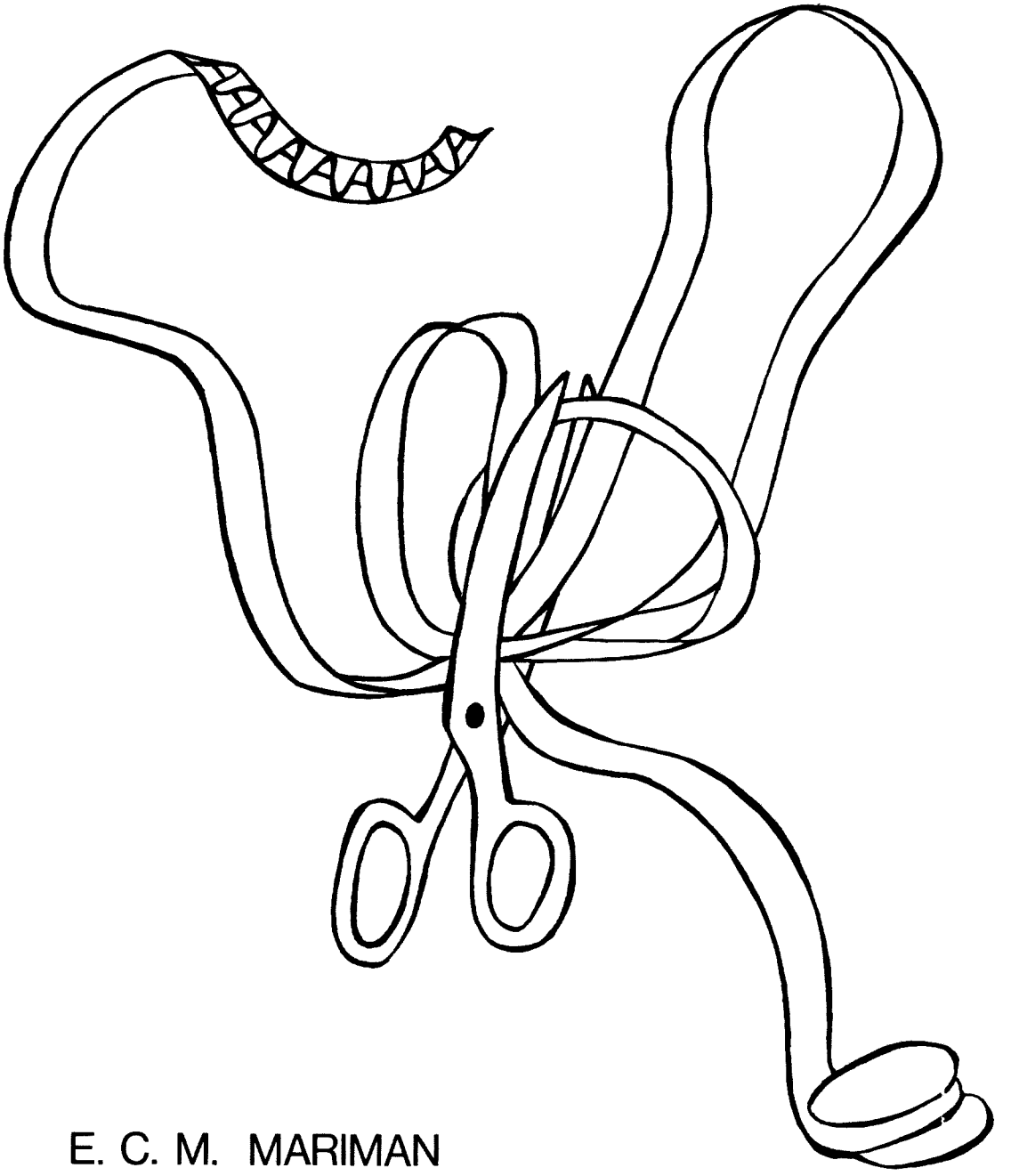
Some Aspects of pre-mRNA Splicing



E. C. M. MARIMAN

SOME ASPECTS OF PRE-mRNA SPLICING

Promotor: Prof.dr.H.Bloemendal

Co-referent: Dr.W.J. van Venrooij

SOME ASPECTS OF PRE-mRNA SPLICING

proefschrift

ter verkrijging van de graad van doctor
in de wiskunde en natuurwetenschappen
aan de katholieke universiteit te nijmegen,
op gezag van de rector magnificus,
prof.dr. j.h.g.i. giesbers
volgens besluit van het college van dekanen
in het openbaar te verdedigen
op woensdag 14 december 1983
des namiddags te 2 uur precies

door

EDWIN CYRIEL MARIA MARIMAN
geboren te Clinge

Nijmegen, 1983

DANKWOORD

Allen die hebben bijgedragen tot het gereedkomen van dit proefschrift wil ik hierbij bedanken. Met name de leden van de werkgroep Biosynthese. In de loop van mijn promotietijd is het onderzoek van Chris van Eekelen en dat van mijzelf steeds dichter bij elkaar gekomen. De discussies met Chris over het werk zijn altijd ondersteunend en stimulerend geweest. Verder dank ik Chris voor het ontwerp van de omslag van dit proefschrift. In het bijzonder dank ik ook Rita van Beek voor de vele praktische bijdragen, waardoor het onderzoek altijd een vlot verloop heeft kunnen hebben. Verder dank ik haar en Els van Genne voor het typewerk in dit proefschrift. In dit dankwoord wil ik eveneens vermelden de personen die tijdens hun stage in onze werkgroep ieder een directe bijdrage hebben geleverd: Virginia Engelaar, Anne-Marie Hagebols en Ton Hermans. Ap Groeneveld dank ik voor het kweken van de cellen. De illustraties werden grotendeels gemaakt door de Medische Tekenkamer en de afdelingen Fotografie (WNF & MF). Het in dit proefschrift beschreven onderzoek is uitgevoerd onder auspiciën van de Stichting Scheikundig Onderzoek in Nederland (SON) met financiële steun van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO).

Woody Allen

door Woody Allen

VOOR MIJN OUDERS

VOOR JUANITA

CONTENTS

Chapter I : Introduction Page

1. RNA processing 9

1.1 Capping 11
1.2 Poly(A)-addition 13
1.3 Methylation 16
- cap methylation
- internal methylation
1.4 Splicing 17
- history
- evolution
- ordering the exons
- consensus sequences
- splicer-RNA mechanism
- correct donor/acceptor site selection
- stepwise splicing
- sequential removal of introns
- function
- diversity of gene expression
- tissue specificity
- species specificity

2. HnRNP and the nuclear matrix 29

2.1 HnRNP 29
2.2 The nuclear matrix 31

3. Adenovirus serotype 2 33

3.1 Early phase 34
3.2 Intermediate phase 35
3.3 Late phase 35
- the major late transcription unit
3.4 VA-RNA 41

4. A brief survey of this thesis 42

5. References 43

Chapter II : Alternative splicing pathways exist in the formation of adenoviral late messengers RNAs. 51

Chapter III: A model for the excision of introns 1 and 2 from adenoviral major late pre-mRNAs. 71

Chapter IV : Some aspects of the excision of intron 3 from adenoviral major late pre-mRNAs. 91

Chapter V : Adenoviral heterogeneous nuclear RNA is associated with the nuclear matrix during splicing. 107

Chapter VI : On the localization and transport of specific adenoviral mRNA-sequences in the late infected Hela cell. 127

Summary 145

Samenvatting 148

Curriculum vitae 153

CHAPTER I

INTRODUCTION

Parts of this chapter are taken from:
"Some aspects of HnRNA Processing" by W.J. van Venrooij and E.C.M. Mariman, Cell Compartmentation and Metabolic Chaneling, 1980, Elseviers/North-Holland Biomedical Press.

1. RNA PROCESSING

In the nucleus of eukaryotic cells 5-15% of the DNA is transcribed into heterogeneous nuclear RNA (hnRNA). HnRNA can be distinguished from other types of nuclear RNA by pulse labeling since it is rapidly labeled before other types of nuclear RNA (as for instance (pre)ribosomal RNA) have accumulated significant label. Part of the hnRNA is processed to translatable mRNA by a series of post-transcriptional modifications (reviewed in 1-6) and is therefore referred to as "pre-mRNA". Pre-mRNA is synthesized by RNA-polymerase II and as such its synthesis can be repressed by α-amanitin, a specific inhibitor of RNA-polymerase II when used in low concentrations (7).

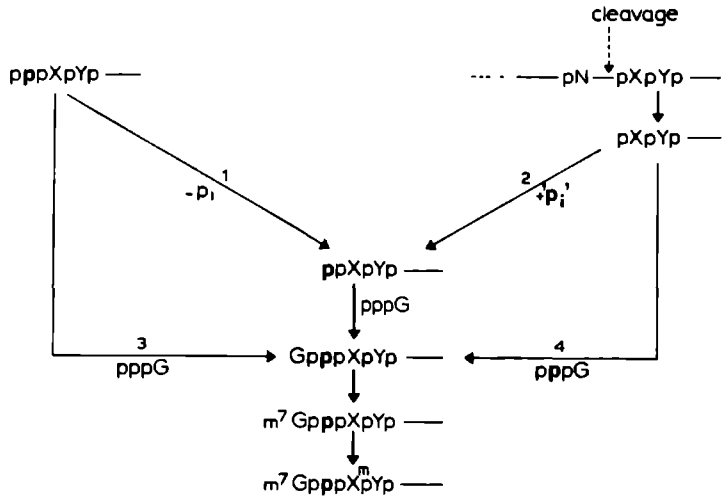
The processing of pre-mRNA to mRNA generally includes several steps: capping, polyadenylation, methylation and splicing. Each of these steps will be described below.

