The EMBO Journal vol.10 no.12 pp.3941 - 3950, 1991

Targeted integration of adeno-associated virus (AAV) into human chromosome 19

R.J.Samulski, X.Zhu, X.Xiao, J.D.Brook[1], D.E.Housman[1], N.Epstein[2] and L.A.Hunter

Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260, [1]Department of Biology and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, and [2]Bldg 10, Rm 7C103, National Institutes of Health, NHLBI, 9000 Rockville Pike, Bethesda, MD 20892, USA

Communicated by T.Kohn

A key feature in adeno-associated virus (AAV) replication is efficient integration of the viral genome into host cell DNA to establish latency when helper virus is absent. The steps involved in this process remain largely uncharacterized, even though AAV integration was first documented 20 years ago. Using a protein-DNA binding method we isolated AAV-cellular junction DNA sequences. The cellular component hybridized to a single restriction fragment in the virus-free parental cell line, and also co-migrated with AAV-specific sequences in numerous latently infected cell lines. Analysis of somatic cell hybrids indicated that this cellular sequence maps to the distal portion of the q arm of human chromosome 19. *In situ* hybridization of AAV DNA to chromosomes from latently infected cells confirms the physical location of AAV integrations to be q13.4-ter of chromosome 19. Sequence analysis of several independent integration sites shows breakpoints occurring within a 100 bp cellular region. This non-pathogenic parvovirus thus appears to establish viral latency by integrating its DNA specifically into one chromosomal region. Such specific integration is so far unique among the eukaryotic DNA viruses. The incorporation of site-specific integration into AAV vector schemes should make this vector system attractive for human gene therapy approaches.

Key words: AAV/adeno-associated virus/chromosome/human/mapping

Introduction

Adeno-associated virus-2 (AAV-2) is a defective parvovirus with a genome of 4675 bp. In a lytic infection, replication requires co-infection with either adenovirus (Atchinson *et al.*, 1965; Hoggan, 1965) or herpes simplex virus (Buller *et al.*, 1981). When no helper virus is available, AAV infection results in efficient integration of the AAV genome into host cell genomic DNA (Berns, 1984; Berns and Bohenzky, 1987). Subsequent super-infection of AAV latently infected cells with helper virus then results in the rescue and replication of the AAV genome (Berns, 1984). Virus integration appears to have no apparent effect on cell growth or morphology (Handa *et al.*, 1977) and was first documented as cryptic infections in primary African green monkey and human embryonic kidney cells (Berns *et al.*, 1982). In these studies, the prevalence of naturally occurring AAV latent infections was determined by monitoring the appearance of AAV antigens after adenovirus challenge; 20% of monkey and 2% of human cells tested scored positive (Hoggan *et al.*, 1972). Latently infected cells established *in vitro* appear to be stable, having maintained the viral DNA for >150 passages (Berns *et al.*, 1982; and this study, for example the H3 cell line).

Studies on the physical structure of integrated AAV genomes suggest that viral insertions are usually in a tandem head to tail orientation via the AAV terminal repeats (Cheung *et al.*, 1980; Berns *et al.*, 1982; McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). When independently derived latent cell lines were previously characterized by Southern hybridization analysis, the viral-cellular junction fragments migrated with different mobilities, suggesting that AAV integration was occurring randomly (Laughlin *et al.*, 1986; McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). In this report, we describe the isolation and characterization of a cellular sequence which is consistently associated with integrated AAV DNA in numerous independent latent cell lines. Analysis of rodent-human panels and *in situ* chromosome hybridization reveals that this sequence maps to the q arm of chromosome 19 and that the AAV DNA integrates specifically into this region.

Results

Strategy for retrieving AAV integrants

To characterize the integration of AAV genomes into cellular DNA, we developed an AAV hybrid virus that would serve as a tool in identifying and retrieving genomic sequences after viral integration. The strategy for retrieving AAV-cellular junctions involved a protein filter binding procedure which made use of the specific interaction between lambda repressor and its operator sequences OR1 and OR2. This strategy is diagrammed in Figure 1 (for further details see X.Zhu and R.J.Samulski, in preparation). The infectious AAV recombinant clone pSM620 (Samulski *et al.*, 1982) was used to generate an AAV-lambda hybrid construct that carried a 47 bp sequence containing the lambda operator subsites OR1 and OR2 (Johnson *et al.*, 1981). These sequences, when inserted into a non-essential region of the AAV genome (Samulski *et al.*, 1987; Figures 1 and 2) did not affect viral DNA replication (Figure 2, lanes 2 and 3) or virus production (data not shown). Before establishing independent latently infected AAV-OR1-OR2 cell lines, we also tested for integration and rescue of the viral genome by infecting HeLa cells with the hybrid virus in the absence of adenovirus (Ad) as described by Samulski *et al.* (1989). The infected cells were passaged until no detectable AAV could be identified in the media or in Hirt extracts (data not shown). The cells were then challenged with Ad and tested for rescue and replication of the integrated AAV hybrid

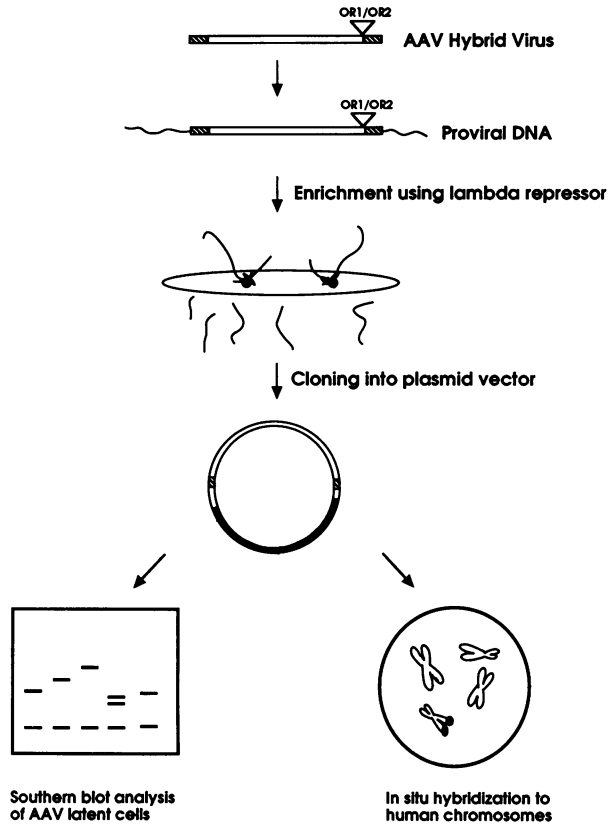

Fig. 1. Schematic diagram of the construction, isolation and characterization of AAV proviral DNA. A 47 bp oligo comprising the lambda operator sequences (OR1-OR2) was cloned into the infectious AAV recombinant plasmid and used to generate hybrid virus stocks which in turn were used to generate AAV latent cell lines. Genomic DNAs isolated from AAV−OR1-OR2 latent cell lines were digested with restriction enzymes that cut outside the hybrid AAV genome, mixed with purified lambda repressor and processed through a nitrocellulose membrane in a filter retention assay. Proviral fragments retained on the filter through repressor specific interactions were eluted, cloned and sequenced. Cellular junction sequences were then used to characterize further various AAV latent cells lines by both Southern hybridization analysis and *in situ* hybridization to human chromosomes.

sequences (Figure 2, lanes 5 and 6). These functions were also indistinguishable from wild type (Figure 2, lanes 4−6). Southern hybridization analysis using a lambda operator-specific probe demonstrated that these viral sequences also maintained the inserted oligos (data not shown). We concluded that the lambda operator sequences were stable and that they had no deleterious effect on the replication, integration or rescue of the AAV genome in cultured cells.

