Molecular Cell, Vol. 13, 403–414, February 13, 2004, Copyright ©2004 by Cell Press

The Nuclease Domain of Adeno-Associated Virus Rep Coordinates Replication Initiation Using Two Distinct DNA Recognition Interfaces

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

Summary

Integration into a particular location in human chromosomes is a unique property of the adeno-associated virus (AAV). This reaction requires the viral Rep protein and AAV origin sequences. To understand how Rep recognizes DNA, we have determined the structures of the Rep endonuclease domain separately complexed with two DNA substrates: the Rep binding site within the viral inverted terminal repeat and one of the terminal hairpin arms. At the Rep binding site, five Rep monomers bind five tetranucleotide direct repeats; each repeat is recognized by two Rep monomers from opposing faces of the DNA. Stem-loop binding involves a protein interface on the opposite side of the molecule from the active site where ssDNA is cleaved. Rep therefore has three distinct binding sites within its endonuclease domain for its different DNA substrates. Use of these different interfaces generates the structural asymmetry necessary to regulate later events in viral replication and integration.

Introduction

Gene therapy is an approach to treating disease in which an exogenous gene is introduced to correct for a defective or missing protein or to affect a biochemical pathway. Few successes have been reported in humans (Hacein-Bey-Abina et al., 2002), as several technical issues limit its broader application. For example, one question is how to deliver DNA to the appropriate cells. Nature provides one solution in the form of viruses, which are in essence protected gene delivery packages with native ability to introduce their genomes into cells. Once the desired gene is delivered to target cells, another issue that arises is the fate of the DNA. Some strategies rely on long-term expression from extrachromosomal DNA, but there are cases, such as dividing cells, where it would be highly beneficial to permanently insert the gene into chromosomes.

Certain viral genomes can be integrated into host DNA by nonhomologous recombination or, in the case of retroviruses, by virally encoded integrases. While integration is not dependent on target sequence, in vivo, retroviruses, such as HIV and murine leukemia virus, integrate preferentially into active genes (Schröder et al., 2002; Wu et al., 2003), introducing the possibility of insertional mutagenesis. The theoretical danger inherent in retrovirus-based gene therapy has been concretely demonstrated in a recent clinical trial in which the modified retrovirus integrated into the *LMO2* locus, causing leukemia in two of the patients (Hacein-Bey-Abina et al., 2003).

To date, only one animal virus—the adeno-associated virus (AAV)—has been identified that integrates its genome into a particular location in human chromosomal DNA. When cells are infected in the absence of helper virus, AAV establishes a latent infection in which the AAV genome integrates into a locus known as AAVS1 on the q arm of chromosome 19 (Kotin et al., 1990, 1992; Samulski et al., 1991; reviewed in Smith and Kotin, 2002). This targeted integration by AAV requires the viral *rep* gene that encodes the nonstructural Rep protein. Rep is an essential viral protein and catalyzes several reactions during viral replication. The N-terminal domain of Rep, consisting of the first ~200 amino acids, has site-specific endonuclease and sequence-specific DNA binding activities (Im and Muzyczka, 1989, 1990; Owens et al., 1993; Chiorini et al., 1994a, 1994b; Davis et al., 2000). The C-terminal region of Rep possesses ATPase and 3′–5′ helicase activities (Im and Muzyczka, 1990; Chiorini et al., 1994a; Wonderling et al., 1995; Smith and Kotin, 1998) and is a member of the SF3 helicase superfamily (Gorbalenya et al., 1990). The structures of the isolated domains have recently been described (Hickman et al., 2002; James et al., 2003).

The varied activities of Rep are needed for reactions on unique DNA structures known as inverted terminal repeats (ITRs) found at the ends of the single-stranded viral genome that serve as the viral origins of replication (Figure 1). Each ITR contains interrupted palindromic sequences that allow the formation of a three-way DNA junction with two short (~9 bp) hairpin arms. Within the ITRs are two sequences required for replication: a *R*ep *b*inding *s*ite (RBS) consisting of several direct repeats of a 5′-GCTC-3′ motif and a *t*erminal *r*esolution *s*ite, or trs. Viral replication requires Repbinding at the RBS and subsequent cleavage of the top strand at the trs to generate the 3′-OH group so that the viral ends can be converted into linear duplex DNA. Cleavage at the trs is strongly stimulated by the presence in *cis* of one of the ITR hairpin arms (Chiorini et al., 1994a, 1994b; McCarty et al., 1994a; Ryan et al., 1996; Wu et al., 1999; Brister and Muzyczka, 2000; Wu et al., 2001), an effect attributed to a five base sequence known as the RBE′, a *R*ep *b*inding *e*lement, at the tip of the hairpin. The endonuclease domain recognizes its trs substrate in the context of ssDNA or a stem loop generated by the Rep helicase activity (Im and Muzyczka, 1990; Snyder et al., 1993; Brister and Muzyczka, 1999; Smith and Kotin, 2000). Two of the important DNA sequences within the ITR, the RBS and the trs, also occur at AAVS1, suggesting that the same Rep-mediated DNA recognition

*Correspondence: dyda@ulti.niddk.nih.gov

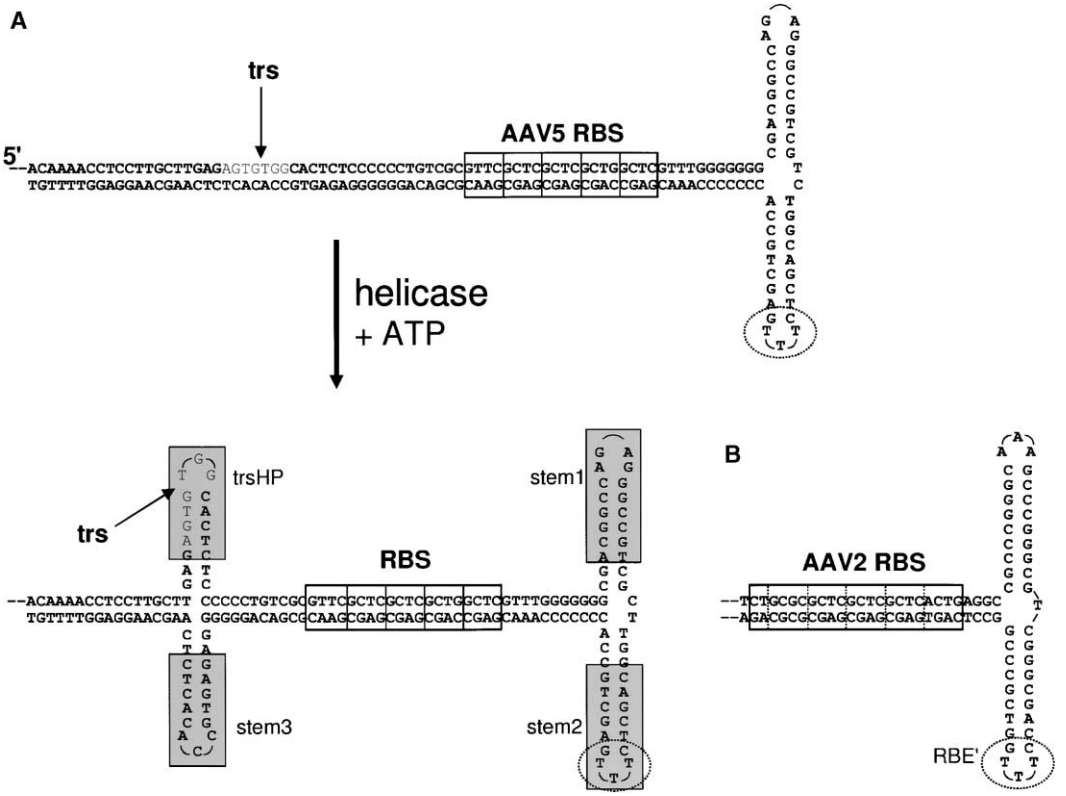

Figure 1. Sequence and Folded Structure of the AAV Inverted Terminal Repeat

(A) One arm of the AAV5 ITR contains the terminal resolution site (trs), at which Rep introduces a strand- and sequence-specific nick, and the Rep binding site (RBS), which contains a series of tetranucleotide repeats (shown in boxes; Chiorini et al., 1999). The other two arms of the three-way DNA junction consist of short hairpins. The helicase activity of Rep generates an alternate structure around the trs in which the cleavage site is single stranded. The four shaded boxes correspond to stem loops tested for Rep binding.

