Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 7919–7923, December 1985
Biochemistry

Direct mapping of adeno-associated virus capsid proteins B and C: A possible ACG initiation codon

(protein sequencing/human parvovirus/translation initiation)

S. PATRICIA BECERRA*, JAMES A. ROSE*, MEDORA HARDY†, BAHIGE M. BAROUDY*, AND CARL W. ANDERSON†

*Laboratory of Biology of Viruses, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20205; and †Department of Biology, Brookhaven National Laboratory, Upton, Long Island, NY 11973

Communicated by Maxine Singer, August 2, 1985

ABSTRACT **The three major capsid proteins of adeno-associated virus type 2 (AAV2) virions are designated A, B, and C and have molecular sizes of 90, 72, and 60 kDa, respectively. These proteins are related, and genetic studies have shown they are encoded by a long open reading frame located in the right half of the genome. The coding capacity distal to the first ATG in this reading frame is only 503 amino acids (i.e., a protein about the size of protein C), but an open frame sequence devoid of ATG codons extends upstream for an additional 184 codons. Although the amino terminus of the C capsid protein is blocked, partial amino acid sequence analyses of peptides from C have confirmed that it is encoded within the portion of the reading frame distal to the first ATG at nucleotide (nt) location 2810. The amino terminus of the B capsid protein is not blocked, and its sequence begins with alanine. The triplet encoding this alanine lies 64 codons upstream from the initiation site for protein C and is immediately preceded by the threonine codon, ACG, at nt 2615. This ACG codon lies in the most favorable sequence context for protein synthesis initiation. All three AAV2 capsid proteins are labeled *in vitro* with formyl[^{35}S]methionyl-tRNA$_f$, indicating that synthesis of each protein is initiated independently. Our data suggest that the nt 2615 ACG codon directs the methionyl-tRNA-dependent initiation of the AAV2 B capsid protein. Proteins B and C may be synthesized from the same mRNA species and their relative abundance could be determined by the efficiencies of their respective initiation codons.**

Adeno-associated viruses (AAV) are defective parvoviruses whose replication requires helper factors supplied by either a coinfecting adenovirus (Ad) or herpesvirus (1, 2). The AAV genome, a plus or minus single DNA strand of ≈4.7 kilobases (kb), is packaged in virions that are constructed with three major capsid polypeptides (A, B, and C) (3, 4). For AAV type 2 (AAV2) the apparent sizes of these components are 90 (A), 72 (B), and 60 (C) kDa, and their proportions in purified virions (≈1:1:10) are similar to those found in total extracts of Ad/AAV-coinfected cells, an indication that their production may be tightly regulated (5).

All three structural proteins are encoded by the right half of the viral genome, and they are specified by at least two mRNA species generated from transcripts initiated by the most rightward promoter (map coordinate 39; Fig. 1) (9). Although it was initially thought that proteins B and C were produced by proteolytic cleavage of protein A (5), it is now clear that the synthesis of B and C does not depend on the synthesis of A (9). This finding, coupled with the observation that proteins A, B, and C share amino acid sequences that correspond to all of C (10), suggests that the three structural

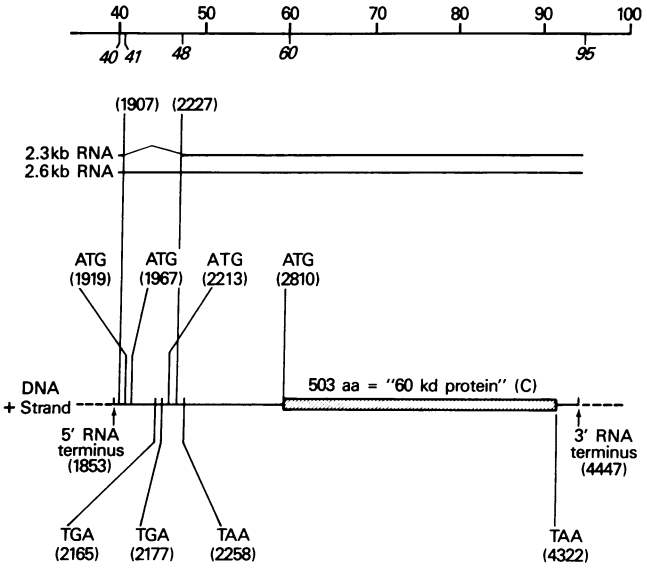

FIG. 1. Diagram showing the longest right-sided ATG-initiated open reading frame in the AAV2 genome. The reported unspliced and spliced RNA species transcribed from the right half of the genome (6, 7) are shown at the top of the diagram, and the major open reading frame (originating at nt 2810) is positioned on the DNA plus (+) strand below. Specific genomic sites are identified by nt number (in parentheses; see ref. 8); the indicated triplets are in-frame with the ATG that maps to nt 2810. There is no ATG triplet in any frame between the splice junction and the ATG at nt 2810.

