JOURNAL OF VIROLOGY, Mar. 1997, p. 2528–2534 Vol. 71, No. 3
0022-538X/97/$04.00+0

Binding Sites for Adeno-Associated Virus Rep Proteins within the Human Genome

RAMANI S. WONDERLING AND ROLAND A. OWENS*

Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received 6 September 1996/Accepted 22 November 1996

The Rep proteins of adeno-associated virus type 2 (AAV) are known to bind to Rep recognition sequences (RRSs) in the AAV inverted terminal repeats (ITRs), the AAV p5 promoter, and the preferred AAV integration site in human chromosome 19, called AAVS1. Integration of the AAV genome into AAVS1 appears to be mediated by an interaction between the Rep proteins of AAV and Rep binding sites within the viral genome and the integration locus. In an attempt to identify potential alternate integration sites, we looked for recognition sites for AAV Rep proteins in the human genome by performing a BLASTN computerized homology search. We used the 16-mer core sequences of the RRSs in the AAV ITRs and AAVS1 separately as query sequences and identified 18 new RRSs in or flanking the genes coding for the following: tyrosine kinase activator protein 1 (TKA-1); colony stimulating factor-1; insulin-like growth factor binding protein 2 (IGFBP-2); histone H2B.1; basement membrane heparan sulfate proteoglycan, also known as perlecan; the *AF-9* gene product, which is involved in the chromosomal translocation t (9:11)(p22:q23); the β_B subunit of the hormone known as inhibin; interleukin-2 enhancer binding factor; an endoplasmic reticulum-Golgi intermediate compartment resident protein called p63; a global transcription activator (hSNF2L); the β-actin repair domain; a retinoic acid-inducible factor, also known as midkine; a breast tumor autoantigen; a growth-arrest- and DNA-damage-inducible protein called gadd45; the cyclin-dependent kinase inhibitor called KIP2, which inhibits several G_1 cyclin–cyclin-dependent kinase complexes; and the hereditary breast and ovarian cancer gene (*BRCA1*). RRSs were also identified in a newly discovered open reading frame on chromosome 10 and in the ERCC1 locus on human chromosome 19. The ability of a maltose binding protein-Rep68 fusion protein to bind to these sequences was confirmed by electrophoretic mobility shift assays. These sites may serve as alternate integration sites for AAV or play a role in Rep-mediated effects on human cells.

Adeno-associated virus type 2 (AAV) is a nonpathogenic human parvovirus. It belongs to the dependovirus subfamily because it depends on helper viruses for efficient replication (4, 44). In the absence of helper virus, the AAV genome can integrate stably into human chromosomal DNA. Sequence analysis of the human DNA flanking the integrated AAV genome in latently infected cells revealed that there was a specific integration locus for AAV in 60 to 70% of the cases (28, 43). The integration locus, designated AAVS1, has been localized to human chromosome 19q13.3-qter (25–27). Its nonpathogenic nature, its ability to preferentially integrate into a specific locus within the human genome (26, 43), and the ability to produce high titers of the virus are some of the attractive features of AAV as a gene therapy vector.

The Rep proteins of AAV, Rep68 and Rep78 in particular, are multifunctional proteins. Their activities include DNA binding (20), site- and strand-specific endonuclease (19), DNA-DNA helicase (19), RNA-DNA helicase (53), and ATPase activities (53). The Rep68 and Rep78 proteins also play a role in gene regulation. These proteins bind to the Rep recognition sequence (RRS) in the AAV p_5 promoter and down-regulate the *rep* gene (29). Rep proteins also down-regulate c-*fos*, c-*myc* (16), and c-H-*ras* gene expression (17). An RRS has been identified in or near most genes regulated by Rep proteins. We have shown recently that Rep68 binds to an RRS in the c-*sis* proto-oncogene and also enhances the expression of the c-*sis* proto-oncogene (54). We have also identified other RRSs in human genes coding for a hepatocyte glucose transporter, α-A-crystallin, and the carcinoma marker GA733-1 (54).

Only the inverted terminal repeats (ITRs) of the AAV genome are required in *cis* for replication and packaging. AAV vectors lacking the *rep* gene failed to preferentially integrate into AAVS1 (47, 55). However, AAV vectors that carry an intact *rep* gene along with the ITRs are capable of integrating into AAVS1, indicating a role for Rep proteins in this site-specific integration (46). An RRS has been identified within AAVS1 (52). Furthermore, Rep proteins have been shown to mediate the formation of a complex between the AAV ITRs and a 109-bp region of AAVS1 which contains the RRS (52). Giraud et al. have shown that this same region of AAVS1 is sufficient to direct the integration of AAV into an episome in vivo (13, 14). Thus, the preferential integration of the AAV genome into AAVS1 appears to be mediated by an interaction between Rep68 or Rep78 and RRSs within the virus and AAVS1. This segment of AAVS1 also contains an endonuclease site for the Rep proteins which may also play a role in Rep-mediated preferential integration into AAVS1 (50).

In an attempt to identify potential alternate, Rep-dependent integration sites within the human genome, a BLASTN computerized homology search (2) was performed with the 16-mer core sequence of the RRS in the AAV ITR (AAV-ITR-RRS) (5′-GAGCGAGCGAGCGCGC-3′) and the 16-mer core sequence in the RRS in AAVS1 (AAVS1-RRS) (5′-CAGCGA GCGAGCGAGC-3′), which were used separately as query sequences. The databases searched were GenBank, EMBL, DDBJ, and PDB, and the search parameters set were as fol-

* Corresponding author. Mailing address: Laboratory of Molecular and Cellular Biology, NIDDK, National Institutes of Health, Bldg. 8, Rm. 309, 8 Center Dr., MSC 0840, Bethesda, MD 20892-0840. Phone: (301) 496-3359. Fax: (301) 402-0053. E-mail: rolando@bdg8.niddk.nih.gov.