1.1. CAPPING

Analysis of the 5' terminus of eukaryotic mRNAs has revealed in almost all cases the presence of a blocked structure, m^7G5'ppp5'N, referred to as "cap" (8,9). Kinetic labeling experiments (10) and the presence of cap structures in prematurely terminated RNA-polymerase II transcripts (11,12) indicate that cap formation is an early event occurring on the hnRNA while transcription proceeds. In the case of reovirus it was found that the mRNA is already capped at the level of the nascent dinucleotide (13,14), while in the case of cytoplasmic polyhedrosis virus mRNA the cap is formed even before the start of transcription (15-17). Generally capping occurs before further processing as polyadenylation and splicing (11,12,22).

At least four mechanism have been suggested for cap-formation. Three of these can be distinguished by the origin of the β-phoshate of the triphosphate bridge of the resulting cap structure (Fig. 1).

1. Dephosphorylation of the 5'-triphosphate terminus of the RNA (by a nucleotide phosphohydrolase) followed by condensation



1 L cells, Reovirus, Cytoplasmic Polyhedrosis virus
2 L cells
3 Vaccinia virus
4 VSV

Fig. 1. Possible reaction schemes for capping. Capping may occur at the 5'-triphosphate terminus of the RNA (mechanisms 1 and 3) or after cleavage at an internal site (mechanisms 2 and 4). The β-phosphate of the triphosphate bridge of the resulting cap structure is provided by the RNA in the mechanisms 1 and 3, by GTP in mechanism 4 or obtained by a kinase-driven reaction in mechanism 2. In mechanism 1, 2 and 3 GTP adds a monophosphate, in mechanism 4 a diphosphate to the cap. In mechanism 3 only one enzyme (guanylyltransferase) is involved.

of the diphosphate terminating derivatives with GTP (by a guanylyl transferase) produces the unmethylated cap structure. Methylating enzymes will subsequently transfer methyl-groups from S-adenosyl-methionine to these structures. The triphosphate bridge of the resulting cap structures is derived from the α-phosphate of GTP and the α- and β-phosphates of the RNA acceptor. This pattern of reactions was observed in reovirus (14) and in cytoplasmic polyhedrosis virus (15), and seems to be operative in L-cells (18), Hela-cells (19) and rat liver cells (20) as well.

It seems likely that most eukaryotic cells follow this mechanism in cap formation.

2. Capping in L-cells may not only occur at the 5' terminus of the primary transcript but also at internal cleavage sites (18). It was proposed that an as yet not detected polynucleotide 5'-monophosphate kinase produces the intermediate diphosphate ppXpYpZp... which subsequently can react with GTP and methylating enzymes.
3. Guanylyl transferase purified from vaccinia virus catalyzed the reaction between GTP and the 5'-triphosphate terminus of RNA (21). Also in this case the triphosphate bridge of the cap is constructed from the α- and β- phosphates of the RNA acceptor (compare mechanism 1).
4. Cap-formation on RNA of vesicular stomatitis virus seems to occur after cleavage of the RNA. The 5'-monophosphate terminating RNA molecules then react with GTP under release of the γ-phosphate of GTP (8)

Although some animal and plant RNAs are not capped, generally cap-formation seems to be closely linked to the initiation of transcription by RNA-polymerase II (22). Analysis of the 5' terminal nucleotide of mRNAs has revealed that cap addition preferentially occurs to A > G >> U > C, and that a 5' end of a specific mRNA may be heterogeneous as if the polymerase looks for pyrimidine residues within a narrow region of the DNA to start initiation of transcription, accompanied by the addition of GMP to the selected ultimate residue (23-25).

The exact function of the cap remains unclear. Early capping may protect the nascent hnRNA chains from 5'-exonucleolytic degradation (26). The cap may also be involved in the binding of the RNA to nuclear or cytoplasmic skeletal structures, in transport and/or translation of the processed mRNA (27).

1.2. POLY(A)-ADDITION

To the 3' end of pre-mRNA, AMP-residues are added by nuclear poly(A)-polymerases. Reported values of completed poly(A)-segments range between 180 and 200 nucleotides (28,29).

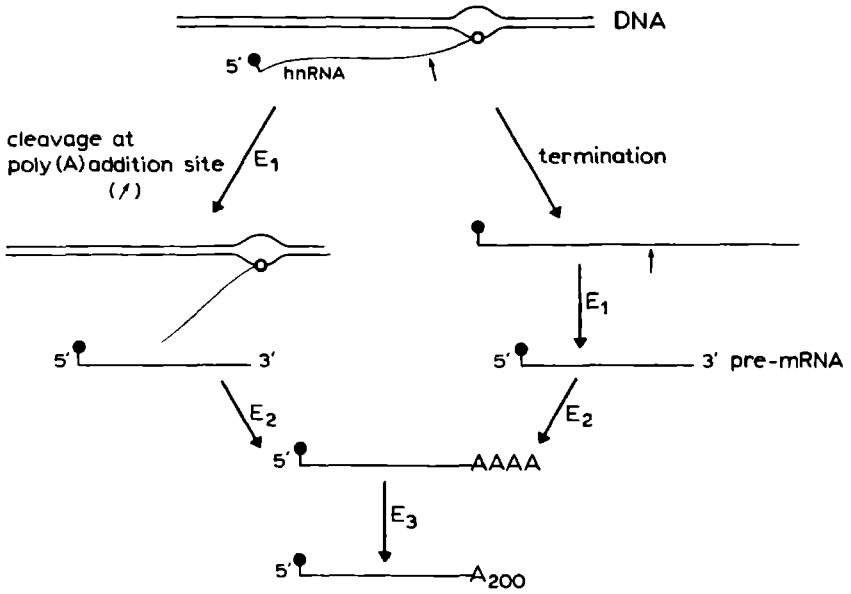


Fig. 2. Possible models for polyadenylation.
Cleavage of the hnRNA molecule at the poly(A)-site may occur before or after termination of transcription by RNA-polymerase (o). In the process of polyadenylation more than one poly(A)-polymerase (E_2, E_3) may be involved.

Polyadenylation may occur before or after termination of transcription (Fig. 2). In some viral and cellular transcription units it has been demonstrated that transcription proceeds beyond the termination site (6,30). In other cases the hnRNA seems to be terminated first and the excess RNA at the 3' end, if present, is removed before addition of the AMP residues (31-33). Anyway, it seems to be a common phenomenon that the polyadenylation site does not coincide with the site for termination of transcription (6). However, in yeast gene expression, and probably in the production of adenoviral L1-mRNAs as well, termination occurs at or close to the polyadenylation site (33,34).

Polyadenylation occurs within 1 or 2 min after transcription has passed beyond the poly(A)-site (35) and it has been suggested that it precedes the process of RNA splicing (30). However, the length of adenoviral L5-specific processing intermediates (36) and studies on in vitro transcription and processing (34) have indicated that this might not always be the case (see also Chapter II of this thesis).

The hexanucleotide 5'-AAUAAA-3' or a variation thereof, AUUAAA (38,141,152), forms part of the recognition site for polyadenylation. The synthesis of poly(A) seems to occur at a fixed distance downstream of this signal, somewhere between 11 and 30 nucleotides beyond the hexanucleotide (33,37). Other signals may also be involved in marking the polyadenylation site, since yeast mRNAs lack the common hexanucleotide (33). Although it is assumed that the poly(A)-addition occurs within a fixed distance beyond the hexanucleotide signal, it was found in prolactin messengers that this distance can vary (39). An interesting fact is that in α-amylase and chicken vimentin RNA-polyadenylation either of two sequential poly(A)-sites may be used resulting in messengers which differ in the length of their 3' untranslated region (40,242). The longer mRNA thus contains a non-used poly(A)-addition site. A similar situation in which either of three sequential poly(A)-sites can be used was observed in chicken ovomucoid mRNA production (141). An adenoviral L5 nuclear RNA species was detected which was polyadenylated at a site downstream from the poly(A)-site of L5 mRNAs (36). If this species is an intermediate in the formation of L5 mRNAs, then further processing must involve cleavage of this large polyadenylated precursor at the mRNA poly(A)-addition site followed by the synthesis of poly(A) at the newly formed 3' end. Such a processing pathway may provide one explanation for the existence of polyadenylated RNA species which do not enter the cytoplasm.

The poly(A)-polymerases catalyze the reaction:

$$\text{primer} + n\text{ATP} \rightarrow \text{primer-(AMP)}_n + n\text{PPi}$$

Characteristics of the poly(A)-polymerases are their absolute Mn^{2+} or Mg^{2+} dependency and their specificity for ATP as substrate (41-43). The whole process of polyadenylation may need

several polyadenylating enzymes. This is also indicated by the fact that in extracts of rat liver cells at last three distinct poly(A)-polymerases were found, one of them having a preference for primers containing 3' oligo(A)-stretches (44). An intriguing finding is that the poly(A)-polymerase can act as a 3'-exonuclease at low ATP concentrations (45). As a consequence the ATP concentration in the nucleus may play an important role in the turnover of nuclear poly(A). It is also conceivable that nuclear poly(A) might function as an adenosine stock which is mobilized when the nuclear ATP concentration runs too low (45). From hybridization studies it has been concluded that a.o. in Hela cells poly(A) is also added to hnRNA molecules that do not function as pre-mRNA (46,47). It is possible that in particular the poly(A)-tails added to this non-precursor class of hnRNA are functioning as an alternative adenosine pool and are subject to rapid turnover. This is in agreement with the observation that poly(A)-segments added to pre-mRNAs, as in adenovirus transcripts (30), are almost completely conserved during processing and transport to the cytoplasm. Finally, the absence of poly(A) from histone messengers, which also lack the common hexanucleotide (33), shows that a poly(A)-tail is not an absolute requirement for either translation or transport of mRNA.

1.3. METHYLATION

Cap methylation

Synthesis of the cap structure involves the addition of one, two or three methylgroups to the 5' end of the RNA. Certain mRNAs from animal and plant viruses contain $m^7GpppXpYpZp...$ (cap 0), which is not methylated further. Other RNAs have an additional metylgroup at the 2' OH position of the penultimate residue to form the cap 1 structure, $m^7GpppX^mpYpZp...$ (8,9). With purified enzymes from Hela cells it was found that the minimal chain extension to make 2'-O-methylation possible is two nucleotides (GpppXpYp) (19).
2'-O-methylation of the third residue of the 5' end, giving rise to cap 2 structures, $m^7GpppX^mpY^mpZp..$, has not been found

to occur on hnRNA and the corresponding methylating enzyme is a specific cytoplasmic enzyme (19,48-50).

