JOURNAL OF VIROLOGY, July 2000, p. 6213–6216 Vol. 74, No. 13
0022-538X/00/$04.00+0

DNA Sequence Motifs Which Direct Adeno-Associated Virus Site-Specific Integration in a Model System

PATRICIO MENESES,[1] KENNETH I. BERNS,[2]* AND ERNEST WINOCOUR[3]

Department of Pathology, Harvard Medical School, Boston, Massachusetts 02215[1]; University of Florida College of Medicine, Gainesville, Florida 32610-0014[2]; and Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel[3]

Received 14 February 2000/Accepted 25 March 2000

The DNA sequence motifs which direct adeno-associated virus type 2 site-specific integration are being investigated using a shuttle vector, propagated as a stable episome in cultured cell lines, as the target for integration. Previously, we reported that the minimum episomal targeting elements comprise a 16-bp binding motif (Rep binding site [RBS]) for a viral regulatory protein (Rep) separated by a short DNA spacer from a sequence (terminal resolution site [TRS]) that can serve as a substrate for Rep-mediated nicking activity (R. M. Linden, P. Ward, C. Giraud, E. Winocour, and K. I. Berns, Proc. Natl. Acad. Sci. USA 93:11288–11294, 1996; R. M. Linden, E. Winocour, and K. I. Berns, Proc. Natl. Acad. Sci. USA 93:7966–7972, 1996). We now report that episomal integration depends upon both the sequence and the position of the spacer DNA separating the RBS and TRS motifs. The spacer thus constitutes a third element required for site-specific episomal integration.

Among the animal viruses, the human parvovirus adeno-associated virus type 2 (AAV2) is unique in its ability to establish latent infection in cell culture by integrating at a specific locus on the q arm of chromosome 19 (the reported specificity ranges from 70 to 100%) (14–16, 22). The AAV2 life cycle is biphasic. When the cell is coinfected with a helper virus, usually an adenovirus or a herpesvirus, AAV2 undergoes a highly efficient cycle of productive, lytic infection. In addition, a low level of helper-independent AAV2 replication can occur in some cells exposed to genotoxic stress (2, 28–30). When no helper virus is present, and the cells are not exposed to genotoxic chemicals or irradiation, AAV2 takes advantage of site-specific chromosomal integration to establish latency. The integrated state persists stably over many cell generations until viral rescue and replication are triggered by an infecting helper virus or other cellular stress conditions (2).

The 4.7-kb linear, single-stranded DNA genome of AAV2 contains two major open reading frames bracketed by 145-nucleotide inverted terminal repeats (ITRs). Due to the presence of palindromic sequences, the ITRs fold into hydrogen-bonded T-shaped hairpinned structures which serve as self-priming origins of replication. The ITRs contain the *cis*-acting replication elements. The *trans*-acting replication proteins (the Rep proteins) are encoded in the left-hand open reading frame, which gives rise to four overlapping polypeptides derived from two promoters by differential splicing (2). The larger Rep 68/78 proteins control major phases of the viral life cycle (2). They resolve an early replication intermediate by binding to the hairpinned stem of the ITR (at the Rep binding site [RBS]) and by introducing a strand- and site-specific nick (at the terminal resolution site [TRS]) (11, 12). The resolution of the hairpinned structure is required for viral DNA synthesis to proceed (21). As discussed below, the Rep 68/78 proteins are also intimately involved in site-specific chromosomal integration (1, 23, 24).

* Corresponding author. Mailing address: University of Florida College of Medicine, 1600 SW Archer Rd., Room H-102, P.O. Box 100014, Gainesville, FL 32610-0014. Phone: (352) 392-2761. Fax: (352) 392-9395. E-mail: kberns@vpha.ufl.edu.

The site of AAV2 integration in chromosome 19 (the preintegration locus is known as AAVS1) has been cloned and sequenced (14). To identify the chromosomal DNA sequences which direct AAV2 to its preferential integration site, AAVS1 DNA has been subcloned in the Epstein-Barr virus-based shuttle vector p220.2 (9). This vector propagates as a stable extrachromosomal episome, at a copy number of 50 to 100, in human 293 cells expressing the Epstein-Barr virus EBNA1 protein (19, 31). p220.2 contains a hygromycin resistance gene for selection of cells propagating the episome and an *Escherichia coli* replicon and ampicillin resistance gene for rescue in bacteria. After the episome-propagating cell line is established, it is infected with AAV2 and integration into the episome carrying AAVS1 DNA is assessed by rescue in *E. coli*. Hybridization is used to identify ampicillin-resistant colonies containing plasmids with inserts of AAV2 DNA (9). Sequencing of the plasmid inserts has defined junctions with AAVS1 DNA and some organizational features reminiscent of those identified when integration occurred at the chromosomal level (10).

