Identification of a DNA-Binding Domain in the Amino Terminus of Adeno-Associated Virus Rep Proteins

ROLAND A. OWENS,* MATTHEW D. WEITZMAN, SIRKKA R. M. KYÖSTIÖ, AND BARRIE J. CARTER†

Laboratory of Molecular and Cellular Biology, Building 8, Room 304, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received 2 September 1992/Accepted 12 November 1992

The Rep78 and Rep68 proteins of adeno-associated virus (AAV) bind to the AAV terminal repeat hairpin DNA and are required for viral replication. We have expressed a series of mutant rep genes from the human immunodeficiency virus type 1 long terminal repeat promoter in human 293 cells and in an in vitro transcription-translation system. Mutant proteins were analyzed for AAV hairpin DNA binding and AAV terminal resolution functions. Deletion of amino acid residues 523 through 621 of Rep78 had no effect on these functions. Amber mutant Rep proteins truncated at either amino acid 237 or amino acid 243 showed no detectable hairpin DNA binding or terminal resolution activity. A frameshift mutant Rep protein which contained Rep78 amino acids 1 through 241 lacked terminal resolution functions but bound specifically to the AAV hairpin DNA. The carboxyl-terminal missense sequence in this mutant appeared to have complemented an AAV-specific DNA-binding domain within the amino terminus of the Rep protein. A mutant Rep protein in which methionine 225 of Rep78 was deleted (M225dl) was reduced threefold in AAV hairpin binding and had no terminal resolution functions. A mutant Rep protein in which a glycine was substituted at position 225 (M225G) was fully functional in these assays. When M225dl extract was mixed with wild-type Rep78 extract, AAV terminal resolution by Rep78 was inhibited. These results suggest that the amino-terminal portion of Rep78 and Rep68 contains a domain which can direct binding to AAV terminal hairpin DNA and that elements within the central region of the protein stabilize binding.

Adeno-associated virus type 2 (AAV) is a human parvovirus which usually requires adenovirus or herpesvirus as a helper to replicate efficiently (4). AAV has a single-stranded, linear DNA genome with terminal repeats at each end which function as origins of replication (13, 25, 30). The left half of the AAV genome contains the rep gene open reading frame, which encodes at least four overlapping Rep proteins (Rep78, Rep68, Rep52, and Rep40) that are required for efficient AAV DNA replication and the production of single-stranded progeny genomes for encapsidation (5, 8, 9, 14, 21, 25, 28, 31, 34). The largest of these four overlapping proteins, Rep78, contains 621 amino acid residues. Rep78 and Rep68 are translated from unspliced and spliced RNA transcripts, respectively, initiated from the p1 promoter. Rep52 and Rep40 are translated from unspliced and spliced transcripts, respectively, initiating from the p19 promoter (for a review, see reference 6). Rep78, Rep68, and Rep40 appear as doublets on denaturing polyacrylamide gels (16, 17).

The AAV terminal repeats are self-complementary and fold over into hairpin structures which provide primers for DNA synthesis (13, 28, 30). It is then necessary to resolve the ends by making a strand-specific, site-specific endonuclease cut at the terminal resolution site (TRS) and then unwinding the hairpin, so that the ends may be replicated and converted to an extended linear duplex (16, 26, 27). This process has been termed terminal resolution (27). Both Rep68 and Rep78 bind specifically to the AAV terminal hairpin DNA (2, 15–17, 23) and can mediate nicking at the TRS in vitro (16, 17, 26). Terminal resolution can be reconstituted in vitro with extracts from uninfected HeLa cells by the addition of purified Rep68 (26) or by the addition of nuclear extracts which contain Rep78 (22).

The two smaller Rep proteins, Rep52 and Rep40, appear to be required for the production of single-stranded progeny genomes for encapsidation but not for duplication of replicative-form AAV DNA (8). Consistent with this, Rep52 and Rep40 can neither bind to AAV hairpin DNA nor perform terminal resolution functions (17). A mutant Rep protein which contained only the amino-terminal 237 amino acid residues—including all of the region which is contained in Rep78 and Rep68 but not in Rep52 or Rep40—has no AAV DNA binding or terminal resolution activities (22, 23). These observations implied that multiple regions of Rep78 and Rep68 were required for these functions. A previously reported mutational analysis of the rep gene also supported this hypothesis (20). In this study, we analyzed a series of Rep mutants for AAV hairpin DNA binding, AAV terminal resolution, and TRS endonuclease activity to define further these functional regions. We present the first observation of a mutant Rep protein lacking the central portion of Rep78 and Rep68 which can bind specifically to AAV hairpin DNA. We also show that the deletion of the methionine at the border between the amino-terminal portion of Rep78 and the central region resulted in a protein which bound weakly to the hairpin and was negative trans dominant for terminal resolution functions.