proteins may originate from independent in-frame initiation sites. Furthermore, staphylococcal V8 proteolysis (10) together with DNA sequence data (8) indicate that protein C should occupy the longest right-sided ATG-initiated open reading frame (map units 60–92; Fig. 1) and should therefore overlap equivalent carboxyl-terminal segments of proteins A and B. This reading frame is sufficient to account for protein C, but it is not long enough to accommodate proteins with the apparent size of either A or B. Thus, the initiation sites and proximal coding sequences for the A and B proteins are not evident from the DNA sequence *per se*. Janik *et al.* (9) have found, however, that protein A is initiated from a site within the intervening sequence, which is intact in the 2.6-kb RNA species (Fig. 1). They proposed that protein A might arise either by a read-through of up to three termination signals or by translation from a hitherto undetected spliced RNA species from which the downstream terminators had been ablated (Fig. 1). On the other hand, in the absence of an obvious AUG initiator for protein B, the possibility that B

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: AAV, adeno-associated virus; nt, nucleotide(s); kb, kilobase(s).

might be derived from a larger precursor remained to be excluded.

In this study, we have directly mapped the origins of both the B and C capsid proteins of AAV2 by aligning their terminal amino acid sequences with corresponding coding sequences of the viral genome. The suspected positioning of C to the open reading frame between map coordinates 60 and 92 was confirmed, but, unexpectedly, B appeared to originate at an in-frame ACG codon 195 bases upstream from the ATG initiator of C. In addition, based on amino-terminal labeling with *N*-formyl[^{35}S]methionyl-tRNA$_f$ (f[^{35}S]Met-tRNA$_f$), we have shown that each of the three capsid polypeptides is separately initiated.

MATERIALS AND METHODS

***In Vivo* and *in Vitro* Synthesis of Radiolabeled AAV Capsid Proteins.** 293-31 cells were coinfected with Ad5 (50 plaque-forming units per cell) and AAV2 (100 focus-forming units per cell) (11). At 12 hr after infection, the cells were suspended for 1 hr in amino acid-free medium supplemented with dialyzed horse serum. Proteins were then labeled with 500 μCi of [^{35}S]methionine per ml (Amersham; 900–1000 Ci/mmol; 1 Ci = 37 GBq) for 2 hr or 2.5 mCi of [^{3}H]leucine, [^{3}H]isoleucine, [^{3}H]alanine, or [^{3}H]proline per ml (Amersham; 56–170 Ci/mmol) for 3 hr. After harvesting, nuclear extracts, containing the bulk of labeled AAV capsid proteins, were prepared as described (9). To label AAV structural proteins *in vitro*, total cytoplasmic RNA from Ad2/AAV2-coinfected KB cells (11), isolated by CsCl centrifugation (9), was used to program AAV structural protein synthesis in micrococcal nuclease-treated rabbit reticulocyte lysates. Reaction mixtures (160 μl) contained 1 mCi of [^{35}S]methionine. To prevent acetylation of nascent α-amino termini, lysates were preincubated with citrate synthase and oxaloacetate as described (12). To prepare f[^{35}S]Met-tRNA$_f$ for amino-terminal labeling of AAV capsid proteins, purified yeast tRNA$_f$ was aminoacylated with [^{35}S]methionine (1440 Ci/mmol), and formylated [^{35}S]Met-tRNA was fractionated from nonformylated tRNA by RPC-5 chromatography as described (13).

HPLC and Amino Acid Sequence Analysis of AAV Proteins and Peptides. AAV capsid proteins, labeled *in vivo* or *in vitro* with [^{35}S]methionine or with a single tritiated amino acid, were purified by immunoprecipitation followed by NaDodSO$_4$/PAGE (9). The individual capsid polypeptides were electroeluted and subjected to amino-terminal sequence analysis (14). Tryptic peptides were fractionated by C-18 reverse-phase chromatography at pH 5.4; individual radiolabeled peptides were partially sequenced as described (15).