					MSP Score
AAV-ITR	GAGC	GAGC	GAGC	GCGC	80
TKA-1	GAGC	GAGC	GAGC	GCGC	80
CSF-1	GAGC	GAGC	GAGC	GaGC	71
IGFBP-2	GAGC	GAGC	GAGC	GgGC	71
Histone H2B.1	GAGC	GAGC	GAGC	GCca	70
Perlecan	GAGC	GAGC	GAGC	GaGa	66
AF-9	aAGC	GAGC	GAGC	GaGC	66
AAVS1	cAGC	GAGC	GAGC	GaGC	66
Inhibin	GAGC	GAGC	GAGC	GgGg	66
ILF	GAGC	GAGC	GAGC	Ggcg	65
p63	agGC	GAGC	GAGC	GCGg	65
hSNF2L	GAGC	GAGC	GAGC	Gtag	65
β-actin	GAGC	GAGC	GAGC	Gtct	65
AAVS1	CAGC	GAGC	GAGC	GAGC	80
CSF-1	CAGC	GAGC	GAGC	GAGC	80
CSF-1	gAGC	GAGC	GAGC	GAGC	75
AF-9	aAGC	GAGC	GAGC	GAGC	75
AAV-ITR	CAGt	GAGC	GAGC	GAGC	71
CSF-1	CAGC	cAGC	GAGC	GAGC	71
ILF	CgGC	GAGC	GAGC	GAGC	71
MK	CAGC	GAGC	GAGt	GAGC	71
CSF-1	gAGC	GAGC	GAGC	GAGg	70
Perlecan	gAGC	GAGC	GAGC	GAGa	70
IGFBP-2	ggGC	GAGC	GAGC	GAGC	70
IGFBP-2	gAGC	GAGC	GAGC	GgGC	66
TKA-1	gAGC	GAGC	GAGC	GcGC	66
BTAA	gAGC	aAGC	GAGC	GAGC	66
gadd45	gAGC	GAGC	GAGC	aAGC	66
KIP2	gAGC	GAGC	GAGC	tAGC	66
AAV-ITR	gAGC	GAGC	GAGC	GcGC	66
BRCA1	gAGg	GAGC	GAGC	GAGC	66
hSNF2L	gAGa	GAGC	GAGC	GAGC	66
ORF	gAGa	GAGC	GAGC	GAGC	66
ERCC1	gcGC	GAGC	GAGC	GAGg	65
KIP2	CctC	GAGC	GAGC	GAGC	65

FIG. 1. Comparison of 16-mer core sequences of the newly identified RRSs in the human genome with their respective query sequences. The query sequence in each case is enclosed in a box. Nucleotides in the newly identified RRSs that are different from nucleotides in the query sequences are indicated by lowercase letters. Some loci contained multiple, overlapping RRSs. The maximal segment pair (MSP) score for each new RRS is indicated to the right of the sequence. BTAA stands for breast tumor autoantigen, and ORF refers to a newly discovered open reading frame in chromosome 10.

lows: expectancy, 1,000; cutoff, 60; matrix, default; strand, both; and filters, none. These parameters were particularly chosen so that each query sequence would be picked up as being similar to the other, as well as to a few additional sequences with slightly less similarity. The BLASTN search results indicated that there were several sequences within the human genes with sequence similarity to the AAV-ITR-RRS or to the AAVS1-RRS. Sequences similar to the query sequences were found in the following: the genes coding for tyrosine kinase activator protein 1 (TKA-1), which selectively activates the platelet-derived growth factor receptor signaling potential (GenBank accession number [GAN] Z50150); colony stimulating factor-1 (CSF-1) (GAN M11038); insulin-like growth factor binding protein 2 (IGFBP-2) (GAN M69237); histone H2B.1 (GAN M60756); basement membrane heparan sulfate proteoglycan (HSPG2), also known as perlecan (GAN L22078); the *AF-9* gene, which is involved in the chromosomal translocation t (9:11)(p22:q23) (GAN L13744); the genes encoding the β_B subunit of the hormone known as inhibin (GAN M31668); interleukin-2 enhancer binding factor (ILF) (GAN U58198); an endoplasmic reticulum-Golgi intermediate compartment resident protein called p63 (GAN X69910); a global transcription activator (hSNF2L) (GAN M88163); the β-actin repair domain (GAN U19758); the retinoic acid-inducible factor also known as midkine (MK) (GAN M94250); a breast tumor autoantigen (GAN U24576); a growth-arrest- and DNA-damage-inducible protein called gadd45 (GAN M60974); the

