Molecular Cell, Vol. 10, 327–337, August, 2002, Copyright ©2002 by Cell Press

# Structural Unity among Viral Origin Binding Proteins: Crystal Structure of the Nuclease Domain of Adeno-Associated Virus Rep

**Alison Burgess Hickman,[1,3] Donald R. Ronning,[1] Robert M. Kotin,[2] and Fred Dyda[1]**
[1]Laboratory of Molecular Biology
National Institute of Diabetes
and Digestive and Kidney Diseases
[2]Laboratory of Biochemical Genetics
National Heart, Lung, and Blood Institute
National Institutes of Health
Bethesda, Maryland 20892

## Summary

**Adeno-associated virus (AAV), unique among animal viruses in its ability to integrate into a specific chromosomal location, is a promising vector for human gene therapy. AAV Replication (Rep) protein is essential for viral replication and integration, and its amino terminal domain possesses site-specific DNA binding and endonuclease activities required for replication initiation and integration. This domain displays a novel endonuclease fold and demonstrates an unexpected structural relationship to other viral origin binding proteins such as the papillomavirus E1 protein and the SV40 T antigen. The active site, located at the bottom of a positively charged cleft, is formed by the spatial convergence of a divalent metal ion and two conserved sequence motifs that define the rolling circle replication superfamily.**

## Introduction

One of the most promising viral vectors for human gene therapy is adeno-associated virus, AAV (Kotin, 1994). Advantages of this nonpathogenic human parvovirus include an apparent lack of toxicity, low immunogenicity, and the ability to infect many dividing and nondividing cell types. AAV is a dependovirus, and its replication requires a simultaneous infection with a helper virus such as adenovirus or herpesvirus. In the absence of helper virus infection, AAV-2—the best characterized dependovirus serotype—integrates its DNA into a well-defined location on human chromosome 19 for later rescue upon helper virus superinfection (Kotin et al., 1990, 1992; Samulski et al., 1991). Disruption of the integration locus, denoted *AAVS1* (Kotin et al., 1991), does not appear to have adverse effects on cell growth, offering the alluring possibility of a safe haven for the insertion of exogenous genes. In addition, the ability to integrate predictably into a defined genetic environment is advantageous since the genomic environment is likely to influence both level and duration of expression.

AAV is a small, icosahedral particle ~22 nm in diameter. Its ~4.7 kb genome is linear single-stranded (ss) DNA that contains only two open reading frames (ORFs). The *cap* ORF encodes for three structural proteins, VP-1, VP-2, and VP-3, produced by alternate splice utilization and an atypical translational initiation codon. The four nonstructural, regulatory proteins encoded by the *rep* ORF are produced by the use of two promoters and a common splice site. Named on the basis of apparent molecular weight, one of the two larger Rep proteins, Rep78 or Rep68, is required for AAV DNA replication (Chejanovsky and Carter, 1990), while the smaller proteins, Rep50 and Rep42, are believed to participate in viral packaging (Chejanovsky and Carter, 1989; King et al., 2001).

The economy exhibited by small viruses such as AAV in exploiting single genes to perform multiple functions is emphasized by the plethora of biochemical properties exhibited by the Rep proteins (reviewed in Smith and Kotin, 2002). For example, Rep78 and Rep68 are site-specific DNA binding proteins (Owens et al., 1991; Snyder et al., 1993; Chiorini et al., 1994a), as well as site-specific endonucleases (Im and Muzyczka, 1990). In addition, they have helicase and ATPase activities (Im and Muzyczka, 1990), properties shared by Rep52 (Smith and Kotin, 1998). A ligase activity has also been demonstrated (Smith and Kotin, 2000). Since expression of Rep78 and Rep68 is necessary for AAV integration, it has been inferred that the DNA binding and endonuclease activities of Rep are required for site-specific chromosomal integration.

The two ends of the AAV genome are characterized by ~145 nucleotide sequences that contain inverted repeats and are capable of folding into hairpin structures (Figure 1A). These inverted terminal repeats (ITRs) serve as the origins for DNA replication. Within each ITR are two sequences required for replication: a Rep binding sequence (RBS) consisting of tetranucleotide direct repeats, and a terminal resolution site (trs) where Rep cleaves the viral DNA. Rep catalyzes DNA cleavage using a tyrosine residue that becomes covalently attached to the 5′ phosphate of the cleaved strand (Figure 1B). This generates a free 3′ end from which the DNA strand can be extended by template-directed synthesis to replicate the terminal palindrome. These steps are a variant of those that occur during rolling circle replication (RCR), an ancient and basic mechanism by which many prokaryotic and eukaryotic replicons replicate (Khan, 2000). A common feature of systems that use RCR is the cleavage of one strand of DNA using a Tyr residue as a nucleophile, resulting in a 5′-phosphotyrosyl linkage and a new 3′ end that serves as an origin for the next round of circle replication.

Some years ago, Koonin and coworkers noted the conservation of primary sequence motifs among bacteriophage proteins catalyzing RCR (such as the gene A protein from $\phi$X174) and a surprisingly large number of proteins previously unsuspected to be related. These included—among others—the endonuclease domains of parvoviral NS1 proteins (such as AAV Rep) and plant geminiviruses (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993). One of the conserved motifs consists of two His residues separated by a large hydrophobic residue (U) and followed by three hydrophobic residues, $HUHU_3$, called here the HUH motif (also known as motif 2). An-

[3]Correspondence: ahickman@helix.nih.gov

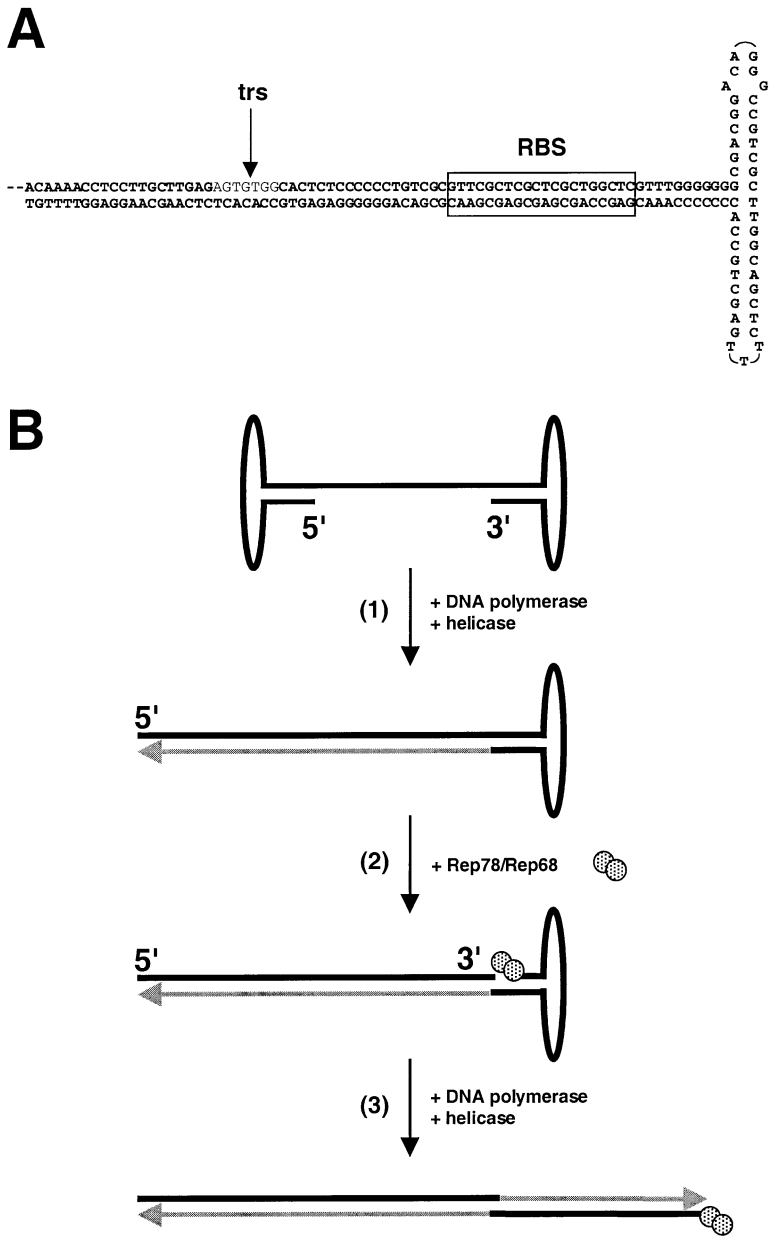


Figure 1. The Inverted Terminal Repeat of AAV-5 and Model for Replication

(A) The predicted T-shaped structure of the inverted terminal repeat (ITR) is the result of optimized base pairing of palindromic sequences. The stem contains the Rep binding site (RBS) and the terminal resolution site (trs) (Chiorini et al., 1999b).