(B) Sequence of the AAV2 ITR. The box indicates the 22 bp binding element identified by Ryan et al. (1996); dotted lines demark tetranucleotide repeats. The circled tip of one of the hairpin arms is designated RBE′ (Brister and Muzyczka, 2000).

and cleavage occur during integration. The 33 bp at AAVS1 that encompass these two sequences are necessary and sufficient for targeted integration (Linden et al., 1996).

In current AAV-based transgene systems, *rep* is deleted since the small viral capsid does not have space for an exogenous gene; retaining *rep* would result in a genome too large to be packaged (Flotte and Carter, 1995). Thus, *rep*$^-$ AAV-based vectors, which retain other inherent advantages of AAV, either do not integrate or integrate inefficiently in a nontargeted manner, a process that has been reported to occur preferentially into active genes (Nakai et al., 2003). Promising alternatives are hybrid vectors that combine AAV ITRs and *rep* with a second virus with larger transgene capacity, such as adenovirus (Recchia et al., 1999), herpes simplex virus (Costantini et al., 1999), or baculovirus (Palombo et al., 1998).

Integration into a specific chromosomal location would be an extremely valuable asset to gene therapy, as it bypasses targeting problems associated with other delivery systems. Continued improvement of gene delivery systems will clearly benefit from a detailed understanding of how AAV Rep recognizes its DNA substrates: its own genome and the site of integration. To this end, we have determined two crystal structures of the N-terminal domain of AAV5 Rep complexed with specific regions of viral ITRs. These structures provide a foundation for a model for Rep assembly on the ITRs, a necessary first step in both AAV replication and integration.

Results and Discussion

Complex Formation, Crystallization, and Structure Determination

Encouraged by observations that the N-terminal domain of Rep contains the determinants for ITR binding (Owens et al., 1993) and specifically binds ITR sequences (Yoon et al., 2001), we have obtained crystals of AAV5 Rep residues 1–197 separately complexed with a double-stranded 26-mer containing the AAV5 Rep binding site ("RBS26") and a 15-mer (denoted stem2) consisting of a stem-loop sequence of one of the ITR hairpin arms. This hairpin arm corresponds to the BB′ palindrome. Although different lengths of DNA incorporating the AAV5 RBS were tried, only the 26-mer yielded crystals.

Stable complex formation between Rep197 and RBS26 was demonstrated by comigration through a Superdex 200 column (data not shown). We similarly attempted to detect complexes of Rep197 with stem-loop sequences (shaded boxes in Figure 1A) representing the two ITR hairpin arms and the proposed hairpins

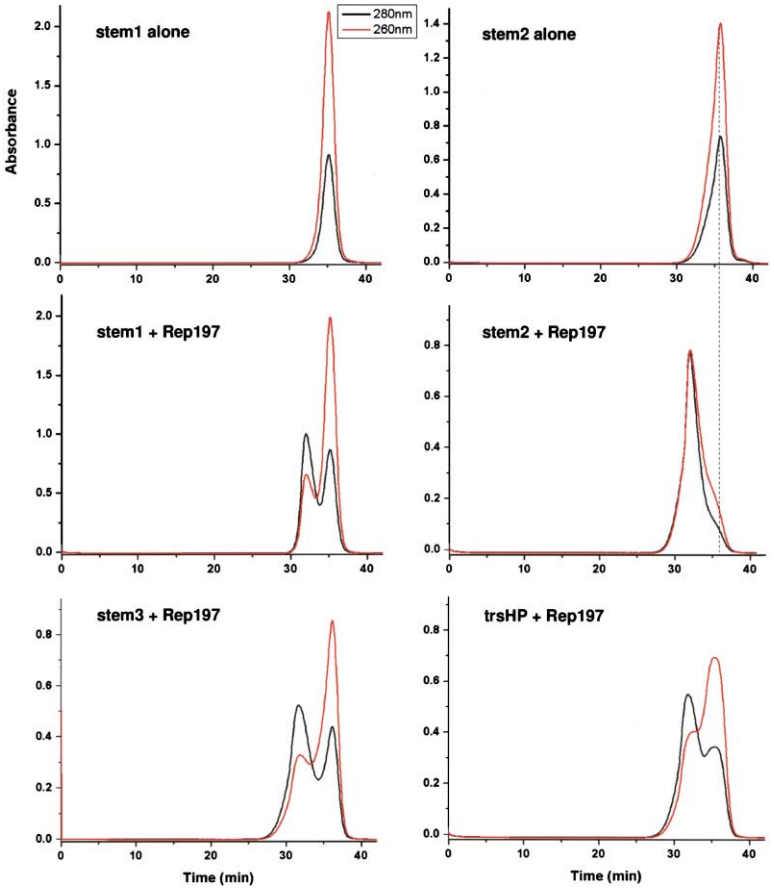

Figure 2. The Rep N-Terminal Domain Specifically Binds One Hairpin Arm of the ITR

Shown are gel filtration elution profiles of AAV5 Rep197 with four possible stem loops of the ITR. DNA and protein were mixed in a 0.5:1 molar ratio, so that binding would result in incorporation of all the DNA into a higher molecular weight complex. Only stem2 (middle, right) comigrates with Rep197. The slight broadening of the trsHP peaks may indicate weak binding accompanied by dissociation during gel filtration.

(Brister and Muzyczka, 1999; Smith and Kotin, 2000) around the trs site. Only the stem loop containing the RBE′ (stem2) formed a stable complex (Figure 2).

The structures of both complexes, Rep197/RBS26 and Rep197/stem2, were solved by molecular replacement using uncomplexed AAV5 Rep197 (Hickman et al., 2002) as the search model (see Experimental Procedures and Table 1).

The Structure of Rep197 Bound to RBS26

The asymmetric unit of the Rep197/RBS26 crystals contains one molecule of RBS26 and five Rep197 monomers. The DNA possesses an overall B-DNA structure and is packed blunt end to blunt end in the crystal lattice. The five Rep197 molecules (designated AA–AE in Figure 3A) are bound to RBS26 and spiral around the DNA axis, offset from one another by four base pairs, or ~138°. As a result, each Rep197 molecule binds RBS26 independently. This binding mode reveals a surprisingly intricate structural basis of the observation that a repeated tetranucleotide sequence is necessary for recognition by Rep (Chiorini et al., 1994a; McCarty et al., 1994a). For purposes of discussion, we refer to the GCTC sequence as a "perfect" tetranucleotide repeat.

Recognition of the GCTC Tetranucleotide Repeat

The two structural elements of Rep197 that are important for RBS binding are the surface loop between strands β4 and β5 (residues 135–144) and helix αC (residues 101–118). These regions of Rep197 are located along one edge of the central β sheet, and the bound monomers are oriented such that all of the active sites face in the direction of the trs (Figure 3A).

The β4/β5 loop interacts with four bases from the

Table 1. Data Collection and Processing Statistics

Data Set	Rep197/RBS26	Rep197/stem2
Wavelength (Å)	1.000	1.5418
Resolution (Å)	3.1	2.0
Total reflections (N)	72,828	158,265
Unique reflections (N)	23,005	38,009
Completeness (%) (for $I/\sigma I > -3$)	87.8	89.7
$I/\sigma I$	6.2	19.7
Rsym	0.10	0.063
Refinement		
Resolution (Å)	20–3.1	30–2.0
Atoms (N)	8965	4112
Reflections $F > 0\ \sigma(F)$	22,393	37,161
R factor (%)	28.7	21.5
Rfree (%)	31.6	25.7
Average B factor (Å^2)	37.6	27.0
Rms ΔB bonded atoms (Å^2)	1.387	3.951
Rms bond lengths (Å)	0.010	0.006
Rms bond angles (°)	1.33	1.35
# water molecules	–	316
# Mg^{2+} ions	–	4

Rsym = $\Sigma|I - \langle I\rangle|/\Sigma\langle I\rangle$; R factor = $\Sigma|FP_o - FP_c|/\Sigma|FP_o|$.