Isolation of AAV−OR1-OR2 DNA from latent cell lines

We established 11 different AAV−OR1-OR2 latent cell lines in human (HeLa and Detroit 6), and monkey (CV1) cells. Single cell clones that contained AAV−OR1-OR2 integrated sequences were identified by PCR analysis of genomic DNA. The viral DNA could be rescued from these cells upon Ad challenge (data not shown). Genomic DNAs isolated from latent cell lines were digested with an enzyme that does not recognize viral sequences, and were then fractionated on an agarose gel (Figure 3A, lanes 1, 3, 5, 7 and 9). An equivalent amount of digested genomic DNA was also used for filter binding with the lambda repressor. DNA sequences retained

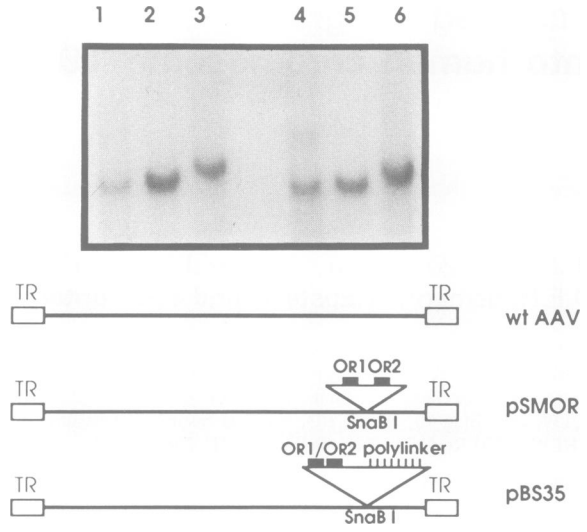

Fig. 2. Analysis of AAV−OR1-OR2 hybrid virus replication and rescue. Wild type AAV and AAV−OR1-OR2 virus stocks were used to infect HeLa cells with (lanes 1−3) or without (lanes 4−6) helper adenovirus (Ad). Cells infected in the absence of helper virus were carried for numerous passages to establish viral latency. Cells which were negative by Southern hybridization analysis for free AAV genomes were then assayed for viral rescue by superinfection with Ad (lanes 4−6). 48 h post-Ad infection low molecular weight DNA was isolated by Hirt procedure, fractionated on a 0.8% agarose gel, and subjected to Southern analysis using wild-type AAV as a specific probe. Lanes 1 and 4, wild type AAV (4675 bp); lanes 2 and 5, pSMOR (4722 bp); lanes 3 and 6, pBS35 (4972 bp).

on the filter via repressor-specific interactions were eluted and fractionated as well (Figure 3A, lanes 2, 4, 6, 8 and 10). Southern hybridization analysis using AAV DNA as a probe demonstrated the specific retention of viral sequences containing the operator sequences OR1-OR2 after filter binding with the lambda repressor when compared with wild type AAV DNA (Figure 3B, lanes 1 and 2 versus lanes 9 and 10). The high molecular weight of these AAV sequences (Figure 3B) suggested that they were associated with cellular DNA.

A total of 25 recombinants isolated from three latent cell lines were characterized. In all cases, the cloned viral−cellular sequences appeared to have rearranged based on a comparison with restriction patterns of genomic DNA, possibly suggesting that these sequences were unstable in bacteria (data not shown). These rearrangements occurred primarily at the AAV−cellular junctions, and seemed to involve the viral terminal repeats. The 145 bp terminal repeat of AAV can form a 125 bp palindromic T shaped structure which appears to be extremely unstable in bacterial cells (Samulski *et al.*, 1982, 1983; Bohenzky *et al.*, 1988; Laughlin *et al.*, 1983). This instability is also observed with the autonomous parvoviruses MVM, JNC, and the pathogenic human parvovirus B19 (Merchlinsky *et al.*, 1983; Jourdan *et al.*, 1990; Deiss *et al.*, 1990). Because of this phenomenon, the exact viral−cellular junction was not elucidated in these preliminary studies (see below, Analysis of PCR-amplified AAV−19 junctions).

Nucleotide comparison among the cloned cellular sequences, however, demonstrated that outside the rearranged region they were identical to one another (data not shown). We further analyzed a 300 bp fragment which was distinct from AAV DNA. This fragment was subcloned

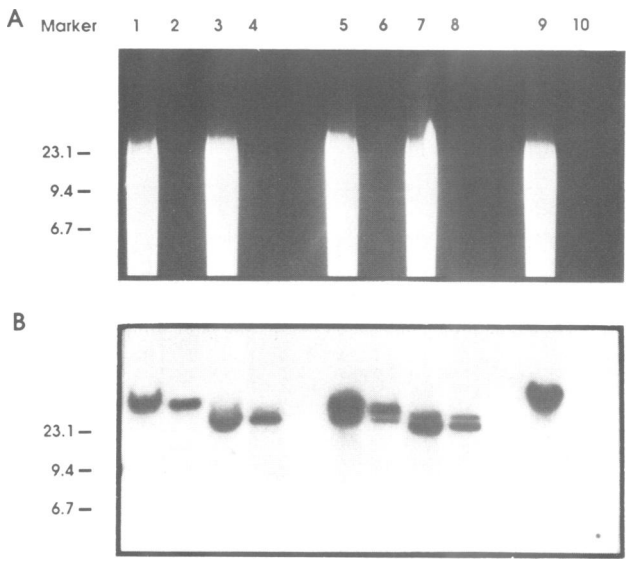

Fig. 3. Selective enrichment of AAV−OR1-OR2 proviral sequences using lambda repressor in a filter binding retention assay. **Panel A.** Ethidium bromide stained agarose gel of digested genomic DNA isolated from AAV−OR1-OR2 latent cell line H3 (lanes 1, 2, 5 and 6), G11 (lanes 3, 4, 7 and 8), or a wild type AAV latent cell line D5 (lanes 9 and 10). Each lane contains 10 μg of genomic DNA either digested (odd numbered lanes), or digested and processed through the filter binding retention assay (even numbered lanes). Molecular weight markers (in kb) are indicated. DNAs in lanes 1, 2, 5 and 6 were digested with *Bgl*II and lanes 3, 4, 7−10, were digested with *Pvu*II. **Panel B.** Southern hybridization analysis of panel A using AAV-specific probe.

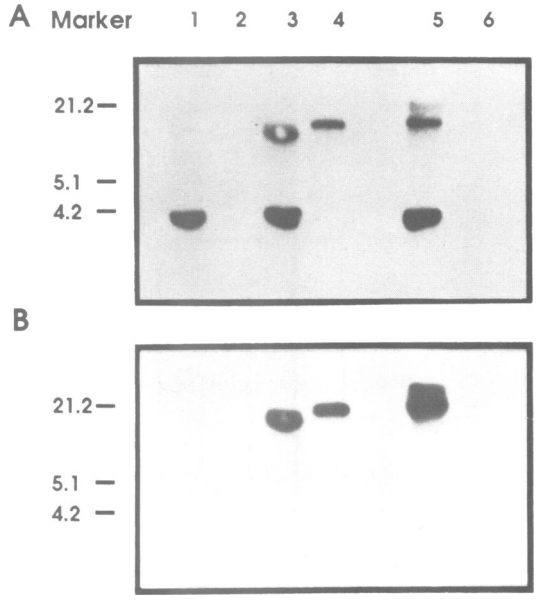

Fig. 4. Analysis of H3 cellular junction sequences in parental and AAV latent cell genomic DNA. **Panel A.** Southern hybridization analysis of genomic DNA from parental cell line HeLa (lanes 1 and 2), AAV−OR1-OR2 latent cell line H3 (lanes 3 and 4) and wild type AAV latent cell line D5 (lanes 5 and 6) using H3 junction-specific probe. Each lane contains 10 μg of genomic DNA digested with an enzyme that does not cut viral sequences (*Pvu*II; odd numbered lanes), or digested and then processed through the repressor filter retention assay (even numbered lanes). **Panel B.** Southern hybridization analysis of the same filter using AAV-specific probe. Molecular weight markers (in kb) are indicated for each panel.

and was used as a probe (referred to as junction probe 1) in Southern hybridization analyses of genomic DNA from the parental cell line (HeLa), the AAV−OR1-OR2 latent cell line (H3) and a previously established AAV latent cell line (D5) (Figure 4, lanes 1, 3 and 5). A single band of 4 kbp was present in the parental cell line (Figure 4A, lane 1). As expected, no parental DNA which hybridized to the probe was retained in the filter binding assay (Figure 4A, lane 2). Hybridization to digested genomic DNA from the latent cell line H3 yielded two bands (Figure 4A, lane 3), one with a mobility identical to the species observed in the parental cell line (4 kbp), and one of a higher molecular weight which co-migrated with the AAV-specific DNA (compare Figure 4A and B, lanes 3). Only the higher molecular weight species was retained by repressor binding (compare Figure 4A and B, lanes 4), indicating the physical linkage of this cellular DNA to viral sequences. The genomic Southern of the D5 cell line (containing wild type AAV DNA and established 20 years ago) detected not only the 4 kbp band, but also a higher molecular weight species that also co-migrated with AAV DNA (compare Figure 4A and B, lanes 5). The D5 fragments were not retained on the repressor filter demonstrating that they were not detected as a result of contamination with AAV−OR1-OR2 DNA (Figure 4A and B, lanes 6). These results suggested that the AAV DNA in both cell lines was associated with the same chromosomal sequence.

