Proc. Natl. Acad. Sci. USA
Vol. 77, No. 11, pp. 6349–6353, November 1980
Biochemistry

Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: Identification and structure of the signal sequence

(mRNA purification/protein structure analysis/binding to transplantation antigens)

HÅKAN PERSSON*, HANS JÖRNVALL†, AND JAN ZABIELSKI*

*Department of Microbiology, The Biomedical Center, Box 581, S-751 23 Uppsala, Sweden; and †Department of Chemistry, Karolinska Institutet, S-104 01 Stockholm 60, Sweden

Communicated by Philip Leder, July 15, 1980

ABSTRACT **Early region 3 of the adenovirus type 2 genome encodes three proteins with molecular weights of 16,000, 14,500, and 14,000 (E3/16, E3/14.5, and E3/14). The E3/16 protein is the precursor to the E3/19 glycoprotein and is around 1500 daltons larger than the unglycosylated $E3/19_0$ protein. The E3/14.5 and E3/14 proteins are structurally related to each other but different from E3/16. Three mRNA species were identified for E3/16; all have common 5′ ends with the same spliced region but with different 3′ ends. E3/14 was translated from a 13S mRNA with the same 5′ structure as the E3/16 mRNA but followed by a second spliced region with a different 3′ end. A partial amino acid sequence was determined for E3/16 after radioactive labeling *in vitro* and this sequence can be aligned with a known DNA sequence. It contains a hydrophobic signal sequence, two presumptive glycosylation sites, and a hydrophobic region close to the COOH terminus.**

The adenovirus (Ad) genome contains five major transcription units (E1A, E1B, E2, E3, and E4) which are expressed early after infection (for review, see ref. 1). A complex spectrum of spliced mRNAs are transcribed from these regions (2–4). Twelve viral-encoded early polypeptides have been identified by cell-free translation of purified early viral mRNA (5–7). A 19,000-dalton glycoprotein (E3/19) encoded by early region 3 (8) has been purified to homogeneity; it contains ≈25% carbohydrates by weight (8, 9). The glycoprotein is associated with the cell membrane where it is complexed with the cell transplantation antigen (8, 10).

The structures of the E3 mRNAs were recently deduced by RNA/DNA heteroduplex analysis (3, 4). Several mRNAs with common 5′ ends were found early after infection. These mRNAs differ in splicing patterns and several of them do not have coterminal 3′ ends.

In this paper, mRNAs encoded in early region 3 were purified and the proteins encoded by the different mRNAs were identified in a cell-free protein-synthesizing system. The primary structure of the E3/19 glycoprotein was also determined by aligning the protein NH_2-terminal sequence with the known DNA sequence (11).

MATERIALS AND METHODS

Procedures for extraction of early viral mRNA, cell-free synthesis in mRNA-dependent reticulocyte lysate, preparation of rough microsomes from dog pancreas, tryptic peptide analysis, immunoprecipitations, and $NaDodSO_4$/polyacrylamide gel electrophoresis have been described (8, 12). Restriction enzyme fragments of Ad 2 DNA were prepared (13) and their purity was determined by agarose gel electrophoresis before use. Ad RNA was purified by hybridization to Ad 2 DNA bound to nitrocellulose filters (14).

Heteroduplex Analysis. Purified Ad mRNA was mixed with 0.25 μg of Ad 2 DNA and the mixture was precipitated with ethanol. The sample was dissolved in 0.1 M Hepes, pH 7.6/0.4 M NaCl/0.01 M EDTA, and purified formamide was added to a final concentration of 80%. The final volume was 50 μl and the samples were heated to 65°C for 5 min followed by immediate cooling on ice. Hybridizations were performed at 53°C for 12–15 hr. Samples were spread from hyperphases of 40% formamide across hypophases containing 10% formamide, or from 45% to 15%. The RNA·DNA hybrids were measured and each value was normalized relative to φX174 DNA (5386 bases) or PM2 DNA (9950 base pairs) included as size markers.

Amino Acid Sequence Analysis. Purified Ad mRNAs were translated *in vitro* in the presence of 200 μCi (1 Ci = 3.7×10^{10} becquerels) of ^3H-labeled amino acids. The translation products were purified by $NaDodSO_4$/polyacrylamide gel electrophoresis with [^{35}S]methionine-labeled proteins as markers. The proteins were eluted from the gel (12), mixed with 1 mg of bovine serum albumin, and precipitated with trichloroacetic acid. Sequence analysis was performed in a Beckman 890C sequencer together with 1 mg of apomyoglobin. Degradations were performed for 20–25 cycles with a 0.1 M Quadrol program in the presence of Polybrene (15). Repetitive yields of both the labeled protein and the carrier ranged between 92% and 95%.

