JOURNAL OF VIROLOGY, Jan. 1983, p. 185–191
0022-538X/83/010185-07$02.00/0

Vol. 45, No. 1

Sequence Arrangement and Protein Coding Capacity of the Adenovirus Type 2 "i" Leader

EILEEN FALVEY† AND EDWARD ZIFF*

Department of Biochemistry, New York University Medical Center, New York, New York 10016

Received 6 August 1982/Accepted 6 October 1982

The adenovirus type 2 (Ad-2) "i" leader is an RNA segment which is preserved in some mRNA species from the Ad-2 late transcription unit. It maps between the second and third segments of the standard tripartite leader. We located the boundaries of the i leader in genomic Ad-2 DNA and determined its nucleotide sequence. The leader contains an ATG initiator near its 5′ boundary, followed by a reading frame which is open for translation. We suggest that the i leader constitutes an Ad-2 coding sequence whose novel position within the leader of major late transcription unit messengers allows it to be translated in preference to coding sequences in mRNA main bodies. The i leader potentially contributes to the coding sequences of a family of proteins. Also, a Northern blot analysis of late mRNAs containing the i leader suggests that it may be retained in the leaders of many different late transcription unit mRNAs. We compare the i leader to the simian virus 40 agnogene.

The adenovirus type 2 (Ad-2) major late transcription unit (T.U.) employs a novel arrangement of genes and RNA processing signals to express upwards of 15 different viral proteins from transcripts initiated at a single promoter (for a review, see reference 21). Transcripts are initiated at the major late promoter at 16.45 map units (m.u.) (22) and at late times extend essentially to the far end of the genome at 100 m.u. (8). Five polyadenylic acid [poly(A)] sites for late mRNA lie within the late T.U. (6). The 3′ ends of late T.U. mRNAs are formed by endonucleolytic cleavage of primary transcripts at one of these sites, followed by 3′ poly(A) addition (15). For RNAs with a 3′ end at any given poly(A) site, a number of different spliced arrangements is possible. Most frequently, the RNAs contain a 5′ tripartite late leader which is transferred by splicing to one of several alternative splice acceptor sequences in the mRNA main body. Messengers differing in the spliced position of the leader, but sharing a common poly(A) end, form a 3′-coterminal family (23). The late T.U. is also active at early times, but only the family with the first poly(A) site is expressed (3, 17).

In most cases, the different members of a 3′-coterminal family have different protein translation products. The translational differences reflect the ribosome mechanism for selecting the translation start site. Most often, translation initiates at the first AUG triplet downsteam from the cap. The ribosome binds to the cap and scans the RNA from 5′ to 3′ until it encounters an AUG, whereupon it initiates translation (see reference 13 for a review). Because the tripartite leader lacks an AUG, it is noncoding (1, 20). The AUG initiation triplet therefore lies beyond the tripartite leader, and in those cases investigated thus far, the start codon lies in the main body of the mRNA (1, 20). Because the alternative splice arrangements juxtapose the leader (and the scanning ribosome) adjacent to different AUGs, each member of a coterminal family has a separate translation product.

Using the electron microscope heteroduplex method, Chow et al. (3) have identified alternative forms of the late leader. Each form contains an extra segment called the "i" leader encoded between m.u. 22 and 23.5, between the standard second and third segments of the tripartite leader. The 3′ boundary and exonic sequences of this leader may, however, vary to include different RNA segments up to and including the third segment (3).

Our initial interest in the i leader was to identify the steps in the RNA splicing pathway that determined whether the i leader RNA is retained or excluded from late T.U. messengers. With the ultimate objective of analyzing splicing pathways, we determined the genomic nucleotide sequence which encodes the i leader and the positions of the acceptor and donor splice boundaries. We find that the major form of the i leader contains an AUG initiation triplet fol-

† Present address: Department of Molecular Biology and Biophysics, Yale University, New Haven, CT 06520.

lowed by an in-phase reading frame. We discuss the coding potential of this open frame with respect to late mRNA main body coding sequences and the multiple i leader forms described by Chow et al. (3) and compare these findings with "agnogene" sequences in simian virus 40.

MATERIALS AND METHODS

Cell culture, virus infection, and RNA isolation. Ad-2 was grown in Spinner-cultured HeLa cells. Late-infection polyadenylated [poly(A)$^+$] cytoplasmic RNA was isolated as described by Fraser and Ziff (6) and Fraser et al. (7). Early RNA was similarly isolated but at 18 h postinfection in the presence of a 25-μg/ml cytosine arabinoside block of DNA replication.

