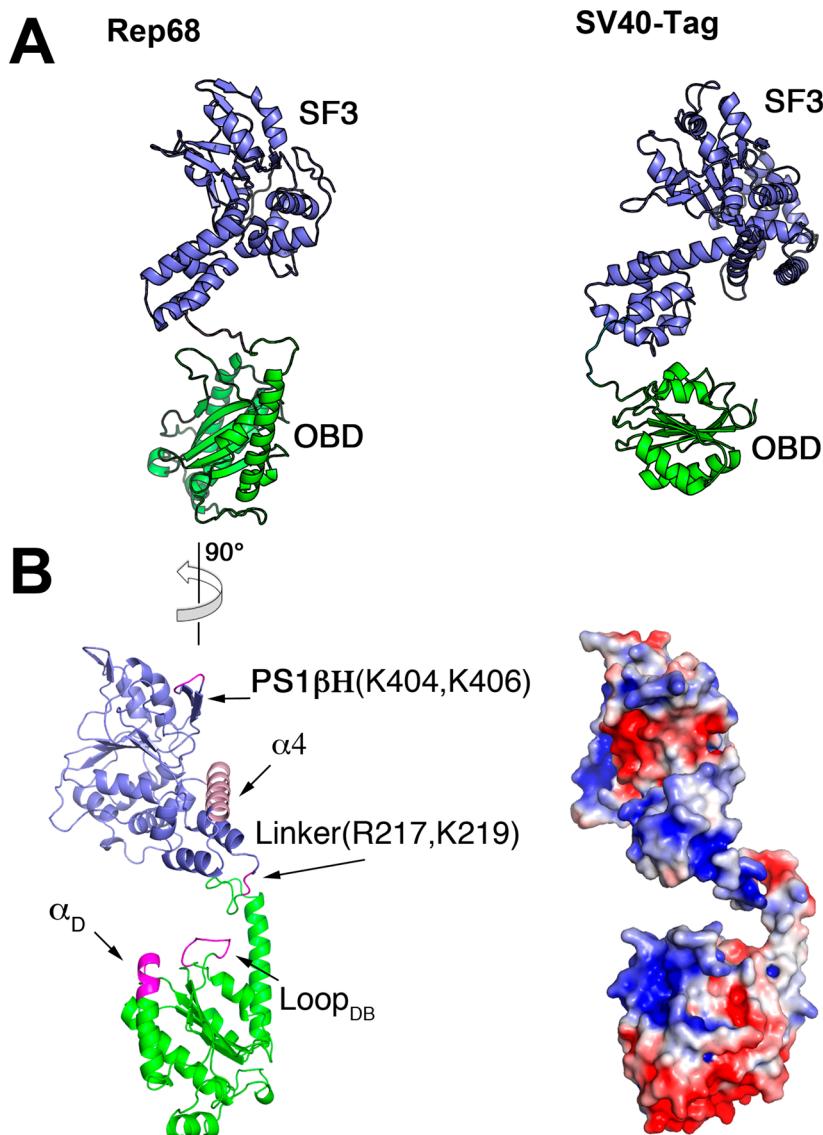


**Figure 7.** SAXS modeling of Rep68. (A) Fit of the experimental SAXS data (black circles) with the theoretical scattering profile obtained from CORAL model (red line) with a  $\chi$  value of 0.8. (B) Two views of the GASBOR averaged molecular envelope for Rep68 $\Delta$  showing the approximate dimensions in angstroms. (C) Docking of OBDL and Rep40 atomic structures to the GASBOR *ab initio* envelope. (D) Porod–Debye plot of Rep68Y224A $\Delta$  SAXS data (black squares) supporting a compact protein with little flexibility. Red line represents the linear plateau. (E) Three views of the CORAL rigid-body final model of Rep68Y224A $\Delta$ .