(B) Schematic of proposed model for the initial steps of AAV replication (adapted from Kotin, 1994). In Step 1, template-directed synthesis proceeds from the free 3′ hydroxyl on the bottom strand to copy one end of the AAV genome. In Step 2, Rep cleaves the top strand at the trs, generating a new 3′ hydroxyl group and a phosphotyrosine bond between Rep and the 5′ phosphate at the cleavage site. In Step 3, the other end of the AAV genome is duplicated. Subsequent separation of the two DNA strands requires a helicase activity.

other motif, (Y)UxxYx$_{2-3}$K (motif 3), contains either one or two conserved Tyr residues that have been shown in many systems to be catalytically essential (e.g., van Mansfeld et al., 1986; Odegrip and Haggård-Ljungquist, 2001). The role of the HUH motif is less clear than that of the active site tyrosines, although it has been suggested that the two histidines could coordinate a metal ion (Ilyina and Koonin, 1992). Another conserved RCR sequence motif that often appears N-terminal to motifs 2 and 3 is not present in AAV Rep. It is believed that all RCR domains evolved from a common ancestral domain (Koonin and Ilyina, 1993), yet the RCR domain does not appear to be related to any proteins involved in host cell replication, suggesting very early branching in pathways that serve to initiate DNA replication. RCR domains are restricted to systems that follow a parasitic lifestyle.

In the presence of ATP and a divalent metal ion, Rep78 and Rep68 bind to and nick linear double-stranded (ds) DNA substrates that contain an RBS and a trs, indicating that the entire ITR is not required for these Rep activities (Chiorini et al., 1994a, 1994b). Rep68 can also cleave ssDNA containing the trs sequence in an ATP-independent reaction, suggesting that the in vivo substrate for Rep is a single-stranded region generated by ATP-dependent DNA unwinding, possibly in the form of a stem-loop structure (Smith and Kotin, 2000; Brister and Muzyczka, 2000). Studies using C-terminally truncated versions of AAV-2 Rep demonstrated that the N-terminal domain, consisting of the first ~200 residues, can bind and specifically cleave DNA that contains RBS and single-stranded trs sequences (Davis et al., 2000; Yoon et al., 2001). This activity is metal dependent and requires Tyr-156, the active site nucleophile (Davis et al., 2000; Smith and Kotin, 2000).

Table 1. Data Collection and Processing Statistics

| Data Set | Zn K Edge | Zn K White Line | Zn Remote | CuKα |
|---|---|---|---|---|
| Energy (keV) | 9.664 | 9.669 | 12.284 | 8.04 |
| Wavelength (Å) | 1.2829 | 1.2823 | 1.0093 | 1.5418 |
| Resolution (Å) | 1.7 | 1.7 | 1.4 | 2.0 |
| Total reflections (N) | 204060 | 204157 | 373647 | 91775 |
| Unique reflection (N) | 43833 | 43822 | 79390 | 26720 |
| Completeness (%) (for $I/\sigma I > -3.0$) | 98.0 (80.9) | 98.0 (81.2) | 99.6 (95.9) | 97.3 (94.1) |
| $I/\sigma I$ | 12.6 (5.46) | 11.5 (5.58) | 10.0 (2.68) | 24.5 (15.8) |
| $R_{sym}$ | 0.058 (0.187) | 0.062 (0.178) | 0.062 (0.404) | 0.033 (0.073) |
| Experimental phase determination at 2.0 Å resolution | | | | |
| $R_{Cullis}$ | | | 0.356 | |
| $R_{Kraut}$ | | 0.06 | 0.020 | |
| Phasing power | | | | |
| Dispersive | | | 4.41 | |
| Anomalous | | 5.01 | | |
| Combined figure of merit before density modification | | | | 0.80 |

| Refinement (against the data set collected at 12.284 keV) | |
|---|---|
| Resolution (Å) | 30–1.4 |
| Atoms (N) | 3848 |
| Reflections $F > 0\sigma(F)$ | 77130 |
| R factor (%) | 17.9 |
| $R_{free}$ (%) | 19.6 |
| Average B factor (Å$^2$) | 14.73 |
| Rms ΔB bonded atoms (Å$^2$) | 2.66 |
| Rms bond lengths (Å) | 0.006 |
| Rms bond angles (°) | 1.36 |

The numbers in parentheses refer to the highest resolution shell: 1.74–1.70 Å for the Zn K edge and white line; 1.43–1.40 Å for the Zn high-energy remote; and 2.05–2.00Å for the Cu Kα data set. $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$. $R_{Cullis} = \Sigma|||FPH_o| \pm |FP_o|| - |FH_C||/\Sigma||FPH_o| \pm |FP_o||$ for centric reflections. $R_{Kraut} = \Sigma||FPH_o| - |FPH_C||/|FPH_o|$ for acentric reflections, isomorphous case; $\Sigma||FPH_o^+| - |FPH_C^+|| + ||FPH_o^-| - |FPH_C^-||/\Sigma(|FPH_o^+| + |FPH_o^-|)$ for acentric reflections, anomalous case. FP is the protein, FPH is the dispersive derivative, FH is the dispersive Zn structure factor. $FPH^+$ and $FPH^-$ denote the Bijvoet mates in the Zn white line (9.669 keV) data set. The phasing power is defined as $FH_c/E$ for the dispersive case and $2FH_C/E$ for the anomalous case, where E is the rms lack of closure. R factor $= \Sigma|FP_o - FP_C|/\Sigma|FP_o|$.

The promise held by AAV as a vector for gene therapy will greatly benefit from a detailed mechanistic understanding of the key steps in site-specific integration, DNA cleavage, and ligation. While AAV integrates predominantly into *AAVS1*, it also integrates with lower frequency into other sites in the genome that are either cryptic recognition sites for Rep or occur independent of Rep. Determining the structural basis of DNA recognition and nicking may allow the rational redesign of Rep to enhance its substrate specificity. Conversely, it may be possible to modify Rep to redirect integration to other sites in the genome. As an approach to understanding the mechanism of DNA cleavage and site-specific integration by AAV Rep, we have determined the structure of the N-terminal endonuclease domain of AAV-5 Rep, consisting of residues 1 to 197, to 1.4 Å resolution.

## Results and Discussion

### Structure Determination

Limited protease digestion of AAV-2 Rep68 with papain and V8 protease yielded several N-terminal fragments which were characterized by N-terminal sequencing and MALDI-TOF mass spectrometry. Crystals were ultimately obtained of the domain consisting of residues 1–197 (Rep1–197) which was monomeric in solution as judged by sedimentation equilibrium (data not shown). Despite intensive efforts to refine the crystallization conditions, these crystals remained poorly diffracting. In contrast, the equivalent fragment from AAV-5 yielded well-diffracting crystals in the presence of $Zn^{2+}$. The amino acid sequences of the AAV-2 and AAV-5 N-terminal domains are 60% identical (Bantel-Schaal et al., 1999; Chiorini et al., 1999a).

The structure of AAV-5 Rep1–197 was solved using Multiwavelength Anomalous Diffraction (MAD) (Hendrickson, 1991) utilizing the anomalous scattering of the bound $Zn^{2+}$ ions at 2.0 Å resolution. The structure has subsequently been refined at 1.4 Å resolution (Table 1). Experimental and refined electron densities are shown in Figure 2A.

