Adeno-Associated Virus Type 5 Utilizes Alternative Translation Initiation To Encode a Small Rep40-Like Protein

K. David Farris and David J. Pintel*
Department of Molecular Microbiology and Immunology, University of Missouri—Columbia, School of Medicine, Bond Life Sciences Center, Columbia, Missouri

Received 15 September 2009/Accepted 27 October 2009

Alternative splicing of adeno-associated virus type 2 (AAV2) P19-generated pre-mRNAs generates the small Rep proteins Rep52 and Rep40 from unspliced and spliced mRNAs, respectively. These Rep proteins (as well as the overlapping large Rep proteins) share a common Walker-type domain common to the superfamily 3 (SF3)-type helicases (3, 4). Both small Rep proteins are required for efficient packaging of the AAV2 genome; Rep52 and Rep40 both possess 3'-to-5' helicase activity and are required for unwinding of the double-stranded replicative forms of AAV for insertion into preformed capsids (5). Rep52 and Rep40 possess some functional redundancy in their helicase activities, ATPase activities, and DNA binding activities; however, their precise mechanisms of action are different. Structural analyses have indicated that when bound to DNA, AAV2 Rep52 remains monomeric (14), while Rep40 assumes a hexameric ring-like structure (1, 4), which has been suggested to be important for its function. Although Rep52 and Rep40 both are capable of unwinding DNA substrates with single-stranded DNA ends in a 3'-to-5' fashion, Rep40, but not Rep52, efficiently unwinds double-stranded DNA (1). Therefore, it has been suggested that during AAV replication, the Rep40 hexamer may be necessary for the initial unwinding of the first few bases of the double-stranded DNA intermediate, and further unwinding may then be accomplished by Rep52 alone or by Rep52 in conjunction with Rep40 and other Rep proteins (1).

In contrast to AAV2, pre-mRNA transcripts derived from the AAV5 and goat AAV P19 promoter are preferentially polyadenylated in the viral central intron and thus spliced at low efficiency (10–12). Thus, these viruses would not be expected to generate significant quantities of Rep40. However, immunoblot analyses have consistently demonstrated the presence of significant levels of a Rep40-like (but not a Rep 68-like) protein during AAV5 and goat AAV coinfection with adenovirus, as well as during transient transfection (7, 8, 11). This suggested that AAV5 may generate a Rep40-like protein using a mechanism distinct from the alternative splicing utilized by AAV2. Here we show that, unlike AAV2 Rep40, the AAV5 Rep40-like protein has the same C terminus as Rep52 and that these two proteins differ in their N-terminal ends. We show further that the AAV5 Rep40-like protein was translated from P19-generated mRNA utilizing an additional in-frame, nested AUG translation initiation codon 50 amino acids downstream of the Rep52 initiating AUG codon.

Although AAV5 and goat AAV P7- and P19-generated pre-mRNAs were inefficiently spliced, high levels of a Rep40-like protein product were produced during viral infection. As mentioned above, the transcription map of AAV5 predicts that it would not generate significant levels of Rep40 (Fig. 1A). However, as shown in Fig. 1B, and as we have shown previously (7, 8, 10–12), substantial levels of a Rep40-like protein product are generated during both AAV5 and goat AAV infections of 293 cells (Fig. 1B, lanes 1 and 2), despite the relatively low levels (<5%) of splicing from upstream transcripts (Fig. 1C, lanes 1 and 2, compare Unsp P7/P19 to Sp P7/P19). Contrary to AAV2, AAV5 Rep52 and the Rep40-like protein differed in their N termini. AAV2 Rep52 and Rep40 have the same N terminus and differ in their C termini by virtue of alternative splicing. As this was likely not to be the case for AAV5 due to inefficient splicing of P19-generated pre-mRNAs, we first attempted to define the N-terminal and C-terminal ends of the AAV5 Rep40-like protein. A hemagglutinin (HA) tag was inserted, using standard previously published techniques (7), within the original AAV5 RepCap plasmid (11) either at the MfeI restriction site at nucleotide (nt) 1980 just upstream of the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt 2158 within the intron (AAV5 RepCap MfeI HA) or at nt.
mRNAs. GoAAV, goat AAV. spliced mRNA; P7/P19, mRNA from all P7- and P19-generated Rep-encoding mRNAs; P41, mRNA from P41-generated capsid-encoding generated from linearized templates by in vitro transcription with SP6 polymerase as described previously (11, 12). Unsp, unspliced mRNA; Sp, donor sites and allows the visualization and separation of spliced and unspliced transcripts from the P41 promoter and upstream P7/P19 promoters.

