Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence

CATHERINE GIRAUD, ERNEST WINOCOUR*, AND KENNETH I. BERNST†

Department of Microbiology, Hearst Microbiology Research Center, Cornell University Medical College, 1300 York Avenue, New York, NY 10021

Communicated by Daniel Nathans, June 9, 1994 (received for review April 8, 1994)

ABSTRACT Different regions of an 8.2-kb cloned DNA segment containing the target for adeno-associated virus (AAV) integration in human chromosome 19q13.3–qter (AAVS1 locus) were subcloned in an Epstein–Barr virus-based shuttle vector and propagated as episomes in a derivative of the 293 human embryonic kidney cell line. Preferential recombination with an infecting AAV genome was assessed by measuring the frequency of recombinants among the shuttle vectors recovered in *Escherichia coli*. The signals which direct recombination with the AAV genome were localized to a 510-nt region at the 5' end of the 8.2-kb AAVS1 DNA. Hence, the results indicate that site-specific integration of AAV is directed by a specific DNA sequence on human chromosome 19. An unusual degree of DNA heterogeneity in the recovered vector was also associated with the 510 nt at the 5' end of AAVS1 DNA, suggesting that the AAV chromosomal integration locus may be involved in genomic instability.

The outstanding feature of the life cycle of the human parvovirus adeno-associated virus 2 (AAV) is the establishment of latent infection by integration into the cellular genome. The significance of latent infection in the life cycle is perhaps best illustrated by the fact that replication of the virus in cell culture normally requires coinfection with an unrelated helper virus, either adenovirus or herpesvirus (1). More recently it has been demonstrated that some lines of cells can be made permissive for productive infection by AAV alone if the cells are exposed to a variety of genotoxic agents (2, 3). The conditions which render a cell permissive for AAV replication, either infection by a cytolytic nuclear DNA virus or exposure to genotoxic agents, have led to a model for the regulation of the viral life cycle (1). The model hypothesizes that if the virus infects a healthy cell, a latent infection will be established. The viral genome will remain relatively quiescent in the integrated state until the cell is severely stressed (e.g., by exposure to a genotoxic agent or virus). Such stress will not only activate cellular repair genes but also activate AAV gene expression, leading to rescue from the integrated state and viral replication.

Integration of the AAV genome occurs at a specific locus, 19q13.3–qter, in a majority of clones when several continuous lines of human cells are latently infected at a high multiplicity of infection (4–7). The preintegration site on chromosome 19q has been cloned and a 4-kb region has been sequenced (6). The sequence, which shows no homology to that of the AAV genome, has several notable features, including an overall G+C content of 65%. The 900 bases at the 5' end have a G+C content of 82% and are close to the site of recombination with viral sequences in a number of clones. The same region also contains binding sites for several transcriptional transactivators. The 900 bases at the 5' end and the 1500 bases at the 3' end have a larger than expected number of direct repeats, while the middle region

contains several open reading frames corresponding to an isolated cDNA from a λ phage library of human foreskin fibroblasts (6). Most interestingly, the 5'-most 500 bases contain a dodecamer, (GCTC)₃, also present in the inverted terminal repeat of the viral DNA, which has been identified as a binding site for the AAV Rep68/78 protein, the primary viral regulatory protein (8, 9). The Rep68/78 protein is both a site-specific DNA nickase and a helicase and may be involved in the integration process (10–12). Indeed, AAV modified to serve as a gene vector in a manner which removes the *rep* gene no longer displays the same degree of site specificity of integration (R. J. Samulski, personal communication; W. G. Kearns, personal communication).

From the above, it appears that the sequence of the preintegration site is likely to be a determining factor in the site specificity of AAV integration. However, the question of whether the sequence of the preintegration site is sufficient, in terms of cellular parameters, to determine site specificity has not been answered. In this paper we describe experiments which directly address this issue by using an Epstein–Barr virus (EBV)-based shuttle vector which has been reported to be highly stable during propagation in mammalian cells (13). DNA containing the AAVS1 preintegration site was placed in the EBV-based shuttle vector and the ability of that DNA, when no longer in the normal context of the long (q) arm of chromosome 19, to direct AAV integration was assessed. We find that AAV integration is still directed by the preintegration DNA even when it is in an episome. The cellular DNA sequence directing integration was localized to a 510-nt region at the 5' end of the AAVS1 locus. The presence of this 510-nt sequence generates instability of the EBV episome in mammalian cells.

