Structure, Vol. 11, 1025–1035, August, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S0969-2126(03)00152-7

# Crystal Structure of the SF3 Helicase from Adeno-Associated Virus Type 2

**J. Anson James,[1,3] Carlos R. Escalante,[2,3] Miran Yoon-Robarts,[1] Thomas A. Edwards,[2] R. Michael Linden,[1,*] and Aneel K. Aggarwal[2,*]**

[1] Carl C. Icahn Center for Gene Therapy and Molecular Medicine

[2] Department of Physiology and Biophysics
Mount Sinai School of Medicine
1425 Madison Avenue
New York, New York 10029

## Summary

**We report here the crystal structure of an SF3 DNA helicase, Rep40, from adeno-associated virus 2 (AAV2). We show that AAV2 Rep40 is structurally more similar to the AAA$^+$ class of cellular proteins than to DNA helicases from other superfamilies. The structure delineates the expected Walker A and B motifs, but also reveals an unexpected "arginine finger" that directly implies the requirement of Rep40 oligomerization for ATP hydrolysis and helicase activity. Further, the Rep40 AAA$^+$ domain is novel in that it is unimodular as opposed to bimodular. Altogether, the structural connection to AAA$^+$ proteins defines the general architecture of SF3 DNA helicases, a family that includes simian virus 40 (SV40) T antigen, as well as provides a conceptual framework for understanding the role of Rep proteins during AAV DNA replication, packaging, and site-specific integration.**

## Introduction

Small DNA viruses have been widely exploited to study cellular processes such as DNA replication and transcription, protein translation, and, in several instances, tumor transformation. Their typically simple genome structures encode proteins that essentially function to (1) orchestrate cellular factors to provide a suitable environment for the virus lifecycle and (2) provide the necessary factors for viral replication, gene expression, and particle assembly. Consequently, the nonstructural proteins of these viruses have evolved to be multifunctional with respect to their inherent biochemical activities. Adeno-associated virus type 2 (AAV2), a human parvovirus, is a representative member of such DNA viruses.

First identified as a contaminant of adenoviral stocks, the name adeno-associated virus reflects the requirement of a helper virus coinfection, typically adenovirus, in order for AAV to undergo productive replication. However, it is now widely accepted that concomitant infection by other viruses, as well as particular cellular states, can provide helper functions in support of productive AAV replication. In the absence of favorable cellular conditions for productive replication (i.e., in healthy cells), AAV establishes a latent infection by site-specifically integrating its genome into a locus on human chromosome 19 (19q13.4), termed *AAVS1* (Kotin et al., 1990; Samulski et al., 1991; Dutheil et al., 2000). This ability to establish latency through targeted integration is an aspect of the AAV lifecycle that is unique among mammalian viruses. To date, AAV infection has not been associated with any apparent pathologies, making it an attractive vector for human gene therapy. This is especially notable in light of the fact that AAV infection is widespread; over 80% of the human population is estimated to be seropositive for anti-AAV antibodies (for a review, see Linden and Berns, 2000 and references therein).

The AAV genome is a 4.7 kb single-stranded DNA molecule that consists of two open reading frames (ORFs), *rep* and *cap*. Together, the ORFs encode four nonstructural and three capsid proteins, respectively. Consistent with their multifunctional role during the viral lifecycle, the products of the *rep* ORF, Rep78, -68, -52, and -40 (designated by their apparent molecular weight), possess a variety of biochemical activities (Im and Muzyczka, 1990). The three major functional domains are all present in the largest of the Rep proteins, Rep78. The Rep amino terminus possesses specific DNA binding and endonuclease activity, the central domain bears motifs necessary for ATPase and helicase activity as well as the nuclear localization sequence, and the carboxy-terminal Zn finger domain has been implicated in interacting with a myriad of cellular factors. The remaining Rep isoforms, Rep68, -52, and -40, are a combination of these functional domains arising from alternative splicing schemes and differential promoter usage within the *rep* ORF. Notably, all four isoforms possess the helicase domain, with Rep40 representing the minimum AAV helicase (see Figure 1A). The majority of biochemical studies focus on the ability of the Rep proteins to mediate viral origin-dependent DNA replication and site-specific integration. However, the only structural information available to date on Rep proteins is a crystal structure of the Rep origin interaction domain of human AAV serotype 5 (Hickman et al., 2002). The lack of structural information has limited the molecular assessment of the AAV lifecycle, including the assembly pathways attributed to ATPase and helicase activities (King et al., 2001), as well as the mechanism underlying site-specific integration.

DNA helicases have been found to be involved in a range of cellular processes, including DNA replication, repair, transcription, translation, and viral packaging (reviewed by Lohman and Bjornson, 1996). Typically, the energy required for the helicase-mediated unwinding of DNA is provided by NTP hydrolysis. Helicases possess a number of conserved motifs necessary for this ATPase function and translocation along the DNA/RNA molecule. A general classification scheme was proposed by Gorbalenya and Koonin (1993) based on the organization of these functional motifs. As a result, five major superfamilies were identified, designated SF1–SF5. The

*Correspondence: michael.linden@mssm.edu (R.M.L.), aggarwal@inka.mssm.edu (A.K.A.)

[3] These authors contributed equally to this work.