FIG. 1. Although AAV5 and goat AAV P7- and P19-generated pre-mRNAs are inefficiently spliced, high levels of a Rep40-like protein product are produced during viral infection. (A) Diagram of the AAV5 transcription map as previously described (11), illustrating predominant polyadenylation of upstream Rep-encoding transcripts from the P7 and P19 promoters. RP, radiolabeled antisense probe (AAV nt 1843 to 2034) (11); ITR, inverted terminal repeat; Inr, transcripts deriving from the ITR region; (pA)p, proximal polyadenylated mRNA; (pA)d, distal polyadenylated mRNA; D, donor; A1, acceptor 1; A2, acceptor 2; TAA, translation termination site for nonstructural proteins Rep78 and Rep52. (B) Immunoblot analysis using anti-Rep antibody of AAV5 or goat AAV coinfections (multiplicity of infection [MOI], 10) with human adenovirus type 5 (Ad5; MOI, 5) into 293 cells. The locations of Rep78, Rep52, and the Rep40-like protein are indicated. (C) Representative RNase protection assay of the same samples shown in panel B with the RP probe as previously described (10, 11). Briefly, 293 cells were coinfected with AAV5 or goat AAV (MOI, 10) and human Ad5 (MOI, 5). Total RNA was harvested 36 to 42 h later using guanidine isothiocyanate as previously described (11, 12). Ten micrograms of total cellular RNA was protected with a 32P-radiolabeled RNA probe (the RP probe, which spans the P41 promoters and intron donor sites and allows the visualization and separation of spliced and unspliced transcripts from the P41 promoter and upstream P7/P19 promoters) generated from linearized templates by in vitro transcription with SP6 polymerase as described previously (11, 12). Unsp, unspliced mRNA; Sp, spliced mRNA; P7/P19, mRNA from all P7- and P19-generated Rep-encoding mRNAs; P41, mRNA from P41-generated capsid-encoding mRNAs. GoAAV, goat AAV.

anti-HA antibody, whereas the tag upstream of the intron (MfeI HA) should have allowed the detection of all three Rep proteins (Rep78, Rep52, and Rep40-like). As shown in Fig. 2B, all three proteins were detected by immunoblot analysis as previously described (7) following the transfection of both plasmids into 293 cells using the polyethyleneimine transfection reagent as previously described (13), suggesting that they shared carboxyl termini (Fig. 2B, compare lanes 3 and 4). In addition, Rep monoclonal antibody 76.3, which for AAV2 recognizes an epitope encoded within intron sequences and does not recognize Rep68 or Rep40 (03-61073; ARP, Belmont, MA), was also able to detect the AAV5 Rep40-like protein, whether in the full RepCap vector (Fig. 2B, lanes 1 and 2) or in a vector derived from RepCap in which the region upstream of P19 had been deleted, and so only expressed P19-generated RNA (Fig. 2B, lane 9).

The AAV5 Rep52 coding region from upstream of the Rep52 AUG to its terminating TAA within the intron was also placed within pcDNA3.1 to generate a splicing-deficient HA-tagged cytomegalovirus Rep52 (CMV-Rep52) expression vector, which was designed to mimic expression of the P19-generated mRNA. This vector was capable of generating both Rep52 and a Rep40-like protein (Fig. 2B, lane 5). When the CMV-Rep52 construct was truncated at the donor site (DKO; this removes Rep52 amino acids 324 to 390, and we then inserted a C-terminal HA tag after amino acid 323), this mutant generated truncated versions of both Rep52 and the Rep40-like protein (Fig. 2B, lane 6), which would not have been expected if the Rep40-like protein were encoded by alternative splicing.