461 (10 000) with random conformations and uses a genetic  
462 algorithm to select an optimized ensemble of configurations  
463 that best describes the SAXS scattering data (Figure 6B).<sup>56,68</sup>  
464 The best ensemble consisted of four models that fit the  
465 experimental curve with a  $\chi$  of 0.5 (Figure 6C). However, the  
466 narrowness of the ensemble  $R_g$  distribution suggests that a  
467 limited number of conformations are preferred, with one  
468 configuration in particular accounting for about 50% of the  
469 population. Interestingly, all of the selected models have the  
470 C-terminal oriented toward the opposite site of the presensor 1  
471  $\beta$ -hairpin (PS1 $\beta$ H), a motif that is involved in DNA interactions  
472 during DNA translocation and unwinding (Figure 6D).<sup>30,69</sup>  
473 Thus, our study shows that the C-terminal tail, although flexible,  
474 has a preferred set of dominant conformations.

475 **SAXS-Based Structural Modeling of Rep68 Shows a**  
476 **Compact Conformation in Solution.** We previously showed  
477 that Rep68 has complex dynamic oligomeric behavior in  
478 solution and is present as a mixture of multiple oligomers,

including heptameric and octameric rings.<sup>45</sup> To obtain a homo- 479 geneous population of Rep68 monomers, we took advantage of 480 the Y224A mutation that affects its tendency to form multiple 481 oligomeric species.<sup>46</sup> The Rep68Y224A protein is present as a 482 monomer at concentrations less than 5 mg/mL and under high 483 salt conditions. Moreover, we have determined that this 484 mutation has helicase activity comparable to that of the wild- 485 type protein and does not cause any significant structural 486 change.<sup>70</sup> To facilitate model building and data interpreta- 487 tion, we used a truncated Rep68 (1–490) to eliminate the 488 C-terminal tail residues.<sup>38</sup> SAXS data was collected at 2, 3, and 489 4 mg/mL and processed as described in the previous sections 490 (Figure 7A). The GASBOR generated model produces an 491 elongated prolate ellipsoid that is slightly curved with 492 dimensions 140  $\times$  73  $\times$  37 Å<sup>3</sup> (Figure 7B). Individual atomic 493 structures of OBDL and Rep40 can be easily docked into the 494 envelope; in particular, one end of the particle has the shape of 495 a flat disk that resembles the GASBOR Rep40 *ab initio* model 496



**Figure 8.** Rep68 DNA interacting face. (A) Comparison of Rep68 model (left) and SV40-LTag (right). Colored equivalently are the OBD domain (green) and the helicase domain (blue). (B) (left) Rep68 structure showing the regions that interact with DNA colored in pink. Orientation of Rep68 is 90° counterclockwise from that in panel A. (right) Surface representation of Rep68 showing electropositive regions in blue and negative regions in red. View is in the same orientation as that on the left.

497 (Figure 7C). We performed rigid-body modeling using the  
498 X-ray structures of OBDL (1–214) and Rep40 (225–490)  
499 with CORAL.<sup>71</sup> The domain boundaries define an unstructured  
500 linker spanning residues 215–224. However, secondary  
501 structure prediction suggests that the region between 220 and  
502 224 could form a  $\alpha$ -helix extending into the first helix of the  
503 helicase domain.<sup>46</sup> Consequently, we generated multiple  
504 models that extended this helix at different positions. The  
505 Rep68 models fit the experimental data very well, with  $\chi$  values  
506 in the range of 1.1–0.8. All models show that the two domains  
507 are positioned such that the long axis of the helicase domain is  
508 almost perpendicular to the OBD, as shown in Figure 7E. The  
509 best model was obtained by extending the N-terminal helicase  
510 helix to residue 223. Two main observations can be drawn from  
511 this model. First, the orientation of the two domains results in a  
512 Rep68 structure where the motifs that interact with DNA  
513 in both the OBD ( $L_{DB}$  and helix C) and helicase domain  
514 ( $\beta$ -hairpin 1) are on the same face of the protein. Second, the  
515 two domains are closely positioned, making an extended linker

516 structure unlikely. To further support this conclusion, we  
517 assessed interdomain flexibility of Rep68 using two method-  
518 ologies. First, a Porod–Debye plot of the scattering data shows  
519 a plateau that is a signature for a compact molecule.<sup>72</sup> The data  
520 fits the linear region with a Porod coefficient of 4, again  
521 consistent with a compact molecule (Figure 7D). In addition,  
522 we carried out BILBOMID and determined that a single model  
523 fits the data equally as well as with multiple conformations  
524 (Figure 7A).

