JOURNAL OF VIROLOGY, Dec. 1985, p. 879–886
0022-538X/85/120879-08$02.00/0

Vol. 56, No. 3

Anatomy of Region L1 from Adenovirus Type 2

GÖRAN AKUSJÄRVI

Department of Medical Genetics, The Biomedical Center, Uppsala University, S-751 23 Uppsala, Sweden

Received 28 May 1985/Accepted 19 August 1985

The structure of r-strand-specific RNAs encoded between coordinates 26 and 32 on the adenovirus type 2 genome was mapped by a combination of S1 endonuclease analysis, primer extension, and in vitro transcription. The region includes the third leader segment (coordinates 26.8 to 27.0), the genes for the low-molecular-weight virus-associated RNAs (VA RNAs) (coordinates 29.5 to 30.7), and the amino-terminal end of the gene for the L1 52,000-55,000 polypeptide (coordinates 30.7 to 32.1). The positions at which the tripartite leader was attached to the three longest L1 mRNAs were mapped at the nucleotide level. The leader splice junction of species $L1_a$ was located at coordinate 26.8 and coincided with the 3′ splice site for the third leader segment, whereas the leader-body splice junction of species $L1_b$ and $L1_c$ were located at coordinates 29.0 and 30.7, respectively. No protein products have so far been assigned to the $L1_a$ and $L1_b$ mRNAs, although it can be predicted from the nucleotide sequence that species $L1_b$ encodes a 8,300 polypeptide. The third RNA, species $L1_c$, encodes the well-characterized 52,000-55,000 polypeptide. It was also shown that a previously unidentified class of VA RNAs exists predominantly in the poly(A)$^-$ fraction of late RNA preparations. These RNAs are heterogenous in length (up to 3,000 nucleotides) because of irregular transcription termination and have 5′ ends which map precisely to the initiation sites for VA RNA_I and VA RNA_{II} transcription. Finally it was shown that an RNA with a 5′ end coinciding with the 5′ splice site for the third leader segment exists in the poly(A)$^-$ fraction of late cytoplasmic RNA. This RNA species might represent an excised intron.

The major late adenovirus transcription unit which extends in the rightward (r-strand) direction from coordinate 16.8 to a position near the right terminal end of the genome generates a primary transcript of approximately 28,000 nucleotides (20, 52). More than 20 cytoplasmic mRNAs are generated by differential processing of this nuclear precursor RNA (Fig. 1A). These mRNAs can be grouped into five families, each consisting of mRNAs with coterminal 3′ ends (L1 to L5; Fig. 1A) and having the common tripartite leader attached to their 5′ termini (9, 19).

Within the major late transcription unit there are also genes coding for two low-molecular-weight RNAs, the virus-associated RNAs VA RNA_I and VA RNA_{II} (28, 36, 39). Both RNAs are about 160 nucleotides long and are transcribed by RNA polymerase III (49) from a position close to coordinate 30 on the adenovirus type 2 (Ad2) genome (1, 29). The VA RNAs functions as positive effectors in mRNA translation (43, 43a), increasing the efficiency by which viral mRNAs are translated late after an adenovirus infection (8, 11, 45).

The VA RNA genes are located close to the L1 3′ cotermination family (Fig. 1B). In fact, the T cluster which terminates VA RNA_{II} transcription is separated by only 11 nucleotides from the 3′ splice site for the mRNA encoding the weight polypeptide L1 52,000-55,000 (52,55K) molecular (2). The mRNAs from region L1 are of particular interest since previous studies have shown that the L1 52,55K mRNA is the only mRNA from the major late transcription unit that is expressed at both early and late times of infection (2, 18, 25, 38, 46). Synthesis of the L1 mRNAs, furthermore, is regulated posttranscriptionally by virus-induced changes in the splicing machinery (2, 33). The L1 nuclear precursor RNA is the same both early and late after infection. However, whereas the $L1_c$ mRNA accumulates exclusively at early times, four differentially spliced L1 mRNAs are generated late after infection (Fig. 1A).

Since the VA RNAs are transcribed from the viral r strand, as a consequence they are included in the precursor RNA originating from the major late promoter. To study the structural and functional relationship between the different types of late transcripts, the r-strand-specific RNA molecules that originate from the vicinity of the VA RNA genes were characterized. The structure of two cytoplasmic poly(A)-containing L1 RNAs which both retain the VA RNAs as part of their hypothetical 3′-noncoding regions are described. Moreover, it is shown that extended forms of the VA RNAs exist in both early and late infected cells and that poly(A)$^-$ RNAs which may correspond to excised intron sequences can be found in cytoplasmic RNA preparations.

MATERIALS AND METHODS

Ad2 DNA. Throughout this report all nucleotide numbers and map coordinates refer to positions in the Ad2 sequence published by Roberts et al., (*in* W. Doerfler, ed., *Developments in Molecular Virology*, vol. 5, in press).

Purification of RNA. HeLa S3 cells were infected with Ad2 virus at a multiplicity of 2,000 particles per cell. Cytoplasmic RNA was extracted as described by Brawerman et al. (13) from infected cells maintained in a medium containing 1-β-D-arabinofuranosylcytosine (25 μg/ml) from 1 to 7 h postinfection (p.i.) (early RNA) or from untreated cells 20 h p.i. (late RNA). The RNA was separated into a poly(A)$^+$ and a poly(A)$^-$ fraction by affinity chromatography on oligo(dT) cellulose columns as described previously (7, 35).