MATERIALS AND METHODS

Human cells and viruses. Stocks of AAV and adenovirus type 5 were grown and assayed as described previously (5). Adenovirus-transformed human cells (293 cells) were grown in Eagle minimum essential medium (Quality Biological Inc., Gaithersburg, Md.) supplemented with 2 mM l-glutamine,
penicillin-streptomycin-neomycin antibiotic mix, and 10% fetal bovine serum (GIBCO/BRL, Gaithersburg, Md.). 293 cells were infected with adenovirus at a multiplicity of 5 PFU per cell and with AAV at a multiplicity of 20 infectious units per cell. The cells were harvested 35 h after infection.

**Plasmids for transfection.** Plasmids pHIVrep, pHIVrepam, pARtat, and pHIVrepNTP have been described previously (1, 23). pHIVrep contains the AAV rep gene expressed from the long terminal repeat promoter of the human immunodeficiency virus (HIV-LTR) and produces Rep78 (1). pHIVrepam is identical to pHIVrep, except that it contains an amber codon at the site which normally encodes serine 238 of Rep78 (1). pHIVrepam encodes the protein designated Repam238 (22). pHIVrepNTP is identical to pHIVrep, except that it contains mutations which change lysine 340 of Rep78 to a histidine and encodes the protein designated K340H (23). pARtat contains the first exon of the sequence encoding the HIV transactivator protein (Tat) expressed from the HIV-LTR (11). Cotransfection with pARtat has been demonstrated to stimulate the production of Rep78 from pHIVrep (1).

Frameshift mutations were generated by cutting pHIVrep with BamHI (for protein Bafl), BstEII (for Bsf2), HindIII (for Hf6), or KpnI (for Kfs1 and Kfs2), which cut only once within the rep gene (Fig. 1). To make Kfs1, the KpnI overhang within the rep gene was ligated to the overhang from the KpnI site between the rep gene and the polyadenylation site of pHIVrep. Otherwise, overhangs were blunt-ended by filling in 5′ overhangs with Klenow fragment (New England Biolabs, Beverly, Mass.) or by digesting 3′ overhangs with T4 DNA polymerase (New England Biolabs), and then the plasmids were sealed with T4 DNA ligase (GIBCO/BRL).

To construct HIV-LTR plasmids encoding the mutant Rep proteins Ba-Bsd1 and Baf2, pHIVRep was simultaneously digested with BamHI and BstEII. The overhangs were filled in, and the larger fragment was ligated. This produced an in-frame deletion (Ba-Bsd1) of the region encoding amino acid residues 244 through 460, inclusive, of the Rep78 protein. With one isolate (Bafl), an unintended exonuclease event resulted in a frameshift mutation at the BamHI site. Table 1 shows the annealed oligonucleotide pairs inserted into the BamHI, BstEII, HindIII, and KpnI sites of pHIVRep to construct plasmids encoding the amber truncation mutant Rep proteins Basm244, Basm463, Ham523, and Kam531, respectively. In the case of Ham523, the HindIII overhangs were filled in prior to the insertion of the oligonucleotide pair. Each oligonucleotide contained amber codons in all three reading frames. The number in the designation of each amber mutant protein is the Rep78 codon which has been replaced with an amber codon.

pSK1 is identical to pHIVrep, except that the right-hand terminal repeat has been deleted by digestion with SnaBI and SpH1 followed by a fill-in reaction and blunt-end ligation. Plasmids pSK3 and pSK4 contain a rep gene with two different mutations. pSK3 has a deletion of the p1 ATG initiator methionine codon (M225df), while in pSK4 this codon has been replaced with a GGG glycine codon (M225g). pSK3 and pSK4 were constructed by replacing the 1.2-kb SstII-HindIII fragment of pSK1 with the corresponding SstII-HindIII fragment of pHIVrep. This plasmid encodes Rep68 (major splice) and Rep68′ (minor splice), respectively. Rep68′ contains an extra 9 amino acid residues near its carboxyl terminus which are not present in Rep68.