## DISCUSSION

525 Our results show new structural features that increase our  
526 knowledge of the architecture of AAV Rep proteins. The  
527 combined X-ray and SAXS studies show that the AAV2 Rep68  
528 linker region is partially structured, with helix F of the OBD  
529 extending until residue 209 and protruding from the main core  
530 at a 45° angle. The presence of two proline residues (209 and  
531 214) seems to impart a certain rigidity to the loop region,  
532 making a small turn before continuing toward the helicase  
533

534 domain (Figure 4). This conformation is important for Rep68  
 535 function, as mutation of one of the proline residues in the linker  
 536 (P214) is sufficient to prevent proper oligomerization and  
 537 DNA binding. In addition, part of the remaining linker may  
 538 extend the first helix of the helical domain of RepS2/40.<sup>46</sup> Both  
 539 of these structural features bring the OBD and helicase domain  
 540 closer to each other. However, although the linker is not  
 541 extended, it allows for a certain degree of flexibility. Our rigid-  
 542 body modeling generated different models with slightly  
 543 different conformations of the two domains that fit the  
 544 scattering data equally well (Figure S3). This suggests a certain  
 545 degree of conformational flexibility between the two domains.  
 546 Structurally, this flexibility originates from the linker, the OBD  
 547 helix F, and the helical bundle of the helicase domain.  
 548 Alignment of the OBD and OBDL structures shows that  
 549 helix F can pivot and move relative to the main core of the  
 550 domain (Figure 4C). Likewise, alignment of the three Rep40  
 551 molecules found in the asymmetric unit of the X-ray structure  
 552 (PDB ID: 1s9h) shows the helical bundle at different positions  
 553 relative to the AAA<sup>+</sup> domain (data not shown). This flexibility  
 554 is important to accommodate changes occurring upon DNA  
 555 binding and oligomerization. Moreover, analysis of the  
 556 flexibility of the C-terminal tail in Rep68 suggest that there is  
 557 a preference in the conformations acquired by the C-terminal  
 558 tail that positions it at the opposite end of the  $\beta$ -hairpin. This is  
 559 important in the context of formation of Rep68 oligomeric  
 560 rings because other conformations may produce steric clashes  
 561 and inhibit their formation.

562 Finally, our Rep68 model closely resembles the configuration  
 563 of the helicase and OBD in one of the molecules seen in the  
 564 X-ray structure of the SV40 large T antigen (SV40-LTag) in  
 565 complex with DNA.<sup>73</sup> This arrangement results in critical  
 566 residues that interact with DNA being located on one face of  
 567 the protein (Figure 8B). In our Rep68 model, these regions  
 568 include loop L<sub>DB</sub> and helix  $\alpha_D$  in the OBD, the helicase domain  
 569 (PS1 $\beta$  residues K404 and K406), and linker residues R217 and  
 570 K219 that have been shown to play a role in complex formation  
 571 and DNA binding in AAVS.<sup>47</sup> In addition, the SV40-LTag  
 572 structure shows that the helical bundle (Zn domain in SV40-  
 573 LTag) interacts with DNA, docking into the major groove.<sup>73</sup>  
 574 Our model shows that the position of the helical bundle is  
 575 similar, and it is intriguing to postulate that the conformation of  
 576 Rep68 is prealigned to interact with DNA in way similar to that  
 577 of SV40-LTag (Figure 8A).

578 In conclusion, we have completed the first structural  
 579 description of full-length AAV2 Rep68 protein, revealing the  
 580 orientation and relative position of its functional domains.  
 581 These observations provide new clues to explain its DNA  
 582 binding mode; however, our model does not discard possible  
 583 changes that should occur upon DNA binding and/or  
 584 oligomerization. These questions require future structural  
 585 studies of high-resolution structures of Rep68 oligomers  
 586 alone and in complex with DNA.

## 587 ■ ASSOCIATED CONTENT

### 588 Supporting Information

589 The Supporting Information is available free of charge on the  
 590 ACS Publications website at DOI: [10.1021/acs.biochem.5b00610](https://doi.org/10.1021/acs.biochem.5b00610).