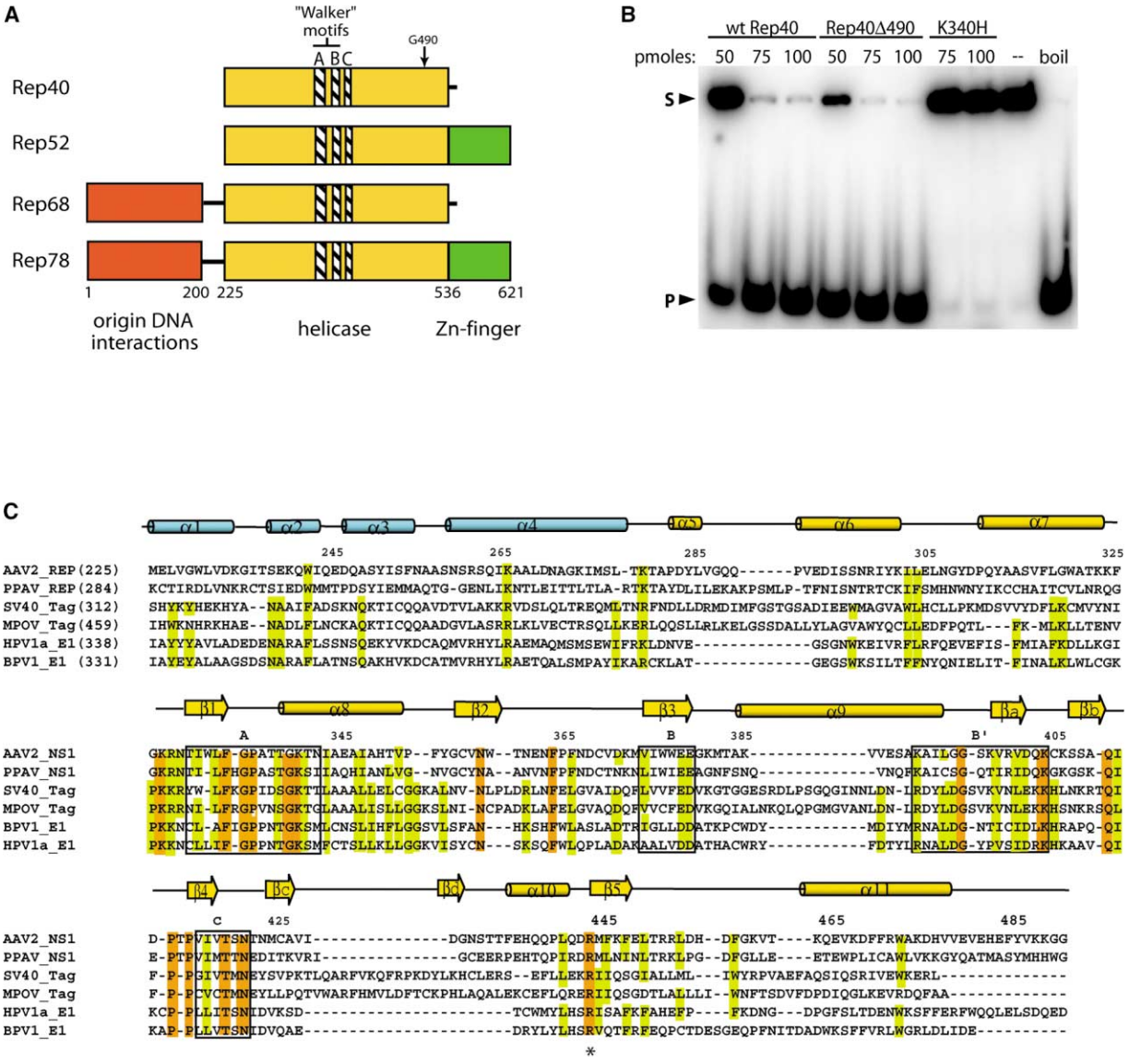


Figure 1. Schematic Representation of the AAV2 Rep Proteins

(A) A schematic diagram of the AAV2 Rep proteins, with N-terminal origin interaction domain (red), helicase domain (gold), and Zn finger domain (green). Dashed rectangles indicate Walker A, Walker B, and motif C. The position of the arrow shows the truncated form (Rep40Δ490) used for crystallization.

(B) Helicase assay. Rep40, Rep40Δ490, and Rep40Δ490 bearing a mutation of the Walker A lysine (K340H) were titrated at various concentrations on partially duplex M13 DNA substrates. Helicase assays were performed as described in Experimental Procedures. The amount of each Rep variant is indicated in pmoles. S, substrate; P, displaced single-stranded product; −, no protein; boil, heat-denatured substrate.

(C) Structure-based sequence alignment of the helicase domains of representative members of the SF3 family. Shown above the alignment is the relative location of α helices and β strands in the Rep40 structure; the N-terminal domain is colored blue, and the C-terminal α/β domain is colored gold. Boxed regions correspond to the SF3 signature motifs. The arginine finger residue is marked with an asterisk. Included in the comparison are: AAV2_Rep, adeno-associated virus 2 Rep40 protein; PPAV_NS1, porcine parvovirus NS1 protein; SV40_TAg, simian virus 40 large T antigen; MPOV, mouse polyomavirus; HPV1a_E1 human papillomavirus type 1a E1 protein; BPV1_E1 bovine papillomavirus type 1 E1 protein. Green, lightly conserved regions; brown, absolutely conserved regions.

AAV Rep helicase falls into the SF3 class, as do many viral helicases including that of papillomaviruses, poliovirus, and simian virus 40 (SV40). Members of the SF3 family contain four highly conserved regions within a short stretch of approximately 100 amino acids: motifs A, B (also known as Walker motifs A and B), B′, and C essentially make up the core of the helicase active site consisting of an NTP binding site, divalent metal cation coordination site, and sensor 1 site, respectively. SF3 helicases have been shown to form hexamers or double hexamers, and exhibit 3′ ⇒ 5′ NTP-dependent helicase activity.

While much is known about the biological activity of several SF3 helicases (particularly SV40 T antigen), a detailed assessment of the DNA unwinding mechanism by SF3 helicases has been hindered by the lack of high-

resolution structural data of a representative from this group. Here, we present the high-resolution structure of AAV2 Rep40, an SF3 helicase. We find that Rep40 is structurally related to the AAA$^+$ (*A*TPases *a*ssociated with diverse cellular *a*ctivities extended family) of cellular ATPases, and find that Rep40 possesses a novel AAA$^+$ module predicted to be shared by other viral SF3 helicases.

## Results and Discussion

### Structure Determination

We determined the structure using the multiwavelength anomalous diffraction (MAD) method (Hendrickson, 1991). Initially, we obtained crystals of full-length Rep40 (amino acids 225–536 relative to the Rep68 sequence), but they diffracted weakly. Based on secondary structure and proteolytic analyses, which suggested an unstructured C terminus, we expressed a shorter version of Rep40 (Rep40Δ490). Rep40Δ490 possesses helicase activity equal to the full-length Rep40 (see Figure 1B), and this activity, despite the requirement for relatively high concentrations of the protein, is attributable to Rep, because a mutant in which the NTP binding site was disrupted no longer displayed activity. Crystals of this shorter construct (both native and a selenomethionine variant) diffracted to 2.4 Å resolution with synchrotron radiation. The crystals belong to space group $p6_5$, with three molecules (A, B, and C) per asymmetric unit (a.u.). The refined model includes residues 225–490 for molecule A and molecule B, residues 225–288, 296–324, and 442–490 for molecule C, and a total of 480 water molecules (Table 1). Each molecule assembles into a helical filament along the $6_5$ crystallographic axis.

### Overall Structure

Rep40 is bimodular; a small helical bundle (α1–α4; residues 225–279) at the N terminus transitions into a large α/β domain (residues 298–490) at the C terminus (Figure 2A). The N-terminal domain is conformationally variable, located at different positions with respect to the α/β domain, between molecules A, B, and C. The α/β domain is the site of ATP binding and is composed of a central five-stranded parallel β sheet (β1–β5) that is flanked by four helices on one side (α6–α8 and α11) and two on the other (α9 and α10). This domain has the classical α/β topology of P loop NTPases, with Walker A and B motifs mapping to the C-terminal end of the central β sheet (β1–β5). The connectivity and the order of β strands (β5β1β4β3β2) establishes a similarity with the AAA$^+$ superfamily of P loop ATPases, distinct from SF1, SF2, and DnaB-like helicases that possess a more general RecA fold (Caruthers and McKay, 2002; Figure 2B). However, the Rep40 AAA$^+$ domain is novel. Instead of a bipartite arrangement in which a C-terminal subdomain drapes over the α/β core, the Rep40 AAA$^+$ domain is unimodular with the C-terminal segment intimately associated with the α/β core (Figure 3A). Furthermore, the α/β core is embellished by two β hairpins that protrude into the solvent, one connecting α9 to β4 and the other linking β4 to α10. Together, the unimodular arrangement and the hairpins lend a "rectangular" (instead of the more common "triangular") look to the Rep40 AAA$^+$ domain. We will refer to this new domain as vAAA$^+$, for viral AAA$^+$, to differentiate it from the "classical" AAA$^+$ domain.

### Rep40, an SF3 Helicase, Defines a Novel AAA$^+$ Architecture

The closest structural homologs to Rep40 are AAA$^+$ proteins. Representative members of AAA$^+$ proteins include the motor protein RuvB, responsible for Holliday junction migration (Putman et al., 2001; Yamada et al., 2001); N-ethylmaleimide-sensitive fusion protein (NSF-D2), which plays a role in membrane fusion (Lenzen et al., 1998; Yu et al., 1998); and the protease-associated ATPase, HslU (Bochtler et al., 2000; Song et al., 2000; Sousa et al., 2000; Wang et al., 2001a, 2001b). AAA$^+$ proteins are structurally defined by two subdomains: an N-terminal α/β ATPase domain encompassing the conserved Walker A, Walker B, and sensor 1 motifs, and a C-terminal helical domain containing the sensor 2 region (Figure 3A). The Rep40 central β sheet and helices α6-α7 and α9-α10 superimpose on equivalent segments in RuvB (189 Ca's), NSF-D2 (127 Ca's), and HslU (134 Ca's), with rmsd's of ~3.5 Å, 4.1 Å, and 3.8 Å, respectively. The SF3 signature A, B, and C motifs map to one end of the Rep40 central β sheet. Specifically, the A, or Walker A motif (GPATTGKT), spans the loop following β1 and the beginning of α9; B, with an atypical Walker B motif (VIWWEE), covers the end of β3 and the loop leading to α9; and C, with an invariant Asn residue, maps to the sensor 1 region on β4.