### Description of the Structure

The endonuclease domain of AAV-5 Rep consists of a central five-stranded antiparallel β sheet flanked by three α helices on one side (helices αA, αB, and αC) and three helices on the other (αD, E [a $3_{10}$ helix], and αF), forming a relatively flat domain approximately 45 Å × 50 Å × 22 Å (Figure 2B). On one side of the sheet, there is a well-defined cleft whose floor is formed by four strands (β1, β3, β2, and β6), while helix αD and the hairpin turn between strands β2 and β3 forms its walls. This cleft is a part of a large surface on the molecule with positive electrostatic potential. One curious feature is a strongly basic protrusion formed by a 10 residue segment between strands β4 and β5 which curves over the surface (Figure 3A). This decreases the opening to the cleft to ~12 Å, restricting access. The protrusion is well ordered, does not appear to be stabilized by pack-

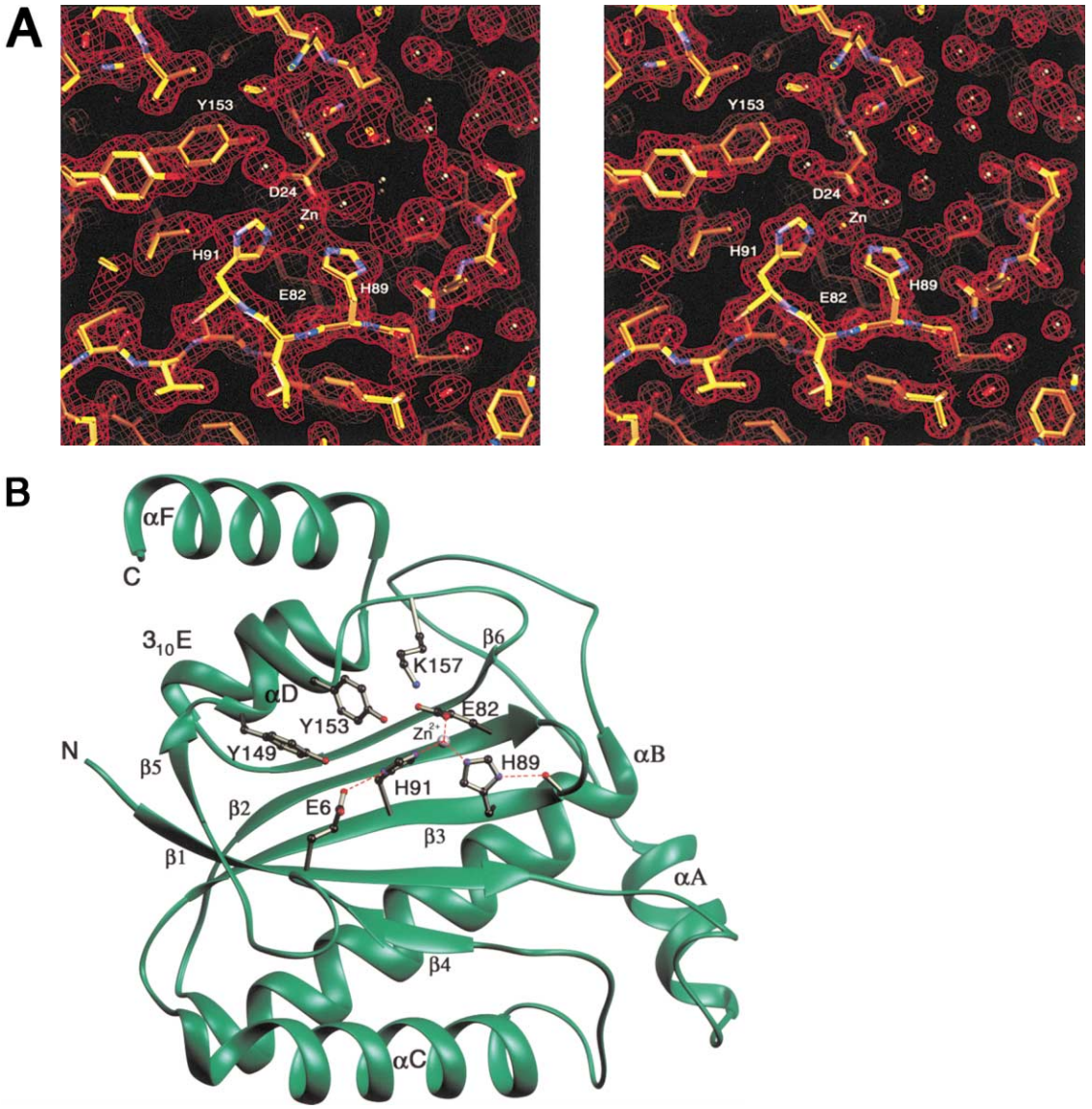


Figure 2. Experimental and Refined Electron Density and the Structure of Rep1–197

(A) Experimental (left) and refined (right) electron density of AAV-5 Rep1–197 in the region of Tyr-153. The electron density is contoured at 1$\sigma$, and the final refined structure is overlayed on both panels. The ligands to the tetrahedrally coordinated $Zn^{2+}$ are Glu-82 (which is behind the plane of the image), His-89, His-91, and Asp-24 which is contributed by an adjacent Rep monomer.

(B) Ribbons (Carson, 1991) representation of AAV-5 Rep1–197. The position of the bound zinc ion is shown in silver.

ing interactions, and is identical in the two crystallographically independent monomers.

On the opposite side of the sheet, helices $\alpha$A, $\alpha$B, and $\alpha$C form a flat, largely hydrophobic surface that is excluded from solvent by an adjacent monomer. Although the crystallographically observed dimer has perfect 2-fold symmetry, the total solvent-excluded surface buried by the dimer is only $\sim$1425 Å$^2$.

## Structurally Related Proteins

The endonuclease domain fold of AAV Rep is unrelated to all other structurally characterized nucleases. Specifically, it bears no topological resemblance to other enzymes that employ nucleophilic active site tyrosines, topoisomerases (e.g., Lima et al., 1994; Berger et al., 1996; Redinbo et al., 1998), and site-specific tyrosine recombinases (Van Duyne, 2001). However, a DALI search (Holm and Sander, 1997) identified 83 protein domains with statistically significant similarity scores that are structurally homologous to the Rep N-terminal domain. According to the SCOP database (Murzin et al., 1995), these proteins share the ferredoxin-like fold, i.e., an $\alpha + \beta$ sandwich with an antiparallel $\beta$ sheet organized $\beta\alpha\beta\beta\alpha\beta$. The structural elements of Rep responsible for the large number of homologs are the first four $\beta$ strands and two intervening helices ($\beta$4, $\beta$1, $\beta$3, $\beta$2, and $\alpha$B and $\alpha$C; Figure 4A), a topology also known as the RNA recognition motif, or RRM (Burd and Dreyfuss, 1994).

Of the many protein homologs of Rep, only three are further distinguished by possessing a homologous fifth

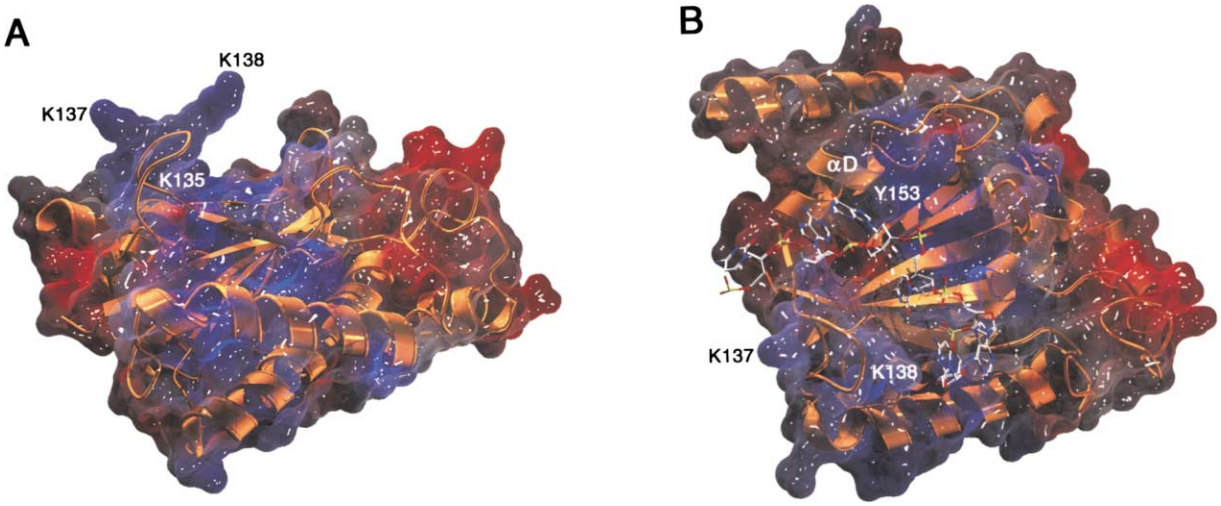


Figure 3. Surface Representations of AAV-5 Rep1–197

(A) Blue and red represent regions of positive and negative surface electrostatic potential, respectively. The protrusion at top left consists of residues 135–144; Lys-135, Lys-137, and Lys-138 are marked.