Using the above-described assay for site-directed integration, the minimum chromosome 19 targeting cassette has been identified as a 16-bp canonical RBS motif, similar to that in the viral ITR, separated by a short spacer DNA from a 6-bp sequence that resembles the viral ITR TRS sequence cleaved by Rep 68/78 (18) (Table 1). Data from cell-free reactions have indicated that Rep 68/78 can form a bound complex comprising both the AAVS1 and viral RBS and TRS sequences (26) and that Rep 68/78 initiates DNA synthesis on a plasmid carrying AAVS1 DNA, suggesting that the bound Rep proteins cleave the chromosomal TRS sequence (25). Recent data from a cell-free integration assay have also highlighted the essential roles of the chromosomal and viral RBS and TRS motifs in the formation of AAV2/AAVS1 recombinant junctions (8).

The present model for site-specific integration (6, 17, 18) proposes that a multimeric form of Rep 68/78, by binding to the RBS motifs on the chromosomal and viral DNAs, positions the incoming infecting genome at the AAVS1 site. Rep-mediated nicking at the adjacent chromosomal TRS sequence, possibly aided by cellular accessory proteins of the HMG1 family (7, 8), mobilizes a polymerase-Rep complex that initiates dis-

TABLE 1. Alignment (5′ to 3′) of TRS, spacer, and RBS motifs in viral ITR DNA and in human AAVS1 DNA

DNA	TRS	Spacer	RBS
AAV2 ITR[a]	AGTTGG	CCACTCCCTCTCT	GCGC GCTC GCTC GCTC
Human AAVS1[b]	GGTTGG	GGCTCGGC	GCTC GCTC GCTC GCTG

[a] The minimum AAV2 TRS sequence for efficient Rep cleavage has recently been described by Brister and Muzyczka (3) as the 7-base core 5′-GTTGGCC-3′. Hence, the 5′ CC dinucleotide of the spacer shown may be part of the TRS motif.
[b] Data are from references 26, 25, and 18.

placement DNA synthesis. As DNA synthesis at the AAVS1 site proceeds, a series of DNA template switches (copy choice) links the viral DNA to the chromosomal DNA. Virus-cell junctions in chromosome 19 are clustered at various distances downstream of the targeting RBS and TRS motifs, and integration is accompanied by rearrangements of the AAVS1 site (13, 16, 20, 22). These features of chromosomal integration can be accounted for in the above-described model if we assume that the order and timing of the template switches (from cellular DNA to viral DNA and back) differ in each integration event.

Although related, the AAVS1 RBS and TRS sequence motifs are not identical to those in the viral ITR (Table 1). Furthermore, the intervening sequence between the AAVS1 RBS and TRS motifs (henceforth called the spacer) is different both in sequence and in size from that of the viral origin. In this study, we have used the functional episome integration test to access the importance of the spacer. We show that integration depends upon both spacer sequence and position.

Oligonucleotides containing the same RBS and TRS motifs derived from the chromosome 19 AAVS1 site, but differing in the intervening spacer sequences, were synthesized and inserted into the p220.2 shuttle vector. Cell lines stably propagating the shuttle vector as an episome were infected with AAV, and integration into the episome was assessed by rescue in *E. coli*. The proportion of colonies hybridizing with an AAV2 DNA probe provided a measure of the episomal integration frequency (9). Each episomal target for AAV2 integration is designated in Tables 2 and 3 by the name of the cell line carrying that episome.

The insert in episome 83 (Table 2) contains the 13-nucleotide sequence which bridges the RBS and TRS elements in the viral ITRs. In episome 102, the 8-nucleotide spacer is that present in the chromosome 19 AAVS1 locus. Episomes 83 and 102 target AAV integration to the same extent (0.35 and 0.33% AAV-positive colonies, respectively). Episomes 80 and 81 also target AAV2 integration to similar extents, indicating that a reduction in spacer length from 13 to 6 nucleotides does not radically affect the frequency of integration.