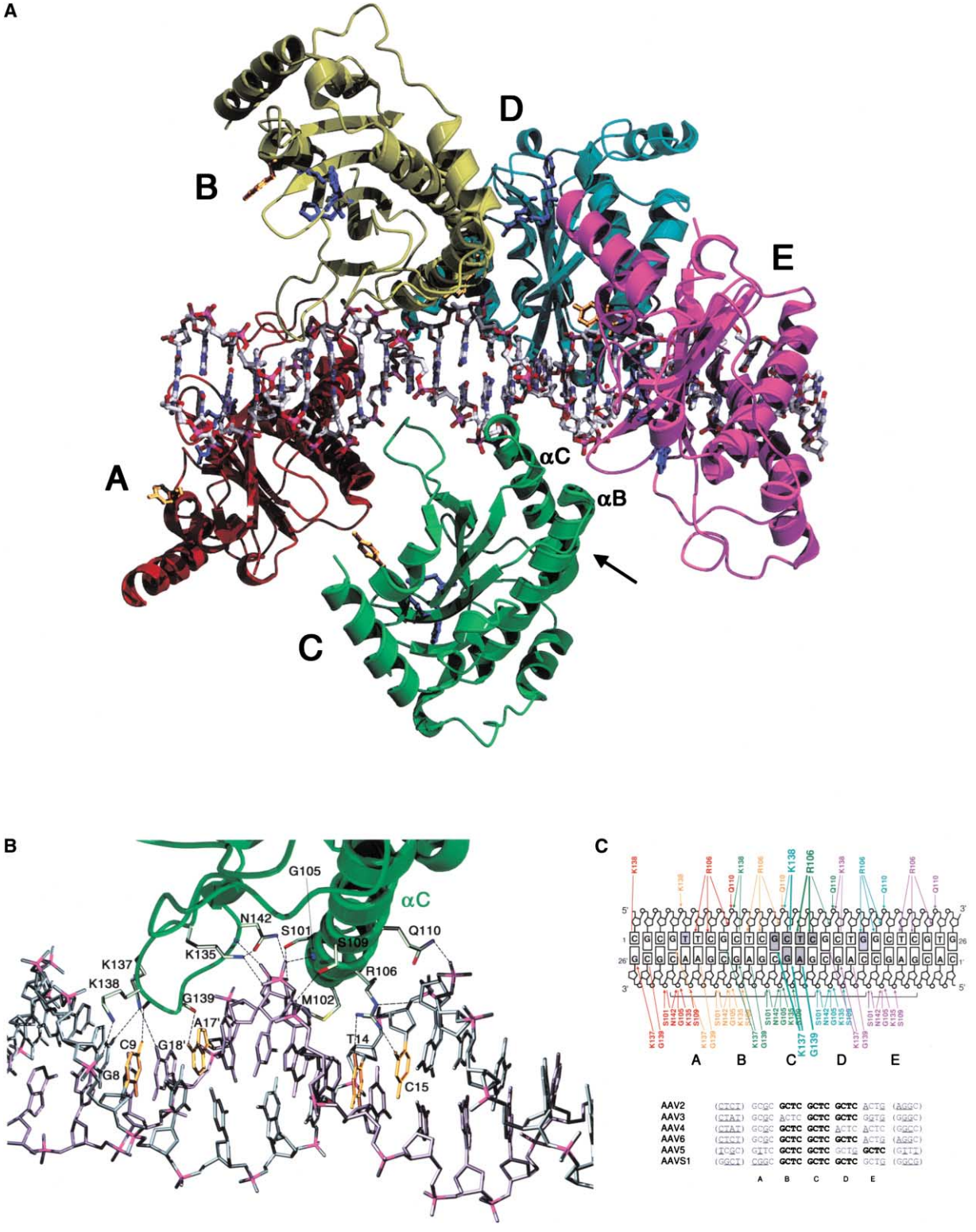

Figure 3. The Rep N-Terminal Domain Binds the RBS with a Four Base Pair Repeat

(A) RBS26 bound by five Rep197 monomers. In each monomer, the active site tyrosine, Tyr-153, is shown in gold and residues that bind the catalytically essential Mg^{2+} ion in dark blue. Two helices (αB and αC) that form the hairpin binding site (indicated by arrow) are labeled in monomer C.

(B) Major and minor groove interactions involving monomer C. Note that the two phosphate groups forming the ridge between the two grooves are contacted by residues from both sides.

(C) (Top) Schematic representation showing base and phosphate contacts between Rep197 and RBS26. The interactions of each Rep monomer are shown in a different color. Straight lines indicate main chain hydrogen bonds, and arrows indicate side chain-mediated interactions. The

major groove side (Figures 3B and 3C). At the tip of the loop, two main chain carbonyl oxygens from Lys-138 and Gly-139 interact directly with diagonal C-A bases (C9 and A17′ for monomer "C," Figure 3B). A conserved feature of the loop is three consecutive amino acids with small side chains: in AAV5 Rep, Gly-139 and Gly-140 are followed by Ala-141, while in serotypes AAV2-6, there are three sequential glycines. Glycines and the lack of residues with bulky side chains are likely important in allowing a backbone conformation in which main chain carbonyls can reach into the major groove. The other two bases recognized within the major groove are the diagonal G bases (G8 and G18′ for monomer C), which are contacted by the side chain amino group of Lys-137. The most notable change when uncomplexed and complexed Rep197 are aligned (overall rmsd of 0.57 Å in Cα positions) is a shift of the β4/β5 loop ~2 Å into the major groove.

In the minor groove, all of the direct DNA contacts involve residues on one face of helix αC (Figures 3B and 3C) whose N-terminal half lies within the groove, consistent with the helix dipole. This secondary structural element is also used by the structurally homologous papillomavirus E1 protein to bind DNA, although the mode of DNA recognition is very different, as E1 contacts only the major groove and forms hydrogen bonds only with the phosphate backbone (Enemark et al., 2002). In Rep197, two residues, Arg-106 and Met-102, reach deeply into the groove: Arg-106 specifically recognizes two sequential pyrimidines (T14 and C15 for monomer C) while Met-102 forms exclusively van der Waals contacts with surrounding bases and the backbone. Other residues that interact exclusively with the phosphate backbone are shown in Figure 3B.

One consequence of helix αC insertion is the distortion of the minor groove, which shows periodic widening and narrowing. The widest point (7.6 Å versus 5.9 Å for B form DNA, calculated using Curves; Lavery and Sklenar, 1988) corresponds to the last bp of one GCTC repeat (i.e., the C recognized by Arg-106) and the first bp of the next repeat. Narrowing occurs (4.5 Å) where the side chain of Met-102 resides and is a result of the β4/β5 loop in the adjacent major groove pushing against a backbone phosphate. The periodic changes in groove dimensions are accompanied by base pair distortions (e.g., in slide, shift, and roll angles) that occur throughout RBS26 with a four bp periodicity. It would be interesting to determine if DNA distortion induced by the binding of one monomer facilitates the binding of subsequent monomers.

Each Rep monomer interacts with DNA by providing a loop to one repeat and an α helix to the adjacent repeat in a manner reminiscent of tweezers grasping DNA. Each four bp repeat is therefore recognized from both the minor and major grooves by two Rep molecules interacting with specific bases.

Repeat Recognition

The chromosomal AAVS1 and RBS sequences of various AAV serotypes are related but not identical (Figure 3C), and the Rep197/RBS26 structure explains how Rep accommodates variations at certain positions within the tetranucleotide repeat. Two dinucleotide sequences within each GCTC repeat provide the foundation of sequence-specific DNA recognition. The first of these is recognized by Rep from the major groove where main chain carbonyls at the tip of the β4/β5 loop bind a diagonal C-A, i.e., the second base (C) on the top strand and the third base (A) on the bottom strand. The second important dinucleotide consists of the third and fourth bases on the top strand (TC), which are recognized by Arg-106 in the minor groove. As each monomer recognizes two adjacent repeats, we propose that for binding to occur, at least one of these repeats must contain one of these two dinucleotide sequences, a pattern observed in all RBS sequences and at AAVS1.

Two variations on a perfect repeat are contained within RBS26. In repeat "A," the C within the diagonal C-A dinucleotide is replaced with T. Although this results in the loss of a hydrogen bond, the overall mode of Rep binding is the same as for perfect repeats. In repeat "D," the C of the -TC- dinucleotide is replaced by G. The structural consequence of this substitution is a change in the side chain orientation of Arg-106: to conserve the total number of protein/DNA bonds, Arg-106 shifts by one base pair to now bridge the fourth base of repeat "D" and the first of repeat "E." Thus, the ability of Rep monomers to bind imperfect repeats reflects not only the fact that two adjacent monomers participate in repeat recognition but also the redundancy in the binding interactions (since interactions can be lost, as in repeat A) and their flexibility (seen in the movement of Arg-106 in repeat D).