Using the same approach, we analyzed eight independently derived AAV−OR1-OR2 latently infected cell lines for the possible association of this junction sequence with the proviral DNA. In all cases we observed co-migration of the same cellular junction with the viral DNA (data not shown). Four of these cell lines were analyzed after enrichment by repressor filter binding and the retained fragment in each case hybridized specifically with the cellular sequence isolated from H3 (data not shown). In this characterization, we observed primarily two types of viral integration pattern. One type appeared to consist of a simple integration—multiple AAV proviral genomes migrated as a single species after digestion with an enzyme that does not cut AAV DNA (Figure 5). The second type of configuration also contains multiple integrants, but in this case, several distinct fragments were produced after digestion with the same enzyme (Figure 5, see G11). In each case, at least one viral fragment was also detected with the H3 junction probe, supporting the use of a common integration sequence. Digestion of all genomic DNAs with an enzyme that cuts once in the AAV genome revealed that single bands observed with a no-cutter contained tandemly repeated copies of the AAV genome (data not shown). However, as previously described for D5 (three to five copies, Cheung *et al.*, 1980) and certain other AAV latent cell lines (McLaughlin *et al.*, 1988), amplified chromosomal sequences between AAV genomes sometimes occurred (for example, G11 and C10 of this study; data not shown).

Chromosome mapping and characterization of the common integration site

The Southern blot analyses of these AAV latent cell lines suggested that the junction sequence originally isolated from H3 was also associated with AAV proviral DNA in other latent cell lines including some of simian origin (data not shown). To determine the extent of conservation of this sequence, cellular DNAs isolated from various species were

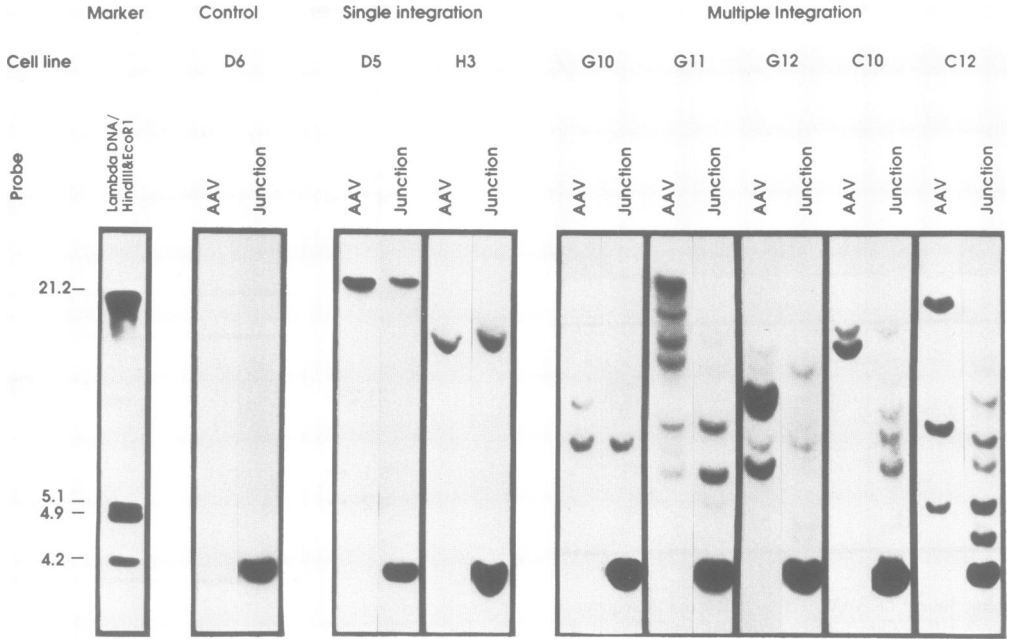

Fig. 5. Analysis of H3 cellular junction sequences in multiple independent latent cell lines. Each lane contains 10 μg of genomic DNA isolated from a parental cell line Detroit 6 (D6), or various independently derived AAV latent cell lines (D5, H3, G10, G11, G12, C10 and C12), digested with *Pvu*II and fractionated on a 0.8% agarose gel. Filters were first probed with H3 cellular sequences (lanes indicated 'junction'), and then stripped and reprobed with AAV-specific sequences (lanes indicated 'AAV'). Molecular weight markers (in kb) are indicated.

analyzed by Southern hybridization using the H3 junction sequence as a probe. The sequence was conserved in human and monkey, but it was not detected in the other species tested including rat, mouse, canine, bovine, rabbit, chicken and yeast (Figure 6C). Although AAV latently infects cells from several of these organisms, its integration has not yet been characterized. Northern hybridization analysis using the H3 cellular probe failed to detect a cytoplasmic RNA species expressed from this region in HeLa cells (data not shown).

To determine the native chromosomal location(s) of the junction sequence in human cells, rodent−human hybrid panels were analyzed by Southern hybridization. Only hybrid panels containing human chromosome 19 hybridized to the junction probe (Figure 6A). Further characterization of human hybrids that contained various portions of chromosome 19 localized this junction sequence to the lower portion of the q arm (Figure 6B).

In situ chromosome analysis of AAV latently infected cell lines

Southern blot analysis of several AAV latent cell lines (Figure 5) revealed numerous high molecular weight fragments positive for AAV. Many of these AAV fragments did not co-migrate with the putative junction fragment; thus, the data could not determine whether these AAV sequences were associated with a common integration locus or were integrated at completely independent sites, perhaps on different chromosomes. Furthermore, localization of the junction probe to chromosome 19 does not confirm that the AAV sequences reside there, as rearrangements subsequent to integration could have occurred. To examine these possibilities, AAV latent cell lines were analyzed by *in situ* chromosome analysis using an AAV-specific probe. In addition, a probe which detects repetitive sequences present in high copy number at q13.2 of chromosome 19 and fewer times on chromosome 4, was used to orient the AAV signal. Figure 7A demonstrates the presence of AAV-specific sequences on the ter portion of only one chromosome from cell line H3 which was identified as 19 by analysis using AAV and chromosome 19-specific probes together (Figure 7D). The two other chromosomes 19 present in this cell line did not show detectable AAV sequences.

In situ chromosome characterization of cell lines that, by Southern analysis appeared to carry multiple different integrations of AAV (for example, G11) also showed viral sequences associated with only one chromosome (Figure 7B). These chromosomes were also identified as 19 by double labeling (Figure 7E). These results suggested that all of the AAV sequences were physically located at one locus even though they sometimes appeared not to be closely associated with the H3 junction sequence by Southern analysis (compare Figure 5, lanes G11 with Figure 7B). Finally, the 20 year old D5 cell line also carried AAV sequences exclusively on chromosome 19 (Figure 7C and F), demonstrating preference by AAV for this chromosomal location, and also ruling out the possibility that the lambda operator sequences played a role in defining the integration site. Furthermore, this result suggests that the AAV DNA which established viral latency some 20 years ago remains stably associated with this locus.