Since the central CTC triplet was common to the spacer sequence in episomes 80, 81, 83, and 102, we next turned our attention to this potential motif. Replacement of CTC by the triplet TTA (episome 82) or GGG (episomes 86 and P-54/55) reduced the frequency of integration 10-fold. Retention of the CTC triplet but alteration of the adjacent 5′ and 3′ dinucleotides resulted in a fivefold decrease in targeting activity (compare the activity of episome 102 to that of 87 or P-58/59). From these data, it appears that the central CTC triplet in the spacer, although necessary, is not by itself sufficient for full targeting activity and may thus be part of a larger motif. One candidate would be the GCTC motif, which is present in the spacers of the efficiently targeting episomes 102, 80, and 81 and which is also the repeating motif in the RBS site (5). However, the GCTC motif is not present in the spacer of episome 83, whose targeting activity is comparable to those of episomes 80, 81, and 102 and whose spacer sequence is that of the AAV2 ITR. Conceivably, it is the two CTC motifs embedded in a pyrimidine-rich tract in the episome 83 spacer which is responsible for full targeting activity. A pyrimidine-rich sequence between the TRS and RBS elements is a common feature of AAV serotypes, including AAV5, whose TRS cleavage specificity differs from that of AAV2 (4).

The importance of the GCTC motif in the AAVS1 spacer is also highlighted by the results with episomes 201, 202, and 203 (Table 3). Deletion of the entire GG CTC GGC spacer sequence reduces episomal integration 35-fold (episome 201). Functionality also depends upon the distance of the GG CTC GGC sequence from both the TRS and RBS sites: insertion of

TABLE 2. DNA sequences which target AAV2 episomal integration[a]

Episome carried by cell line	Sequence between TRS and RBS in p220.2 inserts	AAV-positive colonies (integration)	
		No./total	%
83	TRS-CCA CTC CCTCTCT -RBS	127/36,558	0.35
102	TRS- GG CTC GGC -RBS	119/36,300	0.33
80	TRS- G CTC GGC -RBS	209/33,630	0.62
81	TRS- G CTC GC -RBS	202/46,000	0.44
82	TRS- CA TTA GTC -RBS	17/56,950	0.03
86	TRS- GG GGG GGC -RBS	12/42,000	0.03
P-54/55	TRS- GG GGG GGC -RBS	17/34,728	0.05
87	TRS- CA CTC GTC -RBS	32/57,000	0.06
P-58/59	TRS- CA CTC GTC -RBS	16/24,354	0.07

[a] Cell lines stably propagating p220.2 episomes were established by hygromycin selection (9). The inserts in p220.2 contain the spacer DNA sequence shown (5′ to 3′) linked on one side to the AAVS1 TRS (GGTTGG) and on the other side to the AAVS1 RBS (GCTC GCTC GCTC GCTG) (Table 1). All inserts contain the sequence CTAGAGGC 5′ to the TRS. Episome 83 contains the spacer sequence of the viral ITR. Episome 102 contains the spacer sequence of AAVSI. Forty-eight hours after the cell lines were infected with AAV (multiplicity of 50 infectious units/cell), viral integration into the episomes was assessed by rescue in *E. coli* and colony hybridization with an AAV DNA probe. The integration data are derived from two infections of the same cell line at different passage levels. In independent infections of the same cell line but at different passage levels, the integration frequency varied by a maximum factor of two. The frequencies shown were calculated from the total number of colonies screened in both trials. Independently generated cell lines carrying the same episomes are 86 and P-54/55, 87, and P-58/59. The nature of the episome in each cell line was confirmed by rescue of the shuttle vector in *E. coli* and sequencing of the inserts.

TABLE 3. Deletion of the spacer sequence or alteration of the distance of the spacer from the TRS and RBS motifs abrogates episomal integration[a]

Episome carried by cell line	Sequence between TRS and RBS in p220.2 inserts	AAV-positive colonies (integration)	
		No./total	%
102	TRS- GG CTC GGC-RBS	56/40,000	0.140
201	TRS- -RBS	4/90,500	0.004
202	TRS-50n+GG CTC GGC-RBS	2/74,700	0.003
203	TRS-GG CTC GGC+50n-RBS	1/69,240	0.001

[a] In episome 102, the spacer sequence is that of AAVS1; in episome 201, the spacer is deleted; in episome 202, the spacer sequence is separated from the TRS by 50 random nucleotides (50n); and in episome 203, the spacer sequence is separated from the RBS by 50 random nucleotides. Remaining details are as described in Table 2, footnote *a*.