MATERIALS AND METHODS

Subcloning of AAVS1 in p220.2. The 8.2-kb AAVS1 DNA containing the chromosome 19q preintegration locus was derived by *EcoRI* digestion of pRI-A (6) and subcloned in the *Sal* I site of the EBV-based shuttle vector p220.2 (A. Beaton and K.I.B., unpublished data). p220.2 was a gift of B. Sugden (14, 15). The resulting construct, p220.2/AAVS1(kb 0–8.2), contains the AAVS1 sequence in the B orientation and was used as starting material to generate the following set of AAVS1 deletion constructs (Fig. 1): p220.2/AAVS1(kb 0–0.51), p220.2/AAVS1(kb 0–1.6), p220.2/AAVS1(kb 0–3.5), and p220.2/AAVS1(kb 0–4.4) were derived by digestion of p220.2/AAVS1(kb 0–8.2) with, respectively, *Xba* I (site in the p220.2 polylinker)/*Pvu* II, *Bam*HI, *Xba* I/*Kpn* I, and *Xba* I/*Pvu* II (partial digest) and religation of the digestion products. The 5'-end deletion mutants p220.2/AAVS1(kb 0.51–1.6), p220.2/AAVS1(kb 0.51–3.5), and p220.2/AAVS1(kb 0.51–4.4) were derived, respectively,

Abbreviations: AAV, adeno-associated virus; EBV, Epstein–Barr virus; EBNA-1, EBV-encoded nuclear antigen 1; SV40, simian virus 40.

*Permanent address: Department of Molecular Genetics and Virology, Weizmann Institute of Science, Rehovot, Israel.

†To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

from p220.2/AAVS1(kb 0–1.6), p220.2/AAVS1(kb 0–3.5), and p220.2/AAVS1(kb 0–4.4) digested with *Pvu* II/*Hind* III (a site in the p220.2 polylinker) and religation of the products. p220.2/AAVS1(kb 0–6) was derived by reinsertion of the 6-kb *Hind* III fragment from p220.2/AAVS1(kb 0–8.2) into the *Xba* I polylinker site of p220.2. p220.2/AAVS1(kb 5.2–8.2) was obtained by digestion of p220.2/AAVS1(kb 0–8.2) with *Eco* RV plus *Xba* I and reinsertion of the resulting 3-kb fragment into the *Xba* I polylinker site of p220.2. The control plasmids p220.2/RS-2 kb and p220.2/RS-6 kb contain, respectively, 2-kb and 6-kb DNAs that were selected at random from a human WI-38 (embryonic lung fibroblasts) genomic library by *Xba* I digestion and subcloned in the *Xba* I site of the p220.2 polylinker (R. M. Linden and K.I.B., unpublished data). Neither of these control segments of human DNA hybridize with AAVS1 DNA.

Propagation of EBV p220.2-Based Shuttle Vectors in C17 Cells. The C17 cell line (16), an EBNA-1-expressing derivative of the 293 cell line, was kindly provided by R. F. Margolske. C17 cells were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum and G418 at 600 μ g/ml (to maintain the EBNA-1 gene, which had been cotransfected with pSV2neo). Each vector DNA (10 μ g) was transfected into C17 cells by calcium phosphate coprecipitation (17) or by the cationic lipid *N*-[1-(2,3-dioleoyloxy)propyl]-*N,N,N*-trimethylammonium methyl sulfate (DOTAP) (18, 19) according to the manufacturer's recommendations (Boehringer Mannheim). Three days later, the cells were split 1:10 into DMEM/10% fetal bovine serum with G418 (600 μ g/ml) and hygromycin B (200 μ g/ml). For each construct, 10–30 hygromycin-resistant cell clones arising from a single transfection experiment were pooled and passaged every 4 or 5 days at a split ratio of 1:10 to 1:20 in the constant presence of both G418 and hygromycin. By passage level 2, all of the control mock-transfected C17 cells had been killed by hygromycin. Cells at passage levels 3–45 were used for AAV infection experiments.