Analysis of PCR amplified AAV – 19 junctions

To characterize further the mechanism of integration we isolated and analyzed junction sequences from latent cell lines. Since the bacterial junction clones carried rearrangements, we used the H3 junction sequence to isolate a genomic preintegration site from a cosmid library. A 2651 bp *Bam*HI subclone which hybridized to the junction probe and to chromosome 19 (not shown) was sequenced (Figure 8).

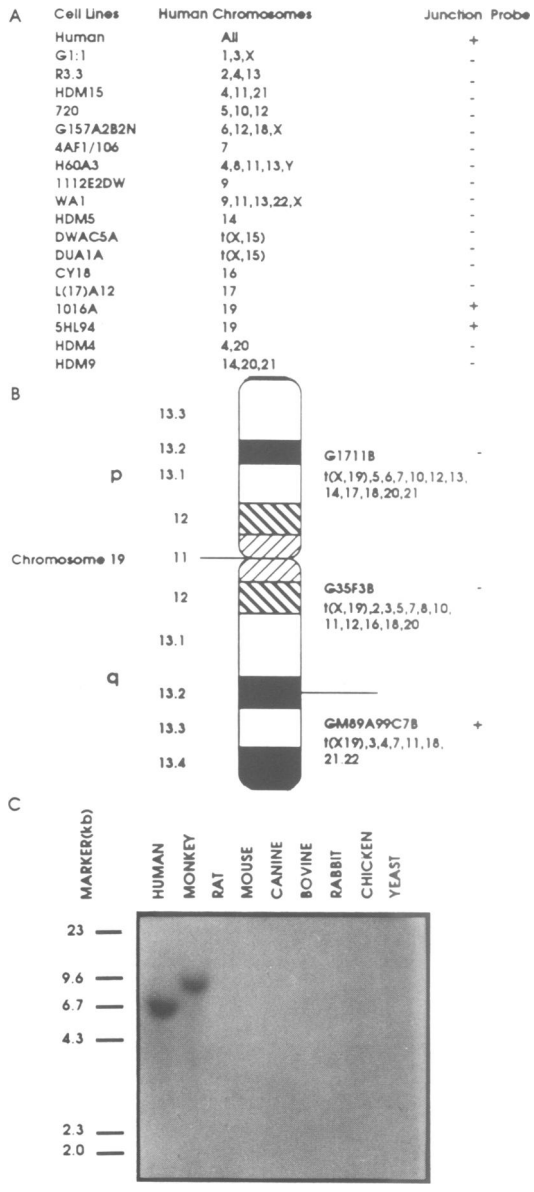

Fig. 6. Chromosome mapping and characterization of the common junction integration site. **Panel A.** Genomic DNAs from a number of rodent−human hybrid cell lines containing various human chromosomes were digested with *Pst*I, fractionated on an agarose gel, transferred to a nylon filter and hybridized with the H3 cellular junction probe. Cell lines positive for H3 junction probe are indicated by plus sign. These cell lines are described in Geissler *et al.* (1991). Cell lines 1016A and 1112E2DW go by the published names of PK-87-19 and PK-87-19 as given in Warburton *et al.* (1990) **Panel B.** Genomic DNAs from three rodent−human hybrid cell lines (G1711B, G35F3B and GM89A99C7B) containing different segments of human chromosome 19 (illustrated by horizontal line) were analyzed as in panel A using H3 junction-specific probe. G1711B and G35F3B retain the region 19pter−19q13.2. GM89A99C7B retains 19q13.3−19qter. All three cell lines contain other human chromosomes and are described in Brook *et al.* (1991). **Panel C.** 10 μg each of genomic DNA isolated from several different organisms as indicated were digested with *Eco*RI and analyzed by Southern hybridization with H3-specific junction probe. Molecular weight markers (in kb) are indicated.

From this sequence, a junction-specific primer designed for PCR amplification when paired with AAV-specific primers (Figure 8). Genomic DNA from latent cell lines H3, C11 or WI38/AAV2 (AAV infected human diploid cells) was used as a template and the amplified product was sequenced and compared with the preintegration sequence (Figure 8). Direct sequencing from asymmetric PCR amplification or from cloned PCR products generated similar results (Figure 8). The majority of the viral−cellular break points lie within a 100 bp sequence. Analysis of the preintegration sequences shows a 21 bp direct repeat overlapping this region. In most cases, portions of the viral terminal repeat were missing and patchy homology of 2−3 bp was present at the viral−cellular crossover. In two clones from independent cell lines (H3 and C11) enough of the viral terminal repeats remained such that the viral orientation could be determined (flip in H3 and flop in C11).

Discussion

Site-specific integration

We utilized a protein DNA binding enrichment technique to isolate AAV proviral DNA from human cell lines. Our characterization of the human-specific DNA and its association with viral DNA indicates that this non-pathogenic human parvovirus integrates site-specifically, a phenomenon which is currently unique among eukaryotic DNA viruses. The initial indication that AAV DNA may integrate in such a manner came from Kotin *et al.* (1990), where high frequency (68%) co-migration between AAV fragments derived from various latent cell lines and a specific AAV junction sequence was demonstrated using Southern hybridization analysis. Given the limitations of this technique, however, it was possible that establishment of viral latency had caused the association of cellular sequences with AAV DNA without stable integration into the site of their origin. In the present study, we provide three lines of evidence which suggest that AAV DNA does indeed become stably integrated into chromosome 19. First, we show the physical linkage between AAV proviral DNA and cellular sequences which map to human chromosome 19 (Figures 4 and 5). Second, using sequence information from a preintegration site for PCR-based amplification, we generated viral/junction sequences that map within 100 bp on chromosome 19. These results were the same whether from latent HeLa cell lines or the diploid WI38 cell line (Figure 8). The third, and strongest line of evidence comes from numerous *in situ* analyses of latently infected cell chromosomes. Using an AAV-specific probe, we demonstrated that viral DNA integrated into only one locus (Figure 7A, B and C). This was true regardless of AAV copy number and apparent complexity of AAV restriction fragment patterns in certain cell lines (compare Figure 5 lanes G11 with Figure 7B). The concomitant use of chromosome 19-specific sequences with AAV DNA in double labeled *in situ* analysis confirms the physical location of AAV integration at q14.3-ter of chromosome 19 (Figure 7D, E and F). It still remains to be determined if the complex integration pattern we observed in some cell lines is an initial event which sometimes occurs during AAV integration, or results from subsequent rearrangement. We anticipate from preliminary examination of AAV integration in normal human diploid cells that this complexity is most likely unique to aneuploid cell lines (X.Zhu, X.Xiao and R.J.Samulski, unpublished observation). Our *in situ* analysis also illustrated that regardless of the number of chromosomes 19 present