The presence of the mutations within the plasmids was confirmed by DNA sequencing (24) with a Sequenase DNA sequencing kit (United States Biochemical Corp., Cleveland, Ohio), according to the manufacturer's instructions. Plasmid pGEM4Z (Promega Corp., Madison, Wis.) was used as negative-control DNA.

**Plasmids for in vitro synthesis of Rep proteins.** The plasmid pMAT4 contains an AvaI fragment (AAV nucleotides 263 to 2233) from pNTC244 (7) inserted into the EcoRV site of the}

---

**FIG. 1.** AAV rep gene and Rep proteins analyzed in this study. The initiation codons for the four major Rep proteins, as well as restriction endonuclease sites used to make mutations, are shown at the bottom. The numbers immediately below are the amino acid residue numbers for Rep78 (1 through 621). The caret indicates the positions of the splice donor and major (solid line) and minor (dashed line) splice acceptors. NTP indicates the conserved lysine residue (K340) within the putative nucleoside triphosphate binding site. The lower portion shows the wild-type and mutant Rep proteins studied. The open and stippled rectangles represent amino acid sequences found within wild-type Rep proteins and missense amino acid sequences, respectively. The designation of each protein and the Rep78 amino acids contained within each are shown at the right. An asterisk indicates that the sequence contains an internal mutation.
vector pBluescriptII SK+ (Stratagene, La Jolla, Calif.) and encodes wild-type type Rep78. Mutant Rep78 plasmids pMAT6, pMAT7, pMAT8, and pMAT9 were also constructed in pBluescriptII SK+ by inserting the corresponding AvaI fragments from pNTC3 (7), pNCTC20 (8), pNCTC23 (9), and pNCTC4 (8), respectively. These plasmids encode mutant Rep  
proteins Repam238, M225G, K340H, and M225Dl, respectively. Plasmid pMAT21 contains the open reading frame for Rep68 (major splice) and was constructed by replacing the  
Psil fragment of pMAT4 with the corresponding fragment from pSK9.  
Plasmid pSK14 contains a gene for Rep40' (minor splice) expressed by a polymerase chain reaction with two primers. The 5′ primer contained an added HindIII site created by a point mutation 15 bases upstream from the p19 ATG codon. The 3′ primer has the natural HindIII site located downstream from  
p19. This polymerase chain reaction product containing AAV sequences from 973 to 1885 was digested with HindIII and inserted into HindIII cleaved pBenNCAT (11) to create pSK11. The chloropanocillic acid transferase gene of pSK11 was removed by digestion with Sall and was replaced with a Sall fragment from pSK10 to create pSK14. A HindIII fragment was replaced in pMAT4 and pMAT21 by the corresponding fragment from pSK14 to create plasmids encoding Rep52 (pMAT20) and Rep40 (pMAT22), respectively.

Transfection of DNA into 293 cells. 293 cells were transfected with plasmid DNA by the calcium phosphate coprecipitation method (12). Monolayers of cells at approximately 2×10^6 cells per 150-cm² flask were transfected with 15 μg of pRta plus 60 μg of the appropriate Rep plasmid. The cells were harvested 48 h after transfection for preparation of nuclear extracts.

Preparation of nuclear extracts. All operations were performed at 0 to 4°C unless otherwise specified. The cells were scraped into the culture medium, centrifuged for 5 min at 800 × g, washed with cold phosphate-buffered saline (Quality Biological Inc.), and resuspended in buffer (1 ml/150-cm² flask of cells) containing 0.25 M sucrose, 10 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 0.5% Nonidet P-40, 1.0 mM phenylmethylsulfonyl fluoride (PMSF), and 1.0 mM dithiothreitol (DTT). The cells were incubated for 10 min and then sheared by six strokes in a Dounce homogenizer with pestle B. The nuclei were collected by centrifugation for 5 min at 1,800 × g and then resuspended in 350 μl of freezing buffer (50 mM N-2-hydroxyethylpiperazine-N'2-2-ethanesulfonic acid–KOH [HEPES-KOH, pH 7.5], 10% sucrose, 0.25 trypsin inhibitory units of aprotinin [Sigma Chemical Co., St. Louis, Mo.]) per ml, 1 mM PMSF, 0.5 mM DTT), frozen on dry ice, and stored at -70°C. The nuclei were subsequently thawed on ice, and 4 M NaCl was added to a final concentration of 600 mM. The samples were rocked gently for 1 h and then centrifuged for 1 h at 100,000 × g. After the addition of glycerol to a final concentration of 20%, the supernatants were frozen on dry ice and stored at -70°C.

