Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 7966–7972, July 1996
Microbiology

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 25, 1995.

The recombination signals for adeno-associated virus site-specific integration

(parvovirus/nonhomologous recombination/integration target sequence/Rep protein/targeted integration)

R. MICHAEL LINDEN, ERNEST WINOCOUR*, AND KENNETH I. BERNS†

Department of Microbiology, Hearst Microbiology Research Center, Cornell University Medical College, 1300 York Avenue, New York, NY 10021; and *Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

Contributed by Kenneth I. Berns, April 25, 1996

ABSTRACT **The adeno-associated virus (AAV) genome integrates site specifically into a defined region of human chromosome 19 (termed *AAVS1*). Using a functional assay for AAV integration into *AAVS1* DNA propagated as an episome, we obtained evidence that a 33-nucleotide *AAVS1* DNA sequence contains the minimum signal required for targeted integration. The recombination signal comprises a DNA-binding motif for the AAV regulatory Rep protein [Rep binding site (RBS)] separated by an eight-nucleotide spacer from a sequence that can act as a substrate for Rep endonucleolytic activity [terminal resolution site (TRS)]. Mutations in either the *AAVS1*-encoded RBS or TRS elements abort targeted integration. Since both the RBS and TRS elements are present in the viral origin of replication and are required for AAV replication, targeted integration into chromosome 19 *AAVS1* DNA may involve a replicative type of recombination that is discussed. An additional chromosome 19 element, which is responsible for DNA rearrangements in episomes propagating *AAVS1* DNA, was identified and shown not to be required for AAV episomal integration, despite its location adjacent to the recombination signal.**

The life cycle of the human parvovirus, adeno-associated virus (AAV), is characterized by persistent infection (1). This is manifest in cell culture by the inability of AAV to undergo productive infection in the absence of a concomitant infection by a helper virus, such as adenovirus (2, 3) or herpesvirus (4, 5), or exposure of the host cell to genotoxic conditions able to induce a stress response (6–8). Although all nuclear DNA viruses commonly cause persistent infections in the intact host, the peculiar properties of AAV have made it a particularly useful model system for study of the molecular mechanisms underlying the establishment of a latent infection. Infection of continuous lines of human cells in culture by AAV at a multiplicity of infection of 20–200 in the absence of a helper virus leads to a high percentage of cells containing the viral genome in a latent chromosomally integrated state (9–11). Activation of the latent AAV genome is readily accomplished by superinfection of the latently infected cell with a helper virus (12).

The AAV genome is a linear, single-stranded DNA of 4.7 kb, which contains two major open reading frames; regulatory proteins are encoded in the left half of the genome, and structural proteins are encoded in the right half (13). Four regulatory proteins with overlapping amino acid sequences are translated from both spliced and unspliced forms of the two mRNAs transcribed from the left side open reading frame (14). The two larger regulatory proteins, Rep 68/78, have been characterized as controlling all phases of the life cycle of the virus by regulating transcription, DNA replication, and integration and excision from the host genome (15).

Our current model of the AAV life cycle has the following elements. (*i*) When AAV infects a healthy, dividing cell, there is a very limited expression of Rep 68/78, which has three biological consequences. Further AAV gene expression is repressed; to a very great extent, AAV DNA replication is inhibited; and integration of AAV DNA into the host cell genome occurs. All three of these are directly dependent on Rep 68/78. (*ii*) Infection of a latently infected cell by a helper virus leads to activation of the integrated AAV genome, rescue from the integrated state, DNA replication, and production of progeny virions. Rep 68/78 is required during this process for gene expression, DNA replication, and rescue from the integrated state. Thus, the phenotype of Rep 68/78 expression is directly dependent on the physiological state of the host cell. When the cell is healthy, the AAV latent state is maintained; when the cell is stressed, the viral genome is activated to allow infection of a new host.

Our goal has been to understand the molecular mechanisms underlying the phenomenon of AAV latency. Although any exogenous DNA introduced into a mammalian cell can recombine in a nonhomologous manner with the host DNA, the AAV genome is the only example of an exogenous DNA that integrates at a specific site (19q13.3-qter) in the human genome (16–19). Our current understanding of this process implicates cis-active signals, both in the viral genome and at the site of integration, as well as Rep 68/78 (AAV vector virions from which the rep gene has been deleted do not integrate in a site-specific manner). Originally, the integration site from chromosome 19 was cloned as an 8.2-kb fragment (20). When the first 4 kb were sequenced, the junctions with integrated viral DNA were identified. Although many features of interest were identified in the sequence, including an expressed open reading frame, short repeats, a CpG island, etc., none appeared to be sufficiently unique to account for the observed specificity of integration. To confirm the notion that it was the DNA sequence of the integration site that was important and to be able to map the critical sequences in finer detail, a model system was developed. The integration site (*AAVS1*) was cloned into an EBV-based shuttle vector (21), which was then used to establish cell lines. Infection of such cell lines demonstrated integration of the AAV selectively into shuttle vector DNA containing the integration site and permitted narrowing down of the critical site to a DNA fragment of 500

Abbreviations: AAV, adeno-associated virus; EBV, Epstein–Barr virus; ITR, inverted terminal repeat; TRS, terminal resolution site; RBS, Rep binding site.