RESULTS

AAV Capsid Protein C Maps to the Expected Open Reading Frame. Each of the gel-purified [^{3}H]proline-labeled AAV capsid proteins that had been synthesized *in vivo* was digested with trypsin, and the resulting peptides were resolved by HPLC as shown in Fig. 2. Not unexpectedly, the chromatograms of proteins A, B, and C show radioactive peaks with the same elution positions (e.g., P1, P8, P9, and P17), whereas some peaks are not present in all three chromatograms (e.g., P5 in the B and A chromatograms does not appear in the C chromatogram). To determine the relationship of protein C to the open reading frame found by nucleotide sequence analysis at map coordinate 60 (Fig. 1; see ref. 8), we determined proline positions in the terminal amino acid sequences of selected tryptic peptides of protein C and then matched these to proline positions in the protein predicted from DNA sequence. As anticipated, sequence analyses of P1, P8, P9, and P17 revealed a pattern of proline

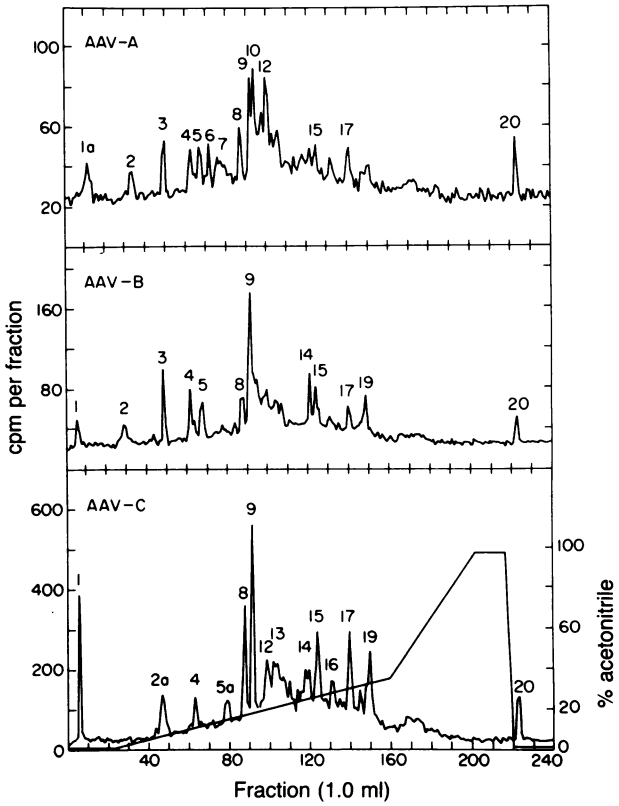

FIG. 2. Reverse-phase HPLC of tryptic peptides derived from AAV2 capsid proteins. AAV capsid proteins labeled with [^{3}H]-proline were individually purified, digested with trypsin, and fractionated by C-18 reverse-phase chromatography. Columns were eluted with a discontinuous linear gradient of acetonitrile (*Lower*). The amount of radioactivity in each fraction was plotted without background subtraction. Peaks of radioactivity representing individual peptides or peptide mixtures have been assigned numbers based on their order of elution. Peaks from each capsid component with the same elution position were assigned the same number. The total radioactivity applied to the column was 17,600 cpm for capsid component A (*Upper*), 29,000 cpm for component B (*Middle*), and 166,000 cpm for component C (*Lower*). Appropriate fractions from the chromatograms of components B and C were pooled and subjected to peptide sequence analysis.

release that matched predicted internal peptides of C [nucleotide (nt) locations 3545–3580 for P1, 3500–3544 for P8, 4145–4192 for P9, and 4058–4117 for P17]. (The positions of these peptides are diagramatically shown in Fig. 5*A*.) To confirm that protein C originates from the ATG at nt 2810, we tried to analyze the amino-terminal sequence of intact protein C. Several attempted analyses of protein C labeled *in vivo* with [^{3}H]leucine, [^{3}H]isoleucine, or [^{3}H]proline suggested that C was blocked at its amino terminus, presumably by acetylation of the predicted amino-terminal alanine residue. Protein C was therefore synthesized *in vitro* under conditions that prevented amino-terminal acetylation (see *Materials and Methods*). Amino-terminal analyses of [^{35}S]methionine- and [^{3}H]alanine-labeled protein C revealed methionine in position 8 and alanine in position 1 (data not shown), results consistent with the predicted origin of C at nt 2810. In addition, low levels of [^{35}S]methionine radioactivity were also found in position 1, presumably reflecting the incomplete removal *in vitro* of the initiating methionine from a small proportion of molecules. We conclude that protein C is encoded by the open reading frame beginning at nt 2810, and that after removal of its first methionine, the following alanine becomes acetylated *in vivo*.