591 Sequence and structural alignment of the AAV OBD  
 592 serotypes, sedimentation velocity profiles of AAV2 OBD  
 593 and OBDL, and superposition of different Rep68 models  
 594 obtained by independent CORAL runs ([PDF](#)).

## ■ AUTHOR INFORMATION

### Corresponding Author

\*Tel.: +1 (804) 628-1202; Fax: +1 (804) 628-3501; E-mail: [cescalante@vcu.edu](mailto:cescalante@vcu.edu).

### Funding

This work was supported by NIH grant R01-GM092854 (R.M.L.) and UK MRC grant 1001764 (R.M.L.).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We would like to thank members of NSLS beamlines X6a (Vivian Stojanoff, Jean Jakoncic, and Edwin Lazo) and X9 (Lin Yang and Marc Allaire) and ALS SIBYLS beamline (Kevin Dyer and Kathryn Burnett) for help in data collection. We thank Vishaka Santosh for helpful comments on the preparation of this manuscript.

## ■ ABBREVIATIONS

AAV, adeno-associated virus; OBD, origin binding domain; OBDL, origin binding domain with linker; HUH, histidine–hydrophobic–histidine; DBD, DNA binding domain; SF3, superfamily 3 helicase; SAXS, small-angle X-ray scattering; ORF, open reading frame; ITR, inverted terminal repeat; RBS, Rep binding site; ssDNA, single-stranded DNA; AAA<sup>+</sup>, ATPases associated with various cellular activities; IPTG, isopropyl- $\beta$ -D-thiogalactopyranoside; LB, Luria–Bertani; TCEP, tris(2-carboxyethyl)phosphine; NTA, nitrilotriacetic acid; MPD, 2-methyl-2,4-pentanediol; NSLS, National Synchrotron Light Source; R<sub>g</sub>, radius of gyration; PEI, polyethylenimine; FACS, fluorescent activated cell sorting

## ■ REFERENCES

- (1) Henckaerts, E., and Linden, R. M. (2010) Adeno-associated virus: a key to the human genome? *Future Virol.* 5, 555–574.
- (2) Weitzman, M. D., and Linden, R. M. (2012) Adeno-associated virus biology. *Methods Mol. Biol.* 807, 1–23.
- (3) Lusby, E., Fife, K. H., and Berns, K. I. (1980) Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA. *J. Virol.* 34, 402–409.
- (4) Samulski, R. J., Srivastava, A., Berns, K. I., and Muzyczka, N. (1983) Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* 33, 135–143.
- (5) Senapathy, P., Tratschin, J. D., and Carter, B. J. (1984) Replication of adeno-associated virus DNA. Complementation of naturally occurring rep- mutants by a wild-type genome or an ori- mutant and correction of terminal palindrome deletions. *J. Mol. Biol.* 179, 1–20.
- (6) Beaton, A., Palumbo, P., and Berns, K. I. (1989) Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. *J. Virol.* 63, 4450–4454.
- (7) Pereira, D. J., McCarty, D. M., and Muzyczka, N. (1997) The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J. Virol.* 71, 1079–1088.
- (8) Surosky, R. T., Urabe, M., Godwin, S. G., McQuiston, S. A., Kurtzman, G. J., Ozawa, K., and Natsoulis, G. (1997) Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J. Virol.* 71, 7951–7959.
- (9) Young, S. M., Jr., and Samulski, R. J. (2001) Adeno-associated virus (AAV) site-specific recombination does not require a Rep-dependent origin of replication within the AAV terminal repeat. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13525–13530.
- (10) McCarty, D. M., Ryan, J. H., Zolotukhin, S., Zhou, X., and Muzyczka, N. (1994) Interaction of the adeno-associated virus Rep