†To whom reprint requests should be addressed.

bases (22). Characterization of the recombinants that had been formed showed several features that had also been identified in analyses of chromosomal integration. These included junctions between viral and cellular DNA at either the viral inverted terminal repeat (ITR) or near the leftward most promoter (map position 5) and within the *AAVS1* sequence; rearrangements of flanking sequences; and evidence for a head-to-tail arrangement of the integrated viral DNA. Thus, the model appeared to closely mimic the natural process (23).

The AAV genome contains an ITR that is palindromic and which serves as the primer-origin for DNA replication (24, 25). The ITR contains two signal sequences that are critical: a Rep binding site (RBS) (26–28) and the terminal resolution site (TRS) where bound Rep 68/78 makes a site-specific nick (29). These two sequences are also present in the 500-bp fragment containing the integration site. Rep 68/78 has been demonstrated to bind to the cellular fragment and to be able to serve as a bridge to join the AAV ITR to *AAVS1* (30). Also a plasmid containing part of the 500 bp sequence can serve as a partial substrate in an *in vitro* assay for AAV-specific DNA replication, indicating that Rep cannot only bind to the sequence but can also cut at the TRS sequence present (31). These data support the notion that the RBS and the TRS present in *AAVS1* play an integral role in site-specific integration. Another signal that is in close proximity to the RBS and TRS is an octamer that has been identified as an enhancer of meiotic recombination in fission yeast (M26) (32, 33). This sequence is also present in the AAV genome near to one ITR.

In this paper we describe a genetic analysis of the region of *AAVS1* indicated to be critical for site-specific integration to directly test whether putative signal sequences are essential and to determine a minimal size of *AAVS1* required for integration. The data show that indeed RBS and TRS are required for site-specific integration but that the M26 recombination enhancer is not. The minimal size of *AAVS1* required has been reduced to 33 bases. Finally, although the M26 sequence was not found to be required for integration, its presence does correlate with genetic instability and rearrangement of the shuttle vector.

MATERIALS AND METHODS

Subcloning of *AAVS1* Target Sequences in the EBV Vector p220.2 (see Fig. 1). pRI-A, which contains the 8.2-kb *AAVS1* DNA (20), was digested with *Eco*RI and *Pvu*II, and the 510-bp sequence from the 5′ end of *AAVS1* was isolated by agarose gel electrophoresis and cloned in pUC18 (generating pUC510). p330 and p110 were generated by *Sma*I digestion of pUC510: two fragments, 336 (*AAVS1* nt 18–353) and 104 bp (*AAVS1* nt 354–461), were isolated and subcloned into the *Xba*I site of p220.2 (an EBV-based shuttle vector kindly provided by B. Sugden) (21). pAatII was constructed by *Eco*RI-*Aat*II digestion of pUC510 and cloning of the resulting 284-bp fragment (*AAVS1* nt 1–284) into the *Xba*I site of p220.2. pHae1 and pHae2 were generated by *Hae*III digestion of the 510-bp *Eco*RI-*Pvu*II fragment of pUC510; two fragments (*AAVS1* nt 347–446; *AAVS1* nt 209–326) were recovered from a 3% NuSieve (FMC BioProducts) agarose gel and recloned into the *Xba*I site of p220.2. Subcloning of the entire 5′ 510-bp *AAVS1* segment into p220.2 (p510) was described previously [p220/*AAVS1*(kb 0–0.51)] (22).

Synthetic Oligonucleotides as Targets for Integration. Synthetic oligonucleotides containing putative target sequences and their respective complimentary strands were purchased from Oligos Etc. (Wilsonville, OR) and cloned into p220.2 previously linearized with *Xba*I and *Sal*I. The sequence and genotypes are described in Fig. 2.

Site-Directed Mutagenesis. pMUT17, which contains a mutant Rep binding site, was generated as follows. The 510-bp fragment isolated from pUC510 by *Eco*RI-*Hin*dIII digestion was cloned into M13mp19 RF. Single-stranded DNA prepared from a single plaque by standard procedures was used as a template for site-directed mutagenesis (Sculptor kit, Amersham, following the protocol supplied by the manufacturer) with the mutant primer RML 17 (Fig. 1C). The mutation was confirmed by sequencing.

Propagation of p220.2-Based Shuttle Vectors in C17 Cells. Transfection of shuttle vectors and propagation in C17 cell line (34–37) was performed as described (22).

Virus Infection. Cells were infected with purified AAV (7) at a multiplicity of infection of 20 as described in ref. 22.

Isolation of Episomal DNA, Recovery of Vector in *Escherichia coli* and Identification of AAV Recombinant Colonies. Extrachromosomal DNA was isolated by the method of Hirt (38), modified as described (22).