Immunoblotting. Samples of nuclear extracts were electrophoresed in a 12.5% polyacrylamide gel containing 0.1% sodium dodecyl sulfate (SDS). The proteins were transferred to a nitrocellulose membrane in a Multiphor II Nova Blot electrophoretic transfer apparatus (Pharmacia LKB Biotechnology, Piscataway, N.J.). Immunodetection was performed as described previously (21). The primary antibody was a rabbit antibody against Rep78 expressed in Escherichia coli (anti-Rep78.93) (34). The secondary antibody was 125I-goat anti-rabbit immunoglobulin G (500 Ci/mmol) (DuPont NEN, Boston, Mass.).

Gel mobility shift assays. Gel mobility shift assays were performed with 32P-end-labeled AAV terminal hairpin DNA as described before (15, 23). Nuclear extracts were diluted at least 10-fold in buffer A (25 mM Tris-HCl [pH 7.5], 0.1 mM EDTA, 1 mM DTT, 0.1 mM PMSF, 20% glycerol) to reduce the NaCl concentration. All samples within a given experiment were diluted to the same NaCl concentration. To control for the influence of cellular proteins, when necessary, the total amount of nuclear extract per sample was equalized with extract from 293 cells transfected with pGEM4Z plus pARtat, which was diluted to the appropriate salt concentration in buffer A. For experiments with antibodies, rabbit antibody against the S18K peptide (Rep78 amino acids 516 through 533) (21) or anti-Rep78.93 (34) was used. Antibodies were added in a volume of 1 μl.

Methylation interference assays. Methylation interference assays were performed as described previously (2) with the modifications indicated below. Briefly, 32P-end-labeled AAV terminal hairpin DNA was partially methylated with dimethyl sulfate (19). The methylated DNA was then incubated with the nuclear extracts and ran on a 4% polyacrylamide gel as described previously for the gel mobility shift assays (15). The radioactive bands containing free DNA or Rep-specific protein-DNA complexes were cut out and eluted by the crush and soak method (19). Each of the three shifted bands in the Rep78 lanes was processed separately. The DNA was then treated with piperidine and lyophilized (19). Samples containing equal numbers of counts were electrophoresed on a 6% polyacrylamide gel containing 7 M urea.

Terminal resolution assays. The in vitro terminal resolution assay was performed with the AAV no-end DNA substrate, HeLa cell cytoplasmic extracts, and nuclear extracts containing Rep proteins, as described previously (22, 27, 29).
RESULTS

In an attempt to identify functional domains within the Rep proteins, we generated a series of frameshift, substitution, internal deletion, and amber truncation mutants. Figure 1 shows the positions of these mutations within the rep gene.

Immunoblot analysis. In order to interpret better the results of the functional assays, we needed an estimate of the relative amounts of each mutant protein contained within the nuclear extracts. Figure 2 shows autoradiographs of immunoblots of wild-type and mutant Rep proteins produced in the HIV-LTR expression system. The primary antibody was rabbit antibody against Rep78 expressed in E. coli (anti-Rep78.93) (34).

A protein of the size predicted for the product of the unspliced Rep transcript from the HIV-LTR was the predominant Rep species detected in each of the nuclear extracts from transfected 293 cells. In the absence of adenovirus infection, AAV transcripts are inefficiently spliced (1, 32) and expression from the p19 promoter is low relative to that of the Tat-induced HIV-LTR promoter (1). Furthermore, Rep52 and Rep40 are isolated mainly in the cytoplasmic fraction (21). Rep68, Rep52, and Rep40 are therefore underrepresented in these nuclear extracts (22, 23).