In vitro transcription. Plasmid pHindB (40), which specifies the genes for the Ad2 VA RNAs, was transcribed in a soluble HeLa whole-cell extract prepared as described pre-

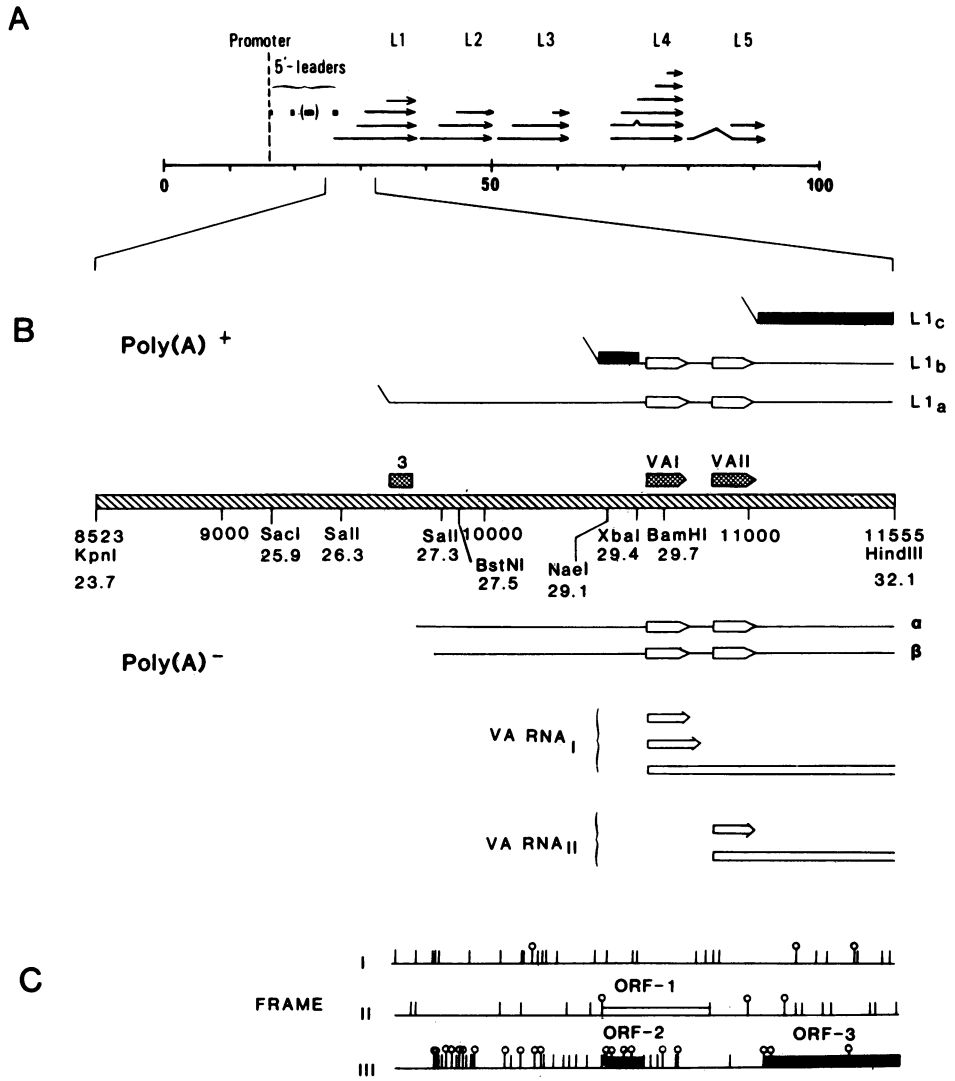

FIG. 1. (A) Organization of the Ad2 genome. Structure of mRNAs originating from the major late transcription unit. (B) Enlargement of the region located between coordinates 23.7 and 32.1 encoding part of the L1 cotermination family. The positions of the third leader segment as well as the VA RNA genes are shown. The RNAs which are specifically found in the poly(A)$^+$ and poly(A)$^-$ fractions are grouped separately in the figure. Black boxes indicate the coding regions for the 8.3K ($L1_b$) and 52,55K ($L1_c$) polypeptides. The position of restriction endonuclease cleavage sites relevant for this study are indicated. (C) The position of ORFs, termination codons (|), and methionines (♀) are indicated for the three translational reading frames. ORF-1 encodes a hypothetical 14.4K polypeptide, which overlaps with the VA RNA_I gene. ORF-2 is the reading frame for the 8.3K polypeptide. ORF-3 denotes the position of the 52,55K polypeptide.

viously (26, 27). A standard 50-μl reaction contained 5 μg of pHindB DNA, 30 μl of cell extract, 12 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [pH 7.9], 60 mM KCl, 7.5 mM $MgCl_2$, 1.2 mM dithiothreitol, and 10.2% glycerol. Incubations were carried out for 1 h at 30°C in the presence of 1 μg of α-amanitin (Sigma Chemical Co., St. Louis, Mo.) per ml. After transcription RNase-free DNase I (Worthington Diagnostics, Freehold, N.J.) was added to a final concentration of 20 μg/ml, and the incubation was continued for an additional 5 min at 30°C. The RNA was subsequently purified by phenol extraction and ethanol precipitated before use in the S1 endonuclease assay.

S1 endonuclease analysis. The protocols described by Berk and Sharp (10) and Weaver and Weissman (48) were essentially followed. Briefly, 0.5 μg of late or 1 μg of early cytoplasmic RNA or the poly(A)$^+$ and poly(A)$^-$ fractions obtained from the equivalent amount of total RNA were hybridized overnight at 57°C to the 5'-end-labeled DNA fragments described in the figure legends. S1 endonuclease cleavage and electrophoretic separation were performed as previously described (44). S1 endonuclease-resistant material was separated in parallel with the nucleotide sequence ladder obtained by chemical degradation of the 5'-end-labeled DNA probe (30).