- 657 protein with a sequence within the A palindrome of the viral terminal  
658 repeat. *J. Virol.* 68, 4998–5006.
- 659 (11) Hauswirth, W. W., and Berns, K. I. (1977) Origin and  
660 termination of adeno-associated virus DNA replication. *Virology* 78,  
661 488–499.
- 662 (12) Spear, I. S., Fife, K. H., Hauswirth, W. W., Jones, C. J., and  
663 Berns, K. I. (1977) Evidence for two nucleotide sequence orientations  
664 within the terminal repetition of adeno-associated virus DNA. *J. Virol.*  
665 24, 627–634.
- 666 (13) Becerra, S. P., Rose, J. A., Hardy, M., Baroudy, B. M., and  
667 Anderson, C. W. (1985) Direct mapping of adeno-associated virus  
668 capsid proteins B and C: a possible ACG initiation codon. *Proc. Natl.  
669 Acad. Sci. U. S. A.* 82, 7919–7923.
- 670 (14) Becerra, S. P., Kocot, F., Fabisch, P., and Rose, J. A. (1988)  
671 Synthesis of adeno-associated virus structural proteins requires both  
672 alternative mRNA splicing and alternative initiations from a single  
673 transcript. *J. Virol.* 62, 2745–2754.
- 674 (15) Cassinotti, P., Weitzman, M., and Tratschin, J. D. (1988)  
675 Organization of the adeno-associated virus (AAV) capsid gene:  
676 mapping of a minor spliced mRNA coding for virus capsid protein.  
677 *Virology* 167, 176–184.
- 678 (16) Trempe, J. P., and Carter, B. J. (1988) Alternate mRNA splicing  
679 is required for synthesis of adeno-associated virus VP1 capsid protein.  
680 *J. Virol.* 62, 3356–3363.
- 681 (17) Muralidhar, S., Becerra, S. P., and Rose, J. A. (1994) Site-  
682 directed mutagenesis of adeno-associated virus type 2 structural  
683 protein initiation codons: effects on regulation of synthesis and  
684 biological activity. *J. Virol.* 68, 170–176.
- 685 (18) Sonntag, F., Schmidt, K., and Kleinschmidt, J. A. (2010) A viral  
686 assembly factor promotes AAV2 capsid formation in the nucleolus.  
687 *Proc. Natl. Acad. Sci. U. S. A.* 107, 10220–10225.
- 688 (19) Green, M. R., and Roeder, R. G. (1980) Transcripts of the  
689 adeno-associated virus genome: mapping of the major RNAs. *J. Virol.*  
690 36, 79–92.
- 691 (20) Green, M. R., Straus, S. E., and Roeder, R. G. (1980)  
692 Transcripts of the adenovirus-associated virus genome: multiple  
693 polyadenylated RNAs including a potential primary transcript. *J.  
694 Virol.* 35, 560–565.
- 695 (21) Laughlin, C. A., Westphal, H., and Carter, B. J. (1979) Spliced  
696 adenovirus-associated virus RNA. *Proc. Natl. Acad. Sci. U. S. A.* 76,  
697 5567–5571.
- 698 (22) Kyostio, S. R., Wonderling, R. S., and Owens, R. A. (1995)  
699 Negative regulation of the adeno-associated virus (AAV) P5 promoter  
700 involves both the P5 rep binding site and the consensus ATP-binding  
701 motif of the AAV Rep68 protein. *J. Virol.* 69, 6787–6796.
- 702 (23) Dutheil, N., Smith, S. C., Agundez, L., Vincent-Mistiaen, Z. I.,  
703 Escalante, C. R., Linden, R. M., and Henckaerts, E. (2014) Adeno-  