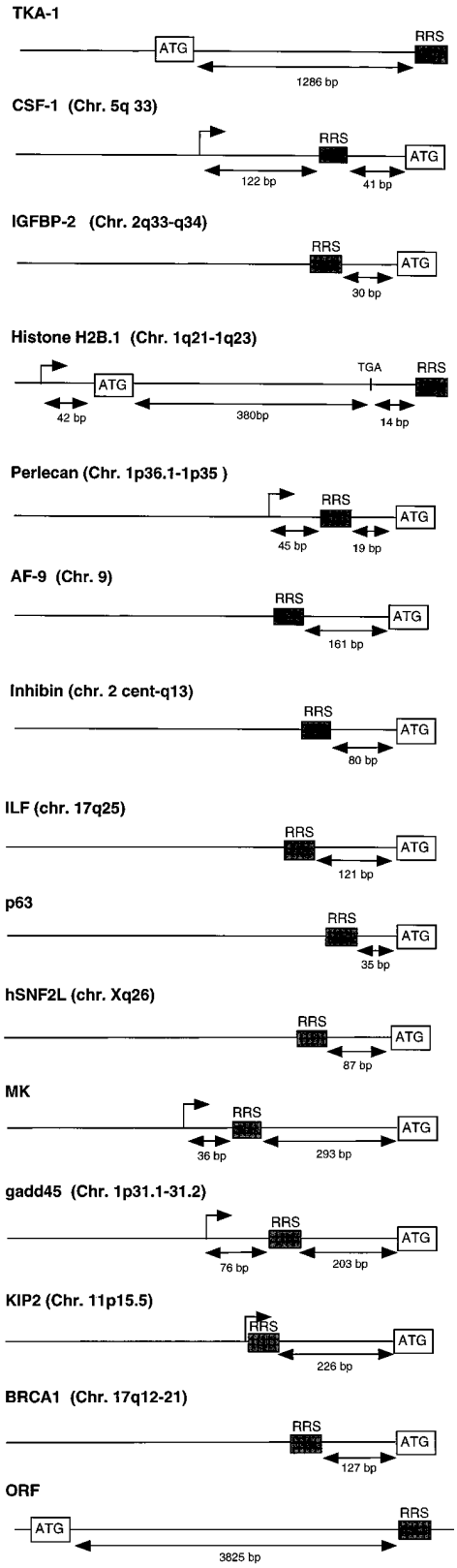

FIG. 2. The locations of the newly identified RRSs relative to their respective genes. ATG and TGA indicate the beginning and end of the coding region, respectively, and the bent arrows indicate transcription start sites. In some cases the transcription start site is not shown, because the sequences were derived from cDNA clones. RRSs are indicated by shaded rectangles. The chromosomal locations of the genes, if known, are also indicated. The figure is not drawn to scale. Chr., chromosome.

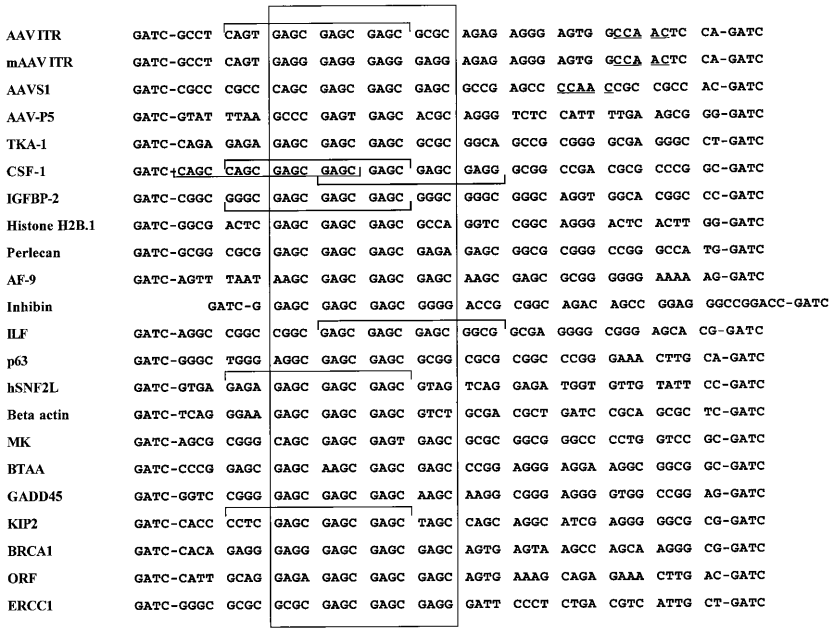

FIG. 3. DNA substrates used in the EMSAs. The sequence of one of the strands (5′ to 3′) of each double-stranded DNA substrate is shown, and the core 16-mer homologs present in these substrates are enclosed in the box. The bases complementary to the consensus cleavage site for the Rep-mediated endonuclease activity have been underlined in both the AAV-ITR and AAVS1 sequences. The GATC sequences at either end of each substrate are derived from the *Bam*HI overhangs incorporated into each oligonucleotide. These were introduced at either end of all the sequences to facilitate cloning if necessary. Additional overlapping RRSs present in some of the substrates are indicated by large brackets. No additional 5′ flanking sequence was available for the inhibin gene. BTAA, breast tumor autoantigen.

cyclin-dependent kinase inhibitor called KIP2, which inhibits several G_1 cyclin–cyclin-dependent kinase complexes (GAN U22398); the hereditary breast and ovarian cancer gene (*BRCA1*) (GAN U11292); a newly sequenced open reading frame on chromosome 10 (which is referred to as ORF in this report) (GAN D38555); and the ERCC1 locus located on chromosome 19 (34) (GAN M89651). The sequence homology between each of these newly identified putative RRSs and the RRS that was used as the query sequence is shown in Fig. 1 along with its maximal segment pair score, which is a measure of the similarity of a particular sequence when compared to the query sequence. It is interesting that several of these genes contained multiple overlapping, putative RRSs, as does the AAV-ITR.

The locations of these putative RRSs in their respective genes along with their chromosomal locations (if known) are shown in Fig. 2. The putative RRSs that were identified by the BLASTN search in IGFBP-2 (11), CSF-1 (23, 30), MK (12, 49), gadd45 (18, 41), KIP2 (36, 48), perlecan (8, 22), *AF-9* (37), ILF (31, 32), *BRCA1* (1), hSNF2L (39), inhibin (35), and the p63 genes (45) are located upstream of the initiator ATG codon. However, the putative RRS in H2B.1 (9, 10) is located at the 3′ end of the message, 14 bp after the stop codon. In TKA-1 and ORF (38) the putative RRSs are located in the coding regions 1,286 bp and 3,825 bp downstream of the initiator ATG codon, respectively.

In order to determine if the putative RRSs identified by the BLASTN search were actual binding sites for AAV Rep proteins, 54-mer double-stranded oligonucleotides (synthesized by Midland Certified Reagent Company, Midland, Tex.) containing the putative RRSs from the different human genes along with their respective flanking regions and *Bam*HI overhangs (shown in Fig. 3) were radiolabeled with Klenow polymerase and [α-^{32}P]dATP and electrophoretic mobility shift assays (EMSAs) were performed with these DNA substrates and a Rep68-maltose binding protein fusion protein (MBP-Rep68Δ) (5). A β-galactosidase-maltose binding protein fusion protein (MBP-LacZ) served as a negative control. A mutant Rep protein, namely, MBP-Rep68ΔNTP (5), which has a lysine-to-histidine substitution at amino acid 340 that has been shown not to affect DNA binding activity (40), was also tested for its ability to bind these putative RRSs. Oligonucleotides 54 bp in length were used in the EMSAs, because previous reports have shown that flanking DNAs in addition to the core sequence are important for Rep protein binding (6, 29, 42). The EMSAs revealed that MBP-Rep68Δ, as well as the mutant MBP-Rep68ΔNTP, bound to all of the linear DNA substrates containing the RRSs identified by the BLAST homology search (Fig. 4A to E). The binding of the Rep proteins to the 54-bp linear DNA substrate containing the AAVS1-RRS was weaker than the binding to the other substrates. Autoradiographs from overnight exposures clearly showed that MBP-Rep68Δ and MBP-Rep68ΔNTP did bind the linear AAVS1 substrate. However, no binding was detected with a mutant version of the AAV-ITR-RRS (mAAV-ITR-RRS) in which the key cytosines within the core binding sequence had been changed to guanines, even in autoradiographs from 10-day exposures (Fig. 4A to E).