Virus Infection. Subconfluent cells (5×10^6) in 10-cm-diameter dishes were infected with a single cesium chloride-purified stock of AAV2 (2) at an input multiplicity of 20 infectious units per cell (400 virus particles per cell). Control cultures were either mock-infected or infected with simian virus 40 (SV40) at 20 plaque-forming units per cell. The infected and control cells were incubated for 48 hr in G418/hygromycin medium.

Isolation of Extrachromosomal DNA. Forty-eight hours after infection, extrachromosomal DNA was isolated by the standard Hirt procedure (20) except that the supernatant fraction, obtained after centrifugation at $48,000 \times g$ for 30 min at 4°C, was digested with proteinase K (100 μ g/ml) for 4 hr at 42°C and for 24 hr at room temperature. After further deproteinization by extractions with phenol and with chloroform containing 2% isoamyl alcohol, the nucleic acids were purified by three successive ethanol (100%) precipitations in the presence of 0.2 M NaCl. The extrachromosomal nucleic acids from a single plate of 10^7 cells were resuspended in 20 μ l of 10 mM Tris-HCl/1 mM EDTA, pH 8.0.

Recovery of Vectors in *Escherichia coli* and Colony Hybridization. One-seventh of the extrachromosomal DNA extracted from 10^7 cells was transfected into the SURE (Stratagene) strain of *E. coli* by electroporation (Bio-Rad *E. coli* Pulser) at an applied voltage of 1.9 kV, using a 0.2-cm-gap cuvette (time constant, 4.0–4.2 msec). The electroporated bacteria (40 μ l) were diluted in 1 ml of SOC medium (Stratagene) and incubated at 37°C for 1 hr. As measured with pUC18 plasmid DNA, the efficiency of transfection was 5×10^9 ampicillin-resistant (Amp^R) colonies per μ g of DNA. The *E. coli* SURE strain (21) minimizes DNA rearrangements, due to mutations in the *recB*, *uvrC*, *umuC*, and *sbcC* genes. High-efficiency electroporation of the recombination-

deficient SURE strain of *E. coli* was an important component of the experimental protocol. One-tenth (100 μ l) of the electroporated bacteria were spread on LB/Amp agar plates to determine the number of Amp^R colonies. A sample of the remaining fraction, adjusted to produce 500–1000 colonies, was diluted with 10 ml of LB medium and plated under suction on a 8.2-cm-diameter BA85 nitrocellulose filter (Schleicher & Schuell) prewashed with LB medium. The filter was then incubated, bacteria-side-up, on top of LB/Amp agar until colonies 1–2 mm in diameter appeared. Imprints ("replicas") made from the "master" filters by standard procedures were incubated on LB/Amp agar until colonies 0.5–1.0 mm in diameter developed. Colony hybridization of the replica filters with a nonradioactive single-stranded AAV DNA probe (prepared from purified virus) was carried out using the Genius system (Boehringer Mannheim). Hybridization was at 68°C for 18–20 hr.

FIG. 1. Subcloning of the AAV preintegration locus AAVS1 in the EBV-based vector p220.2. (A) The 8.9-kb p220.2 plasmid, containing the herpes simplex virus (HSV) thymidine kinase (tk) promoter and polyadenylation signal (poly a) flanking a hygromycin-resistance gene for selection in mammalian cells. EBNA-1, EBV-encoded nuclear antigen 1. (B) Schematic representation of the 8.2-kb AAVS1 DNA sequence and the p220.2 vectors containing various segments of AAVS1. The AAVS1 sequence (top bold horizontal line) shows a CpG island (nt 1–900; stippled box), a region corresponding to a partial cDNA clone (nt 1620–2318; hatched box), a minisatellite repetitive DNA sequence (nt 3660–4021; open bar), crossover sites associated with AAV integration (arrows), and a binding site for the AAV rep gene products within nt 353–468 (asterisk) (8). The restriction sites on the AAVS1 sequence and on the polylinker used for the construction of the p220.2 vectors are shown on the lower scale (thin and dotted horizontal line). p220.2/RS-2 kb and p220.2/RS-6 kb contain random fragments of human DNA from a WI-38 library, inserted in the polylinker.