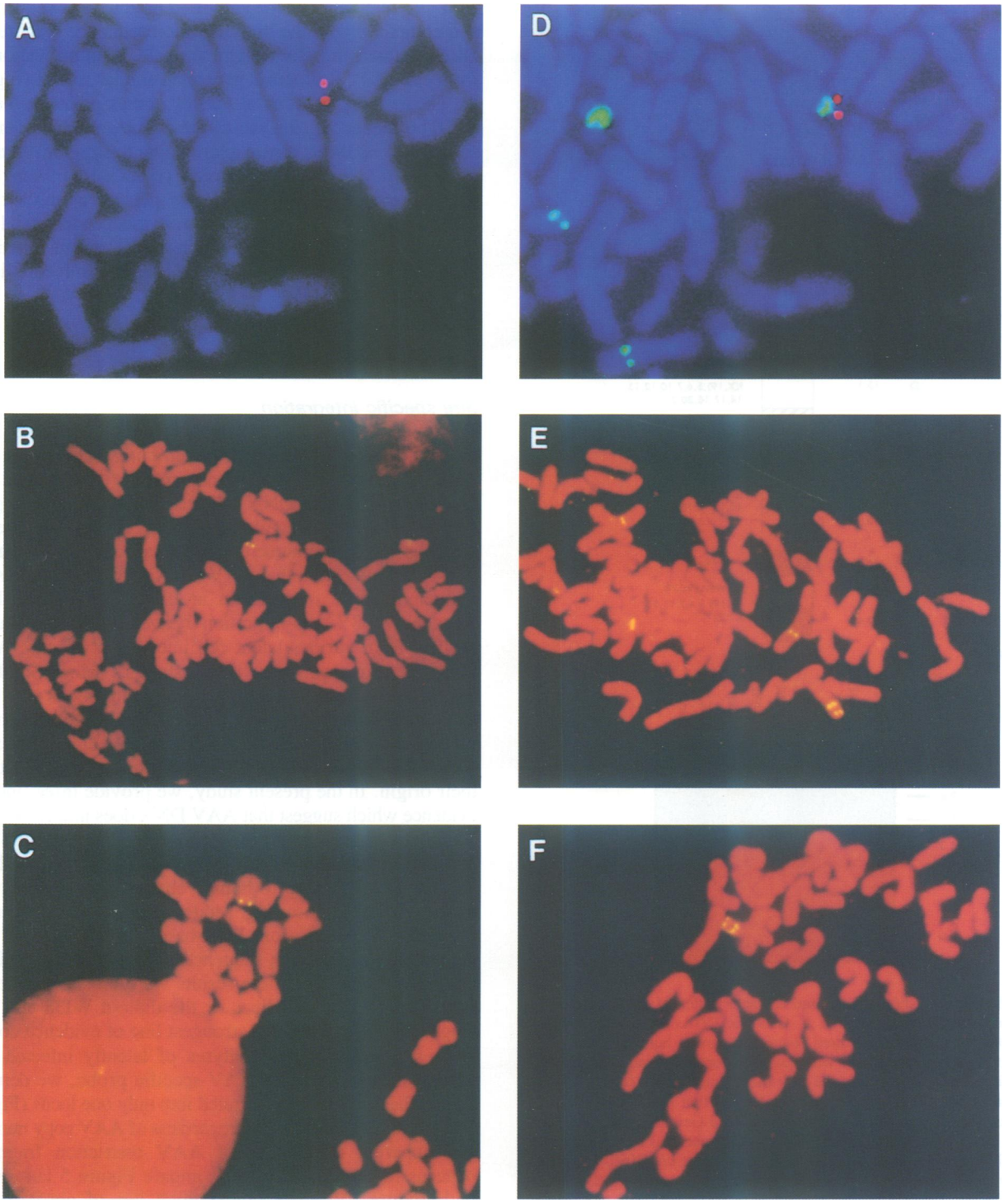

Fig. 7. *In situ* chromosome analysis of AAV latent cell lines. Detection of AAV provirus at the tip of one chromosome in the latent cell line H3 (**panel A**, red signal), G11 and D5 (**panels B** and **C**, yellow signal) using AAV-specific probe. Colocalization of AAV provirus to one chromosome 19 by *in situ* analysis using both AAV and 19-specific probe (**panels D, E** and **F**, double dots). Panel A and D chromosomes were counterstained with DAPI (blue), whereas panels B, E, C and F were counterstained with propidium iodide (red). All DNA probes were indirectly labeled with FITC, except the AAV probes used in panels A and D which were labeled with Cy3 (a rhodamine-like dye with red color).

within a cell (HeLa contained three copies of 19, whereas D5 and WI38 contained two), AAV DNA was associated with only one chromosome. It is possible that AAV integration into only one chromosome 19 was observed because of selective pressure (cell viability may require one unrearranged chromosome 19) or because the input levels of virus were limiting. Further analyis should resolve these possibilities.

Viral targeting

The high specificity described here for AAV DNA integration has not previously been demonstrated with other eukaryotic DNA viruses. However, evidence for non-random integration has been obtained in some cases for retroviruses (Cohen and Murphey-Corb, 1983; Shih *et al.*, 1988). Common integration sites were documented in cells transformed by many non-acute retroviruses (Gallahan *et al.*,

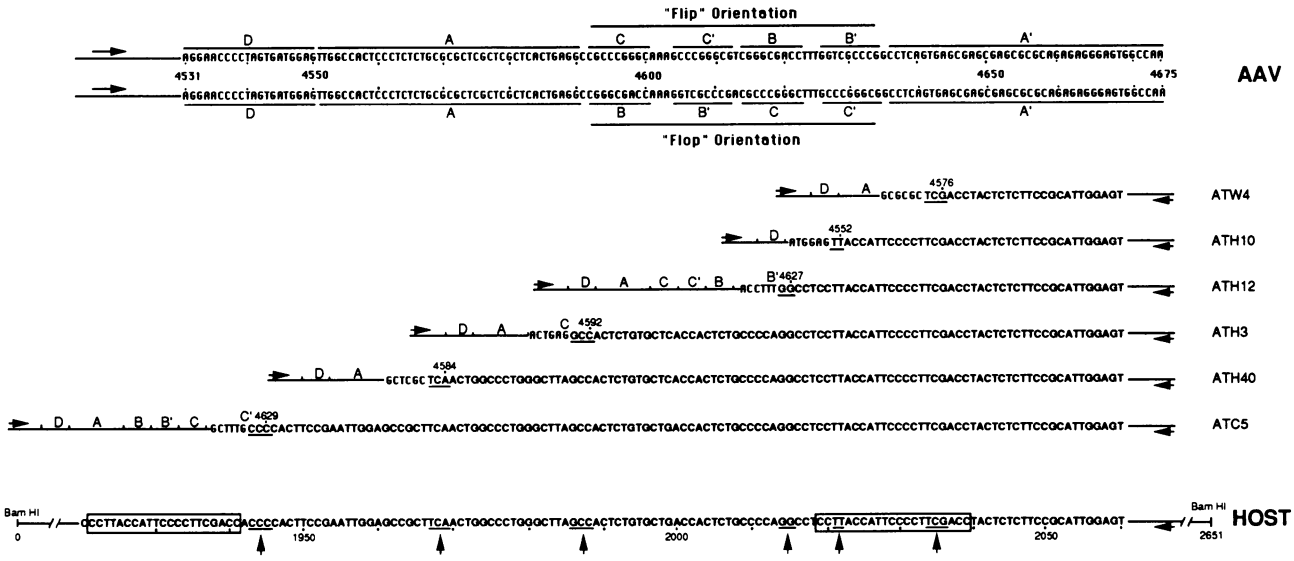

Fig. 8. Analysis of PCR amplified AAV−host junctions. Latent cell lines H3 and C11, and human diploid cells WI38 were subjected to PCR amplification using a chromosome 19-specific oligo (◄) in combination with AAV-specific primers (►). Asymmetrically amplified DNA (H3 cell line ATH10, ATH12) or amplified DNA which had been cloned (WI38 DNA ATW4, H3 DNA ATH3 and ATH40, and C11 DNA ATC5) were sequenced and compared with host chromosome 19 sequences ('HOST'). The AAV 145 bp terminal repeat (TR) sequences ('AAV') are presented with the nucleotide numbering for the right end. Palindromic sequences within the TRs are indicated by the lettering D, A, C, C′, B, B′, A′ and the two possible sequence orientations Flip and Flop are indicated. For each viral/junction sequence isolated, the amount of viral terminal sequences are illustrated by lettering D, A, C, C′, B, B′, A′ and the last six unique nucleotides. DNA overlap between virus and cellular sequences is indicated by underline, and the last discernible viral nucleotide is noted by its position in the AAV sequence. The 21 bp direct repeat of the preintegration sequence is shown in boxes, with the position of the viral insertions indicated by (▲).