Protein B Originates Upstream from Protein C. To locate the initiation position of an open reading frame for protein B, we analyzed the amino-terminal amino acid sequence of intact protein B labeled *in vivo* with [^{3}H]proline or [^{3}H]-alanine and HPLC tryptic peptide (P5) of B (not present in C) labeled with [^{3}H]proline. Sequence analyses of protein B are shown in Fig. 3. The [^{3}H]proline protein B analysis shows that the yield of radioactivity in residue 2 was 56% of that theoretically expected from an unblocked protein containing 41 proline residues (Fig. 3 *Upper*). Moreover, [^{3}H]alanine protein B prepared *in vivo* also gave a high percentage yield of radioactivity in only residue 1, in agreement with the amino acid sequence for protein B (Fig. 3 *Upper*). This sequence corresponds to a protein being initiated at an ACG codon (nt 2615) upstream and in-frame with protein C. Fig. 3 (*Lower*) shows the analysis of HPLC tryptic peptide P5, which was only present in tryptic digests of proteins A and B. The matching amino acid sequence for P5 begins at the fifth residue with respect to the amino terminus of B, and the entire P5 peptide mapped from nt 2630 to 2686. As was the case for C, amino-terminal analysis of *in vitro*-synthesized [^{35}S]methionine-labeled protein B revealed a small peak of radioactivity in residue 1. Presumably, this also resulted from incomplete removal of the initiating amino acid residue *in vitro* and suggests that the putative ACG initiator specifies methionine as the initiating residue.

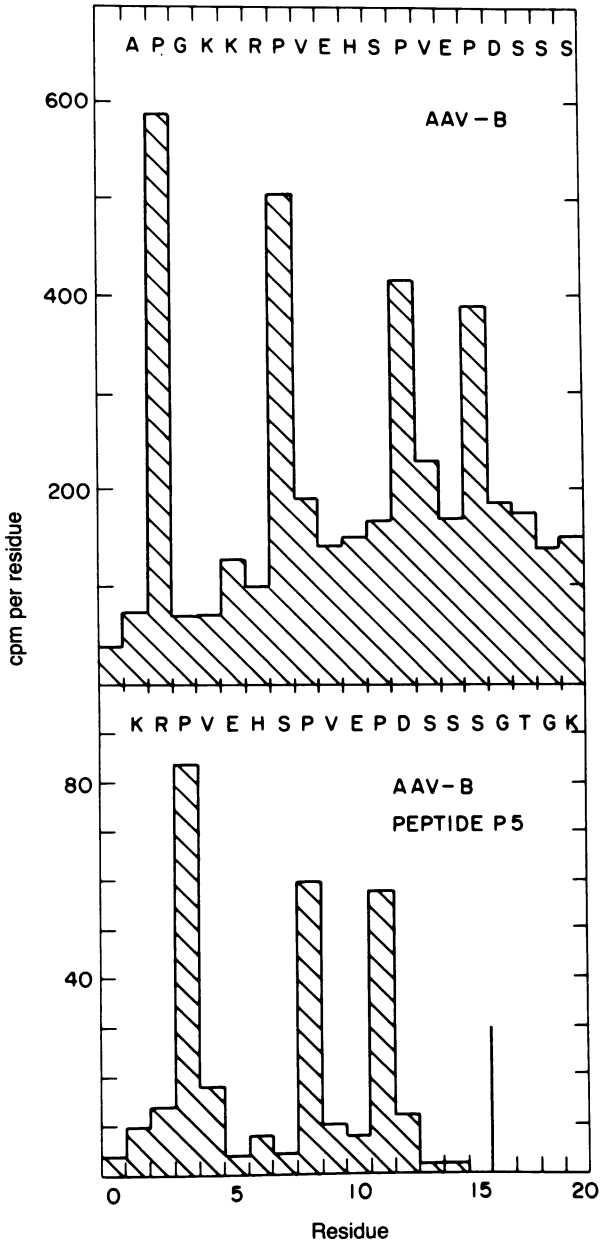

FIG. 3. Sequence analysis of AAV capsid protein B. (*Upper*) Amino-terminal analysis of purified [^{3}H]proline-labeled AAV B protein synthesized *in vivo* (44,500 cpm were applied to the sequencer). (*Lower*) Analysis of tryptic peptide P5 (360 cpm applied) derived from the HPLC chromatogram in Fig. 2 (*Middle*). The matching amino acid sequence (identified by the single-letter code) derived from the DNA sequence of AAV is given at the top of each panel. Note that peptide P5 begins at the fifth residue with respect to the amino terminus of capsid component B.