Primer extension. The *Nae*I-*Xba*I and *Sal*I-*Bst*NI DNA fragments were strand separated by electrophoresis through denaturing polyacrylamide gels as previously described (4, 6) and used as primers for cDNA synthesis. Hybridization mixtures (6 μl of 75 mM Tris hydrochloride [pH 7.5], 0.2 M NaCl) contained 2 μg of total cytoplasmic RNA isolated from late Ad2-infected cells and approximately 2×10^4 cpm

of the DNA primer fragments. The samples were heated for 2 min at 100°C in a sealed capillary and incubated at 65°C for 2 h. After annealing, the hybridization mixtures were emptied into a 50-μl solution consisting of 50 mM Tris hydrochloride (pH 8.3), 10 mM $MgCl_2$, 20 mM dithiothreitol, and 0.7 mM of the four deoxyribonucleotide triphosphates. Fifteen units of avian myeloblastosis virus reverse transcriptase (Life Sciences, Inc., St. Petersburg, Fla.) was added, and the samples were incubated at 37°C for 2 h. To stop cDNA synthesis 50 μl of 0.2 M NaOH was added, and the incubation was continued for an additional 20 min at 65°C. After alkaline hydrolysis, the reaction mixtures were neutralized, extracted with phenol, and precipitated with ethanol. The electrophoretic separation was through 8% polyacrylamide gels containing 8 M urea.

RESULTS

Poly(A)-containing RNA from region L1. The L1 mRNAs have a common poly(A) addition site at coordinate 39.3 (Fig. 1A). Four mRNAs with 3'-splice sites located at coordinates 26.8 ($L1_a$), 29.0 ($L1_b$), 30.7 ($L1_c$), and 34 ($L1_d$) have previously have been identified (2, 18). In the present study the complexity of the L1 RNAs was first analyzed. Early cytoplasmic RNA isolated from Ad2-infected cells treated with 1-β-D-arabinofuranosylcytosine from 1 to 7 h p.i., and late cytoplasmic RNA isolated 20 h p.i. was fractionated by oligo(dT) cellulose chromatography (7) into poly(A)$^+$ and poly(A)$^-$ molecules. The structure of the RNA species present in the two fractions was subsequently studied by S1 endonuclease mapping by using 5'-end-labeled DNA fragments as hybridization probes (10, 48).

Total RNA was hybridized to a 3-kilobase-pair *Kpn*I-*Hin*dIII fragment (coordinates 23.7 to 32.1; Fig. 1B) which was 5' end labeled at the *Hin*dIII cleavage site and digested with endonuclease S1. The resistant material was resolved by electrophoresis through a denaturing polyacrylamide gel. In agreement with previously reported results (2, 18, 33, 38, 46), only one major mRNA species, with a size corresponding to the 52,55K mRNA (species $L1_c$, 520-nucleotide band; Fig. 2), was detected in the poly(A)$^+$ fraction of early RNA. In contrast, in late RNA preparations, three poly(A)$^+$ RNA were detected. In addition to the $L1_c$ mRNA species, $L1_a$ and $L1_b$ (bands of 1,900 and 1,120 nucleotides, respectively; Fig. 2) were also present. Fine mapping studies of RNA species $L1_a$ and $L1_b$ will be discussed in detail below. A protected fragment of approximately 550 nucleotides was also apparent when the late poly(A)$^+$ RNA fraction was analyzed. This band may represent an artifact, since it was not reproducible. However, the alternative that it results from the protection of a spliced mRNA cannot be excluded. A sequence arrangement typical of 3' splice junctions (14, 32) was found at the predicted position.

Species $L1_a$, $L1_b$, and $L1_c$ were also present in the poly(A)$^-$ fraction of late RNA preparations. Repeated oligo(dT) cellulose selections of the poly(A)$^-$ fraction brought the majority of these RNAs into the poly(A)$^+$ fraction (data not shown), suggesting that they in fact are poly(A)$^+$. The RNAs giving rise to the 1,900-nucleotide-long, S1-protected fragment (Fig. 2), on repeated oligo(dT) selections, was more complex since it consisted of a mixture of three RNA species ($L1_a$, α, and β; Fig. 1C; see below).

Position of the splice site between the tripartite leader and the body of mRNA species $L1_b$. To map the leader-body splice junction of mRNA $L1_b$ (Fig. 1B and 2) at the nucleotide level, a combination of S1 nuclease cleavage and primer extension was used. Cytoplasmic RNA selected by two rounds of oligo(dT) cellulose chromatography was hybridized to a 2-kilobase-pair *Kpn*I-*Xba*I DNA fragment (position 8523 to 10579; Fig. 1B) which was 5' end labeled at the *Xba*I cleavage site and treated with S1 endonuclease. Protected DNA fragments were separated on a denaturing polyacrylamide gel in parallel with the corresponding DNA sequence ladder (30). The fragment protected by the $L1_b$ RNA was 150