With the exception of the mutants described below, the extracts produced Rep protein bands which bound an amount of 125I-antibody within fourfold of that bound by the Rep78 band. Frameshift mutant proteins Kf61 (data not shown) Hfs, and Bfs were as well as amber mutant Bsaam463 were significantly underproduced. Both the M225dl/K340H double-mutant protein (data not shown) and the K340H mutant protein were overproduced. Immunoblot analysis may underestimate the relative amounts of the shorter mutant proteins (Bfs, Bsaam463, Baf61, Baf62, Baam244, and Repam238), because they lack at least one polypeptide sequence which is known to be immunogenic (21).

Gel mobility shift assays. We first wished to determine which portions of the Rep proteins were necessary or sufficient for binding to AAV hairpin DNA. Figure 3 shows AAV hairpin DNA gel mobility shift assays with a representative group of mutant and wild-type Rep extracts. Rep78 showed the previously observed pattern comprising three shifted bands (23). Kf62 (data not shown), Rep68, Rep68', Kam531, Ham523, and M225G all showed similar patterns of
Each sample contained 4,000 cpm of 32P-labeled AAV terminal repeat hairpin DNA and either no extract or 62-nl equivalents of the nuclear extract from 293 cells transfected with pARat plus either control plasmid pGEM4Z (Gem), pHIVrep (Rep78), or a pHIVrep-derived plasmid encoding the indicated protein. Samples contained 1 µl of either bovine serum albumin (BSA, 5 mg/ml), anti-Rep78.93 antibody (78.93, diluted 1:3 in 5-mg/ml BSA), or undiluted anti-RepS18K antibody (S18K).

three bands. The migration rates of these shifted bands increased as the size of the protein decreased. The amount of band shift activity for each of these extracts was proportional to the amount of Rep protein detected by immunoblotting. No Rep-specific bands were seen with Kfs1, Hfs, Bfs5, Bsm463 (data not shown), Bafs2, Baam244, or Repam238 extracts. In the cases of Kfs1, Hfs, Bfs5, and Bsm463, the amounts of Rep protein were so low that an equivalent amount of Rep78 would not have shown any shifted bands. The use of higher amounts of extracts resulted in an acceptably high level of nonspecific DNA binding by cellular proteins. Ba-Bfs7 extract produced broad, faint smears of shifted DNA which were seen only in some experiments.

M225dl extract produced a single, weak, shifted band which comigrated with the fastest-migrating of the three Rep78 shifted bands. The M225dl/K340H protein also produced a single shifted band. Because the M225dl/K340H protein is overexpressed in this system, the single shifted band with this extract is much more intense. Both of these extracts shifted about one-third as much AAV hairpin DNA as Rep78 compared with the relative amounts of Rep proteins, as determined by immunoblots.

The Bafs1 extract produced a single shifted band. However, Bafs1 binding was similar to that previously described for Rep78 (23) in its preference for binding to the hairpin, as opposed to the linear conformation of AAV terminal repeat DNA, and in its failure to bind to a 155-bp fragment of the bacterial chloramphenicol acetyltransferase gene (data not shown).

Effects of antibodies on hairpin DNA binding. To confirm that a truncated Rep protein was actually a component of the Bafs1 shifted band, two different anti-Rep antibodies were added to the binding assay mixture after the addition of the labeled hairpin DNA. Anti-Rep78.93 (34), which recognizes at least one epitope in the amino-terminal portion of Rep78, reduced the intensity of the normal Bafs1 shifted band and produced a novel, more slowly migrating band (Fig. 4). Anti-Rep78.93 also supershifted a portion of the Rep78 shifted bands. Higher concentrations of anti-Rep78.93 blocked binding by both Rep78 and Bafs1 (data not shown).

An antibody made against a specific oligopeptide (anti-RepS18K) found near the carboxyl end of all four wild-type Rep proteins (21), but not in Bafs1, failed to affect Bafs1 binding. Anti-RepS18K did, however, supershift the Rep78 shifted bands, as reported previously (Fig. 4) (23).