(B) Rep has been rotated forward ~90° relative to (A), and six nucleotides of ssDNA (adapted from PDB ID 1a9n) modeled into the active site cleft. The model was made by three-dimensional superposition of the spliceosomal U2B″-U2A′ protein-RNA complex (Price et al., 1998) to AAV Rep. The two proteins can be superimposed with an rmsd of 1.8 Å over 44 Cα positions. The RNA 2′-OH groups have been removed. Features of the modeled DNA, shown in stick representation, include binding across the β sheet as seen in RRM complexes (Conte et al., 2000) and curving around helix αD to avoid steric clashes. The position of the active site Tyr-153 is marked. Figures were generated with SPOCK (Christopher, 1998) and MOLSCRIPT (Kraulis, 1991), and rendered with PovRay (http://www.povray.org).

strand (β6 in Rep). Interestingly, two of these are also viral proteins: the DNA binding domain of the simian virus 40 (SV40) T antigen (Luo et al., 1996), and the DNA binding domain of the replication initiation protein E1 from bovine papillomavirus (BPV) (Enemark et al., 2000). When these proteins are structurally superimposed (Jones et al., 1991), the Rep endonuclease domain aligns with the DNA binding domain of the SV40 T antigen with an rmsd of 1.8 Å over 59 α-carbon positions, and that of BPV E1 with an rmsd of 2.1 Å over 73 α-carbon atoms. The structural homology is not reflected at the amino acid level as there is no significant primary sequence identity between the three proteins (Figure 4B). In particular, the viral homologs have neither the HUH motif nor the conserved tyrosine motif that are hallmarks of the RCR superfamily, and have no catalytic activities. SV40 T antigen and BPV E1 function by binding to their respective origins of replication (*ori*), where they initiate the assembly of higher order complexes to facilitate local distortion and DNA unwinding and recruit other proteins to the replication origin.

One marked structural distinction of the Rep endonuclease domain when compared with the origin binding domains of SV40 T antigen and BPV E1 appears to be associated with its capacity for catalysis. In Rep, there is an easily recognizable cleft harboring the active site. This cleft is missing from the SV40 and BPV origin binding domains because N-terminal extensions of about 25–30 residues not present in Rep fill the space that would otherwise be the active site cavities (Figure 5). In the case of E1, two helices occupy the cavity, while in the SV40 T antigen, there is one additional helix, and a shift of the structural equivalent of Rep helix αD (shown in orange in Figure 5) into the active site cleft.

Another distinct feature of the Rep endonuclease domain is a long meandering region between strand β1 and helix αB (residues 12 to 49), whose only regular secondary structural element is the short αA helix. This region is not present in the homologous viral origin binding proteins (see Figure 4B); rather, in these proteins there is a short, direct connection between the structural equivalents of β1 and helix αB. This region is also not conserved among RCR proteins. In all the AAV serotypes, however, this insert is unusually rich in acidic residues, giving rise in AAV-5 Rep to an extensive negatively charged surface (Figure 3).

The only other structural homolog of Rep that has an extended β sheet by virtue of a fifth C-terminal strand is the RRM-3 domain of the polypyrimidine tract binding protein (Conte et al., 2000), which is a regulator of pre-mRNA splicing. Structures of RRM proteins determined in the presence of mRNA show that mRNA is bound across the exposed face of the β sheet. In AAV Rep, a portion of the positively charged cleft is formed by one face of its central β sheet, suggesting that RRM proteins and Rep may share a common mode of nucleic acid binding.

The structure of another topologically related origin binding protein domain has been determined, that of the DNA binding domain of the initiator protein from the Epstein-Barr virus, EBNA1 (Bochkarev et al., 1996). However, EBNA1 does not possess the fifth β strand that distinguishes AAV Rep and its dsDNA viral origin binding domain homologs.

### The Active Site

The Rep endonuclease active site is established by the location of the active site tyrosine, Tyr-153 (Tyr-156 in

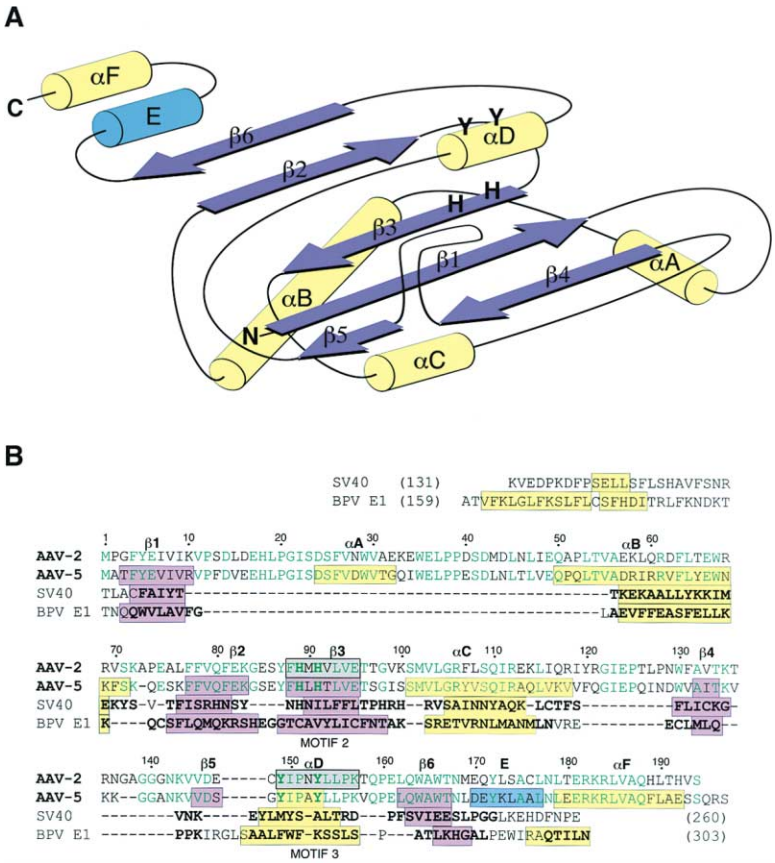


Figure 4. The AAV-5 Fold and Sequence Alignment with Related Viral Proteins

(A) Folding topology of AAV-5 Rep1–197. Strands are shown in purple and α helices in yellow. The two His residues of the HUH motif are located on strand β3, in close proximity to two conserved Tyr residues on helix αD.

(B) Structure-based sequence alignment of AAV-5 Rep1–197 with SV40 T antigen and BPV E1. The top sequence corresponds to AAV-2, and residues that are identical in AAV-5 are colored green. The numbering is that of AAV-5 Rep. RCR motifs 2 and 3 are highlighted in gray. The purple and yellow boxes indicate β strands and α helices, respectively; the blue box corresponds to the $3_{10}$ helix. The DNA binding domains of the SV40 T antigen (residues 131–260; Luo et al., 1996) and BPV E1 (residues 159–303; Enemark et al., 2000) were aligned in DALI (Holm and Sander, 1997). Residues in bold letters are those that structurally align with residues in AAV-5 Rep.

AAV-2), as shown in Figure 2B. Clustered nearby are the two histidine residues, His-89 and His-91, which comprise the HUH motif of the RCR superfamily. His-89 and His-91 contribute two of the four ligands to a zinc ion which is located 5.7 Å from the hydroxyl group of Tyr-153. The tetrahedrally coordinated zinc ion is bound in the bottom of the positively charged cleft, where its other two ligands are a side-chain carboxylate oxygen from Glu-82 and another from Asp-24 from a crystallographically related monomer. Thus, the active site is assembled by the spatial convergence of a divalent metal ion and amino acids from the two conserved RCR motifs.