A major difference between the Rep40 core domain and that of AAA$^+$ family members is the presence of an additional subdomain in the latter which is necessary for ATPase activity. This subdomain is predominantly helical in AAA$^+$ proteins and is the site of a sensor 2 residue (generally an arginine, on a sensor 2 helix) that extends to the Walker A motif in the α/β core and interacts with the bound ATP. The interaction not only provides binding energy for ATP, but also translates ATP hydrolysis into interdomain movement. In NSF-D2 and HslU, for instance, this subdomain sits atop the α/β core and "sandwiches" the adenine base of the bound ATP. Rep40 lacks this C-terminal subdomain: residues subsequent to β5 form a long α helix (α11) and a segment without secondary structure, both of which are closely packed against the α/β core. The lack of a corresponding sensor 2 region or helix in Rep40 is compensated for by a spatially overlapping loop connecting strand β5 to β11 (residues 449–463).

To gain insight into ATP binding and catalysis, we superimposed Rep40 onto HslU, structures of which are available bound to ADP, ATP, and AMP-PNP. The superposition of the Walker A, Walker B, and sensor 1 regions allows for a modeling of ATP in the Rep40 active site (Figure 3B). Walker A residues Lys340 and Thr341 are predicted to form hydrogen bonds to an ATP molecule much like the Walker A motif in NSF-D2 and HslU. The catalytic core of the enzyme resides in residues of the Walker B motif. In most AAA$^+$ proteins, the motif is characterized by a conserved aspartate that is involved in coordination of a metal ion, followed by a

Table 1. Data Collection and Refinement Statistics

| Data Collection | Se-Edge | Se-Peak | Se-Remote | Native |
|---|---|---|---|---|
| Wavelength (Å) | 0.9812 | 0.9810 | 0.9583 | 0.9786 |
| Resolution (Å) | 2.6 | 2.6 | 2.8 | 2.4 |
| Number of reflections measured | 128,785 | 207,984 | 153,004 | 203,768 |
| Unique | 26,828 | 26,361 | 21,125 | 35,212 |
| (%) Completeness | 97.9 (91.6) | 96.5 (91.3) | 96.7 (99.0) | 96.8 (84.0) |
| $R_{merge}$ (%)[a,b] | 6.5 (39.1) | 7.5 (34.9) | 7.0 (27.2) | 5.9 (37.3) |
| $I/\sigma$ | 12.7 (3.1) | 16.2 (3.2) | 21.3 (3.9) | 16.8 (1.42) |
| **Mad Phasing Statistics** | | | | |
| Number of sites | 11 | 11 | 11 | |
| FoM (centric/acentric) 3.5 Å | 0.808/0.701 | | | |
| FoM (DM) 2.4 Å | 0.924 | | | |
| Phasing power | 2.58 | 2.18 | 1.57 | |
| **Refinement Statistics** | | | | |
| Resolution range (Å) | | | | 20–2.4 |
| Reflections, $F > 2\sigma$ (F) | | | | 29,685 |
| $R_{crys}$ (%) | | | | 21.92 |
| $R_{free}$ (%) | | | | 28.90 |
| Nonhydrogen atoms | | | | |
| Protein | | | | 5616 |
| Water | | | | 480 |
| Rms deviations | | | | |
| Bonds (Å) | | | | 0.00733 |
| Angles (°) | | | | 1.28 |
| Average B factor (Å$^2$) | | | | 51.34 |

Values in parentheses are for the outermost shell.
$R_{merge} = \Sigma|I - \langle I \rangle|/\Sigma|$, where I is the integrated intensity of a given reflection.
FoM = mean figure of merit at 3.5 Å.
FoM (DM) = overall mean figure of merit at 2.4 Å after density modification.
$R_{crys} = \Sigma||F_o| - |F_c||/\Sigma|F|$.
$R_{free}$ was calculated using 10% of data excluded from the refinement.

glutamate that is believed to be the catalytic base. Both residues are found at the end strand of β3. Surprisingly, the main chain atoms of the equivalent acidic residues in Rep40 (VIWWEE) are displaced by as much as ~2 Å from those in HslU and NSF-D2. This displacement reflects the necessity of a conformational shift in Rep40 strand β3, in order to bring these residues into optimal alignment for catalysis. We find that strands β1, β4, and β5 overlay with a low rmsd of ~0.6 Å onto equivalent strands in NSF-D2 and HslU, but strands β2 and β3 superimpose with a high rmsd of about ~4 Å. This discrepancy may be caused by the absence of an α helix present in most AAA$^+$ proteins, connecting β2 to β3. Instead, Rep40 has a small quasihelical segment of 4 residues that provides only limited packing interactions, causing β2 and β3 to "fall off" the plane of the β sheet. To bring β3 and the glutamates into position for catalysis may thus require ATP and/or DNA binding. In AAA$^+$ proteins, the sensor 1 region plays an important role in distinguishing ATP from ADP via interactions between a polar residue on strand β4 (generally, Thr or Asn) and the γ-phosphate of the bound ATP. Similarly, an invariant Asn (Asn421) in motif C or the sensor 1 region in Rep40 is poised to form hydrogen bonds with the γ-phosphate of the modeled ATP (but not ADP). The lack of a sensor 2 helix in Rep40 is compensated for by a 13 residue loop (see above) containing hydrophobic residues (LDHDF) that could sandwich the adenine base of the bound nucleotide. We note that several SF3 helicases contain a similar LxxxHy motif at this position, where Hy is a bulky hydrophobic residue, generally a phenylalanine or a tryptophan, and may define a new sensor 2 loop (Figure 3B).

### A Putative Rep40 Hexamer

A hallmark of AAA$^+$ proteins is that they function as oligomers, in most cases by forming hexameric rings (Vale, 2000). The ability to multimerize is also a characteristic of AAV Rep, and it is presumed that different oligomeric states exist in support of Rep's various biochemical functions throughout the viral lifecycle. The largest of the Rep proteins, Rep78, has been shown to multimerize both in vitro and in vivo, and preferentially forms hexamers in the presence of AAV origin-derived DNA (Smith et al., 1997). AAA$^+$ proteins that contain a single "AAA$^+$ module," as we observe in the Rep40 structure, are monomeric in solution but oligomerize upon substrate and nucleotide binding. Such induced oligomerization has been demonstrated for AAA$^+$ proteins katanin (Hartman and Vale, 1999) and DnaA (Erzberger et al., 2002). Interestingly, Rep52, differing from Rep40 only in the presence of the C-terminal Zn finger domain, has been shown to be monomeric in solution despite incubation with single- or double-stranded DNA (Smith and Kotin, 1998).

To gain insights into the likely hexameric state of Rep proteins, we generated a model based on superposition with hexameric HslU. The Rep40 hexamer model is "mushroom" shaped with the vAAA$^+$ domain forming the dome and the N-terminal helical domain forming the

