Polypeptide Structure and Encoding Location of the Adenovirus Serotype 2 Late, Nonstructural 33K Protein

ELIZABETH A. OOSTEROM-DRAGON† AND CARL W. ANDERSON*

Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Received 19 July 1982/Accepted 4 October 1982

Radiochemical microsequence analysis of selected tryptic peptides of the adenovirus type 2 33K nonstructural protein has revealed the precise region of the genomic nucleotide sequence that encodes this protein. The initiation codon for the 33K protein lies 606 nucleotides to the right of the *EcoRI* restriction site at 70.7 map units and 281 nucleotides to the left of the postulated carboxyterminal codon of the adenovirus 100K protein. The coding regions for these two proteins thus overlap; however, the 33K protein is derived from the +1 frame with respect to the postulated 100K reading frame. Our results contradict an earlier published report suggesting that these two proteins share extensive amino acid sequence homology (N. Axelrod, *Virology* 87:366-383, 1978). The published nucleotide sequence of the Ad2 *EcoRI*-F fragment (70.7 to 75.9 map units) cannot accommodate in a single reading frame the peptide sequences of the 33K protein that we have determined. Sequence analysis of DNA fragments derived from virus has confirmed the published nucleotide sequence in all critical regions with respect to the coding region for the 33K protein. Consequently, our data are only consistent with the existence of an mRNA splice within the coding region for 33K. Consensus donor and acceptor splice sequences have been located that would predict the removal of 202 nucleotides from the transcripts for the 33K protein. Removal of these nucleotides would explain the structure of a peptide that cannot otherwise be directly encoded by the *EcoRI*-F fragment. Identification of the precise splice points by peptide sequencing has permitted a prediction of the complete amino acid sequence for the 33K protein.

Adenoviruses have nonenveloped, icosahedral virions composed of a linear double-stranded DNA of about 23×10^6 M_r encased in a shell constructed from 10 to 12 different virus-encoded proteins (29). In addition to these structural proteins, late viral transcripts are believed to encode at least three nonstructural proteins that are believed to function in virion assembly and maturation. A maturation proteinase activity has been described (9) that processes at least four virion polypeptides and that is believed to be encoded within late mRNA region L3 (see Fig. 2) (1, 17, 21), but the polypeptide corresponding to this activity has not yet been observed. Two nonstructural proteins are synthesized in large amounts late after infection from mRNAs produced from late region L4. The 100,000-molecular-weight (100K) polypeptide is believed to function in hexon capsmer assembly and transport (14-16, 26, 27). A 33K polypeptide

is also encoded by this region, but its function is not known, and its possible relationship to the 100K protein sequence has been suggested. Before functions can be definitively ascribed to either the 100K or 33K polypeptides, their relationship to each other and to their genome-encoding nucleotide sequences must be determined.

Although abundant at late times after infection of HeLa cells, the 33K protein was overlooked by early studies of adenovirus-specific proteins (4, 36), probably because of its low methionine content. The first description of a ca. 33,000 M_r polypeptide specific to extracts of adenovirus-infected cells was by Russell and Blair (31), who described a major virus-specific phosphorylated polypeptide of 39,000 M_r found in the nuclei of adenovirus 5-infected cells late after infection. This component was subsequently observed by Levinson et al. (22) and by Axelrod (6). Axelrod compared the tryptic peptides from the three major adenovirus-specific nonstructural phos-

† Present address: CODON, Brisbane, CA 94005.

phoproteins, the early 72,000- M_r DNA-binding protein, the 100K protein, and the 33K protein and concluded that the 100K protein and the 33K protein were closely related in amino acid sequence. Lewis et al. (23) mapped by in vitro translation of hybridization-selected adenovirus mRNA a 38,000- M_r polypeptide to the *EcoRI*-F (70.7 to 75.9 map units [m.u.]) and -D (75.9 to 83.4 m.u.) restriction fragments of adenovirus 2. But not until the much later hybridization-translation studies of Ricciardi et al. (30) and Miller et al. (25) was the 33K polypeptide shown to be an independently coded adenovirus product. Even these studies did not show that the in vivo and in vitro 33K products were necessarily related in sequence.

Recently Gambke and Deppert (14, 15) prepared antiserum to purified adenovirus 2 33K protein and found that this serum did not cross-react with the 100K protein. This result led Gambke and Deppert to reexamine the conclusion of Axelrod that these proteins were closely related; their own tryptic mapping studies showed that these proteins were probably unrelated in amino acid sequence. Using their antiserum we have shown that the in vitro product mapped to the *EcoRI*-F fragment of the adenovirus 2 genome is indeed the same as the in vivo 33K product described by Axelrod, Gambke and Deppert, and others. Radiochemical microsequence analysis of the major tryptic peptides of 33K protein isolated from adenovirus 2-infected HeLa cells has allowed us to predict the complete amino acid sequence of this protein by comparing partial peptide sequences with the known nucleotide sequence of the *EcoRI*-F fragment (13). Surprisingly, this analysis showed that the 33K protein is encoded by an mRNA that is spliced within the coding region. This result explains why the 33K protein could not previously be related correctly to the nucleotide sequence.

MATERIALS AND METHODS

Cells and virus. The origins of our HeLa cells and of serotype 2 adenovirus stocks and the propagation of these have been described previously (2, 32). The adenovirus stock used for these experiments is directly related to those used by the Cold Spring Harbor Laboratory for DNA sequence studies (38). Adenovirus DNA was prepared from purified virions as described previously (32).

Purification of radioactively labeled 33K protein. Nearly confluent 90-mm culture dishes of HeLa cells were infected with 250 PFU of adenovirus per cell. At 24 to 26 h after infection, the infected cultures were placed into medium containing the appropriate ^3H -amino acid (1 mCi per dish) or [^{35}S]methionine (250 μCi per dish). After incubation at 37°C for 4 to 6 h, the radiolabeled cultures were fed with 5 ml of normal Dulbecco modified Eagle medium supplemented with 2% calf serum per dish, and incubation was continued

until approximately 48 h after infection. Harvested cultures were washed with ice-cold phosphate-buffered saline and then with 0.01 M Tris-hydrochloride (pH 7)–0.15 M NaCl–0.002 M MgCl_2 (TSM buffer). After suspension in 1.0 ml of ice-cold TSM buffer per dish, the cultures were lysed by the addition of Nonidet P-40 to 0.5% and phenylmethylsulfonyl fluoride to 300 $\mu\text{g}/\text{ml}$. The nuclear pellet, obtained by centrifugation, was washed once in TSM buffer containing 0.5% Nonidet P-40 suspended in sodium dodecyl sulfate (SDS) sample buffer (4), and heated for 5 min in a boiling-water bath. Nuclear extracts were fractionated by electrophoresis through slab-type SDS-polyacrylamide gels (4). Proteins were located by staining with Coomassie brilliant blue or by autoradiography of the dried gels. A dried gel slice corresponding to the position of the 33K protein was excised and eluted electrophoretically with an ISCO sample concentrator (3, 33).