Internal methylation

Internal methylations in hnRNA or mRNA are only found on adenine. The m^6Ap residues are not part of the poly(A)-segment and most RNAs contain an average of 2-3 base methylations per 5'-terminus (49-52). In Hela cell mRNA, and in other cell types as well, the m^6A residues are preceded by a purine and followed exclusively by cytosine (52-55). This indicates that this short sequence, albeit not the only requirement (55), is an important part of the signal recognized by the methylating enzyme.

Most of the cap structures as well as the internal adenine methylations seem to be maintained during processing (48,53, 56,57). The internal methylations may be located in the protein-coding region of the messengers (55). Until now no function could be ascribed to the m^6A residues. Some messengers may even lack internal methyl groups (58). Several workers have suggested a role for the methylated residues in marking the places for other processing steps (for example splicing) (55,56) a principle which has already been established for rRNA and tRNA precursors.

1.4. SPLICING

History

In 1977 electron microscopic studies on the expression of DNA tumor viruses led to the discovery that viral mRNAs are composed of pieces of RNA which are transcribed from non-contiguous regions in the DNA (59,60,61). Although originally several mechanisms were proposed to explain this phenomenon, further studies showed that the pieces of RNA, which are destined to constitute the messenger (exonic-regions, exons) are spliced with the removal of the intervening sequences (intronic-regions, introns) from a precursor molecule, which is colinear to the DNA. Soon after the discovery of splicing

in viral mRNAs it was also found to occur in the formation of cellular mRNAs, and nowadays it seems that splicing is a common step in the expression of eukaryotic genes.

Evolution

As far as evolution is concerned the general idea has arisen that the presence of intervening sequences in genes is a reflexion of a very early evolutionary state of a genome (62). The presence of intervening sequences may facilitate the reshuffling of exonic sequences by crossing-over within the introns (63). Combined with the idea that exons code for functional units of proteins (64), this presents a mechanism in which the combination of different exons may lead to the formation of new proteins. In this way the intervening sequences give an advantageous quality to the genome and may, therefore, have been conserved through evolution. However, only in a few cases, as for instance globin (65,66), lysozyme (67), ovomucoid (144) and immunoglobin (68), exons can be clearly correlated with functional units. This discrepancy might in part be due to the loss of introns from genes. In such cases exons are composed of more than one structural unit which are called "modules" (69). Loss of introns from genes has been demonstrated, for example, in the two rat insulin genes, one of which has two introns and the other contains only one of the introns (70). Thus in eukaryotes there might exist a mechanism by which intervening sequences could be accurately removed from genes. Some pseudogenes demonstrate extreme cases of this phenomenon since they have lost all intervening sequences (71,72) and are therefore an exact copy of the mRNA. These pseudogenes simply may be reintegrations of cDNAs derived from reverse transcription of totally spliced RNAs (73,74). This opens the possibility that in the case of rat insulin the gene lacking one intron may have arisen via the reintegration of a cDNA of a partly spliced RNA molecule. The surprising homology between ovalbumin and α1-antitrypsin has led to the suggestion that their genes have evolved via the selective loss of introns from a duplicated ancestral gene. The exon-sequences in both genes are related but the intron-sequences and their number

(7 and 3 respectively) are not and they are present in different locations in the gene (75). Therefore, selective loss of introns may have occurred in their formation from an ancestral gene which contained 10 introns. A smaller degree of intron variability has been observed in the actin gene family. Although the genes of insulin, actin, ovalbumin and α1-antitrypsin may have been formed by the selective introduction of introns in intron-less ancestral genes (76), it is more likely that a reintegration of cDNAs of (partly) spliced RNAs is responsible for the loss of introns. The use of retroviruses as cloning vectors, in which loss of introns from inserts can occur, can be seen as an argument in favor of the latter mechanism (77).

Ordering the exons

The average gene constitutes approximately 20% exon- and 80% intron-sequences. The size of introns may vary considerably from about 30 nucleotides, as for instance in the late SV40 pre-mRNAs, to about 21000 nucleotides in adenoviral L5 pre-mRNAs. The length of an intron might be related to the rapidity of removal of the intron but no proof exists for a direct relation. Irrespective of the length of an intron all the exon parts appear in the spliced product in the same order as they occur in the DNA. This obviously is a consequence of the mechanism by which the introns are removed. Every donor site (5' end of an intron) seems in some way to be linked to the exact acceptor site (3' end of an intron). Selecting the wrong donor and acceptor site would lead to the excision of one or more exons. Loss of an exon by aberrant splicing was observed in the expression of a mutant κ-immunoglobulin light chain gene. The absence of a donor site from one intron caused the removal of the V-segment since splicing occurred between the acceptor site of the incomplete intron, following the V-exon, and the donor site of the preceding intron (78). The simplest way for bringing the correct splice sites together would be via base pairing of the splice regions to one another. Although incidentally this might be the case (79-82) it is not a general mechanism since there seems to be hardly any complementarity between donor and acceptor regions.

Consensus sequences

In comparing donor sites as well as acceptor sites it was found that introns start with 5'GU and end with AG3'. These consensus dinucleotides are referred to as the "Chambon rules" (83). When more gene sequences became known and the splice site regions were lined up the consensus could be extended to Pu/GUAAGU for the ideal donor site and PyPyXPyAG/ for acceptor sites (84,95). The nucleotides with the exception of GU and AG can obviously deviate from the consensus without preventing splicing. The importance of specific intron sequences for splicing was determined to some extent by studying the expression of mutant genes containing deletions or point mutations. The results indicate that most of the internal intron-(85) (and exon-(86)) sequences can be deleted without affecting correct splicing. Complete deletion of the donor site inhibits splicing (87) or leads to the selection of the wrong donor site (78). Deletion of the 5' site of an intron of the yeast actin gene including the U of the consensus GU dinucleotide reduces splicing (85). Similarly, deletion of UGAGG from GUGAGG in the 5' splice site of the first intron of α2-globin leads to the absence of α2-mRNA and α-thalassemia (88). Severe reduction of splicing was observed in the expression of an adenovirus mutated region E1a (89). Here, a point mutation changing the GU dinucleotide to GG prevented the formation of the 12S mRNA. Another change in this donor site, GUGAGG to GUGAAU, also led to a splicing defect probably caused by the G to A change in the 5th nucleotide (90). A change from GU to GC, together with a six nucleotide deletion in the intervening sequence, was present in an ADH negative mutant of Drosophila melanogaster. This mutant produced no mature mRNA indicating that splicing was inhibited (153). Little data are yet published concerning the importance of the nucleotides in the 3' splice sequence. It was shown that deletion of the acceptor site prevents splicing (122). The AG dinucleotide is very important as a recognition site for the splicing enzymes. This is demonstrated by the aberrant splicing of β-globin transcripts in patients suffering from β^+-thalassemia. A single G to A base substitution obviously has created a recognizable acceptor site in the small

intron, CUAUUGG/ to CUAUUAG/ (91-94). Further, there has been suggested that the pyrimidine rich region, which precedes the consensus acceptor sequence (84), also may contribute to the efficiency of splicing. Deletion of this region accompanied by an AG to GG change in the conserved dinucleotide prevented splicing in an ADH negative mutant of Drosophila melanogaster (153). Although at the present time the GU and AG dinucleotides seem to be strict requirements, one case is found in which the donor site is changed from GU to GC without complete inhibition of splicing. In mouse αA-crystallin RNA-processing such a site seems to be used for splicing with about 10% efficiency (96). This indicates that other factors, which compensate for the deviation in the primary sequence, are also involved in the recognition of splice sites.

Splicer-RNA mechanism

The existence of the "Chambon" dinucleotides and the consensus sequences leads us to another proposed mechanism for bringing splice sites together, i.e. the use of a so-called "splicer-RNA" which base-pairs to both the donor and acceptor region. It was proposed that the splicer-RNA might be an integral part of the mRNA precursor. As such the poly(A)-tail was presented (97). By forming a triple RNA-helix the poly(A) might bring splice sites in close proximity. This mechanism, of course, requires that poly(A)-addition precedes splicing. Although it was originally supposed that polyadenylation indeed occurs before splicing (30), recent findings strongly suggest that splicing can occur on non-polyadenylated RNAs as well (34, Chapter II of this thesis). Furthermore, inhibition of poly(A)-addition does not prevent splicing (154). Altogether, these results argue against an important role for poly(A) as splicer RNA.

Instead of being part of the primary transcript the splicer RNA might exist as an individual RNA species. Based on hybridization experiments (98) and a base pairing model (99,100) it was suggested that VA-RNA(I), a 160 nucleotides long adenoviral RNA which is synthesized in large quantities during the late stage of infection, is involved as splicer-RNA in the formation of virus-specific late mRNAs. However, no direct evidence for

such a role of VA-RNA(I) in vivo has been obtained. Further, mutants of the adenovirus which do not synthesize VA-RNA do not show an alteration in splicing, but rather in the initiation of translation of the late mRNAs (101).Therefore, the function of VA(I) as splicer-RNA is at least doubtful.

Experiments do, however, point to a class of small RNAs, which are present in the nucleus of eukaryotic cells as the most likely candidates for splicer-RNA and probably for other functions in RNA-processing as well. These RNAs are referred to as snRNAs (small nuclear RNAs/ stable nuclear RNAs) and are associated with a set of proteins(snRNP)(102).The main snRNAs are called U1 to U6, of which U3 seems to be confined to the nucleolus (103). When nuclei of early adenovirus infected Hela cells were treated with antiserum directed against the U1-associated proteins, splicing was inhibited (104). Introduction of antibodies directed against U1-snRNP into adenovirus infected Hela cells reduced the expression of two late genes (140). Using psoralen for in vivo RNA-RNA cross-linking it was found that U1, and also U2, could be cross-linked to hnRNA (105,106). From these experiments it was deduced that U1 may function as splicer-RNA. A base pairing model was presented in which U1 brings two exons together (107,108). Nuclease experiments have indicated which parts of the U1-sequence are left un-protected by the associated proteins and are, therefore, available for base pairing (109). It seems that the 5' end of U1 contains a purine-rich sequence which can form a stable hybrid with the pyrimidine-rich region preceding the consensus acceptor site (107,108). As mentioned above this base pairing might be involved in regulating the speed of excision of an intron. Although the role for U1 as splicer-RNA is more or less accepted, it may well be that other snRNAs are involved in RNA splicing as well together with U1 in a more complex structure (110).