Next, we wanted to determine if binding of Rep proteins to the substrates containing the newly identified RRSs was specific. We did preliminary binding-competition assays using radiolabeled AAV-ITR-RRS along with 10-, 40-, 160-, and 640-fold molar excesses of either unlabeled AAV-ITR-RRS or mAAV-ITR-RRS as substrates. As seen in Fig. 5A, a 160-fold molar excess of unlabeled AAV-ITR nearly abolished the binding of Rep proteins to the radiolabeled AAV-ITR-RRS whereas a 160-fold excess of unlabeled mAAV-ITR-RRS had little or no effect. Then the competition assays were done with all of the other radiolabeled substrates along with a 160-fold excess of either unlabeled AAV-ITR-RRS or mAAV-ITR-

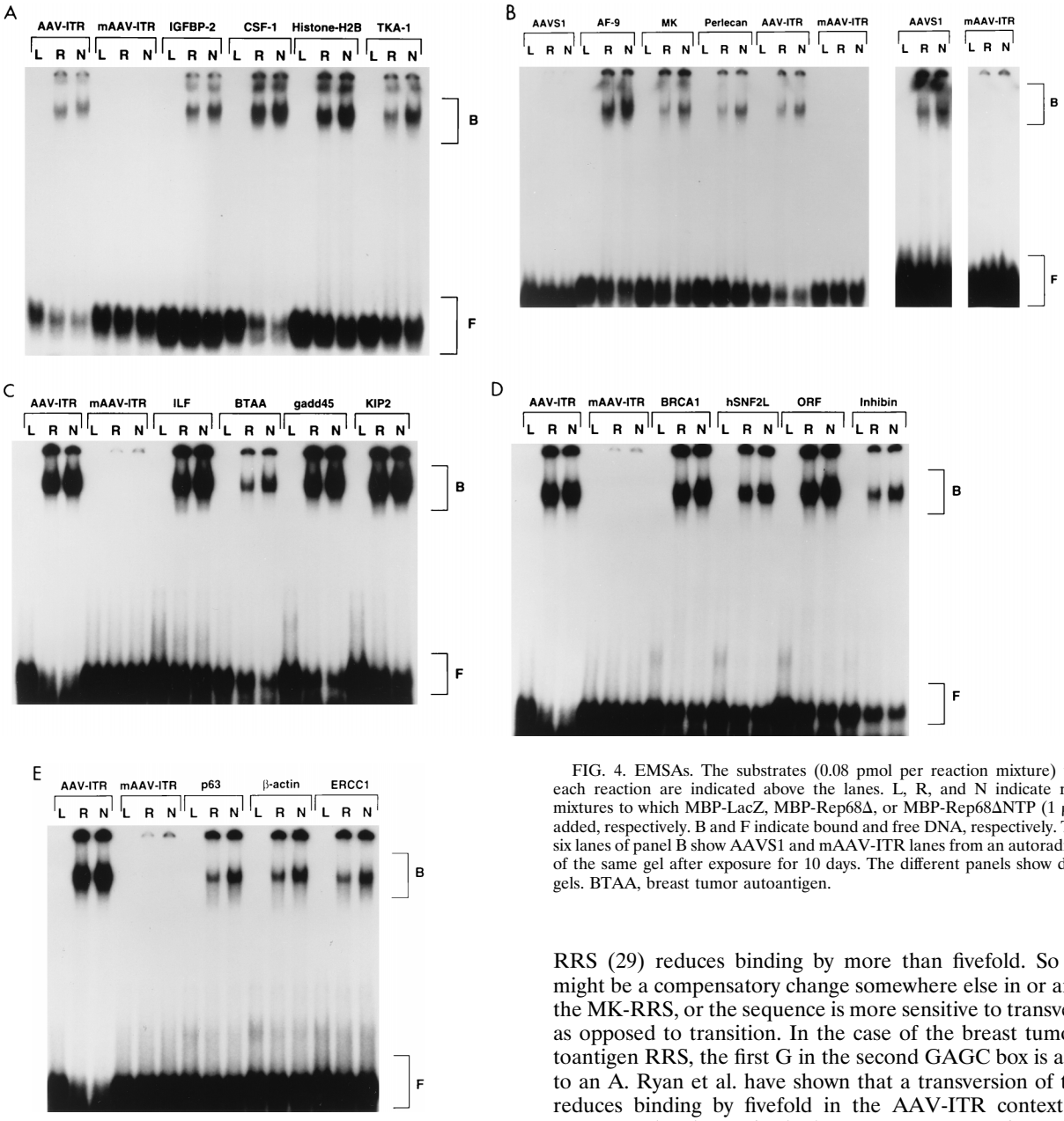

FIG. 4. EMSAs. The substrates (0.08 pmol per reaction mixture) used in each reaction are indicated above the lanes. L, R, and N indicate reaction mixtures to which MBP-LacZ, MBP-Rep68Δ, or MBP-Rep68ΔNTP (1 μg) was added, respectively. B and F indicate bound and free DNA, respectively. The last six lanes of panel B show AAVS1 and mAAV-ITR lanes from an autoradiograph of the same gel after exposure for 10 days. The different panels show different gels. BTAA, breast tumor autoantigen.

RRS. When unlabeled AAV-ITR-RRS was present in 160-fold molar excess, the binding of the Rep proteins to all the substrates was greatly reduced, indicating that the binding to all the substrates was specific (Fig. 5B to F). However, unlabeled mAAV-ITR-RRS failed to compete for the binding of the Rep proteins.