Digestion with trypsin and peptide fractionation by reverse-phase (C18) high-pressure liquid chromatography. Samples to be digested with trypsin were first extracted with acetone-acetic acid-triethanolamine (90:5:5) three times and acetone twice to remove SDS (18). After the addition of 1 mg of apomyoglobin, samples were performic acid oxidized (4) and lyophilized. For digestion with trypsin, samples were dissolved in 0.9 ml of water and 0.1 ml of 1 M NH_4HCO_3 ; 10 μg of tosylamido 2-phenylethyl chloromethyl ketone-treated trypsin (Worthington Diagnostics, Freehold, N.J.) was added. After 2 h of incubation at 37°C, a second 10 μg of trypsin was added, and incubation was continued for a total of 6 h. Samples were then frozen and lyophilized.

Fractionation by reverse-phase chromatography was performed by using a 250- by 4.6-mm, 10- μm RP-18 column (E. M. Reagents, Inc.) equilibrated with 95% buffer A (10 mM ammonium acetate, pH 5.4) and 5% buffer B (100% acetonitrile). Samples were dissolved in 0.3 ml of 20% formic acid and centrifuged at $12,000 \times g$ for 15 min before application to the column. The column was developed with a discontinuous linear gradient of acetonitrile (see Fig. 3) at a flow rate of 2.0 ml/min; 1.0-ml fractions were collected. The position of apomyoglobin peptides was monitored at 214 nm; these positions were used to facilitate the alignment of radioactive peptide profiles from separate experiments. In general, separations varied by less than ± 1 fraction. A portion of each fraction (0.05 to 0.15 ml) was assayed for radioactivity with 3.5 ml of scintillation fluid (Aqualog; New England Nuclear Corp., Boston, Mass.). Appropriate fractions were pooled, 1 mg of apomyoglobin was added, and pooled fractions were lyophilized before amino acid sequence analysis.

Amino-terminal sequence analysis. Radiochemical microsequence analysis was performed with a Beckman 890 C sequencer and a 0.1 M Quadrol buffer program as described previously (3, 5, 33). Polybrene (3 mg) was included with all peptide samples. Sequencer performance was routinely monitored by analysis of carrier apomyoglobin.

DNA sequence analysis. DNA sequence analysis was performed by the method of Maxam and Gilbert (12, 24). Appropriate restriction enzyme-generated fragments were labeled with [^{32}P]ATP and polynucleotide kinase.

Immunoprecipitation. Immune complexes were adsorbed to Formalin-fixed *Staphylococcus aureus* cells (The Enzyme Center, Boston, Mass.) and collected by centrifugation (19). Complexes were released by boiling in 2% SDS-0.1 M dithiothreitol sample buffer as previously described (26).

RESULTS

Identification of the adenovirus 33K protein. The adenovirus 33K protein is a major nonstructural virus polypeptide produced at late times after infection; it can readily be identified in the pattern of Coomassie brilliant blue-stained bands after SDS-polyacrylamide gel electrophoresis of infected cell nuclear extracts (Fig. 1A). To confirm the identity of the 33K protein, radioactively labeled nuclear extracts were immunoprecipitated with antiserum specific for the adenovirus 33K protein (the kind gift of W. Deppert), and the precipitates were analyzed by SDS-polyacrylamide gel electrophoresis. Figure 1B shows that the only product immunoprecipitated from [³⁵S]methionine-labeled or ³²PO₄-labeled nuclear extracts of infected cells is a product of ca. 33,000 *M_r* that corresponds in mobility to the major nuclear component identified as the 33K protein in Fig. 1A. Figure 1B also shows that the electrophoretically purified product used for the peptide and sequence analysis described here is radiochemically pure and can be immunoprecipitated by 33K protein-specific antiserum.

In vitro 33K product labeled with [³⁵S]methionine was prepared by translation of RNA selected by hybridization to the *Eco*RI-F (70.7 to 75.9 m.u.) or *Hind*III-H (73.6 to 79.9 m.u.) fragment of adenovirus 2 DNA (28, 30). Among the products from these translations was a major component with a mobility by SDS-polyacrylamide gel electrophoresis essentially identical to that of 33K protein from infected cells (data not shown). This component was excised from a preparative SDS-polyacrylamide gel, digested with trypsin, and compared with similarly prepared 33K protein from infected cells by reverse-phase liquid chromatography (data not shown). Both preparations had single major methionine-containing peptides that eluted from the column at identical positions. Subsequent studies have shown that this elution position corresponds to that of the single methionine-containing peptide predicted to be present in the 33K protein (see below and Table 1). Thus, the in vivo 33K product must be encoded within the region from 70 to 80 m.u. on the adenovirus 2 genome (Fig. 2).

Inspection of the nucleotide sequence corresponding to late mRNA within the *Eco*RI-F fragment of adenovirus 2 (13) reveals four moderately long open reading frames that could

potentially encode the 33K protein. The first open reading frame begins at the left-hand boundary of the *Eco*RI-F fragment (70.7 m.u.), is 296 codons long, contains one methionine codon, and is believed to encode the carboxy-terminal portion of the adenovirus 100K late, nonstructural protein (13). The second open frame begins at nucleotide 307 (N307; to the right of 70.7), is 295 codons long, overlaps the first open frame, and contains four methionine codons. The third open frame begins at N1006, partially overlaps the second open frame, and contains no methionine codons. The fourth open frame begins at N1519 and has been found to encode the precursor to virion component VIII (Oosterom-Dragon and Anderson, unpublished data).