Correct acceptor/donor site selection

As mentioned above, correct splicing requires that the correct donor and acceptor site are brought together. The splicer-RNA mechanism cannot by itself affect such a specificity. One way to achieve this specificity is through the binding of

the splicing enzymes to either the donor or the acceptor site. Such a recognition might be provided by the secondary structure of the RNA around the splice point (99). The enzymes then scan the RNA sequence until the other end of the intron is reached. Subsequently, the looped-out intron sequences are cleaved off and the exons spliced. A triplet motion, like ribosomes pass over the messenger, was proposed by Ohno (111). Although the basic idea is very interesting, in relation to more recent data this triplet motion model is far from being acceptable. Another possibility for bringing correct splice sites together could be through the association of RNA with proteins. In the RNP structure the RNA may be folded in such a way that the correct donor and acceptor sites are in close contact. Studies on SV40 RNA have indicated that intron sequences may be less comprised in an RNP structure than the exon sequences (112). Splicing would then simply be the digestion of the "naked" RNA pieces and linkage of the protein-associated sequences. Studies on nuclease sensitivities of adenoviral RNA have indicated a special RNP structure for viral pre-mRNA (113). In vivo cross-linking experiments performed by van Eekelen et al. (114), however, do not demonstrate a clear difference in protein-association between intron- and exon-sequences. The importance of the RNP structure, in particular in connection with the nuclear matrix, will be dealt with in this thesis.

The exact nature of the splicing mechanism remains unclear, but it may well turn out to be a combination of all the aspects discussed above, i.e. primary-, secondary- and RNP-structure, splicer-RNA, and the RNA-scanning model.

Stepwise splicing

Introns can have the extensive length of several Kb as for example in the adenoviral L5 pre-mRNAs. The presence of extra leader sequences in certain L5 mRNAs, especially those produced in monkey cells, suggested that excision of (larger) introns might be a stepwise process (115). Further, heteroduplex mapping of major late adenoviral nuclear RNAs has indicated that removal of the second intron from these pre-mRNAs probably is a step-

wise process as well (116). Also from the study of cellular RNAs arguments for stepwise splicing can be obtained, as for instance form the β-globin pre-mRNAs. In this case evidence has been presented that both introns may be excised in more than one step (117-119). Other examples are the excision of introns from α-collagen (120) and fibroin (121) pre-mRNAs. The few experimental data obtained so far indicate that the removal of an intron follows the 5' to 3' direction, i.e.pieces at the 5' end of an intron are removed first. It may soon turn out that stepwise splicing is a general aspect in the removal of introns (see also Chapter II & III of this thesis). A stepwise splicing process requires the presence of splice signals located within the intron sequences. In the case of α-collagen and β-globin the internal splice sites were mapped (119,120). Most of the detected internal sequences deviate, however, from the consensus splice sequences.

Sequential removal of introns

The number of introns present in eukaryotic genes may extend 50, as is the case in the α-collagen gene (120). When more than one intron has to be removed from a primary transcript, a mechanism may be involved in regulating the successive removal of the introns. Only a few cases are known in which a certain directionality can be detected in the removal of introns. The pre-mRNA of the adenoviral DNA-binding protein contains two intervening sequences. Kinetic measurements have indicated that the 5' located intron is removed before the 3' intron (123). However, heteroduplex mapping of nuclear RNA has contradicted data (124). In the formation of the tripartite leader, present in late adenoviral messengers, two introns have to be removed. Kinetic analysis of the ligation reaction, involved in splicing the three leaders, shows that splicing follows the 5' to 3' direction (125). β-globin RNA molecules isolated from rabbit bone-marrow cells suggest that splicing of this pre-mRNA also follows the 5' to 3' direction (118). Of the two introns in the cellular pre-mRNA the 5' located intron seems to be removed before the 3' intron. However, results obtained from studying the expression of the mouse β-globin gene contradict

these data, since here the 3' intron is partly removed before the 5' intron is attacked (117). This difference may be due to species-specific gene expression with the existence of alternative splicing pathways in the processing of β-globin pre-mRNAs. The structure of splicing intermediates of the Xenopus vitellogenin gene indicates that the splicing order in this gene does not follow a single direction (126). Taken together, a 5' to 3' direction for the removal of multiple introns seems likely for some genes, but cannot be established as a common rule. Observations in β^+-thalassemia, a defect caused by aberrant splicing of β-globin pre-mRNA, indicate that the removal of the small intron may facilitate the removal of the large intron (91). Another example of non-random removal of introns was observed in ADH negative mutants of Drosophila melanogaster (153). Preventing the removal of one intron seemed to inhibit the removal of other introns. This could in general explain the regulation of the sequential removal of intervening sequences.

Function

In looking for a function of intervening sequences in genes and for the splicing process some interesting ideas have been proposed. As already mentioned above intervening sequences in genes may provide the genome with a certain flexibility since reshuffling of exons and the creation of new exons from introns (155) may lead to the formation of new polypeptides. This may be regarded as a positive aspect in the course of evolution (63).

When the intron present in VP1 pre-mRNA of SV40 was deleted from the DNA, no stable mRNA accumulated in infected monkey cells (127). The same result was found when the intron with both splice junctions was deleted (128). When the expression of SV40 transducing viruses carrying part of the mouse β-globin gene was studied, similar results were obtained. Preventing the splicing process prevented the accumulation of stable RNA (129). Splicing may, therefore, have a stabilizing influence on RNA. To explain these findings it has been proposed that splicing and nucleocytoplasmic transport are coupled processes (127, 129). In this case splicing is regarded as the last step in

RNA-maturation. The last event of splicing would be performed near the nucleopores. With the removal of the last intron-sequences the completed product of processing would be ejected in the cytoplasm (129,130). However, preventing splicing in adenoviral E1a RNAs did lead to the accumulation of unspliced products in the cytoplasm rather than in the nucleus (122). In the expression of a gene chimaera of SV40 and mouse β-globin, in which the donor site of SV40 is spliced to the acceptor site of globin, the spliced RNA was not transported to the cytoplasm (131). This indicates that splicing is not a requirement for transportation of messengers. Still, it is likely that there is some linkage between processing and transport since normally pre-mRNAs are confined to the nucleus while processed mRNAs are predominantly cytoplasmic (132). Further, in patients suffering from β^+-thalassemia there seems to be a delay in the transport of β-globin mRNA to the cytoplasm (133). The existence of unspliced messengers is a problem in finding a specific function for RNA splicing. Since this class of messengers does not undergo splicing it demonstrates that the cell can handle mRNA-sequences efficiently without the involvement of the splicing process. Examples of unspliced messengers are the mRNAs of histone (134), interferon (135,136), cytochrome-c (137) and the adenoviral polypeptide IX (138). Of these mRNAs the histone messengers differ in another processing step because they also lack the poly(A)-tail (139). The possibility exists that the cell treats histone messengers, and unspliced messengers in general, in a way different from spliced mRNAs with respect to RNA processing, transport and cytoplasmic localization. A comparison of the transport-rates of spliced and unspliced adenoviral messengers will be described in this thesis (Chapter VI).

Diversity of gene expression

Rather than functioning in the mechanism of processing and transport, splicing contributes to the diversity of gene expression. Splicing is the linkage of two pieces of RNA with the removal of the intervening sequences. The diversity of gene expression arises when a pre-mRNA can be spliced in several ways. Four situations can be distinguished as shown in Figure 3.

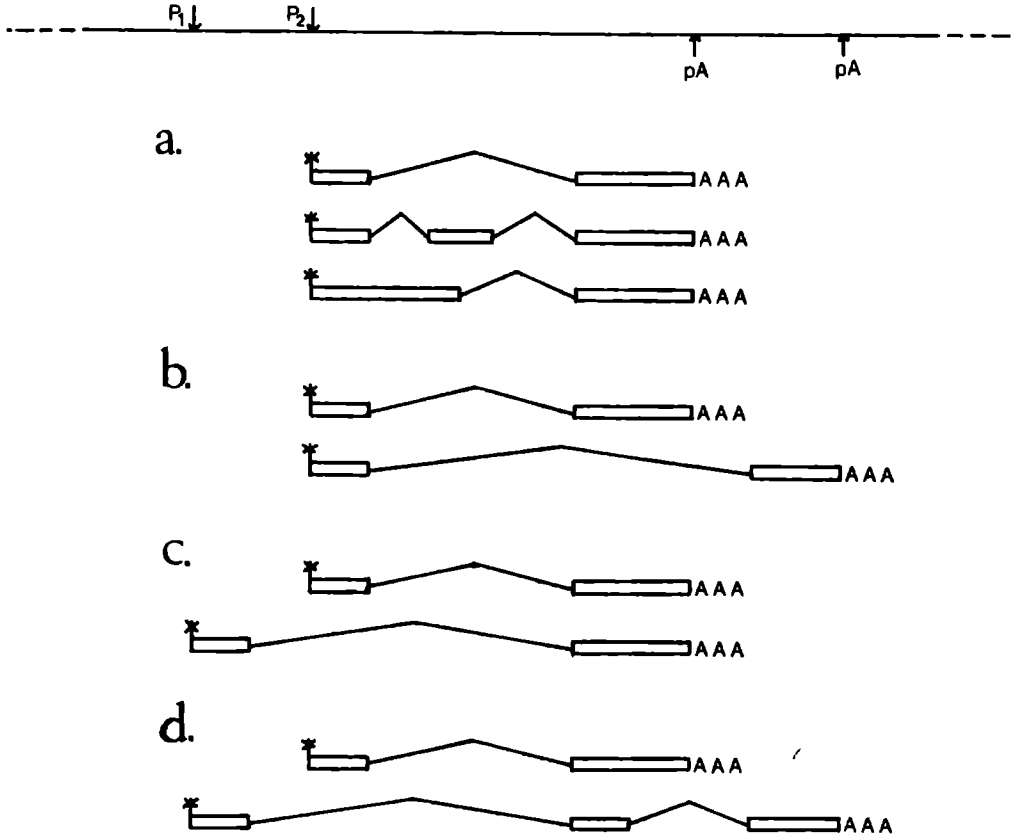


Fig. 3. Diversity of gene expression via alternative splicing patterns. Four situations can be distinguished in which mRNAs may differ in their 5', 3' and/or internal nucleotide sequences. P_1 and P_2, alternative promotor sites. pA, polyadenylation site. Boxes represent exon sequences while caret symbols represent the excised intervening sequences.