50 random nucleotides at either end essentially abolishes integration (episomes 202 and 203). (The absolute distance between TRS and RBS sites may also be a factor.)

The above-described experiments show that the spacer sequence separating the TRS from the RBS is critical for episomal integration. Either a GCTC motif or two CTC triplets embedded in a pyrimidine-rich sequence were found to provide an optimum integration frequency. It is noteworthy that the GCTC motif is also the main repeat of the RBS consensus. Positioning also plays a role; increasing the distance between the AAVS1 spacer and the TRS or RBS motifs dramatically reduces integration. The AAVS1 spacer may thus be viewed as an extension of the canonical RBS consensus in that it adds an extra GCTC motif. The minimum targeting sequence for episomal integration can now be reduced to 30 bp: the 16-bp RBS separated by GCTCGC from the 6-bp TRS (Table 2, episome 81).

The step in integration blocked by changes in the spacer is unknown. Four of the spacers used were checked for the ability to bind Rep and to be nicked. These were spacers in episomes 102, 81, 83, and P-54/55 (Table 2). AAV was able to integrate into plasmids containing the first three but not into the last. All four of the spacers bound Rep, and all four were nicked to an extent which was within a factor of two of that observed for the AAVS1 sequence (spacer 102). The actual ratios were 0.64 for episome 81, 0.45 for episome P-54/55, and 1.93 for episome 83, which contained the spacer in the AAV ITR. Thus, Rep nicking showed greater tolerance with respect to spacer sequence than did integration in the model system.

The human genome contains numerous binding sites for the AAV2 Rep proteins, as judged by data bank inspection and by biochemical tests (27). However, when an appropriately positioned AAV2 TRS motif is included in the search parameters, the available data bank reveals only a single site for targeted AAV2 integration, consistent with biological experiments at the chromosomal level (15, 16). The episomal integration assay can provide additional parameters to search for other potential integration sites that might be exploited by different AAV serotypes in different hosts. The present data indicate that the spacer sequence and position should be included in these parameters.

We thank Nenita Cortez and Bernard Danovitch for excellent technical assistance.

This work was supported by Public Health Service grants AI 122251 ("Molecular Biology of Adeno-Associated Virus") and GM50032 ("Regulation of AAV DNA Replication").