33K protein has a blocked amino terminus. The 33K protein is known to map to the left of virion component VIII (25). The five methionine codons located in the first two open reading frames are each followed within 20 codons by one or more leucine or alanine codons. Consequently, if the amino terminus of the 33K protein were not blocked, direct amino-terminal radiochemical sequence analysis of [³H]leucine- or [³H]alanine-labeled 33K protein would yield radioactivity in a characteristic pattern of residues that would identify the encoding location of the amino terminus of the protein. Several such experiments were performed with [³⁵S]methionine-, [³H]leucine-, or [³H]alanine-labeled 33K protein isolated as described above, and no significant release of radioactivity was obtained from any of the first 20 Edman cycles. Analysis of carrier apomyoglogin included with the radioactive samples assured us that the sequencer was functioning properly. Since we have rarely failed to obtain sequence information by this method for proteins without blocked amino termini (3, 5, 33), we conclude that the 33K protein, like most of the late virion proteins, has a blocked amino terminus that precludes direct sequence determination.

Sequence analysis of 33K protein tryptic peptides. Adenovirus 2 33K protein, labeled separately with different ³H-amino acids or [³⁵S]methionine, was purified by gel electrophoresis, performic acid oxidized, and digested with trypsin as described above. The 20% formic acid-soluble peptides were applied to a reverse-phase column for fractionation by high-pressure liquid chromatography. Examples of the patterns of radioactive tryptic peptides obtained are given in Fig. 3. Each different amino acid label produced a unique pattern of well-resolved peptides; some peptides appeared to contain only one of the radioactive amino acids used, and others were detected by several different amino acid labels. Appropriate fractions from individ-

FIG. 1. Identification of the adenovirus 33K protein. (A) Coomassie brilliant blue R-250-stained 17.5% SDS-polyacrylamide gel upon which were electrophoresed the following: 1, nuclear extract of mock-infected HeLa cells; 2, nuclear extract of adenovirus 2-infected HeLa cells harvested 32 h after infection; 3, the corresponding cytoplasmic extract of adenovirus 2-infected HeLa cells. The position of the adenovirus 33K protein as well as the positions of several other viral proteins are indicated at the left of the figure. A cellular protein with a mobility similar to the adenovirus 33K protein is found in nuclear extracts of uninfected HeLa cells, but this cellular protein is not recognized by antibody to the viral 33K protein. Cellular protein synthesis would have been largely inhibited at the time infected cells were labeled for the preparation of radioactive 33K protein. (B) Autoradiogram of an SDS-polyacrylamide gel upon which were fractionated immunoprecipitates obtained with antibody specific for the 33K protein. The samples were as follows: 1, [^{35}S]methionine-labeled nuclear extract of adenovirus 2-infected HeLa cells labeled ca. 26 to 32 h after infection; 2, immunoprecipitate of the [^{35}S]methionine-labeled nuclear extract obtained with normal rabbit serum; 3, as in lane 2, but obtained with serum specific for the adenovirus 33K protein; 4, nuclear extract of adenovirus 2-infected HeLa cells labeled with ^{32}P ; 26 to 32 h after infection; 5, immunoprecipitate of the ^{32}P -labeled nuclear extract with normal rabbit serum; 6, as in lane 5, but with antiserum specific for the adenovirus 33K protein (Fig. 3); 8, immunoprecipitate of [^3H]alanine 33K protein with normal rabbit serum; as in lane 8, but with antiserum specific for the adenovirus 33K protein. Immune complexes were collected with Formalin-fixed *S. aureus* cells as previously described (26). The polyacrylamide gel was treated with En 3 Hance (New England Nuclear Corp.) before fluorography on Kodak AXR-5 film.

ual experiments were pooled and lyophilized for subsequent sequence analysis.

Individual peptides were applied to the spinning cup sequencer for automated sequential Edman degradation. This analysis gives the position(s) of the radioactive amino acid with respect to the amino-terminal trypsin cleavage site

that produced the peptide. Examples of the analysis of several 33K peptides are given in Fig. 4 and 5. The patterns of [^3H]leucine and [^3H]proline released from peptides designated L12 and P10 are given in Fig. 4A and B, respectively. These two peptides eluted at the same position during chromatography. Leucine was found in

TABLE 1. Predicted amino acid sequence of adenovirus 2 33K protein

Open Reading frame	Nucleotide no. ^a	HPLC fraction no. ^b	HPLC peak names ^c	Peptide no.	Peptide sequence ^d																			
					1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	609	ND		11	B- Ala	Pro	Lys																	
2	618	FT		12	Lys																			
2	621	FT		13	Lys																			
2	624	179	L12, P10, A9, M3, V7, P9, K8	14	LEU	Gln	LEU	PRO	PRO	PRO	PRO	PRO	Thr	Asp	Glu	Glu	Tyr	Trp	Asp	Ser	Gln	Ala	Glu	
					Glu	Val	Leu	Asp	Glu	Glu	Glu	Met	Met	Glu	Asp	Trp	Asp	Ser	Leu	Asp	Glu	Ala	Ser	
					Glu	Ala	Glu	Glu	Val	Ser	Asp	Glu	Thr	Pro	Ser	Pro	Ser	Val	Ala	Phe	Pro	Ser	Pro	Ala
					Pro	Gln	Lys																	
2	813	97	A7, P7, V5, L10	15	Leu	ALA	Thr	Val	PRO	Ser	ILE	ALA	Thr	Thr	Ser	ALA	PRO	Gln	ALA	PRO	PRO	ALA	Leu	PRO
					Val	Arg																		
2	879	ND		16	Arg	Pro	Asn	Arg																
2	891	FT		17	Arg																			
2	894	ND		18	Trp	Asp	Thr	Thr	Gly	Thr	Arg													
2/3	915/1120	88	A6, P5, V4	19	ALA	VAL	PRO	Thr	ALA	PRO	ALA	ALA	ALA	ALA	ALA	ALA	ALA	Thr	ALA	ALA	VAL	Thr	Gln	Lys
3	1177	FT		110	Gln	Arg																		
3	1183	ND		111	Arg	Pro	Asp	Ser	Lys															
3	1198	ND		112	Thr	Leu	Thr	Lys	Pro	Lys														
3	1216	FT		113	Lys																			
3	1219	44	A3, L3	114	Ser	Thr	ALA	ALA	ALA	ALA	ALA	Gly	Gly	Gly	ALA	Leu	Arg							
3	1258	47	P4, V3, L4	115	Leu	Ala	PRO	Asn	Glu	PRO	VAL	Ser	Thr	Arg										
3	1288	ND		116	Glu	Leu	Arg																	
3	1297	FT		117	Asn	Arg																		
3	1303	142	F8, P9, A8	118	ILE	PHE	Pro	Thr	Leu	Tyr>Ala	ILE	PHE	Gln	Gln	Ser	Arg								
3	1342	ND		119	Gly	Gln	Glu	Gln	Glu	Leu	Lys													
3	1363	ND		120	ILE	Lys																		
3	1369	FT		121	Asn	Arg																		
3	1375	ND		122	Ser	Leu	Arg																	
3	1384	ND		123	Ser	Leu	Thr	Arg																
3	1396	ND		124	Ser	Cys	Leu	Tyr	Arg															
3	1414	ND		125	Ser	Glu	Asp	Gln	Leu	Arg														
3	1432	FT		126	Arg																			
3	1435	62	A5, L5, F6	127	Thr	LEU	Glu	Asp	ALA	Glu	ALA	LEU	PHE	Ser	Lys									
3	1468	ND		128	Tyr	Cys	Ala	Leu	Thr	Leu	Lys													
3	1489	FT		129	Asp																			