RESULTS

Cell-Free Synthesis with Purified mRNAs. Restriction enzyme fragments from the E3 region of the viral genome were prepared and used for selection of viral mRNAs (Fig. 1). The selected mRNAs were translated *in vitro* and the [^{35}S]methionine-labeled products were analyzed by $NaDodSO_4$/polyacrylamide gel electrophoresis. The *Eco*RI D and *Hin*dIII H fragments selected mRNAs for two early viral polypeptides with sizes of 16,000 and 14,000 daltons (E3/16 and E3/14) (Fig. 2). The *Hin*dIII L fragment selected the mRNA for the E3/16 protein; the mRNA for the E3/14 protein was predominantly selected by the *Eco*RI E fragment. This fragment also selected the mRNA for a 14,500-dalton protein (E3/14.5) as well as small amounts of the E3/16 protein. The E3/16 protein is ≈1500 daltons larger than the unglycosylated $E3/19_0$ protein synthesized in the presence of tunicamycin (Fig. 2, lane b). Restriction enzyme fragments derived from early regions 2 and 4 (*Eco*RI B + C, Fig. 2) selected mRNAs encoding a different spectrum of early proteins (E2/75, E4/26, E4/18.5, E4/17.5, E4/12.5, and E4/11). The cell-free system was also supplemented with rough microsomes before the start of incubation and the translation products were immunoprecipitated with an anti-E3/19 antiserum. mRNA selected by hybridization to

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.

Abbreviation: Ad, adenovirus.

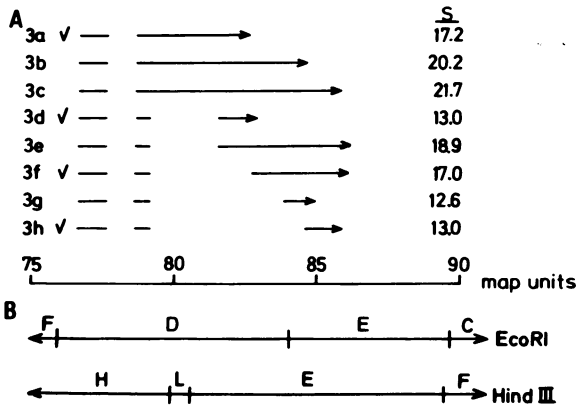

FIG. 1. Diagram of Ad2 early region 3. (*A*) Early region 3 transcripts (3, 4). Calculated sedimentation coefficients are shown to the right, and the major RNA species are indicated by check marks. (*B*) Restriction enzyme fragment map of early region 3.

the *Eco*RI D and *Hin*dIII L fragments directed the synthesis of the mature glycoprotein under these conditions (Fig. 2, lanes c–e).

Tryptic Peptide Analysis. The mature E3/19 glycoprotein contained five [^{35}S]methionine-labeled tryptic peptides (Fig. 3). The E3/16 protein contained the same tryptic peptides and one additional peptide (peptide 6). This peptide was more heavily labeled with [^{35}S]methionine, suggesting that it contained more than one methionine residue. The E3/14.5 and the E3/14 polypeptides contained [^{35}S]methionine-labeled tryptic peptides that were similar to each other but different from any of the peptides found in the E3/16 protein.

Fractionation of mRNA from Region 3. mRNA selected by hybridization was further purified by sucrose gradient centrifugation. Labeled RNA selected on the *Eco*RI D fragment sedimented in two peaks (Fig. 4*A*). The major peak had a sedimentation coefficient of ≈17 S; the minor peak sedimented at ≈13 S. The 17S RNA directed the *in vitro* synthesis of the E3/16 protein; RNA from the 13S peak synthesized the E3/14 protein. The E3/16 protein was also synthesized to a lesser extent from a mRNA sedimenting slightly faster than the 18S ribosomal RNA marker.