To confirm the presence of the ACG codon at the start site for protein B, we sequenced (16, 17) a 927-base-pair restriction fragment of AAV DNA [*Tth111* I (nt 2912) and *Eco*RI (nt 1985); see Fig. 5*A*] derived from an AAV2 DNA clone (map units 3–97) contained in pBR325 (pLH1). This clone has been shown to specify protein B in transfection experiments (9). Our results agree with the previous AAV DNA sequence determination (8) and corroborate the occurrence of an ACG triplet at nt 2615.

The A, B, and C Proteins Arise from Individual Initiations. Purified f[^{35}S]Met-tRNA$_f$ was used to incorporate methionine at the amino-terminal position of AAV polypeptide chains synthesized *in vitro* in a rabbit reticulocyte lysate. This tRNA species initiates polypeptide synthesis by transferring its methionine to the amino-terminal position of the peptide chain (18). Formylation of the methionine residue blocks cleavage from nascent polypeptides (19) and f[^{35}S]Met-tRNA$_f$ should not be incorporated at positions corresponding to internal AUG codons. The electrophoretic pattern of immunoprecipitated AAV capsid proteins synthesized *in vitro* with f[^{35}S]Met-tRNA$_f$ is shown in Fig. 4 (lane 3). Three capsid polypeptides were produced that correspond to the A, B, and C components of purified virions (lane 4). There is also correspondence with the A, B, and C proteins labeled either *in vitro* (lane 1) or *in vivo* (lane 2) with [^{35}S]methionine. To rule out internal incorporation of f[^{35}S]Met, f[^{35}S]Met-labeled A, B, and C capsid proteins were separately digested with staphylococcal V8 protease and were electrophoresed to see if more than 1 peptide fragment contained radioactive label. In each case, only 1 of at least 10 fragments was found labeled (i.e., the initial peptide; refs. 10 and 20). The A, B, and C proteins, therefore, were labeled only by their initial fMet residues, and we conclude that each must originate from an independent initiation site.

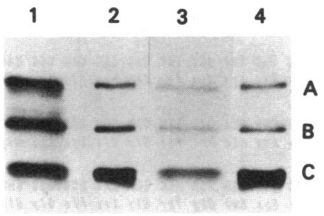

FIG. 4. Electrophoresis of AAV2 capsid proteins labeled with f[^{35}S]Met-tRNA$_f$. Lanes: 1, immunoprecipitated AAV capsid proteins labeled with [^{35}S]methionine in a rabbit reticulocyte lysate programmed with total RNA from Ad/AAV-coinfected KB cells; 2, AAV capsid proteins labeled with [^{35}S]methionine and immunoprecipitated from a nuclear extract of Ad/AAV-coinfected 293 cells; 3, AAV capsid proteins labeled with f[^{35}S]Met-tRNA$_f$ and immunoprecipitated from an *in vitro* reaction as in lane 1. f[^{35}S]Met is only incorporated into the amino-terminal position of newly initiated polypeptide chains. Lane 4, CsCl-purified AAV virions labeled with [^{35}S]methionine.

DISCUSSION

We have directly mapped the locations encoding two of the three AAV2 capsid proteins. AAV protein C, as expected, maps to the right-sided open reading frame. Although virion AAV protein C appears to have a blocked amino terminus, analysis of *in vitro* synthesized C has shown that C synthesis initiates with the AUG encoded at nt 2810 (Fig. 5). From analogy with similar initiation sequences, we suggest that the initiating methionine is removed and the new amino-terminal alanine becomes acetylated (12).

The major right-sided open reading frame ends with a TAA termination codon at nt 4322. We have not analyzed the carboxyl-terminal sequence of the capsid proteins, but we have identified two AAV C peptides (P17 and P9) near the carboxyl-terminal position predicted by this termination codon (Fig. 5). All three AAV capsid proteins contain these two peptides. If the capsid polypeptides terminate with the same sequence specified by the nt 4322 TAA codon, the AAV C protein would be predicted to consist of 503 amino acids and to have a molecular mass (including acetyl group) of 56,278 Da.

Direct sequencing of the AAV B protein has positioned its amino-terminal alanine-encoding sequence at nt 2618. This alanine codon is in the same reading frame as the C protein, and no termination codons occur in this frame between nt 2618 and the beginning of the C coding sequence; therefore, B is predicted to consist of 568 amino acids, 65 of which do not occur in C, and to have a size of 62,792 Da. The codon for the amino-terminal alanine is directly preceded by the threonine codon, ACG, rather than the expected ATG codon.