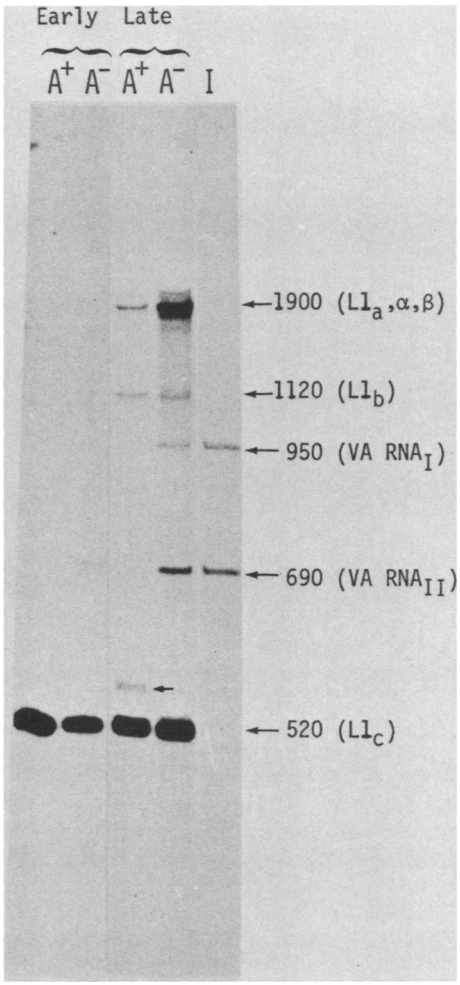

FIG. 2. Complexity of RNAs encoded from the L1 cotermination family. Analysis of the structure of RNA isolated from early and late Ad2-infected cells and RNA synthesized in a soluble whole-cell extract programmed with plasmid pHindB (coordinates 17.3 to 32.1). The in vivo-synthesized RNA was fractionated into poly(A)$^+$ and poly(A)$^-$ by oligo(dT) cellulose chromatography, and equal fractions were analyzed by the S1 nuclease method. RNA was hybridized to a 3-kilobase-pair *Kpn*I-*Hin*dIII restriction fragment (coordinates 23.7 to 32.1), 5' end labeled at the *Hin*dIII site, and treated with S1 nuclease. Resistant material was separated by electrophoresis through a 4% polyacrylamide gel containing 8 M urea. The sizes of the S1 nuclease-protected fragments are indicated in nucleotides, and the RNAs to which they correspond are shown in brackets. The arrow indicates a ghost band produced under certain conditions (see text). Early indicates RNA isolated from cells treated with 1-β-D-arabinofuranosylcytosine from 1 to 7 h p.i. Late indicates RNA isolated 20 h p.i. I is RNA synthesized in vitro.

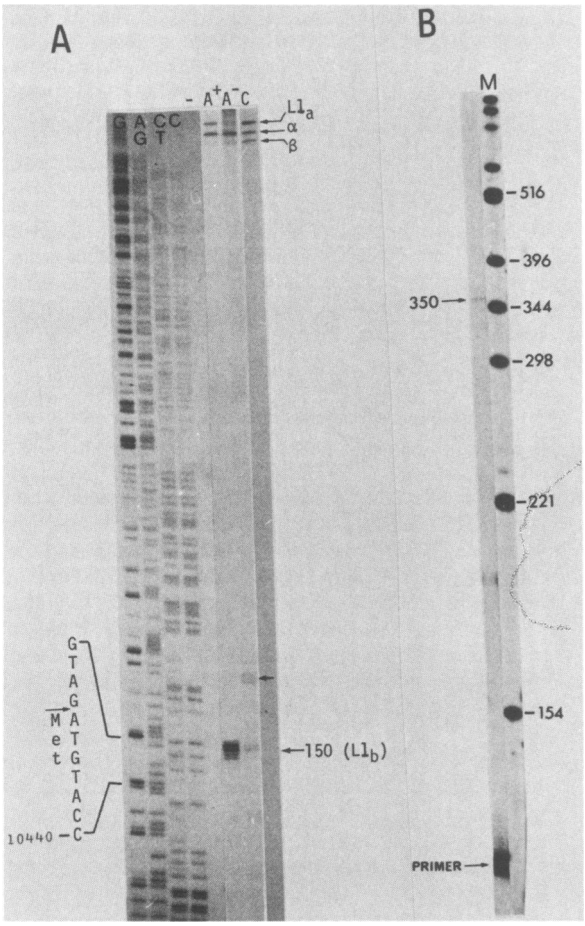

FIG. 3. Spliced structure of RNA species $L1_b$. Cytoplasmic RNA isolated from early and late Ad2-infected cells was hybridized to a *Kpn*I-*Xba*I DNA fragment (coordinates 23.7 to 29.11; Fig. 1B) 5′ end labeled at the *Xba*I cleavage site. S1 nuclease-resistant material was separated in parallel with the corresponding DNA sequence ladder on a 8% denaturing polyacrylamide gel. The positions of the predicted 3′ splice site for RNA species $L1_b$ (long arrow) as well as the hypothetical initiator AUG for the 8.3K polypeptide are indicated. The short arrow indicates the position of a ghost band observed in some experiments. A^+ and A^- denote the poly(A)$^+$ and poly(A)$^-$ fractions, respectively, obtained after oligo(dT) cellulose chromatography of late RNA. C, Total cytoplasmic RNA (10 μg) isolated from early Ad2-infected cells; −, total cytoplasmic RNA isolated from uninfected HeLa cells. (B) Analysis of the spliced structure of RNA species $L1_b$ by primer extension. A 124-nucleotide-long *Nae*-*Xba*I DNA primer, 5′ end labeled at the *Xba*I site, was used to prime cDNA synthesis with total late cytoplasmic RNA used as a template (2 μg). Size fractionation of cDNA products was on a 6% polyacrylamide gel containing 8 M urea. 350 indicates the position of the major extension product; M indicates pBR322 marker fragments.

nucleotides long, which positions the splice junction close to position 10435 on the Ad2 genome (Fig. 3A). At this site the sequence reads CGTAGATGT.

Since S1 endonuclease cleavage does not discriminate between splice sites and ends of RNA molecules, I analyzed the structure of the $L1_b$ RNA also by primer extension. A 124-nucleotide-long, single-stranded DNA primer (*Nae*I-*Xba*I; position 10460 to 10588), 5′ end labeled at the *Xba*I cleavage site, was hybridized to late poly(A)$^+$ RNA and extended with avian myeloblastosis virus reverse transcriptase. Size fractionation of the extension products showed a 350-nucleotide-long cDNA product (Fig. 3B). Since no 150-nucleotide-long component was detected among the cDNAs, it was concluded that the 150-nucleotide-long S1-protected fragment (Fig. 3A) corresponds to a 3′ splice site at which noncontiguous RNA segments were joined.

Since RNA splicing followed the GT-AG rule (14, 32), and only one AG dinucleotide was present in the vicinity of the $L1_b$ splice junction (Fig. 3A), it was concluded that the 3′ splice junction of the $L1_b$ RNA body is located at position 10434 on the Ad2 genome.

