Volume 9 Number 1 1981 Nucleic Acids Research

The sequence of the 3' non-coding region of the hexon mRNA discloses a novel adenovirus gene

Göran Akusjärvi*, Jan Zabielski, Michel Perricaudet[1] and Ulf Pettersson

Department of Microbiology, The University of Uppsala, The Biomedical Center, Box 581, S-751 23 Uppsala, Sweden, and [1]Institut Pasteur, 28, rue du Dr. Roux, 75724, Paris, France

Received 20 November 1980

SUMMARY

We report the sequence of a 1164 nucleotide long DNA segment, located between map positions 59.5 and 62.8 on the adenovirus type 2 genome. The sequence comprises the 701 nucleotides long 3' non-coding region of the hexon mRNA as well as several important processing signals. The sequence revealed unexpectedly that the 3' non-coding region of the hexon mRNA contains a 609 nucleotide long uninterrupted translational reading frame following a potential initiator AUG. A late 14S mRNA, corresponding to the open reading frame, could be identified by S1 nuclease mapping and electronmicroscopy. The mRNA shares a poly(A) addition site with the hexon and pVI mRNAs, and carries a leader sequence which is related, and probably identical, to the tripartite leader, found in late adenovirus mRNAs. The junction between the leader and the body of this novel mRNA is located within the coding part of the hexon gene.

INTRODUCTION

The double-stranded adenovirus type 2 (ad2) genome has a molecular weight of 23×10^6 containing approximately 35,000 base pairs (1,2). Our knowledge about the structure and organization of the adenovirus DNA is expanding rapidly. Clusters of early and late genes have been mapped with respect to certain restriction enzyme cleavage sites and the two complementary strands (3,4,5). Electronmicroscopy and *in vitro* translation of mRNAs selected by hybridization have, even more accurately, identified the location of many early and late adenovirus genes (6,7,8). Recently, nucleic acid sequencing methods have been applied to the adenovirus genome and more than 25% of the nucleotide sequence of DNA from subgroup C adenovirus has already been reported (for a review see ref.9). So far, the genes for more than 20 different polypeptides have been positioned on the adenovirus genome by electronmicroscopy, S1 nuclease mapping and molecular cloning (7,10,11,12). Undoubtedly, many more genes will be revealed when more sophisticated methods are used to examine the adenovirus genome.

We and others have recently applied DNA sequencing methods to locate genes and important processing signals on the adenovirus genome. Much of our attention has been focused on the hexon gene and its surroundings and in recent reports (13,14) we have identified the amino- as well as the carboxy-terminal ends of the hexon gene. In the present investigation we have extended this analysis to include the entire 3' non-coding region of the hexon mRNA. Our analysis reveals a hitherto unrecognized adenovirus gene which overlaps the 3' non-coding region of the hexon and pVI mRNAs. The corresponding mRNA could also be identified by electronmicroscopy and S1 nuclease mapping.

MATERIALS AND METHODS

Cells and Virus: Adenovirus type 2 was propagated in suspension cultures of HeLa S2 cells as described before (15) and viral DNA extracted from purified virus as described by Pettersson and Sambrook (16).

Restriction endonucleases and cleavage of DNA: Restriction enzymes were purchased from Biolabs Inc. or Bethesda Research Laboratories and used according to specifications from the manufacturers.

Preparation of RNA: HeLa cells were infected at a multiplicity of 10,000 particles per cell. The cells were harvested 20 hours post infection and the cytoplasmic fraction was prepared with 0.65% Nonidet P40. The cytoplasmic RNA was extracted according to the method of Brawermann et al. (17). The extracted RNA was collected by ethanol precipitation and fractionated by oligo(dT) cellulose affinity chromatography as described before (5). The poly(A) containing fraction was collected and further purified by sucrose gradient centrifugation. Gradients containing 15-30% sucrose, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.1 M NaCl and 0.5% Sarcosyl were used. The gradients were centrifuged at 25,000 rpm and 4°C for 20 hrs in the SW27 rotor.

DNA sequence analysis: The method of Maxam and Gilbert (18) was used. For most experiments fragment BglII-J was prepared and cleaved with a suitable restriction endonuclease. The 5'-termini were labeled with γ-ATP and polynucleotide kinase before making a second cleavage with a restriction enzyme to generate fragments which were uniquely labeled at one end. Sequencing gels containing 8 or 20% polyacrylamide were prepared as described by Sanger and Coulson (19). Our sequencing strategy is outlined in Fig. 1.

Endonuclease S1 analysis: A modification of the method originally described by Berk and Sharp (10) was used. 2 µg of ad2 DNA was mixed with RNA and varying amounts of tRNA so that the final amount of RNA in the sample was 20

µg. The mixture was precipitated with ethanol, and dissolved in 10 µl of a buffer containing 80% formamide, 50 mM Pipes pH 6.4, 0.4 M NaCl and 1 mM EDTA. The subsequent denaturations, hybridizations and S1 treatments were as described by Akusjärvi and Persson (20). The S1 resistant hybrids were collected and dissolved in 20 µl of 50 mM NaOH with 1 mM EDTA and 2% Ficoll and separated electrophoretically on a 2% alkaline agarose gel (21). The DNA was transferred to a nitrocellulose sheet and hybridized with a DNA probe which was ^{32}P-labeled by nick translation (22). For many hybridization experiments a cDNA clone, designated clone 2 (Perricaudet et al., unpublished) was used. The cDNA insert in this clone was approximately 800 nucleotides long and maps between positions 60 and 62 on the ad2 genome. The cDNA in the clone does, not however include the poly(A) tail of the hexon and pVI mRNAs.

Hybridization selection of RNA: Cytoplasmic RNA, fractionated by sucrose gradient centrifugation, was divided into five pools. Pool "4" (see fig. 5A) was further purified by hybridization to nitrocellulose filters on which 20 µg equivalents of restriction fragments of ad2 DNA had been immobilized. The hybridized RNA was thermally eluted as described by McGrogan et al. (23) and used for electron microscopy and S1 analysis.