DNA fragment pareparation. Ad-2 DNA clones in pBR322 propagated with *Escherichia coli* HB101 host were used for nucleotide sequence analysis and preparation of nick-translated probes. Plasmids were p*Hin*dIII-B (17 to 31.5 m.u., *Hin*dIII fragment B at the pBR322 *Hin*dIII site); p*Bam*-D (29 to 41 m.u., *Bam*HI fragment D at the pBR322 *Bam*HI site); p*Hin*dIII-I (31.5 to 37.3 m.u., *Hin*dIII fragment I cloned at the pBR322 *Hin*dIII site); and p*Sma*-B (18.5 to 35.4 m.u., *Sma*I fragment B cloned at the pBR322 *Pst*I site by guanine-cytosine tailing). The former two plasmids were the gift of M. Mathews, the next was from J. Nevins, and the latter was from U. Pettersson. Restriction enzymes were from New England Biolabs or Bethesda Research Laboratories and were used according to the manufacturer's instructions.

DNA nucleotide sequencing and S1 nuclease mapping of the i leader boundary. DNA sequencing was by the method of Maxam and Gilbert (14). DNA was 5′ labeled by the polynucleotide kinase method. The *Xho*I site at 22 m.u. was 3′ labeled by incorporating α-^{32}P-triphosphates through repair synthesis catalyzed by the Klenow fragment of DNA polymerase I. The i leader splice boundaries were mapped by a modification of the S1 nuclease hybridization procedure of Berk and Sharp (2). Plasmid p*Sma*I-B (18.5 to 35.4 m.u.) was restricted with *Hin*cII, and 5′-phosphates were removed with bacterial alkaline phosphatase. The 5′ termini were $^{32}PO_4$ labeled by the polynucleotide kinase method, and fragments were resolved by agarose gel electrophoresis. One fragment, approximately 930 nucleotides long, which overlapped the 5′ boundary of the i leader, was eluted and rerestricted with *Alu*I. One labeled band, approximately 240 nucleotides long, overlapped the 5′ boundary of the leader and was eluted for S1 hybridization to poly(A)$^+$ cytoplasmic RNA from late infected cells under the conditions of Berk and Sharp (2). Hybrids of DNA 3′ labeled at the 22-m.u. *Xho*I site were also prepared. An *Xho*I digest of p*Hin*dIII-B was 3′ labeled by repair synthesis catalyzed by the Klenow fragment of DNA polymerase I and utilizing α-^{32}P-triphosphate. The 22.0- to 28.1-m.u. *Xho*I fragment was then isolated. This was rerestricted with *Kpn*I at 23.5 m.u., and the fragment labeled at 22.0 m.u. was employed for S1 mapping. DNA-RNA hybrids were S1 nuclease digested to remove single-stranded regions and electrophoresed on thin acrylamide sequencing gels. The neighboring lanes contained Maxam-Gilbert digests of the same labeled fragment for nucleotide sequence analysis. The S1 nuclease bands migrated near the bands in the sequencing lane corresponding to the i leader 5′ or 3′ splice boundaries.

Northern blot analysis of nuclear and cytoplasmic poly(A)$^+$ RNA. RNA was electrophoresed in agarose gels with running buffer (0.02 M morpholinepropanesulfonic acid, 0.005 M sodium acetate, 0.001 M EDTA [pH 7.0]) and 1.1 M formaldehyde. Before being loaded, RNA was heated to 60°C in 50% formamide–3% formaldehyde plus running buffer and then applied to the gel. After electrophoresis, RNA was transferred to nitrocellulose sheets by the Southern blotting technique. RNA bands were located by hybridization for 12 h to α-^{32}P-labeled nick-translated probes at 42°C with hybridization buffer (50% formamide, 0.05 M Tris-hydrochloride [pH 7.4], 0.1% sodium dodecyl sulfate, 10× Denhardt solution, 1 M NaCl, 50 μg of denatured salmon sperm DNA per ml). After being washed, blots were visualized by autoradiography employing intensifying screens.