To determine whether AAV5 Rep52 and Rep40 differed in their N termini, CMV-Rep52 was recloned such that an HA tag was inserted after amino acid 10 [CMV-Rep52 (10 HA)] or after amino acid 60 [CMV-Rep52 (60 HA)]. Both AAV5 Rep52 and the Rep40-like protein were detected by immunoblot analysis using anti-HA antibody when the HA tag was at position 60 (Fig. 2B, lane 8); however, only Rep52 was detected when the HA tag was placed at position 10 (Fig. 2B, lane 7), suggesting that these proteins differed at their amino termini.

Taken together, these data indicated that, unlike AAV2, the AAV5 Rep40-like protein is not generated by alternative splicing and that the AAV5 Rep52 and Rep40-like proteins have the same C terminus and differ in their N termini. In addition, the Rep40-like protein was efficiently derived from the same P19 mRNA that also generates Rep52.

Genetic analysis of the N-terminal region of AAV5 Rep52 revealed that the Rep40-like protein was likely generated by the usage of an internal AUG initiation codon. One of the possibilities for the origin of the AAV5 Rep40-like protein was that it was the result of a specific protease cleavage of Rep52. However, because the accumulation of the Rep40-like protein was unaffected by the addition of either a pan-caspase inhibitor, protease inhibitor cocktails, or by the proteasome inhibitor MG132 (data not shown), we entertained the possibility that the Rep40-like protein might be the result of a specific protease cleavage of Rep52.

Attempts to directly determine the amino acid terminal sequence of the Rep40-like protein were not successful. Therefore, to determine the initiation site of the Rep40-like mole-
cule, we proceeded to examine protein expression from a CMV-Rep52 construct in which we had generated a set of nonbiased deletion mutations within the putative initiation site for the smaller protein (Fig. 3A). As expected for an alternative initiation within this region, a large deletion within Rep52 of amino acids 31 to 60 (ΔH9004 31-60) resulted in the elimination of the Rep40-like protein, as well as the expected noticeable shift in the size of Rep52 (Fig. 3B, compare lanes 1 and 2). However, none of a set of smaller deletion mutations tested, each of which deleted five amino acids (ΔH9004 31-35, ΔH9004 36-40, ΔH9004 41-45, ΔH9004 45-49, ΔH9004 50-54, or ΔH9004 55-59), resulted in elimination of the Rep40-like protein, although they all resulted in a slight shift in the size of Rep52 (Fig. 3B, lanes 3 to 8). Surprisingly, the ΔH9004 50-54 mutation, which eliminated the internal AUG at position 50, also did not result in elimination of the Rep40-like protein from the CMV-Rep52 construct (Fig. 3B, lane 7). This was unexpected; however, as shown below, it is most likely that this mutant generated its Rep40-like protein from alternate initiation at a non-AUG initiation codon (as also observed for AAV2 VP2).

Interestingly, AAV2 also contains the internal methionine at the same location; however, it does not appear to be used. AAV2 and AAV5 show only 78% amino acid identity between the Rep52 initiating methionine and Met50, and their nucleotide sequences are even more divergent (73% identity). Replacement of the first 40 (v2 1-40) or 50 (v2 1-50) amino acids of Rep52 in AAV5 with those of AAV2 prevented detectable expression of the Rep40-like protein (Fig. 3C, lanes 10 and 11). This suggested that the choice of the use of this internal methionine as an initiator was not due merely to the nucleotide makeup immediately surrounding the Met50 AUG and that

![Diagram showing the AAV5 Rep52 and Rep40-like proteins and their N termini.](http://jvi.asm.org/index.php/jvi/article/viewFile/1195)

FIG. 2. The AAV5 Rep52 and Rep40-like proteins differ in their N termini. (A) Diagram of AAV5 P19-driven Rep52-encoding mRNAs. Locations of the initiating methionine (AUG 1) and an internal in-frame methionine 50 amino acids downstream of the initiating AUG (AUG 2) are indicated, as are the locations of the inserted HA tags (HA 10, HA 60, HA MfeI, and HA 380) used in this study. Residue numbers indicate amino acids in Rep52. The previously described (3, 4) helicase domain common to SF3-type helicases (underlined) and the Walker A, B, and C motif residues required for helicase activity of all Rep proteins are indicated. D, donor site; A1, acceptor 1; A2, acceptor 2. (B) Immunoblot assays of extracts following transfection with HA-tagged constructs as described in the text. Lanes 1 and 2 were immunoblotted with Rep antibody 76.3 (03-61073; ARP, Belmont, MA), and lanes 3 through 8 were immunoblotted with an antibody against the HA epitope (H3663; Sigma-Aldrich, St. Louis, MO). Migration of Rep78, Rep52, the Rep40-like protein, and truncated (Δ) Rep52 and Rep40-like proteins are indicated.