^a Nucleotides numbered from *EcoRI* cleavage site at 70.7 m.u. by the method of Galibert et al. (13).^b Approximate fraction number of indicated radioactive peaks. Chromatography conditions are described in the text and in the legend to Fig. 3. FD, Peptide expected in flow-through (fractions 5 through 7) as judged from a comparison of peptide composition with model peptides; ND, position of peptide has not positively been determined.^c Radioactive peaks are named for the labeled amino acid by the single letter code (e.g., A for alanine) and by number in order of elution. The positions of some additional peaks are indicated in Fig. 2; not all peaks have been identified by sequence analysis.^d Amino acid sequence of the expected tryptic peptides of adenovirus 2 33K protein predicted from the nucleotide sequence. The asterisk indicates predicted splice junction; B indicates blocked amino terminus; > indicates the position of an apparent illegitimate trypsin cleavage site (see text). Residues capitalized have been verified by radiochemical microsequence analysis.

FIG. 2. Map of the adenovirus 2 genome. The adenovirus 2 genome is represented by two thin parallel lines in the middle of the picture. Map units are marked below these lines; the positions of *EcoRI* cleavage sites are indicated by solid triangles above these lines; the letter designation of the *EcoRI* fragments is given between the lines. Lines above and below the genome representation indicate RNA transcripts or mRNAs; promoter positions are indicated by a bracket. Individual late mRNA body sequences from the 5 late regions L1 through L5 are indicated by thick lines (10). The approximate location encoding several proteins is indicated near the top of the figure. The lower part of the figure shows a map of just the *EcoRI*-F fragment (70.7 to 75.9 m.u.). The regions encoding the 100K protein (13), the 33K protein (this manuscript), and the precursor to virion component VIII (Oosterom-Dragon and Anderson, unpublished data) are indicated. The exact location of the amino terminus of P-VIII has not been determined. Solid diamonds indicate the positions of *HpaII* restriction sites; the open circle designates the position of the *HindIII* restriction site at 72.8 m.u. Numbers along the bottom of the *EcoRI*-F fragment representation indicate distance in nucleotides from the *EcoRI* site at 70.7 m.u.

amino acid positions 1 and 3 of this peptide, whereas proline was found in positions 4, 5, 6, 7, and 8. Inspection of the *EcoRI*-F nucleotide sequence revealed only one location that could encode a peptide with this pattern of leucines and prolines. The location begins 624 nucleotides from the *EcoRI*-F cleavage site at 70.7 m.u. in the second open reading frame. Immediately preceding the leucine codon at N624 are three lysine codons, a condition consistent with cleavage by trypsin at this location in the protein. Similar analysis of the peptide A7-P7 (Fig. 4C and D) shows it to be uniquely encoded by the sequence beginning at N813, also within the second open reading frame. These results prove that a portion of the 33K protein is encoded by the open reading frame between N307 and N1190. Since the 100K protein is believed to be encoded by the first open reading frame (13, 25), this result strongly suggests that the 33K and 100K proteins cannot be directly related by amino acid sequence. To confirm this expecta-

tion, we prepared [^3H]leucine-labeled 100K tryptic peptides and chromatographed them under conditions identical to those used for separation of the 33K tryptic peptides. No significant coincidence of radioactive profiles was observed (data not shown).

Partial sequence analysis of the four peptides for which data are presented in Fig. 5 revealed surprisingly that these peptides are all encoded in the third open reading frame. Peptide P4-V3 is encoded beginning at N1258, peptide A3 is encoded beginning at N1219, peptide F8 is encoded beginning at N1303, and peptide L5-A5 is encoded beginning at N1435 (Table 1). The data presented in Table 1, therefore, prove that the 33K protein is also partially encoded by the third open reading frame.

Peptide A6 was found to have alanines at positions 1, 2, 5, 7, 8, 9, 10, 11, 12, 13, 15, and 16 (Fig. 4). This pattern of 11 alanines cannot be encoded by a contiguous sequence within the *EcoRI*-F fragment; however, the last 10 alanines

FIG. 3. High-pressure reverse-phase, liquid chromatography of the tryptic peptides of the adenovirus 2 33K protein. SDS-polyacrylamide gel-purified 33K protein, labeled in vivo with [^3H]alanine (A), [^3H]proline (B), [^3H]valine (C), or several other amino acid isotopes ([^3H]isoleucine, [^3H]leucine, [^3H]lysine, [^{35}S]methionine, [^3H]phenylalanine, or [^3H]tyrosine; data not shown) was digested with trypsin as described in the text, and the tryptic digests were fractionated by reverse-phase high-pressure liquid chromatography. The programmed gradient profile used for all experiments is shown in (A); the flow rate was 2.0 ml/min, and 1.0-ml (30-s) fractions were collected. For the three experiments shown, 100 μl of each fraction was counted with 3.5 ml of Aquasol per sample in a Searle Mark III scintillation counter. The counts per minute from these 100- μl samples have been plotted directly without background subtraction. The effluent was also monitored at 214 nm to detect tryptic peptides derived from the apomyoglobin present as carrier in each sample (profiles not shown). Each radioactive peak has been assigned a name derived from the isotope used (single-letter amino acid code) and the relative elution order for that peptide. Appropriate fractions were pooled, mixed with apomyoglobin, and lyophilized in preparation for radiochemical sequence analysis. Approximate positions of tryptic peptides predicted from the nucleotide sequence (Table 1) that have

exactly match the sequence predicted beginning at N1120 in the third reading frame. Furthermore, prolines found in peptide P5 and the valine found in peptide V4, two peptides with elution positions identical to the position of A6, also exactly matched this sequence (Table 1). This result strongly suggested that the coding region for the carboxy-terminal portion of the 33K protein begins at least as far left in the third open frame as N1120, and that an event connecting the second and third open reading frames might occur immediately to the left of this point. Analysis of peptides from open reading frame 2 indicated that the last confirmed residue in this frame is the proline (residue 20) in peptide P7 (Fig. 4D) encoded by N870 to N872. Therefore, the connecting event must occur to the right of this site and within the next 249 nucleotides.