RESULTS

Figure 1 shows the arrangement of i leader RNA sequences in late T.U. mRNAs as reported by Chow et al. (3). The cap site and first leader of late mRNA map at coordinate 16.4. The second and third leader segments map at 19.6 and 26.5 m.u., respectively. Heteroduplex mapping of early RNA from the late T.U. by Chow et al. (3) placed the major i leader species (Fig. 1) between m.u. 22 and 23.5, approximately 900 residues from the second leader segment, l_2. From contour lengths measured by Chow et al. (3), this form of the i leader was expected to be approximately 400 nucleotides long.

Figure 2a shows the location of the i leader in greater detail. To map the position of the i

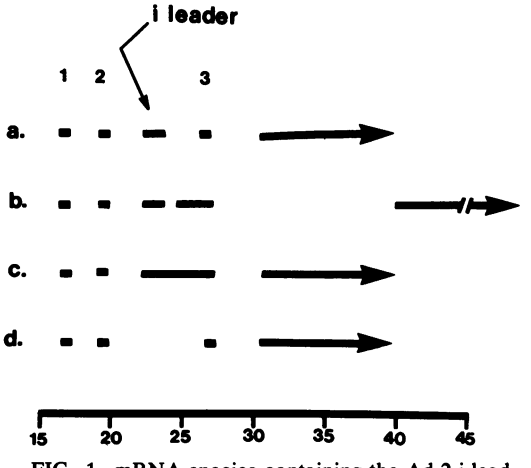

FIG. 1. mRNA species containing the Ad-2 i leader. Species a, b, and c are mRNAs containing the i leader or variants of this leader. Species a is the form analyzed in this report. Species d has the standard tripartite leader with leader segments 1, 2, and 3. This figure is redrawn from Chow et al. (3).

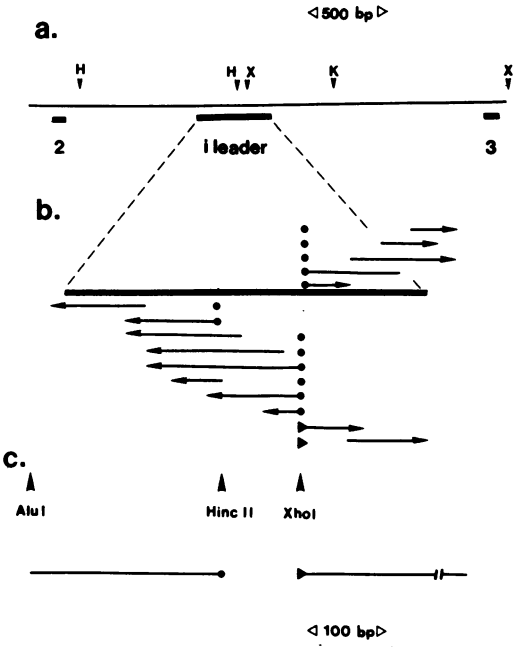

FIG. 2. Structural analysis of the i leader. (a) i leader region of Ad-2 DNA and the location of *Xho*I (X), *Kpn*I (K), and *Hin*cII (H) restriction enzyme cleavage sites. Only those *Hin*cII sites employed in these studies are shown. The positions of the leader 2 and leader 3 genomic sequences with respect to the i leader are given. (b) Regions of i leader DNA analyzed by specific Maxam-Gilbert DNA sequencing gels. The arrows correspond to the sequence analyzed in individual gels. The solid circles represent DNA labeled at the 5′ termini, and the solid triangles are DNA labeled at the 3′ termini, all as described in the text. (c) DNA fragments employed for S1 nuclease analysis of RNA-DNA hybrids as described in the text. One fragment labeled in the 5′ terminus (solid circle) extended from a *Hin*cII site to an *Alu*I site and was used to map the 5′ boundary to the i leader. The other fragment labeled at the 3′ terminus (solid triangle) extended from the *Xho*I site to a *Kpn*I site (a) and was used to map the 3′ boundary of the i leader.