704 associated virus Rep represses the human integration site promoter by  
705 two pathways that are similar to those required for the regulation of  
706 the viral p5 promoter. *J. Virol.* 88, 8227–8241.
- 707 (24) Im, D. S., and Muzyczka, N. (1990) The AAV origin binding  
708 protein Rep68 is an ATP-dependent site-specific endonuclease with  
709 DNA helicase activity. *Cell* 61, 447–457.
- 710 (25) Chiorini, J. A., Weitzman, M. D., Owens, R. A., Urcelay, E.,  
711 Safer, B., and Kotin, R. M. (1994) Biologically active Rep proteins of  
712 adeno-associated virus type 2 produced as fusion proteins in  
713 Escherichia coli. *J. Virol.* 68, 797–804.
- 714 (26) Weitzman, M. D., Kyostio, S. R., Kotin, R. M., and Owens, R. A.  
715 (1994) Adeno-associated virus (AAV) Rep proteins mediate complex  
716 formation between AAV DNA and its integration site in human DNA.  
717 *Proc. Natl. Acad. Sci. U. S. A.* 91, 5808–5812.
- 718 (27) Urcelay, E., Ward, P., Wiener, S. M., Safer, B., and Kotin, R. M.  
719 (1995) Asymmetric replication in vitro from a human sequence  
720 element is dependent on adeno-associated virus Rep protein. *J. Virol.*  
721 69, 2038–2046.
- 722 (28) Lamartina, S., Ciliberto, G., and Toniatti, C. (2000) Selective  
723 cleavage of AAVS1 substrates by the adeno-associated virus type 2  
724 rep68 protein is dependent on topological and sequence constraints.  
725 *Journal of virology* 74, 8831–8842.
- (29) King, J. A., Dubielzig, R., Grimm, D., and Kleinschmidt, J. A. 726  
727 (2001) DNA helicase-mediated packaging of adeno-associated virus  
728 type 2 genomes into preformed capsids. *EMBO J.* 20, 3282–3291.
- (30) Yoon-Robarts, M., Blouin, A. G., Bleker, S., Kleinschmidt, J. A., 729  
730 Aggarwal, A. K., Escalante, C. R., and Linden, R. M. (2004) Residues  
731 within the B' motif are critical for DNA binding by the superfamily 3  
732 helicase Rep40 of adeno-associated virus type 2. *J. Biol. Chem.* 279, 732  
50472–50481.
- (31) Hickman, A. B., Ronning, D. R., Perez, Z. N., Kotin, R. M., and 733  
734 Dyda, F. (2004) The nuclease domain of adeno-associated virus rep  
735 coordinates replication initiation using two distinct DNA recognition  
736 interfaces. *Mol. Cell* 13, 403–414.
- (32) Yoon-Robarts, M., and Linden, R. M. (2003) Identification of 738  
739 active site residues of the adeno-associated virus type 2 Rep  
739 endonuclease. *J. Biol. Chem.* 278, 4912–4918.
- (33) Owens, R. A., Weitzman, M. D., Kyostio, S. R., and Carter, B. J. 741  
742 (1993) Identification of a DNA-binding domain in the amino terminus  
743 of adeno-associated virus Rep proteins. *J. Virol.* 67, 997–1005.
- (34) Hickman, A. B., Ronning, D. R., Kotin, R. M., and Dyda, F. 744  
745 (2002) Structural unity among viral origin binding proteins: crystal  
745 structure of the nuclease domain of adeno-associated virus Rep. *Mol.  
746 Cell* 10, 327–337.
- (35) Koonin, E. V., and Ilyina, T. V. (1993) Computer-assisted 748  