All of the previously identified RRSs contain imperfect $(GAGC)_4$ motifs. Ryan et al. had shown that changes in the middle two GAGC boxes of the 16-mer core AAV-ITR-RRS were most likely to inhibit binding of Rep proteins (42). With the exception of the MK and breast tumor autoantigen RRSs, all the core sequences in our newly identified RRSs have the middle two GAGC boxes intact (Fig. 3). In the MK-RRS, a C in the third GAGC box is altered to a T. Previous studies have shown that mutation of this C to a G within the context of an AAV-ITR-RRS (42) or within the context of the AAV-p5-RRS (29) reduces binding by more than fivefold. So there might be a compensatory change somewhere else in or around the MK-RRS, or the sequence is more sensitive to transversion as opposed to transition. In the case of the breast tumor autoantigen RRS, the first G in the second GAGC box is altered to an A. Ryan et al. have shown that a transversion of this G reduces binding by fivefold in the AAV-ITR context (42). However, the change in the breast tumor autoantigen RRS is only a transition and therefore the change might not be as detrimental as a transversion. The fact that this RRS still has three perfect GAGC boxes may also be relevant, since this is the same number of perfect GAGC boxes as in the intact AAV-ITR-RRS.

Chiorini et al. have employed a technique called random oligonucleotide selection to identify the consensus Rep binding site (7). By performing EMSAs they have isolated oligonucleotide probes containing Rep binding sites from a pool of randomly generated oligonucleotides. Their consensus sequence has two GAGC boxes in the center. Most of the new RRSs that we have identified here have the two intact GAGC boxes. However, it is interesting that none of the 16-mer core sequences of the RRSs that we have reported in this paper match exactly any of their oligonucleotide probes containing Rep binding sequences.

The binding of MBP-Rep68Δ to AAVS1 appears to be rel-

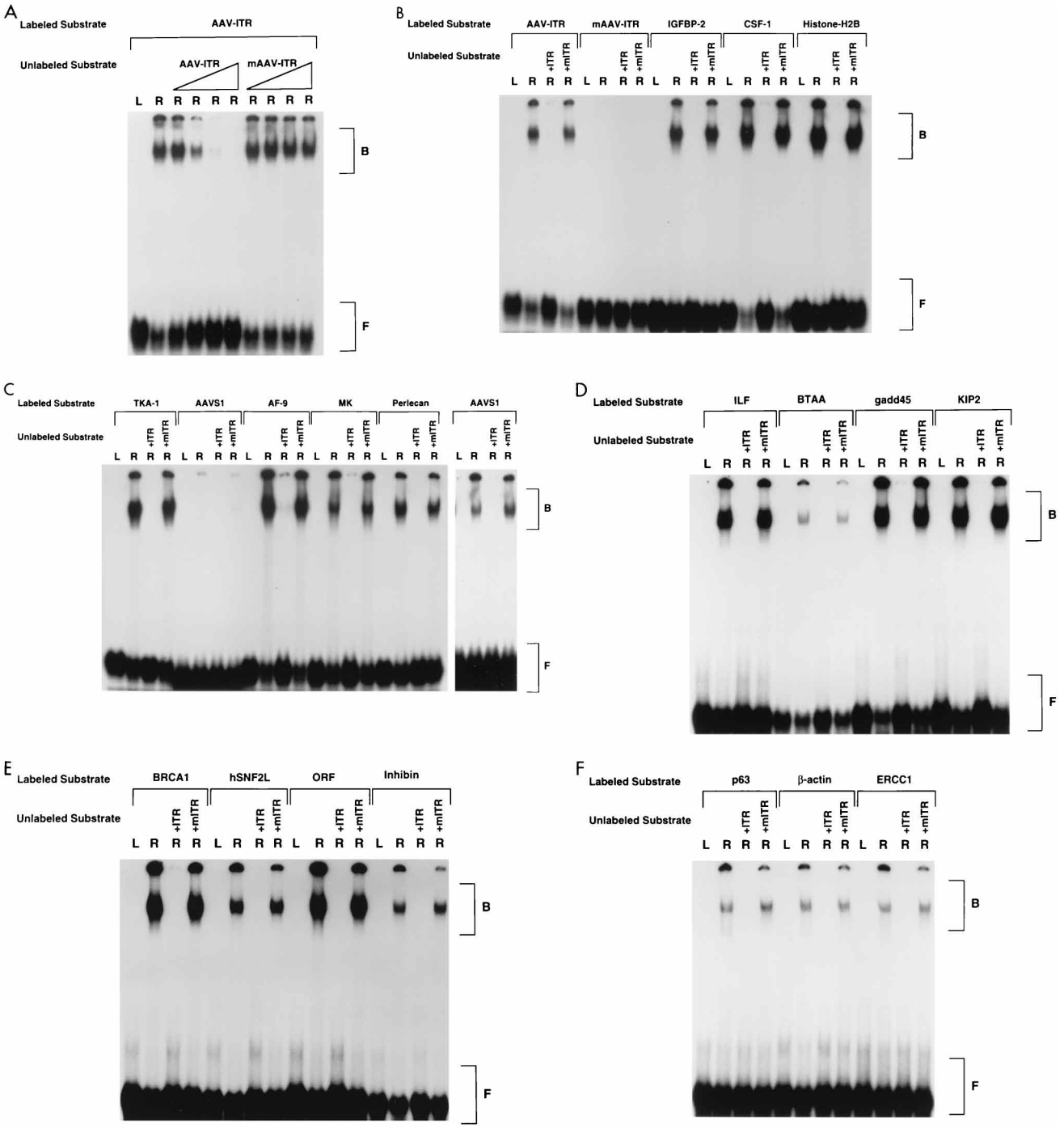

FIG. 5. Unlabeled AAV-ITR substrate inhibits the binding of MBP-Rep68Δ to the radiolabeled substrates. L and R represent reactions to which MBP-LacZ or MBP-Rep68Δ (1 μg) was added, respectively. B and F indicate bound and free DNA, respectively. (A) EMSAs were performed with the radiolabeled AAV-ITR substrate (0.04 pmol) along with either a 10-, 40-, 160-, or 640-fold molar excess of unlabeled AAV-ITR or the unlabeled mAAV-ITR substrate. (B to E) EMSAs were done with different radiolabeled substrates (0.04 pmol per reaction mixture) along with either a 160-fold molar excess of unlabeled AAV-ITR (indicated by ITR) or unlabeled mAAV-ITR (indicated by mITR). The last four lanes of panel C show AAVS1 lanes from an autoradiograph of the same gel after exposure for 10 days. The different panels show different gels. BTAA, breast tumor autoantigen.