All three AAV capsid proteins can be initiated *in vitro* with fMet-tRNA$_f$; thus, B does not arise by proteolytic processing of A. Furthermore, f[^{35}S]Met-labeled B protein migrates identically to authentic virion-derived B protein during NaDodSO$_4$/polyacrylamide gel electrophoresis (Fig. 4), and the f[^{35}S]Met-labeled staphylococcal V8 protease-derived amino-terminal B peptide and an internally labeled V8 peptide derived from [^{35}S]methionine-labeled B from virions also comigrate. These results show that, at most, only a very few amino acids could be removed from the nascent B polypeptide to generate the alanine amino-terminus of mature B. We conclude that the initiation site for B must reside very close to the amino-terminal alanine codon (nt 2618) in the B protein mRNA; this site could be the adjacent nt 2615 ACG codon.

In higher eukaryotes, only AUG has been found to function as an initiation codon for protein synthesis. Before proposing that the nt 2615 ACG codon serves to initiate synthesis of the AAV B protein, we considered two alternative explanations. One possibility is that during transcription, the nt 2615 codon is misread as AUG with a low frequency. Because of the favorable context surrounding the nt 2615 site, the resulting AUG would be expected to function efficiently as an initiation codon. We know of no precedent for polymerase misreading at a frequency as high as 1–5%, the amount required to account for the observed ratio of capsid polypeptides. Nevertheless, rigorous exclusion of this possibility is difficult.

The second possibility for introducing an AUG codon adjacent to the nt 2618 alanine codon is mRNA splicing. Two separate studies of AAV mRNA species have been reported