RESULTS

Localization of the Recombination Initiating DNA. Using the EBV shuttle vector model system described previously, it was shown that a 510-bp sequence at the 5′ end of *AAVS1* contains all of the genetic information necessary to mediate AAV integration into the episomal vector carrying this DNA (22, 23). To determine the recombinogenic segment of the 510-bp DNA, subfragments (graphically described in Fig. 1) were inserted into p220.2 and propagated as episomes in C17 cells under selective pressure. After five or more passages, the cells were infected with wild-type AAV; 48 h later, the vectors were rescued in *E. coli*, and Amp^R colonies were screened for the presence of AAV DNA by colony hybridization to virion DNA. As previously, the relative number of AAV-positive

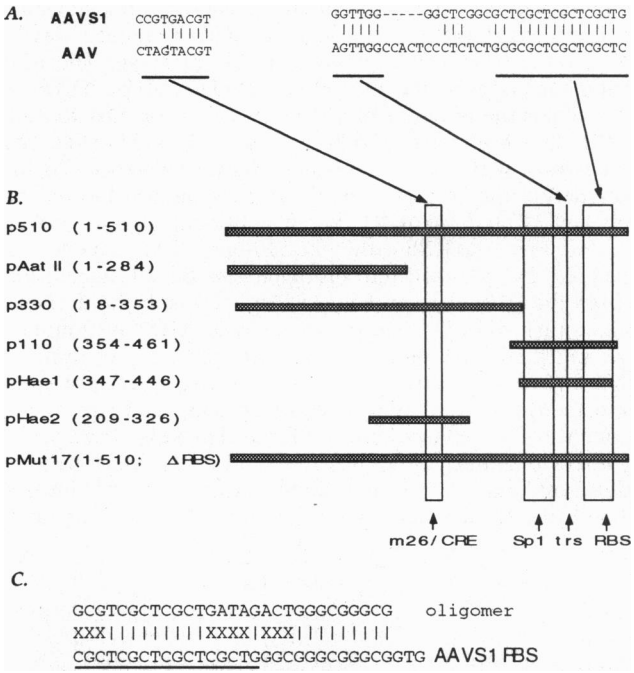

FIG. 1. Putative recombinogenic portions of the 510-bp 5′ end of *AAVS1*. (*A*) Alignment of putative recombinogenic elements present in the AAV terminal repeats and in *AAVS1*. (*B*) Portions of *AAVS1* cloned into p220.2 for assessment of their ability to serve as targets for AAV integration. The *AAVS1* nucleotide positions are indicated in parentheses. Vertical lines indicate the positions of the putative recombination signals M26/cAMP response element (nt 275–283), Sp1 recognition site (nt 354–370), TRS (nt 384–389), and RBS (nt 398–413). (*C*) The oligonucleotide used for generating mutations within the RBS in pMUT17 is shown in alignment with the corresponding *AAVS1* sequence. Seven nucleotides within the RBS were exchanged. The recombinogenic activity of the above *AAVS1* segments is given in Table 1.

Table 1. Genotypes and recombination frequencies of AAVS1 DNA subfragments in p220.2 vectors.

DNA	Nucleotide position in *AAVS1*	Genotype			Recombination (AAV-positive colonies, no./total)
		M26/CRE	TRS	RBS	
p510	1–510	+	+	+	24/7908
pMUT17 (ΔRBS)	1–510	+	+	−	0/7020
p330	18–353	+	−	−	0/13468
pAatII	1–279	+	−	−	0/2376
p110	354–461	−	+	+	5/3815
pHae1	347–446	−	+	+	4/3895
pHae2	209–326	+	−	−	0/3180

See Fig. 1 for description of the AAVS1 DNA subfragments listed in the first column.

colonies was taken to represent the recombination frequency between AAV virion DNA and the p220.2 vectors carrying the defined segment of *AAVS1* DNA. The results obtained are presented in Table 1.

The 510-bp *AAVS1* DNA insert recombined with AAV DNA at a frequency of ~0.3% (24/7908 colonies). This result is in accordance with previously published data (22) showing a recombination frequency of 0.4–0.8% for the same *AAVS1* DNA. The inserts pAatII (*AAVS1* nt 1–279), pHae2 (*AAVS1* nt 209–326), and p330 (*AAVS1* nt 18–353) were recombination inactive; 0/2376 colonies, 0/3180 colonies, and 0/13,468 colonies, respectively. In contrast, the p110 insert (*AAVS1* nt 354–461) was active in recombination; five out of 3815 colonies were shown to contain AAV sequences resulting in an estimated frequency of 0.13%. The localization of the recombinogenic DNA was further narrowed by the activity of the pHae1 insert (*AAVS1* nt 347–446), which recombined at a frequency of 0.1% (4/3895 colonies). These results indicated, therefore, that the recombinogenic signal sequence was located on *AAVS1* DNA between nt 347 and 446. The most striking feature of the nt 100-bp cellular *AAVS1* DNA, nt 347–446, is the presence of a RBS in close proximity to a site (TRS) shown previously to be nicked by Rep (31) (Fig. 1*B*).