A. mRNAs can have the same 5' and 3' end but may differ internally. This kind of alternative processing patterns is known to exist in the expression of many DNA- and RNA-viruses, but also of cellular genes. Examples are chicken ovomucoid (141, 144), rat γ-fibrinogen (145), human growth hormone (146) and αA-crystallin (96).

B. Common 5' located exons are spliced to different 3' exons. The polyadenylation site of one mRNA is located in the intron of another mRNA. Examples of genes in which the expression

follows this pattern are the adenovirus major late trancription unit, adenovirus early regions IIa and IIb, formation of μ- and δ-heavy chains in GCL2.8 cells (147), formation of μ-chains of the membrane-bound (μ_m) and secreted (μ_s) forms of IgM in B lymphocytes (148).

C Different 5' exons can be spliced to common 3' exons. In the formation of these messengers transcription starts from different promotors, but the mRNAs use the same polyadenylation site. Known examples are adenovirus regions EIIb/IVa2, regions EIb/pIX, region EIIa early/late and α-amylase (142,143).

D. 5' and 3' exons can differ resulting from the use of different promotors and different polyadenylation sites, but the mRNAs have some internal sequences in common. An example is formed by the adenovirus EIII and L5 mRNAs.

Tissue-specificity.

The proteins produced from mRNAs formed via alternative splicing patterns may either have a closely related function or may be functionally different. Alternative splicing patterns are also observed in the expression of the calcitonin gene (149,150). Although it is not yet clear how the splicing patterns differ, different splicing pathways correspond to different cell types. In the thyroid the calcitonin mRNA predominates while the calcitonin gene-related peptide (CGRP) is the major product from this gene in the hypothalamus. These findings demonstrate that tissue-specific gene expression can occur as a result of different splicing pathways. Another example is the formation of mouse α-amylase mRNAs in liver and salivary gland. Results suggest that here in each tissue a different promotor is active and that a promotor-proximal leader sequence and common body sequences are spliced. The mRNAs from each tissue have therefore a different 5' untranslated region (142,143).

Species-specificity.

As already mentioned above, in SV40 transducing viruses which contain part of the β-globin gene of mouse the globin part is spliced correctly in monkey cells (129). In the

expression of a gene chimaera of SV40 and mouse β-globin the donor site of SV40 is used with the acceptor site of mouse globin (131). Similar results were obtained with an Ad2-SV40 hybrid virus, Ad2^{+}ND4. Although splicing can be tissue specific as observed in some cases, these results suggest that the mechanism of splicing is the same for all cell types. However, there seems to be some species-specificity in respect of splicing among eukaryotes since transformation in yeast cells of gene pieces from other species showed an abnormal expression and splicing did not occur (151,156). Therefore, it may well be that yeast cells differ from other eukaryotes regarding the mechanism of splicing.

2. HnRNP AND THE NUCLEAR MATRIX

2.1. HnRNP

From the preceding paragraphs it may be clear that the primary sequence of a pre-mRNA molecule by itself is not enough to permit correct processing, in particular RNA splicing. Other factors are obviously involved as well. The structural organization of heterogeneous nuclear RNA (reviewed in 157, 158) has been studied extensively in order to understand the mechanism of RNA processing and its regulatory role in the control of gene expression (159). All hnRNA,containing mRNA sequences or not, is associated with protein to form ribonucleoprotein particles (157,158,160). HnRNA-protein complexes, generally referred to as hnRNP particles, can be isolated from nuclei by extraction at elevated pH or sonication. The size of the isolated particles depends on the extent of ribonuclease degradation during the isolation procedure. Under nuclease-free conditions large 30-250 S hnRNP particles (polyparticles) can be recovered, whereas after mild nuclease digestion 30-40 S monoparticles accumulate (113,158). The polypeptide composition of the isolated complexes is also dependent on the isolation procedure used. Polypeptide compositions ranging from almost all non-histone nuclear proteins to a few distinct polypeptides have been reported (157,158,161-165). The protein

composition of purified monoparticles, which are considered as the structural building blocks of hnRNP, is relatively simple. 30 S particles from Hela cells contain 6 major polypeptides, subdivided into three groups called A,B and C proteins and ranging in molecular weight from 32,000 - 44,000 dalton (165-167). Each group consists of two polypeptides.

An important experimental approach in studying hnRNP structure involves limited nuclease digestion, which has been applied to determine the RNA sequences protected by and thus associated with proteins. Although no discrete sequence specificity was found using a total cellular hnRNP population (168), Munroe and Pederson (169) found, following this strategy, that mRNA sequences in hnRNP are complexed with proteins. No discrete lengths of protein-protected RNA were obtained upon limited nuclease digestion distinguishing hnRNP organization from the repeating subunit structure of chromatin (170). Even after mild nuclease treatment of hnRNP, pre-mRNA is preferentially present in large hnRNP particles and not randomly distributed over small and large particles (113). Large particles might therefore be a distinct subclass of hnRNP in which splicing takes place, whereas monoparticles contain RNA confined to the nucleus.

In order to study hnRNP specifically and not be limited to the detection of only structural features, it is mandatory to isolate hnRNA-protein complexes containing specific RNA sequences. It is technically feasable to isolate intact pre-mRNA from hnRNP particles (171,172) and from nuclear matrix preparations, in which the hnRNA is also associated with a specific set of hnRNP proteins (173, chapter V of this thesis). One approach follows the purification of photochemically cross-linked hnRNA-protein complexes by hybridization to immobilized DNA probes (171,174,175).

Electron microscopy has also been used to study the structural organization of specific transcripts. Chromatin actively transcribed by RNA polymerase II obviously contains fibrillar hnRNP structures resembling "beads on a string" which extend away from the DNA-protein axis. The RNP configuration for products from the same transcription unit is similar and non-random with respect to the RNA sequence (176,177).

Steitz and Kamen (112) have studied RNP organization in 30S particles containing late polyoma transcripts and found that from these particles the polyoma intervening sequences were preferentially lost upon nuclease digestion. Also the experiments of Ohlsson et al. (171) indicate that within a specific adenoviral RNA precursor there is a non-random protection of RNA sequences by RNP proteins. Similar results have been obtained by van Eekelen et al. (114) who demonstrated that the cross-linked hnRNP proteins are non-randomly distributed over late adenoviral pre-mRNA sequences. No difference in protein-association between exon- and intron-sequences, however, could be found.

The secondary structure of hnRNA has also been studied. Double-stranded RNA regions were found to exist in native hnRNP and to be relatively accessible to enzymatic digestion (178-180). The function of these double-stranded structures is still unresolved.

2.2. THE NUCLEAR MATRIX

A higher level of structural organization of hnRNA has also been encountered. When isolated nuclei are depleted of their membranes, soluble molecules and chromatin by subsequent treatments with detergents, nucleases and high salt, a structural framework, referred to as nuclear matrix, remains (181,183). The isolated nuclear matrix consists of three morphologically distinguishable structural elements: (a) a peripheral layer, which represents the remainder of the nuclear envelope and contains pore complexes in association with a lamina, (b) residual nucleoli and (c) internal grano-fibrillar structures.

It has been suggested that aggregation or precipitation of nuclear proteins or disulphide bridge formation between nuclear protein during cell fractionation could cause the artificial formation of the internal matrix structures (184,185). However, evidence is now accumulating that the internal matrix represents a predominantly proteinaceous structure existing in the intact cell (186-194). Important arguments in favor of the existence of the nuclear matrix in vivo are that the internal matrix does not depend on disulphide bridge formation and that specific proteins are localized internally in the isolated matrix in a

similar way as in the intact cell. Recent studies in which the three-dimensional structural organization of isolated matrices is viewed using electron microscopy on whole mount preparations instead of thin sections, corroborate the conclusion that the internal nuclear protein structure is a reality (187).

Additional indications for the existence of an intranuclear structural framework can be deduced from studies of its functional aspects. One of these possible functions is the attachment of replicating DNA to fixed sites inside the nucleus which is now amply documented (195-200). The association of specific receptor proteins for hormones that modulate gene expression to the matrix seems to support the idea that transcriptional activity is localized at fixed sites inside the nucleus, for the hormone-receptor complexes are presumed to bind to specific sites near genes that regulate transcription (201-203). Nuclear "cage" associated RNA synthesis has been reported by Jackson et al. (204,205). Based on their studies they propose a model, in which RNA transcription takes place at fixed sites, implicating that the transcibed DNA moves through these fixed complexes. Several years before, it had already been reported that completed transcripts, i.e. hnRNA as well as nuclear (pre)ribosomal RNA, were not present in freely floating RNP particles in nuclei, but are bound to nuclear structures. When RNAse treatment during matrix isolation was omitted hnRNA and nuclear rRNA could almost quantitatively be recovered from the isolated structures (186,190, 206,209).