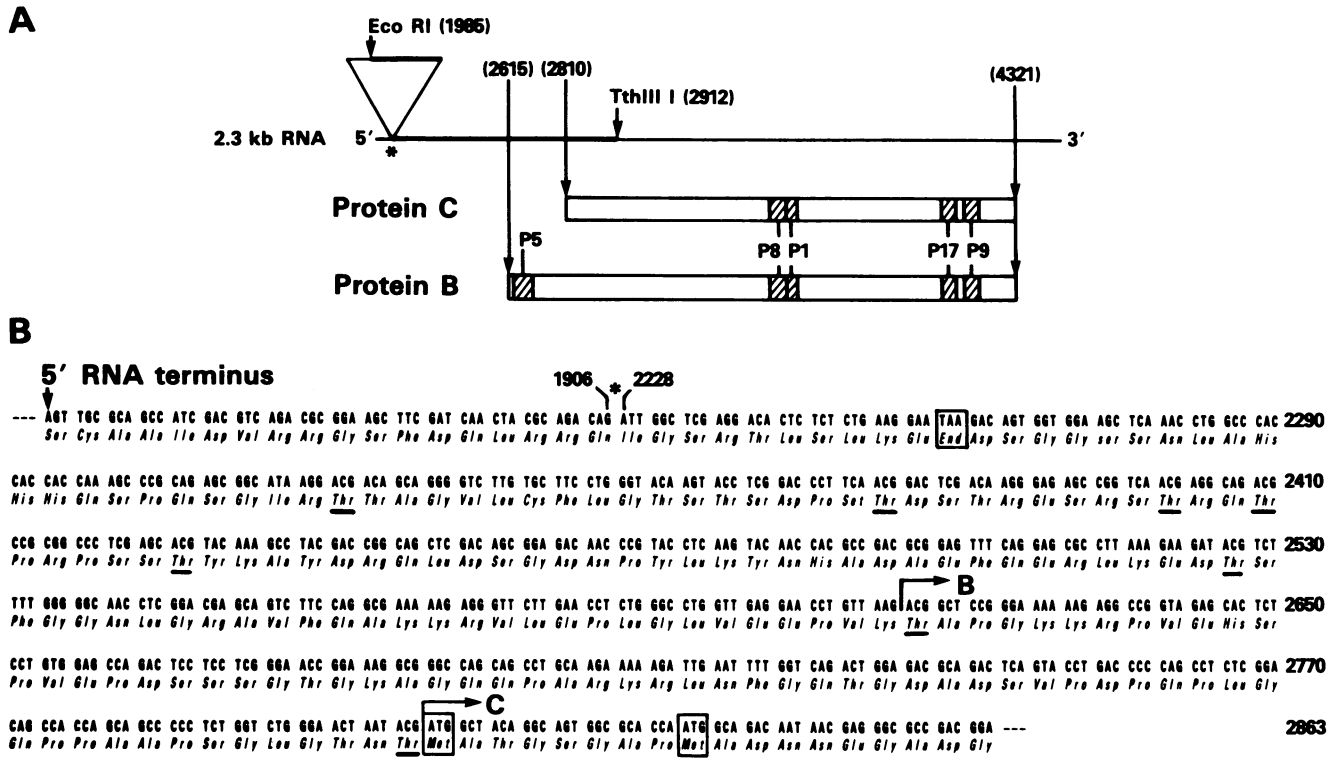

FIG. 5. (*A*) Schematic representation of coding regions for AAV capsid proteins B and C as deduced from amino acid sequence analysis. Positions of internal tryptic peptides are shown (hatched boxes). The DNA sequence between the *Eco*RI and *Tth111* I sites (nt 2598–2635) was repeated to confirm the putative ACG initiation codon for B. The smallest and most abundant AAV RNA species (spliced 2.3-kb RNA) is represented and may be responsible for the synthesis of proteins B and C. (*B*) DNA plus strand sequences that encode the 5′ half of 2.3-kb AAV RNA. The splice junction is identified with an asterisk. The indicated amino acid sequence lies in-frame with the major open reading frame initiated at nt 2810. A termination signal (TAA) and methionine codons (ATG) are enclosed in squares. ACG triplets (threonine codons) are underlined with bars. Initiation codons for proteins B and C are indicated with arrows. In contrast to the other upstream ACG triplets, the presumptive B initiator lies in the most favorable sequence context for a functional initiation codon (21).

(6, 22). Both found only two transcripts (of 2.3 and 2.6 kb) corresponding to the capsid region (Fig. 1). We believe that a third mRNA, spliced at or near nt 2615 to provide an in-frame AUG codon, is unlikely to have been missed if it were at an abundance in AAV-infected cells ≥5% with respect to the 2.3-kb mRNA.

Further arguments against a novel spliced mRNA responsible for B protein synthesis are derived from an analysis of the nucleotide sequences around nt 2615 and at upstream sites (Fig. 5). A compatible splice acceptor site downstream from nt 2620 is unlikely because no upstream ATG codons are followed by a sequence that could encode the amino-terminal sequence of the B protein. A splice acceptor upstream from nt 2600 is unlikely because the resulting B protein precursor (or its V8 proteolytic amino-terminal peptide) would be expected to migrate slower than the authentic B equivalent. Eleven ATG codons are found between the 5′ end of the 2.6-kb RNA (at nt 1854) and nt 2618, but all are in the intron (between nt 1907 and 2227) that is removed from the 2.3-kb message. Although many examples of alternative splice site usage have been reported, we are not aware of an example where alternative spliced mRNAs are produced simultaneously from the same transcript by using both different donor and different acceptor sites. Finally, while splice donor and acceptor sites display considerable sequence variability, one rule has emerged. Introns always begin with the dinucleotide sequence G-T and end with A-G (23). The sequence A-G occurs twice immediately upstream from nt 2615 (at nt 2601–2602 and nt 2613–2614). Neither A-G lies in a striking splice acceptor context [$(Y)_{11}$NYAG:G]. A splice at either site would maintain the frame of B, and thus normal initiation would require an in-frame AUG prior to the donor site (no likely candidate donor has been found); a splice to either site would require additional amino-terminal processing to produce the alanine amino terminus of B.

Given the above considerations, we believe the most likely explanation consistent with our amino-terminal sequence of B is that initiation of protein synthesis occurs at the nt 2615 ACG codon. As noted above, ACG has not been reported to function as an initiation codon for protein synthesis in eukaryotic cells; however, the ACG triplet does promote (weakly) the binding of fMet-tRNA$_f$ to *Escherichia coli* ribosomes (24) and ACG, in a synthetic message, was found to promote initiation of protein synthesis (by *E. coli* ribosomes) *in vitro* (25).

If ACG serves as an initiation codon, this activity may require its placement in the most favorable initiation context (ANNXXXG; see ref. 21). In-frame ACG triplets occur at six positions between the termination codon at nt 2258 and the B start site (nt 2615; Fig. 5), and three ACG triplets also occur in other reading frames. Of these, none has the most favored context, although one, at nt 2321, is in the frequently found context: AGGACGACA. An ACG triplet in a similar context (AATACGATG) also occurs immediately prior to the initiation site for the C protein. We have no evidence that any of these ACG triplets promote initiation. It is perhaps relevant that a rare thalassemia has been found to result from a mutation that has converted the ATG initiation codon for a human α-2-globin gene to ACG (26). This ACG codon, like the putative AAV B protein initiator lies in the most favorable initiation context, a feature that suggests the mutation might not have entirely abolished expression of the α-2-globin gene.