Two separate explanations for the above results are possible. An error might have been made in the sequence between N872 and N1120, or the mRNA for the 33K protein might be spliced within the coding region.

Partial nucleotide sequence analysis of the adenovirus 2 *EcoRI*-F fragment. To address the possibility that a sequencing ambiguity may have caused an apparent shift in the 33K protein reading frame, we independently determined the sequence of selected portions of the *EcoRI*-F fragment. The sequence of 249 nucleotides around the *HindIII* site at 72.8 m.u. was determined by labeling *HindIII*-cleaved genomic DNA with [^{32}P]ATP and polynucleotide kinase, isolating the relevant *EcoRI* subcut fragments (70.7 to 72.8 m.u., 72.8 to 75.9 m.u.), and treating these fragments by the method of Maxam and Gilbert (12, 24). The 153 nucleotides to the left of N736 (to N583) and the 196 nucleotides to the right of N736 (to N932) were found to be identical to the published sequence (13). In a similar way, except starting with the isolated genomic *EcoRI*-F fragment, the 188 nucleotides to the left of the *HpaII* site at N1554 (to N1365) and the 160 nucleotides to the left of the *HpaII* site at N1134 (to N975) were also found to be identical to the published sequence. The 697-nucleotide sequence independently confirmed by us include the sequence immediately before the alanine codon at N1120. Thus the two codons immediately preceding this alanine must be Ser-Thr- and not (Lys/Arg)-Ala- as is suggested by our direct peptide sequence analysis. Our sequence analysis has confirmed all but 43 of the nucleotides between N872 (the Pro in P7) and N1120 (the second Ala in A6).

been identified by sequence analysis are indicated at the top of (A). Predicted tryptic peptides are numbered from the amino terminus (Table 1).

TRYPTIC PEPTIDES OF THE ADENOVIRUS 2 33K PROTEIN

FIG. 4. Radiochemical microsequence analysis of adenovirus 2 33K protein tryptic peptides. Tryptic peptides of radioactively labeled 33K protein were fractionated by reverse-phase chromatography as described in the legend to Fig. 3. Appropriate pooled fractions were applied to a Beckman 890C sequencer together with 2 mg of apomyoglobin and 3 mg of polybrene. Panel pairs depict the total yield of radioactivity in the amino acid fraction after each Edman degradation cycle for corresponding peptides. Peptides are named as described in the legend to Fig. 3, and the isotope with which each peptide was labeled is indicated in the panel. Peptide elution positions can be ascertained from Fig. 3 or Table 1. The amino acid sequence predicted from the nucleotide sequence corresponding to each peptide pair is given near the top of the upper panel in the single-letter amino acid code. Peptides A6 and P5 (panels E and F) correspond to the tryptic peptide that spans the internal splice junction in the 33K mRNA (see text). The amount of radioactivity applied to the sequencer was as follows: (A) 3,900 cpm of [3 H]leucine in L12; (B) 5,300 cpm of [3 H]proline in P10; (C) 1,520 cpm of [3 H]alanine in A7; (D) 13,500 cpm of [3 H]proline in P7; (E) 36,300 cpm of [3 H]alanine in A6; and (F) 6,650 cpm of [3 H]proline in P5.

33K mRNA is internally spliced. The structure of peptide A6 cannot readily be explained by a frameshift event in the vicinity of N1120. Therefore, we examined the possibility that the 33K mRNA is spliced in a manner that connects the two reading frames and explains the structure of peptide A6. The nucleotide sequence between N1112 and N1120 (TCTCTACAG) conforms well to the consensus splice acceptor sequence YNYYYNCAG (34) (Fig. 6). Examination of the sequence to the left of N1120 revealed a potential splice donor site at N918. A splice from this position to the C at N1121 would result in the removal of 202 nucleotides from the transcript for the 33K protein, would preserve the alanine codon that appeared to be encoded by N1120 to N1122, and would cause this codon to be preceded in the mRNA by an Arg-Ala codon pair. This predicted splice is precisely consistent with the observed structure of peptide A6. No better potential splice donor site was found by exami-

nation of the nucleotide sequence between N872 and N1120.

If the 33K mRNA is indeed spliced as suggested above, the 202-nucleotide splice should have been seen by heteroduplex or S1 nuclease analysis of region L4 mRNAs, and certain peptides predicted to be encoded by either the second or third open reading frame should not be identifiable. In fact, a spliced L4 mRNA has been observed both by S1 analysis (8; S. Berget, unpublished data quoted in reference 25) and by heteroduplex analysis (10). Furthermore, we have not been able to find several peptides that we would have expected to observe during reverse-phase chromatography of 33K tryptic peptides. For example, the proline-containing peptide Gln-Pro-Pro-Leu-Ala-Gln-Glu-Gln-Gln-Gln-Arg, predicted by the second open reading frame before the start of the third open frame, has not been found among the proline- or alanine-labeled tryptic peptides.

TRYPTIC PEPTIDES OF THE ADENOVIRUS 33K PROTEIN

FIG. 5. Amino-terminal sequence analysis of adenovirus 2 33K protein tryptic peptides. Sequence analysis was as described in the legend to Fig. 4. The amount of radioactivity applied to the sequencer was as follows: (A) 8,300 cpm of [^3H]proline in P4; (B) 4,550 cpm of [^3H]valine in V3; (e) 9,500 cpm of [^3H]alanine in A3; (D) 1,200 cpm of [^3H]phenylalanine in F8; (E) 5,450 cpm of [^3H]leucine in L5; (F) 5,150 cpm of [^3H]alanine in A5. Additional sequencing experiments with other labels and other peptides are not shown; a summary of all amino acid positions directly confirmed by sequencing is given in Table 1.