Labeled RNA selected by hybridization to the *Eco*RI E fragment also sedimented in two peaks (Fig. 4*B*), corresponding to 13 S for the major and 19–20 S for the minor peak. The E3/14

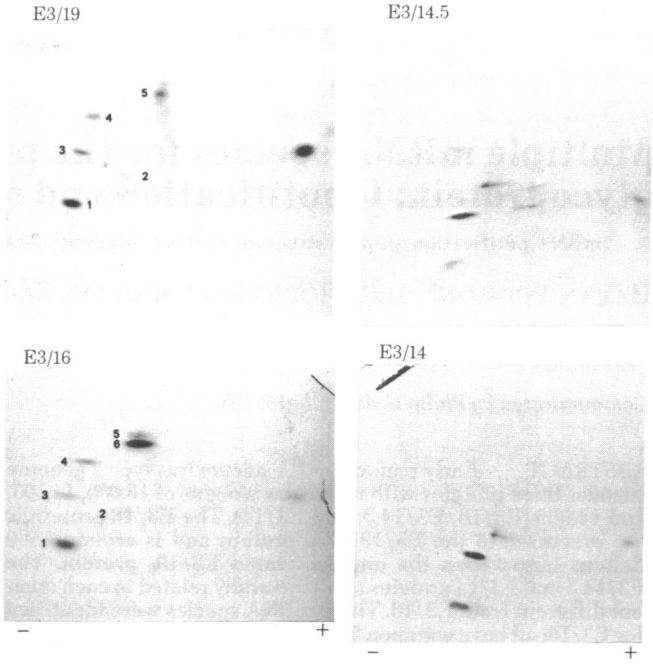

FIG. 3. Tryptic peptide analysis. The E3/19 glycoprotein was immunopreciptiated after labeling with [^{35}S]methionine *in vivo*. E3/16, E3/14.5, and E3/14 were synthesized *in vitro* in the presence of [^{35}S]methionine. The proteins were purified by NaDodSO$_4$/polyacrylamide gel electrophoresis before tryptic digestion. Chromatography in the second dimension was from bottom to top.

protein was translated from the 13S peak, and E3/16 protein synthesis was directed by the 19–20S peak. The 19–20S peak also contained mRNA encoding the DNA-binding E2/75 protein. This mRNA may initiate transcription at a low frequency at the right-hand end of the *Eco*RI E fragment (3). It is therefore possible that a minor species of the E2/75 mRNA was selected on the *Eco*RI E fragment.

Heteroduplex Analysis of RNA·DNA. Purified early region 3 mRNAs were hybridized to Ad 2 DNA or restriction fragments thereof and the resulting heteroduplexes were visualized in the electron microscope. Ad 2 mRNA selected on the *Hin*dIII L fragment had its 5′ end located at coordinate 76.7(±0.2) with an intron of ≈430 nucleotides from 77.8(±0.2) to 79.0(±0.2) (Fig. 5*A*). About two-thirds of the molecules had a 3′ terminus located at position 82.8(±0.2); the remaining third had a 3′ end

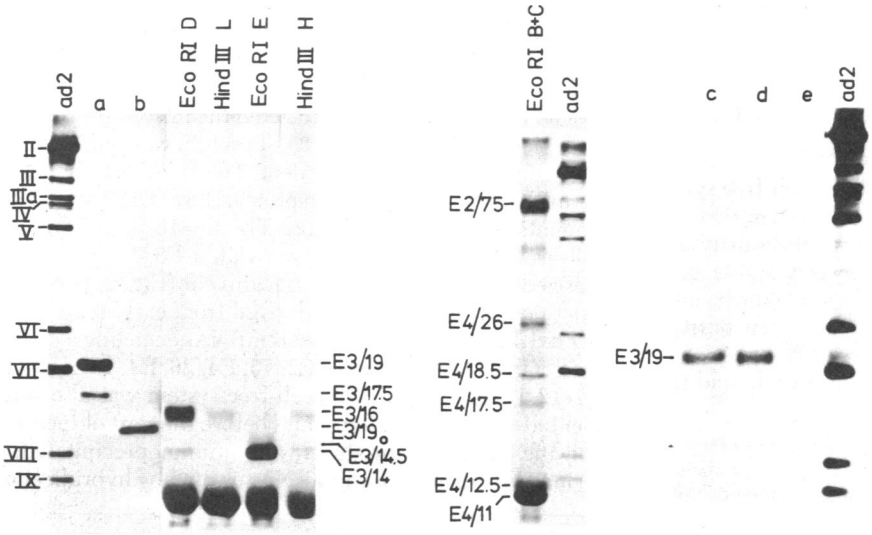

FIG. 2. Ad 2-encoded early region 3 proteins. Early viral RNA was hybridized to restriction enzyme fragments of Ad 2 DNA and the purified RNA was translated *in vitro*. The translation products were analyzed by NaDodSO$_4$/polyacrylamide gel electrophoresis followed by fluorography. Lanes a and b, mature E3/19 glycoprotein and E3/19$_0$ protein synthesized in the presence of tunicamycin after immunoprecipitation from cells labeled *in vivo*. The cell-free system was supplemented with rough microsomes and the translation products were immunoprecipitated with the anti-E3/19 antiserum (lanes c and d) or a preimmune serum (lane e). The cell-free system was programmed with mRNA selected by hybridization to *Eco*RI D (lanes c and e) or *Hin*dIII L (lane d). Here and in subsequent figures, ad2 denotes a [^{35}S]methionine-labeled Ad 2 marker.