749 dissection of rolling circle DNA replication. *BioSystems* 30, 241–268.
- (36) Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., 750  
751 Moncalian, G., and Ton-Hoang, B. (2013) Breaking and joining single-  
751 stranded DNA: the HUH endonuclease superfamily. *Nat. Rev.  
752 Microbiol.* 11, 525–538.
- (37) Gonzalez-Prieto, C., Agundez, L., Linden, R. M., and Llosa, M. 754  
755 (2013) HUH site-specific recombinases for targeted modification of  
755 the human genome. *Trends Biotechnol.* 31, 305–312.
- (38) James, J. A., Escalante, C. R., Yoon-Robarts, M., Edwards, T. A., 757  
758 Linden, R. M., and Aggarwal, A. K. (2003) Crystal structure of the SF3  
758 helicase from adeno-associated virus type 2. *Structure* 11, 1025–1035.
- (39) Gorbaleyna, A. E., Koonin, E. V., and Wolf, Y. I. (1990) A new 760  
761 superfamily of putative NTP-binding domains encoded by genomes of  
761 small DNA and RNA viruses. *FEBS Lett.* 262, 145–148.
- (40) Yang, Q., Kadam, A., and Trempe, J. P. (1992) Mutational 763  
764 analysis of the adeno-associated virus rep gene. *J. Virol.* 66, 6058–  
6069.
- (41) Walker, S. L., Wonderling, R. S., and Owens, R. A. (1997) 766  
767 Mutational analysis of the adeno-associated virus type 2 Rep68 protein  
767 helicase motifs. *J. Virol.* 71, 6996–7004.
- (42) Wonderling, R. S., Kyostio, S. R., and Owens, R. A. (1995) A 769  
770 maltose-binding protein/adeno-associated virus Rep68 fusion protein  
770 has DNA-RNA helicase and ATPase activities. *J. Virol.* 69, 3542–3548.
- (43) Smith, R. H., and Kotin, R. M. (1998) The Rep52 gene product 772  
773 of adeno-associated virus is a DNA helicase with 3'-to-5' polarity. *J.  
773 Virol.* 72, 4874–4881.
- (44) Mansilla-Soto, J., Yoon-Robarts, M., Rice, W. J., Arya, S., 775  
776 Escalante, C. R., and Linden, R. M. (2009) DNA structure modulates  
776 the oligomerization properties of the AAV initiator protein Rep68.  
777 *PLoS Pathog.* 5, e1000513.
- (45) Zarate-Perez, F., Mansilla-Soto, J., Bardelli, M., Burgner, J. W., 779  
780 2nd, Villamil-Jarauta, M., Kekilli, D., Samso, M., Linden, R. M., and  
780 Escalante, C. R. (2013) Oligomeric properties of adeno-associated  
781 virus Rep68 reflect its multifunctionality. *Journal of virology* 87, 1232–  
782 1241.
- (46) Zarate-Perez, F., Bardelli, M., Burgner, J. W., 2nd, Villamil-  
784 Jarauta, M., Das, K., Kekilli, D., Mansilla-Soto, J., Linden, R. M., and  
785 Escalante, C. R. (2012) The interdomain linker of AAV-2 Rep68 is an  
786 integral part of its oligomerization domain: role of a conserved SF3  
787 helicase residue in oligomerization. *PLoS Pathog.* 8, e1002764.
- (47) Maggin, J. E., James, J. A., Chappie, J. S., Dyda, F., and Hickman, 789  
790 A. B. (2012) The amino acid linker between the endonuclease and  
790 helicase domains of adeno-associated virus type 5 Rep plays a critical  
791 role in DNA-dependent oligomerization. *Journal of virology* 86, 3337–  
792 3346.