Methylation interference assays. Since the mutant proteins Bafs1 and M225dl generated gel mobility shift patterns different from that of Rep78 (Fig. 3), methylation interference analysis was performed to determine whether the same G residues within the AAV hairpin were required for the binding of mutant and wild-type Rep proteins. Figure 5 shows that the methylation sensitivity patterns were virtually identical. Methylation of any one of four G residues within the A′ region of the AAV terminal repeat hairpin greatly inhibited binding by Rep78, Bafs1, or M225dl. Methylation of G residues within the D region of the hairpin clearly had no effect on the binding of any of these proteins. This result is consistent with the results of previous Rep footprinting analyses (2, 15). Each of the three components from the Rep78 shifted bands showed the same general pattern of methylation sensitivity, although the bottom band may have been slightly less sensitive to methylation in the A′ region.

Terminal resolution assays. The mutant and wild-type extracts from the HIV-LTR expression system were tested for their ability to complement uninfected HeLa cell cyto-
plasmic extracts in the in vitro terminal resolution assay (Fig. 6). Rep78 and Repam238 extracts were included as positive and negative controls, respectively (22). Kfs2, Hfs (data not shown), Rep78, Rep68, Rep68', Kam531, Ham523, and M225G extracts all had terminal resolution activity, as indicated by the appearance of the linearized AAV terminal DNA band (26, 27). For all extracts which showed activity, the level was roughly proportional to the amount of Rep protein detected in immunoblots. No terminal resolution activity was detected in Bsfs (data not shown), Bafs1, Bafs2, Baam244, Repam238, Ba-Bsdf, M225d1, M225d1/K340H, or Bsam463 extracts. With the Bsfs and Bsam463 samples, the small amount of Rep protein in these extracts makes the interpretation of these negative results ambiguous.

Mutants which were negative for terminal resolution activity were tested for the ability to inhibit terminal resolution by wild-type Rep78. K340H extract, which was previously shown to inhibit in vitro terminal resolution (22), was included as a control. In addition to K340H, both M225d1 and M225d1/K340H showed reproducible inhibition of in vitro terminal resolution by Rep78 (Fig. 7). The degree of inhibition was proportional to the ratio of mutant to wild-type protein (reference 22 and data not shown). As a result, greater inhibition was seen with K340H and M225d1/K340H, which were overproduced, than with M225d1. AAV genomes containing the K340H mutation inhibited the replication of AAV wild-type genomes in in vivo cotransfection assays in adenovirus-infected cells (9). An M225d1 mutant AAV genome exhibited a similar in vivo trans dominance (10).

Analysis of in vitro-synthesized Rep proteins. As an alternate expression system for Rep proteins, we used an in vitro transcription-translation system. Figure 8 shows a typical SDS-polyacrylamide gel analysis of 35S-labeled, in vitro-synthesized Rep proteins. This expression system had two advantages. First, the level of Rep protein production was much more consistent from mutant to mutant than is expression from the HIV-LTR with transfected DNA. Second, we did not have to assume equal cross-reactivity of an antibody to compare the relative amounts of all proteins produced. We used this system to produce Rep78, Rep68, Rep52, Rep40, Repam238, M225G, M225d1, and K340H.

As reported previously (20), in vitro-synthesized Rep78 bound specifically to AAV hairpin DNA (data not shown) and generated the site-specific, strand-specific nick in the TRS endonuclease assay (Fig. 9). As seen in earlier experiments (15–17, 20, 23) (Fig. 3), M225G, K340H, and Rep68 were positive for binding; M225d1 bound weakly; and Rep52, Rep40, and Repam238 were negative for binding (data not shown). Also in agreement with published results (16, 17), we found that Rep68 had TRS endonuclease activity and Rep52 and Rep40 did not (data not shown). Additionally,
M225G was positive and Repam238, K340H, and M225dl were negative for TRS endonuclease activity (Fig. 9).

In vitro-synthesized mutant Rep proteins which were negative for TRS endonuclease activity were tested for their ability to inhibit the endonuclease activity of Rep78 (Fig. 9). Both K340H and M225dl were inhibitory. Rep52, Rep40 (data not shown), and Repam238 were not inhibitory.

**DISCUSSION**

We generated a series of mutant Rep proteins in an attempt to define functional domains with respect to AAV hairpin DNA binding and terminal resolution functions. Table 2 summarizes the results of analyses of these mutants.