The three AAV capsid polypeptides are synthesized in about the same proportions as are found in virions. If these proportions are required to enhance assembly of functional virions, a mechanism must exist for regulating their relative rates of synthesis. The possible use of ACG as the initiation codon for protein B suggests an interesting mechanism that could regulate the relative synthesis of proteins B and C *in vivo*. If these proteins are translated from the same mRNA species (i.e., the spliced 2.3-kb RNA; Figs. 1 and 5; ref. 9), the initiation complex should first encounter the putative ACG initiator for protein B. Assuming this initiation signal is relatively weak, owing to the presence of cytosine instead of uracil, only occasional initiations might ensue (in perhaps 5–10% of instances), with the complex usually moving on to the stronger AUG initiator for protein C. The B/C ratio would thus be fixed and independent of message concentration. The use of an inefficient initiation codon in this manner might have further relevance in other eukaryotic systems.

We thank Drs. Wei-May Ching and Arthur Wittwer for providing the *E. coli* aminoacyl tRNA synthetase, Dr. U. L. Rajbhandary for purified yeast tRNA$_f$, and Dr. Dolf Hatfield for his assistance in purifying f[^{35}S]Met-tRNA$_f$. We also thank Dr. Edwin Sebring and Mr. Keith Thompson for computer analyses and Joyce Glover for typing the manuscript. This work was supported in part by the United States Department of Energy.

1. Rose, J. A. (1974) in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum, New York), Vol. 3, pp. 1–61.
2. Buller, R. M. L., Janik, J. E., Sebring, E. D. & Rose, J. A. (1981) *J. Virol.* **40,** 241–247.
3. Rose, J. A., Berns, K. I., Hoggan, D. & Koczot, F. J. (1969) *Proc. Natl. Acad. Sci. USA* **64,** 863–869.
4. Rose, J. A., Maizel, J. V., Jr., Inman, J. K. & Shatkin, A. J. (1971) *J. Virol.* **8,** 766–770.
5. Buller, R. M. L. & Rose, J. A. (1978) *J. Virol.* **25,** 331–338.
6. Green, M. R. & Roeder, R. G. (1980) *J. Virol.* **36,** 79–92.
7. Green, M. R. & Roeder, R. G. (1980) *Cell* **22,** 231–242.
8. Srivastava, A., Lusby, E. W. & Berns, K. I. (1983) *J. Virol.* **45,** 555–564.
9. Janik, J. E., Huston, M. M. & Rose, J. A. (1984) *J. Virol.* **52,** 591–597.
10. McPherson, R. A. & Rose, J. A. (1983) *J. Virol.* **46,** 523–529.
11. Janik, J. E., Huston, M. M. & Rose, J. A. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 1925–1929.
12. Palmiter, R. D. (1977) *J. Biol. Chem.* **252,** 8781–8783.
13. Kelmers, A. D. & Heatherly, D. E. (1971) *Anal. Biochem.* **44,** 486–495.
14. Anderson, C. W. (1982) in *Genetic Engineering*, eds. Setlow, J. K. & Hollaender, A. (Plenum, New York), Vol. 4, pp. 147–167.
15. Oosterom-Dragon, E. A. & Anderson, C. W. (1983) *J. Virol.* **45,** 251–263.
16. Volgestein, B. & Gillespie, D. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 615–619.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
18. Housman, D., Jacobs-Lorena, M., Rajbhandary, U. L. & Lodish, H. F. (1970) *Nature (London)* **227,** 913–918.
19. Lodish, H. F., Housman, D. & Jacobsen, M. (1971) *Biochemistry* **10,** 2348–2356.
20. Becerra, S. P., Rose, J. A. & Anderson, C. W. (1985) in *Sequence Specificity in Transcription and Translation*, eds. Calendar, R. & Gold, L. (Liss, New York), Vol. 30, in press.
21. Kozak, M. (1983) *Microbiol. Rev.* **47,** 1–45.
22. Laughlin, C. A., Westphal, H. & Carter, B. J. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5567–5571.
23. Mount, S. M. (1982) *Nucleic Acids Res.* **10,** 459–472.
24. Ghosh, H. P., Soll, D. & Khorana, H. G. (1967) *J. Mol. Biol.* **25,** 275–298.
25. Thach, R. E., Sundararajan, T. A., Dewey, K., Brown, T. C. & Doty, P. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31,** 85–97.
26. Pirastu, M., Saglio, G., Chang, J. C., Cao, A. & Kan, Y. W. (1984) *J. Biol. Chem.* **259,** 12315–12317.