Interestingly, the distribution of perfect repeats within RBS sequences is not random: repeat "C" is always perfect and is flanked on at least one side by another perfect repeat. This pattern may be a tactic to dictate tight binding to the center of the RBS and weaker binding to the repeats on the flanks, which are generally imperfect. It is worth noting that AAVS1 contains three perfect repeats followed by the GCTG variation seen in repeat D of the AAV5 RBS; thus, our structure provides the structural information necessary to model Rep binding to the chromosomal integration site. To date, only AAV2 has been shown to integrate, but the high sequence similarity between Rep proteins from different serotypes indicates that the mode of DNA recognition will likely be the same.

Conservation of Interacting Residues of Rep197

Of the eight Rep197 residues that form direct side chain interactions with the RBS, seven are strictly conserved among serotypes AAV2-6. Lys-137 of AAV5 is substi-

six bases in gray boxes in repeat C correspond to those bases within a tetranucleotide repeat that are recognized by Rep197 (also highlighted with thick lines/arrows). The light gray boxes mark those bases that deviate from the perfect repeat sequence. The top strand has unprimed numbering (1 to 26 in 5′ to 3′ direction), and its complementary strand has primed numbering (26′ to 1′ in 3′ to 5′ direction). (Bottom) AAV2, 3, 4, and 6 RBS sequences are aligned as in Chiorini et al. (1999) while the AAVS1 perfect repeats and the five repeats of AAV5 bound by Rep197 are centered on the tandemly repeating sequence. Perfect repeats are shown in bold, bases that deviate from the perfect repeat sequence are underlined, and flanking regions are bracketed.

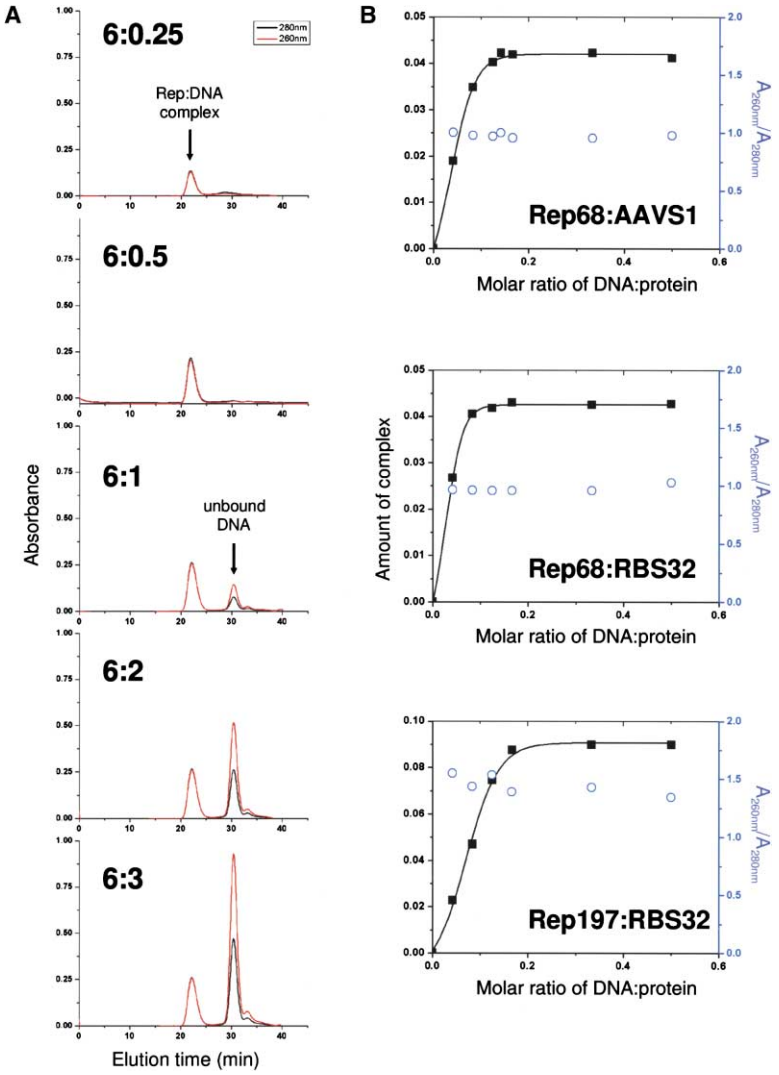

Figure 4. Binding of AAV2 Rep68 and AAV2 Rep197 to AAVS1 and RBS Sequences

(A) Gel filtration elution profiles demonstrate the formation of a single, stable Rep68/AAVS1 complex at protein:DNA ratios ranging from 6:0.25 to 6:3. Using a fixed amount of protein, increasing amounts of DNA were added resulting in increasing amounts of a single complex and no free DNA, until saturation was reached at ~6:1 Rep197:DNA and ~7.5:1 Rep68:DNA. Thereafter, the addition of more DNA did not result in the formation of higher-order complexes. Excess unbound protein precipitates under these experimental conditions and, therefore, is not seen in the chromatograms.

(B) Data obtained from gel filtrations as described in (A) was plotted as the amount of complex formed (filled squares represent integrated peak area in arbitrary units) as a function of the DNA:protein ratio. The A260:A280 nm ratio for the complex formed at each protein-to-DNA ratio was also calculated (blue open circles) based on integrated peak areas. Top and middle plots correspond to AAV2 Rep68 binding to 32-mers containing the AAVS1 Rep binding site and the RBS, respectively; the bottom plot represents binding of AAV2 Rep197 to the RBS 32-mer. Binding of AAV2 Rep197 to the AAVS1 32-mer is essentially identical to the bottom plot (data not shown).

tuted by Arg in the other serotypes. The contribution toward RBS binding has been assessed for a few of these residues (Urabe et al., 1999). Mutation of three charged residues in the N terminus of AAV2 Rep resulted in complete loss of RBS binding: Arg-107 (Arg-106 in AAV5), Lys-136 (Lys-135), and Arg-138 (Lys-137). This is consistent with the Rep197/RBS26 structure which reveals that both Arg-106 and Lys-137 provide important base contacts, and Lys-135 forms a salt bridge to the phosphate backbone.

Assembly of a Multimeric Complex on Viral RBS Sequences

In the Rep197/RBS26 complex, five Rep197 monomers are observed bound to the five direct repeats of RBS26. Although several lines of biochemical evidence indicate that two consecutive tetranucleotide repeats are sufficient for Rep binding as long as flanking DNA is present (Chiorini et al., 1995; Wonderling and Owens, 1997), the presence of at least four repeats within the RBS sequences of serotypes AAV2-6 (Figure 3C) suggests that two repeats are not enough to assemble a functional complex. Up to six different bound species can be detected by gel shift (McCarty et al., 1994b) and chemical crosslinking (Smith et al., 1997). Furthermore, dinucleotide transversion mutations anywhere within the five repeats in the AAV2 RBS result in an increase in K_d for Rep68 binding; while the most marked effects are centered over the middle three perfect repeats, the limits correspond to the 20 bp of the five repeats (Ryan et al., 1996). Our structural result that each Rep monomer binds to two sequential repeats and each repeat is recognized by elements from two monomers suggests that, at the edges of the RBS, one recognizable repeat next to a random tetranucleotide sequence might be bound by two Rep monomers.

To address the issue of stoichiometry of Rep binding at RBS sequences and to confirm that our structure of AAV5 Rep197/RBS26 is relevant to the more thoroughly biochemically characterized AAV2 Rep, we performed a series of binding studies using purified AAV2 Rep68 (Chiorini et al., 1994b) and AAV2 Rep197 on 32-mers containing AAVS1 and AAV2 RBS sequences. A single, stable complex was detected when a fixed amount of protein and increasing amounts of dsDNA were mixed and subsequently analyzed by gel filtration (Figure 4).