leader, we determined the nucleotide sequence of the i leader region by the method of Maxam and Gilbert (14; Fig. 2b). We next prepared restriction fragments with one labeled endpoint mapping within the predicted sequence of the i leader. For each fragment, this endpoint was labeled either at the 5′ end by the polynucleotide kinase method or, when possible, at the 3′ end by the repair of single-stranded tails using α-^{32}P-deoxytriphosphates and the Klenow fragment of DNA polymerase I. DNA fragments overlapping the 5′ boundary of the i leader were 5′ labeled, and those overlapping the 3′ boundary were 3′ labeled, both in the i leader template strand. These are shown in Fig. 2c. Both types of labeled fragments were reannealed with unlabeled poly(A)$^+$ mRNA isolated from late-stage Ad-2-infected cells. Hybridization was under the R loop conditions described by Berk and Sharp (2). Hybrids were then treated with S1 nuclease to digest single-stranded regions and yield end-labeled DNA stretching from a defined labeled position within the i leader genomic sequence to one or the other splice boundary. The DNA fragments employed for S1 analysis were the very same fragments sequenced by the method of Maxam and Gilbert (14). Fragments obtained from the S1 digestions were electrophoresed on thin polyacrylamide gels. The Maxam-Gilbert sequencing reactions were fractionated in adjacent gel lanes for comparison. Bands from the S1 digest extending from the labeling point to the splice point comigrated with the band in the sequencing ladder corresponding to the splice junction.

An autoradiogram of the gel fractionation of the S1 hybrids is shown in Fig. 3. Figure 3A gives the pattern with a 5′-labeled *Hin*cII-*Alu*I DNA fragment that defines the 5′ boundary, and Fig. 3B shows the trimmed hybrids of a 3′-labeled *Xho*I-*Kpn*I DNA fragment, giving the 3′ boundary. Note that these hybrids have some fraying of the terminus resulting from S1 cleavage, and thus the splice boundary is localized only to a region of approximately five nucleotides. In Fig. 3, specific bands in the sequencing ladders which migrate near the S1 hybrids are marked and correlated with the DNA sequence as described in the legend to Fig. 3.

Figure 4 shows the nucleotide sequence of the i leader region DNA deduced from the gels described in Fig. 2b. Scanning the positions in this sequence corresponding to the bands obtained from the S1 nuclease hybrid experiment of Fig. 3 reveals sequence homologies to consensus sequences at donor and acceptor splice points. Although the S1 hybrid of Fig. 3 defining the 5′ side of the i leader gave a band spread resulting from 3′ fraying of the hybrid, the DNA sequence specified by this spread in Fig. 3A contains the sequence TCGTAG (seen as its complement on the sequencing gel), which corresponds, at four of six specified positions, to the splicing homology YYNCAG (Y = pyrimidine, N = any) reviewed by Sharp (16).

Likewise, on the 3′ boundary, the sequence AGGTCAG (also seen as its complement in the sequencing gel) appears adjacent to the S1 hybrid bands of Fig. 3B, corresponding at six of the seven specified positions to the consensus sequence AGGTRAG (R = purine). Although our S1 digests do not precisely determine the splice boundaries, they are sufficient to unambiguously locate the consensus sequences, which in turn give a precise boundary.