To test the importance of the methionine at position 50 directly, we created a series of point mutations in which the individual amino acids between positions 50 and 54 were mutated to arginine. Only the point mutation at residue 50 (M50R; ATG to AGG) resulted in clear, reproducible elimination of the Rep40-like molecule (Fig. 3C, compare lanes 1 and 2). Mutation of this methionine to alanine (M50A; ATG to GCC) also resulted in loss of the Rep40-like protein (Fig. 3C, lane 9), suggesting that the loss of this protein was not merely due to instability of the protein as a consequence of the N-end rule (6, 15). Other mutations in this region (S51R, AGT to AGG; L52R, CTG to AGG; T53R, ACA to AGG; K54R, AAA to AGG) had no effect on the accumulation of the Rep40-like protein (Fig. 3C, lanes 3 to 6). Additionally, we did not detect a compensatory increase in the levels of Rep52 generated by M50R, compared to the wild type, which would have been expected if the loss of the Rep40-like protein was due to the loss of a specific protease cleavage event rather than alternative initiation.
important differences exist between AAV2 and AAV5 in this region. It is possible that the usage of Met50 depends on the upstream mRNA secondary structure, as is observed in Alevian mink disease parovirus (9).

The potential use of the internal methionine was tested further by a series of frameshift and ochre termination mutations around this region. The introduction of premature termination codons at Rep52 position 10, 48, or 49 prior to Met50 resulted in the loss of Rep52 but not the Rep40-like protein (Fig. 4B, lanes 2 to 4). As expected, introduction of premature termination codons at position 50 or 51 relative to Met50 resulted in the loss of both proteins (Fig. 4B, lanes 5 and 6). In addition, the insertion of frameshift mutations at positions 46 (data not shown; GCGACC to GCGAACC [frameshift mutation in italics]), 48 (FS 48; AAAATT to AAAATT), and 49 (FS 49; ATTATG to ATTAATG) upstream of Met50 resulted in the loss of Rep52 but not the Rep40-like protein (Fig. 4C, lanes 3 and 4), while the frameshift insertions at positions 50 (FS 50; ATGAGT to ATGAAGT) and 51 (FS 51; AGTCTG to AGTAATG) relative to Met50 resulted in the loss of both proteins.

**FIG. 3.** The Rep40-like protein is likely initiated at an internal AUG codon. (A) Diagram of the CMV-Rep52 mutants used. The locations of the CMV promoter, the Rep52-initiating AUG (AUG 1), and a nested in-frame internal AUG (AUG 2) are indicated. The locations of deletion mutants are marked. D, donor site location in CMV-Rep52 mRNA, which lacks the intron acceptors and is inactive in this truncated construct. (B) Immunoblot analyses (with anti-HA antibody) of extracts following CMV-Rep52 deletion mutant constructs explained in the text. The locations of Rep52 and the Rep40-like protein are indicated. (C) Immunoblot analyses (using anti-HA antibody) of extracts following transfection with CMV-Rep52 and specific arginine/alanine/AAV2 substitution constructs as described in the text. The locations of Rep52 and the Rep40-like protein are indicated.