REFERENCES

1. **Balague, C., M. Kalla, and W.-W. Zhang.** 1997. Adeno-associated virus Rep78 protein and terminal repeats enhance integration of DNA sequences into the cellular genome. J. Virol. **71:**3299–3306.
2. **Berns, K. I.** 1996. Parvoviridae: the viruses and their replication, p. 2173–2198. *In* B. N. Fields (ed.), Virology, 3rd ed. Lippincot-Raven, New York, N.Y.
3. **Brister, J. R., and N. Muzyczka.** 1999. Rep-mediated nicking of the adeno-associated virus origin requires two biochemical activities, DNA helicase activity and transesterification. J. Virol. **73:**9325–9336.
4. **Chiorini, J. A., S. Afione, and R. M. Kotin.** 1999. Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. J. Virol. **73:**4293–4298.
5. **Chiorini, J. A., L. Yang, B. Safer, and R. M. Kotin.** 1995. Determination of adeno-associated virus Rep68 and Rep78 binding sites by random sequence oligonucleotide selection. J. Virol. **68:**7334–7338.
6. **Chiorini, J. A., S. M. Wiener, L. Yang, R. H. Smith, B. Safer, N. P. Kilcoin, Y. Liu, E. Urcelay, and R. M. Kotin.** 1996. The role of AAV Rep proteins in gene expression and targeted integration. Curr. Top. Microbiol. Immunol. **218:**25–33.
7. **Costello, E., P. Saudan, E. Winocour, L. Pizer, and P. Beard.** 1997. High mobility group protein 1 binds to the adeno-associated virus replication protein (Rep) and promotes Rep-mediated site-specific cleavage of DNA, ATPase activity and transcriptional repression. EMBO J. **16:**5843–5954.
8. **Dyall, J., P. Szabo, and K. I. Berns.** 1999. Adeno-associated virus (AAV) site-specific integration: formation of AAV-AAVS1 junctions in an *in vitro* system. Proc. Natl. Acad. Sci. USA **92:**12849–12854.
9. **Giraud, C., E. Winocour, and K. I. Berns.** 1994. Site specific integration by adeno-associated virus is detected by a cellular sequence. Proc. Natl. Acad. Sci. USA **91:**10039–10043.
10. **Giraud, C., E. Winocour, and K. I. Berns.** 1995. Recombinant junctions formed by site-specific integration of adeno-associated virus into an episome. J. Virol. **69:**6917–6924.
11. **Im, D.-S., and N. Muzyczka.** 1990. The AAV origin binding protein Rep68 is an ATP dependent site-specific endonuclease with DNA helicase activity. Cell **61:**447–457.
12. **Im, D.-S., and N. Muzyczka.** 1992. Partial purification of adeno-associated virus Rep78, Rep68, Rep52, and Rep40 and their biochemical characterization. J. Virol. **66:**1119–1128.
13. **Kotin, R. M., and K. I. Berns.** 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. Virology **170:**460–467.
14. **Kotin, R. M., R. M. Linden, and K. I. Berns.** 1992. Characterization of a preferred site on chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. **11:**5071–5078.
15. **Kotin, R. M., J. C. Menninger, D. C. Ward, and K. I. Berns.** 1991. Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. Genomics **10:**831–834.
16. **Kotin, R. M., M. Siniscalco, R. J. Samulski, X. D. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns.** 1990. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA **87:**2211–2215.
17. **Linden, R. M., P. Ward, C. Giraud, E. Winocour, and K. I. Berns.** 1996. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. USA **93:**11288–11294.
18. **Linden, R. M., E. Winocour, and K. I. Berns.** 1996. The recombination signals for adeno-associated virus site-specific integration. Proc. Natl. Acad. Sci. USA **93:**7966–7972.
19. **Margolskee, R. F.** 1992. Epstein-Barr virus based expression vectors. Curr. Top. Microbiol. Immunol. **158:**67–95.
20. **McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka.** 1988. Adeno-associated virus general transduction vectors: analysis of proviral structures. J. Virol. **62:**1963–1973.
21. **Muzyczka, N.** 1991. *In vitro* replication of adeno-associated virus DNA. Semin. Virol. **2:**281–290.
22. **Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter.** 1991. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J. **10:**3941–3950.
23. **Shelling, A. N., and M. G. Smith.** 1994. Targeted integration of transfected and infected adeno-associated virus vectors containing the neomycin resistance gene. Gene Ther. **1:**165–169.
24. **Surosky, R. T., M. Urabe, S. K. Godwin, S. A. McQuiston, G. J. Kurtzman, K. Ozawa, and G. Natsoulis.** 1997. Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. J. Virol. **71:**7951–7959.
25. **Urcelay, E., P. Ward, S. M. Wiener, B. Safer, and R. M. Kotin.** 1995. Asymmetric replication in vitro from a human sequence element is dependent on adeno-associated virus Rep protein. J. Virol. **69:**2038–2046.
26. **Weitzman, M. D., S. R. M. Kyostio, R. M. Kotin, and R. A. Owens.** 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.
27. **Wonderling, R. S., and R. A. Owens.** 1997. Binding sites for adeno-associated virus Rep proteins within the human genome. J. Virol. **71:**2528–2534.

28. **Yakobson, B., T. Koch, and E. Winocour.** 1987. Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. J. Virol. **61:**972–981.
29. **Yakobson, B., T. A. Hrynko, M. J. Peak, and E. Winocour.** 1989. Replication of adeno-associated virus in cells irradiated with UV light at 254 nm. J. Virol. **63:**1023–1030.
30. **Yalkinoglu, A. O., R. Heilbronn, A. Burkle, J. R. Schlehhofer, and H. zur Hausen.** 1988. DNA amplification of adeno-associated virus as a response to cellular genotoxic stress. Cancer Res. **48:**3123–3129.
31. **Yates, J. L., N. Warren, and B. Sugden.** 1985. Stable replication of plasmids derived from Epstein-Barr virus in various mammalian cells. Nature (London) **313:**812–815.