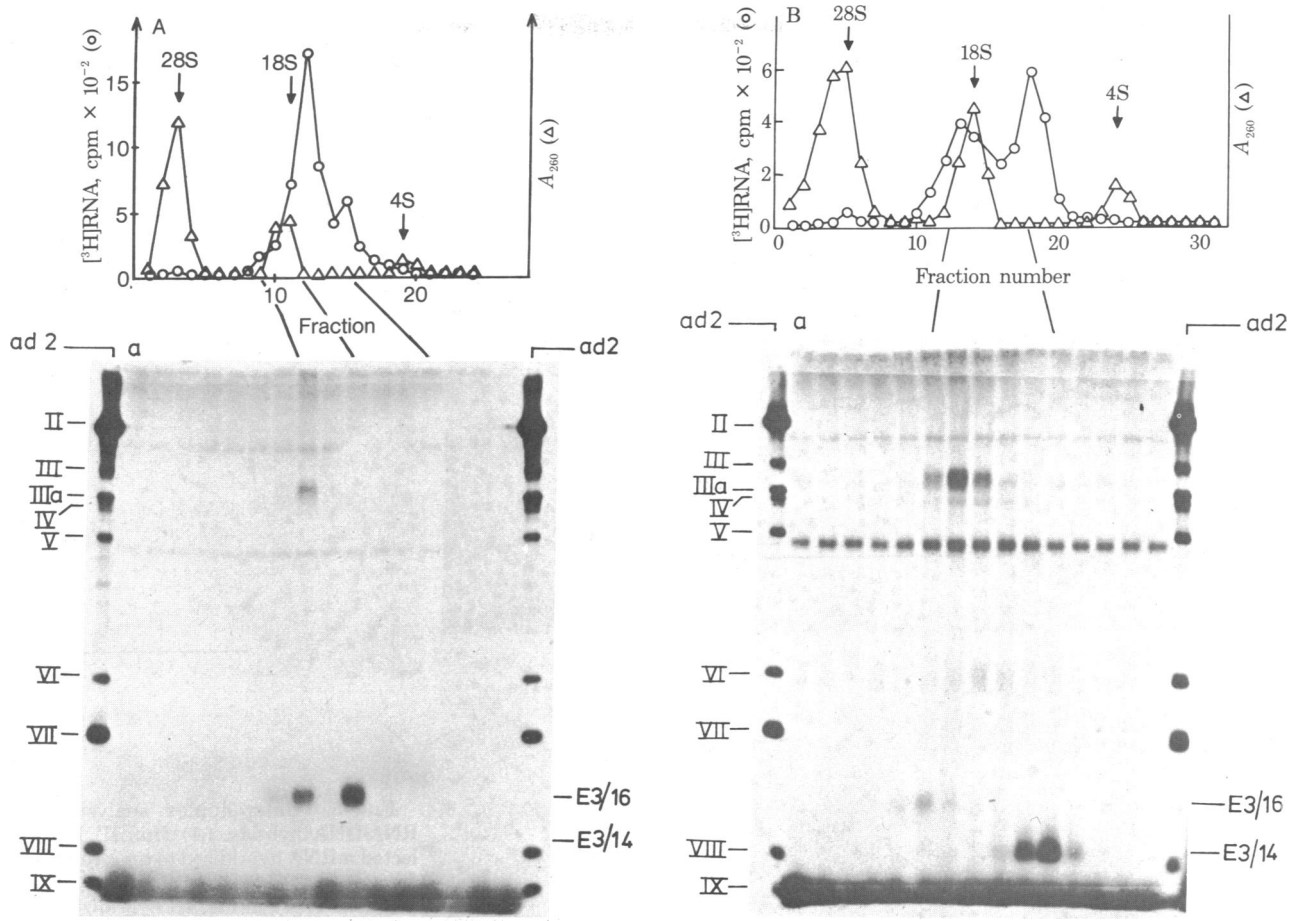

FIG. 4. Size fractionation of early region 3 mRNAs. Early viral mRNA labeled with [^{3}H]uridine and purified by hybridization to restriction enzyme fragments *Eco*RI D (*A*) and *Eco*RI E (*B*) was sedimented on a 15–30% sucrose gradient. RNA from each fraction was precipitated with ethanol and translated *in vitro*. The [^{35}S]methionine-labeled products were analyzed by NaDodSO$_4$/polyacrylamide gel electrophoresis followed by fluorography. Lane a, no mRNA added. Cytoplasmic RNA prepared from mock-infected cells was used as size markers.