Analysis of the Recombinogenic Signal Sequence. To further define the recombinogenic DNA sequence located between nt 347 and 446 of *AAVS1*, an additional set of targets was created. First, the RBS within the 510-bp *AAVS1* sequence was changed by site-directed mutagenesis. Seven nucleotides within the RBS were exchanged (Fig. 1*C*), and the mutation was confirmed by DNA sequence analysis. After subcloning of the mutated 510-bp fragment into p220.2 (resulting in pMUT17, Fig. 1), integration into this target vector was not detected (0/7020 colonies) in contrast to the wild-type 510-bp sequence (24/7908 colonies) (Table 1). Next, the specific synthetic oligonucleotides described in Fig. 2 were cloned into p220.2 and assessed for their abilities to target AAV integration (Table 2). The recognition site for the transcription factor Sp1 was included since the sequence is an imperfect inverted repeat of the RBS and binds Rep 68 *in vitro* (unpublished data). The 62-nt STR oligomer contains the RBS, the TRS, and the putative DNA recognition site for Sp1 (Fig. 2). In the episomal assay, AAV DNA integrated into p220.2 vectors carrying STR DNA at a frequency of 0.12% (17 of 14,000 colonies). pT1 and pT2 (identical oligonucleotides inserted independently into p220.2) represent a 33-bp target lacking the putative Sp1 site but containing both the RBS and the TRS sequences. The 33-bp insert in p220.2 recombined with AAV DNA at a frequency of 0.07% (22/31,000 colonies). To assess the role of the TRS site in AAV integration, pG1 and pG2 were constructed; pG1/G2 are similar to pT1/T2, except for a four-nucleotide alteration in the TRS, which renders it dysfunctional for Rep-mediated nicking activity (G1 and G2 are identical oligos inserted independently into p220.2). Clearly, the G1 and G2 targets have lost the ability to support AAV integration (0/29,800 colonies compared with 23/31,000 colonies for pT1 and pT2). Further evidence for the essential role of the TRS sequence in the target for AAV integration is provided by the inserts SR5 and SP1.6. SR5 is a 62-nt oligomer similar to STR, except that the TRS site is absent; in contrast to pSTR (17/14,000 colonies), pSR5 does not support AAV integration (0/30,000 colonies). SP1.6 is a 40-nt oligomer that includes the Sp1 recognition site but no TRS sequence appropriately positioned with respect to the Sp1 site; pSP1.6 does not recombine with AAV DNA (0/18,000 colonies examined). The data therefore indicate that a target sequence of 33 nucleotides, comprising the TRS and the RBS sequence elements separated by an 8-nt spacer (T1 and T2, Fig. 2) contains the minimal necessary and sufficient elements that direct site-specific AAV integration. Elimination either of the TRS or RBS elements from the target aborts integration.

The frequency of recombination between *AAVS1* DNA sequences and AAV DNA was influenced by the length of the target DNA in which the TRS and RBS elements are embedded. Thus the 510-bp *AAVS1* segment recombines at a fre-

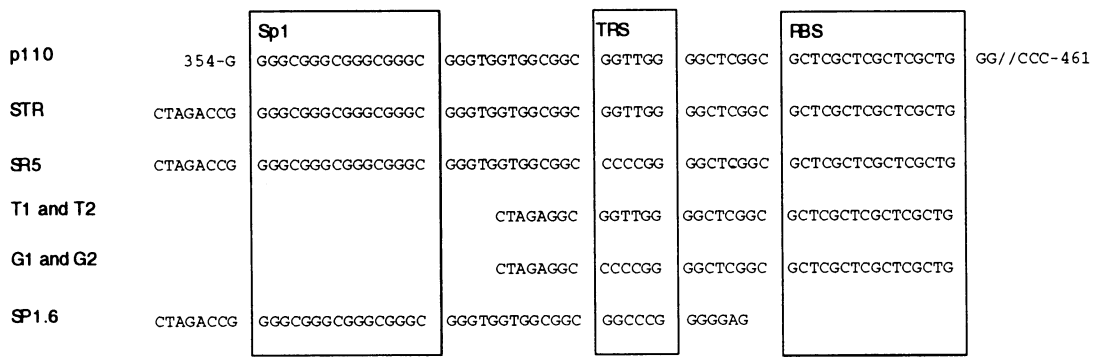

FIG. 2. Oligonucleotides inserted into p220.2 to assess their recombinogenic activity. The oligo sequences are aligned to those of p110 described in Fig. 1. Boxes indicate the sequences of the Sp1 recognition site, TRS, and RBS. The recombinogenic activity of the oligonucleotides in p220.2 is given in Table 2.

Table 2. Genotype and recombination frequencies of AAVS1 oligonucleotides in p220.2 vectors.

DNA	oligonucleotide length/position in *AAVS1*	Genotype SP1	TRS	RBS	Recombination (AAV-positive colonies, no./total)
pSTR	62/352–413	+	+	+	17/14,000
pSR5	62/352–413	+	−	+	0/30,000
pSP1.6	40/352–391	+	−	−	0/18,000
pT1+pT2	33/381–413	−	+	+	22/31,000
pG1+pG2	33/381–413	−	−	+	0/29,800

See Fig. 2 for description of the oligonucleotides, which were inserted into p220.2 and are listed in the first column. The oligonucleotide length in the second column does not include the five-nucleotide cloning site (*Xba*I and *Sal*I) for insertion into p220.2.

quency of 0.3%, whereas the 110-bp *AAVS1* segment recombines at a frequency of 0.13% (Table 1). Similarly, the 62-bp pSTR target recombines at a frequency of 0.12%, whereas the 33-bp pT1 (or pT2) target recombines at a frequency of 0.07% (Table 2). The reason for the effect of target length in episomal integration is not known.