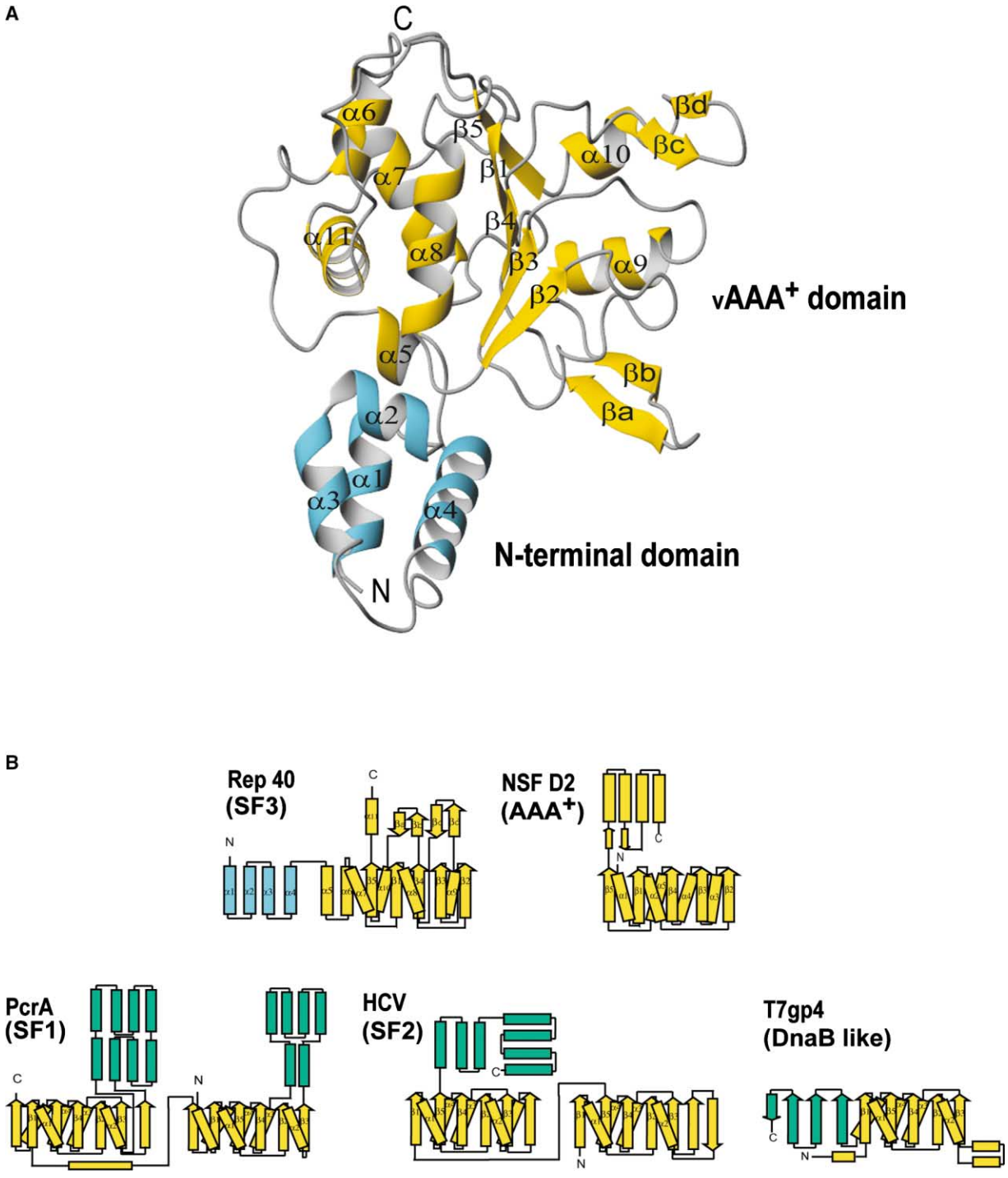


Figure 2. Structure and Topology of AAV2 Rep40

(A) Ribbon representation of the AAV2 Rep40 structure. The N-terminal α-helical domain is shown in blue. The C-terminal α/β (vAAA$^+$) domain is colored gold. The α helices (α1–α11) are labeled sequentially from the N to the C terminus. β strands are labeled according to their belonging to the central β core (β1–β5) or to the β hairpins (βa–βd).

(B) Topology diagrams of representative members of helicase and AAA$^+$ superfamilies. Gold, NTPase cores; green, extra functional domains; blue, Rep40 N-terminal domain. NSF-D2, D2 domain of N-ethylmaleimide-sensitive fusion protein; PcrA, DNA helicase from *B. stereothermophilus*; HCV, hepatitis C virus nonstructural protein 3 protease-helicase; T7gp4, bacteriophage T7 gene 4 primase-helicase.

root (Figure 4A). The ring has an overall diameter of ~115 Å with an internal channel ~18 Å in diameter. These dimensions are strikingly similar to those obtained by electron microscopy for SV40 T antigen (SV40 TAg), which reveals a "ring-like" particle with an outer diameter of 120 Å and a 20 Å inner channel (Valle et al.,

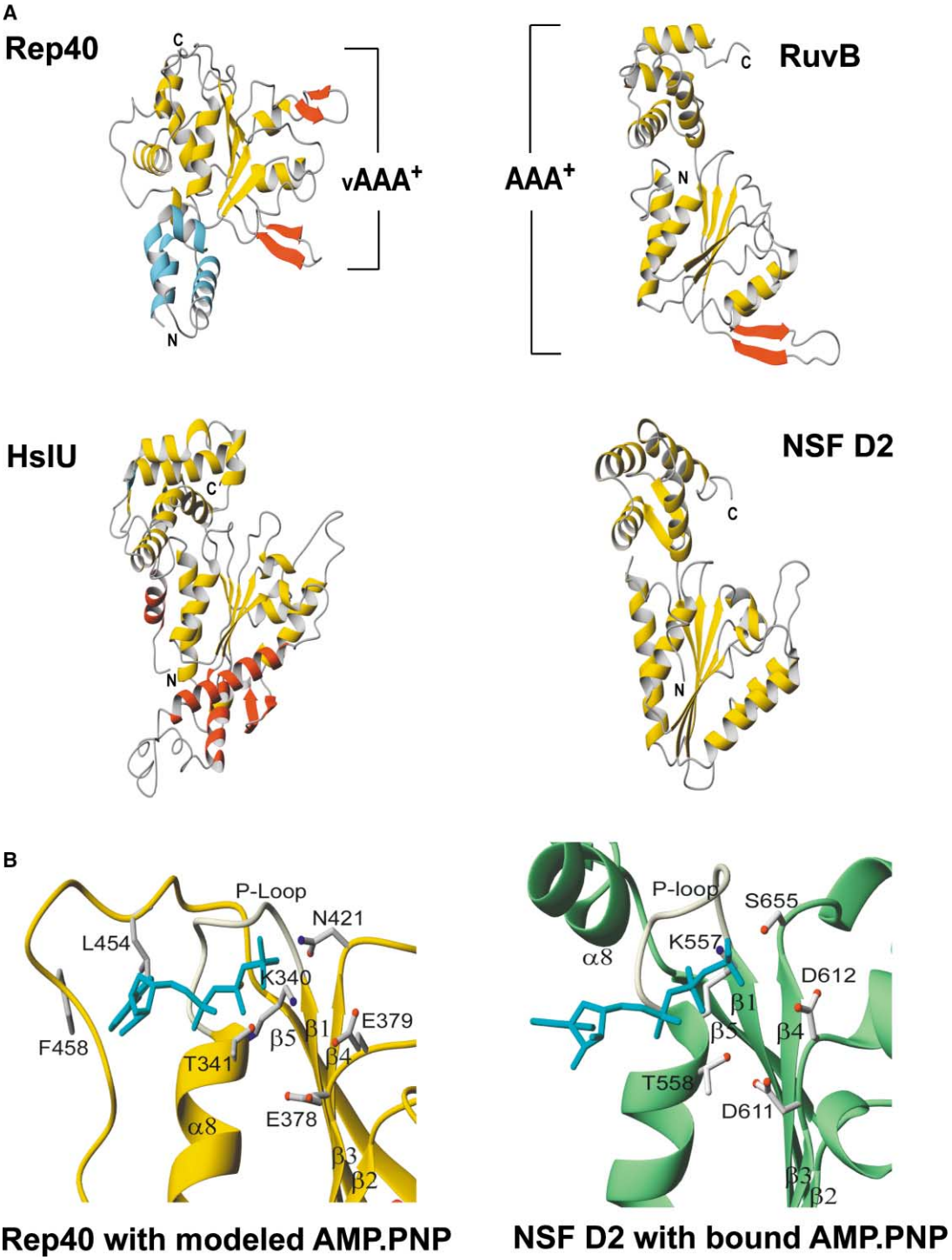


Figure 3. Structural Comparison of AAV2 Rep40 and AAA$^+$ Proteins

(A) Ribbon representation of AAV2 Rep40 and the AAA$^+$ domains from: *T. thermophilus* RuvB, *H. influenzae* HslU, and Chinese hamster ovary NSF-D2. The AAA$^+$ modules are colored gold. Extra embellishments on the minimal AAA$^+$ module are colored red. The N-terminal domain of Rep40 is shown in blue.