To exclude the unlikely possibility that all predicted peptides from the intervening sequence were lost during reverse-phase chromatography, we prepared tryptic peptides of [^3H]isoleucine-labeled 33K protein and subjected the complete mixture to amino-terminal sequence analysis (Fig. 7). The second open reading frame can potentially code for five isoleucines, three of which would come from the predicted intervening sequence. The third open frame could potentially encode six isoleucines, three of which also would come from the intervening sequence. The isoleucine codons in the predicted intervening sequence are situated so that radioactivity would be expected in the 2nd, 3rd, and 14th residues from second reading frame codons, and the 16th, 27th, and 31st residues from third reading frame codons. Radioactivity would be expected in residues 1 (twice), 7, 8, and possibly 3 from the remaining

isoleucine codons. Figure 7 shows the presence of isoleucine in residues 1, 2, 7, and possibly 8, but not significant radioactivity in residues 3, 14, and 16. The presence of isoleucine in residue 2 is explained by an unusual trypsin cleavage that occurs after the tyrosine in peptide t18 (Table 1). Trypsin is known to occasionally cleave after tyrosine, and the existence of the carboxy-terminal fragment of this peptide has been independently confirmed by reverse-phase chromatography and partial sequence analysis. The carboxy-terminal fragment elutes at approximately the same position as t27 (fraction 62, data not shown). Thus, at least two isoleucine peptides from the expected intervening sequence peptides were not found, and all of the expected isoleucines from the peptides predicted to be encoded beginning at N570, except for the isoleucine at position 3 (see below), were observed.

Locations encoding the amino terminus and

PREDICTED ADENOVIRUS 2 33K PROTEIN INTERNAL mRNA SPLICING POINTS

FIG. 6. Predicted internal splice sites in the 33K protein mRNA. Portions of the nucleotide sequence within the *EcoRI*-F fragment are shown. The corresponding open reading frame amino acid sequence is also given. Nucleotides are numbered from the *EcoRI* cleavage site at 70.7 m.u. (13). The consensus splice site sequences are from Sharp (34).

carboxy terminus of the 33K protein. The amino-terminal and carboxy-terminal sequences of the 33K protein have not been determined directly, but several indirect arguments can be applied to deduce their encoding locations. None of the expected peptides predicted by the sequence between N411 and N606 has been observed, including the isoleucine-containing tryptic peptide predicted by N570 to N587. Only one major methionine-containing 33K tryptic peptide was observed by reverse-phase chromatography, and this peptide eluted at the position of peptide t4 (Fig. 3). This 33-amino-acid peptide is expected to contain two methionine residues at positions 29 and 30 from the amino terminus (Table 1). Sequence analysis of the mixture of [35 S]methionine-labeled 33K tryptic peptides did not reveal a methionine in the first 20 residues. Sequence analysis of cyanogen bromide-cleaved, [3 H]alanine-labeled 33K mixed peptides revealed no significant radioactivity in the first 20 residues (data not shown). Thus cleavage after methionine residues produced no unblocked amino termini. These results all suggest that the 33K protein contains only the two methionines expected in peptide t4. Cleavage after these methionines by cyanogen bromide would leave an amino-terminal glutamic acid residue which would be expected to cyclize under the strong acid conditions used. The methionine codon at N606 to N608 is followed by an alanine codon. Where this sequence occurs at known eucaryotic protein synthesis initiation sites (5, 37), the initiation methionine has been found to be removed, and the amino-terminal alanine becomes acetylated. Thus, our findings are compatible with initiation of 33K translation at the N606-608 methionine codon. The sequence surrounding this AUG also matches the most frequently found eucaryotic translation initiation sequence, CANNAUGG (20). If initiation were to occur at the only other available

methionine codon in the second open frame (N411 to N413), we would not expect it to be removed post-translationally, and we would have expected to find evidence for the methionine postulated to be encoded by N606 to N608. Furthermore, the sequences surrounding this AUG codon suggest that it might not be a strong initiation site for translation even if present in the mRNA (20).

The last codon confirmed by peptide sequence analysis in the third open frame is the Phe codon

FIG. 7. Amino-terminal sequence analysis of the mixed [3 H]isoleucine-labeled 33K tryptic peptides. Adenovirus 2 33K protein labeled with [3 H]isoleucine was digested with trypsin. The lyophilized digest containing 20,000 cpm was dissolved in 88% formic acid and applied to the sequencer together with 2 mg of apomyoglobin and 3 mg of polybrene. Sequencing was as described in the text and in the legend to Fig. 4.

TABLE 2. Predicted amino acid composition of adenovirus 2 33K protein^a

Predicted amino acid composition of proteins								
Nonpolar ^b			Noncharged, polar ^b			Charged, polar ^b		
Amino acid	No. of residues	%	Amino acid	No. of residues	%	Amino acid	No. of residues	%
Gly	5	2.2	Ser	18	7.9	Asp	12	5.3
Ala	34	15.0	Thr	19	8.4	Glu	22	9.7
Val	7	3.1	Asn	4	1.8	Lys	14	6.2
Ile	4	1.8	Gln	11	4.8	Arg	17	7.5
Leu	20	8.8	Cys	2	0.9	His	1	0.4
Met	2	0.9	Tyr	4	1.8			
Phe	4	1.8						
Trp	3	1.3						
Pro	24	10.6						

^a Predicted number of amino acids, 227; calculated polypeptide chain molecular weight, 24,892.

^b Nonpolar residues, 45.4%; noncharged, polar residues, 25.6%; charged, polar residues, 29.1%.

from t27 at N1459 to N1461. Eleven codons after this Phe codon, the sequence predicts an UAG termination codon. The existence of this termination codon was confirmed by our nucleotide sequence analysis. Because no obvious splice donor site is located between N1461 and N1493, we conclude that the 33K protein is likely to terminate with aspartic acid as predicted by the nucleotide sequence.

Composition of the 33K protein. The predicted sequence of the 33K protein is given in Table 1 in the form of the expected tryptic peptide sequences. The predicted amino acid composition is given in Table 2. The protein has a high content of alanine (15%), proline (10.6%), and glutamic acid (9.7%). The 33K protein has previously been shown to contain phosphoserine (6), and a preliminary analysis of the ³²PO₄-labeled tryptic peptides by reverse-phase chromatography indicates that a peptide with the mobility of peptide t4 is phosphorylated. This 33-amino-acid peptide contains six serine residues; we have not yet determined which one(s) is phosphorylated. Additional ³²P radioactivity was observed in the column flow-through. We have not yet determined whether this material represents a second phosphorylation site or phosphate released during trypsin digestion. A potential cyclic AMP-dependent kinase phosphorylation site (37) exists in peptide t11 (Table 1). No other peptide that is retarded during reverse-phase chromatography was found to be labeled by ³²PO₄ (data not shown). At least one potential glycosylation site (37) exists in the carboxyterminal portion of the 33K protein (Asn-Arg-Ser-); however, we have no experimental evidence to suggest that this site is actually glycosylated.