Table 1. Recombination between AAV and the p220.2 shuttle vector carrying the 8.2-kb insert of chromosome 19q preintegration DNA (AAVS1)

moi	Passage no. of cell line	Hours after infection	Recombination	
			AAV-hybridizing colonies, no./total	%
20	4	48	16/500	3.2
	8	48	13/500	2.6
	17	48	8/540	1.5
	45	48	22/1000	2.2
50	11	10	1/1305	0.08
		23	41/2142	1.9
		46	65/2943	2.2

C17 cells carrying the vector p220.2/AAVS1(kb 0–8.2) were infected with AAV (at 20 or 50 infectious units per cell, as indicated by the multiplicity of infection, moi) at the passage shown (no. of cell passages after the start of hygromycin selection). At the designated times after infection, extrachromosomal DNA was isolated and electroporated into *E. coli*. Recombination is given by the percentage of Amp^R colonies which hybridized with an AAV DNA probe.

RESULTS

Recombination Hotspot Within the AAVS1 Preintegration Locus. To assess preferential recombination between the AAV genome and its integration site on the q arm of human chromosome 19, the following experiment was devised. Defined portions of a cloned 8.2-kb DNA segment containing the preintegration locus AAVS1 were inserted into the EBV-based shuttle vector p220.2 and propagated as extrachromosomal elements in C17 cells (Fig. 1). p220.2 contains the EBV latent origin of replication *oriP*, the EBNA-1 gene, whose product transregulates *oriP*, a hygromycin-resistance gene for the selection of C17 cells propagating this shuttle vector, and the pBR322 replication origin and ampicillin-resistance gene for selection in *E. coli* (14, 15). The C17 derivative of the adenovirus 5-transformed human 293 cell line expresses EBNA-1 constitutively, which may facilitate the propagation of p220.2 as an extrachromosomal nuclear episome (16, 22). C17 cells carrying p220.2/AAVS1 episomes were infected with AAV at various cell passage levels. Forty-eight hours later, recombination between the infecting viral genome and the resident p220.2/AAVS1 episomes was measured by electroporating the extrachromosomal DNA fraction into the recombination-deficient SURE strain of *E. coli* and by colony hybridization with a single-stranded AAV virion DNA probe. The number of AAV-positive Amp^R colonies expressed as a percentage of the total number of Amp^R colonies was taken as the measure of the frequency of recombination. The results are presented in Tables 1 and 2.

p220.2 containing the entire 8.2-kb AAVS1 DNA [p220.2/AAVS1(kb 0–8.2)], when propagated as an episome in C17 cells, recombined with the infecting AAV genome at every cell passage tested (Table 1). The recombination values, in the range 1.5–3.2%, did not decrease with increasing cell-passage levels, indicating that the recombination targets on p220.2/AAVS1 were conserved during continuous passage of the cells. As a control, we examined the recombinogenic capacity of p220.2 vectors containing 2-kb and 6-kb random segments of human DNA. No evidence for recombination between those vectors propagated in C17 cells and the infecting AAV genome was obtained (>4000 Amp^R colonies were screened). Similarly, no recombination between the parental p220.2 vector and the infecting AAV genome was detected. The time course of accumulation of recombinants showed that near-maximum yields were obtained by 23 hr postinfection (Table 1). Only a rare recombinant was detected at 10 hr postinfection; this observation supports our assumption that the primary recombination events occurred in C17 cells and not in the recombination-deficient *E. coli*

Table 2. Recombination between AAV and the p220.2 shuttle vector containing different portions of AAVS1 DNA

AAVS1 insert in p220.2 vector, kb	Passage no. of cell line	Recombination	
		AAV-hybridizing colonies, no./total	%
0–6	11	3/240	1.2
0–4.4	14	1/504	0.2
0–1.6	9	8/957	0.8
0–1.6	12	5/288	1.7
0–0.51	11	17/2140	0.8
0–0.51	12	2/459	0.4
0.51–4.4	12	0/1000	<0.1
0.51–4.4	14	0/1640	<0.06
0.51–3.5	14	0/2740	<0.04
0.51–1.6	9	0/3800	<0.03
0.51–1.6	12	0/648	<0.15
5.2–8.2	11	0/4000	<0.03

See Fig. 1 for the derivation of the various AAVS1 inserts.

SURE bacteria. A time interval of 48 hr postinfection was chosen, however, for the routine isolation of the shuttle vectors from the infected cells.