A Sequence Responsible for Episome DNA Rearrangements Maps to a 117-bp Segment of *AAVS1* That Is Not Required for AAV Integration. As shown previously, a subpopulation of p220.2 vectors containing *AAVS1* DNA were rearranged upon passage in C17 cells (22). To address the question whether the DNA rearrangements are dependent on the presence of the recombination signals involved in AAV DNA integration, several of the episomal shuttle vectors used in this study were analyzed for their stability in C17 cells. p220.2 vectors from AAV-infected as well as mock-infected C17 cells propagating p110, p330, pHae1, pHae2, and pSTR *AAVS1* DNA inserts were rescued by electroporation of the respective extrachromosomal DNA extract into *E. coli* cells. The DNA from 20 colonies (all AAV negative) rescued from each cell line was subjected to an analytical *Sma*I digest and compared with the *Sma*I digest of the same vector DNA before passage in C17 cells. As summarized in Table 3, the vectors pHae1, p110, and pSTR were not rearranged after passage in C17 cells. In contrast, 20–30% of the vectors p330 and pHae2 (rescued from infected as well as mock-infected cells) showed rearrangements upon passaging in C17 cells. Inspection of the *AAVS1* DNA in these vectors (Table 3) indicates that episomal DNA rearrangements in C17 cells are associated with the presence of a 117-bp DNA segment (*AAVS1*, nt 209–326 present in pHae2) that is not required for AAV integration (cf. pHae1 and pHae2, Table 1).

DISCUSSION

In this paper we present data showing that two signal sequences (TRS and RBS) are required and that a 33-nucleotide sequence including these signals found in *AAVS1* is sufficient for AAV site-specific integration. This is a unique system for viral DNA integration and represents the only case of integration by a mammalian virus at a single site in the host genome. Several questions are raised by these observations. What role, if any, does this region play in the physiology of the cell? Why has the cell maintained the sequence if it can serve as an integration site for a virus? Analogous signal sequences within the AAV genome serve as an origin for viral DNA replication; does it serve a similar function in the cell? How has the virus evolved to recognize the site? The origins of all viruses are uncertain; in the case of AAV, however, one issue is whether the virus evolved to be able to use the site or has somehow evolved from the sequence at the site. Why does the site occur only once in the human genome; is it simply by random chance?

Although the two defined signals, TRS and RBS, combined within the integration site apparently occur only once in the human genome in the required orientation, RBS occurs more than 15 times. In all but one of the cases represented in the current data bases, the sequence is in the 5′-untranslated region of identified genes. Thus, it seems that RBS may serve as a binding site for a cellular regulatory protein in addition to the AAV Rep protein.

Because some p220.2 vectors carrying the 510-bp 5′ end of *AAVS1* show evidence of DNA rearrangements after passage in C17 cells (22), the question arose of the connection between vector DNA instability and AAV integration. The present data clearly show that the *AAVS1* sequence responsible for episomal DNA rearrangements and the one that signals targeted integration are separate elements (Tables 1 and 3). pHae1 (*AAVS1* nt 347–446), while a target for AAV integration, does not induce episomal DNA rearrangements; in contrast, the pHae2 insert (*AAVS1* nt 209–326) is not a target for AAV integration, but it does induce vector DNA instability. Similarly, the p330 insert (*AAVS1* nt 18–353) induces episomal DNA rearrangements, but it is not a target for AAV integration. Interestingly, the *AAVS1* DNA nt 209–326 carries a motif, M26, which has been characterized as an enhancer of meiotic gene conversion in fission yeast (32, 33). This motif also partly overlaps a cAMP response element (Fig. 1). We are currently testing whether it

Table 3. Frequency of DNA rearrangements in p220.2 vectors propagated in C17 cells.

	Inserts that include RBS and TRS sites						Inserts that include M26/CRE sites			
	p110 (354–461)		pHae 1 (347–446)		pSTR (352–413)		p330 (18–353)		pHae 2 (209–326)	
	Inf	Mock	Inf	Mock	Inf	Mock	Inf	Mock	Inf	Mock
colonies with rearranged p220.2 DNA (no./total)	0/20	0/16	0/20	0/12	0/18	0/15	6/17 24%	6/20 30%	3/11 27%	3/15 20%

The p220.2 vectors containing the indicated *AAVS1* DNA inserts (*AAVS1* nucleotide positions are in parenthesis) were serially passaged five times in C17 cells and then rescued and amplified in *E. coli*. The figures refer to the number of p220.2 isolates whose *Sma*I digestion pattern differed from that of the parental p220.2 vectors before passage in C17 cells. Although p220.2 vectors were also rescued from AAV infected cultures of C17 cells (Inf) none of these vectors contained AAV DNA.