All these data indicate that several important processes in the nucleus do not take place in the nucleoplasm but are localized at fixed sites inside the nucleus. Regarding gene expression, it is likely to assume that the nuclear matrix is not only involved in transcription but also in mRNA maturation, a process which is confined to the nucleus, since only precursors of RNA are tightly bound to this structure (109,110),

3. ADENOVIRUS SEROTYPE 2

The adenoviruses form a subgroup of the DNA viruses. The viral genome consists of a linear, double-stranded DNA molecule with a molecular weight of about 23 x 10^6. The ends of the DNA are marked by an inverted repeat of about 150 base pairs. (211). To each 5' end a protein of 55 kD is covalently linked (212). this protein or its 80 kD precursor may function as a primer for DNA replication (213). Both strands are transcribed and are referred to as r- and l-strand according to the direction in which they are transcribed. Conventionally the DNA is divided in 100 map units (m.u.) starting with 0 at the 5' end of the l-strand and ending with 100 at the 3' end of this strand. Coordinate 0 is called the "left end" of the genome, while 100 is referred to as the "right end". Each map unit represents 1% of the DNA and equals $\pm$ 360 base pairs. The division of the genome in map units is used to locate the transcription units on the genome.

The major disease caused by human adenoviruses is an acute infection of the respiratory tract. Nearly a hundred different serotypes have now been characterized (214). Based on a number of properties of the virus, as for instance oncogenicity, the different serotypes have been divided in six groups A to F (214). Although none of the virus strains is associated with human malignancy, at least some of them can cause tumors in rodents.

The adenovirus has been widely used to study the processing of pre-mRNA to mRNA in vivo as well as in vitro. The use of adenovirus infected cells has certain advantages over the use of uninfected eukaryotic cells. In the past, before cloning techniques were well developed, viral systems were chosen because at the late phase of infection large quantities of virus-specific pre-mRNAs and mRNAs are produced, which can be individually identified by hybridization methods in which restriction fragments of the viral genome are used as a probe. The relative "simplicity" of the viral genome can be considered to as an advantage since the virus contains a limited number of transcription units. The rather extensive knowledge that already has been gathered on both the transcription and the

translation products contributes to the usefulness of the adenovirus in studying RNA processing.

The working hypothesis for using eukaryotic DNA viruses to study RNA processing is that the virus makes use of the host processing machinery and host enzymes for its replication (215). Therefore, conclusions about RNA processing obtained from viral systems are assumed to provide general information about the processing of cellular pre-mRNAs. Until now all results obtained from the study of the expression of cellular genes have confirmed the validity of this hypothesis. However, it has become clear that certain steps in adenovirus replication, which were thought to be performed exclusively by cellular enzymes, as for example DNA replication (215), may well be catalyzed by virus-coded enzymes (213).

The infectious cycle is divided into three periods: early, intermediate and late phase of infection. Each phase is characterized by the activation of certain RNA-polymerase II promoters and consequently the appearance of a new set of mRNAs and the synthesis of new proteins in the cytoplasm.

3.1. EARLY PHASE

The early phase begins with the introduction of the viral DNA in the cell and lasts till the start of the Ad-DNA replication about 8 h after infection. During this stage the cell seems to maintain its normal functions. The Ad-specific mRNA amounts to only about 1% of the total cytoplasmic mRNA. The viral proteins which are produced seem to play a regulatory role in the infectious cycle rather than being structural components of the virion (216,217,218,219). Host DNA synthesis starts to decline and gradually keeps declining during the infectious cycle (220).

Three transcription units are located on the r-strand: EIa (1.4-4.8 m.u.), EIb (4.9-11.2 m.u.) and EIII (76.6-86.0 m.u.). Both EIa and EIb are responsible for the transforming activity of the virus (221,222,223). Three other transcription units are located on the l-strand: EIIa (75.1-61.6 m.u.), EIIb (75.1-11.2 m.u.) and EIV (99.1-91.5 m.u.). Region EIIa codes for the

72 kD DNA-binding protein which may be involved in the turn-on of the Ad-DNA replication (224). Region EIIb codes a.o. for the "Bellet" protein which is covalently linked to the 5' end of the DNA.

3.2. INTERMEDIATE PHASE

The intermediate phase lasts from 8 h to 12 h after infection. During this period the Ad-DNA replication begins and two promoters are activated. One transcription unit on the l-strand (15.8-11.2 m.u.) codes for polypeptide IVa2, while the other transcription unit on the r-strand (9.7-11.2 m.u.) is involved in the synthesis of polypeptide IX, a structural component of the viral capsid. Some characteristics of the expression of the pIX transcription unit, which is located in the early region EIb, are that the expression of this gene is not inhibited by the drug DRB (dichloro-ß-D-ribofuranosylbenzimidazole (225)), which is known to promote premature termination of transcription,and that the pIX mRNA is colinear to the DNA, i.e. that the messenger is not formed via splicing (138).

3.3. LATE PHASE

The late phase of infection is the final period in the infectious cycle. It starts about 12 h after infection and ends with the lysis of the infected cell. At this stage the cellular functions are inhibited. Host proteins are no longer synthesized and although transcription of host genes seems to continue, only virus-specific mRNAs enter the cytoplasm (226). This suggests that transportation of host messengers and/or processing of host pre-mRNAs are blocked. It may also be that these functions are still normally performed, but that host mRNA is rapidly degraded as soon as it enters the cytoplasm. Ribosomal 18 and 28 S RNAs are no longer produced which gradually leads to a decrease in the number of ribosomal subunits in the cytoplasm. The production of tRNAs seems unchanged.

Two promoters are activated in this stage. One region,

transcribed from the l-strand, is called EIIa/late (72.0-61.6 m.u.) since the messengers contain the same information as those produced from the early region EIIa, but differ in the 5' leader sequence which is derived from coordinate 72.0 instead of 75.1 (EIIa/early) (227). At coordinate 16.2 on the r-strand another promoter, called the major late promoter, is activated. Transcription extends all the way to the right end of the genome. The major late transcription unit (16.2-100.0 m.u.) directs the production of structural proteins most of which make up the building blocks of the viral capsid, as for instance pentons, hexons and fibers.

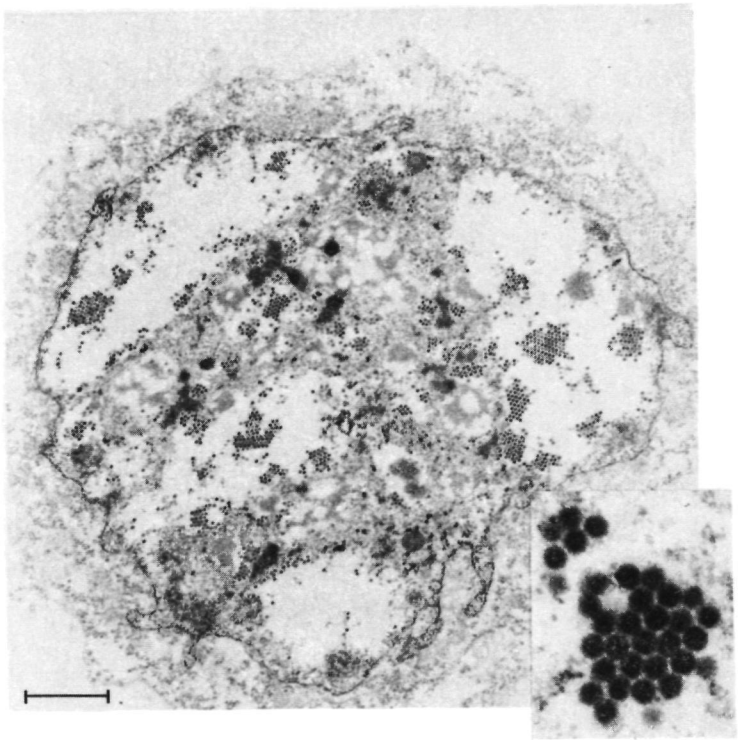


Fig. 4. Accumulation of newly formed adenovirus in the Hela cell nucleus 18 h after infection. Bar represents 1 µm. Inset shows virus particles, enlargement 54,000x. Electron micrographs made by Drs. P. Sillekens.

Other proteins derived from the major late transcription unit may function on the post-translational modification of structural proteins. Encapsidation of newly replicated Ad-DNA, starting with the left end, is performed in the nucleus. The progeny virions which are assembled in the nucleus (Fig. 4) are released by the lysis of the expiring host cell.

In Figure 5 the locations of the promoters and the transcription units are summarized. Since most of the results presented in this thesis are obtained with RNA derived from the major late transcription unit the expression of this region will be described in more detail.

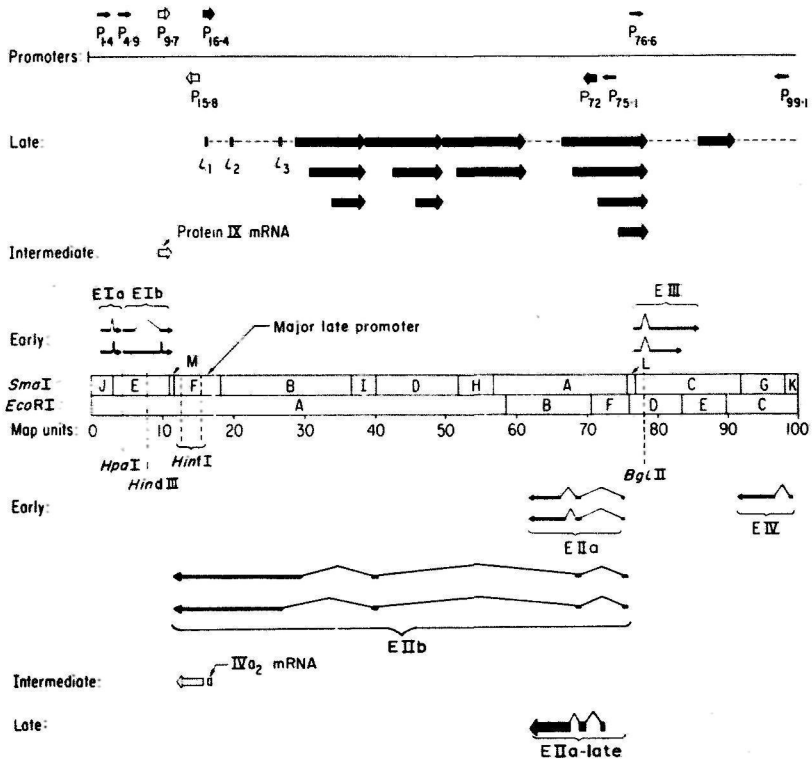


Fig. 5. Expression of the adenovirus transcription units. Ad2 RNA polymerase II promoters (P) for early (thin solid arrow), intermediate (open arrow) and late (thick solid arrow) transcription units are shown together with the major mRNA species produced by each transcription unit. Promoter subcripts indicate the cap site coordinate.