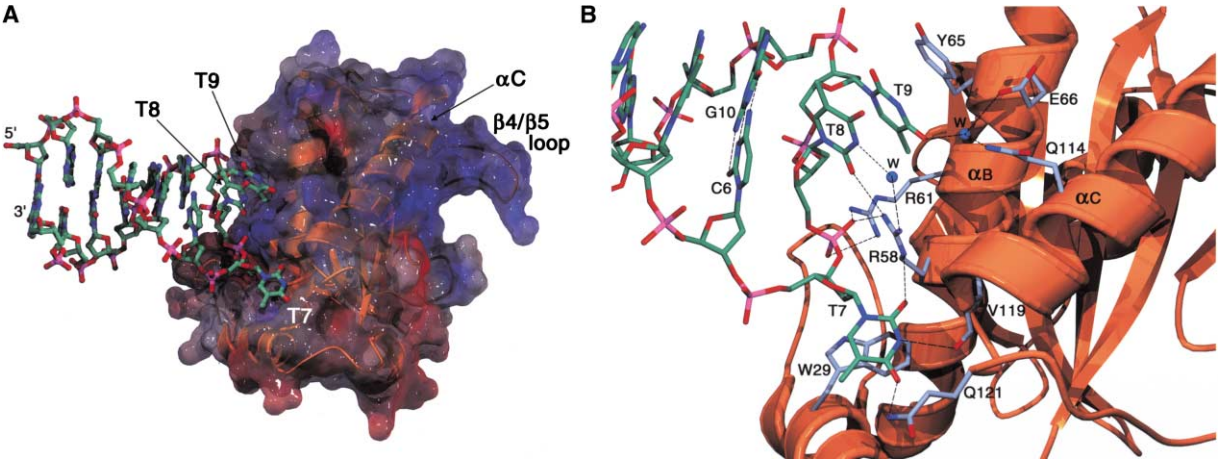

Figure 5. The N-Terminal Domain of Rep Specifically Recognizes the Tip of One of the Hairpin Arms

(A) Surface representation of Rep197/stem2. For reference, the two structural elements responsible for RBS binding, αC and the β4-β5 loop, are also indicated.

(B) Bases C6-G10 showing hydrogen bonding interactions. Water molecules are shown as blue spheres.

As the amount of DNA was increased, the amount of complex increased until a saturating level was reached at a protein:DNA molar ratio of ~1:0.17 (or or 6:1) for Rep197 and ~1:0.13 (7.5:1) for Rep68; this difference is likely within the error of this nonequilibrium method, but may also reflect slight degradation of Rep68 (data not shown). The A260:A280 nm of each complex was constant at all protein-to-DNA ratios examined, suggesting that under these experimental conditions a single species was present throughout, a result also suggested by the constant elution position of the complexes. These results are therefore consistent with the notion that both the isolated N-terminal domain and full-length Rep assemble as multimeric complexes, most likely hexamers, on viral and AAVS1 sequences. The observation of ~6:1 complexes with the N-terminal domain alone demonstrates that hexamerization is not driven by the C-terminal domain, which, as a member of the SF3 family of helicases, likely forms hexameric rings on DNA (James et al., 2003). The crystallographic observation of a 5:1 complex on RBS26 may reflect the use of an oligonucleotide too short to bind its full complement of Rep197 monomers, and it seems likely that a 6:1 complex will simply have another monomer spirally bound at one end of the assembly.

The Structure of Rep197 Bound to stem2

Rep197 and the DNA stem loop corresponding to the RBE′-containing hairpin arm of the AAV5 ITR (stem2) form a 1:1 complex in which stem2 is bound to an extensive positively charged surface on the opposite side of Rep197 from the active site (Figure 5A). Only three consecutive bases, T7–T9, interact directly with the protein. The base-paired stem of the hairpin extends away from Rep197, orthogonal to the protein surface. Apart from some side chain rotamer changes, stem2 binding does not induce significant conformational changes in Rep, and superposition of uncomplexed and complexed Rep results in an rmsd of 0.58 Å in Cα positions.

Description of the Interface

As shown in Figure 5B, stem2 contains a three-membered loop in which T7–T9 are unpaired and the flanking bases C6 and G10 form a Watson-Crick bp, which stacks on the remaining base pairs of the stem. T7 is flipped out, while T8 and T9 are in conformations that allow continued helical stacking of the stem into the loop in the 5′-3′ direction. This differs substantially from the NMR structure of the free AAV2 stem loop (Chou et al., 2000) in which the first T of the loop is folded into the minor groove, and the second and third T bases point in parallel toward the major groove. Since the RBE′ sequences are identical between AAV2 and AAV5 and the amino acid residues at the Rep197/stem2 interface are highly conserved among serotypes, it is likely that Rep binds a conformation that is not the predominant conformation in solution.

The flipped-out base T7 is inserted into a preformed surface pocket of Rep197 assembled by the convergence of several residues, reminiscent of the mode of substrate binding by methyltransferases and uracil-DNA glycosylases (Roberts and Cheng, 1998). One wall of the pocket is formed by Trp-29 onto which T7 is stacked at a distance of 3.5 Å. All of the potential hydrogen bond donors and acceptors in T7 are used, as T7 forms hydrogen bonds with Nε of Arg-58, the carbonyl oxygen of Val-119, and the amide side chain of Gln-121. The other two T bases in the loop have far fewer protein interactions. T8 is stacked between bases C6 and T9 and forms only a hydrogen bond with Arg-58. T9, sitting in a shallower pocket than T7, is stacked on Tyr-65 at a distance of 3.4 Å and there is a single water-mediated hydrogen bond between O4 and the side chains of Glu-66 and Gln-114. The DNA backbone conformation is sharply twisted between C6 and T7 as indicated by large changes in α and γ backbone torsion angles. Although the C6:G10 bp retains Watson-Crick hydrogen bonding, it deviates substantially from coplanarity with an inclination of 27° and buckle of 18°.

Four residues of Rep197 (Trp-29, Tyr-65, Arg-58, and

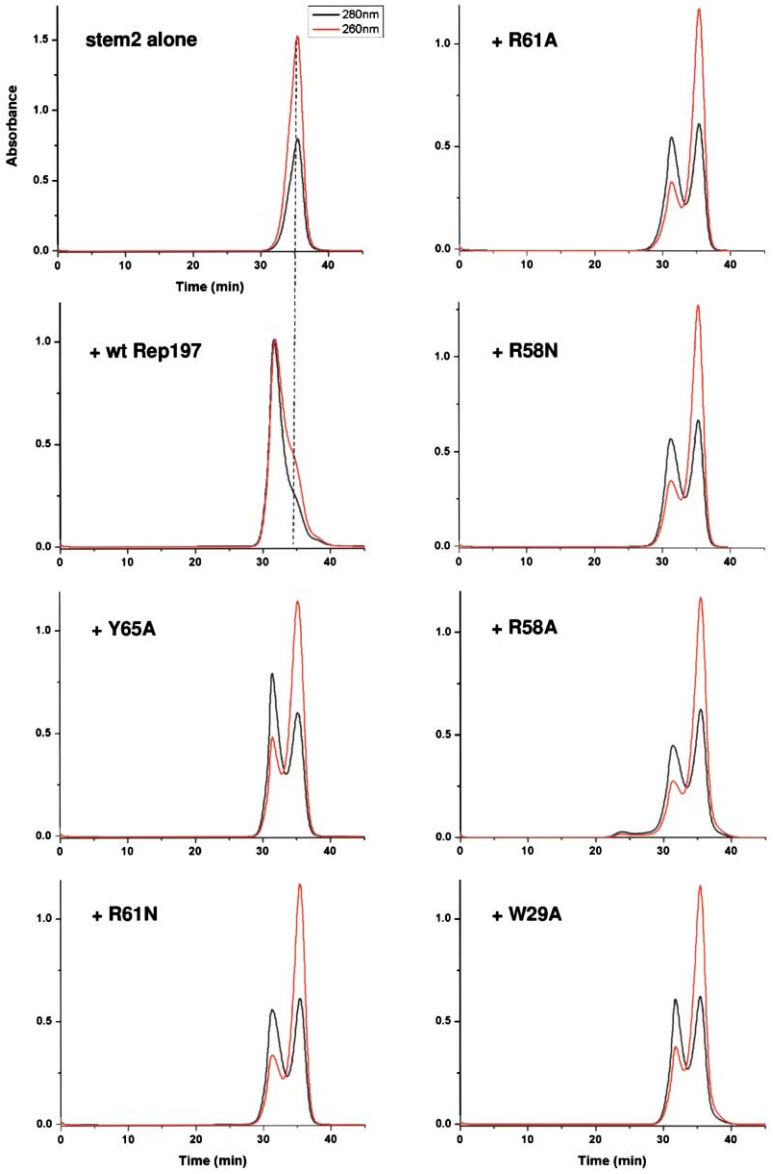

Figure 6. Mutations in the Rep197/Stem2 Interface Eliminate Stem-Loop Binding

Gel elution profiles for 0.5:1 molar ratio mixtures of stem2 and Rep197 (wild-type and mutants).