Electronmicroscopy

The EcoRI-B fragment of ad2 (map coordinates 58.5-70.7) was melted and re-annealed to check for homogeneity. No deletion or insertion loops were seen. The following hybridization conditions were used:
2 µg/ml EcoRI-B DNA, 10 µg/ml mRNA, 80% formamide (twice re-crystallised), 0.1M Hepes, pH 7.6, 0.4 M NaCl and 10 mM EDTA. This hybridization cocktail had a volume of 35 µl. The entire mixture was heated to 65°C for 5 minutes to separate the strands, followed by cooling on ice. Hybridization was for 12 hrs at 56°C. Small aliquots were spread either immediately or the reaction was arrested by the addition of 4 volumes of ice cold water and the sample kept at 4°C. Isodenaturing conditions were used, whereby the hyperphase contained 45% formamide,in 0.1M Tris pH 8.5, 0.01 M EDTA, 0.2 µg DNA/ml and 50 µg/ml cytochrome C, upon a hypophase of 15% formamide, 0.01 M Tris pH 8.5 and 0.001 M EDTA. PM2 was used as a double stranded DNA marker (9950 base pairs) and ØX174 SS as the single stranded DNA marker (5383 bases).

RESULTS

The sequence of the 3' non-coding region of the adenovirus type 2 hexon mRNA

We have previously determined a 210 base pairs long sequence of ad2 DNA located between map positions 59.5 and 60.1 and shown that the coding part of the hexon gene ends 135 base pairs to the right of the cleavage site for endonuclease BamHI at position 59.5 (13). In order to extend our previously determined sequence we made use of fragment BglII-J which is located between positions 60.1 and 63.6. The fragment was purified by agarose gel electrophoressis and recleaved with either one of the endonucleases BalI, EcoRII, Hinf, Hph, KpnI MboI or SacII. The generated fragments were endlabeled with polynucleotide kinase and sequenced according to the protocol of Maxam and Gilbert (18) after recleavage with the appropriate enzymes. Our sequencing strategy is illustrated in Fig. 1. The sequence which is shown in Fig. 2 starts at the BamHI cleavage site (map position 59.5), ends at the endonuclease BalI cleavage site at map position 62.8 and comprises 1164 base pairs. More than 90% of the sequence has been determined by analysis of both the complementary strands.

The first 135 nucleotides of the established sequence represent an uninterrupted translational reading frame which ends with the codons for the amino acids -Ala-Thr-Thr. This sequence corresponds to the carboxy-terminal end of the hexon polypeptide as we have reported earlier (13,24). The remaining part of the sequence would then be expected to contain the 3' non-coding region of the hexon mRNA.

The hexanucleotide sequence AAUAAA, which is found near the poly(A) junction of most eukaryotic mRNAs (25) is found between nucleotides 812 and 817 (Fig. 2). The poly(A) addition site for the hexon mRNA would thus be expected to occur some 20 nucleotides downstream from that hexanucleotide

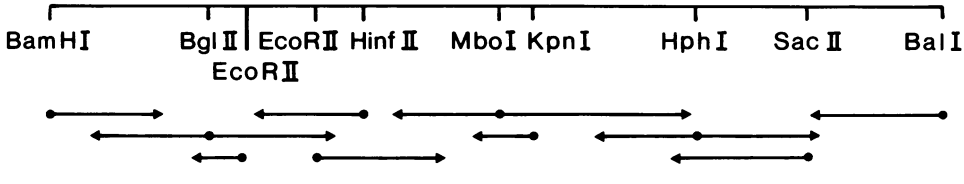


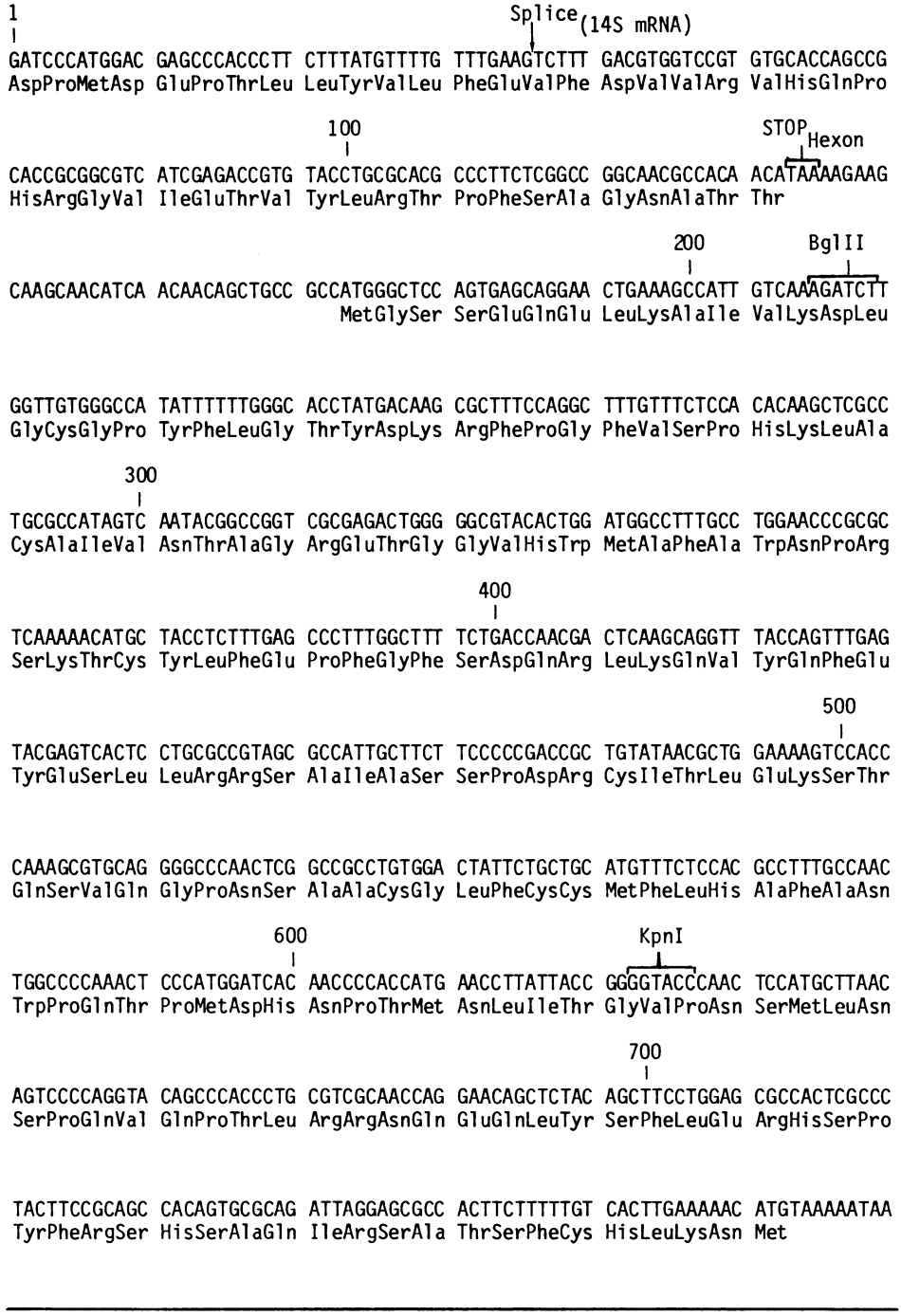
Figure 1. The sequencing strategy. The figure indicates the positions of the restriction enzyme cleavage sites which were used to establish the reported sequence. The arrows indicate the amounts of sequence information collected from each site.