DISCUSSION

We have shown that the adenovirus 2 33K protein is a unique, late, nonstructural virus protein that is encoded by two separate portions

of the *Eco*RI-F restriction fragment (70.7 to 75.9 m.u.). Sequence analyses of radioactively labeled peptides and comparison with predictions made from the published nucleotide sequence of this region (13) have allowed us to deduce the precise sites for an internal mRNA splice, the initiation site for translation, and the termination site for 33K protein synthesis. This information was not (13–15) and could not have been readily deduced from the nucleotide sequence alone given our present understanding of the molecular signals required for gene expression.

The proposed sequence of the 33K protein is 227 amino acids long; the calculated molecular weight of this amino acid chain is 24,892, a number considerably smaller than the size estimated from SDS-polyacrylamide gel electrophoresis. Similarly anomalous behavior has been noted for the adenovirus 2 E1A proteins (35). We believe that the high proline and glutamic acid content of these proteins may contribute to their lower-than-expected migration rates during gel electrophoresis. In fact, we estimate the molecular weight of the 33K protein by gel electrophoresis to be closer to 39,000 *M_r* than to 33,000 *M_r*. To avoid confusion, we have retained the name "33K protein" until an appropriate functional designation can be given to this product.

The amino acid sequence of the 33K protein has some unusual features. The region between amino acid residues 103 and 118 (within t9, Table 1) contains 12 alanines, 7 of which are consecutive. This sequence is encoded immediately after the internal mRNA splice and is followed at residue 139 to 146 by another run of five alanines and three glycines. Regions very rich in proline (residues 9 to 13), glutamic acid (residues 16 to 53) or basic amino acids (residues 3 to 5, 90 to 95, and 122 to 136) are also found. Analysis of the protein sequence by the methods of Chou and Fasman (11) suggests the presence

of several short regions of probable α -helix. Two of these are in the amino-terminal portion (residues 22 to 38, and residues 41 to 52) before the mRNA splice point. The corresponding region of the nucleotide sequence probably also codes for the carboxy-terminal portion of the 100K protein; 103 codons are predicted to simultaneously code for both proteins. This is the only region of the adenovirus 2 genome thus far identified that simultaneously encodes two products unrelated in amino acid sequence. The carboxy terminus of the 100K protein is predicted to occur just before the 33K mRNA splice point.

Why is the 33K mRNA spliced? As far as is known, neither strand of the sequence between N918 and N1120 contains signals for other events that might be incompatible with a functional coding sequence for the 33K protein. A promoter sequence for region E2 transcripts does occur on the opposite strand (7), but at a position corresponding to coding sequences for the carboxy terminus of the 33K protein (Fig. 2). No termination codons in either reading frame used by the 33K protein occur within the region removed by splicing. Thus we have found no obvious explanation for the internal splice in the 33K mRNA except to connect the two reading frames and to alter the amino acid sequence that would otherwise result. Perhaps the 33K protein was once two separate functions, or perhaps the region between N918 and N1120 performs some function which has not yet been recognized. While many cellular protein mRNAs are spliced within coding sequences, the only other adenovirus products with mRNAs known to be spliced within coding sequences are the early region 1A products (35). It has been suggested, however, that the mRNAs for component IVa2 and for the region E2B products may be spliced just within the amino-terminal coding region (Jeff Engler and Bruce Stillman, personal communication).

The 33K mRNA presumably corresponds to the L4 mRNA in which a splice has been observed by heteroduplex analysis (10). The structure of this mRNA is not known precisely, but presumably another splice also connects sequences to the left of N606 (the initiation codon) to the tripartite leader sequence. The nucleotide triplet CAG found in the splice acceptor consensus sequence occurs six times between N606 and N411; the codon for the first AUG in the amino-terminal open reading frame is at N411. Four of these CAG sequences are preceded by pyrimidine-rich sequences which are very similar to the consensus splice acceptor sequence. In contrast, the four CAG sequences that occur between N606 and the double-methionine codon at N708 are preceded by purine-rich sequences and do not appear to resemble splice sites.

The function of the 33K protein is unknown. To our knowledge, no mutants have been identified which map within the predicted 33K coding sequences. Several temperature-sensitive mutations are known that map within sequences encoding the 100K protein. Two of these, H5ts115 and H5ts116, have very similar phenotypes, belong to a single complementation group, but map about 1 m.u. apart (27; E. Oosterom-Dragon, Ph.D. thesis, Yeshiva University, Bronx, N.Y., 1980). H5ts116 maps to the left of the *EcoRI* restriction site at 70.7 m.u., whereas H5ts115 maps within the *EcoRI*-F fragment (70.7 to 75.9 m.u.). The similar phenotypes of these two mutants suggest that they are defective in only the same function. Thus we believe that the H5ts115 mutation is likely to be found before N606 and the start of sequences which also encode the 33K protein. Although we have not been able to deduce the function of the 33K protein from knowledge of the nucleotide sequences which encode it, this knowledge may allow the construction of mutants which will help to elucidate its function.

ACKNOWLEDGMENTS

We thank Medora Hardy and Sydel Lamb for excellent technical assistance. We are grateful to John Dunn and William Crockett for assistance with the sequencing of DNA fragments, and to Marshall Elzinga, Nicholas Alonzo, and Jeanne Wysocki for assistance with peptide separation techniques and peptide sequence analysis. We thank Keith Thompson for computer analysis of nucleotide and protein sequences.

This work was supported by Public Health Training grant T32CA09121 from the National Cancer Institute and by the United States Department of Energy.

LITERATURE CITED

1. Akusjärvi, G., J. Zabielski, M. Perricaudet, and U. Pettersson. 1981. The sequence of the 3' non-coding region of the hexon mRNA discloses a novel adenovirus gene. *Nucleic Acids Res.* 9:1-17.
2. Anderson, C. W. 1981. Spontaneous mutants of the adenovirus-simian virus 40 hybrid, Ad2+ND3, that grow efficiently in monkey cells. *Virology* 111:263-269.
3. Anderson, C. W. 1982. Partial sequence determination of metabolically labeled radioactive proteins and peptides, p. 147-167. In J. K. Setlow and A. Hollaender (ed.), *Genetic engineering: principles and methods*, vol. 4. Plenum Publishing Corp., New York.
4. Anderson, C. W., P. R. Baum, and R. F. Gesteland. 1973. Processing of adenovirus 2-induced proteins. *J. Virol.* 12:241-252.
5. Anderson, C. W., and J. B. Lewis. 1980. Amino-terminal sequence of adenovirus type 2 proteins: hexon, fiber, component IX, and early protein 1B-15K. *Virology* 104:27-41.
6. Axelrod, N. 1978. Phosphoproteins of adenovirus 2. *Virology* 87:366-383.
7. Baker, C. C., T. Herissé, G. Courtois, F. Gallibert, and E. Ziff. 1979. Messenger RNA for the Ad2 DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end. *Cell* 18:569-580.
8. Berget, S. M., and P. A. Sharp. 1979. Structure of late adenovirus 2 heterogeneous nuclear RNA. *J. Mol. Biol.* 129:547-565.