(B) Comparison of Rep40 and NSF-D2 active sites. Rep40 (gold) modeled with AMP-PNP (cyan). NSF-D2 (green) complexed with AMP-PNP (PDB code 1D2N). The P loop is colored light yellow. Active site residues are shown as all-bonds representation.

2000; VanLoock et al., 2002). The electrostatic potential of the hexamer shows that most of the surface is negatively charged, but the inner channel has an electropositive character in accordance with its possible role in ssDNA translocation (Figure 4A). Bearing in mind the limitations of modeling, several interesting observations

emerge. First, the oligomerization interface is largely due to interactions between the vAAA$^+$ domains. Specifically, the interface is generated by helices $\alpha$9 and $\alpha$10, $\beta$ hairpin 2, and the loop that precedes strand $\beta$1, while the adjoining surface is made up of helix $\alpha$8, strand $\beta$2, and loops preceding helix $\alpha$11 and $\beta$ hairpin 1. Second, the oligomerization interface is small when compared to NSF-D2 and HslU and may explain the existence of Rep52 and Rep40 as monomers in solution (Smith and Kotin, 1998, and data not shown). In NSF-D2 hexamer, for instance, the interfaces between subunits are made up of interactions between both the $\alpha/\beta$ cores as well as a C-terminal subdomain, resulting in a total of $\sim$2500 Å$^2$ of solvent-accessible surface (SAS) buried per protomer. Because Rep40 lacks this equivalent C-terminal subdomain, the potential SAS is limited to $\sim$1000 Å$^2$. Third, the increased tendency of Rep68/Rep78 to oligomerize suggests that regions outside the central helicase domain contribute to stable oligomerization. Indeed, a region within the amino-terminal origin interaction domain identified as a 3,4-heptad repeat (residues 162–181 in Rep68/78) has been implicated in contributing to Rep-Rep associations (Davis et al., 1999; Smith et al., 1997). Helicase assays performed with Rep40 demonstrate activity at only relatively high concentrations of the protein relative to Rep68 (compare Figure 1B to Figure 4D). Indeed, a titration curve of the helicase activity against Rep40 concentration shows a sigmoidal, rather than linear relationship, suggesting that oligomerization is necessary for unwinding activity (data not shown).

The dynamic oligomeric nature of the Rep protein seems to be a common property of SF3 family members. For example, the papillomavirus E1 helicase is a monomer in solution but binds as a dimer to viral origin DNA through interactions with the virally encoded transcription factor E2 (Chen and Stenlund, 1998). Subsequently, additional E1 molecules are added to the E1-DNA complex, forming hexamers on each strand of the consequently "melted" origin. SV40 TAg is also a monomer in solution but readily forms hexamers upon binding of ATP. However, preformed hexamers are not competent in SV40 origin DNA binding or helicase activity, suggesting that a functional complex must first be assembled from a pool of monomeric subunits (Dean et al., 1992). Analogous to SV40 TAg, ATP binding to the nonstructural protein NS1 of the related parvovirus MVM (minute virus of mice) has also been implicated in promoting NS1 oligomerization necessary for its interaction with the viral p38 promoter (Christensen et al., 1995). Consistent with this observation is the enhancement of Rep78 oligomerization in the presence of both ATP and the nonhydrolyzable analog ATP$\gamma$S. Further, these studies also imply that ATP binding, and not its hydrolysis, is responsible for promoting self-association (Smith et al., 1997).

### Rep40 Has an Arginine Finger

A characteristic of AAA$^+$ proteins, pertinent to their oligomerization, is the presence of a so-called arginine finger. Within the context of the hexamer, the arginine penetrates the active site of a neighboring subunit and thus plays a critical role in cooperative ATP hydrolysis. Analogous to AAA$^+$ proteins, Arg444 in our Rep40 hexamer model, located at the N terminus of $\beta$5, is in an excellent position to make hydrogen bonds with the $\gamma$-phosphate of an ATP modeled in the adjacent subunit (Figure 4C). To test its role in catalysis, we mutated Arg444 to alanine in the context of Rep68$\Delta$490. Figure 4D shows that the mutant is deficient in helicase activity. Of note is the fact that truncation of the Rep68 protein at residue G490 does not result in an alteration of activity when compared to full-length wild-type Rep68. A similar loss of ATPase activity is observed when the arginine finger is mutated in AAA$^+$ proteins, such as RuvB and HslU (Song et al., 2000; Putnam et al., 2001). Together, this is the first indication of the existence of an arginine finger in SF3 helicases, which was not discerned in the original classification of SF3 helicases (Gorbalenya and Koonin, 1993) or in subsequent studies. The presence of a functional arginine finger directly implicates the requirement for oligomerization in order to create a competent catalytic center.

### Role of Motif B′ in Rep40

In addition to motifs A, B, and C, SF3 helicases are characterized by motif B′, an $\sim$14 amino acid region containing a highly conserved glycine at its center and a positively charged residue at each end (Koonin, 1993). In our structure, motif B′ spans the region from the C-terminal end of helix $\alpha$9 to $\beta$ hairpin 1 (Figure 1C). The importance of motif B′ is demonstrated by mutations that diminish Rep68 helicase and ATPase activities. For instance, alanine substitution at Lys391 at one end of the motif completely abolishes helicase and ATPase activities, while alanine substitution at Lys404 at the opposite end eliminates helicase activity but not all of the ATPase activity (Walker et al., 1997). Importantly, in our Rep40 hexamer model, Lys391 is in proximity to the phosphates of the modeled ATP molecule, consistent with a role in catalysis or ATP binding (Figure 4C). Lys404 (on $\beta$ hairpin 1), on the other hand, is distal to the ATP binding site and may be involved in DNA binding.

### Rep40-DNA Interaction

Structures of hexameric ring helicases reveal loops facing into the center of the pore whose residues are involved in single-stranded DNA binding (Singleton et al., 2000; Sawaya et al., 1999; Niedenzu et al., 2001). The Rep40 hexamer model shows the presence of similar loops protruding into the central pore that may play an analogous role (Figure 4A). Of particular interest is the loop in $\beta$ hairpin 1. Two lysine residues, K404 and K406, have the potential to interact with ssDNA passing through the central channel. Moreover, Lys404 is conserved throughout the SF3 family (Figure 1C). Mutation of Lys404 to alanine in Rep68 results in loss of helicase activity but not the ability to hydrolyze ATP (Walker et al., 1997). This phenotype is strikingly similar to that of T7 DNA helicase, in which mutants of residues within the putative DNA binding loops abrogate DNA unwinding but retain some ATPase activity. Furthermore, the T7 mutants were later shown to indeed be deficient in binding single-stranded DNA (Sawaya et al., 1999). By