sequence. In a separate study (Perricaudet et al., in preparation) it is shown by sequence analysis of a cDNA copy of the hexon mRNA that the poly(A) tract follows nucleotide 836 in our established sequence (Fig. 2).

It was thus surprising to find that an additional, 609 nucleotide long, open reading frame is apparent in the interspace between the end of the hexon coding sequence and the poly(A) addition site (Fig. 2 and 3C).

The poly(A) addition site for the 72K mRNA

The mRNA for the 72K DNA binding protein is transcribed from the viral l-strand (4,5,7,26) and its poly(A) addition site has also been mapped around position 62.0 (Fig. 3B). It was therefore interesting to note that our established sequence read TTTATT between position 828 and 833 (Fig. 2). When transcribed from the l-strand the sequence will read AAUAAA and could thus be the poly(A) addition signal for the 72K mRNA. The following approach was chosen in order to map the poly(A) addition site for the 72K mRNA: a clone carrying fragment BalI-J of ad2 DNA which is located between map coordinates 58.7 and 62.8 was labeled in vitro by nick translation. The labeled clone was denatured and hybridized to early ad2 mRNA under conditions that prevented the reassociation of the two DNA strands (27). The mixture was subsequently treated with the single-strand specific nuclease S1 and the nuclease resistant hybrids were analyzed by electrophoresis on an 8% sequencing gel using HindII and HaeIII fragments of ØX174 RFII DNA as markers. The results revealed a 350 nucleotide long band (Fig. 4, slot A) which should have originated from hybrids between the 72K mRNA and the l-strand of the BalI-J fragment since the 72K mRNA is the only mRNA which is transcribed from this region early after infection (fig. 3B). The results thus suggest that the poly(A) addition site for this mRNA is located 350 nucleotides to the left of the cleavage site for endonuclease BalI at position 62.8. To locate the poly(A) site more precisely the BalI-J fragment was cleaved with endonuclease SacII (cleavage site at map position 62.4) prior to hybridization with early mRNA. The hybrids were then treated with nuclease S1 and analyzed by electrophoresis on an 8% sequencing gel with the same markers as in the previous experiment. In this case two fragments, 172 and 175 nucleotides long, were observed (Fig. 4, slot B). Since we know that the distance between the SacII and the BalI site is 175 nucleotides (Fig. 2) the results show that the poly(A) addition site is located 172 nucleotides from the cleavage site for SacII at position 62.4. Since poly(A) addition usually occurs where one or more A residues are present in the sequence (28,29) we can predict that the poly(A) addition site for the 72K mRNA is located


1
Splice
(14S mRNA)
GATCCCATGGAC GAGCCCACCCTT CTTTATGTTTTG TTTGAAGTCTTT GACGTGGTCCGT GTGCACCAGCCG
AspProMetAsp GluProThrLeu LeuTyrValLeu PheGluValPhe AspValValArg ValHisGlnPro
100
STOP
Hexon
CACCGCGGCGTC ATCGAGACCGTG TACCTGCGCACG CCCTTCTCGGCC GGCAACGCCACA ACATAAAAGAAG
HisArgGlyVal IleGluThrVal TyrLeuArgThr ProPheSerAla GlyAsnAlaThr Thr
200
BglII
CAAGCAACATCA ACAACAGCTGCC GCCATGGGCTCC AGTGAGCAGGAA CTGAAAGCCATT GTCAAAGATCTT
MetGlySer SerGluGlnGlu LeuLysAlaIle ValLysAspLeu
GGTTGTGGGCCA TATTTTTTGGGC ACCTATGACAAG CGCTTTCCAGGC TTTGTTTCTCCA CACAAGCTCGCC
GlyCysGlyPro TyrPheLeuGly ThrTyrAspLys ArgPheProGly PheValSerPro HisLysLeuAla
300
TGCGCCATAGTC AATACGGCCGGT CGCGAGACTGGG GGCGTACACTGG ATGGCCTTTGCC TGGAACCCGCGC
CysAlaIleVal AsnThrAlaGly ArgGluThrGly GlyValHisTrp MetAlaPheAla TrpAsnProArg
400
TCAAAAACATGC TACCTCTTTGAG CCCTTTGGCTTT TCTGACCAACGA CTCAAGCAGGTT TACCAGTTTGAG
SerLysThrCys TyrLeuPheGlu ProPheGlyPhe SerAspGlnArg LeuLysGlnVal TyrGlnPheGlu
500
TACGAGTCACTC CTGCGCCGTAGC GCCATTGCTTCT TCCCCCGACCGC TGTATAACGCTG GAAAAGTCCACC
TyrGluSerLeu LeuArgArgSer AlaIleAlaSer SerProAspArg CysIleThrLeu GluLysSerThr
CAAAGCGTGCAG GGGCCCAACTCG GCCGCCTGTGGA CTATTCTGCTGC ATGTTTCTCCAC GCCTTTGCCAAC
GlnSerValGln GlyProAsnSer AlaAlaCysGly LeuPheCysCys MetPheLeuHis AlaPheAlaAsn
600
KpnI
TGGCCCCAAACT CCCATGGATCAC AACCCCACCATG AACCTTATTACC GGGGTACCCAAC TCCATGCTTAAC
TrpProGlnThr ProMetAspHis AsnProThrMet AsnLeuIleThr GlyValProAsn SerMetLeuAsn
700
AGTCCCCAGGTA CAGCCCACCCTG CGTCGCAACCAG GAACAGCTCTAC AGCTTCCTGGAG CGCCACTCGCCC
SerProGlnVal GlnProThrLeu ArgArgAsnGln GluGlnLeuTyr SerPheLeuGlu ArgHisSerPro
TACTTCCGCAGC CACAGTGCGCAG ATTAGGAGCGCC ACTTCTTTTTGT CACTTGAAAAAC ATGTAAAAATAA
TyrPheArgSer HisSerAlaGln IleArgSerAla ThrSerPheCys HisLeuLysAsn Met