9. Bhatti, A. R., and J. Weber. 1979. Protease of adenovirus type 2: partial characterization. *Virology* 96:478-485.
10. Broker, T. R., and L. T. Chow. 1979. Alternative RNA splicing patterns and the clustered transcription and splicing signals of human adenovirus 2, p. 611-635. *In* R. Axel, T. Maniatis, and C. F. Fox (ed.), *Eukaryotic gene regulation*. Academic Press, Inc., New York.
11. Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Annu. Rev. Biochem.* 47:251-276.
12. Dunn, J. J., and F. W. Studier. 1981. Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol.* 148:303-330.
13. Galibert, F., J. Herissé, and G. Courtols. 1979. Nucleotide sequence of the *EcoRI*-F fragment of adenovirus 2 genome. *Gene* 6:1-22.
14. Gambke, C., and W. Deppert. 1981. Late nonstructural 100,000- and 33,000-dalton proteins of adenovirus type 2. I. Subcellular localization during the course of infection. *J. Virol.* 40:585-593.
15. Gambke, C., and W. Deppert. 1981. Late nonstructural 100,000- and 33,000-dalton proteins of adenovirus type 2. II. Immunological and protein chemical analysis. *J. Virol.* 40:594-598.
16. Ginsberg, H. S., and C. S. H. Young. 1977. Genetics of adenovirus, p. 27-86. *In* H. Frankel-Conrat and R. R. Wagner (ed.), *Comprehensive virology*, vol. 9. Plenum Publishing Corp., New York.
17. Hassell, J. A., and J. Weber. 1978. Genetic analysis of adenovirus type 2. VIII. Physical locations of temperature-sensitive mutations. *J. Virol.* 28:671-678.
18. Henderson, L. E., S. Oroszlan, and W. Konigsberg. 1979. A micromethod for complete removal of dodecyl sulfate from proteins by ion-pair extraction. *Anal. Biochem.* 93:153-157.
19. Kessler, S. W. 1975. Rapid isolation of antigens from cells with a staphylococcal protein A-antibody adsorbent: parameters of the interaction of antibody-antigen complexes with protein A. *J. Immunol.* 115:1617-1624.
20. Kozak, M. 1981. Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* 9:5233-5252.
21. Kruijer, W., F. M. A. van Schaik, and J. S. Sussenbach. 1980. Nucleotide sequence analysis of a region of adenovirus 5 DNA-encoding a hitherto unidentified gene. *Nucleic Acids Res.* 8:6033-6042.
22. Levinson, A. D., E. H. Postel, and A. J. Levine. 1977. *In vivo* and *in vitro* phosphorylation of the adenovirus 5 single-strand specific DNA-binding protein. *Virology* 79:144-159.
23. Lewis, J. B., J. F. Atkins, C. W. Anderson, P. R. Baum, and R. F. Gesteland. 1975. Mapping of late adenovirus genes by cell-free translation of RNA selected by hybridization to specific DNA fragments. *Proc. Natl. Acad. Sci. U.S.A.* 72:1344-1348.
24. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* 65:499-560.
25. Miller, J. S., R. P. Ricciardi, B. E. Roberts, B. M. Paterson, and M. B. Mathews. 1980. Arrangement of messenger RNAs and protein coding sequences in the major late transcription unit of adenovirus 2. *J. Mol. Biol.* 142:455-488.
26. Oosterom-Dragon, E. A., and H. S. Ginsberg. 1979. Purification and preliminary immunological characterization of the type 5 adenovirus, nonstructural 100,000-dalton protein. *J. Virol.* 33:1203-1207.
27. Oosterom-Dragon, E. A., and H. S. Ginsberg. 1981. Characterization of two temperature-sensitive mutants of type 5 adenovirus with mutations in the 100,000-dalton protein gene. *J. Virol.* 40:491-500.
28. Palmiter, R. D. 1977. Prevention of NH₂-terminal acetylation of proteins synthesized in cell-free systems. *J. Biol. Chem.* 252:8781-8783.
29. Philipson, L., U. Pettersson, and U. Lindberg. 1975. Molecular biology of adenoviruses. *In* S. Gard and C. Hallaver (ed.), *Virology monographs*, vol. 14. Springer-Verlag, New York.
30. Ricciardi, R. P., J. S. Miller, and B. E. Roberts. 1979. Purification and mapping of specific mRNAs by hybridization-selection and cell-free translation. *Proc. Natl. Acad. Sci. U.S.A.* 76:4927-4931.
31. Russell, W. C., and G. E. Blair. 1977. Polypeptide phosphorylation in adenovirus-infected cells. *J. Gen. Virol.* 34:19-35.
32. Schechter, N. M., W. Davies, and C. W. Anderson. 1980. Adenovirus coded deoxyribonucleic acid binding protein. Isolation, physical properties, and effects of proteolytic digestion. *Biochemistry* 19:2802-2810.
33. Semler, B. L., C. W. Anderson, N. Kitamura, P. G. Rothberg, W. L. Wishart, and E. Wimmer. 1981. Poliovirus replication proteins: RNA sequence encoding P3-1b and the sites of proteolytic processing. *Proc. Natl. Acad. Sci. U.S.A.* 78:3464-3468.
34. Sharp, P. A. 1981. Speculations on RNA splicing. *Cell* 23:643-646.
35. Smart, J. E., J. B. Lewis, M. B. Mathews, M. L. Harter, and C. W. Anderson. 1981. Adenovirus type 2 early proteins: assignment of the early region 1A proteins synthesized *in vivo* and *in vitro* to specific mRNAs. *Virology* 112:703-713.
36. White, D. O., M. D. Scharff, and J. V. Maizel. 1969. The polypeptides of adenovirus III. Synthesis in infected cells. *Virology* 38:395-406.
37. Wold, F. 1981. *In vivo* chemical modifications of proteins. *Annu. Rev. Biochem.* 50:783-814.
38. Zain, S., J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn. 1979. Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. *Cell* 16:851-861.