atively weak compared to the binding to the AAV-ITR. Because this result was somewhat unexpected, we performed several confirmatory experiments. In order to confirm the sequence of the AAVS1 substrate used in this report, the annealed oligonucleotide pair containing the AAVS1-RRS was cloned into the *Bam*HI site of plasmid pBluescriptII SK(+) (Stratagene) and sequenced, and the sequence was shown to be correct. Additionally, we cut out the 223-bp *Bss*HII fragment of this plasmid, which contained the 46-bp segment of AAVS1, and found that MBP-Rep68Δ bound this fragment much better than the 54-bp-long annealed oligonucleotide pair containing the AAVS1-RRS. In fact, with about the same

affinity MBP-Rep68Δ bound a 200-bp *Bss*HII fragment containing a cloned core 16-mer AAVS1-RRS that was used by Weitzman et al. (52) (data not shown). The results from other EMSAs using fragments 60 to 70 bp in length that contained the cloned core 16-mer AAVS1-RRS or the AAV-p5-RRS as substrates and nuclear extracts from human 293 cells transfected with plasmids encoding wild-type Rep68 indicated that the AAVS1-RRS is a relatively weak binding site compared to the AAV-p5-RRS binding site (data not shown). Our conclusion is that the AAVS1-RRS is a weaker Rep protein binding sequence than either the AAV-ITR-RRS or the AAV-p5-RRS and that this weakness is accentuated when the AAVS1-RRS is in the context of a shorter DNA segment (the 54-bp annealed oligonucleotides, as opposed to a 200-bp restriction fragment).

We speculate that the Rep proteins actually recognize a secondary structure rather than a primary base sequence within the RRSs. This is based on the observation by Chiorini et al. (7) that Rep proteins bind extremely diverse sequences with comparable affinities and that many of the binding sequences identified contain no perfect GAGC boxes. Core sequences which contain changes that would drastically reduce binding in the AAV-ITR context (42) could still bind well, presumably because of compensatory deviations from the AAV-ITR sequence which would restore binding.

AAV preferentially integrates into AAVS1 approximately 70% of the time (28, 43). However, there are other sequences in the human genome that we have identified in this report that bind MBP-Rep68Δ significantly better than AAVS1. Therefore, it is intriguing to contemplate why AAV prefers AAVS1 as opposed to other sequences. One possible explanation could be that the presence of a Rep nicking site nearby (Fig. 3) may make AAVS1 a preferred Rep-dependent integration site (50). Linden et al. (33) have recently demonstrated that a 33-bp segment of AAVS1, which includes the RRS and the nicking site, is sufficient to direct integration of AAV into an episome in vivo. They also showed that mutation of either the RRS or the nicking site abolishes targeted integration. It is also important to note that we could not identify a consensus nicking site (5′-GTTGG-3′ or its complement, 5′-CCAAC-3′) for Rep close to the RRS core sequence of any of the newly identified RRSs. However, we could not rule out the possibility that a nicking site that is further away from the RRS core might be brought closer during chromatin formation. It has also recently been documented that the linker histones can affect the pattern of digestion of single-strand-specific endonucleases (21). Thus, in the context of chromatin, AAVS1 might be the optimal Rep-dependent nicking site, and that could explain why Rep-mediated AAV integration is into AAVS1 most of the time.

In this report we have identified 18 novel RRSs in the human genome. We have also recently reported four more RRSs, akin to the RRS in the p5 promoter of AAV, in different human genes, namely, the c-*sis* proto-oncogene (chromosome 22q12.3 to -q13.1), the hepatocyte glucose transporter gene (chromosome 22q13.1), the gene for carcinoma marker GA733-1 (chromosome 1p32), and the α-A-crystallin gene (chromosome 21) (54). Thus far, we have identified 22 sites in the human genome to which AAV Rep proteins bind. Since AAV integrates into AAVS1 only 70% of the time, these may serve as alternate integration sites for AAV. It is interesting that Walz and Schlehofer (51) detected integration of wild-type AAV into the long (q) arm of human chromosome 17, where we have identified two RRSs (in the ILF and *BRCA1* genes). Xiao et al. (55) noted integration of AAV into human chromosome 2, where we have identified two RRSs (in the IGFBP-2 and inhibin genes). Moreover, Rep proteins exhibit a wide variety of effects on human cells: Rep proteins have antineoplastic (15, 24), cytostatic (56), and antiviral (3) effects and down-regulate or up-regulate the expression of several cellular proto-oncogenes (16, 17, 54). These recognition sites in the human genome may play a role in these varied Rep-mediated effects. It is also important that many of these RRSs are within genes that are either directly or indirectly involved in cell proliferation.

We thank Jeffery Miller and Nancy Nossal for critically reviewing the manuscript. We also thank Timothy Wiggins for technical assistance.