1987; Garcia *et al.*, 1986; Peters *et al.*, 1986; Nusse and Varmus, 1982; Berns *et al.*, 1989). The proviral DNA found in tumors appears to be associated preferentially with cellular proto-oncogenes and chromosome fragile sites (van Lohuizen *et al.*, 1989). For example, in Cas-Br-E murine leukemia virus-induced tumors, a cluster of 16 integrations, with the same orientation, in < 100 bp was observed (Bergeron *et al.*, 1991). This specificity appears to reflect molecular constraints required to produce an oncogenic effect, such as the insertion of a viral regulatory element near a cellular proto-oncogene which results in activation, or removal of negative elements effected by insertion, or both (Bergeron *et al.*, 1991). AAV has no apparent impact on cell growth or morphology (Handa *et al.*, 1977). Thus, the specificity we have observed with AAV integration was not selected for by cell transformation or by any other means. Similar unselected analysis of retrovirus insertions indicates that most occur at different sites in the host genome (Hughes *et al.*, 1978; Steffen and Weinberg, 1978; Ringold *et al.*, 1979), although retrovirus integraton often results in provirus insertions preferentially near DNase 1-hypersensitive sites and apparently transcriptionally active regions (Rohdewohld *et al.*, 1987; Scherdin *et al.*, 1990; Vijaya *et al.*, 1986). While we see no evidence for expression of the AAV target site in HeLa cells, we have not yet tested a large portion of the locus, and examination of the chromatin configuration has not been carried out. What role if any virus integration may play in the host remains to be determined.

Studies by Shih *et al.* (1988) using Rous sarcoma virus (RSV) have demonstrated the use of highly preferred target sites for integration in the absence of selective pressure. These target sites appear to be present about once in 2.5 million bp (on average) or 800 such sites per avian genome. Compared with random sites of integration, use of these specific sites occurred at a frequency more than a million-fold greater than expected, suggesting a mechanism of preferred integration. Another retrovirus, Baboon endogenous virus (BEV), appears to integrate more specifically (Cohen and Murphey-Corb, 1983). BEV requires the presence of the BEVI (baboon endogenous virus infection) locus located on chromosome 6 for virus replication in human cells (Lemons *et al.*, 1977, 1978). Although documented as a relatively inefficient process (Huang *et al.*, 1989), after many rounds of DNA replication, integration occurs resulting in approximately five copies/cell. Characterization of this retrovirus integration revealed a common cellular sequence in the proximity of one of the junctions between the cellular DNA and the integrated virus (Cohen and Murphey-Corb, 1983). From these results the authors proposed that a short cellular sequence repeated on chromosome 6 and separated by unique DNA sequences presents a high affinity target for the BEV integration in human cells. AAV integration appears to be more specific compared with RSV and more efficient compared with BEV: 11 of 11 integrated genomes were observed to be associated with a single locus. Interestingly, the AAV terminal nucleotides are the same as those of retroviruses (5′-TG . . . CA-3′). Whether additional features of these viral integration mechanisms are similar can now be addressed with the AAV system described here.

Another documented possibility for targeted integration involves chromosomal fragile sites. The human homologue of the int-2 locus, a murine mammary tumor virus integration site, is located at 11q13, a hereditary site of chromosome fragility (Berger *et al.*, 1986; Casey *et al.*, 1986; Sutherland and Mattei, 1987). Evidence for human papilloma virus (HPV) DNA integration into sites of known heritable chromosome fragility has also been demonstrated in two cervical carcinoma cell lines (Popescu *et al.*, 1987a,b). Recently, Romani *et al.* (1990) showed preferential

integration of an Ad5−SV40 hybrid into a highly recombinogenic region of human chromosome 1. We assayed all available cell lines from the human genetic mutant cell repository (GM89, GM271 and GM4188) known for translocation between the chromosome 19 q region and other chromosomes and saw no evidence for rearrangement of the AAV target site (X.Zhu and R.J.Samulski, unpublished results). While we cannot rule out the presence of an undetected fragile site at this location, these results indicate that AAV integration does not involve any of the break-points currently mapped to this region of chromosome 19.

To define better the AAV target sequences, we initiated studies to characterize the cellular preintegration site. Recently, we isolated two overlapping cosmid clones that hybridize to the H3 junction from chromosome 19 (N.Epstein and R.J.Samulski, unpublished results). The sequence of a 2651 bp *Bam*HI subclone shows two direct repeats of 21 bp flanking the predominant viral integration site (Figure 8). Further analysis of this region should be illuminating for both the integration mechanism and for understanding what role if any this virus may play in the host cell. This host sequence may also facilitate the establishment of an *in vitro* integration system for AAV.

AAV vectors containing only the viral terminal repeat sequences carrying heterologous internal genes integrate at high frequency (Hermonat and Muzyczka, 1984; Tratschin *et al.*, 1985; McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). In our preliminary studies, similar mutant AAV viruses have been analyzed for targeted integration. Initial results suggest that AAV sequences in addition to the terminal repeats may be required for site-specific integration (X.Xiao, X.Zhu and R.J.Samulski, manuscript in preparation). One possibility is that AAV *trans*-acting factors are required to achieve site-specificity. Numerous labs have shown that the retrovirus-encoded INT protein is essential when establishing the proviral state (for reviews see Skalka, 1988 and Brown *et al.*, 1989). Recently, *in vitro* systems have been developed to characterize this molecular event in detail (see review by Grandgenett and Mumm, 1990). AAV proteins which may be potential candidates for a role in virus integration are the AAV 78 kDa and 68 kDa replication (rep) proteins. These proteins, which are essential for viral DNA replication *in vivo*, have been characterized biochemically *in vitro* (Snyder *et al.*, 1990a, b; Im and Muzyczka, 1989, 1990 and 1991). They interact with AAV terminal sequences only if the termini exist in a T shaped secondary conformation (Im and Muzyczka, 1989; Ashktorab and Srivastava, 1989), a conformation which is believed to exist in virion DNA. Both rep 78 and rep 68 contain ATP-dependent site-specific and strand-specific endonuclease activities which recognize a sequence in the AAV terminal repeat referred to as the terminal resolution site (trs) (Snyder *et al.*, 1990b; Im and Muzyczka, 1991). DNA helicase activity is also associated with these viral proteins (Im and Muzyczka, 1990). Both enzymatic activities are required in the first steps of AAV DNA replication, and we are currently testing the possibility that they are involved in integration by assaying AAV rep mutants for site-specific integration.

AAV as a vector

Beyond the basic interest in the molecular mechanism of site-specific integration, the AAV system described here has tremendous potential for applications in human gene therapy approaches. AAV vectors have been developed which appear to provide a viable alternative to currently utilized vector systems (Hermonat and Muzyczka, 1984; Tratschin *et al.*, 1985; LaFace *et al.*, 1988; Mendelson *et al.*, 1988; Srivastava *et al.*, 1989; Samulski *et al.*, 1989). These AAV vectors make use of an infectious recombinant clone comprised of the *cis*-acting AAV terminal repeats, and a helper plasmid that carries the essential AAV *trans*-acting factors (Hermonat and Muzyczka, 1984; Tratschin *et al.*, 1985; McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). Recent results demonstrate transduction frequencies as high as 80% (McLaughlin *et al.*, 1988; Samulski *et al.*, 1989). Thus, incorporation of site-specific integration into AAV vector schemes should provide tremendous advantages for targeted delivery of genes, particularly in the case of gene therapy where the possibility of detrimental effects resulting from random integration must be minimized.

Materials and methods

Cells and viruses

HeLa, Detroit 6 and CV1 monolayer cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% calf serum, 100 U/ml penicillin and 0.1 mg/ml streptomycin. Cells were transfected using the DEAE−dextran method (McCutchan and Pagano, 1968) as previously described (Samulski *et al.*, 1983). Adenovirus type 5 stocks and helper virus infections were carried out as described (Samulski *et al.*, 1987). The AAV virus lysate was heated at 56°C for 30 min to inactivate Ad. The latently infected cell lines were established as described in Samulski *et al.* (1989).