The 200-nucleotide length difference between the primer extension product (350-nucleotide band; Fig. 3B) and the S1-protected fragment (150-nucleotide band; Fig. 3A) suggests that the complete tripartite leader (203 nucleotides [5, 51]) is attached to the body of the $L1_b$ RNA. In support of this conclusion it was shown that the 350-nucleotide-long cDNA hybridized to fragment *Hin*dIII C (coordinates 7.7 to 17.1) which encodes the major late promoter and first leader segment (data not shown).

Structure of spliced and unspliced RNAs in the vicinity of the third leader segment. The 5′ end of the $L1_a$ mRNA maps very close to the third leader segment at map coordinate 26.8 (1900-nucleotide band; Fig. 2). By using a DNA probe with a labeled 5′ end located closer to the third leader segment, it was possible to separate these RNAs into three components (Fig. 3A and 4): species $L1_a$, which is enriched in the poly(A)$^+$ fraction, and species α and β, which select inefficiently in the poly(A)$^+$ fraction (Fig. 3A). A repeated passage of the poly(A)$^+$ fraction on a oligo(dT) cellulose column distributed the α and β RNAs between the poly(A)$^+$ and poly(A)$^-$ fractions (data not shown), suggesting that they lack or have a very short poly(A) tail. To determine the precise structure of transcripts $L1_a$, α, and β, I used a combination of S1 nuclease mapping and primer extension. A 354-base-pair *Sac*I-*Bst*NI DNA fragment (positions 9296 to 9895) was used as a DNA probe for the S1 endonuclease cleavage, and a 59-nucleotide-long single-stranded *Sal*I-*Bst*NI fragment (positions 9836 to 9895) was used as a primer for the cDNA synthesis. Both fragments were 5′ end labeled at the *Bst*NI cleavage site. RNA species $L1_a$ generated a S1-protected fragment which mapped precisely to the 3′ splice site for the third leader segment (position 9634; Fig. 4). The result thus suggests that the $L1_a$ RNA represents a partially processed RNA, in which the tripartite leader has been assembled but no further splicing has taken place between the third leader segment and any of the mRNA bodies in the major late transcription unit. This hypothesis was supported by the observation that the primer extension product was approximately 110 nucleotides longer (leader 1 [41 nucleotides] plus leader 2 [72 nucleotides]) than the S1-protected fragment (Fig. 4) and hybridized to restriction endonuclease fragments encoding the first leader segment (data not shown).

Species α and β, in contrast, both generated identical bands in the primer extension and S1 endonuclease cleavage experiments, thus suggesting that they represent colinear RNAs with 5′ ends mapping close to positions 9723 and 9786 on the Ad2 genome. Neither of these RNAs are made in a HeLa whole-cell extract programmed with plasmid pHindB (data not shown), suggesting that they do not correspond to initiation sites of transcription. It is noteworthy that, with the accuracy obtained by the S1 nuclease mapping technique, the 5′ end of the α species maps to the 5′ splice site for the third leader segment (Fig. 1C and 4).

Novel VA RNA transcripts. Two S1-protected DNA fragments of 690 and 950 nucleotides (Fig. 2) are specific for the poly(A)$^-$ fraction of late RNA. The size of the protected fragments positions the 5′ ends for these transcripts close to the initiation sites for VA RNA_I (950-nucleotide band) and VA RNA_{II} (690-nucleotide band) transcription, respectively. To demonstrate that these RNAs, indeed, correspond to long readthrough products initiated at the VA RNA promoters, in vitro transcription in a HeLa whole-cell extract (26, 27) was performed. To inhibit RNA polymerase II activity, 1 μg of α-amanitin per ml was included in the transcription mixture (50). The 690- and 950-nucleotide-long RNA species were also synthesized in a whole-cell extract programmed with plasmid pHindB (coordinates 17.1 to 32.1, 38) which encodes the Ad2 VA RNAs (1) (Fig. 2). The combined data suggest that these novel poly(A)$^-$ RNAs represent RNA polymerase III readthrough transcripts which extend far beyond their normal transcription termination signals (Fig. 1B).

The two species of VA RNA accumulate with different kinetics during a lytic infection; VA RNA_{II} is predominantly made early during the infectious cycle, whereas the synthesis of VA RNA_I continues to increase at late times of infection (11, 39). In agreement with these kinetic studies, it was observed that small amounts of the VA RNA_{II} readthrough transcript also were present in the early RNA preparations (Fig. 2).