We next investigated the recombinogenic potential of different regions of the 8.2-kb AAVS1 DNA inserted into the p220.2 vector (Table 2). p220.2 vectors carrying the AAVS1 segments spanning kb 0–6.0, 0–4.4, 0–1.6, and 0–0.51 all recombined with AAV. However, deletion of the 510 nt at the 5' end of AAVS1 completely abrogated the capacity of the AAVS1 DNA segments (kb 0.51–4.4, 0.51–3.5, 0.51–1.6) to recombine with the AAV genome. In these cases, not a single colony out of ≈9800 Amp^R colonies screened reacted with the AAV DNA probe. A 3-kb segment at the 3' end of AAVS1 (kb 5.2–8.2) did not recombine with the AAV genome (≈4000 Amp^R colonies were screened). Together, >14,000 Amp^R colonies screened gave a negative hybridization result, highlighting the specificity of the positive reactions between vectors carrying the 5'-most 510 nt of AAVS1 and AAV DNA. These data clearly localized the recombinogenic signals to 510 nt at the 5' end of AAVS1.

To determine whether the AAVS1 hotspot for AAV recombination is specific for that virus, we infected the C17 lines propagating p220.2 vectors containing AAVS1 DNA (kb 0–0.51, 0–1.6, and 0–8.2) with SV40 and screened, by colony hybridization, for p220.2/AAVS1/SV40 recombinants among the vectors recovered in *E. coli*. No recombinants were detected out of the 928 colonies screened. The 293 parent line of C17 is permissive for SV40 infection, as determined by replication of the viral DNA (23). It appears, therefore, that the AAVS1 hotspot for AAV recombination is not a comparable hotspot for SV40 recombination.

Episomes Containing the 5' End of AAVS1 Are Unstable.

During the course of the above experiments, it was noted that the recovery of p220.2 vectors in *E. coli*, measured by the number of Amp^R colonies resulting from the electroporation of the extrachromosomal DNA fraction of 10⁷ C17 cells, varied widely, depending upon the segment of AAVS1 DNA present in the vector (Table 3). When the insert in p220.2 was derived from the 3' end of AAVS1 (kb 5.2–8.2), the recovery in *E. coli* (e.g., 120,000 Amp^R colonies per 10⁷ C17 cells; Table 3, Exp. 1) was comparable to that of p220.2 vectors containing a 2-kb random segment of human DNA (140,000 Amp^R colonies per 10⁷ C17 cells) or no DNA insert (184,000 colonies per 10⁷ C17 cells). In contrast, when the inserts contained the 5' end of AAVS1 (kb 0–8.2, 0–4.4, 0–1.6, 0–0.51) the efficiency of recovery in *E. coli* was reduced by a factor of 50–100. Deletion of the 510 nt at the 5' end (kb 0.51–1.6, 0.51–3.5, or 0.51–4.4) showed an increase of the

Table 3. Recovery of p220.2 vectors in *E. coli* depends upon the DNA of their inserts

Exp.	Insert in p220.2 vector	Infection	Amp ^R colonies per extract of 10 ⁷ cells
1	AAVS1		
	kb 0–8.2	AAV	2,345
	kb 0–8.2	SV40	2,200
	kb 0–1.6	AAV	2,211
	kb 0–0.51	AAV	4,958
	kb 0.51–1.6	AAV	13,400
	kb 5.2–8.2	AAV	120,000
	Random human DNA, 2 kb	SV40	140,000
	None	AAV	184,000
2	AAVS1		
	kb 0–8.2	AAV	3,015
	kb 0–1.6	AAV	2,144
	kb 0–0.51	AAV	3,417
	kb 0–0.51	SV40	3,484
	kb 0–0.51	None	1,600
	kb 0.51–1.6	AAV	4,824
	kb 0.51–1.6	None	5,829
	kb 0.51–4.4	AAV	26,800
	kb 0.51–4.4	None	26,000
	kb 5.2–8.2	None	268,000
3	AAVS1		
	kb 0–4.4	AAV	1,698
	kb 0–4.4	None	1,447
	kb 0–1.6	None	1,206
	kb 0.51–4.4	AAV	19,025
	kb 0.51–3.5	AAV	31,356
	kb 0.51–3.5	None	31,658

efficiency of recovery by a factor 5–10. These data show that the recovery of the shuttle vectors in *E. coli* decreases in the presence of sequences from the first 4.4 kb at the 5' end of AAVS1. This negative influence on the recovery of the vectors in *E. coli* was observed irrespective of the infecting virus (AAV or SV40), as well as in mock-infected cell lines.