cont. fig. 2

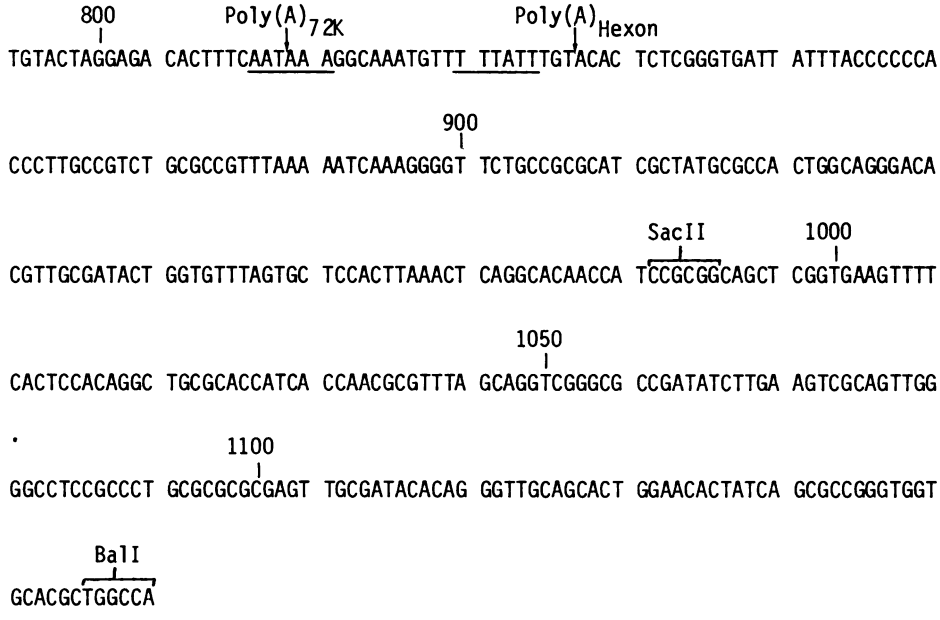


Figure 2. The established sequence starts where endonuclease BamHI cuts the l-strand at position 59.5 and ends with the recognition sequence for endonuclease BalI at position 62.8. Selected restriction enzyme cleavage sites are indicated as well as the poly(A) addition site for the hexon and 72K mRNA. The position where the leader is attached to the body of the 14S mRNA is also shown.

either at position 810 or 814 in our established sequence. Although position 814 corresponds with our size estimates our data does not permit an entirely unambiguous choice between these two alternatives. However, in a separate study (Stålhandske et al., in preparation) we show by sequence analysis of a cDNA-containing plasmid that the poly(A) junction is located at nucleotide 814. From the results it is clear that the 3'-ends for the hexon and the 72K mRNAs overlap for a short distance,

Detection of a novel mRNA by S1 nuclease mapping

The presence of a long uninterrupted translational reading frame in the 3' non-coding region of the hexon mRNA prompted a search for a novel mRNA species which is expressed by this region of the genome. For this purpose we chose to use the S1 nuclease mapping procedure of Berk and Sharp (10); cyto-