Since both Rep68 and Rep78 bind to AAV hairpin DNA (15–17, 20, 23), we had reason to believe that the portion of Rep78 encoded by the RNA spliced out of the message for Rep68 is not involved in hairpin DNA binding. The possibility remained, however, that the carboxyl-terminal portion of Rep68 encoded by RNA downstream from the splice might contain a DNA-binding element which could substitute for a DNA-binding element in the carboxyl-terminal portion of Rep78. We created a version of Rep68 with an alternative carboxyl terminus (Rep68'), as well as several mutant proteins which are truncated at or just before the region encoded by the RNA which is spliced out of the message for Rep68. Our data suggest that if the previously reported Rep68 doublet (16, 17) is due to this alternative splicing, there would be no functional differences between the two forms with respect to AAV hairpin DNA binding and terminal resolution activities. Furthermore, our analysis of the Ham523 mutant protein clearly shows that the amino-terminal 522 amino acid residues of Rep78 and Rep68 are sufficient for AAV hairpin DNA binding and terminal resolution functions.

Several lines of evidence indicated that the amino-terminal portion of Rep78 and Rep68 (amino acid residues 1 through 224) contain elements which are essential for AAV hairpin DNA binding. The first is that Rep78 and Rep68 can bind AAV hairpin DNA but that Rep52 and Rep40, which lack this region, cannot bind (15–17, 20, 23). The second line of evidence is that several 10-amino-acid deletions within this region of Rep78 result in the loss of both AAV hairpin DNA binding and TRS endonuclease activity (20). There have been no reports of any mutant Rep proteins which have lost the ability to bind but retained TRS endonuclease activity. This suggests that binding to the hairpin DNA is required for terminal resolution.

**TABLE 2. Wild-type and mutant AAV Rep proteins and their phenotypes**

<table>
<thead>
<tr>
<th>Rep protein</th>
<th>AAV hairpin DNA binding</th>
<th>Multiple shifted bands</th>
<th>Terminal resolution</th>
<th>TRS endonuclease</th>
<th>Negative trans dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep78</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>NA</td>
</tr>
<tr>
<td>Rep68</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>NA</td>
</tr>
<tr>
<td>Rep68'</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>Rep52</td>
<td>-</td>
<td>NA</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Rep40</td>
<td>-</td>
<td>NA</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Kfs1</td>
<td>?</td>
<td>?</td>
<td>ND</td>
<td>+</td>
<td>ND</td>
</tr>
<tr>
<td>Kfs2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Hfs</td>
<td>?</td>
<td>?</td>
<td>ND</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>Bfs</td>
<td>?</td>
<td>?</td>
<td>ND</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>Bafs1</td>
<td>-</td>
<td>-</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bafs2</td>
<td>-</td>
<td>-</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ham531</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ham523</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Bsam463</td>
<td>?</td>
<td>?</td>
<td>ND</td>
<td>-</td>
<td>?</td>
</tr>
<tr>
<td>Baam244</td>
<td>-</td>
<td>NA</td>
<td>ND</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Repam238</td>
<td>-</td>
<td>NA</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>K340H</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>M225G</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td>M225dl</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Ba-Bsd</td>
<td>+/-</td>
<td>+</td>
<td>-</td>
<td>ND</td>
<td>-</td>
</tr>
<tr>
<td>M225dl/K340H</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
<td>ND</td>
<td>+</td>
</tr>
</tbody>
</table>

*NA, not applicable; ND, not determined; ?, protein level too low to be certain; +/-, significantly lower than Rep78 per unit protein as determined by immunoblotting.

* Multiple band pattern is abnormal.
The ability of the Baf51 mutant Rep protein to bind to the AAV hairpin suggested that there is a region within the first 241 amino acid residues of Rep78 and Rep68 which can direct binding specifically to the AAV terminal repeat hairpin. However, Baf52 and Bafs2, which each contain the first 243 amino acid residues, did not bind AAV hairpin DNA. This apparent paradox may be resolved by an examination of the missense sequence of Baf51 (Table 3), which contains 18 amino acid residues. Four of these residues are positively charged, and none have negatively charged side groups. In contrast, the Baf52 missense region contains 35 amino acid residues, four of which are negatively charged. It seems unlikely that the Baf51 missense sequence alone is sufficient for specific binding to the AAV hairpin DNA, because this binding was sensitive to methylation of the same G residues as binding by wild-type Rep78. It is more probable that binding directed by elements within the first 241 amino acid residues of Rep78 is stabilized by the interaction of the positively charged missense amino acid residues with the phosphate backbone of the DNA. This suggests that the AAV hairpin DNA binding of Rep78 and Rep68 may be specified by the amino-terminal region and stabilized by elements within the central regions of the proteins.