Restriction digests of p220.2/AAVS1 vectors recovered from mock-infected C17 cells by electroporation of the extrachromosomal DNA fraction into the recombination-deficient SURE strain of *E. coli* are shown in Fig. 2. In each case, five independent Amp^R colonies were isolated and the plasmids were analyzed by restriction analysis. When the 5'-most 510 nt of AAVS1 were present in the vector, propagation as an episome in C17 cells generated an unusual degree of heterogeneity, as judged by the restriction digests of the recovered vectors compared with those of the input p220.2/AAVS1 DNAs transfected into C17 cells (Fig. 2 *a–c*). In striking contrast, when the 510 nt at the 5' end of AAVS1 were deleted from the insert in p220.2, the digests of the recovered vectors were identical to those of the parental DNAs transfected into C17 cells (Fig. 2 *d* and *e*), as well as with p220.2 plasmids carrying random human sequences (Fig. 2*f*). Interestingly, the altered pattern, compared with that of the parental plasmid, of episomes rescued from the p220.2/AAVS1(kb 0–1.6) and -(kb 0–0.51) cell lines was identical for several of the clones (Fig. 2 *b* and *c*).

DISCUSSION

The site-specific integration of AAV into the q arm of chromosome 19 in established lines of human cells implies a specific recognition signal, possibly encoded in the chromosomal DNA sequence or possibly acting indirectly via the conformation or some unique activity of the integration locus. A major conclusion of the results reported here is that AAV integration is directed by signals encoded in the primary DNA sequence of chromosome 19. These signals were localized to a 510-bp DNA segment at the 5' end of the AAVS1

FIG. 2. Restriction pattern of p220.2 vectors rescued from mock-infected C17 cell lines. (*a* and *c*) *Cla*I/*Hind*III digest of five plasmids recovered from p220.2/AAVS1(kb 0–8.2) and p220.2/AAVS1(kb 0–0.51) cell lines, respectively. (*b*, *e*, and *f*) *Pst*I digest of five plasmids recovered from p220.2/AAVS1(kb 0–1.6), p220.2/AAVS1(kb 0.51–1.6), and p220.2/RS-6 kb cell lines, respectively. (*d*) *Bam*HI/*Sal*I digest of five plasmids recovered from a p220.2/AAVS1(kb 5.2–8.2) cell line. Lane C, the parental plasmid DNA (before transfection in C17 cells) digested by the same enzymes; lane M, DNA molecular size marker (MIII or MVII of Boehringer Mannheim).

preintegration locus. Virus–cell recombinant junctions in one clone of latently infected human cells have been mapped to AAVS1 nt 1026–1030 and 1144–1146 (6); in several other latently infected human cell lines, the recombinant junctions were mapped to the same region, plus or minus 100 nt (7). Hence, the data suggest that the signals directing AAV integration are to the 5' side of the actual recombination crossover points.

A recent discovery of particular interest is that the 5' end of AAVS1 DNA contains a binding site for the AAV *rep* gene products, Rep 68 and Rep 78; this binding site is located within a DNA segment bordered by nt 353–468 (8) within the 510-bp region that directs recombination with AAV (Fig. 3*a*). Rep68/78 has been shown to bind specifically to AAV

FIG. 3. (*a*) Characteristics of the 510 nt at the 5' end of AAVS1 preintegration sequence. The 510 nt are part of a CpG island (nt 1–900) and a region with high frequencies of short repeats (6). Binding elements for upstream binding factor 1 (UBF1) or the reverse complement (CGGCC, GGCCG) and the Sp1 binding sites are indicated by U and S, respectively (6). The cAMP-responsive element (TGACGTCA) is designated by CRE (6). The enlarged part of the sequence shows the Rep binding site (8), a putative terminal resolution site (*trs*), and the M26 motif from yeast (28). (*b*) AAV2 DNA sequence; the nucleotide numbers refer to the complementary strand of the AAV sequence. *trs* is indicated by a vertical arrow (27).