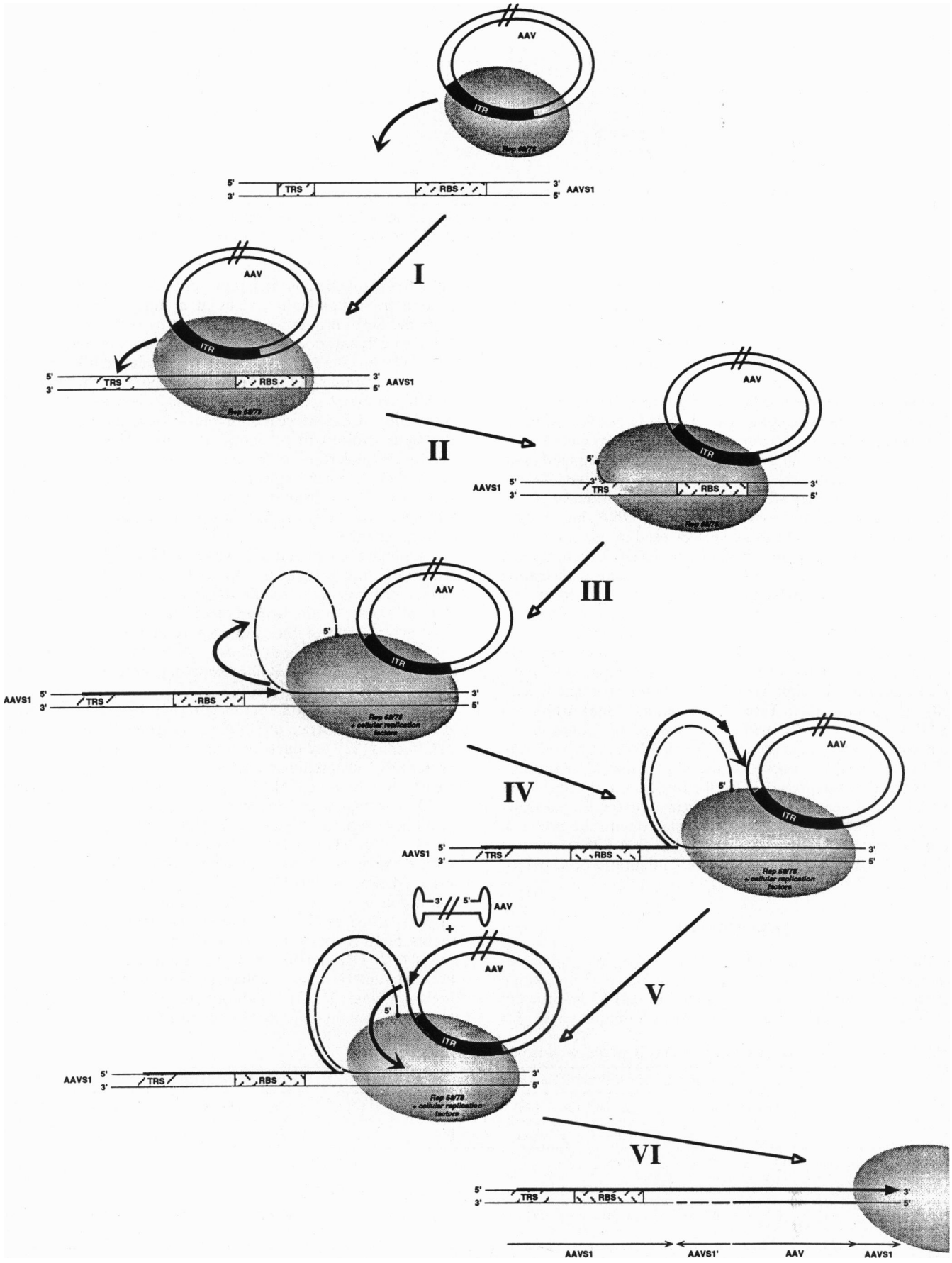

FIG. 3. A model for the AAV site-specific DNA integration. Parallel lines in the AAV molecules are indicating that the sizes of DNA structures in the figure are not drawn in their actual proportions. A thick grey line indicates the newly synthesized strand; (*Fig. 3 legend continues on the opposite page)*

is the M26 sequence itself that is required for the instability observed. The q arm of chromosome 19 contains a fragile site and has been refractory to cloning in yeast artificial chromosomes and cosmids because of frequent rearrangements. It is possible that the region in *AAVS1* is among those responsible for the observed instability. In the chromosomal site fragility or instability may promote a more open structure that could facilitate integration of AAV.

It is important to note that the rearrangement observed with the M26 signal and associated sequences was not dependent on AAV infection. However, rearrangement is also associated with AAV integration. In the shuttle vector *AAVS1*, sequences upstream of the junction with AAV were either deleted or rearranged in all the recombinant structures that have been analyzed in detail (23). Portions of the AAV genome were frequently deleted. Similarly, flanking sequences in chromosome 19 were found to be rearranged after AAV integration, and internal AAV sequences are also rearranged (16). Whether these rearrangements are due to the same mechanism involved in the instability of the shuttle vector remains to be determined.

A model of AAV site-specific integration must be able to account for the features observed for AAV integration. The TRS and the RBS must play an integral role, and junctions must be within the ITR or near the leftward most promoter at map position five (the two sites seen with equal frequency in AAV–*AAVS1* junctions derived from both the shuttle vector system and in chromosome 19). The potential for DNA rearrangements of *AAVS1* during integration should be present but may be too complex for presentation in a simple form. The presence of the TRS and RBS in the integration site strongly suggests that some form of Rep-mediated DNA replication is involved in the integration process. The data of Urcelay *et al.* (31) showing that AAV-specific DNA replication can be initiated on a plasmid containing the essential elements of the integration site support this notion. On the other hand, the presence of head-to-tail junctions in the integrated copies of the AAV genome cannot be explained by the accepted model of AAV DNA replication, which is based on events occurring during productive infection. This model predicts only head-to-head and tail-to-tail junctions. For head-to-tail junctions to occur, AAV DNA replication involved in the integration process must be based on a mechanism involving some form of circular AAV template.