Gln-121) form direct side chain contacts with the three T bases in the loop, and the side chains of Arg-58 and Arg-61 also participate in backbone phosphate binding. Among serotypes AAV2-6, Trp-29, and Arg-61 are strictly conserved; Arg-58 and Gln-121 are conservatively substituted by Lys and Arg, respectively, in all other serotypes, while Tyr-65 is the most variable residue, appearing as Thr in AAV2 and Val in AAV3, 4, and 6. This lack of conservation is consistent with residue 65 providing van der Waals contacts and generic bulk to pack against the third base of the loop. To confirm the importance of residues Trp-29, Arg-58, Arg-61, and Tyr-65 in contributing to stem2 binding, we mutated each individually to Ala; Arg-58 and Arg-61 were also mutated to Gln. As shown in Figure 6, none of the point mutants was able to bind stem2.

Significance of the RBE′ Sequence

Methylation interference experiments and DNaseI protection assays indicate that AAV2 Rep preferentially contacts one of the two hairpin arms of the ITR and that this arm corresponds to the one furthest away from the trs, whether the ITR is in the flip or the flop orientation (Ashktorab and Srivastava, 1989; Im and Muzyczka, 1989). Ryan et al. (1996) showed that five bases, 5′-CTTTG-3′, at the tip of one of the hairpin arms particularly contribute to Rep binding. The low efficiency of trs cleavage on linear duplex substrates compared to that on substrates with intact ITRs is unaffected by the addition of the hairpin arms in *trans* (Chiorini et al., 1994b), suggesting that the Rep/RBE′ interaction stimulates cleavage by imposing a geometric constraint on Rep assembled on the RBS. It is possible that this stimulation is indirect and that the Rep/RBE′ interaction is important primarily to generate a single-stranded substrate rather than to stimulate the chemical steps of DNA cleavage (Brister and Muzyczka, 2000). This is consistent with the structure, which shows that stem-loop binding does not induce any conformational changes at the enzyme active site.

The fine tuning of the cleavage activity by the RBE′ plays an important role in Rep's ability to discriminate between fully replicated dsDNA viral genomes—which still possess RBS and trs sequences—and genomes whose terminal hairpin arms identify them as unreplicated. Nevertheless, the Rep/RBE′ interaction must be modulatory rather than essential; otherwise, Rep would not be able to cleave the trs sequence present at the chromosomal AAVS1 site and the virus would be unable to integrate.

Model of Proposed Assembly of Rep on ITRs

For its size (~20 kDa), the AAV Rep endonuclease domain exhibits a remarkable capacity to interact with different DNA substrates. There is an active site for trs cleavage located on one surface, a hairpin binding site on the opposite side, and a region in between that is used for sequence-specific RBS binding. Satisfyingly, Rep molecules bind RBS DNA such that all of the active sites are oriented in the direction of the trs, and the surface that binds stem2 is necessarily oriented toward its hairpin arm substrate. This arrangement also provides an elegant explanation for the results of RBS polarity mutation studies, which demonstrated that various orientations of the RBS sequence do not significantly affect Rep binding, yet only the wild-type orientation is able to direct efficient trs cleavage (Brister and Muzyczka, 2000).

It has been established that DNA unwinding at the trs is catalyzed by the Rep helicase activity, that trs cleavage occurs on ssDNA, and that Rep must remain bound to the RBS to cleave the trs efficiently (Brister and Muzyczka, 2000). AAV2 Rep can hexamerize when bound to dsDNA containing the RBS, and each integral protein-DNA complex can be detected (Smith et al., 1997). The Rep197/RBS26 structure and the binding data in Figure 4 reveal that this reflects Rep N-terminal domains binding to RBS DNA and does not necessarily imply anything about the helicase multimeric state.

Our structure of the Rep197/RBS26 complex suggests that binding of several Rep N-terminal domains to the RBS is the initiating step in the assembly of an active complex on the viral ITR, as shown in Figure 7. We propose that the binding of N-terminal domains stimulates the subsequent assembly of the C-terminal helicase domains into a hexamer at the site where unwinding activity is required to generate ssDNA at the trs. The observation that Rep binds more tightly to the RBS when it is flanked by nonspecific DNA (Chiorini et al., 1994b; Weitzman et al., 1994) is consistent with the need for room to accommodate another DNA binding domain. DNaseI footprinting data indicating protection of the region between the trs and the RBS (Im and Muzyczka, 1989) clearly suggests that this is where the helicase assembles. Once the helicase is assembled, DNA can be unwound, possibly in steps that resemble the "iris" mechanism recently proposed on the basis of the SV40 large T antigen helicase structure (Li et al., 2003). Experimental support for this is provided by the observation that AAV2 Rep68 can bind and unwind fully duplex blunt-ended DNA provided it contains an RBS (Zhou et al., 1999). A particularly appealing aspect of the proposed parallel with the SV40 T antigen origin

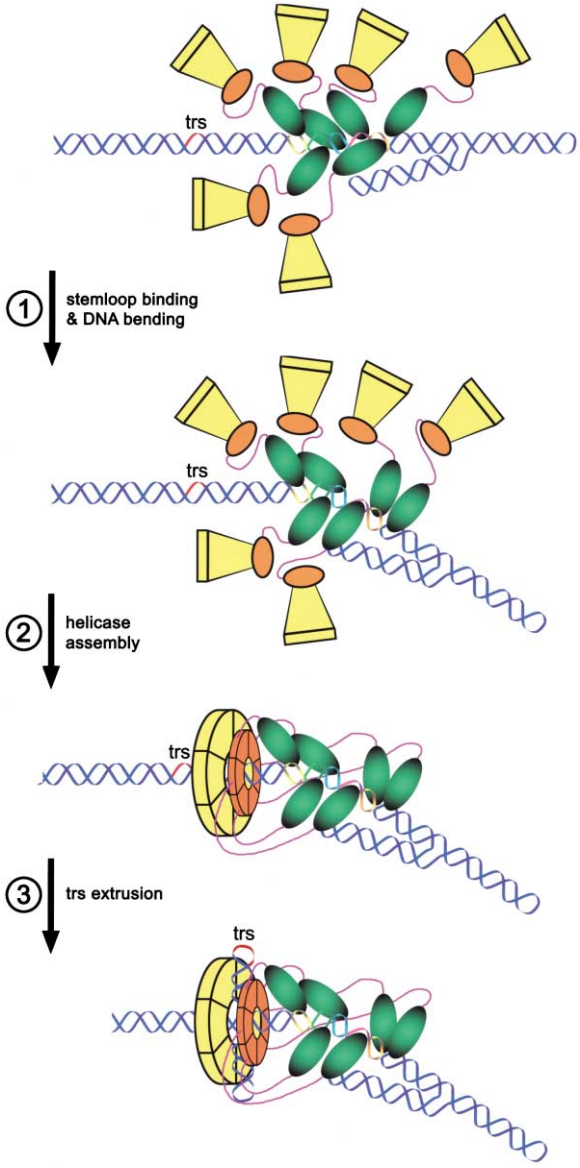

Figure 7. Model of Rep Assembly on the Viral Inverted Terminal Repeat

In step 1, six N-terminal domains of Rep (green) bind to tetranucleotide repeats within the RBS. At the viral ITR, this is accompanied by binding of one monomer to the tip of one of the hairpin arms. At AAVS1, there is no evidence for stem-loop structures; therefore, this is not a necessary step in complex formation. In step 2, RBS binding by Rep leads to the assembly of the helicase (orange and yellow, corresponding to the two domains) close to the trs site. In step 3, helicase activity generates a single-stranded substrate at the trs (red) that can be cleaved by Rep. For clarity, the two hairpin arms of the three-way DNA junction are not shown to scale.

melting mechanism is that extrusion of ssDNA through a side channel of the helicase would cause the trs sequence to pass close to the active sites of the Rep N-terminal domains during its transit from the interior.