terminal hairpin DNA (10–12, 24–26). The AAVS1 Rep-binding sequence motif, (GCTC)₃GCTG, is similar to the GCGC(GCTC)₃ motif identified in the AAV terminal repeat (8) (Fig. 3). In cell-free reactions, Rep68/78 forms a complex between the AAVS1 DNA segment bordered by nt 353–468 and the AAV genome (8). These findings complement the results reported herein and suggest a model for site-specific integration in which the initial event is the Rep-mediated binding of the AAV termini to the Rep-binding motif on AAVS1 DNA. A putative terminal resolution site (*trs*) (27) 14 bp upstream of the Rep binding site has been identified within the 510-bp fragment, as well as in the M26 motif from yeast, which has been characterized as an enhancer element for recombination (Fig. 3) (28, 29). If Rep were able to nick at the *trs* in AAVS1, this might serve as a step to initiate recombination. That Rep may be able to nick the *trs* sequence is suggested by the report that Rep can initiate replication *in vitro* on a plasmid construct in which is inserted a short stretch (<200 bp) of AAVS1 containing the Rep binding sequence and the *trs* sequence (E. Urcelay *et al.*, personal communication).[†] The model is consistent with the possibility that the AAV Rep68/78 protein may be important for site-specific integration. It is likely that the Rep binding site occurs at more than one site throughout the human genome (as confirmed by database analysis). Therefore, several sequences within the 510-bp fragment of AAVS1 may function in concert to direct site-specific integration.

p220.2, the vector that we used for propagating AAVS1 preintegration DNA outside of its normal chromosome 19q context, replicates as a nuclear episome in the C17 cells. The EBNA-1 gene expressed constitutively, and by the plasmid itself, transactivates the EBV *oriP* sequence necessary in cis for maintenance of the vector as an episome in cells (14, 30–32). EBV vectors can be introduced with high efficiency into mammalian cells and recovered in *E. coli*. The actual copy number per cell is influenced by the amount of DNA initially transfected and has been reported to be in the range 50–100 (14, 33). Each episome replicates once per cell cycle (33) in step with the chromosomal DNA. These vectors are maintained as unrearranged nuclear episomes with inserts up to 35 kb in length (13). In agreement with those reports, we found that p220.2 vectors containing no inserts, inserts of random human DNA, or inserts of AAVS1 DNA that lacked the 510 bp at the 5' end, were stable in C17 cells, as judged by comparing the restriction digestion patterns of the vectors recovered in *E. coli* with those of the parental constructs transfected into C17 cells. Hence, it was unexpected to find that p220.2 vectors with inserts of AAVS1 DNA which included the 5'-terminal 510-bp segment frequently exhibited DNA rearrangements. The same vectors cloned and amplified in the SURE strain of *E. coli* prior to their propagation as episomes in C17 cells did not exhibit DNA rearrangements. Consequently, the genomic rearrangements specifically associated with the presence of the 5' AAVS1 510-bp segment in the vector must have occurred in the C17 cells. The question of whether the rearrangements are clustered in particular regions of the vector DNA is complicated by two selective requirements: (i) the retention of the hygromycin-resistance gene and *oriP* required for propagation in C17 cells (passaged in the constant presence of hygromycin) and (ii) the retention of the pBR322 replication origin and Amp^R gene required for selection in *E. coli*. The cells propagating the 220.2 vectors with inserts of AAVS1 DNA which included the 5'-terminal 510 bp showed no signs of diminished growth over

many passages in culture. Evidently, some of the intracellular episomes in each cell generation must retain the hygromycin-resistance gene and *oriP* in a functionally unaltered form. Conceivably, the events responsible for vector DNA rearrangements occur during each cell cycle but do not affect the entire episomal population.

To further characterize the mechanisms involved in site specific integration, we are analyzing the structure of the recombinants in detail. Preliminary data indicate that AAV integration is 5' to the site of AAVS1 insertion in the p220.2 polylinker but, because the sequences on the 3' side are required for selection in *E. coli* and in C17 cells, we cannot make a definitive statement with regard to the polarity of the recombination.

We thank E. Falck-Pedersen, N. Hackett, C. Leonard, R. M. Linden, and P. Ward for helpful discussions and for critical reading of the manuscript. We thank N. Cortez for excellent technical assistance. This work was supported by Grant AI22251 from the U.S. Public Health Service.