Six stages in a possible model for AAV site-specific integration are illustrated in Fig. 3. In stage one, a circular, double-stranded form of the AAV genome is brought into register with the specific integration site. This is accomplished by a complex involving several copies of Rep 68/78, which can bind both to the integration site and the circular AAV genome (either within the AAV ITR or at a known site close to the p5 promoter) and thus serve to link the AAV genome to the integration site. This has been demonstrated *in vitro* by the work of Weitzman *et al.* (30). The structure of the AAV genome at this stage is not yet determined. However, to account for structures of integrated AAV genomes identified from latently infected cell lines as well as from the p220.2 shuttle vector system, we assume that AAV is in a circular duplex form. Duplex AAV DNA could be circularized either by end to end joining (known to occur in mammalian cells) or by recombination between the palindromic ITRs.

In stage two a nick is introduced by Rep 68/78 at the TRS in *AAVS1*, and DNA synthesis is initiated by cellular enzymes (for the sake of simplicity we ignore the possibility that AAV DNA synthesis could also initiate at the AAV origin). *AAVS1* DNA synthesis then proceeds by a single-strand displacement mechanism, so that the parental strand complementary to the template is simply displaced. It should be noted that the 5′ end of the displaced strand is covalently attached to Rep 68/78, and, thus, it may be in a complex with the Rep linkage of AAV DNA to *AAVS1* (29).

In stage three, the elongating strand switches templates. Template strand switching appears to be a feature of AAV DNA replication. During productive replication, e.g., when one strand is completed, the replication complex can reinitiate to copy the complementary strand. Strand switching can also occur prematurely before the template strand has been completely copied, causing the synthesis of defective DNA molecules, and has been observed both *in vivo* and *in vitro* (25, 39). Analysis of integrants from the shuttle vector system leads us to propose that the strand switch is likely to occur near the RBS and that the new template is most likely to be the displaced single strand. At this step, the replication complex is creating an inversion of the displaced *AAVS1* sequence. When synthesis approaches the end of this new template, which is covalently linked to Rep 68/78, a second strand switch occurs. This switch can be either onto AAV, creating a link of the newly made inverted repeat with AAV (stage four) or back onto *AAVS1*, in which case it will not be detectable in our system. If the switch has been onto AAV, replication of AAV sequences, now linked to *AAVS1*, can occur until a third strand switch occurs, back onto *AAVS1* (stage five). (Replication on the AAV template may be greater than unit length to account for concatamers of the AAV genome which are observed as integrants.) The point of switch from AAV is again likely to be at or near an RBS (either in the ITR or near the p5 promoter). However, recombinants analyzed also indicate that this final switch, creating the second link between AAV and *AAVS1* sequences, can occur at other places within the AAV genome. Finally, stage six involves repair of DNA structures containing noncomplementary strands created by the proposed replication-mediated integration mechanism.

The model has features that predict a rearrangement of *AAVS1* at one junction with AAV and places the potential junctions near RBS in *AAVS1* and in the ITR or near p5 in AAV. These features are in accord with our observations and serve to extend a model of integration based on DNA synthesis that has also been proposed by Kotin *et al.* (40).

In this paper we describe a recombination event that is initiated by recognition of a specific DNA signal (TRS/RBS) by a viral protein (Rep 68/78). With the exception of gene rearrangements observed in the immune system, this remains the only known example of site-specific nonhomologous recombination in multicellular organisms (for review, see ref. 41). Our current data indicate, however, that a different mechanism is used. We expect that the model as presented here is not likely to be correct in detail. However, it does take into account the particular features observed for recombinants in which AAV has inserted into *AAVS1*. It is particularly necessary to suggest why junctions occur at specific sites in the AAV genome and are likely to be near RBS in *AAVS1* in the model system used in this paper. The model also takes into account that *AAVS1* contains an AAV origin and offers some

the dashed line indicates the displaced strand of *AAVS1*. (*I*) Complex formation between AAV and *AAVS1* is mediated by Rep 68/78. (*II*) Introduction of a strand-specific nick at the TRS in *AAVS1* by Rep 68/78 and assembly of cellular replication factors. (*III*) DNA synthesis by single-strand displacement originating at the TRS is followed by template strand switch onto the displaced strand. (*IV*) A second strand switch occurs onto AAV creating a link between *AAVS1* and AAV sequences. (*V*) After synthesis of AAV DNA sequences, a third template strand switch back onto *AAVS1* results in a second link between viral and host DNA sequences. (*VI*) Repair of DNA structures containing noncomplementary strands by cellular enzymes results in integrated copies of AAV DNA within *AAVS1*.

suggestion as to how rearrangements in the flanking sequences can occur.

We thank Dr. Peter Ward for his original contribution regarding the integration model; Nenita Cortez and Bernard Danovitch for excellent technical assistance; and Drs. S. Cotmore, W. Holloman, T. Kelly, N. Lue, M. Malkinson, P. Tattersall, and M. Wiedmann for critical review of the manuscript. This work was supported by Grant AI22251 from the National Institute of Allergy and Infectious Diseases and by Grant GM50032 from the National Institute of General Medical Sciences. R.M.L. was supported in part by a fellowship from the Norman and Rosita Winston Foundation.