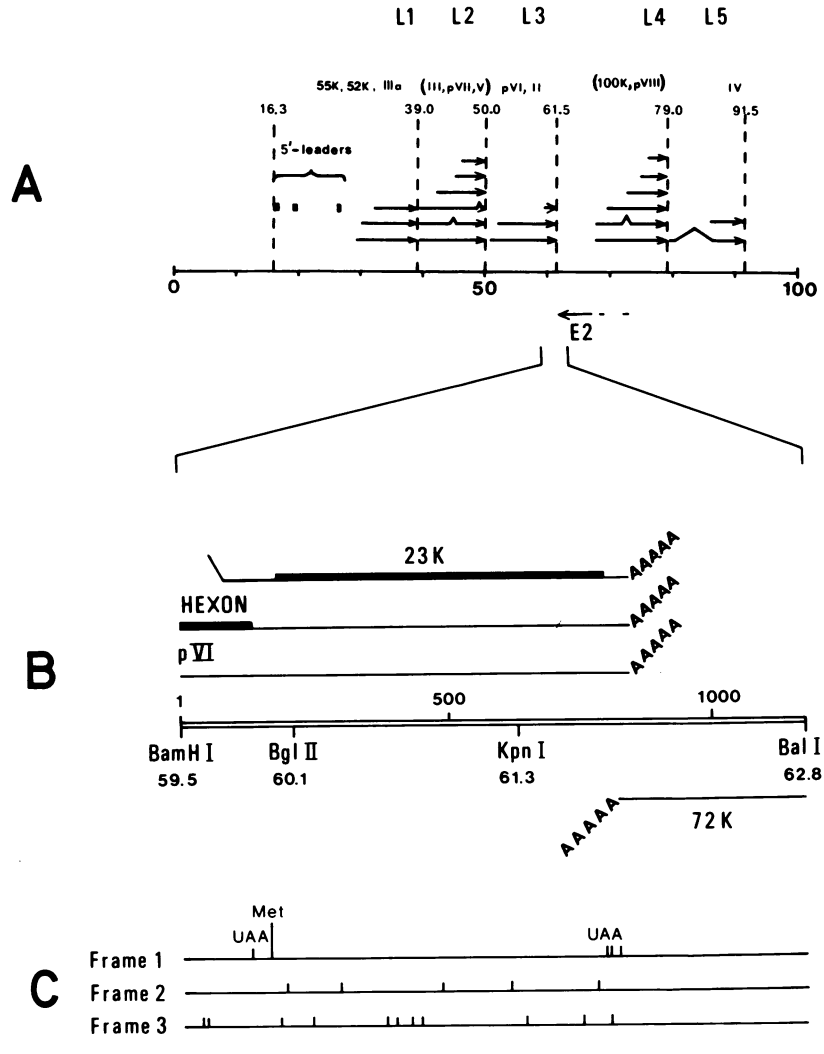


Figure 3. (A) A schematic drawing which indicates the different co-termination families of late mRNAs which are transcribed from the major adenovirus promoter.
B) A schematic drawing indicating the 3'-ends of different mRNAs which terminate around map position 62. The 23K mRNA indicates the novel 14S mRNA which is described in the present study. Coding areas (——) as well as non-coding regions (__) are indicated for each mRNA. The positions of selected restriction enzyme cleavage sites are also indicated.
C) The positions of termination codons in the three reading frames of the established sequence. Termination codons are indicated by vertical bars. The presumptive initiator AUG for the product of the 14S mRNA is indicated as well as the two UAA triplets which terminate the hexon polypeptide and the product of the 14S mRNA.

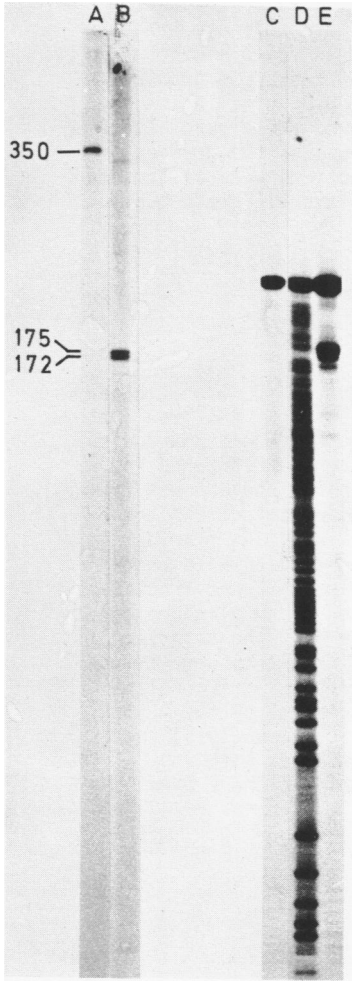


Figure 4. Location of the poly(A) addition site for the 72K mRNA and the splice site for the 14S mRNA. A) A plasmid clone carrying the BalI-J fragments of ad2 was labeled in vitro by nick translation and hybridized with pool "4" RNA. The resulting hybrids were treated with S1 nuclease and analyzed on an 8% sequencing gel. HaeIII and HindII fragments of ØX 174 RFII DNA were used as size markers. B) The same experiment with ^{32}P-labeled fragment BalI-J which was cleaved with endonuclease SacII prior to hybridization. C,D,E) Location of the splice junction between the body and the leader sequence of the 14S mRNA. Fragment BglI-D was endlabeled with (γ)^{32}P-ATP and polynucleotide kinase before cleavage with endonuclease BamHI. In panel E the 210 nucleotides long BamHI-BglII fragment was isolated and hybiridized to pool "4" RNA (see Fig. 5A). The resulting hybrids were treated with endonuclease S1 and separated on an 8% sequencing gel. Panel C shows fragment BamHI-BglII annealed in he absence of RNA and panel D shows a Maxam and Gilbert (18) sequencing ladder from the same end-labeled fragment. Only a purine-specific reaction is shown.

plasmic RNA was extracted from HeLa cells late after infection and fractionated by sucrose gradient centrifugation. The gradient fractions were collected in five pools, 1, 2,3,4 and 5 as illustrated in Fig. 5A. Pool "1" contained 28-35S RNA, pool "2" 23-27S RNA, pool "3" 18-22S RNA, pool "4" 12-17S RNA and pool "5" 5-11S RNA. RNA from each of the five pools was hybridized to denatured ad2 DNA under conditions favouring RNA/DNA hybrid formation (see Materials and Methods section) and the resulting hybrids were treated with endonuclease S1. The digestion products were separated on an alkaline agarose gel and transferred to nitrocellulose by the method of

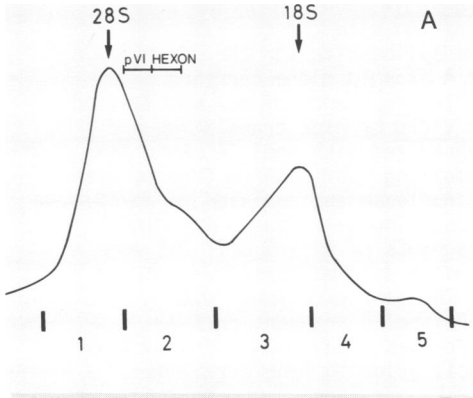


Fig. 5.A) Fractionation of late poly(A)containing ad2 RNA on a linear 15-30% sucrose gradient. The O.D. profile is shown. The figure also shows how the gradient fractions were collected in five different pools.

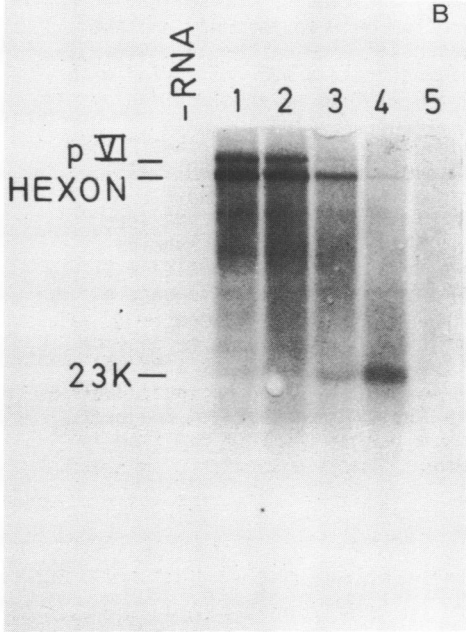


Figure 5. B) S1 nuclease analysis of RNA from the five different RNA pools which are described in panel A. The RNA was hybridized to ad2 DNA and the resulting hybrids were treated with endonuclease S1 and fractionated on an alkaline agarose gel. The separated DNA fragments were transferred to a nitrocellulose sheet by the method of Southern (1975) and hybridized with ^{32}P-labeled clone 2 DNA.