- Berns, K. I. (1990) *Microbiol. Rev.* **54**, 316–329.
- Yakobson, B., Koch, T. & Winocour, E. (1987) *J. Virol.* **61**, 972–981.
- Yalkinoglu, A. O., Heilbronn, R., Burkle, A., Schlehofer, J. R. & zur Hausen, H. (1988) *Cancer Res.* **48**, 3123–3129.
- 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.
- Kotin, R. M., Menninger, J. C., Ward, D. C. & Berns, K. I. (1991) *Genomics* **10**, 831–834.
- Kotin, R. M., Linden, R. M. & Berns, K. I. (1992) *EMBO J.* **11**, 5071–5078.
- Samulski, R. J., Zhu, X., Xiao, X., Brook, J. D., Housman, D. E., Epstein, N. & Hunter, L. A. (1991) *EMBO J.* **10**, 3941–3950.
- Weitzman, M. D., Kyöstö, S. R. M., Kotin, R. M. & Owens, R. A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4808–4812.
- Chiorini, J. A., Weitzman, M. D., Owens, R. A., Urcelay, E., Safer, B. & Kotin, R. M. (1994) *J. Virol.* **68**, 797–804.
- Im, D. S. & Muzyczka, N. (1989) *J. Virol.* **63**, 3095–3104.
- Im, D. S. & Muzyczka, N. (1990) *Cell* **61**, 447–457.
- Im, D. S. & Muzyczka, N. (1992) *J. Virol.* **66**, 1119–1128.
- Margolske, R. F. (1992) *Curr. Top. Microbiol. Immunol.* **158**, 67–95.
- Yates, J. L., Warren, N. & Sugden, B. (1985) *Nature (London)* **313**, 812–815.
- DuBridge, R. B., Tang, P., Hsia, H. C., Leong, P.-M., Miller, J. H. & Calos, M. P. (1987) *Mol. Cell. Biol.* **7**, 379–387.
- Canfield, V., Emanuel, J. R., Spickofsky, N., Levenson, R. & Margolske, R. F. (1990) *Mol. Cell. Biol.* **10**, 1367–1372.
- Graham, F. L. & Van der Eb, A. J. (1973) *Virology* **52**, 456–467.
- 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.
- Stamatatos, L., Leventis, R., Zuckermann, M. J. & Silviu, J. R. (1988) *Biochemistry* **27**, 3917–3925.
- Hirt, B. (1967) *J. Mol. Biol.* **26**, 365–369.
- Greener, A. (1990) *Strategies* **3**, 5–6.
- Swirski, R. A., Van Den Berg, D., Murphy, A. J. M., Lambert, C. M., Friedberg, E. C. & Schimke, R. T. (1992) *Methods Compan. Methods Enzymol.* **4**, 133–142.
- Lewis, E. D. & Manley, J. L. (1985) *Nature (London)* **317**, 172–175.
- Ashktorab, H. & Srivastava, A. (1989) *J. Virol.* **63**, 3034–3039.
- Snyder, R. O., Im, D. S. & Muzyczka, N. (1990) *J. Virol.* **64**, 6204–6213.
- Owens, R. A., Weitzman, M. D., Kyöstö, S. R. M. & Carter, B. J. (1993) *J. Virol.* **67**, 997–1005.
- Snyder, R. O., Im, D. S., Ni, T., Xiao, X., Samulski, R. J. & Muzyczka, N. (1993) *J. Virol.* **67**, 6096–6104.
- Schuchert, P., Langsford, M., Käslein, E. & Kohli, J. (1991) *EMBO J.* **10**, 2157–2163.
- Ponticelli, S. A. & Smith, G. R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 227–231.
- Yates, J. L., Warren, N., Reisman, D. & Sugden, B. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3806–3810.
- Lupton, S. & Levine, A. J. (1985) *Mol. Cell. Biol.* **5**, 2533–2542.
- Su, W., Middleton, T., Sugden, B. & Echols, H. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10870–10874.
- Yates, J. L. & Guan, N. (1991) *J. Virol.* **65**, 483–488.

[†]Urcelay, E., Weitzman, M. D., Safer, B. & Kotin, R. M., Fifth International Parvovirus Workshop, Nov. 10–14, 1993, Crystal River, FL, p. 4.3.