to the right of the *Eco*RI D fragment. The E3/14 mRNA was obtained from the sucrose gradient shown in Fig. 4*B* (fractions 17–19). This mRNA had a 5′ end at coordinate 76.4(±0.2) followed by an intron at coordinates 77.3(±0.2) to 78.5(±0.2) (Fig. 5*B*). The molecules also had a second intron between coordinates 79.0(±0.3) and 84.3(±0.3). The 3′ end of the mRNA was located at coordinate 85.3(±0.3). The 19–20S mRNA (fractions 11–13 in Fig. 4*B*) was also analyzed by RNA·DNA heteroduplex formation. This mRNA was a mixture of the 3b and 3c species shown in Fig. 1*A* with a 5′ end at coordinate 76.7(±0.2), an intron between 77.8(±0.2) and 79.0(±0.2), and 3′ ends at coordinate 84.6(±0.2) (3b) or 85.8(±0.3) (3c) (Fig. 5*C*).

Alignment of the E3/16 Protein with the DNA Sequence. Protein NH_2-terminal sequence analysis was performed on the *in vitro*-synthesized E3/16 protein in order to align the protein in the DNA sequence of the *Eco*RI D fragment (11). The E3/16 protein was synthesized *in vitro* in the presence of different radiolabeled amino acids, one at each time, and the NH_2-terminal sequence was determined by liquid-phase sequencer analysis. The protein contained methionine at positions 1 and 4, leucine at positions 6, 8, 9, and 11, isoleucine at position 5, and lysine at positions 19, 20, and 24 (Fig. 6).

Positions analyzed could be completely aligned with the DNA sequence and show that the protein initiates at coordinate 80.3. From here the DNA sequence contains an open reading frame for 159 amino acids (11). The sequence contains a hydrophobic presumptive signal sequence of maximally 18 amino acids including the initiating methionine residue (Fig. 7).

DISCUSSION

The present results show that the 3a mRNA shown in Fig. 1*A* (3, 4) encodes a 16,000-dalton polypeptide (E3/16). The E3/16 protein is also translated to a minor extent from the 3b and 3c mRNAs having the same 5′ end as the 3a species but with alternative 3′ ends. The 3b and 3c mRNAs are less frequent and represent together about one-third of the E3/16 mRNAs (3, 4). The presence of alternative 3′ ends in mRNA molecules encoding the same protein is unique to region 3 but a possibly analogous situation is found in transcripts of the early region of polyoma virus where a minor class of the three differentially spliced mRNAs have an alternative site of 3′-polyadenylylation located just after the termination codon for the middle-size tumor antigen (16). The coding function of these mRNAs in productively infected cells is not known, but certain transformed lines whose integrated genomes lack the principal polyadenylylation sequence express predominantly the alternatively polyadenylylated mRNA class and synthesize both small and middle-size tumor antigens (16).

The E3/14 protein was translated from the 3h mRNA shown in Fig. 1*A*. This mRNA has the same 5′ structure as the E3/16 mRNAs but contains a second intron and a 3′ end different from that of the E3/16 mRNAs. Early region 3 mRNAs are transcribed from a promoter at coordinate 76.5 (17) which is very close to the 5′ ends of all region 3 mRNAs. Because the E3/16 mRNAs only contain one intron sequence they might serve as precursors to the E3/14 mRNA. In accordance with this, the E3/19 glycoprotein is synthesized ≈2 hr earlier after infection than is the E3/14 protein (9, 12).