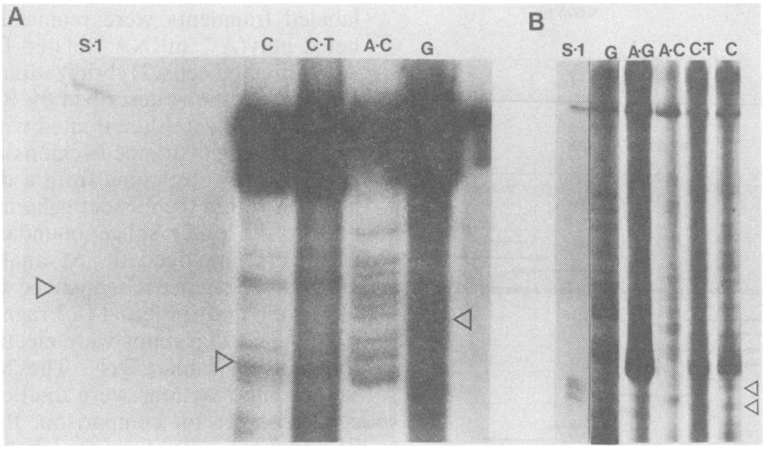

FIG. 3. Mapping the i leader boundaries by S1 nuclease analysis of RNA-DNA heteroduplexes. Poly(A)$^+$ cytoplasmic RNA from late-infected cells was hybridized in (A) to the 5′-labeled *Hin*cII-*Alu*I fragment of Fig. 2c and in (B) to the 3′-labeled *Xho*I-*Kpn*I fragment of Fig. 2c. Hybrids were treated with S1 nuclease and electrophoresed on thin gels as described in the text. Maxam-Gilbert sequence analyses of the same fragments were electrophoresed in adjacent lanes. Standard C, C-T, A-C, A-G, and G chemistries were employed as noted. The boundaries of the i leader were identified as described in the text. First, Maxam-Gilbert bands comigrating with S1 bands in Fig. 5 were noted. Then, the equivalent Maxam-Gilbert bands were located in the primary sequencing gels described in Fig. 2. Specifically, in (A), the triangle in the S1 lane identifies the S1-resistant hybrid mapping the i leader 5′ boundary. The triangles in the C and G lanes mark bands flanking the sequence CTACGATG (bands are read in an upward direction because DNA is 5′ labeled). The complement CATCGTAG appears in the i leader genomic sequence at the 5′ boundary (Fig. 4). In (B), the two open triangles in the C lane mark CC doublets which flank the sequence CCTGACC (read down the gel because DNA is 3′ labeled), whose complement GGTCAGG appears in the i leader genomic sequence at the 3′ boundary (Fig. 4).

Inspection of the DNA sequence of the i leader reveals an ATG triplet 26 nucleotides from the 5′ boundary. The presence of this ATG distinguishes the i leader from the tripartite version of the late leader, which lacks any ATG sequence.

Figure 4 also shows the amino acid sequence of a hypothetical translation product commencing at the ATG. Although nine termination codons are present in the i leader, they are confined to the two reading frames out of phase with the ATG. The frame specified by the ATG is open for translation. In fact, no terminator is encountered in the i leader itself, and the translation product predicted by the DNA must terminate outside the i leader in a different RNA segment.

These results suggest that the i leader constitutes the 5′-terminal segment of an Ad-2 gene.

We next assayed for i leader sequences in viral mRNA present during the late stage of infection. Poly(A)$^+$ cytoplasmic RNA was isolated from Ad-2-infected cells either 18 h postinfection with a cytosine arabinoside block of DNA replication to maintain the early state or 24 h postinfection, electrophoresed on formaldehyde denaturing agarose gels, and then blotted onto nitrocellulose. The i leader was first detected by Chow et al. (3) in mRNA from the first, L-1, 3′-coterminal family of late T.U. mRNAs. Figure 5, lanes A and B, show Northern blot analyses of mRNA from this family at early and late times. The probe was p*Bam*-D, which contains 29- to 41-m.u. DNA and is specific for the L-1 family. At early times (Fig. 5, lane B), a single predominant mRNA band was seen, but at late times (Fig. 5, lane A), two additional shorter mRNAs were also seen. This change in splicing patterns (early versus late) has been observed for several messengers of Ad-2 (3, 21).

We next analyzed a Northern blot of late mRNA with an i leader probe to estimate the number of late mRNA species which contain the i leader. Figure 5, lane C, shows the autoradiogram of late mRNA probed with the i leader-specific DNA. The probe was the *Hin*cII fragment which overlaps the i leader 5′ boundary (Fig. 2a). The positions of migration of late L-1 family mRNAs are noted alongside the blot pattern. At least three size classes of mRNA containing the i leader must be present in at least several different late mRNAs. Also, late mRNA with the i leader can be seen in Fig. 5, lane C, to migrate more rapidly than the early L-1 mRNA of Fig. 5, lane B. Thus, the i leader cannot, at late times, be confined to the L-1 species noted at early times by Chow et al. (3). Finally, the band pattern of late mRNAs containing the i

```
                                                  Initiate Translation:
                                                        met arg ala asp arg glu glu
ctc tac atc gta gGT GAC AAA GAG ACG CTC GGT GCG AGG ATG CGA GCC GAT CGG GAA GAA
        YY NCA GGT G : Consensus Homology at 5' Boundary

leu asp leu pro pro pro val gly gly val ala val asp val val lys val glu val pro
CTG GAT CTC CCG CCA CCA GTT GGA GGA GTG GCT GTT GAT GTG GTG AAA GTA GAA GTC CCT


ala thr gly arg thr leu val leu ala phe val lys thr cys ala val leu ala ala val
GCG ACG GGC CGA ACA CTC GTG CTG GCT TTT GTA AAA ACG TGC GCA GTA CTG GCA GCG GTG


his gly leu tyr ile leu his glu val asp leu thr thr ala his lys glu ala glu trp
CAC GGG CTG TAC ATC CTG CAC GAG GTT GAC CTG ACG ACC GCG CAC AAG GAA GCA GAG TGG


glu phe glu pro leu ala trp arg val trp leu val val phe tyr phe gly cys leu ser
GAA TTT GAG CCC CTC GCC TGG CGG GTT TGG CTG GTG GTC TTC TAC TTC GGC TGC TTG TCC


leu thr val trp leu leu glu gly ser tyr gly gly ser asp his his ala ala arg ala
TTG ACC GTC TGG CTG CTC GAG GGG AGT TAT GGT GGA TCG GAC CAC CAC GCC GCG CGA GCC


gln ser pro asp val arg ala arg arg ser glu leu asp asp asn ile ala gln met gly
CAA AGT CCA GAT GTC CGC GCG CGG CGG TCG GAG CTT GAT GAC AAC ATC GCG CAG ATG GGA


ala val his gly leu glu leu pro arg arg gln val arg arg glu leu leu gln val tyr
GCT GTC CAT GGT CTG GAG CTC CCG CGG CGA CAg gtc agg cgg gag ctc ctg cag gtt tac
                                          AG GTR AG : Consensus Homology
                                                        at 3' Boundary

     Terminator:
leu ala ***
ctc gca tag ccg ggt
```