The two imidazoles of the HUH motif, His-91 and His-89, are hydrogen-bonded to Glu-6 and the main chain carbonyl of Ala-94, respectively. This results in tautomerization states for the imidazole groups such that the NE1 nitrogens that point toward each other are not protonated and therefore are able to ligate the metal ion. Alanine scanning mutational studies of AAV-2 Rep have shown that Glu-6, Glu-83 (equivalent to AAV-5 Glu-82), and the two HUH histidines are not required for DNA binding, yet are essential for in vitro nicking activity and integration of an ITR-containing plasmid into *AAVS1* (Urabe et al., 1999). From the structure, we conclude that Glu-6 and Glu-82 are essential for the correct orientation of the two active site imidazoles. Both Glu residues are strictly conserved among AAV serotypes.

The two tyrosine residues of RCR motif 3 are located on the same face of a helix (αD) as previously hypothesized (van Mansfeld et al., 1986), pointing into the active site cleft. Although there is no convincing evidence for a catalytic role for the first Tyr in AAV Rep (Tyr-149 in AAV-5), both Tyr residues are required in other RCR systems such as phage $\phi$X174 gene A protein and protein A of phage P2, where they are believed to alternate in the cleavage and joining reactions (Hanai and Wang, 1993; Odegrip and Haggård-Ljungquist, 2001).

Another conserved component of RCR motif 3 is a Lys located three or four residues C-terminal of the second Tyr. In AAV-5 Rep, this residue, Lys-157, points into the active site. The equivalent residue in AAV-2 Rep, Lys-160, is not required for DNA binding, yet the Lys-to-Ala mutation is catalytically inactive (Urabe et al., 1999). Other nucleases that use nucleophilic tyrosines such as topoisomerases and tyrosine recombinases contain arrays of conserved basic residues that are crucial for catalysis (see Lima et al., 1994; Berger et al., 1996; Redinbo et al., 1998; Van Duyne, 2001). From the structure, Lys-157 does not appear associated with the metal binding site. Rather, Nϵ is hydrogen-bonded to the nearest oxygen atom which is a carboxylate oxygen of Asp-24 from an adjacent Rep monomer. In the absence of this crystal packing interaction, Lys-157 would be free to coordinate a phosphate oxygen of the substrate. Alternatively, it could serve as a general acid to reprotonate the 3′-hydroxy leaving group.

### The Requirement for a Divalent Metal Ion

Divalent metal ions such as $Mg^{2+}$ and $Mn^{2+}$ are often used as cofactors by nucleases, usually to activate the water nucleophile (Cowan, 1998). However, in nuclease

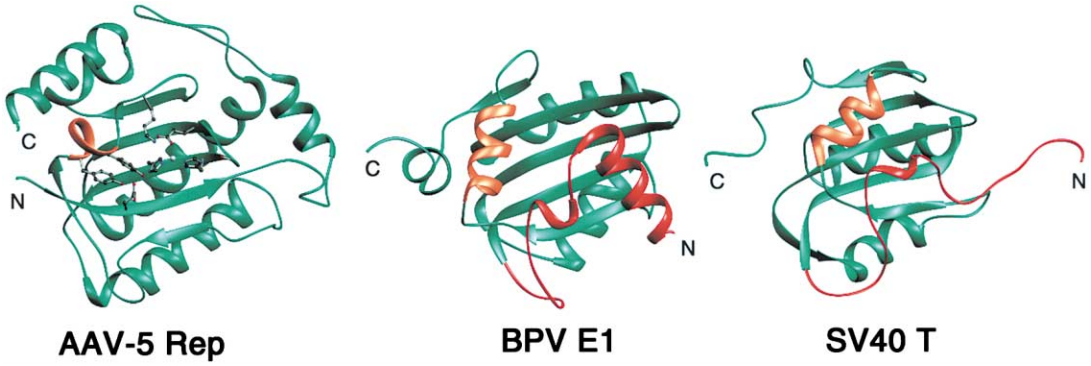


Figure 5. Comparison of the Folds of AAV-5 Rep1–197 and the DNA-Binding Domains of BPV E1 and SV40 T Antigen

In AAV-5 Rep1–197 (left), the active site residues shown in ball-and-stick are clustered in a cleft on one surface of the central β sheet. In BPV E1 (center) and the SV40 T antigen (right), N-terminal extensions (shown in red) occupy this region instead.

reactions where the -OH group of a tyrosine or a serine serves as the nucleophile, there is apparently no chemical need of a metal ion for activation. The simultaneous catalytic requirements of Rep for an active site tyrosine and a divalent metal ion for DNA cleavage is therefore puzzling.

To investigate the observed metal ion binding site, we soaked crystals for 24 hr in mother liquor containing 40 mM $MnCl_2$ ($\Delta f''$ of Mn at Cu K$\alpha$ = 2.808 $e^-$) in place of $Zn^{2+}$ ($\Delta f''$ at Cu K$\alpha$ = 0.702 $e^-$) prior to data collection. As an internal standard, we used the anomalous difference peak height of a $Cl^-$ ion bound close to the active site. The anomalous difference peak heights for the bound metal in the two crystallographically independent monomers in the $Mn^{2+}$-soaked crystals were 1.47 and 1.58 times higher than in the $Zn^{2+}$ crystals, suggesting that $Mn^{2+}$ substitution had occurred at the same site.

To directly confirm $Mn^{2+}$ binding to the N-terminal domain of Rep in solution, we performed isothermal titration calorimetry (ITC) experiments. Solutions of $MnCl_2$ were titrated against wild-type Rep1–197 and several active site point mutants; the results are shown in Figure 6A. $Mn^{2+}$ binds to Rep1–197 with a $K_D$ of 2.7 ± 0.6 μM, and the data can be modeled as a single binding site. Binding is exothermic with $\Delta H = -7.6$ kcal/mol; there is only a small entropic ($-T\Delta S = 0.1$ kcal/mol) contribution to $\Delta G$, the Gibbs free energy of binding, suggesting that metal binding does not cause major structural rearrangements. When either His-89 or His-91 is mutated to Ala or Gln, $Mn^{2+}$ binding is not detectable. $Mn^{2+}$ binding is also disrupted when Gln is substituted for Glu-82, demonstrating the importance of a carboxylate group in this position for metal binding, consistent with the structure. In contrast, mutating the catalytically essential Lys-157 to Ala resulted in a protein that retains $Mn^{2+}$ binding, suggesting a role for Lys-157 in stabilizing the pentavalent intermediate or possibly acting as a general acid, rather than participating in the assembly of the metal binding site. Binding of $Mg^{2+}$ to Rep1–197 could not be detected by ITC, consistent with the reported inability of $Mg^{2+}$ to support ssDNA cleavage by C-terminally truncated versions of Rep (Yoon et al., 2001).

DNA cleavage assays were also performed to assess the contribution of the active site residues to the DNA cleavage reaction. Although alanine scanning mutagenesis had previously demonstrated the importance of His-89 and His-91 for DNA cleavage and integration (Urabe et al., 1999), we sought to establish if conservative substitution of His with the nearly isosteric residue Gln could support cleavage. We used AAV-5 Rep1–489 rather than Rep1–197 to more readily detect low levels of activity. Although the H89Q and E82Q mutants of Rep1–489 are catalytically inactive, we could measure a low level of $Mn^{2+}$-dependent cleavage activity with the H91Q mutant (Figure 6B). This is consistent with the proposed role of His-91 in metal chelation since Gln retains the ability to coordinate $Mn^{2+}$ (Harding, 2002). The low activity correlates with our inability to detect $Mn^{2+}$ binding to this mutant by ITC.

As with AAV Rep, type IA topoisomerases form a 5′-phosphotyrosine linkage upon DNA cleavage. In these enzymes, a constellation of conserved basic residues together with the nucleophilic tyrosine are sufficient for cleavage; a metal ion is not required. A detailed view of this is provided by the structure of an *E. coli* type IA enzyme complexed with ssDNA, in which two positively charged residues, Lys-8 and Arg-330, coordinate the scissile phosphate from opposite sides of the active site (Changela et al., 2001). The two basic residues are similarly oriented relative to the active site Tyr, as are Lys-157 and $Zn^{2+}$ in AAV Rep. It is tempting to speculate that the metal cation seen bound to Rep could play a role in catalysis equivalent to that of Lys-8 in the topoisomerase, i.e., by contributing to the proper positioning of the scissile phosphate or to transition state stabilization. Thus, we favor the notion that the role of $Mn^{2+}$ is structural. It should be noted that, other than the residues of Rep explicitly discussed here, there are no other residues in the vicinity of Tyr-153 that could obviously contribute to catalysis. For example, we cannot identify a residue that might serve as a general base to deprotonate the catalytic tyrosine.