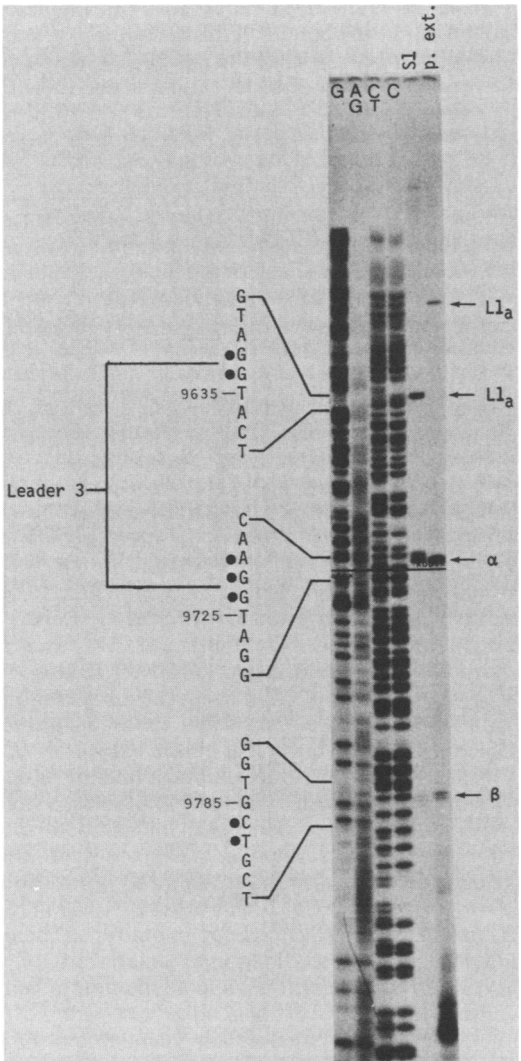

FIG. 4. Structure of spliced and unspliced RNAs near coordinate 27. Total cytoplasmic RNA isolated from late Ad2-infected cells was used for this analysis. S1, 0.5 μg of late RNA was hybridized to a 354-base-pair *Sac*I-*Bst*NI DNA fragment and treated with nuclease S1; p.ext., a 59-nucleotide single-stranded *Sal*I-*Bst*NI fragment was used to prime cDNA synthesis with 2 μg of late RNA. Electrophoresis was on a 6% polyacrylamide gel containing 8 M urea. Relevant parts of the nucleotide sequence are shown expanded in the figure, and the positions of species $L1_a$, α, and β are indicated.

DISCUSSION

By a combination of S1 endonuclease cleavage, primer extension, and in vitro transcription, the structure of several r-strand-specific RNAs encoded by the Ad2 genome between coordinates 26 and 32 were determined. An unexpectedly high complexity of both spliced and unspliced transcripts was found in the cytoplasm of adenovirus-infected cells. The complexity is in part due to the fact that the VA RNAs which are transcribed by RNA polymerase III are encoded within this region. The complexity would be even greater if the early region 2B mRNAs which are transcribed from the opposite strand (41) were taken into account.

Region L1 has previously been shown to encode at least four spliced mRNAs (2, 18). The shortest RNA (Fig. 1A) encodes virion polypeptide III, whereas species $L1_c$ (Fig. 1B) encodes the 52,55K polypeptide (2, 25, 31).

No translational products have been assigned so far to either of species $L1_a$ or $L1_b$. Since the tripartite leader lacks methionine codons (5, 51), all mRNAs originating from the major late adenovirus transcription unit are expected to use the first AUG triplet following the common tripartite leader for the initiation of translation. Based on this assumption, species $L1_b$ would be predicted to code for a polypeptide with a molecular weight of 8.3K (open translational reading frame 2[ORF-2], Fig. 1C). The initiator codon in this reading frame follows directly downstream of the 3′ splice site for this RNA (Fig. 3A). A similar sequence organization has previously been noted in the fiber mRNA (51), the 52,55K mRNA (2), and the pVI mRNA (3). The termination codon of ORF-2 is located only 26 nucleotides upstream of the major G start for VA RNA_I transcription (17, 47). The predicted protein product corresponding to the 8.3K polypeptide has so far not been detected either in vivo or by in vitro translation of hybridization-selected RNA (25, 31). This is probably due to the fact that the $L1_b$ RNA is present in low amounts in infected cells and that the protein, because of its small size, has escaped detection.

Species $L1_a$ represents an RNA in which the tripartite leader has been assembled but the leader-body splice has not yet taken place. Whether it corresponds to an RNA-processing intermediate or an abnormally spliced end product is not known. However, it seems unlikely that the $L1_a$ RNA encodes a functional protein since no ORF of significant length is found in the vicinity of the third leader segment (Fig. 1C). In fact, if the first AUG triplet encountered in the $L1_a$ RNA was used for translation, the product would be a dipeptide, Met-Met. Although initiation of translation can occur at internal AUG sequences (for a review, see refer-

ence 23), it seems unlikely that species $L1_a$ encodes for a protein with a significant length (Fig. 1C).

No candidate mRNA encoding the 14.4K polypeptide, specified by ORF-1 (Fig. 1C), could be detected either in early or in late RNA preparations (Fig. 3). Thus, it seems unlikely that the VA RNA_I sequence is translated into a protein.

The L1 nuclear precursor RNA has the same structure both early and late after infection. However, at early times, one L1 mRNA, species $L1_c$, accumulates almost exclusively, whereas four mRNAs mature from the same nuclear precursor RNA at late times of infection (2, 18) (Fig. 2). The shift in the splicing pattern of the L1 mRNAs has been shown to be caused by a posttranscriptional regulation of splicing rather than resulting from changes in mRNA stability or transport (2, 33). By analyzing large amounts of RNA (Fig. 3A), I showed that small amounts of species $L1_a$, α, and β are also present in the early cytoplasmic RNA preparations. However, also under these conditions, it was not possible to detect species $L1_b$ (Fig. 3A). The presence of low levels of species $L1_a$ but not $L1_b$, in early cytoplasmic RNA supports the notion that species $L1_a$ represents a processing intermediate which has escaped into the cytoplasm or contaminated the cytoplasmic fraction during RNA preparation.

The organization of genes and transcripts in the vicinity of the VA RNA genes is, from a structural point of view, of particular interest. In both RNA species $L1_a$ and $L1_b$, which are cytoplasmic poly(A)$^+$ RNA species, the VA RNAs are present as internal structures (Fig. 1B). The existence of transcripts which normally are synthesized by RNA polymerase III as integral components of eucaryotic mRNAs is, however, not unique to the adenovirus VA RNAs. So-called identifier sequences (42) and Alu or Alu-like sequences, which in many cases have been shown to function as RNA polymerase III transcription units, have also been found in the 3′-noncoding region of cellular mRNAs (for a review, see reference 22). As is the case for the VA RNAs, they are always oriented in the plus direction with respect to the mRNA.

Multiple forms of the two VA RNAs have been found in adenovirus-infected cells. For example, VA RNA_I transcription has been shown to initiate both at a major G start and a minor A start located three nucleotides upstream (17, 47). VA RNA_I transcription terminates, furthermore, within a run of T residues, thus generating a four-nucleotide-long heterogeneity at the 3′ end (16). An alternative form of VA RNA_I, V200 (Fig. 1B), in which the RNA is extended approximately 35 nucleotides at its 3′ end, also has been observed (21, 49). The transcript of VA RNA_{II}, in contrast, appears to be initiated at a unique G residue but most likely is also heterogenous at the 3′ end due to imprecise termination of RNA synthesis (1). The T cluster terminating VA RNA_{II} transcription, in fact, is separated by only 11 nucleotides from the 3′ splice site which is used in maturation of the $L1_c$ mRNA (2). Thus, it is likely that the recognition signals for these two processes overlap each other at this position.