FIG. 4. DNA nucleotide sequence and coding potential of the i leader region. The nucleotide sequence was deduced from gels indicated in Fig. 2b. The positions of 5′ and 3′ splice boundaries are deduced from the experiments of Fig. 3. The splice consensus homologies at these boundaries are given, with Y = pyrimidine, R = purine, and N = any nucleotide. The predicted amino acid sequence of the translation product of an open reading frame commencing with ATG at nucleotide 27 of the leader is given. A terminator appears just beyond the 3′ boundary of the leader. The coding capacity of this sequence in the mRNAs of Fig. 1 is discussed in the text.

leader appears to be more complex than that of the late L-1 family mRNAs shown in Fig. 5, lane A.

DISCUSSION

The data presented in this report identify the 5′ and 3′ boundaries of the major form of the i leader and describe the genomic DNA nucleotide sequence of the leader and flanking residues. We found that the leader contains an AUG initiation triplet 26 residues within its 5′ boundary, followed by an in-phase reading frame which continues 11 codons beyond the 3′ splice junction. After the completion of these experiments, we became aware of a similar study by Virtanen et al. (18), who have also identified the open reading frame in the i leader.

As discussed above, the major late T.U. of Ad-2 expresses upwards of 15 different proteins by forming mRNAs with the tripartite leader juxtaposed to the mRNA coding sequence. Because the tripartite late leader of Ad-2 mRNA lacks an AUG initiator, translation starts within the main body. Thus, the diversity of the translation products (see reference 21 for a review) relies upon the diversity of splice acceptor positions for the tripartite leader. The presence of an AUG in the i leader suggests that this translation start may be employed by any mRNA bearing the i leader segment. Because this AUG lies to the 5′ side of main bodies, translation from the

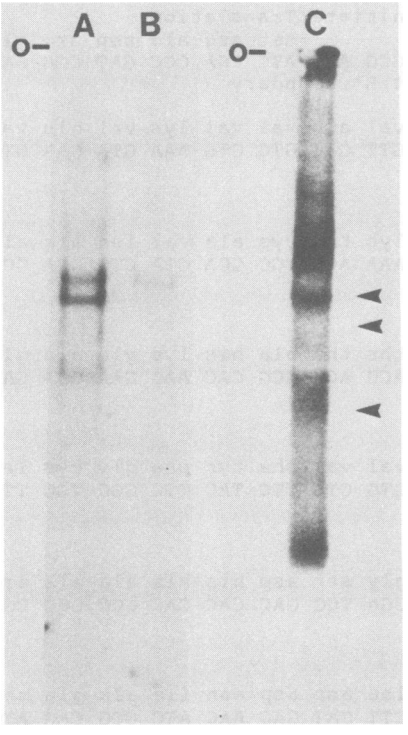

FIG. 5. Northern blot analysis of mRNAs. The L-1 family mRNAs at late times (lane A) and early times (lane B) were analyzed with a p*Bam*-D probe as described in the text. mRNA from late infected cells containing the i leader was analyzed in lane C by probing with the *Hin*cII fragment overlapping the i leader 5′ boundary (Fig. 2a). In lane C, the three arrowheads mark the position of the migration of the L-1 family bands seen in lane A.

main body start would be preempted by the i leader start. Virtanen et al. (18) have identified a 16-kilodalton polypeptide translation product from mRNAs enriched for i leader sequences. These workers have further sequenced cDNA copies of late T.U. mRNA and shown that the i leader reading frame terminates within leader 3, yielding a peptide close to the size of the observed product. Thus, it is likely that the i leader is, in fact, a coding sequence.