1. Berns, K. I. & Linden, R. M. (1995) *BioEssays* **17,** 237–245.
2. Atchison, R. W., Casto, B. C. & Hammon, W. (1965) *Science* **149,** 754–756.
3. Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. (1966) *Proc. Natl. Acad. Sci. USA* **55,** 1467–1471.
4. Buller, R. M., Janik, J. E., Sebring, E. D. & Rose, J. A. (1981) *J. Virol.* **40,** 241–247.
5. Weindler, F. W. & Heilbronn, R. (1991) *J. Virol.* **65,** 2476–2483.
6. Yakinoglu, A. O., Heilbronn, R., Burkle, A., Schlehofer, J. R. & zur Hausen, H. (1988) *Cancer Res.* **48,** 3123–3129.
7. Yakobson, B., Koch, T. & Winocour, E. (1987) *J. Virol.* **61,** 972–981.
8. Yakobson, B., Hrynko, T. A., Peak, M. J. & Winocour, E. (1989) *J. Virol.* **63,** 1023–1030.
9. Berns, K. I., Pinkerton, T. C., Thomas, G. F. & Hoggan, M. D. (1975) *Virology* **68,** 556–560.
10. Cheung, A. K., Hoggan, M. D., Hauswirth, W. W. & Berns, K. I. (1980) *J. Virol.* **33,** 739–748.
11. Handa, H., Shiroki, K. & Shimojo, H. (1975) *J. Gen. Virol.* **29,** 239–242.
12. Hoggan, M. D., Thomas, G. F. & Johnson, F. B. (1972) *Continuous "carriage" of Adenovirus-Associated Virus Genomes in Cell Cultures in the Absence of Helper Adenovirus.* (North-Holland, Amsterdam), pp. 243–249.
13. Srivastava, A., Lusby, E. W. & Berns, K. I. (1983) *J. Virol.* **45,** 555–564.
14. Mendelson, E., Smith, M. G., Miller, I. L. & Carter, B. J. (1988) *Virology* **166,** 612–615.
15. Berns, K. I. (1996) in *Parvoviridae: The Viruses and Their Replication*, eds. Fields, B. N., Knipe, D. M. & Howley, P. M. (Lippincott-Raven, Philadelphia), Vol. 2, pp. 2173–2220.
16. Kotin, R. M. & Berns, K. I. (1989) *Virology* **170,** 460–467.
17. Kotin, R. M., Siniscalco, M., Samulski, R. J., Zhu, X. D., Hunter, L., Laughlin, C. A., McLaughlin, S., Muzyczka, N., Rocchi, M. & Berns, K. I. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2211–2215.
18. Kotin, R. M., Menninger, J. C., Ward, D. C. & Berns, K. I. (1991) *Genomics* **10,** 831–834.
19. Samulski, R. J., Zhu, X., Xiao, X., Brook, J. D., Housman, D. E., Epstein, N. & Hunter, L. A. (1991) *EMBO J.* **10,** 3941–3950.
20. Kotin, R. M., Linden, R. M. & Berns, K. I. (1992) *EMBO J.* **11,** 5071–5078.
21. Yates, J. L., Warren, N. & Sugden, B. (1985) *Nature (London)* **313,** 812–815.
22. Giraud, C., Winocour, E. & Berns, K. I. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10039–10043.
23. Giraud, C., Winocour, E. & Berns, K. I. (1995) *J. Virol.* **69,** 6917–6924.
24. Straus, S. E., Sebring, E. D. & Rose, J. A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 742–746.
25. Hauswirth, W. W. & Berns, K. I. (1979) *Virology* **93,** 57–68.
26. Owens, R. A., Weitzman, M. D., Kyostio, S. R. & Carter, B. J. (1993) *J. Virol.* **67,** 997-1005.
27. Im, D. S. & Muzyczka, N. (1989) *J. Virol.* **63,** 3095–3104.
28. Snyder, R. O., Im, D. S., Ni, T., Xiao, X., Samulski, R. J. & Muzyczka, N. (1993) *J. Virol.* **67,** 6096–6104.
29. Im, D. S. & Muzyczka, N. (1990) *Cell* **61,** 447–457.
30. Weitzman, M. D., Kyostio, S. R., Kotin, R. M. & Owens, R. A. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 5808–5812.
31. Urcelay, E., Ward, P., Wiener, S. M., Safer, B. & Kotin, R. M. (1995) *J. Virol.* **69,** 2038–2046.
32. Grimm, C., Bahler, J. & Kohli, J. (1994) *Genetics* **136,** 41–51.
33. Schuchert, P., Langsford, M., Kaslin, E. & Kohli, J. (1991) *EMBO J.* **10,** 2157–2163.
34. Canfield, V., Emanuel, J. R., Spickofsky, N., Levenson, R. & Margolskee, R. F. (1990) *Mol. Cell. Biol.* **10,** 1367–1372.
35. Graham, F. L. & Eb, A. J. v. d. (1973) *Virology* **52,** 456–467.
36. Felgner, P. L., Gadek, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M. & Danielsen, M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7413–7417.
37. Stamatatos, L., Leventis, R., Zuckermann, M. J. & Silvius, J. R. (1988) *Biochemistry* **27,** 3917–3925.
38. Hirt, B. (1967) *J. Mol. Biol.* **26,** 365–369.
39. Ward, P. J. & Berns, K. I. (1996) *J. Virol.*, in press.
40. Kotin, R. M. (1995) *Hum. Gene Ther.* **5,** 793–801.
41. Lewis, S. M. (1994) *Adv Immunol* **56,** 27–150