REFERENCES

1. **Albertsen, H. M., S. A. Smith, S. Mazoyer, E. Fujimoto, J. Stevens, B. Williams, P. Rodriguez, C. S. Cropp, P. Slijepcevic, M. Carlson, M. Robertson, P. Bradley, E. Lawrence, T. Harrington, Z. Mei Sheng, R. Hoopes, N. Sternberg, A. Brothman, R. Callahan, B. A. J. Ponder, and R. White.** 1994. A physical map and candidate genes in the BRCA1 region on chromosome 17q12-21. Nat. Genet. **7:**472–479.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Antoni, B. A., A. B. Rabson, I. L. Miller, J. P. Trempe, N. Chejanovsky, and B. J. Carter.** 1991. Adeno-associated virus Rep protein inhibits human immunodeficiency virus type 1 production in human cells. J. Virol. **65:**396–404.
4. **Carter, B. J.** 1990. Adeno-associated virus helper functions, p. 255–282. *In* P. Tijssen (ed.), Handbook of parvoviruses, vol. I. CRC Press, Inc., Boca Raton, Fla.
5. **Chiorini, J. A., M. D. Weitzman, R. A. Owens, E. Urcelay, B. Safer, and R. M. Kotin.** 1994. Biologically active Rep proteins of adeno-associated virus type 2 produced as fusion proteins in *Escherichia coli*. J. Virol. **68:**797–804.
6. **Chiorini, J. A., S. M. Wiener, R. A. Owens, S. R. M. Kyöstiö, R. M. Kotin, and B. Safer.** 1994. Sequence requirements for stable binding and function of Rep68 on the adeno-associated virus type 2 inverted terminal repeats. J. Virol. **68:**7448–7457.
7. **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. **69:**7334–7338.
8. **Cohen, I. R., S. Grässel, A. D. Murdoch, and R. V. Iozzo.** 1993. Structural characterization of the complete human perlecan gene and its promoter. Proc. Natl. Acad. Sci. USA **90:**10404–10408.
9. **Collart, D., A. Ramsey-Ewing, R. Bortell, J. Lian, J. Stein, and G. Stein.** 1991. Isolation and characterization of a cDNA from a human histone H2B gene which is reciprocally expressed in relation to replication-dependent H2B histone genes during HL60 cell differentiation. Biochemistry **30:**1610–1617.
10. **Collart, D., P. L. Romain, K. Huebner, S. Pockwinse, S. Pilapil, L. A. Cannizzaro, J. B. Lian, C. M. Croce, J. L. Stein, and G. S. Stein.** 1992. A human histone H2B.1 variant gene, located on chromosome 1, utilizes alternative 3′ end processing. J. Cell. Biochem. **50:**374–385.
11. **Ehrenborg, E., S. Vilhelmsdotter, S. Bajalica, C. Larsson, I. Stern, J. Koch, K. Brøndum-Nielsen, and H. Luthman.** 1991. Structure and localization of the human insulin-like growth factor-binding protein 2 gene. Biochem. Biophys. Res. Commun. **176:**1250–1255.
12. **Fairhurst, J. L., P. J. Kretschmer, E. Kovacs, P. Böhlen, and I. Kovesdi.** 1993. Structure of the gene coding for the human retinoic acid-inducible factor, MK. DNA Cell Biol. **12:**139–147.
13. **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.
14. **Giraud, C., E. Winocour, and K. I. Berns.** 1994. Site-specific integration by adeno-associated virus is directed by a cellular DNA sequence. Proc. Natl. Acad. Sci. USA **91:**10039–10043.
15. **Hermonat, P. L.** 1989. The adeno-associated virus Rep78 gene inhibits cellular transformation induced by bovine papillomavirus. Virology **172:**253–261.
16. **Hermonat, P. L.** 1994. Down-regulation of the human c-*fos* and c-*myc* proto-oncogene promoters by adeno-associated virus Rep78. Cancer Lett. **81:**129–136.
17. **Hermonat, P. L.** 1991. Inhibition of H-*ras* expression by the adeno-associated virus Rep78 transformation suppressor gene product. Cancer Res. **51:**3373–3377.
18. **Hollander, M. C., I. Alamo, J. Jackman, M. G. Wang, O. W. McBride, and A. J. Fornace, Jr.** 1993. Analysis of the mammalian gadd45 gene and its response to DNA damage. J. Biol. Chem. **268:**24385–24393.
19. **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.
20. **Im, D. S., and N. Muzyczka.** 1989. Factors that bind to adeno-associated virus terminal repeats. J. Virol. **63:**3095–3104.