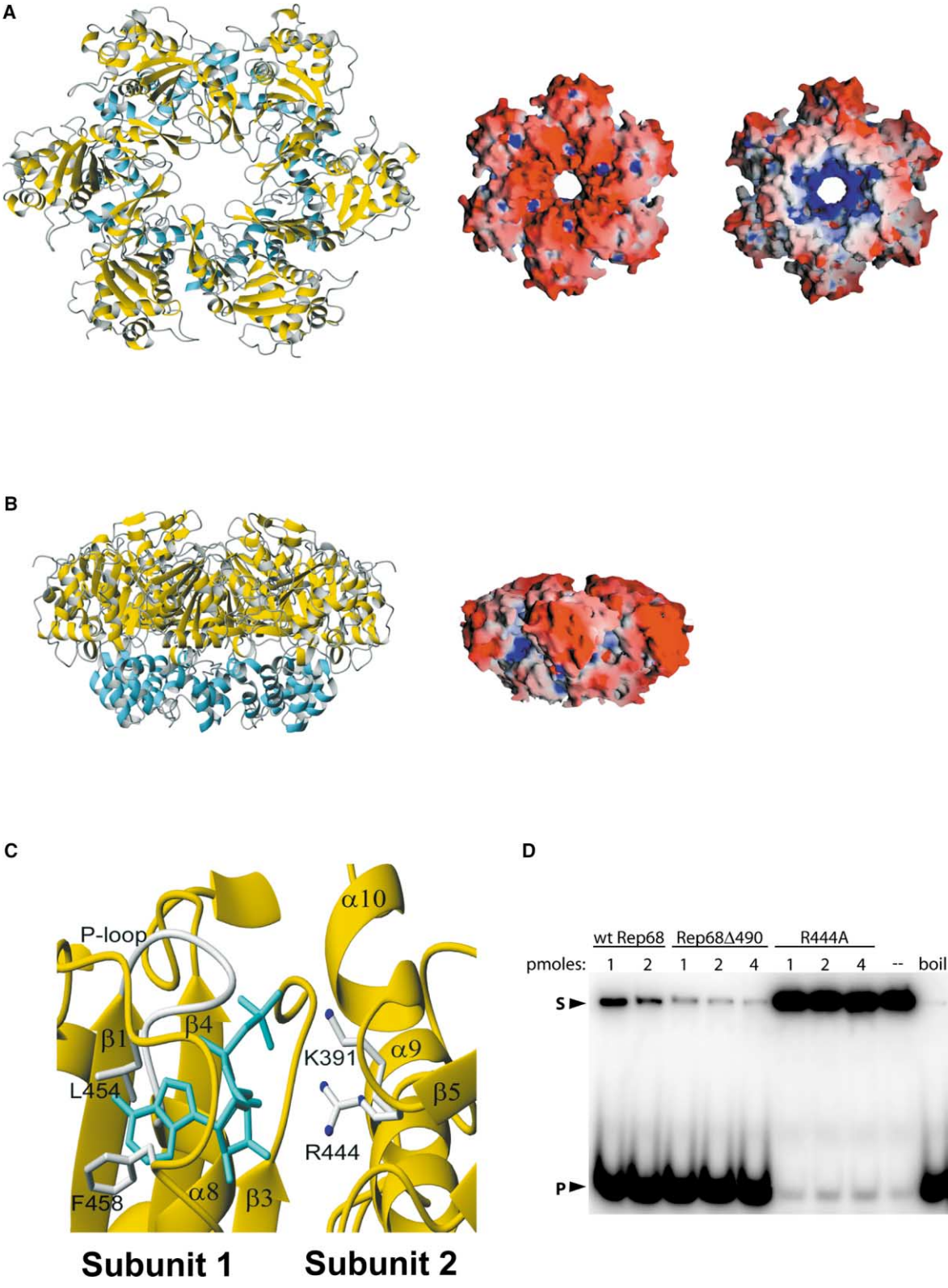


Figure 4. Hexamer Model

The Rep40 hexamer model was generated from a superposition with HslU (PDB ID code 1G3I).

(A) Left: ribbon representation of the hexamer model; N-terminal domains are colored blue and the vAAA$^{+}$ domains are colored gold. The view is from the "top," or the C-terminal end. Electrostatic surface potentials of top (center) and "bottom" (right) views. The positively charged regions are colored blue and negatively charged regions are colored red. A positively charged region at the center of the ring indicates potential DNA interaction regions and is the location of β hairpin 1. The C-terminal face is mostly electronegative.

analogy, the Rep proteins may translocate DNA using the same general mechanism as other hexameric ring helicases. The recent structure of the minichromosome maintenance (MCM) complex from *M. thermoautotrophicum* shows that the use of β hairpin "fingers" containing arginine/lysine residues for binding DNA may be a general feature of some hexameric helicases (Fletcher et al., 2003). Inspection of the electrostatic surface potential shows an electropositive groove at the interface between the N-terminal and vAAA$^+$ domains in Rep40. The groove varies in width due to the intrinsic mobility of the N-terminal domain, but is large enough to accommodate single-stranded DNA (Laskowski, 1995). In particular, several conserved and semiconserved basic residues emanate from helix α4, allowing for a possible interaction with DNA. Examination of the potential role of these residues in DNA binding is underway.

## Conclusions

Several viral proteins involved in replication initiation, such as AAV Rep and SV40 TAg, have been proposed to be distant relatives of the AAA$^+$ superfamily (Neuwald et al., 1999). We show here that Rep40 contains a novel, unimodular AAA$^+$ domain (vAAA$^+$) that defines for the first time the general architecture of SF3 DNA helicases. A common denominator of AAA$^+$ proteins is their ability to utilize ATP hydrolysis to remodel the quaternary structure of macromolecular complexes, including unfolding of both proteins and DNA molecules (Patel and Latterich, 1998; Ogura and Wilkinson, 2001; Lupas and Martin, 2002). Rep's structural relationship to AAA$^+$ proteins provides us with a conceptual framework to try to understand the role of AAV Rep proteins during the viral lifecycle. The fact that all Rep proteins possess the central vAAA$^+$ domain suggests that the additional Rep domains act as "adaptor" regions specific to particular Rep functions. In the context of the larger Rep proteins that contain the amino-terminal origin interaction domain (Rep68/Rep78), the vAAA$^+$ domain is used as a motor to melt the double-stranded origin. During DNA synthesis, Rep can act as a processive helicase. The smaller Rep proteins, when associated with capsids late in infection, function as molecular pumps to introduce the single-stranded viral genome into preformed capsids (Dubielzig et al., 1999; King et al., 2001). In this context, a potential site for Rep-capsid interaction could reside in the second β hairpin loop, whose hydrophobic character and position on "top" of the hexamer ring is similar to the hairpin used by RuvB for interaction with RuvA (Han et al., 2001; Yamada et al., 2002). Perhaps interaction with the capsid itself induces Rep40/52 oligomerization, which in turn activates ATPase activity. Further, the structural similarity between Rep and RuvB brings to light the possibility of a branch migration event during the establishment of AAV site-specific integration.

The use of the "AAA$^+$ module" as a molecular DNA melting motor is conserved from bacteria to eukaryotes (Giraldo, 2003; Davey et al., 2002). In *E. coli*, for instance, the initiator protein DnaA contains an AAA$^+$ module, and binds to DnaA boxes present in the origin of replication, oriC. Upon DNA binding, DnaA oligomerizes, wrapping the DNA around the protein core. In contrast to DnaA, however, which needs to recruit the replicative hexameric helicase DnaB, the vAAA$^+$ module in Rep proteins appears capable of not only providing the necessary ATPase activity for melting the viral origin, but also of functioning as a replicative helicase. Many AAA$^+$ proteins such as RuvB exhibit a deficient helicase activity that can only unwind very short DNA fragments (Tsaneva et al., 1993). The novel Rep vAAA$^+$ module has somehow optimized the AAA$^+$ scaffold to function as a standalone processive helicase.

Since the submission of this manuscript, the structure of the SV40 TAg helicase fragment has been reported, crystallized as a hexamer in the absence of ATP (Li et al., 2003). The SV40 structure has many of the features anticipated here, including a vAAA$^+$ domain preceded by an N-terminal helical subdomain. The N-terminal helical subdomain is, however, larger and comprises a novel Zn binding module. The corresponding vAAA$^+$ domain is also larger by the inclusion of a "helical bulge" that packs closely against one side of the α/β core. Several helices contributing to this bulge stem from the C terminus, whereas the equivalent segment in AAV Rep40 lacks secondary structure and instead wraps around the vAAA$^+$ domain. However, the most striking difference occurs in the orientation of the N-terminal domain, which is swung inward by ~23° compared to Rep40. Consequently, the "root" of the mushroom-shaped SV40 hexamer is narrower (~25 Å) than in Rep40 (~35 Å) and defines a channel continuous with that circumscribed by the vAAA$^+$ domain. The additional intersubunit contacts contributed by the N-terminal helical domain may explain why, in contrast to Rep40, the SV40 TAg crystallized as a hexamer. Interestingly, the conformational variability of the N-terminal domain observed between the three Rep40 molecules in the asymmetric unit here suggests that the Rep40 N-terminal domain may twist in a similar fashion upon DNA binding and/or NTP binding, and perhaps stabilize the hexameric state. Overall, it is tempting to think that the mechanism of DNA unwinding by SF3 helicases, coupled to ATP hydrolysis, is linked to the conformational sensitivity of the N-terminal domain.