The weakened DNA binding of the M225dl mutant Rep protein is consistent with this concept of a multipartite DNA binding region. This model also suggests how a relatively nonconservative alteration in the M225G mutant of the same methionine to a glycine (3) could have had no effect on any of the assayed Rep functions. Alteration of an amino acid between interacting domains may be much less deleterious than deletion of a residue. Similarly, the existence of an internal stabilization domain is supported by the poor DNA binding and lack of terminal resolution exhibited by the Baf5dl mutant. This is consistent with the previously reported observation that mutations between amino acids 415 and 490 of Rep78 eliminated AAV hairpin DNA binding (20).

Our data also suggest that multiple functional Rep proteins (Rep78 and/or Rep68) must bind to the AAV terminal repeat hairpin DNA for terminal resolution to take place. Whereas Rep78 or Rep68 produces three major shifted bands, the M225dl mutant Rep protein produces a single band which comigrates with the fastest-migrating Rep78 band. Baf51, which lacks the central part of the protein (as well as the carboxyl terminus), also forms only one shifted band with AAV hairpin DNA. The TRS endonuclease assay shows that both K340H and M225dl can inhibit Rep78 endonuclease activity in the presence of a gross excess of substrate. This is consistent with the hypothesis that multiple Rep molecules interacting with an individual hairpin DNA molecule are required for nicking and that if any one of these Rep molecules lacks TRS endonuclease activity, then the reaction is blocked.

In summary, of the 621 amino acid residues of the largest of the overlapping Rep proteins, Rep78, only residues 523 through 621 were clearly nonessential for AAV hairpin DNA binding and terminal resolution functions. The amino-terminal portion of Rep78 and Rep68 (amino acid residues 1 through 224), which is missing from Rep52 and Rep40, is essential for terminal resolution functions and probably contains elements which confer AAV hairpin DNA-binding specificity. The central portion of Rep78 (amino acid residues 225 through 529), which is present in all wild-type Rep proteins, is essential for terminal resolution functions and appears to contain elements which stabilize binding to AAV hairpin DNA. We also suggest that elements within the central region are responsible for the characteristic threeshifted-band pattern in the DNA binding assay. It remains to be rigorously proven whether the three shifted bands represent different numbers of Rep protein molecules (possibly held together by protein–protein interactions) bound to a single hairpin rather than different conformations of the DNA–protein complex. The observation that all of the three Rep78 shifted bands showed the same methylation interference pattern is consistent with each of these models.

---

**TABLE 3. Missense sequences of Baf51 and Baf52**

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Position of first amino acid shown</th>
<th>Nucleotide and amino acid sequencesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baf51</td>
<td>241</td>
<td>CAG TGT CAC GAA GCA GGA AGT CAA AGA TCT TTT CGG GTG GCC AAA GGA + CAG TGT GGT TGA Ser Arg Gly Opal +</td>
</tr>
<tr>
<td>Baf52</td>
<td>243</td>
<td>ATC GAT CCA GGA GCA GCA GCC CTC ATA CAT CTC CTT CAA TCG GCC CTC Ile Asp Pro Gly Pro Gly Leu Ile His Leu Leu Gln Cys Gly Leu - CAA CTC GCC GTC CAA AAT CAA GGC TCG CTT GCA CAA TGC GGG AAA GAT Gln Leu Ala Val Pro Asn Gln Gly Cys Leu Gly Gln Cys Gly Lys Asp + -</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TAT GAG CCT GAC TAA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tyr Glu Pro Asp Ochr</td>
</tr>
</tbody>
</table>

a The last sense codon and sense amino acid residue are underlined. + and −, charges of the amino acid side groups.
ACKNOWLEDGMENTS

We thank Takami Oka and Nancy Nossal for critical reading of the manuscript.

REFERENCES