21. **Ivanchenko, M., J. Zlatanova, P. Varga-Weisz, A. Hassan, and K. van Holde.** 1996. Linker histones affect patterns of digestion of supercoiled plasmids by single-strand-specific nucleases. Proc. Natl. Acad. Sci. USA **93:**6970–6974.
22. **Kallunki, P., and K. Tryggvason.** 1992. Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD protein containing multiple domains resembling elements of the low density lipoprotein receptor, laminin, neural cell adhesion molecules, and epidermal growth factor. J. Cell Biol. **116:**559–571.
23. **Kawasaki, E. S., M. B. Ladner, A. M. Wang, J. Van Arsdell, M. K. Warren, M. Y. Coyne, V. L. Schweickart, M. T. Lee, K. J. Wilson, A. Boosman, E. R. Stanley, P. Ralph, and D. F. Mark.** 1985. Molecular cloning of a complementary DNA encoding human macrophage-specific colony-stimulating factor (CSF-1). Science **230:**291–296.
24. **Khleif, S. N., T. Myers, B. J. Carter, and J. P. Trempe.** 1991. Inhibition of cellular transformation by the adeno-associated virus rep gene. Virology **181:** 738–741.
25. **Kotin, R. M., and K. I. Berns.** 1989. Organization of adeno-associated virus DNA in latently infected Detroit 6 cells. Virology **170:**460–467.
26. **Kotin, R. M., R. M. Linden, and K. I. Berns.** 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. **11:**5071–5078.
27. **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.
28. **Kotin, R. M., M. Siniscalco, R. J. Samulski, X. 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.
29. **Kyöstiö, S. R. M., R. S. Wonderling, and R. A. Owens.** 1995. Negative regulation of the adeno-associated virus (AAV) P_5 promoter involves both the P_5 Rep binding site and the consensus ATP-binding motif of the AAV Rep68 protein. J. Virol. **69:**6787–6796.
30. **Ladner, M. B., G. A. Martin, J. A. Noble, D. M. Nikoloff, R. Tal, E. S. Kawasaki, and T. J. White.** 1987. Human CSF-1: gene structure and alternative splicing of mRNA precursors. EMBO J. **6:**2693–2698.
31. **Li, C., C. Lai, D. S. Sigman, and R. B. Gaynor.** 1991. Cloning of a cellular factor, interleukin binding factor, that binds to NFAT-like motifs in the human immunodeficiency virus long terminal repeat. Proc. Natl. Acad. Sci. USA **88:**7739–7743.
32. **Li, C., A. J. Lusis, R. Sparkes, A. Nirula, and R. Gaynor.** 1992. Characterization and chromosomal mapping of the gene encoding the cellular DNA binding protein ILF. Genomics **13:**665–671.
33. **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.
34. **Martin-Gallardo, A., W. R. McCombie, J. D. Gocayne, M. G. FitzGerald, S. Wallace, B. M. B. Lee, J. Lamerdin, S. Trapp, J. M. Kelley, L.-I. Liu, M. Dubnick, L. A. Johnston-Dow, A. R. Kerlavage, P. de Jong, A. Carrano, C. Fields, and J. C. Venter.** 1992. Automated DNA sequencing and analysis of 106 kilobases from human chromosome 19q13.3. Nat. Genet. **1:**34–39.
35. **Mason, A. J., L. M. Berkemeier, C. H. Schmelzer, and R. H. Schwall.** 1989. Activin B: precursor sequences, genomic structure and in vitro activities. Mol. Endocrinol. **3:**1352–1358.
36. **Matsuoka, S., M. C. Edwards, C. Bai, S. Parker, P. Zhang, A. Baldini, J. W. Harper, and S. J. Elledge.** 1995. $p57^{KIP2}$, a structurally distinct member of the $p21^{CIP1}$ Cdk inhibitor family, is a candidate tumor suppressor gene. Genes Dev. **9:**650–662.
37. **Nakamura, T., H. Alder, Y. Gu, R. Prasad, O. Canaani, N. Kamada, R. P. Gale, B. Lange, W. M. Crist, P. C. Nowell, C. M. Croce, and E. Canaani.** 1993. Genes on chromosomes 4, 9, and 19 involved in 11q23 abnormalities in acute leukemia share sequence homology and/or common motifs. Proc. Natl. Acad. Sci. USA **90:**4631–4635.
38. **Nomura, N., T. Nagase, N. Miyajima, T. Sazuka, A. Tanaka, S. Sato, N. Seki, Y. Kawarabayasi, K. Ishikawa, and S. Tabata.** 1994. Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. DNA Res. **1:**223–229.
39. **Okabe, I., L. C. Bailey, O. Attree, S. Srinivasan, J. M. Perkel, B. C. Laurent, M. Carlson, D. L. Nelson, and R. L. Nussbaum.** 1992. Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast S. cerevisiae. Nucleic Acids Res. **20:**4649–4655.
40. **Owens, R. A., J. P. Trempe, N. Chejanovsky, and B. J. Carter.** 1991. Adeno-associated virus Rep proteins produced in insect and mammalian expression systems: wild-type and dominant-negative mutant proteins bind to the viral replication origin. Virology **184:**14–22.
41. **Papathanasiou, M. A., N. C. K. Kerr, J. H. Robbins, O. W. McBride, I. Alamo, Jr., S. F. Barrett, I. D. Hickson, and A. J. Fornace, Jr.** 1991. Induction by ionizing radiation of the *gadd45* gene in cultured human cells: lack of mediation by protein kinase C. Mol. Cell. Biol. **11:**1009–1016.
42. **Ryan, J. H., S. Zolotukhin, and N. Muzyczka.** 1996. Sequence requirements for binding of Rep68 to the adeno-associated virus terminal repeats. J. Virol. **70:**1542–1553.
43. **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.
44. **Schlehofer, J. R., M. Ehrbar, and H. zur Hausen.** 1986. Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus. Virology **152:**110–117.
45. **Schweizer, A., J. Rohrer, P. Jenö, A. DeMaio, T. G. Buchman, and H. P. Hauri.** 1993. A reversibly palmitoylated resident protein (p63) of an ER-Golgi intermediate compartment is related to a circulatory shock resuscitation protein. J. Cell Sci. **104:**685–694.
46. **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.
47. **Sun, X.-L., B. R. Murphy, Q.-J. Li, S. Gullapalli, J. Mackins, H. N. Jayaram, A. Srivastava, and A. C. Antony.** 1995. Transduction of folate receptor cDNA into cervical carcinoma cells using recombinant adeno-associated virions delays cell proliferation in vitro and in vivo. J. Clin. Invest. **96:**1535–1547.
48. **Tokino, T., T. Urano, T. Furuhata, M. Matsushima, T. Miyatsu, S. Sasaki, and Y. Nakamura.** 1996. Characterization of the human $p57^{KIP2}$ gene: alternative splicing, insertion/deletion polymorphisms in VNTR sequences in the coding region, and mutational analysis. Hum. Genet. **97:**625–631.
49. **Uehara, K., S. Matsubara, K. Kadomatsu, J. Tsutsui, and T. Muramatsu.** 1992. Genomic structure of human midkine (MK), a retinoic acid-responsive growth/differentiation factor. J. Biochem. **111:**563–567.
50. **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.
51. **Walz, C., and J. R. Schlehofer.** 1992. Modification of some biological properties of HeLa cells containing adeno-associated virus DNA integrated into chromosome 17. J. Virol. **66:**2990–3002.
52. **Weitzman, M. D., S. R. M. Kyöstiö, 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.
53. **Wonderling, R. S., S. R. M. Kyöstiö, and R. A. Owens.** 1995. A maltose-binding protein/adeno-associated virus Rep68 fusion protein has DNA-RNA helicase and ATPase activities. J. Virol. **69:**3542–3548.
54. **Wonderling, R. S., and R. A. Owens.** 1996. The Rep68 protein of adeno-associated virus type 2 stimulates expression of the platelet-derived growth factor B c-*sis* proto-oncogene. J. Virol. **70:**4783–4786.
55. **Xiao, X., W. deVlaminck, and J. Monahan.** 1993. Adeno-associated virus (AAV) vectors for gene transfer. Adv. Drug Delivery Rev. **12:**201–215.
56. **Yang, Q., F. Chen, and J. P. Trempe.** 1994. Characterization of cell lines that inducibly express the adeno-associated virus Rep proteins. J. Virol. **68:**4847–4856.