In conclusion, we hypothesize that small DNA viruses such as HPV, SV40, and AAV have adapted the AAA$^+$ protein architecture in their nonstructural proteins to act as molecular motors involved in their wide range of activities from DNA replication, transcription, packaging and—in the case of AAV—site-specific integration.

(B) Hexamer model rotated 90° from (A) and its corresponding surface electrostatic representation.
(C) Interface between adjacent monomers in the modeled hexamer. The Rep40 model is shown in gold with the arginine finger residue (R444) and motif B′ residue K391 represented as all-bonds at the interface. A modeled AMP-PNP molecule is shown in cyan.
(D) The helicase activities of the Rep68 protein and mutants were performed as described in Experimental Procedures. Rep68Δ490 and a Rep68Δ490 carrying an R444A mutation were tested for helicase activity. Numbers represent picomoles of protein; S and P are substrate and product, respectively; and boil indicates heat-denatured substrate.

## Experimental Procedures

### Expression, Purification, and Crystallization

The sequence encoding AAV2 Rep40 protein was generated via PCR and subcloned into pET-15b vector (Novagen). His-tagged protein was expressed in *E. coli* BL21(DE3) Gold cells and purified over an $Ni^{+2}$-agarose column. After removal of the histidine tail by digestion with thrombin, the protein was further purified by MonoQ (HR10/10) anion exchange column followed by size exclusion chromatography on a Superdex-200 column. Selenomethionine (SeMet)-substituted Rep40 was expressed in *E. coli* B834 methionine auxotrophic strain in M9 minimal medium supplemented with all amino acids except methionine. SeMet-labeled Rep40 was purified in the same manner as native protein. A final concentration of 32 mg/ml of protein was used to grow crystals by the hanging drop vapor diffusion method. Crystals appeared overnight from a solution containing 0.2 M sodium acetate, 10% PEG 8K, and 0.1 mM Tris (pH 8.5) and grew to an average size of 0.6 × 0.07 × 0.04 mm over several days. Conditions were subsequently refined to give native crystals that diffracted to 2.4 Å. Crystals belong to space group $P6_5$ with cell dimensions of a = b = 126.74 Å, c = 98.24 Å; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$), with three molecules per asymmetric unit. For data collection, crystals were cryoprotected with a solution containing 30% glycerol and 15% PEG 8K.

### Helicase Assays

The M13-based partially duplex substrate used in these assays was generated by first annealing M13(-20) sequencing primer (5′-GTAAAACGACGGCCAGT-3′) to M13mp18 DNA, followed by klenow-mediated extension of the 3′ end in the presence of dGTP and α-[32P]-dCTP. Unincorporated radiolabel was removed by passage over a Sephadex G-50 column. The resulting duplex region of the substrate is 20 bp. Helicase assays were performed as previously described (Yoon-Robarts and Linden, 2003), with some modifications. Briefly, Rep protein was incubated in the presence of 20 fmol radiolabeled DNA substrate, 20 mM Tris-Cl (pH 7.5), 20 mM $MgCl_2$, 8 mM DTT, 1 mM ATP, 5 μg/ml BSA, and 4% sucrose in a total volume of 15 μl. Reactions were incubated for 45 min at 37°C and terminated by the addition of 4 μl loading buffer (10 mM Tris-Cl [pH 7.5], 1 mM EDTA, 0.5% SDS, 0.1% each bromophenol blue and xylene cyanol, and 20% glycerol). Products of the helicase reaction were resolved on 8% native polyacrylamide gels in 1× Tris-borate-EDTA (pH 8.3) buffer. Following electrophoresis, gels were fixed in 10% trichloroacetic acid, dried, and visualized by phosphorimager (Molecular Dynamics).

### Data Collection, Phasing, and Refinement

Native data were collected at the National Synchrotron Light Source (beamline X6a). MAD data were measured from a single frozen crystal at the Argonne National Laboratory Advanced Photon Source (APS) on beamline 14BM. The data were measured at three wavelengths, corresponding to the edge and peak of the selenium K edge absorption profile plus an additional remote point (Table 1). The data were processed with the HKL2000 package (Otwinowski and Minor 1997). The positions of 11 of the expected 18 total Se sites were located using SOLVE (Terwilliger and Berendzen, 1999) and were refined and phased using CNS (Brunger et al., 1998). The initial phases (3.0 Å) were improved by density modification (Cowtan, 1994) and extended to 2.4 Å. The resulting electron density map was of excellent quality and was used to build the initial model in O (Jones et al., 1991). After several iterative rounds of refinement with CNS and manual rebuilding with the program O, the model converged to an R factor of 24% ($R_{free}$ of 28.3%) and contained all residues for molecules A and B, residues 225–288, 296–324, and 442–490 for molecule C, and 480 water molecules. The model has good stereochemistry, with no residues in the disallowed regions of a Ramachandran plot.

### Sequence Alignments and Modeling

Initially, a Complete Database search for closely related family members was performed in RPS-BLAST using the Rep40 amino acid sequence. The sequences generated in BLAST were used to perform multiple family sequence alignments using CLUSTAL-X (Thompson et al., 1997) with a BLOSUM matrix for proteins. Sequences were then manually aligned based on the Rep40 structure. Electrostatic surface representations were calculated using GRASP (Nicholls et al., 1991). Models for the Rep40 hexamer were generated from a least square superposition with HslU (Protein Data Bank ID code 1G3I) using the residues from Walker A, Walker B, and sensor 1 motifs.

## Acknowledgments

We thank the staff at beamlines 14BM (APS) and X6A (NSLS) for facilitating X-ray data collection. A.K.A. is supported by NIH grant R01 AI41706, and R.M.L. is supported by NIH grant R01 GM62234.

Received: May 15, 2003
Revised: June 26, 2003
Accepted: July 3, 2003
Published online: July 10, 2003

## References

Bochtler, M., Hartmann, C., Song, H.K., Bourenkov, G.P., Bartunik, H.D., and Huber, R. (2000). The structures of HslU and ATP-dependent protease HslU-HslV. Nature *403*, 800–805.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nigles, M., Punnu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. *D54*, 905–921.

Caruthers, J.M., and McKay, D.B. (2002). Helicase structure and mechanism. Curr. Opin. Struct. Biol. *12*, 123–133.

Chen, G., and Stenlund, A. (1998). Characterization of the DNA-binding domain of the bovine papillomavirus replication initiator E1. J. Virol. *72*, 2567–2576.

Christensen, J., Cotmore, S.F., and Tattersall, P. (1995). Minute virus of mice transcriptional activator protein NS1 binds directly to the transactivation region of the viral P38 promoter in a strictly ATP-dependent manner. J. Virol. *69*, 5422–5430.

Cowtan, K. (1994). DM: an automated procedure for phase improvement by density modification. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography *31*, 34–38.

Davey, M.J., Jeruzalmi, D., Kuriyan, J., and O'Donnell, M. (2002). Motor and switches: $AAA^+$ machines within the replisome. Nat. Rev. Mol. Cell Biol. *3*, 1–10.

Davis, M.D., Wonderling, R.S., Walker, S.L., and Owens, R.A. (1999). Analysis of the effects of charge cluster mutations in adeno-associated virus Rep68 protein in vitro. J. Virol. *73*, 2084–2093.

Dean, F.B., Borowiec, J.A., Eki, T., and Hurwitz, J. (1992). The simian virus 40 T antigen double hexamer assembles around the DNA at the replication origin. J. Biol. Chem. *267*, 14129–14137.

Dubielzig, R., King, J.A., Weger, S., Kern, A., and Kleinschmidt, J.A. (1999). Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. J. Virol. *73*, 8989–8998.

Dutheil, N., Shi, F., Dupressoir, T., and Linden, R.M. (2000). Adeno-associated virus site-specifically integrates into a muscle-specific DNA region. Proc. Natl. Acad. Sci. USA *97*, 4862–4866.

Erzberger, J.P., Pirruccello, M.M., and Berger, J.M. (2002). The structure of bacterial DnaA: implications for general mechanisms underlying DNA replication initiation. EMBO J. *21*, 4763–4773.