Southern (30). Fragments originating from the 3' non-coding region of the hexon were detected by hybridization of a ^{32}P-labeled cDNA clone (clone 2 DNA), which contains 800 nucleotides adjacent to the poly(A) addition site for the hexon mRNA. Analysis of RNA from pools "1" and "2" revealed two prominent bands, approximately 3600 and 4400 nucleotides long (Fig. 5B). These correspond to the mRNA bodies of the hexon and pVI mRNAs, respectively. Analysis of pool "3" RNA revealed a band in the same position as the hexon mRNA body. RNA from pool "4" gave a different pattern; a prominent, approximately 790 nucleotides long band was observed as well as a faint band

resulting from the body of the hexon mRNA. These results demonstrate the presence of a mRNA species (designated 14S) in pool "4" RNA, the main body of which maps close to the common poly(A) junction of the hexon and pVI mRNAs.

In order to map the novel 14S RNA more precisely we made use of the cleavage site for endonuclease BglII at map position 60.1. If the 14S RNA indeed corresponded to the open translational reading frame in the 3' non-coding part of the hexon gene, it would be exptected to span the BglII site at position 60.1 (Fig. 2). Therefore ad2 DNA was first cleaved with endonuclease BglII and subsequently hybridized to the 12-17S fraction of late ad2 mRNA. The hybrids were treated with endonuclease S1 before separation on alkaline agarose gels. The DNA was transferred to a nitrocellulose sheet which was then hybridized to nick translated clone 2 DNA. The results showed that the 790 nucleotides long band which was present after hybridization to uncleaved ad2 DNA was replaced by two bands, 170 and 620 nucleotides long (data not shown), clearly showing that the recognition sequence for endonuclease BglII is present in the novel 14S RNA.

The novel RNA contains a 5' leader sequence

In order to obtain additional evidence for the existence of a novel mRNA around map position 61 of the ad2 genome we used electronmicroscopy. Pool "4" RNA, selected by hybridization to clone 2 DNA was hybridized to fragment EcoRI-B (map coordinates 58.5 to 70.7) and the resulting hybrids were examined in the electronmicroscope. Numerous hybrids were observed which start from one end of the fragment and terminate with a tail 1045 nucleotides from the same end (type I molecules) (Fig. 6A). These hybrids presumably represent the 3' ends of broken hexon and pVI mRNA molecules since such truncated mRNAs would be selected on clone 2 DNA and would be expected to form a hybrid area which starts from one end of the fragment and terminate 1045 nucleotides further downstream (7). In addition to these predominant structures we observed another type of hybrid molecule (type II molecules) which is depicted in Fig. 6B. This molecule consists of characteristic structures with a 760 nucleotides long hybrid area flanked on both sides by short single stranded tails. One tail probably represents a 3' poly(A) tract and the other a 5' leader sequence. Ten hybrid molecules with the same structure as shown in Fig. 6B were observed. The average length of the hybrid area which starts 404±110 nucleotides from one end of the EcoRI-B fragment was 759±56 nucleotides, and the tails were 100-200 nucleotides long. The results

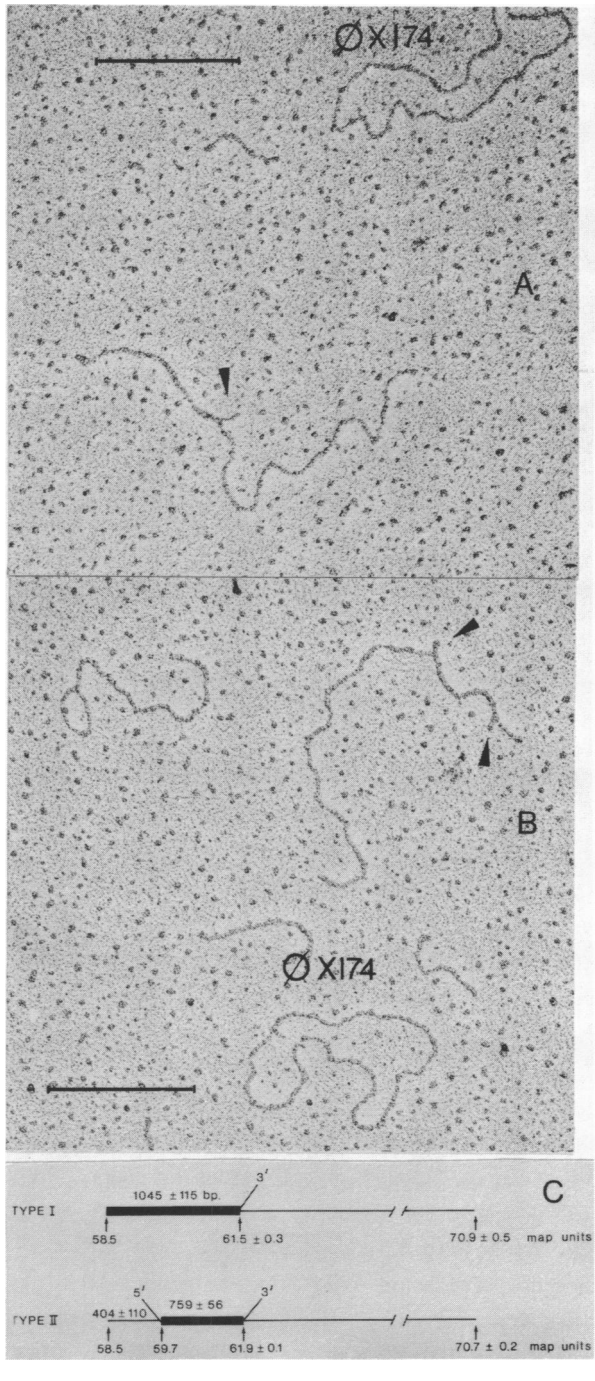


Figure 6. Electronmicroscopy of hybrids between fragment EcoRI-B and pool "4" RNA. The RNA was purified by hybridization selection on clone 2 DNA prior to electronmicroscopy. Type I (panel A) and type 2 molecules (panel B) were observed. ØX174 single stranded circular DNA and PM2 double stranded circular DNA were used as references. The magnification bar represents 1000 base pairs of double-stranded nucleic acid. Panel C summarizes data collected by measuring 15 type I molecules and 10 type II molecules.

thus demonstrate the presence of a short mRNA within fragment EcoRI-B which presumably carries a spliced 5' leader sequence. The electronmicroscopic data together with the S1 nuclease mapping show that the body of this mRNA is located between map positions 59.7 and 62.2. The position of the 3'-end suggests that the novel mRNA shares its poly(A) addition site with the hexon and the pVI mRNAs.