In the present study I showed that a novel class of long readthrough transcripts exists which are initiated at the VA RNA_I and VA RNA_{II} promoters (Fig. 2). Preliminary data indicate that rare VA RNA transcripts with a length exceeding 3,000 nucleotides can be found both in vivo and in RNA synthesized in a HeLa whole-cell extract (data not shown). However, this type of molecule does not seem to have a discrete length, probably because of transcription termination which occurs irregularly at several T clusters encountered in the DNA sequence. A stretch of four T residues, preferably bordered by GC base pairs, is sufficient for efficient termination of RNA polymerase III transcription (12). Similar long RNA polymerase III transcripts, which also appears to be poly(A)$^+$, have been shown to extend from an Alu-like sequence into the structural gene for the human β-globin (15). Since the VA RNA readthrough transcripts are not poly(A)$^+$, I suspect that they do not extend beyond the L1 poly(A) addition site which is located at coordinate 39.3 (24) (Fig. 1A). Whether these long VA RNA transcripts play any functional role during an adenovirus infection is unclear. However, it should be noted that since the VA RNAs accumulate to massive amounts late after infection (10^5 to 10^6 copies per cell), these readthrough transcripts, although they represent only a minor fraction of the VA RNA population (below 1%), accumulate to a substantial amount in the infected cells.

The origin of species α and β (Fig. 1B) is unknown. Both species appear to be unspliced, lacking a poly(A) tail, and have 5′ termini located at positions 27.0 and 27.2, respectively (Fig. 1B). I was not able to demonstrate the existence of RNA polymerase II or III polymers at either of these positions, suggesting that the 5′ ends do not correspond to initiation sites for transcription. However, the possibility that promoters, inactive under my in vitro transcription conditions, exist at these sites cannot be excluded. Since the 5′ end of species α appears to coincide with the 5′ splice site for the third leader segment, the hypothesis that this RNA represents a spliced out intron is attractive. Recent experiments (34, 37) have shown that introns are excised as lariats during pre-mRNA splicing. These lariat structures may be relatively stable since they have a circular structure. It is possible that both species α and β represent intron sequences which have been excised during the maturation of one of the late adenovirus mRNAs (Fig. 1A). Whether their presence in the cytoplasm reflects a specific function for these RNAs or simply is due to leakage from the nucleus is unknown.

ACKNOWLEDGMENTS

I am grateful to Ulf Pettersson and Göran Magnusson for comments on the manuscript, to Marie Lager for technical assistance, and to Elisabeth Sandberg for secretarial help.

This work was supported by grants from the Swedish National Science Research Council and the Swedish Cancer Society.

LITERATURE CITED

1. **Akusjärvi, G., M. B. Mathews, P. Andersson, B. Vennström, and U. Pettersson.** 1980. Structure of genes for virus-associated RNA_I and RNA_{II} of adenovirus type 2. Proc. Natl. Acad. Sci. USA **77**:2424–2428.
2. **Akusjärvi, G., and H. Persson.** 1981. Controls of RNA splicing and termination in the major late adenovirus transcription unit. Nature (London) **292**:420–426.
3. **Akusjärvi, G., and H. Persson.** 1981. Gene and mRNA for precursor polypeptide VI from adenovirus type 2. J. Virol. **38**:469–482.
4. **Akusjärvi, G., and U. Pettersson.** 1978. Nucleotide sequence at the junction between the coding region of the adenovirus 2 hexon messenger RNA and its leader sequence. Proc. Natl. Acad. Sci. USA **75**:5822–5826.
5. **Akusjärvi, G., and U. Pettersson.** 1979. Complete nucleotide sequence of the spliced 5′ noncoding region of adenovirus 2