## DNA Binding

A particularly notable feature of AAV Rep is the extensive positively charged region that includes the active site cleft (Figure 3). This surface is gated by the unusual

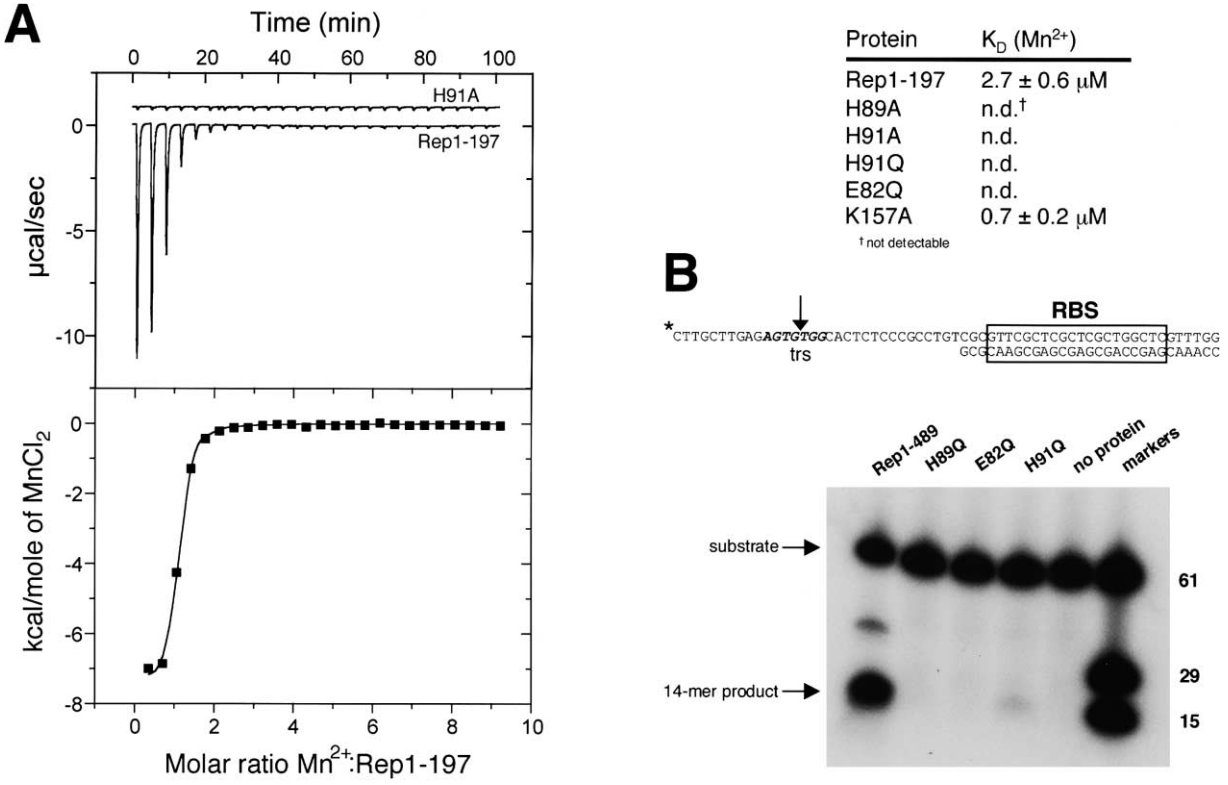


Figure 6. Biochemical Characterization of AAV-5 Rep

(A) Binding of $Mn^{2+}$ to the Rep endonuclease domain measured by isothermal titration calorimetry. Top: Raw data for injection of $MnCl_2$ into Rep1–197 and the point mutant H91A (the data for H91A are offset slightly for clarity). Bottom: Integrated heats of injection for the wild-type data, with the solid line showing the best fit.

(B) Cleavage activity of AAV-5 Rep1–489 on partially single-stranded substrates. Reactions were carried out in a volume of 10 µl containing 40 mM Tris-HCl (pH 8), 50 mM NaCl, 0.1 mg/ml BSA, 1 mM DTT, 4 mM $MnCl_2$, 0.34 µM protein, and 1 pmol substrate. After incubation for 1 hr at 37°C, reactions were quenched at 90°C for 5 min after addition of 20 µl formamide gel loading buffer. Reaction products were resolved on a 10% polyacrylamide/7 M urea denaturing gel. The minor cleavage product seen with Rep1–489 likely represents cleavage at a suboptimal site.

protrusion consisting of residues 135–144 between strands β4 and β5 which appears to restrict access to the active site to ssDNA. The protrusion includes three Lys residues. The basic residues Lys-135 and Arg-137 in this region of AAV-2 Rep are required for DNA binding (Urabe et al., 1999). Facing this protrusion from the opposite side of the cleft is Arg-10, whose equivalent in AAV-2, Lys-10, is also needed for DNA binding (Urabe et al., 1999). A similar protruding structure has been observed in complexes of spliceosomal U1A and U2B″/U2A′ proteins with snRNA hairpins (Oubridge et al., 1994; Price et al., 1998). In these complexes, the protruding region is inserted into the loop of the RNA stem-loop structure.

The structure of AAV Rep is consistent with a model for substrate binding in which the positively charged surface of Rep interacts with DNA, some portion of which is likely to be single-stranded to be accommodated in the active site, as shown in Figure 3B. The path that DNA takes across the face of the β sheet may well be similar to the mode in which RNA is bound by RRM domains (Conte et al., 2000) and is probably influenced by the prominent surface loop, supporting the notion that AAV Rep cleaves the trs in the context of ssDNA or a stem-loop structure (Smith and Kotin, 2000; Brister and Muzyczka, 2000). This model for ssDNA binding is speculative; the recent structures of BPV E1 bound to dsDNA (Enemark et al., 2002) provide tantalizing clues to the way in which Rep may bind the other required component of its substrate, the RBS sequence.

## Conclusion

It has been a decade since an analysis of the primary sequences of plasmids and viruses revealed the existence of a superfamily of proteins sharing three sequence motifs which were proposed to be united by the mechanism of rolling circle replication (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993). The wide dispersal of the RCR domain among virtually all forms of life, from phages that infect bacteria to animal viruses which tend to coexist benignly with their hosts, indicates that it is one of the most ancient of all protein domains (Shadan and Villarreal, 1996). The three-dimensional structure of a catalytically active RCR domain presented here offers the possibility to follow the evolution of an ancient protein domain; it also unexpectedly unites the RCR superfamily with another group of replication proteins, dsDNA viral origin binding proteins.

The unanticipated link between the endonuclease domain of AAV Rep and the origin binding proteins of other viruses demonstrates how a common fold that has been widely employed can be fine-tuned in different contexts to perform specialized tasks in replication. For example, in the case of the ssDNA virus AAV, the unique hairpin structures at the ends of its genome allow AAV to bypass the priming step of DNA replication: at one end of the genome, the free 3′-OH is in place as a consequence of folding, while at the other end, the virally encoded Rep protein generates the necessary 3′-OH. In contrast, the homologous domain of small dsDNA viruses has only a specific DNA binding function. This is permitted since DNA cleavage activity is not needed in the context of complementary separated strands where two replication forks can readily assemble.

The SV40 T antigen and papillomavirus E1 protein form hexamers on their DNA recognition sites (Patel and Picha, 2000), as does AAV Rep on substrates containing RBS and trs sequences (Smith et al., 1997). This propensity for hexamerization likely reflects a second common structural feature of these proteins, a helicase domain that appears C-terminal to their DNA binding domains. The helicase domains of small DNA and RNA viruses are evolutionarily related and form a distinct helicase superfamily: superfamily III, or SF3 (Gorbalenya et al., 1990; Koonin, 1993). SF3 helicases share only three conserved motifs contained within a region of ~100 amino acids; two of these motifs are the Walker A and B sequences of NTP binding proteins. The structure of the Rep endonuclease domain reveals that the N-terminal domains of certain small DNA viruses, which exhibit a range of functions, are also evolutionarily related.