The presence of a 5' leader sequence could also be confirmed by hybridization. For this purpose we used a cloned cDNA copy of the 100K mRNA (Perricaudet, unpublished data) which carried the tripartite leader sequence for hybridization selection. The 12-17S RNA fraction from late ad2 mRNA was hybridized to nitrocellulose filters carrying DNA from the above mentioned clone. The RNA was eluted from the filters and subsequently analyzed by the S1 nuclease mapping technique (Fig. 7). The results showed that the novel 14S RNA was specifically selected by the cDNA clone which contained the tripartite leader. We can, however, not conclude from these results that its leader sequence has precisely the same composition as the tripartite leader of other late mRNAs.

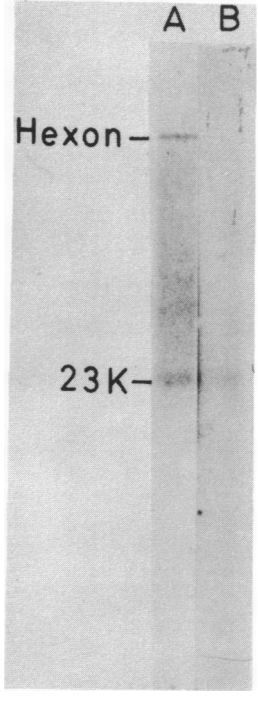


Figure. 7. Selection of the 14S mRNA by hybridization with a clone that carries the tripartite leader sequence. Panel A shows an endonuclease S1 analysis of hybrids between pool 4 RNA and ad2 DNA. The bands were visualized by hybridization with clone 2 DNA as described in the legend to Fig. 5B. Panel B shows an identical analysis of RNA which first was purified by hybridization selection on a cloned cDNA copy which contains the tripart ite leader sequence.

The position of the splice

In order to map the position where the leader sequence is attached to the main mRNA body we applied a modification of the S1 nuclease mapping method of Weaver and Weissman (31). Fragment BglII-D of ad2 DNA (map coordinates 45.3 to 60.1) was end labeled with γ-ATP and polynucleotide kinase and recleaved with endonuclease BamHI which cuts at position 59.5. The 210 base pairs long BamHI-BglII fragment (59.5 to 60.1) was isolated and hybridized to the pool "4" fraction of late ad2 mRNA (Fig. 5A). The hybrids were treated with endonuclease S1 and analyzed on an 8% sequencing gel. At the same time a fraction of the end-labeled BamHI-BglII fragment was treated according to the Maxam and Gilbert sequencing procedure (18) and selected cleavage products were separated on the same gel as the S1 nuclease digestion products (Fig. 4C-E). In this way we could pin point the splice site within 3 nucleotides of the established DNA sequence. The results predict that the splice site is located around position 43 in the sequence shown in Fig. 2. At this position the sequence reads AAGTC. Since almost all intervening sequences are surrounded by the GT....AG dinucleotides (32), and since no alternative AG dinucleotide is found within the surrounding 20 nucleotides, we would like to propose that the splice point is located between nucleotides 43 and 44 in the estalished sequence (Fig. 2). This prediction is in perfect agreement with the electronmicroscopic observations.

DISCUSSION

We report a 1164 base pair long sequence of ad2 DNA which is an extension in the rightwards direction of a recently reported sequence (13). In our previous study we showed that the coding region of the hexon gene ends 135 nucleotides to the right of the cleavage site for endonuclease BamHI at position 59.5. The termination codon for the hexon polypeptide occupies positions 136-138 in the sequence depicted in Fig. 2. The purpose of the current investigation was to determine the 3' non-coding part of the hexon mRNA and the so called strand switch point where the hexon and 72K mRNAs terminate. Our results clearly show that the 3' ends of the mRNAs overlap for a short distance as is the case for the IVa_2 and polypeptide IX mRNAs at the strand switch point near position 10.5 (28). Both poly(A) junctions around position 62 are preceded by the hexanucleotide sequence AAUAAA which is found to be adjacent to the poly(A) tails of nearly all eukaryotic mRNAs (for a review see ref 29). A comparison between the sequences at the two strand switch points which are located at position 10.5 and 62.0 reveals

that several short regions of almost perfect homology can be found (Fig. 8).

When analyzing the established DNA sequence it became apparent that it contained an unexpected open translational reading frame beyond the termination codon for the hexon polypeptide. Since initiation of translation on eucaryotic mRNAs appears to take place at an AUG triplet near the 5' terminal cap (33,34) one would not expect that the presumptive coding sequence in the 3' non-coding part of the hexon mRNA is translated from intact hexon or pVI mRNAs. Therefore, due to this observation a search for a novel spliced mRNA having its body around position 61 on the ad2 genome was appropriate. The presence of this mRNA could be demonstrated by electronmicroscopy and S1 nuclease mapping and it is clear from our mapping data that the novel 14S RNA has the same poly(A) addition site as the pVI and hexon mRNAs. It can thus be classified as a member of the L3 co-termination family (Fig. 3A). Like other late mRNAs which are transcribed from the viral r-strand, the 14S mRNA carries a 5' leader sequence which probably is identical to the tripartite leader found on other late adenovirus mRNAs (34); hence the 14S mRNA is thus likely to be controlled by the major late promoter at position 16.3.

The mRNAs which are transcribed from the major late adenovirus promoter appear mostly to encode structural proteins of the virion. One previously observed exception is the 100K mRNA which encodes a late non-structural polypeptide of hitherto unidentified function. The product of the novel 14S mRNA is probably another exception since no corresponding polypeptide has been found in the virion. Unlike most other mRNAs which are transcribed from the major late promoter, the 14S mRNA is present in very few copies per cell. Our electronmicroscopic analysis indicated that the ratio between type I and type II hybrid molecules was greater than 100:1.