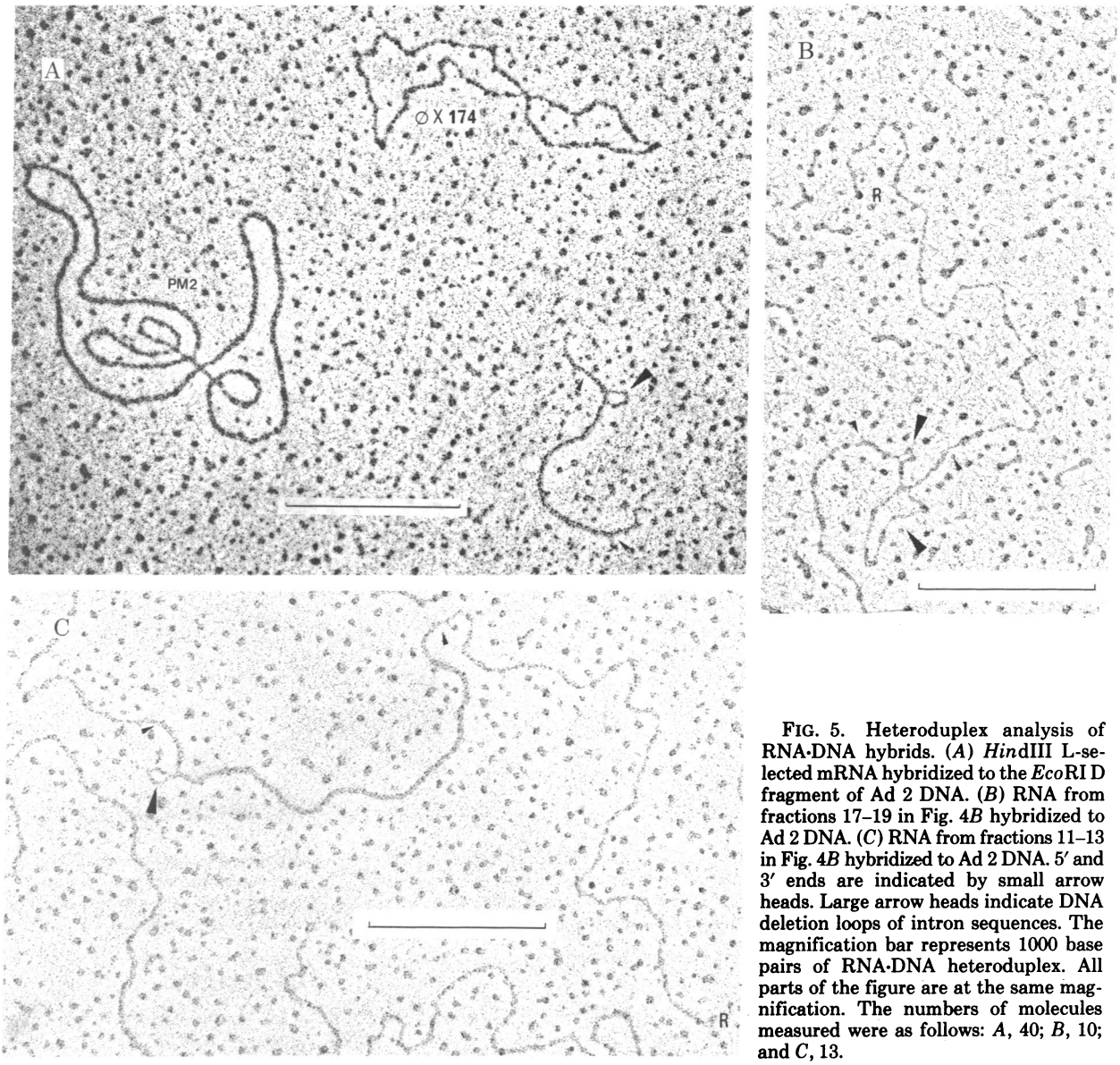

FIG. 5. Heteroduplex analysis of RNA·DNA hybrids. (*A*) *Hin*dIII L-selected mRNA hybridized to the *Eco*RI D fragment of Ad 2 DNA. (*B*) RNA from fractions 17–19 in Fig. 4*B* hybridized to Ad 2 DNA. (*C*) RNA from fractions 11–13 in Fig. 4*B* hybridized to Ad 2 DNA. 5′ and 3′ ends are indicated by small arrow heads. Large arrow heads indicate DNA deletion loops of intron sequences. The magnification bar represents 1000 base pairs of RNA·DNA heteroduplex. All parts of the figure are at the same magnification. The numbers of molecules measured were as follows: *A*, 40; *B*, 10; and *C*, 13.

The E3/14 protein is structurally related to the E3/14.5 protein. The latter protein is a minor polypeptide synthesized *in vitro* and it may either be encoded by the minor RNA species 3e, f, and g, read in the same frame as the 3h mRNA, or constitute a posttranslationally modified form of the E3/14 protein.