One challenge in reconciling the Rep197/RBS26 structure with the properties of a helicase is that it is difficult to envisage a planar hexameric ring of C-terminal domains physically linked to six bound N-terminal

domains spiraling along the DNA axis, and we assume that DNA bending must occur. We also note that the expected conformation of the three-way junction in the ITR—coaxial stacking between two arms and an antiparallel orientation of the third arm (Altona et al., 1996; see also Figure 7)—would in principle allow one of the bound Rep197 monomers to simultaneously bind the tip of one of the terminal hairpins. This would provide a mechanism to explain the reported stimulation of trs cleavage by the RBE′ sequence, as this might introduce asymmetry, leading to enhanced DNA bending and unwinding that might facilitate helicase activity and subsequent cleavage.

The model in Figure 7 is specific for complex assembly at ITRs during viral replication. The role played by the stem loop is not needed in other contexts; for example, at AAVS1 where there are no stem-loop sequences and yet where a Rep complex also assembles to direct integration. It is possible that the N-terminal domains are not irreversibly bound to the RBS and that, upon helicase domain multimerization, N-terminal domains can transiently dissociate from the DNA. Thus, one monomer within the complex might be bound only to the stem loop rather than simultaneously to both the RBS and the hairpin sequence. More intriguingly, liberated N-terminal domains would be free to bind to other DNA molecules. Our identification of two interfaces on the surface of the N-terminal domain capable of binding DNA provides an obvious mechanism for bridging two pieces of DNA that contain RBS sequences or stem-loop sequences, a necessary step during the process of site-specific integration.

Experimental Procedures

Protein Purification and Crystallization

AAV5 Rep1-197 (Rep197) was purified as previously described (Hickman et al., 2002), dialyzed against 20 mM Tris (pH 7.5), 0.5 M NaCl, and concentrated to ~10 mg/ml. Oligonucleotides were from Integrated DNA Technologies, Inc. (Coralville, IA). Site-specific point mutations were introduced using the QuikChange kit (Stratagene), and the entire coding region was sequenced to confirm that no additional mutations had been introduced. The elution times on size-exclusion chromatography were essentially identical for each of the point mutants and corresponded to that of wild-type Rep197, suggesting that the overall structure of the protein was unperturbed. Histidine-tagged versions of AAV2 Rep68Δ and AAV2 Rep 1–197 were similarly purified.

For the RBS26 substrate, PAGE-purified oligonucleotides (5′-CGCGTTCGCTCGCTCGCTGGCTCGTG-3′ and 5′-CACGAGCCAGC GAGCGAGCGAACGCG-3′) were resuspended in TE (10 mM Tris [pH 8], 1 mM EDTA), mixed in a 1:1 molar ratio, heated to 90°C for 10 min and then cooled slowly to 20°C. Rep197 and RBS26 were mixed in a 3:1.1 molar ratio and dialyzed into 20 mM Tris (pH 7.5), 0.12 M NaCl. Initial crystals were grown at 4°C by the hanging drop method by mixing the protein/DNA complex in a 1:1 ratio with 20% (w/v) PEG 3000, 0.1 M sodium citrate (pH 5.5). Two rounds of microseeding into 14% (w/v) PEG 3350, 0.1 M sodium citrate yielded diffraction-quality crystals. Crystals were cryoprotected by transfer into 10 mM Tris (pH 7.5), 60 mM NaCl, 18% (w/v) PEG 3350, 50 mM sodium citrate (pH 5.5), 50 μM RBS26, 5% (v/v) ethylene glycol, and the ethylene glycol concentration then slowly increased to 20% (v/v). Crystals were flash cooled by immersion in liquid propane. The space group was $P2_1$ (a, 78.67 Å; b, 131.71 Å; c, 82.16 Å; β, 112.7°).

For the stem2 substrate, the 15-mer oligonucleotide (5′-CAGCTC TTTGAGCTG-3′) was resuspended in TE, heated to 90°C for 15 min, and then rapidly cooled on ice. Rep197 and stem2 were mixed in a 1:1.1 molar ratio and dialyzed into 20 mM Tris (pH 7.5), 0.12 M NaCl. Crystals were grown at 4°C by the hanging drop method by mixing the protein/DNA complex in a 1:1 ratio with 0.2 M magnesium acetate, 0.1 M cacodylate (pH 6.5), 20% w/v PEG 8000. Crystals grew to full size in approximately 3 weeks. Crystals were cryoprotected by dipping briefly in 10 mM Tris (pH 7.5), 60 mM NaCl, 0.1 M magnesium acetate, 50 mM cacodylate (pH 6.5), 25% w/v PEG 8000, 5% w/v glycerol, and then into liquid propane. The space group was $P2_12_12_1$ (a, 41.83 Å; b, 78.71 Å; c, 184.28 Å).

Data Collection and Structure Determination

X-ray diffraction data on the Rep197/RBS26 crystals were collected at the Southeast Regional Collaborative AccessTeam (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory, using a Mar165 CCD detector at 100 K. The usable data ($I/\sigma(I) > 2.0$) extended to 3.1 Å resolution. The data were integrated and scaled internally using the HKL suite (Otwinowski and Minor, 1997). The structure was solved by molecular replacement using a stand-alone version of AMoRe (Navaza, 2001) with diffraction data between 15 and 4 Å. The search model was the crystal structure of unbound Rep197 (Hickman et al., 2002; pdb code 1m55). The five peaks of the rotation function that corresponded to actual solutions had correlation coefficients on structure factor scale 23.4%, 19.6%, 19.6%, 18.5%, and 18.5%, respectively, while the highest peak corresponding to a nonsolution was at 17.9%. With all five molecules placed in the asymmetric unit, the structure had an R factor of 40.6% after rigid body refinement. Further rigid body refinements, molecular dynamics and energy minimizations, and B factor refinements were carried out with CNS (Brünger et al., 1998). After rigid body refinement of the five molecules at 3.1 Å resolution, the difference electron density clearly indicated the positions of most of the backbone phosphate groups of RBS26. A standard B form DNA model corresponding to the RBS26 sequence was built with Sybyl 6.9 (Tripos, Inc.) and fit manually to the difference density using O (Jones et al., 1991) in several segments to accommodate the needed deviations from the canonical structure. Noncrystallographic (NCS) restraints between the five molecules of Rep197 were used throughout the refinement and were gradually relaxed as the model improved. The final NCS restraint weight of 50 kcal/mole $Å^{-2}$ was chosen as it resulted in the lowest R_{free} (Brünger, 1992b) computed over 5% of reflections selected randomly. Simulated annealed omit maps indicated deviations in the individual monomers from each other in segments that contacted the DNA, and these were excluded from the NCS restraints. Final refinement statistics are shown in Table 1; there are no residues of the final model in the disallowed regions of the Ramachandran plot.

Diffraction data on the Rep197/stem2 complex to 2.0 Å were collected on a RUH3R rotating anode source equipped with multilayer focusing optics using CuKα radiation and an RaxisIV image plate detector at 95 K. The data were integrated and scaled internally with the HKL suite and the structure was solved with molecular replacement using AMoRe. The search model was again the 1m55 structure; in this case, there were two complexes in the asymmetric unit. The two rotation function peaks corresponding to the solutions had correlation coefficients 14.1% and 12.4% while the highest nonsolution was 10.1% on the structure factor scale using the 15 to 4 Å resolution range. Refinement was carried out with Xplor 3.1 (Brünger, 1992a) using molecular dynamics, energy minimization, and restrained B factor refinement, bulk solvent correction, and the TNT B factor restraint library (Tronrud, 1996) at 2.0 Å resolution. Monitoring R_{free} indicated that NCS restraints were not required. Difference electron density clearly showed both stem-loop DNA structures bound to the individual monomers and were built into the density using O. Further refinements and the addition of solvent molecules resulted in the model with statistics shown in Table 1. The geometry of the final refined structure is excellent with the exception of a short loop in one monomer between Phe-13 and Glu-17 that is disordered and remote from the stem-loop DNA.

The molecular figures were created with Spock (Christopher, 1998) and Molscript (Kraulis, 1991) and rendered with Povray (www.povray.org).