- 794 (48) Otwinowski, Z., and Minor, W. (1997) Processing of X-ray  
795 diffraction data collected in oscillation mode. *Methods Enzymol.* 276,  
796 307–326.  
797 (49) Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T.  
798 R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter,  
799 N. K., and Terwilliger, T. C. (2002) PHENIX: building new software  
800 for automated crystallographic structure determination. *Acta Crystal-  
801 logr., Sect. D: Biol. Crystallogr.* 58, 1948–1954.  
802 (50) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools  
803 for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 60,  
804 2126–2132.  
805 (51) Schuck, P. (2000) Size-distribution analysis of macromolecules  
806 by sedimentation velocity ultracentrifugation and lamm equation  
807 modeling. *Biophys. J.* 78, 1606–1619.  
808 (52) Konarev, P. V., Petoukhov, M. V., Volkov, V. V., and Svergun,  
809 D. I. (2006) ATSAS 2.1, a program package for small-angle scattering  
810 data analysis. *J. Appl. Crystallogr.* 39, 277–286.  
811 (53) Svergun, D. (1992) Determination of the regularization  
812 parameter in indirect-transform methods using perceptual criteria. *J.  
813 Appl. Crystallogr.* 25, 495–503.  
814 (54) Svergun, D. I. (1999) Restoring low resolution structure of  
815 biological macromolecules from solution scattering using simulated  
816 annealing. *Biophys. J.* 76, 2879–2886.  
817 (55) Svergun, D. I., Petoukhov, M. V., and Koch, M. H. (2001)  
818 Determination of domain structure of proteins from X-ray solution  
819 scattering. *Biophys. J.* 80, 2946–2953.  
820 (56) Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and  
821 Svergun, D. I. (2007) Structural characterization of flexible proteins  
822 using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129, 5656–5664.  
823 (57) Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and  
824 flexibility within proteins as identified through small angle X-ray  
825 scattering. *Gen. Physiol. Biophys.* 28, 174–189.  
826 (58) Amaralunga, M., and Lohman, T. M. (1993) Escherichia coli rep  
827 helicase unwinds DNA by an active mechanism. *Biochemistry* 32,  
828 6815–6820.  
829 (59) Zolotukhin, S., Potter, M., Hauswirth, W. W., Guy, J., and  
830 Muzychka, N. (1996) A "humanized" green fluorescent protein cDNA  
831 adapted for high-level expression in mammalian cells. *J. Virol.* 70,  
832 4646–4654.  
833 (60) Grimm, D., Kern, A., Rittner, K., and Kleinschmidt, J. A. (1998)  
834 Novel tools for production and purification of recombinant  
835 adenoassociated virus vectors. *Hum. Gene Ther.* 9, 2745–2760.  
836 (61) Yoon, M., Smith, D. H., Ward, P., Medrano, F. J., Aggarwal, A.  
837 K., and Linden, R. M. (2001) Amino-terminal domain exchange  
838 redirects origin-specific interactions of adeno-associated virus rep78 in  
839 vitro. *Journal of virology* 75, 3230–3239.  
840 (62) Chiorini, J. A., Afione, S., and Kotin, R. M. (1999) Adeno-  
841 associated virus (AAV) type 5 Rep protein cleaves a unique terminal  
842 resolution site compared with other AAV serotypes. *J. Virol.* 73, 4293–  
843 4298.  
844 (63) Fogolari, F., Corazza, A., Yarra, V., Jalaru, A., Viglino, P., and  
845 Esposito, G. (2012) Blukes: a program for the analysis of the  
846 electrostatic properties of proteins based on generalized Born radii.  
847 *BMC Bioinf.* 13, S18.  
848 (64) Hewitt, F. C., and Samulski, R. J. (2010) Creating a novel origin  
849 of replication through modulating DNA-protein interfaces. *PLoS One*  
850 5, e8850.  
851 (65) Hewitt, F. C., Li, C., Gray, S. J., Cockrell, S., Washburn, M., and  
852 Samulski, R. J. (2009) Reducing the risk of adeno-associated virus  
853 (AAV) vector mobilization with AAV type 5 vectors. *Journal of virology*  
854 83, 3919–3929.  
855 (66) Franke, D., Kikhney, A. G., and Svergun, D. I. (2012)  
856 Automated acquisition and analysis of small angle X-ray scattering  
857 data. *Nucl. Instrum. Methods Phys. Res., Sect. A* 689, 52–59.  
858 (67) Kozin, M. B., and Svergun, D. I. (2001) Automated matching of  
859 high- and low-resolution structural models. *J. Appl. Crystallogr.* 34, 33–  
860 41.  
861 (68) Tria, G., Mertens, H. D. T., Kachala, M., and Svergun, D. I. 861  
862 (2015) Advanced ensemble modelling of flexible macromolecules 862  
863 using X-ray solution scattering. *IUCrJ* 2, 207–217. 863  
864 (69) Hickman, A. B., and Dyda, F. (2005) Binding and unwinding: 864  
865 SF3 viral helicases. *Curr. Opin. Struct. Biol.* 15, 77–85. 865  
866 (70) Bardelli, M. e. a. Manuscript in preparation. 866  
867 (71) Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body 867  
868 modeling of macromolecular complexes against small-angle scattering 868  
869 data. *Biophys. J.* 89, 1237–1250. 869  
870 (72) Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible 870  
871 and intrinsically unstructured biological macromolecules by SAS using 871  
872 the Porod-Debye law. *Biopolymers* 95, 559–571. 872  
873 (73) Chang, Y. P., Xu, M., Machado, A. C., Yu, X. J., Rohs, R., and 873  
874 Chen, X. S. (2013) Mechanism of origin DNA recognition and 874  
875 assembly of an initiator-helicase complex by SV40 large tumor antigen. 875  
876 *Cell Rep.* 3, 1117–1127. 876  
877 (74) Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2010) 877  
878 FoXS: a web server for rapid computation and fitting of SAXS profiles. 878  
879 *Nucleic Acids Res.* 38, W540–544. 879