It is generally accepted that among the small DNA viruses, virus evolution is tightly linked to host evolution (Shadan and Villarreal, 1996). It is therefore interesting to note the changes that have accompanied the RCR domain in the context of the evolution of higher organisms. One of the simplest RCR proteins, gene A protein of phage $\phi$X174 which infects bacteria, includes all three conserved sequence motifs, requires both active site tyrosines of motif 3 for replication, and must recruit a separate cellular helicase to its replication origin. The plant geminiviral Rep proteins which retain the three RCR motifs have acquired a helicase domain and lost the requirement for the first Tyr residue of motif 2 (Laufs et al., 1995). Parvoviruses such as AAV which infect vertebrates have lost the first conserved RCR motif while retaining the helicase domain. We propose, based on our structural result and bolstered by functional homology, that evolution of an ancient RCR protein domain has also led to the catalytically inactive origin binding proteins represented structurally by the DNA binding domains of the SV40 T antigen and papillomavirus E1. This newly identified link between these virus classes will provide a framework for further investigation into the mechanism and regulation of site-specific AAV integration.

## Experimental Procedures

### Protein Purification and Crystallization

The Rep gene of AAV-5 was cloned (Veritas, Inc., Rockville, MD) from pGEM-3Z, a gift of J.A. Chiorini, into pET-15b, and a stop codon introduced after Ser-197 (QuikChange, Stratagene). Optimal soluble protein expression was obtained in *E. coli* strain BL21(DE3) by growth at 37°C in LB broth until $OD_{600nm} = 0.6$, followed by induction at 19°C by addition of IPTG to a final concentration of 0.1 mM. Cells were harvested 16 hr later and resuspended in 25 mM Tris (pH 7.5), 0.1 mM EDTA. Subsequent steps were at 4°C unless otherwise noted. After sonication in the presence of 0.5 M NaCl and centrifugation, soluble material was loaded onto a POROS MC/M column (PerSeptive Biosystems) previously equilibrated with $NiSO_4$. The column was washed extensively with 20 mM Tris (pH 7.5), 2 mM imidazole (Im), and 0.5 M NaCl followed by the same buffer containing 10 mM Im. Rep1–197 was eluted from the column using a gradient from 10–400 mM Im. Fractions containing Rep were combined and dialyzed with several changes against 20 mM Tris (pH 7.5), 1 mM EDTA, 0.5 M NaCl, and 10% (w/v) glycerol. To remove the polyhistidine tag, 6 units thrombin (Sigma) were added per mg protein. Thrombin was subsequently removed by passage over a 0.5 ml benzamidine Sepharose 4B (Pharmacia) column. The recovered protein was concentrated prior to size-exclusion chromatography on a TSK-Gel G3000SW column. Fractions containing Rep were collected, dialyzed against 20 mM Tris (pH 7.5), 0.5 M NaCl, and 1 mM EDTA, and concentrated to ~8 mg/ml. The final yield from 4 liters cells was ~25 mg Rep. Crystals were grown at 20°C by the hanging drop method by mixing Rep1–197 in a 1:1 ratio with 10% (v/v) PEG 400, 80 mM sodium cacodylate (pH 6.5), 24 mM zinc acetate, and 10% (v/v) PEG 8000; the well contained 100 mM cacodylate, 30 mM zinc acetate, and 12.5% PEG 8000. Crystals grew to full size in ~2 days and were cryoprotected by transferring the crystals into 20 mM zinc acetate, 50 mM cacodylate (pH 6.5), 10 mM Tris (pH 7.5), 0.25 M NaCl, 0.5 mM EDTA, 12.5% PEG 8000, and 20% (w/v) glycerol for 20–30 s prior to flash-cooling in liquid propane. The crystals are of space group $P2_1$ ($a = 42.90$ Å, $b = 73.47$ Å, $c = 65.60$ Å, $\beta = 94.48°$) with two molecules in the asymmetric unit.

### Structure Determination and Refinement

Diffraction data were collected at beamline ID-19 at the Advanced Photon Source, on the APS-2 mosaic CCD detector. Additional data sets using Cu K$\alpha$ (8.04 keV) were collected for the $Mn^{2+}$ soaking experiments. Three X-ray energies (9.664, 9.669, and 12.284 keV) were used to take advantage of the dispersive and anomalous components of the $Zn^{2+}$ structure factor for phase determination. (The computed energy of the K absorption edge of unbound, elemental Zn is 9.661 keV). The six $Zn^{2+}$ positions in the asymmetric unit were located with direct methods using SHELXD (Sheldrick, 1998) and utilizing the magnitudes of the $F_a$ (X-ray energy independent) components of the Zn structure factors. MAD was treated as a special case of Single Isomorphous Replacement with Anomalous Scattering (Ramakrishnan et al., 1993), with the data set collected at 9.664 keV used as native, and the Bijvoet differences in the data set collected at 9.669 keV as a source of anomalous scattering phase information. The dispersive differences between the 12.284 and 9.664 keV data sets were the source of isomorphous phase information. Positional occupancy and thermal parameters of the $Zn^{2+}$ ions were refined with phase integrating least squares as implemented in PHASES (Furey and Swaminathan, 1997). The solvent-flattened (Wang, 1985) experimental map calculated at 2.0 Å resolution was readily interpretable for both monomers in the asymmetric unit. The model was built with O (Jones et al., 1991). The structure was refined using the data set collected at 12.284 keV with molecular dynamics, energy minimization, and individual B factor refinement at 1.4 Å resolution. 1143 reflections selected randomly were used to monitor $R_{free}$, and were never used in refinement. Bulk solvent correction and the TNT B restraint library (Tronrud, 1996) were used throughout. The current model contains all residues between Met-1 and Asp-193, six $Zn^{2+}$ ions, two bound $Cl^-$ ions, and 682 water molecules that were added at the final stages of the refinement. 91% of the residues are in the most favored region of the Ramachandran plot, with no residues in the disallowed regions.

### DNA Cleavage Assays

Cleavage assays were performed using AAV-5 Rep residues 1–489. Point mutations were introduced using the QuikChange kit, and the entire Rep gene sequenced to confirm that no additional mutations

had been introduced. Proteins were purified as described for Rep1–197. To prepare the partially single-stranded substrate, 200 pmoles of a 61-mer (Figure 6B) containing the AAV-5 trs and RBS sites (Chiorini et al., 1999b) was 5′ end labeled with 50 μCi γP-32 ATP and T4 polynucleotide kinase (New England Biolabs). Unincorporated nucleotide was removed using a Sephadex G25 spin column, and the 61-mer recovered in 100 μl TE. A 29-mer containing the complementary RBS sequence was annealed to the labeled oligonucleotide in 200 μl buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, and 1 mM DTT [pH 7.9]) by heating to 65°C and slow cooling. Assays conditions are described in the legend to Figure 6.

### Isothermal Titration Calorimetry

Binding of $Mn^{2+}$ to Rep1–197 and the point mutants H89A, H91A, H91Q, E82Q, and K157A was measured by isothermal titration calorimetry using an MCS titration calorimeter (MicroCal Inc.). Rep1–197 was used to prevent any interference by metal binding to the helicase domain. Aliquots of 5 μl 10 mM $MnCl_2$ or 50 mM $MgCl_2$ (Sigma-Ultra grade) were titrated at 20°C by injection into protein (25–105 μM in 1.347 ml) in 20 mM HEPES (pH 7.6), 0.5 M NaCl, 10% (w/v) glycerol. Calorimetric data were analyzed using the Origin package. The $K_D$ value reported for Rep1–197 is the average of four measurements from two independent protein preparations.

### Acknowledgments

We thank Z. Dauter for assistance with data collection at NSLS X9B, S. Chacko for help with databases, J. Chiorini for providing pET-28 Rep68Δ and AAV-5 pGEM-3Z clones, R. Ghirlando for performing the sedimentation equilibrium experiments, and T. Obsil for assistance with ITC. We are grateful to L. Yang for performing the activity assays.

Received: March 15, 2002
Revised: May 29, 2002

### References

Bantel-Schaal, U., Delius, H., Schmidt, R., and zur Hausen, H. (1999). Human adeno-associated virus type 5 is only distantly related to other known primate helper-dependent parvoviruses. J. Virol. *73*, 939–947.

Berger, J.M., Gamblin, S.J., Harrison, S.C., and Wang, J.C. (1996). Structure and mechanism of DNA topoisomerase II. Nature *379*, 225–232.

Bochkarev, A., Barwell, J.A., Pfuetzner, R.A., Bochkareva, E., Frappier, L., and Edwards, A.M. (1996). Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA. Cell *84*, 791–800.