It is of interest to consider the coding potential of the i leader in terms of the sequence arrangement of the leaders identified by Chow et al. (3), given Fig. 1. All i leader sequences contain identical 5′ boundaries, but differ in the sequence arrangement on the 3′ side. The major leader structures of the late T.U. mRNA are tripartite and noncoding and correspond to the mRNA of Fig. 1d. The major i leader form identified by Chow et al. (3) and analyzed in this report is species a of Fig. 1. With this form of the i leader, translation terminates within the third leader segment (18). Note that the terminator is joined to the i leader by a splice event.

Chow et al. (3) describe two additional forms of late leader containing the i leader sequences. In one, species c of Fig. 1, the entirety of the RNA sequences between the i leader and the third leader is retained. With this arrangement, the i leader open reading frame translation would continue across the i leader 3′ boundary for 11 codons, where it would encounter the UAG terminator shown in Fig. 4. Thus, the species c translation product would be a peptide 149 amino acids long, which is amino coterminal with the 16-kilodalton polypeptide reported by Virtanen et al. (18), but which differed in the carboxy terminus. With species b of Fig. 1, a short intron follows the i leader, but other sequences on the 5′ side of the third leader are retained. With this splicing pattern, the termination codon employed by species a would be spliced out, but translation would be restored in sequences upstream from leader 3, to yield an additional form of the i leader translation product. Again, this would be amino coterminal with the Fig. 1 species a and c translation products, but would have a third structure at or near the carboxy terminus.

The alternative 3′ boundaries of the i leader reported by Chow et al. (3) are similar to the splicing difference between region Ia mRNAs also reported by these workers. As is possible for the i leader translation products, peptides from the early Ia mRNAs are amino coterminal and differ in internal or carboxy-terminal sequences.

Simian virus 40 late mRNA for the viral coat protein VPI contains a coding sequence upstream from the VPI AUG initiator which encodes an agnogene polypeptide (4, 11, 12). The agnoprotein is evidently translated from the very same mRNA as VPI, making this mRNA polycistronic. The secondary structure of the mRNA is proposed to determine whether the agnogene AUG is, in fact, available to the ribosome (9, 10). Further, the agnogene polypeptide has been proposed to act as a regulator of the transcription termination of simian virus 40 late transcripts, thus controlling the level of VPI mRNA (10). It is not known whether downstream main body sequences can be translated from mRNA containing the i leader. However, attenuation or transcription termination of late promoter RNA has been reported (5), although only a small proportion of this attenuated RNA may reach the i leader genomic sequences.

The different spliced versions of the early Ia mRNAs shift their steady-state proportions during infection (19). It is not known whether the frequency of presence of the i leader or the proportion of the different i leader forms detect-

ed by Chow et al. (3) changes in a similar manner. Splice alterations for the L-1 mRNAs are known, and this change is seen in Fig. 5, lanes A and B. It was reasonable that the i leader is confined to L-1 mRNAs early in infection because these are the only late T.U. species made early (17). It remained possible that at late times the i leader might be either absent or confined to a simple subset of L-1 mRNAs or present on all L-1 forms. Finally, its production might be unrelated to the final main body splicing events, in which case it would be present on all or any late T.U. species. To distinguish these possibilities, we determined the complexity of mRNAs containing the i leader (Fig. 5, lane C). The observed pattern has a complexity greater than the L-1 family of late mRNA, suggesting that the i leader can be present in other late T.U. family mRNAs and possibly all late T.U. mRNAs.

What could be the selective advantage for the virus to position a gene or gene family such that its coding sequences could appear in many mRNAs of the late T.U? Such a gene could be expressed in proportion to the level of synthesis of late T.U. transcripts. As the rate of mRNA synthesis initiated at the late promoter changed, for example, during the early-late shift when a 1,000-fold increase per cell occurs (17), the level of mRNA for the i leader gene could also increase in proportion. This form of control would be advantageous if the i leader gene product participated in the expression of late T.U. messengers. Thus, it will be interesting to determine whether the i leader translation products function in the synthesis or translation of late T.U. mRNAs.