DNA Binding Assays

For stem-loop assays, HPLC-purified oligonucleotides representing the four possible stem-loop sequences at the AAV5 ITR (stem1: 5′-

ACGGCCAGAGGGCCGT-3′; stem2: 5′-CAGCTCTTTGAGCTG-3′; stem3: 5′-GAGTGCCACACTC-3′; trsHP: 5′-GAGTGTGGCACTC-3′) were resuspended in TE and rapidly annealed by heating to 90°C for 15 min followed by rapid cooling on ice. Binding was assessed by mixing the oligonucleotide with purified AAV5 Rep197 (69.5 μM; 1.60 mg/ml) at a 0.5:1.0 molar ratio of DNA: protein in 0.5 M NaCl-containing buffer. The solution was then dialyzed into 20 mM Tris (pH 7.5), 0.2 M NaCl and 50 μl applied to an analytical gel filtration Superdex 200 column (Pharmacia SmartSystem) equilibrated at 4°C in the same buffer. Binding was assessed by the ability to form a complex sufficiently stable to persist under gel filtration conditions.

For the AAVS1 (5′-TTGCGGCTCGGCGCTCGCTCGCTCGCTGG GCG-3′ and its reverse complement) and AAV2 RBS32 (5′-CCC TCTCTGCGCGCTCGCTCGCTCACTGAGGC-3′ and its reverse complement) assays, oligonucleotides were annealed and binding assessed as described above except that the buffer contained 0.12 M NaCl. Purified AAV2 Rep68Δ was held constant at 57.8 μM; AAV2 Rep197 at 66.9 μM.

Acknowledgments

We thank P. Musingarimi for the gift of AAV2 Rep197; S. Chacko for helpful discussions; W. Yang and E. Greene for insightful suggestions; N. Leyarovska for assistance with data collection at SER-CAT; and T. Chiu for help with Curves. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Office of Science under Contract No. W-31-109-Eng-38. This work was supported in part by the NIH Intramural AIDS Targeted Antiviral Program. Coordinates have been deposited with the Protein Data Bank.

Received: July 30, 2003
Revised: December 8, 2003
Accepted: December 9, 2003
Published: February 12, 2004

References

Altona, C., Pikkemaat, J.A., and Overmars, F.J.J. (1996). Three-way and four-way junctions in DNA: a conformational viewpoint. Curr. Opin. Struct. Biol. *6*, 305–316.

Ashktorab, H., and Srivastava, A. (1989). Identification of nuclear proteins that specifically interact with adeno-associated virus type 2 inverted terminal repeat hairpin DNA. J. Virol. *63*, 3034–3039.

Brister, J.R., and Muzyczka, N. (1999). Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J. Virol. *73*, 9325–9336.

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

Brünger, A.T. (1992a). X-PLOR Version 3.1. A system for X-ray crystallography and NMR. Version 3.1. (New Haven, CT: Yale University Press).

Brünger, A.T. (1992b). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature *355*, 472–475.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. *54*, 905–921.

Chiorini, J.A., Weitzman, M.D., Owens, R.A., Urcelay, E., Safer, B., and Kotin, R.M. (1994a). 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., Wiener, S.M., Owens, R.O., Kyöstiö, S.R.M., Kotin, R.M., and Safer, B. (1994b). 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., Yang, L., Safer, B., and Kotin, R.M. (1995). Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J. Virol. *69*, 7334–7338.

Chiorini, J.A., Kim, F., Yang, L., and Kotin, R.M. (1999). Cloning and characterization of adeno-associated virus type 5. J. Virol. *73*, 1309–1391.

Chou, S.-H., Tseng, Y.-Y., and Chu, B.-Y. (2000). Natural abundance heteronuclear NMR studies of the T_3 mini-loop hairpin in the terminal repeat of the adenoassociated virus 2. J. Biomol. NMR *17*, 1–16.

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

Costantini, L.C., Jacoby, D.R., Wang, S., Fraefel, C., Breakefield, X.O., and Isacson, O. (1999). Gene transfer to the nigrostriatal system by hybrid herpes simplex virus/adeno-associated virus amplicon vectors. Hum. Gene Ther. *10*, 2481–2494.

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., 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.

Flotte, T.R., and Carter, B.J. (1995). Adeno-associated virus vectors for gene therapy. Gene Ther. *2*, 357–362.

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.

Hacein-Bey-Abina, S., Le Deist, F., Carlier, F., Bouneaud, C., Hue, C., De Villartay, J.P., Thrasher, A.J., Wulffraat, N., Sorensen, R., Dupuis-Girod, S., et al. (2002). Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy. N. Engl. J. Med. *346*, 1185–1193.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., et al. (2003). *LMO2*-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415–419.

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., and Muzyczka, N. (1989). Factors that bind to adeno-associated virus terminal repeats. J. Virol. *63*, 3095–3104.

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.

James, J.A., Escalante, C.R., Yoon-Robarts, M., Edwards, T.A., Linden, R.M., and Aggarwal, A.K. (2003). Crystal structure of the SF3 helicase from adeno-associated virus type 2. Structure *11*, 1025–1035.

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.

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., Linden, R.M., and Berns, K.I. (1992). Characterization of a preferred site on chromosome 19Q for integration of adeno-associated virus DNA by nonhomologous 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.

Lavery, R., and Sklenar, H. (1988). The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. J. Biomol. Struct. Dyn. *6*, 63–91.

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.

Linden, R.M., Winocour, E., and Berns, K.I. (1996). The recombina-

tion signals for adeno-associated virus site-specific integration. Proc. Natl. Acad. Sci. USA *93*, 7966–7972.

McCarty, D.M., Pereira, D.J., Zolotukhin, I., Zhou, X., Ryan, J.H., and Muzyczka, N. (1994a). Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J. Virol. *68*, 4988–4997.

McCarty, D.M., Ryan, J.H., Zolotukhin, S., Zhou, X., and Muzyczka, N. (1994b). Interaction of the adeno-associated virus Rep protein with a sequence within the A palindrome of the viral terminal repeat. J. Virol. *68*, 4998–5006.

Nakai, H., Montini, E., Fuess, S., Storm, T.A., Grompe, M., and Kay, M.A. (2003). AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat. Genet. *34*, 297–302.

Navaza, J. (2001). Implementation of molecular replacement in AMoRe. Acta Crystallogr. D Biol. Crystallogr. *57*, 1367–1372.

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

Owens, R.A., Weitzman, M.D., Kyöstiö, S.R.M., and Carter, B.J. (1993). Identification of a DNA-binding domain in the amino terminus of adeno-associated virus Rep proteins. J. Virol. *67*, 997–1005.

Palombo, F., Monciotti, A., Recchia, A., Cortese, R., Ciliberto, G., and La Monica, N. (1998). Site-specific integration in mammalian cells mediated by a new hybrid baculovirus-adeno-associated virus vector. J. Virol. *72*, 5025–5034.

Recchia, A., Parks, R.J., Lamartina, S., Toniatti, C., Pieroni, L., Palombo, F., Ciliberto, G., Graham, F.L., Cortese, R., La Monica, N., et al. (1999). Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. Proc. Natl. Acad. Sci. USA *96*, 2615–2620.

Roberts, R.J., and Cheng, X. (1998). Base flipping. Annu. Rev. Biochem. *67*, 181–198.

Ryan, J.H., Zolotukhin, S., and Muzyczka, N. (1996). Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J. Virol. *70*, 1542–1553.

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.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521–529.

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 as a transposable element. 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.

Weitzman, M.D., Kyöstiö, S.R.M., Kotin, R.M., and Owens, R.A. (1994). Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc. Natl. Acad. Sci. USA *91*, 5808–5812.

Wonderling, R.S., and Owens, R.A. (1997). Binding sites for adeno-associated virus Rep proteins within the human genome. J. Virol. *71*, 2528–2534.

Wonderling, R.S., Kyöstiö, S.R.M., and Owens, R.A. (1995). A maltose-binding protein adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. J. Virol. *69*, 3542–3548.

Wu, J., Davis, M.D., and Owens, R.A. (1999). Factors affecting the terminal resolution site endonuclease, helicase, and ATPase activites of adeno-associated virus type 2 Rep proteins. J. Virol. *73*, 8235–8244.

Wu, J., Davis, M.D., and Owens, R.A. (2001). A Rep recognition sequence is necessary but not sufficient for nicking of DNA by adeno-associated virus type-2 Rep proteins. Arch. Biochem. Biophys. *389*, 271–277.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749–1751.

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.

Zhou, X., Zolotukhin, I., Im, D.-S., and Muzyczka, N. (1999). Biochemical characterization of adeno-associated virus Rep68 helicase and ATPase activities. J. Virol. *73*, 1580–1590.

Accession Numbers

Coordinates have been deposited with the Protein Data Bank with accession codes 1uut for the Rep/stem2 complex and 1rz9 for the Rep/RBS26 complex.