Brister, J.R., and Muzyczka, N. (2000). Mechanism of Rep-mediated adeno-associated virus origin nicking. J. Virol. *74*, 7762–7771.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. Science *265*, 615–621.

Carson, M. (1991). RIBBONS 2.0. J. Appl. Crystallogr. *24*, 958–961.

Changela, A., DiGate, R.J., and Mondragón, A. (2001). Crystal structure of a complex of a type 1A DNA topoisomerase with a single-stranded DNA molecule. Nature *411*, 1077–1081.

Chejanovsky, N., and Carter, B.J. (1989). Mutagenesis of an AUG codon in the adeno-associated virus rep gene: effects on viral DNA replication. Virology *173*, 120–128.

Chejanovsky, N., and Carter, B.J. (1990). Mutation of a consensus purine nucleotide binding site in the adeno-associated virus *rep* gene generates a dominant negative phenotype for DNA replication. J. Virol. *64*, 1764–1770.

Chiorini, J.A., Wiener, S.M., Owens, R.O., Kyöstiö, S.R.M., Kotin, R.M., and Safer, B. (1994a). Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. *68*, 7448–7457.

Chiorini, J.A., Weitzman, M.D., Owens, R.A., Urcelay, E., Safer, B., and Kotin, R.M. (1994b). Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in *Escherichia coli*. J. Virol. *68*, 797–804.

Chiorini, J.A., Kim, F., Yang, L., and Kotin, R.M. (1999a). Cloning and characterization of adeno-associated virus type 5. J. Virol. *73*, 1309–1319.

Chiorini, J.A., Afione, S., and Kotin, R.M. (1999b). Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. J. Virol. *73*, 4293–4298.

Christopher, J.A. (1998). SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual) (College Station, TX: Texas A&M University).

Conte, M.R., Grüne, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., and Curry, S. (2000). Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. EMBO J. *19*, 3132–3141.

Cowan, J.A. (1998). Metal activation of enzymes in nucleic acid biochemistry. Chem. Rev. *98*, 1067–1087.

Davis, M.D., Wu, J., and Owens, R.A. (2000). Mutational analysis of adeno-associated virus type 2 Rep68 protein endonuclease activity on partially single-stranded substrates. J. Virol. *74*, 2936–2942.

Enemark, E.J., Chen, G., Vaughn, D.E., Stenlund, A., and Joshua-Tor, L. (2000). Crystal structure of the DNA binding domain of the replication initiation protein E1 from papillomavirus. Mol. Cell *6*, 149–158.

Enemark, E.J., Stenlund, A., and Joshua-Tor, L. (2002). Crystal structures of two intermediates in the assembly of the papillomavirus replication initiation complex. EMBO J. *21*, 1487–1496.

Furey, W., and Swaminathan, S. (1997). PHASES-95: a program package for processing and analyzing diffraction data from macromolecules. Methods Enzymol. *277*, 590–620.

Gorbalenya, A.E., Koonin, E.V., and Wolf, Y.I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. FEBS Lett. *262*, 145–148.

Hanai, R., and Wang, J.C. (1993). The mechanism of sequence-specific DNA cleavage and strand transfer by ϕX174 gene A protein. J. Biol. Chem. *268*, 23830–23836.

Harding, M.M. (2002). Metal coordination groups in proteins. Some comments on geometry, constitution and B-values. CCP4 Newsletter on Protein Crystallography *40*.

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

Holm, L., and Sander, C. (1997). Dali/FSSP classification of three-dimensional protein folds. Nucleic Acids Res. *25*, 231–234.

Ilyina, T.V., and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Res. *20*, 3279–3285.

Im, D., 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.A., 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. A *47*, 110–119.

Khan, S.A. (2000). Plasmid rolling-circle replication: recent developments. Mol. Microbiol. *37*, 477–484.

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.

Koonin, E.V., and Ilyina, T.V. (1993). Computer-assisted dissection of rolling circle DNA replication. Biosystems *30*, 241–268.

Kotin, R.M. (1994). Prospects for the use of adeno-associated virus as a vector for human gene therapy. Hum. Gene Ther. *5*, 793–801.

Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X., 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.

Kotin, R.M., Menninger, J.C., Ward, D.C., and Berns, K.I. (1991). Mapping and direct visualization of a region-specific viral-DNA integration site on chromosome 19q13-qter. Genomics *10*, 831–834.

Kotin, R.M., Linden, R.M., and Berns, K.I. (1992). Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. *11*, 5071–5078.

Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F., and Gronenborn, B. (1995). Geminivirus replication: genetic and biochemical characterization of Rep protein function, a review. Biochim. *77*, 765–773.

Lima, C.D., Wang, J.C., and Mondragón, A. (1994). Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. Nature *367*, 138–146.

Luo, X., Sanford, D.G., Bullock, P.A., and Bachovchin, W.W. (1996). Solution structure of the origin DNA-binding domain of SV40 T-antigen. Nat. Struct. Biol. *3*, 1034–1039.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. *247*, 536–540.

Odegrip, R., and Haggård-Ljungquist, E. (2001). The two active-site tyrosine residues of the A protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage P2. J. Mol. Biol. *308*, 147–163.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.-H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. Nature *372*, 432–438.

Owens, R.A., Trempe, J.P., Chejanovsky, N., and Carter, B.J. (1991). Adeno-associated virus rep proteins produced in insect and mammalian expression systems: wild-type and dominant-negative mutant proteins bind to the viral replication origin. Virology *184*, 14–22.

Patel, S.S., and Picha, K.M. (2000). Structure and function of hexameric helicases. Annu. Rev. Biochem. *69*, 651–697.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA. Nature *394*, 645–650.

Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L., and Sweet, R.M. (1993). Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. Nature *362*, 219–223.

Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G.J. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. Science *279*, 1504–1513.

Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Housman, 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.

Shadan, F.F., and Villarreal, L.P. (1996). The evolution of small DNA viruses of eukaryotes: past and present considerations. Virus Genes *11*, 239–257.

Sheldrick, G.M. (1998). SHELX: applications to macromolecules. In Direct Methods for Solving Macromolecular Structures, S. Fortier, ed. (Dordrecht: Kluwer Academic Publications), pp. 401–411.

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

Smith, R.H., and Kotin, R.M. (2000). An adeno-associated virus (AAV) initiator protein, Rep78, catalyzes the cleavage and ligation of single-stranded AAV *ori* DNA. J. Virol. *74*, 3122–3129.

Smith, R.H., and Kotin, R.M. (2002). Adeno-associated virus. In Mobile DNA II, N.L. Craig, R. Craigie, M. Gellert, and A.M. Lambowitz., eds. (Washington, DC: ASM Press).

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

Snyder, R.O., Im, D.S., Ni, T.H., Xiao, X., Samulski, R.J., and Muzyczka, N. (1993). Features of the adeno-associated virus origin involved in substrate recognition by the viral Rep protein. J. Virol. *67*, 6096–6104.

Tronrud, D.E. (1996). Knowledge-based B-factor restraints for the refinement of proteins. J. Appl. Crystallogr. *29*, 100–104.

Urabe, M., Hasumi, Y., Kume, A., Surosky, R.T., Kurtzman, G.J., Tobita, K., and Ozawa, K. (1999). Charged-to-alanine scanning mutagenesis of the N-terminal half of adeno-associated virus type 2 Rep78 protein. J. Virol. *73*, 2682–2693.

Van Duyne, G.D. (2001). A structural view of Cre-*loxP* site-specific recombination. Annu. Rev. Biophys. Biomol. Struct. *30*, 87–104.

van Mansfeld, A.D.M., van Teeffelen, H.A.A.M., Baas, P.D., and Jansz, H.S. (1986). Two juxtaposed tyrosyl-OH groups participate in $\phi$X174 gene A protein catalysed cleavage and ligation of DNA. Nucleic Acids Res. *14*, 4229–4238.

Wang, B.-C. (1985). Resolution of phase ambiguity in macromolecular crystallography. Methods Enzymol. *115*, 90–112.

Yoon, M., Smith, D.H., Ward, P., Medrano, F.J., Aggarwal, A.K., and Linden, R.M. (2001). Amino-terminal domain exchange redirects origin-specific interactions of adeno-associated virus Rep78 in vitro. J. Virol. *75*, 3230–3239.

## Accession Numbers

Coordinates have been deposited with the Protein Data Bank (accession number 1M55).