ACKNOWLEDGMENTS

This work was supported by grants from the American Cancer Society and the National Institutes of Health. E.Z. is the recipient of an American Cancer Society Faculty Research Award.

LITERATURE CITED

1. **Akusjarvi, G., and U. Pettersson.** 1979. Sequence analysis of adenovirus DNA. The genomic sequences encoding the common tripartite leader of late adenovirus messenger RNA. J. Mol. Biol. **134**:143–158.
2. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell **12**:721–732.
3. **Chow, L., T. Broker, and J. Lewis.** 1979. Complex splicing patterns of RNAs from the early regions of adenovirus-2. J. Mol. Biol. **134**:265.
4. **Dhar, R., K. N. Subrammanian, J. Pan, and S. M. Weissman.** 1977. Structure of a large segment of the genome of simian virus 40 that does not encode known proteins. Proc. Natl. Acad. Sci. U.S.A. **74**:827–831.
5. **Evans, R., J. Weber, E. Ziff, and J. E. Darnell.** 1979. Premature termination during adenovirus transcription. Nature (London) **278**:367–370.
6. **Fraser, N., and E. Ziff.** 1978. RNA structures near poly(A) of adenovirus-2 late messenger RNAs. J. Mol. Biol. **124**:27–51.
7. **Fraser, N. W., C. C. Baker, M. A. Moore, and E. B. Ziff.** 1982. Poly(A) sites of adenovirus serotype 2 transcription units. J. Mol. Biol. **155**:207–233.
8. **Fraser, N. W., J. R. Nevins, E. Ziff, and J. E. Darnell.** 1979. The major late adenovirus type-2 transcription unit: termination is downstream from the last poly(A) site. J. Mol. Biol. **129**:643–656.
9. **Ghosh, P. K., V. B. Reddy, J. Swinscoe, P. Lebowitz, and S. M. Weissman.** 1978. The 5′ terminal leader sequence of late mRNA from cells infected with simian virus 40. J. Biol. Chem. **253**:3643–3647.
10. **Hay, N., H. Skolnik-David, and Y. Aloni.** 1982. Attenuation in the control of SV40 gene expression. Cell **29**:183–193.
11. **Jackson, V., and R. Chalkley.** 1981. Use of whole-cell fixation to visualize replicating and maturing simian virus 40: identification of new viral gene product. Proc. Natl. Acad. Sci. U.S.A. **78**:6081–6085.
12. **Jay, G., S. Nomura, C. W. Anderson, and G. Khoury.** 1981. Identification of the SV40 agnogene products: a DNA binding protein. Nature (London) **291**:346–349.
13. **Kozak, M.** 1978. How do eukaryotic ribosomes select initiation regions in messenger RNA? Cell **15**:1109–1123.
14. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
15. **Nevins, J. R., and J. E. Darnell.** 1978. Steps in the processing of Ad2 mRNA: poly(A)$^+$ nuclear sequences are conserved and poly(A) addition precedes splicing. Cell **15**:1477–1494.
16. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell **23**:643–646.
17. **Shaw, A. R., and E. B. Ziff.** 1980. Transcripts from the adenovirus-2 major late promoter yield a single early family of 3′ coterminal mRNAs and five late families. Cell **22**:905–916.
18. **Virtanen, A., P. Alestrom, H. Persson, M. G. Katze, and U. Pettersson.** 1982. An adenovirus agnogene. Nucleic Acids Res. **10**:2539–2547.
19. **Wilson, M. C., J. R. Nevins, J.-M. Blanchard, H. S. Ginsberg, and J. E. Darnell, Jr.** 1980. Metabolism of mRNA from the transforming region of adenovirus 2. Cold Spring Harbor Symp. Quant. Biol. **44**:447–455.
20. **Zain, A., 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.
21. **Ziff, E. B.** 1980. Transcription and RNA processing by the DNA tumour viruses. Nature (London) **287**:491–499.
22. **Ziff, E. B., and R. M. Evans.** 1978. Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit. Cell **15**:1463–1476.
23. **Ziff, E. B., and N. Fraser.** 1978. Adenovirus type 2 late mRNA's: structural evidence for 3′-coterminal species. J. Virol. **25**:897–906.