The major late transcription unit

At the late stage of infection a promoter, located at coordinate 16.2 on the r-strand, is activated. In the nucleotide sequence at this location a TATA-box was found to precede the 5' ultimate nucleotide of the RNA, to which the cap is added (23,228). Transcription proceeds to the right end of the genome, where termination occurs somewhere between coordinates 98.2 and 100.0 (229). However, premature termination of transcription, also referred to as attenuation, was found to occur with the major late promoter as well. Termination then occurs within the first 2000 nucleotides (230), but not at a single distinct location so that multiple discrete RNA species can be detected (231). The drug DRB enhances attenuation in such a way that more transcripts are prematurely terminated and that termination occurs even closer to the promotor site as compared to drug-free conditons (231). The products of premature termination are already capped (11,12,231). This and other findings (10) indicate that capping is a very early event in RNA processing and is closely linked to the initiation of transcription. In the nucleus the late transcripts are internally methylated and these methylgroups are conserved during processing (56).

When attenuation does not occur, each transcription initiation event leads to the formation of one out of several possible messengers (30). About 15 different mRNAs can be derived from the major late transcription unit and as already mentioned most of these mRNAs code for structural components of the virion. The messengers can be divided in five classes according to the map location of their 3' end. These 3' co-terminal mRNA families are formed by polyadenylation of the precursor RNA at one of 5 different polyadenylation sites and are referred to as: L1 mRNAs (38.5 m.u.), L2 mRNAs (49.5 m.u.), L3 mRNAs (61.5 m.u.), L4 mRNAs (78.3 m.u.) and L5 mRNAs (91.5 m.u.) (232). Polyadenylation occurs very rapidly (233). As soon as the RNA-polymerase has passed a few thousands of nucleotides beyond a polyadenylation site the nascent transcript is cleaved at this site and on the newly formed 3' end a poly(A)-stretch is added (30). Polyadenylation is assumed to be performed before splicing (30). However, the structure of certain L5-specific nuclear RNAs (36) and the

study on in vitro transcription and processing (34) indicate that this may not always be the case (see also Chapter II of this thesis). Although L2 to L5 RNAs seem to be polyadenylated by cleavage of the nascent transcript, in vitro studies have suggested that in the formation of L1 mRNAs termination of transcription may occur at 38.5 m.u. accompanied by rapid polyadenylation of the 3' terminus (34). This finding is not unlikely since at the early stage of infection transcription from the major late promotor seems to be terminated before coordinate 61.5 (234,235). It is possible that the major late transcription unit contains more than 5 polyadenylation sites. An L5-specific nuclear RNA has been reported which was found to be polyadenylated at 96.9 m.u., downstream from the L5 co-terminal poly(A)-site (36). Polyadenylation at this site, as there may be others, is confined to nuclear RNA. These species may be intermediates in the production of L5 mRNA.

Polyadenylation is the mechanism which determines to which family the matured mRNA will belong. Nothing is known yet by which mechanism a particular polyadenylation site is selected.

All the messengers derived from the major late transcription unit have three common leader segments linked to the protein-coding body (Fig 5). The length of the leaders are 41 nt (leader 1), 72 nt (leader 2) and 90 nt (leader 3), and their map locations are 16.2, 19.3 and 26.3 m.u., respectively (236). Kinetic measurements of the ligation reaction between the leaders as part of RNA splicing has shown that leader 1 is rapidly linked to leader 2 before linkage of leader 2 to leader 3 (125). The cap is added to leader 1 and, further, this segment contains the"Shine and Dalgarno sequence" (243). Both features are known to be involved in the initiation of translation. No functions have yet been ascribed to leaders 2 and 3. Cross-linking experiments performed by van Eekelen et al. (114) have shown that especially leader 3 sequences are closely associated with hnRNP proteins. As such leader 3 may be involved in the binding of late RNAs to a nuclear matrix and possibly also to the cytoskeletal framework.

The messengers of a 3' co-terminal family differ in the length of their body sequences. This length depends on the splicing pattern of the RNA family. Between two polyadenylation sites

of succesive co-terminal families a number of acceptor sites for splicing can be found. In the formation of a particular messenger one acceptor site is selected to which the tripartite leader will be spliced and in this way splicing determines which messenger will be formed within a particular 3' co-terminal family. The longer messengers thus have a number of non-used acceptor sequences. Except from their own protein-coding region the longer messengers encompass all the coding regions of the smaller messengers belonging to the same family. Therefore, most of the major late mRNAs are polycistronic, only the smallest ones being monocistronic. Only the 5' protein-coding region seems to be translated. This may be due to the fact that this region is nearest to the tripartite leader, which contains the requirements for the initiation of translation (79).

In short, both polyadenylation and splicing determine which messenger and consequently which protein will be formed in each transcription event. Polyadenylation selects the co-terminal family. Splicing selects which member of a particular family will be produced. Both processing steps are, therefore, involved in the regulation of adenovirus gene expression.

In some messengers additional leader segments may be present. In predominantly L1-specific messengers a sequence called the "i-leader" (21.5-22.6 m.u.) may be inserted between leaders 2 and 3 (236). In particular, the i-leader is present in those L1 mRNAs which are the most important products during the early expression of the major late transcription unit (234). The i-leader contains an open reading frame which extends into leader 3. Since the i-leader is closer to leader 1 than the protein-coding body, polycistronic messengers which have the i-leader probably are translated into a 16 kD viral protein (237). Up to three extra leader sequences may be present between leader 3 and the body of L5 mRNAs. These leader segments are called X, Y and Z (238) and are also involved in the formation of EIII messengers (236). It has been suggested that these leaders may be residual parts of the larger intervening sequence separating leader 3 and the body sequences, which have not been excised in a stepwise splicing process. A further indication in favor of this hypothesis is the existence of L5 messengers, especially in Ad-infected monkey cells, which contain even larger regions of the intron between coordinates 66 and 78 (115).

3.4. VA-RNA

Except for RNA-polymerase II transcription units two RNA-polymerase III genes are located on the genome near coordinates 29 and 30, respectively (239). The genes are encompassed by the major late transcription unit and transcription is from the r-strand. The RNA products are referred to as virus-associated RNAs (VA-RNAs). The major ones are called VA(I) and VA(II) and are about 160 nucleotides in length. VA(I) is accumulated in large quantities during the late stage of infection. Both RNAs have an extensive secondary structure (240) and part of the VA-RNA is associated with a protein of about 50 kD (175). A third virus-associated RNA, V200, is a read-through product of the VA(I) gene in which about 40 nucleotides are added to the 3' end (230).

The function of the VA-RNAs remains unknown. It has been suggested that they might be active as splicer-RNA in the processing of pre-mRNAs transcribed from the major late transcription unit (98,99,100). Recent results do, however, point to a role in the initiation of translation rather than in splicing (101,241).

4. A BRIEF SURVEY OF THIS THESIS

Splicing as a step in pre-mRNA processing was reported in 1977. Although since that time many data have been gathered, in many of its aspects the splicing phenomenon remains obscure. The aim of this study is to contribute to the understanding of RNA splicing. At the start of this study, DNA cloning facilities were not readily available. Therefore, based on the working hypothesis that viral genes are expressed similarly as cellular genes, we studied splicing in vivo as one of the steps in adenovirus gene expression.

The pattern of cleavages occurring during the excision of introns 1 and 2 from viral pre-mRNAs is described in Chapter II. In Chapter III the excision of intron 1 is studied in more detail. Based on the results of Chapters II and III a model for RNA splicing is presented. This model and the data obtained from studying the excision of intron 3, as described in Chapter IV, underline the need for a higher organization of the pre-mRNA. A possible answer to this requirement is presented in Chapter V, where the interaction between hnRNA and a nuclear matrix during splicing is examined. The possibility that an interaction with a nuclear matrix could be essential only for RNA molecules which have to be spliced and the idea that splicing might be coupled with nucleocytoplasmic transport have been considered in Chapter VI, in which the localization and nucleocytoplasmic transport rates of spliced and unspliced viral RNAs are compared.