The E3/16 protein is structurally related to the E3/19 glycoprotein. Addition of rough microsomes to the translation mixture programmed with purified E3/16 mRNAs allowed the synthesis of the mature glycoprotein (Fig. 2). These results show that the E3/16 protein is the precursor to the glycoprotein. The amino acid sequence of the E3/16 protein shows a 17-amino-acid-long hydrophobic sequence at its NH_2 terminus (ref. 11; Fig. 7). The following results suggest that this region constitutes a signal sequence.

i. The E3/16 protein is ≈1500 daltons larger than the tunicamycin product ($E3/19_0$).

ii. In accordance with the predicted amino acid sequence for the E3/16 protein, showing methionine at positions 1 and 4, the tryptic peptides of this protein contain one [^{35}S]methionine-labeled peptide (peptide 6 in Fig. 3) not found in the E3/19 glycoprotein.

iii. Addition of rough microsomes before the start of cell-free synthesis followed by protease treatment showed that the E3/19 protein was inserted into the vesicles (not shown).

iv. Sequence analysis of the E3/19 glycoprotein labeled *in vivo* with [^{3}H]lysine indicated a possible cleavage after position 17 in the E3/16 protein (not shown).

The signal sequence is contiguous with the sequence present in the mature protein and it is not encoded in a separate exon as observed for some other proteins [mouse immunoglobulin light chain (18) or conalbumin (19)]. There are two potential sites for carbohydrate attachment in the protein based on the presence of the sequence -Asn-Val-Thr- which is typical for glycosylation. The COOH terminus of the protein contains a 23-amino-acid-long region of uncharged predominantly hydrophobic amino acids, followed by a hydrophilic region of 15 amino acids. This resembles the structure for the HLA antigens and glycophorin (20, 21), two proteins that span the cell membrane and have a hydrophilic COOH terminus on the cytoplasmic side of the membrane. The predicted structure for the E3/19 protein confirms that this glycoprotein is associated with the membranes in a similar fashion as demonstrated earlier (8). Both glycosylation sites are then located on the exterior side

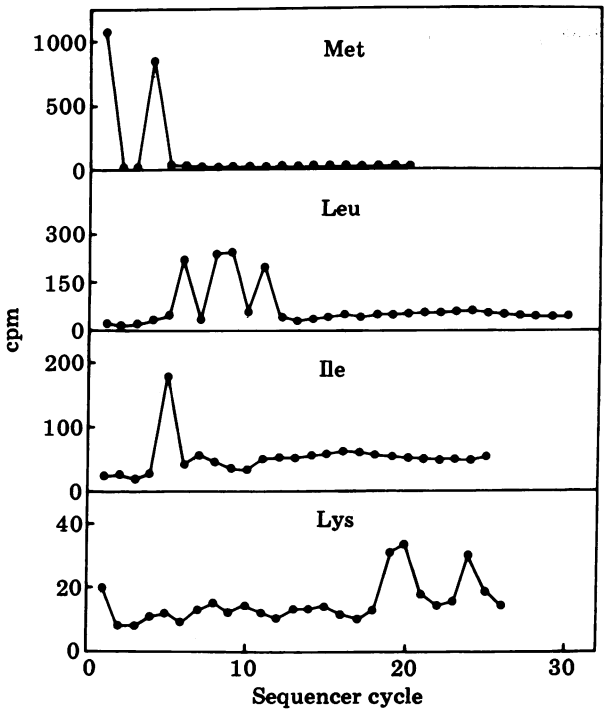

FIG. 6. Sequence analysis of the E3/16 protein. Radioactive E3/16 protein was prepared and the radioactivity in each cycle of sequence degradation was determined. The amino acids used for labeling are indicated.

of the membrane. The E3/19 glycoprotein is complexed with the major transplantation antigen of the host cell both during lytic infection in human cells as well as in one Ad 2-transformed rat cell line (10, 22). This complex appears to be recognized by cytotoxic T cells causing lysis of Ad-infected and -transformed cells after incubation with sensitized T lymphocytes from syngeneic hosts (10, 22). In the transformed cell almost all of the viral glycoprotein is associated with the heavy chain of the transplantation antigen whereas the E3/19 glycoprotein appears to be in a molar excess during lytic infection. The similar orientation of the two glycoproteins in the membranes suggests that complex formation occurs within the membrane. Defined peptides can now be prepared from the viral glycoprotein in order to identify the region that interacts with the heavy chain of the transplantation antigens.

We thank Dr. F. Galibert for communicating the DNA sequence of the *Eco*RI D fragment before publication. We also thank Dr. G. Akusjärvi, L. Philipson, and U. Pettersson for helpful discussions. C. Sjöholm and M. Gustafson provided expert secretarial help. This work was supported by grants from the Swedish Cancer Society and Knut and Alice Wallenberg's Foundation.