hexon messenger RNA. Cell **16**:841–850.
6. **Akusjärvi, G., and U. Pettersson.** 1979. The genomic sequences encoding the common tripartite leader of late adenovirus mRNA. J. Mol. Biol. **134**:143–158.
7. **Aviv, H., and P. Leder.** 1972. Purification of active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
8. **Babich, A., L. T. Feldman, J. R. Nevins, J. E. Darnell, and C. Weinberg.** 1983. Effect of adenovirus on metabolism of specific host mRNAs: transport control and specific translational discrimination. Mol. Cell. Biol. **3**:1212–1221.
9. **Berget, S. M., C. Moore, and P. A. Sharp.** 1977. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc. Natl. Acad. Sci. USA **74**:3171–3175.
10. **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.
11. **Bhat, R. A., and B. Thimmappaya.** 1984. Adenovirus mutants with DNA sequence perturbations in the intragenic promoter of VA RNA_I gene allow the enhanced transcription of VA RNA_{II} gene in HeLa cells. Nucleic Acids Res. **12**:7377–7388.
12. **Bogenhagen, D. F., and D. D. Brown.** 1981. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell **24**:261–270.
13. **Brawerman, G. J., J. Mendecki, and S. Y. Lee.** 1972. A procedure for the isolation of mammalian ribonucleic acid. Biochemistry **2**:637–641.
14. **Breatnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
15. **Carlsson, D. P., and J. Ross.** 1983. Human β-globin promoter and coding sequences transcribed by RNA polymerase III. Cell **34**:857–864.
16. **Celma, M. L., J. Pan, and S. M. Weissman.** 1977. Studies of low molecular weight RNA from cells infected with adenovirus 2: the sequences at the 3′ end of VA RNA 1. J. Biol. Chem. **252**:9032–9042.
17. **Celma, M. L., J. Pan, and S. M. Weissman.** 1977. Studies of low molecular weight RNA from cells infected with adenovirus 2: heterogeneity at the 5′ end of VA RNA 1. J. Biol. Chem. **252**:9043–9046.
18. **Chow, L. T., T. R. Broker, and J. B. Lewis.** 1979. Complex splicing patterns of RNAs from the early regions of adenovirus-2. J. Mol. Biol. **134**:265–303.
19. **Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts.** 1977. An amazing sequence arrangement at the 5′-ends of adenovirus 2 messenger RNAs. Cell **12**:1–8.
20. **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.
21. **Harris, B., and R. G. Roeder.** 1978. Structural relationships of low molecular weight viral RNAs synthesized by RNA polymerase III in nuclei from adenovirus-2 infected cells. J. Biol. Chem. **253**:4120–4127.
22. **Jelinek, W. R., and C. W. Schmid.** 1982. Repetitive sequences in eucaryotic DNA and their expression. Ann. Rev. Biochem. **51**:813–844.
23. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
24. **LeMoullec, J. M., G. Akusjärvi, P. Stålhandske, U. Pettersson, B. Chambraud, P. Gilardi, M. Nasri, and M. Perricaudet.** 1983. Polyadenylic acid addition sites in the adenovirus type 2 major late transcription unit. J. Virol. **48**:127–134.
25. **Lewis, J. B., and M. B. Mathews.** 1980. Control of adenovirus early gene expression: a class of immediate early products. Cell **21**:303–313.
26. **Manley, J. L.** 1984. Transcription of eucaryotic genes in a whole-cell extract. *In* B. D. Hames and S. J. Higgins (ed.), Transcription and translation; a practical approach. IRL Press, Oxford.
27. **Manley, J. L., A. Fire, A. Cano, P. A. Sharp, and M. L. Gefter.** 1981. DNA dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Natl. Acad. Sci. USA **77**:3855–3859.
28. **Mathews, M. B.** 1975. Genes for VA-RNA in adenovirus 2. Cell **6**:223–229.
29. **Mathews, M. B., and U. Pettersson.** 1978. The low molecular weight RNAs of adenovirus 2 infected cells. J. Mol. Biol. **119**:293–328.
30. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
31. **Miller, J. S., R. Ricciardi, B. E. Roberts, B. 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.
32. **Mount, S. M.** 1982. A catalouge of splice junction sequences. Nucleic Acids Res. **10**: 459–472.
33. **Nevins, J. R., and M. C. Wilson.** 1981. Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. Nature (London) **290**:113–118.
34. **Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp.** 1984. Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. Science **225**:898–903.
35. **Pettersson, U., C. Tibbetts, and L. Philipson.** 1976. Hybridization maps of early and late mRNA sequences on the adenovirus type 2 genome. J. Mol. Biol. **101**:479–502.
36. **Reich, P. R., B. G. Forget, S. H. Weissman, and J. A. Rose.** 1966. RNA of low molecular weight in KB cells infected with adenovirus type 2. J. Mol. Biol. **17**:428–439.
37. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell **38**:317–331.
38. **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.
39. **Söderlund, H., U. Pettersson, B. Vennström, L. Philipson, and M. B. Mathews.** 1976. A new species of virus-coded low molecular weight RNA from cells infected with adenovirus type 2. Cell **7**:585–593.
40. **Stenlund, A., M. Perricaudet, P. Tiollais, and U. Pettersson.** 1980. Construction of restriction enzyme libraries containing DNA from human adenovirus types 2 and 5. Gene **10**:47–52.
41. **Stillman, B. W., J. B. Lewis, L. T. Chow, M. B. Mathews, and J. E. Smart.** 1981. Identification of the gene and mRNA for the adenovirus terminal protein precursor. Cell **23**:497–508.
42. **Sutcliffe, J. G., R. J. Milner, F. E. Blom, and R. A. Lerner.** 1982. Common 82 nucleotide sequence unique to brain RNA. Proc. Natl. Acad. Sci. USA **79**:4942–4946.
43. **Svensson, C., and G. Akusjärvi.** 1984. Adenovirus VA RNA_I: a positive regulator of mRNA translation. Mol. Cell. Biol. **4**:736–742.
43a. **Svensson, C., and G. Akusjärvi.** 1985. Adenovirus VA RNA_I mediates a translational stimulation which is not restricted to the viral mRNAs. EMBO J. **4**:957–964.
44. **Svensson, C., U. Pettersson, and G. Akusjärvi.** 1983. Splicing of adenovirus 2 early region 1A mRNAs is non-sequential. J. Mol. Biol. **165**:475–499.
45. **Thimmapaya, B., C. Weinberg, R. J. Schneider, and T. Schenk.** 1982. Adenovirus VAI RNA is required for efficient translation of viral mRNAs at late times after infection. Cell **31**:543–551.
46. **Thomas, G. P., and M. B. Mathews.** 1980. DNA replication and the early to late transition in adenovirus infection. Cell **22**:523–533.
47. **Vennström, B., U. Pettersson, and L. Philipson.** 1978. Two initiation sites for adenovirus 5,5S RNA. Nucleic Acids Res. **5**:195–204.
48. **Weaver, R. F., and C. Weissman.** 1980. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15S β-globin mRNA precursor and mature 10S β-globin mRNA mRNA have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
49. **Weinmann, R., H. J. Raskas, and R. G. Roeder.** 1976. Low

molecular weight viral RNAs transcribed by RNA polymerase III during Ad2-infection. Cell **7**:557–566.

50. **Weinmann, R., and R. G. Roeder.** 1974. Role of DNA-dependent RNA polymerase III in the transcription of the tRNA and 5S RNA genes. Proc. Natl. Acad. Sci. USA **71**:1790–1794.
51. **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.
52. **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**:1463–1475.