Fletcher, R.J., Bishop, B.E., Leon, R.P., Sclafani, R.A., Ogata, C.M., and Chen, X.S. (2003). The structure and function of MCM from archael *M. thermoautotrophicum*. Nat. Struct. Biol. *10*, 160–167.

Giraldo, R. (2003). Common domains in the initiators of DNA replication in *Bacteria*, *Archaea* and *Eukarya*: combined structural, functional and phylogenetic perspectives. FEMS Microbiol. Rev. *26*, 533–554.

Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. Curr. Opin. Struct. Biol. *3*, 419–429.

Han, Y.-W., Iwasaki, H., Miyata, T., Mayanagi, K., Yamada, K., Mori-

kawa, K., and Shinagawa, H. (2001). A unique β-hairpin protruding from AAA$^+$ ATPase domain of RuvB motor protein is involved in the interaction with RuvA DNA recognition protein for branch migration of Holliday junctions. J. Biol. Chem. *276*, 35024–35028.

Hartman, J.J., and Vale, R.D. (1999). Microtubule disassembly by ATP-dependent oligomerization of the AAA enzyme katanin. Science *286*, 782–785.

Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchroton radiation. Science *254*, 51–58.

Hickman, A.B., Ronning, D.R., Kotin, R.M., and Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Mol. Cell *10*, 327–337.

Im, D.S., and Muzyczka, N. (1990). The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell *61*, 447–457.

Jones, T., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. *A49*, 1110–1119.

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.

Koonin, E.V. (1993). A common set of conserved motifs in a vast variety of putative nucleic acid-dependent ATPases including MCM proteins involved in the initiation of eukaryotic DNA replication. Nucleic Acids Res. *21*, 2541–2547.

Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A., McLaughlin, S., Muzyczka, N., Rocchi, M., and Berns, K.I. (1990). Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA *87*, 2211–2215.

Laskowski, R.A. (1995). SURFNET: a program for visualization molecular surfaces, cavities and intermolecular interactions. J. Mol. Graph. *13*, 323–330.

Lenzen, C.U., Steinmann, D., Whiteheart, S.W., and Weis, W.I. (1998). Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. Cell *94*, 525–536.

Linden, R.M., and Berns, K.I. (2000). Molecular biology of adeno-associated viruses. Contrib. Microbiol. *4*, 68–84.

Li, D., Zhao, R., Lilyestrom, W., Gai, D., Zhang, R., DeCaprio, J.A., Fanning, E., Jochimiak, A., Szakonyi, G., and Chen, X.S. (2003). Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen. Nature *423*, 512–518.

Lohman, T.M., and Bjornson, K.P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. *65*, 169–214.

Lupas, A.N., and Martin, J. (2002). AAA proteins. Curr. Opin. Struct. Biol. *12*, 746–753.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E. (1999). AAA$^+$: a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res. *9*, 27–43.

Nicholls, A., Sharp, K.A., and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins *11*, 281–296.

Niedenzu, T., Roleke, D., Bains, G., Scherzinger, E., and Saenger, W. (2001). Crystal structure of the hexameric replicative helicase RepA of plasmid RSF1010. J. Mol. Biol. *306*, 479–487.

Ogura, T., and Wilkinson, A.J. (2001). AAA$^+$ superfamily ATPases: common structure—diverse function. Genes Cells *6*, 575–597.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Patel, S., and Latterich, M. (1998). The AAA team: related ATPases with diverse functions. Trends Cell Biol. *8*, 65–71.

Putnam, C.D., Clancy, S.B., Tsuruta, H., Gonzalez, S., Wetmur, J.G., and Tainer, J.A. (2001). Structure and mechanism of RuvB Holliday junction branch migration motor. J. Mol. Biol. *311*, 297–310.

Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Houssman, D.E., Epstein, N., and Hunter, L.A. (1991). Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J. *10*, 3941–3950.

Sawaya, M.R., Guo, S., Tabor, S., Richardson, C.C., and Ellenberger, T. (1999). Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. Cell *99*, 167–177.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., and Wigley, D.B. (2000). Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides. Cell *101*, 589–600.

Smith, R.H., and Kotin, R.M. (1998). The Rep52 product of adeno-associated virus is a DNA helicase with 3′-to-5′ polarity. J. Virol. *72*, 4874–4881.

Smith, R.H., Spano, A.J., and Kotin, R.M. (1997). The Rep78 gene product of adeno-associated virus self-associates to form a hexameric complex in the presence of AAV ori sequences. J. Virol. *71*, 4461–4471.

Song, H.K., Hartmann, C., Ramachandran, R., Bochtler, M., Behrendt, R., Moroder, L., and Huber, R. (2000). Mutational studies of HslU and its docking mode with HslV. Proc. Natl. Acad. Sci. USA *97*, 14103–14108.

Sousa, M.C., Trame, C.B., Tsuruta, S., Wilbanks, S.M., Reddy, R.S., and McKay, D. (2000). Crystal and solution structures of an HslVU protease-chaperone complex. Cell *103*, 633–643.

Terwilliger, T.C., and Berendzen, J. (1999) Automated MAD and MIR structure solution. Acta Crystallogr. D Biol. Crystallogr. *55*, 849–861.

Thompson, J.D., Gibson, T.J., Plewinak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL-X interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876–4882.

Tsaneva, I.R., Muller, B., and West, S.C. (1993). RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity in vitro. Proc. Natl. Acad. Sci. USA *90*, 1315–1319.

Vale, R.D. (2000). AAA proteins: lords of the ring. J. Cell Biol. *150*, F13–F19.

Valle, M., Gruss, C., Halmer, L., Carazo, J.M., and Donate, L.E. (2000). Large T-antigen double hexamers imaged at the simian virus 40 origin of replication. Mol. Cell. Biol. *20*, 34–41.

VanLoock, M.S., Alexandrov, A., Yu, X., Cozzarelli, N.R., and Egelman, E.H. (2002). SV40 large T antigen hexamer structure: domain organization and DNA-induced conformational changes. Curr. Biol. *12*, 472–476.

Walker, S.L., Wonderling, R.S., and Owens, R.A. (1997). Mutational analysis of the adeno-associated virus type 2 Rep68 protein helicase motifs. J. Virol. *71*, 6996–7004.

Wang, J., Song, J.J., Franklin, M.C., Kamtekar, S., Im, Y.J., Rho, S.H., Seong, I.S., Lee, C.S., Chung, C.H., and Eom, S.H. (2001a). Crystal structures of the HslVU peptidase-ATPase complex reveal an ATP-dependent proteolysis mechanism. Structure *9*, 177–184.

Wang, J., Song, J.J., Seong, I.S., Franklin, M.C., Kamtekar, S., Eom, S.H., and Chung, C.H. (2001b). Nucleotide-dependent conformational changes in a protease-associated ATPase HslU. Structure *9*, 1107–1116.

Yamada, K., Kunishima, N., Mayanagi, K., Ohnishi, T., Nishino, T., Iwasaki, H., Shinagawa, H., and Morikawa, K. (2001). Crystal structure of the Holliday junction migration motor protein RuvB from *Thermus thermophilus* HB8. Proc. Natl. Acad. Sci. USA *98*, 1442–1447.

Yamada, K., Miyata, T., Tsuchiya, D., Oyama, T., Fujiwara, Y., Ohnishi, T., Iwasaki, H., Shinagawa, H., Ariyoshi, M., Mayanagi, K., et al. (2002). Crystal structure of the RuvA-RuvB complex: a structural basis for the Holliday junction migrating motor machinery. Mol. Cell *10*, 671–681.

Yoon-Robarts, M., and Linden, R.M. (2003). Identification of active site residues of the adeno-associated virus type 2 Rep endonuclease. J. Biol. Chem. *278*, 4912–4918.

Yu, R.C., Hanson, P.I., Jahn, R., and Brunger, A.T. (1998). Structure of the ATP-dependent oligomerization domain of N-ethylmaleimide sensitive factor complexed with ATP. Nat. Struct. Biol. *5*, 803–811.