DNA cloning, sequencing and filter binding assay

High molecular weight DNA was extracted as described (Sambrook *et al.*, 1989), digested with restriction endonucleases and subjected to the DNA fragment retention assay essentially as described by Fuller *et al.* (1984). 20 μg digested genomic DNA were incubated with 25 ng repressor in 25 μl binding buffer (10 mM Tris−Cl pH 7.4, 50 mM KCl, 0.1 mM EDTA, 1 mM $CaCl_2$, 10 mM DTT, 60 mg/ml BSA) on ice. After incubation for 15 min, reactions were filtered under gentle suction (0.4 ml/min) on pre-soaked filters (Millipore, Type HA, 0.45 μm). The filters were washed under the same suction with 250 μl washing buffer (10 mM Tris−Cl pH 7.4, 50 mM KCl, 0.1 mM EDTA, 1 mM $CaCl_2$, 10 mM DTT, 5% DMSO) at room temperature. The DNA was eluted by soaking filters in 250 μl elution buffer (50 mM Tris−Cl pH 7.4, 10 mM EDTA, 0.5 M sodium acetate, 0.5% SDS) at 65°C for 30 min with constant shaking. The eluted DNA fragments were either analyzed on agarose gels or ligated to plasmid vector pGEM (Promega), transformed into *Escherichia coli* (SURE strain, Stratagene), and screened for positive clones by hybridizing to ^{32}P-labeled AAV terminal repeat probe. PCR products were cloned into plasmid pCR™ 1000 according to the manufacturer's instructions (Invitrogen). Plasmid preparations, restriction enzyme digestions and other routine DNA manipulations were performed using standard procedures (Sambrook *et al.*, 1989). DNA was sequenced by the chain termination method (Sanger *et al.*, 1977).

Southern hybridization analysis and polymerase chain reaction

High molecular weight chromosomal DNA was digested with various restriction endonucleases and electrophoresed in a 0.7% agarose gel, transferred to Gene Screen Plus (New England Nuclear) as recommended by the manufacturer and hybridized at 68°C with random primed ^{32}P-labeled probes. Filters were washed in 2 × SSC, 1% SDS at 65°C for 30 min, and then 0.2 × SSC, 1% SDS at room temperature for another 30 min, then exposed to X-ray film with an intensifying screen overnight.

PCR was carried out using 100 ng of genomic DNA added to 25 μl of total PCR reaction cocktail (Promega), consisting of 50 mM KCl, 10 mM Tris−HCl pH 9.0, 1.5 mM $MgCl_2$, 0.01% gelatin (w/v), 0.1% Triton X-100, 200 mM each of dATP, dGTP, dCTP and TTP, 0.5 U of Taq polymerase (Promega) and 5 pmol of each amplification primer. For screening AAV positive cell lines, p1: 5′-GCGCATCAGAATTGGGATTC-3′ hybridized to AAV nt960−941, p2: 5′-GAACGCGCAGCCGCC-3′ hybridized to AAV nt305−320. For amplifying the AAV−cellular junction, p1: 5′-ATAAGTAGCATGGCGGGTTA-3′ hybridized to AAV nt4498−4517, p2: CGCATAAGCCAGTAGAGC-3′ hybridized to chromosome 19

sequences. The reaction proceeded for 30 cycles in a PCR 1000 Thermal Cycler (Perkin-Elmer Cetus) programmed for 1 min at 94°C, 2 min at 50°C, 1 min at 72°C per cycle. The PCR products were analyzed on a 1.2% agarose gel, stained with ethidium bromide then transferred and hybridized with AAV or junction-specific probe.

In situ hybridization

Two plasmids, psub 201 (Samulski *et al.*, 1987) and plasmid sst repeat (Epstein *et al.*, 1987) containing the AAV genome and a fragment localized to chromosome 19 respectively, were labeled by digoxigenin-11-dUTP (Boehringer Mannheim) and/or biotin-16-dATP (BRL, Bethesda, MD) according to manufacturer's protocols and used as probes in chromosome analysis. The chromosome spreads from AAV latently infected cells were prepared according to typical cytogenetic techniques. The fluorescent *in situ* hybridization protocol described by Lawrence *et al.* (1988) with slight modification was used to detect AAV provirus in latent cells. For each slide, 30 μl hybridization mix containing 150 ng AAV probe, 50 ng chromosome 19 probe, or both were mixed with 15 μg sonicated salmon sperm DNA, 60 μg yeast tRNA, 50% formamide, 10% dextran sulfate, 0.1% BSA and 2 × SSC, denatured at 70°C for 10 min and then applied to the 37°C prewarmed slide. The samples were hybridized overnight and washed according to the procedure of Lawrence *et al.* (1988).

To locate the hybridization signals, anti-digoxin monoclonal antibody (Sigma Chemical) and a Cy3 conjugated sheep anti-mouse antibody (gift from Dr Alan Waggoner, Carnegie Mellon University Fluorescence Center) were used to detect AAV and chromosome 19 probe, respectively. The slides were incubated with avidin (5 μg/ml) and anti-digoxin antibody (1:25000 dilution) in 4 × SSC, 1% BSA, 0.1% Tween 20 for 30 min at 37°C and washed three times in 4 × SSC, 0.1% Tween 20 at room temperature. Then the slides were incubated with Cy3 conjugated sheep anti-mouse antibody in 4 × SSC, 1% BSA, 0.10 Tween 20 for 30 min at 37°C and washed as before. Finally, the samples were counterstained with DAPI (diamino-phenylindole, 0.1 μg/ml in PBS) or propidium iodide (0.1 μg/ml PBS), for 5 min and mounted in an antifade solution (Johnson and Aroujo Nogueira, 1981).

A Zeiss Axiophot epifluorescent microscope equipped with an image processor or a Nikon labophot-2-epifluorescent microscope was used to visualize FITC (exciter 485 nm, dichroic 510, barrier 420), Cy3 (exciter 546 nm, dichroic 580, barrier 590) and DAPI (exciter 365 nm, dichroic 390, barrier 420). Fluorescence observed using the Nikon labophot-2 was photographed using Kodacolor Gold ASA400 film.

Acknowledgements

This manuscript is dedicated to Ted P.Samulski. We thank Dr Terry Van Dyke for critical reading and comments which improved the manuscript. AAV primers for PCR analysis were kindly provided by Nick Muzyczka. Cell line 5HL94 was kindly supplied by Dr Larry Thompson (Lawrence Livermore Laboratory). Lambda operator oligo was a kind gift from Dr Raquel Sussman. *In situ* chromosome images taken on a Zeiss Axiophot epifluorescent microscope equipped with an image processor and the use of Cy3 conjugated sheep anti-mouse antibody were kindly provided by Dr Alan Waggoner, Carnegie Mellon University Fluorescence Center, with technical help from Kevin Ryan. Thanks to Christine Berliner and Crystal Petrone for proofreading and preparation of the manuscript. This study was supported by National Institutes of Health grant AI 25530-03.