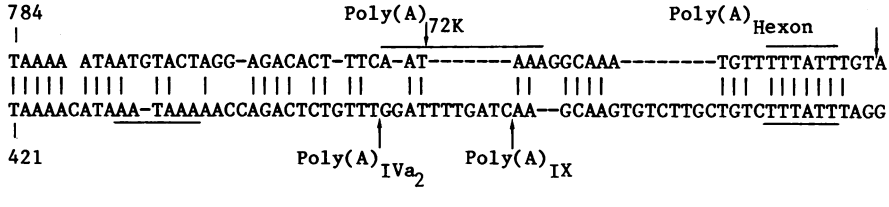


Figure. 8. A comparison of sequences at two so-called strand switch points on the ad2 genome. The upper sequence is located around the poly(A) addition sites for the hexon and the 72K mRNAs (nucleotides 784 to 837 in Fig. 2). The lower sequence includes the poly(A) addition sites for the polypeptide IX and IVa_2 mRNAs (nucleotides 421-490 in Figure 3 in ref. 28).

If we assume that the first AUG is used for initiation of translation the 14S mRNA would encode a 23K polypeptide. On translating pool "4" mRNA in a cell free protein synthesizing system, a band with a molecular weight around 20,000 was observed together with several other smaller polypeptides (data not shown). This heterogeneity could be caused either by proteolysis of the newly synthesized polypeptide or by initiation of protein synthesis on the numerous truncated pVI and hexon mRNAs which are present in pool "4" RNA.

It is not, of course, yet possible to correlate any specific function to the product of the 14S mRNA but it is, however, interesting that the ts-1 mutation in ad2 which results in a defective cleavage of the precursor polypeptides of the virion has been mapped to the 3'-end of the hexon gene (35). It has been suggested that this mutation affects a viral protease and we are currently sequencing DNA from ts-1 virus to locate the mutation.

ACKNOWLEDGEMENTS

We would like to thank Monica Pettersson and Elsy Johnsen for competent technical assistance and Marianne Gustafson and Christina Sjöholm for excellent secretarial help. Our special gratitude is due to Per Stålhandske for preparing a 72K cDNA clone. We would also like to thank Lennart Philipson and Margaret Bywater for valuable comments on the manuscript. The investigation was supported by grants from the Swedish Medical Research Council and the Swedish Cancer Society.

*Present address: The Biological Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

REFERENCES

1 Green, M., Pina, M., Kimes, R.C., Wensink, P.C., MacHattie, L.A. and Thomas, C.A. Jr. (1967) Proc. Natl. Acad. Sci. USA 57, 1302-1309.
2 van der Eb, A.J., Kestern, L.W. and van Bruggen, E.F.J. (1969) Biochim. Biophys. Acta 182, 530-541.
3 Philipson, L., Pettersson, U., Lindberg, U., Tibbetts, C., Vennström, B. and Persson, T. (1974) Cold Spring Harbor Symp. Quant. Biol. 39, 447-456.
4 Sharp, P.A., Gallimore, P. and Flint, S.J. (1974) Cold Spring Harbor Symp. Quant. Biol. 39, 457-474.
5 Pettersson, U., Tibbetts, C. and Philipson, L. (1976) J. Mol. Biol. 101, 479-501.
6 Kitchingman, G.R., Lai, S.P. and Westphal, H. (1977) Proc. Natl. Acad. Sci. USA 74, 4392-4396.
7 Chow, L.T., Broker, T. and Lewis, J.B. (1979) J. Mol. Biol. 134, 265-303.

8 Lewis, J.B., Atkins, J.F., Baum, P.R., Solem, R., Gesteland, R.F. and Andersson, C.W. (1976) Cell 7, 141-151.
9 Tooze, J. (editor) Molecular Biology of Tumor Viruses, part 2. Cold Spring Harbor Laboratory 1980, Appendices D-F.
10 Berk, A.J. and Sharp, P.A. (1978) Cell 14, 695-711.
11 Perricaudet, M., Akusjärvi, G., Virtanen, A. and Pettersson, U. (1979) Nature (London) 281, 694-696.
12 Perricaudet, M., Le Moullec, J.M. and Pettersson, U. (1980) Proc. Natl. Acad. Sci. USA 77, 3778-3782.
13 Akusjärvi, G. and Pettersson, U. (1978) Virology 91, 477-480.
14 Akusjärvi, G. and Pettersson, U. (1978) Proc. Natl. Acad. Sci. USA 75, 5822-5826.
15 Pettersson, U. and Mathews, M. (1977) Cell 11, 741-750.
16 Pettersson, U. and Sambrook, J. (1973) J. Mol. Biol. 73, 125-130.
17 Brawermann,G., Mendecki, J. and Lee, S.Y. (1972) Biochemistry 11, 637-641.
18 Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
19 Sanger, F. and Coulson, A.R. (1978) FEBS Letters 87, 107-110.
20 Akusjärvi, G. and Persson, H. (1980) Nature (London) in press.
21 Favaloro, J., Treisman, R. and Kamen, R. (1980) Methods in Enzymology, in press.
22 Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1970) J. Mol. Biol. 113, 237-251.
23 McGrogan, M., Spector, D.J., Goldenberg, C.J., Halbert, D. and Raskas, H.J. (1979) Nucl. Acids Res. 6, 593-607.
24 Jörnvall, H., Pettersson, U. and Philipson, L. (1974) Eur. J. Biochem. 48, 179-192.
25 Proudfoot, N.J. and Brownlee, G.G. (1976) Nature (London) 263, 211-214.
26 Chow, L.T., Roberts, J.W., Lewis, J.B. and Broker, T.R. (1977) Cell 11, 819-836.
27 Casey, J. and Davidson, N. (1977) Nucl. Acids Res. 4, 1539-1552.
28 Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D.F. and Pettersson, U. (1980) Cell 19, 671-681.
29 Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl.Acid Res. 8, 127-142.
30 Southern, E.M. (1975) J. Mol. Biol. 98, 503-518.
31 Weaver, R.F. and Weissmann, C. (1979) Nucl. Acids Res. 7, 1175-1193.
32 Breathnach, R., Benoist, C., O'Hare, K., Gannon, G. and Chambon, P. Proc. Natl. A ad. Sci. USA 75, 4853-4857.
33 Kozak, M. (1978) Cell 15, 1109-1123.
34 Akusjärvi, G. and Pettersson, U. (1979) Cell 16, 841-850.
35 Hassell, J.A., and Weber, J. (1978) J. Virol. 28, 671-678.