**FIG. 4.** The Rep40-like protein is likely generated by the usage of an internal AUG initiation codon. (A) Diagram of the CMV-Rep52 mutants used in this study. The diagram demonstrates a frameshift of Rep52 into an open reading frame that terminates downstream of Met50 (AUG 2). The actual locations of frameshifts are described in the text. The locations of the CMV promoter, the Rep52-initiating AUG (AUG 1), and a nested in-frame internal AUG (AUG 2) are indicated. D, donor site location in CMV-Rep52 mRNA, which lacks the intron acceptors and is inactive in this truncated construct. (B) Immunoblot analysis (with anti-HA and anti-tubulin [T4026; Sigma-Aldrich, St. Louis, MO] antibodies) of extracts following the transfection of CMV-Rep52 early-termination (ochre) mutant constructs as described in the text. TAA, premature termination mutant (numbers indicate amino acid locations of ochre mutations, with the initiation codon of Rep52 as residue 1). (C) Immunoblot analysis (using anti-HA antibody) of extracts following transfection with CMV-Rep52 and specific arginine/alanine/AAV2 substitution constructs as described in the text. The locations of Rep52 and the Rep40-like protein are indicated. (D) Immunoblot analysis of extracts following the transfection of frameshift and deletion mutant CMV-Rep52 or AAV5 RepCap constructs as described in text. The locations of the Rep78, Rep52, and Rep40-like proteins are indicated.
proteins (Fig. 4C, lanes 5 and 6). Importantly, mutation of Rep52-initiating Met1 (ATG KO; ATG to GCC) also resulted in exclusive expression of the Rep40-like protein (Fig. 4C, lane 2). Taken together, these results strongly indicate that the AAV5 Rep 40-like protein is an independent translation product initiating at internal AUG Met50.

As mentioned above, it was surprising that deletion of amino acids 50 to 54 within the CMV-driven Rep52 expression vector, which removed the putative alternative AUG, still generated a Rep40-like protein (Fig. 3B, lane 7). Upon further investigation, we found that when this deletion mutation (Δ50-54) was incorporated into the full AAV5 RepCap plasmid, which has the native P19 promoter rather than the CMV promoter, this mutant was capable of eliminating detectable expression of the Rep40-like protein (Fig. 4D, compare lanes 5 and 6). Perhaps this difference is related to the stronger CMV promoter or the slightly different leader sequence present in the CMV-driven plasmid. Additionally, introduction into the CMV-driven Rep52 (Δ50-54) expression vector of a frameshift mutation at position 40, 10 amino acids upstream of the deletion mutation (FS 40; AAGGCC to AAGAGGCC), resulted in loss of Rep52 but not loss of the Rep40-like protein (Fig. 4D, compare lanes 3 and 4), suggesting that the Rep40-like protein generated by this mutant utilized a nontraditional initiation codon. It is unclear if the difference between the two constructs in utilizing a downstream AUG is important to AAV replication.

Given the small genome size of AAV and that Rep40 expression appears to be a conserved feature of all AAVs, including other animal AAVs, such as goat AAV (which displays a transcription profile similar to that of AAV5), it seems unlikely that the AAV5 Rep40-like protein is dispensable during the AAV life cycle. The conserved production of Rep40 in all AAVs appears to demand its presence, perhaps for proper packaging. However, because of overlap with the essential protein Rep78, we have not yet been able to decipher the specific role of the Rep40-like protein in AAV5 replication. We hypothesized that defects in packaging due to the loss of the Rep40-like protein might be observed by a significant reduction in the production of recombinant AAV (rAAV). Mutations in RepCap that successfully eliminated the Rep40-like protein (M50R, M50A, Δ50-54) were used to generate rAAV as previously described (2), and as expected, these mutations caused significant (approximately 8- to 10-fold) reductions in the production of rAAV (data not shown). However, efforts to recover rAAV levels, either by adding the Rep40-like protein in trans from another vector or by creating secondary mutations within the original mutants that recovered the Rep40-like protein, were unable to complement these effects. Therefore, we could not directly attribute the deficiency in rAAV packaging to the loss of Rep40 rather than potential secondary effects on either Rep78 or Rep52. Thus, it is still unknown whether the AAV5 Rep40-like protein functions as the true Rep40 does for AAV2. Mechanistic comparisons between the small Rep proteins of AAV2 and AAV5 are currently being investigated.

We thank Lisa Burger for excellent technical help and Ramnath Nayak for advice and initial participation in this study. This work was supported by PHS grants AI46458 and AI56310 from NIAID to D.J.P.

REFERENCES

12. Qiu, J., and D. J. Pintel. 2004. Alternative polyadenylation of adeno-associated virus type 5 RNA within an internal intron is governed by the distance between the promoter and the intron and is inhibited by U1 small nuclear RNP binding to the intervening donor. J. Biol. Chem. 279:14889–14898.