1. Philipson, L. (1979) *Adv. Virus Res.* **25**, 357–405.
2. Berk, A. J. & Sharp, P. A. (1978) *Cell* **14**, 695–711.
3. Chow, L. T., Broker, T. R. & Lewis, J. B. (1979) *J. Mol. Biol.* **134**, 265–303.
4. Kitchingman, G. R. & Westphal, H. (1980) *J. Mol. Biol.* **137**, 23–48.
5. Lewis, J. B., Atkins, J. F., Baum, P. R., Solem, R., Gesteland, R. F. & Anderson, C. W. (1976) *Cell* **7**, 141–151.
6. Harter, M. L. & Lewis, J. B. (1978) *J. Virol.* **26**, 736–749.
7. Persson, H., Perricaudet, M., Tolun, A., Philipson, L. & Pettersson, U. (1979) *J. Biol. Chem.* **254**, 7999–8003.
8. Persson, H., Jansson, M. & Philipson, L. (1980) *J. Mol. Biol.* **137**, 375–394.
9. Persson, H., Signäs, C. & Philipson, L. (1979) *J. Virol.* **29**, 938–948.
10. Kvist, S., Östberg, L., Persson, H., Philipson, L. & Peterson, P. A. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5674–5678.
11. Hérissé, J., Courtois, G. & Galibert, F. (1980) *Nucleic Acids Res.* **8**, 2173–2192.
12. Persson, H., Pettersson, U. & Mathews, M. B. (1978) *Virology* **90**, 67–79.
13. Pettersson, U., Tibbetts, C. & Philipson, L. (1976) *J. Mol. Biol.* **101**, 479–501.
14. McGrogan, M., Spector, D. J., Goldenberg, C. J., Halbert, B. & Raskas, H. J. (1979) *Nucleic Acids Res.* **6**, 593–607.
15. Jörnvall, H. & Philipson, L. (1980) *Eur. J. Biochem.* **104**, 237–247.
16. Kamen, R., Favaloro, J., Parker, J., Treisman, R., Lania, L., Fried, M. & Mellor, A. (1980). *Cold Spring Harbor Symp. Quant. Biol.* **44**, 63–75.
17. Baker, C. C. & Ziff, E. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **44**, in press.
18. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* **15**, 1133–1144.
19. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. & Chambon, P. (1979) *Nature (London)* **282**, 567–574.
20. Springer, T. A. & Strominger, J. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2481–2485.
21. Jokinen, M., Gahmberg, C. G. & Andersson, L. C. (1979) *Nature (London)* **279**, 604–607.
22. Persson, H., Kvist, S., Östberg, L., Peterson, P. A. & Philipson, L. (1980) *Cold Spring Harbor Symp. Quant. Biol.* **44**, in press.

[(Met) Arg Tyr (Met) (Ile) (Leu) Gly (Leu) (Leu) Ala (Leu) Ala Ala Val Cys Ser Ala] Ala (Lys) (Lys)

Val Glu Phe (Lys) Glu Pro Ala Cys Asn Val Thr Phe Lys Ser Glu Ala Asn Glu Cys Thr Thr

Leu Ile Lys Cys Thr Thr Glu His Glu Lys Leu Ile Ile Arg His Lys Asp Lys Ile Gly Lys

Tyr Ala Val Tyr Ala Ile Trp Gln Pro Gly Asp Thr Asn Asp Tyr Asn Val Thr Val Phe Gln

Glu Asn Arg Lys Gly Thr Phe Met Tyr Lys Phe Pro Phe Tyr Glu Met Cys Asp Ile Thr Met Tyr

Ser Lys Gln Tyr Met Lys Leu Trp Pro Pro Gln Lys Cys Leu Glu Asn Thr Gly Thr Phe Cys

Thr Ala Leu Leu Ser Ile Thr Ala Leu Ala Leu Val Cys Thr Leu Leu Tyr Leu Lys Tyr Lys Ser

Arg Ser Phe Ile Arg Asp Glu Lys Lys Met Pro

FIG. 7. Predicted amino acid sequence for the E3/16 protein as deduced by alignment with the DNA sequence (11). Residues identified by sequence analysis are indicated by circles. A presumptive signal sequence in the protein is indicated in brackets. The two tripeptides -Asn-Val-Thr- are underlined as well as a predominant hydrophobic noncharged region close to the COOH terminus of the protein.