References

Ashktorab,H. and Srivastava,A. (1989) *J. Virol.*, **63**, 3034–3039.
Atchinson,R.W., Castro,B.C. and Hammon,W.McD. (1965) *Science*, **149**, 754–756.
Berger,R., Bloomfield,C.D. and Sutherland,G.R. (1986) *Cytogenet. Cell Genet.*, **40**, 490–535.
Bergeron,D., Poliquin,L., Kozak,C.A. and Rassart,E. (1991) *J. Virol.*, **65**, 7–15.
Berns,K.I. (1984) *The Parvoviruses*. Plenum Publishing Corp., New York.
Berns,K.I. and Bohenzky,R.A. (1987) *Adv. Virus. Res.*, **32**, 485.
Berns,K.I., Cheung,A., Ostrove,J. and Lewis,M. (1982) In Mahy,B.W.J., Minson,A.C. and Darby,G.K. (eds), *Virus Persistence*. Cambridge University Press, Cambridge, UK, pp. 249.
Berns,A., Breuer,M., Verbeek,S. and van Lohuizen,M. (1989) *Int. J. Cancer Suppl.*, **4**, 22–25.
Bohenzky,R.A., Lefebvre,R.B. and Berns,K.I. (1988) *Virology*, **166**, 316–327.
Brook,J.D., Harley,H.G., Walsh,K.V., Rundle,S.A., Siciliano,M.J., Harper,P.S. and Shaw,D.J. (1991) *J. Med. Genet.*, **28**, 84–88.
Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2525–2529.
Buller,R.M.L., Janik,E., Sebring,E.D. and Rose,J.A. (1981) *J. Virol.*, **40**, 241–247.
Casey,G., Smith,R., McGillivray,D., Peters,G. and Dickson,C. (1986) *Mol. Cell. Biol.*, **6**, 502–510.
Cheung,A., Hoggan,M.D., Hauswirth,W.W. and Berns,K.I. (1980) *J. Virol.*, **33**, 739–748.
Cohen,J.C. and Murphey-Corb,M. (1983) *Nature*, **301**, 129–132.
Deiss,V., Tratschin,J.-D., Weitz,M. and Siegl,G. (1990) *Virology*, **175**, 247–254.
Epstein,N.D., Karlsson,S., O'Brien,S., Modi,W., Moulton,A. and Nienhuis,A.W. (1987) *Nucleic Acids Res.*, **15**, 2327–2341.
Fuller,R.S., Funnell,B.E. and Kornberg,A. (1984) *Cell*, **38**, 889–900.
Gallahan,D., Kozak,C. and Callahan,R. (1987) *J. Virol.*, **61**, 218–220.
Garcia,M., Wellinger,R., Vessaz,A. and Diggelmann,H. (1986) *EMBO J.*, **5**, 127–134.
Geissler,E.N., Liao,M., Brook,J.D., Martin,F.H., Zsebo,K.M., Housman,D.E. and Galli,S.J. (1991) *Somat. Cell Mol. Genet.*, **17**, 207–214.
Grandgenett,D.P. and Mumm,S.R. (1990) *Cell*, **60**, 3–4.
Handa,H., Shiroki,K. and Shimojo,H. (1977) *Virology*, **82**, 84–88.
Hermonat,P.L. and Muzyczka,N. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6466–6470.
Hoggan,M.D. (1965) *J. Virol.*, **1**, 171–180.
Hoggan,M.D., Thomas,G.F. and Johnson,F.B. (1972) *Proceedings of the Fourth Lepetit Colloquium, Coyocac, Mexico*. North-Holland Publishing Co., Amsterdam.
Huang,L.-H., Silberman,J., Rothschild,H. and Cohen,J.C. (1989) *J. Biol. Chem.*, **264**, 8811–8814.
Hughes,S.H., Shank,P.R., Spector,D.H., Kung,H.-J., Bishop,J.M., Varmus,H.E., Vogt,P.K. and Breitman,M.L. (1978) *Cell*, **15**, 1397–1410.
Im,D.-S. and Muzyczka,N. (1989) *J. Virol.*, **63**, 3095–3104.
Im,D.-S. and Muzyczka,N. (1990) *Cell*, **61**, 447–457.
Im,D.-S. and Muzyczka,N. (1991) *J. Virol.*, in press.
Johnson,A.D., Poteete,A.R., Lauer,G., Sauer,R.T., Ackers,G.K. and Ptashne,M. (1981) *Nature*, **294**, 217–223.
Johnson,G.D. and Aroujo Nogueira,G.M. (1981) *J. Immunol.*, **43**, 349–350.
Jourdan,M., Jousset,F.X., Gervais,M., Skory,S., Bergoin,M. and Dumas,B. (1990) *Virology*, **179**, 403–409.
Kotin,R.M., Siniscalco,M., Samulski,R.J., Zhu,X., Hunter,L., Laughlin, C.A., McLaughlin,S., Muzyczka,N., Rocchi,M. and Berns,K.I. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2211–2215.
LaFace,D., Hermonat,P., Wakeland,E. and Peck,A. (1988) *Virology*, **162**, 483.
Laughlin,C.A., Tratschin,J.D., Coon,H. and Carter,B.J. (1983) *Gene*, **23**, 65–73.
Laughlin,C.A., Cardellichio,C.B. and Coon,H.C. (1986) *J. Virol.*, **60**, 515–524.
Lawrence,J.B., Villnave,C.A. and Singer,R.H. (1988) *Cell*, **52**, 51–61.
Lemons,R.S., O'Brien,R.E. and Sherr,C.J. (1977) *Cell*, **12**, 251–262.
Lemons,R.S., Nash,W.G., O'Brien,S.J., Benveniste,R.E. and Sherr,C.J. (1978) *Cell*, **14**, 995–1005.
McCutchan,J.H. and Pagano,J.S. (1968) *J. Natl. Cancer Inst.*, **41**, 351–357.
McLaughlin,S.K., Collis,P., Hermonat,P.L. and Muzyczka,N. (1988) *J. Virol.*, **62**, 1963–1973.
Mendelson,E., Smith,M.G., Miller,I.L. and Carter,B.J. (1988) *Virology*, **166**, 612–615.
Merchlinsky,M.J., Tattersall,P.J, Leary,J.J., Cotmore,S.F., Gardiner,E.M. and Ward,D.C. (1983) *J. Virol.*, **47**, 227–232.
Nusse,R. and Varmus,H.E. (1982) *Cell*, **31**, 99–109.
Peters,G., Lee,A.E. and Dickson,C. (1986) *Nature*, **320**, 628–631.
Popescu,N.C, Amsbaugh,S.C. and DiPaolo,J.A. (1987a) *J. Virol.*, **61**, 1682–1685.
Popescu,N.C., DiPaolo,J.A. and Amsbaugh,S.C. (1987b) *Cytogenet. Cell Genet.*, **44**, 58–62.
Ringold,G.M., Shank,P.R., Varmus,H.E., Ring,J. and Yamamoto,K.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 665–669.
Rohdewohld,H., Weiher,H., Reik,W., Jaenisch,R. and Breindl,M. (1987) *J. Virol.*, **61**, 336–343.
Romani,M., De Ambrosis,A., Alhadeff,B., Purrello,M., Gluzman,Y. and Siniscalo,M. (1990) *Gene*, **95**, 231–241.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Samulski,R.J., Berns,K.I., Tan,M. and Muzyczka,N. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2077–2081.
Samulski,R.J., Srivastava,A., Berns,K.I. and Muzyczka,N. (1983) *Cell*, **33**, 135–143.
Samulski,R.J., Chang,L.S. and Shenk,T. (1987) *J. Virol.*, **61**, 3096–3101.
Samulski,R.J., Chang,L.S. and Shenk,T. (1989) *J. Virol.*, **63**, 3822–3828.
Sanger,F., Nicklen,S. and Coulson,A.R (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Scherdin,U., Rhodes,K. and Breindl,M. (1990) *J. Virol.*, **64**, 907–912.
Shih,C.C., Stoye,J.P. and Coffin,J.M. (1988) *Cell*, **53**, 531–537.
Skalka,A.M. (1988) In Smith,G.R. and Kucherlapati,R. (eds), *Genetic Recombination.* American Society for Microbiology Publications, Washington, DC, pp. 701–724.
Snyder,R.O., Im,D.-S. and Muzyczka,N. (1990a) *J. Virol.*, **64**, 6204–6213.
Snyder,R.O., Samulski,R.J. and Muzyczka,N. (1990b) *Cell*, **60**, 105–113.
Srivastava,C.H., Samulski,R.J., Lu,L., Larsen,S.H. and Srivastava,A. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 8078–8082.
Steffen,D. and Weinberg,R.A. (1978) *Cell*, **15**, 1003–1010.
Sutherland,G.R. and Mattei,J.F. (1987) *Cytogenet. Cell Genet.*, **46**, 316–324.
Tratschin,J.D., Miller,I.L., Smith,M.G. and Carter,B.J. (1985) *Mol. Cell Biol.*, **5**, 3251.
Vijaya,S., Steffen,D.L. and Robison,H.L. (1986) *J. Virol.*, **60**, 683–692.
Warburton,D., Gerson,S., Yu,M.-T., Jackson,C., Handelin,B. and Housman,D. (1990) *Genomics*, **6**, 358–366.
van Lohuizen,M., Breuer,M. and Berns,A. (1989) *EMBO J.*, **8**, 133–136.

Received on May 28, 1991; revised on August 6, 1991

Note added in proof

Since submission of this manuscript, Berns *et al.* have also visualized AAV integration on the q arm of chromosome 19 [Berns *et al.* (1991) *Genomics*, **10**, 831–834].