5. REFERENCES

1. Revel, M. and Groner, Y. (1978) Ann.Rev.Biochem. 47, 1079-1126.
2. Rosen, J.M. (1977) in: Receptors and Hormone Action Vol I, pp 237-266, Academic Press, N.Y.
3. Brown, D.D. (1981) Science 211, 667-674.
4. Perry, R.P. (1981) J.Cell.Biol. 91, 285-385.
5. Sharp, P.A. (1981) Cell 23, 643-646.
6. Darnell, J.E. (1982) Nature 297, 365-371.
7. Sawadogo, M. (1981) Biochem.Biophys.Res.Comm. 98, 261-267.
8. Shatkin, A.J. (1976) Cell 9, 645-653.
9. Filipowicz, W. (1978) FEBS letters 96, 1-11.
10. Babich, A., Nevins, J.R. and Darnell, J.E. (1980) Nature 287, 246-248.
11. Salditt-Georgieff, M., Harpold, M., Chen-Kiang, S and Darnell, J.E. (1980) Cell 19, 69-78.
12. Tamm, I. and Seghal, P.B. (1979) in: Effect of drugs on the cell nucleus, pp 251-275, Acad.Press.Inc.
13. Furuichi, Y. and Shatkin, A.J.(1976) Proc.Natl.Acad.Sci.USA 73, 3448-3452.
14. Furuichi, Y., Muthukrishnan, S., Tomasz, J. and Shatkin, A.J. (1976) J.Biol.Chem. 251, 5043-5053.
15. Shimotohno, K. and Miura, K. (1976) FEBS letters 64, 204-208.
16. Furuichi, Y. (1974) Nucleic Acids Res. 1 809-822.
17. Furuichi, Y. (1978) Proc.Natl.Acad.Sci.USA 75, 1086-1090.
18. Schibler, U. and Perry, R.P. (1976) Cell 9, 121-130.
19. Langberg, S.R. and Moss, J.B. (1981) J.Biol.Chem. 256, 10054-10060.
20. Mizumoto, K., Kaziro, Y. and Lipmann, F. (1982) Proc.Natl. Acad.Sci.USA 79, 1693-1697.
21. Monroy, G., Spencer, E. and Hurwitz, J. (1978) J.Biol.Chem. 253, 4490-4498.
22. Jove, R. and Manley, J.L. (1982) Proc.Natl.Acad.Sci.USA 79, 5842-5846.
23. Baker, C.C. and Ziff, E.B. (1981) J.Mol.Biol. 149, 189-221.
24. Hashimoto, S., Pursley, M.H. and Green, M. (1981) Nucleic Acids Res. 9, 1675-1689.
25. Malek, L.T., Eschenfeldt, W.H., Munns, T.W. and Rhoads, R.E. (1981) Nucleic Acids Res. 9, 1657-1673.
26. Furuichi, Y., La Fiandra, A. and Shatkin, A.J. (1977) Nature 266, 235-239.
27. Clemens, M.J. (1979) Nature 279, 673-674.
28. Brawerman. G. (1976) Prog.Nucl.Acid Res.Mol.Biol. 17, 117-148.
29. Sheiness, D., Puckett, L. and Darnell, J.E. (1975) Proc. Natl.Acad.Sci.USA 72, 1077-1081.
30. Nevins, J.R. and Darnell, J.E. (1978) Cell 15, 1477-1493.
31. Perry, R.P., Kelley, D.E. and La Torre, J. (1974) J.Mol.Biol. 82, 315-331.
32. Brandhorst, B.P. and Mc-Conkey, E.H. (1975) Proc.Natl.Acad. Sci.USA 72, 3580-3584.
33. Proudfoot, N. (1982) Nature 298, 516-517.
34. Manley, J.L., Sharp, P.A. and Gefter, M.L. (1982) J.Mol. Biol. 159, 581-599.

35. Salditt-Georgieff, M., Sawicki, S., Nevins, J. and Darnell, J.E. (1980) J.Cell Biol. 86, 844-848.
36. Carlson, D.P. and Raskas, H.J. (1980) J.Mol.Biol. 141, 249-265.
37. Fitzgerald, M. and Shenk, T. (1981) Cell 24, 251-260.
38. Fraser, N.W. , Baker, C.C., Moore, M.A. and Ziff, E.B. (1982) J.Mol.Biol. 155, 207-233.
39. Sasavage, N.L., Smith, M., Gillam, S., Woychik, R.P. and Rottman, F.M. (1982) Proc.Natl.Acad.Sci.USA 79, 223-227.
40. Tosi, M., Young, R.A., Hagenbüchle, O. and Schibler, U. (1981) Nucleic Acids Res. 9, 2313-2323.
41. Rose, K.M. and Jacob, S.T. (1976) Eur.J.Biochem. 67, 11-21.
42. Antoniades, D. and Antonoglou, O. (1978) Biochim.Biophys. Acta 519, 447-460.
43. Milchev, G.I. and Hadjiolov, A.A. (1978) Eur.J.Biochem. 84, 113-121.
44. Niessing, J. (1975) Eur.J.Biochem. 59, 127-135.
45. Abraham, A.K. and Jacob, S.T. (1978) Proc.Natl.Acad.Sci.USA 75, 2085-2087.
46. Herman, R.C., Williams, J.G. and Penman, S. (1976) Cell 7, 429-437.
47. Berger, S.L. and Cooper, H.L. (1978) Biochim.Biophys. Acta 517, 84-98.
48. Perry, R.P. and Kelley, D.E. (1976) Cell 8, 433-442.
49. Pery, R.P., Kelley, D.E., Frederici, F. and Rottman, F.M. (1975) Cell 6, 13-19.
50. Rottman, F.M., Desrosiers, R.C. and Frederici, K. (1976) Prog.Nucl.Acid Res.Mol.Biol. 19, 21-38.
51. Salditt-Georgieff, F., Jelinek, W., Darnell, J.E., Furuichi, Y., Morgan, M. and Shatkin, A.J. (1976) Cell 7, 227-237.
52. Wei, C.M., Gershowitz, A. and Moss, B. (1976) Biochemistry 15, 397-401.
53. Schibler, U., Kelley, D.E. and Perry, R.P. (1077) J.Mol. Biol. 115, 695-714.
54. Dimock, K. and Stoltzfus, C.M. (1977) Biochemistry 16, 471-478.
55. Canaani, D., Kahana, C., Lavi, S. and Groner, Y. (1979) Nucleic Acids Res. 6, 2879-2899.
56. Chen-Kiang, S., Nevins, J.R. and Darnell, J.E. (1979) J.Mol. Biol. 135, 733-752.
57. Lavi, U., Fernandez-Munoz, R. and Darnell, J.E. (1977) Nucleic Acids Res. 4, 63-69.
58. Revel, M. and Groner, Y. (1978) Ann.Rev.Biochem. 47, 1079-1126.
59. Berget, S.M., Moore, C. and Sharp, P.A. (1977) Proc.Natl. Acad.Sci.USA 74, 3171-3175.
60. Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (1977) Cell 12, 1-8.
61. Klessig, D.F. (1977) Cell 12, 9-21.
62. Chambon, P. (1981) Scientific American 5, 48-59.
63. Gilbert, W. (1978) Nature 271, 501.
64. Blake, C.C.F. (1979) Nature 277, 598.
65. Craik, C.K., Buchman, S.R. and Beychok, S. (1980) Proc. Natl.Acad.Sci.USA 77, 1384-1388.
66. Craik, C.S., Buchman, S.R. and Beychok, S. (1981) Nature 291, 87-90.
67. Artymiuk, P.J., Blake, C.C.F. and Sippel, A.E. (1981) Nature 290, 287-288.

68. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15, 1133-1144.
69. Lewin, R. (1982) Science 217, 921-922.
70. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.
71. Proudfoot, N. (1980) Nature 286, 840-841.
72. Moos, M. and Gallwitz, D. (1982) Nucleic Acids Res. 10, 7843-7849.
73. Lueders, K., Leder, A., Leder, P. and Kuff, E. (1982) Nature 295, 426-428.
74. Goff, S.P., Gilboa, E., Witte, O. and Baltimore, D. (1980) Cell 22, 777-785.
75. Leicht, M., Long, L., Chandra, T., Kurachi, K., Kidd, V.J., Mace, M., Davie, E.W. and Woo, S.L. (1982) Nature 297, 655-659.
76. Cornish-Bowden, A. (1982) Nature 297, 625.
77. Shimotohno, K. and Temin, H.M. (1982) Nature 299, 265-268.
78. Seidman, J.G. and Leder, P. (1980) Nature 286, 779-783.
79. Akusjärvi, G. and Pettersson, U. (1979) Cell 16, 841-850.
80. Akusjärvi, G. and Pettersson, U. (1979) J.Mol.Biol. 134, 143-158.
81. Akusjärvi, G. and Pettersson, U. (1978) Proc.Natl.Acad.Sci. USA 75, 5822-5826.
82. Trapnell, B.L., Tolstoshew, P. and Crystal, R.G. (1980) Nucleic Acids Res. 8, 3659-3672.
83. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc,Natl.Acad.Sci.USA 75, 4853-4857.
84. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
85. Gallwitz, D. (1982) Proc.Natl.Acad.Sci.USA 79, 3493-3497.
86. Subramanian, K.N. (1979) Proc.Natl.Acad,Sci.USA 76, 2556-2560.
87. Khoury, G., Gruss, P., Dhar, R. and Lai, C.J. (1979) Cell 18, 85-92.
88. Orkin, S.H., Goff, S.C. and Hechtman, R.L. (1981) Proc.Natl. Acad.Sci.USA 78, 5041-5045.
89. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1982) Nature 295, 380-384.
90. Solnick, D. (1981) Nature 291, 508-510.
91. Busslinger, M., Moschonas, N. and Flavell, R.A. (1981) Cell 27, 289-298.
92. Fukumaki, Y., Gosh, P.K., Benz, E.J., Reddy, V.B., Lebowitz, P., Forget, B.G. and Weissman, S.M. (1982) Cell 28, 585-593.
93. Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J.T., de Riel, J.K., Manley, J.L., Gefter, M.L., Forget, B.G. and Weissman, S.M. (1981) Proc.Natl.Acad.Sci. USA 78, 2455-2459.
94. Ley, T.J., Anagnou, N.P., Pepe, G. and Nienhuis, A.W. (1982) Proc.Natl.Acad,Sci.USA 79, 4775-4779.
95. Seif, I., Khoury, G. and Dhar, R. (1979) Nucleic Acids Res. 6, 3387-3398.
96. King, C.R. and Piatigorsky, J. Cell, in the press.
97. Bina, M., Feldmann, R.J. and Reeley, R.G. (1980) Proc.Natl. Acad.Sci.USA 77, 1278-1282.
98. Mathews, M.B. (1980) Nature 285, 575-577.
99. Zain, S., Gingeras, T., Bullock, P., Wong, G. and Gelinas, R.E. (1079) J.Mol.Biol. 135, 413-433.
100. Murray, V. and Holiday, R. (1979) FEBS letters 106, 5-7.

101. Thimmappaya, B., Weinberger, C., Schneider, R.J. and Shenk, T. (1982) Cell 31, 543-551.
102. Brunel, C., Widada, J.S., Lelay, M-N., Jeanteur, P. and Liautard, J-P. (1981) Nucleic Acids Res. 4, 815.
103. Reddy, R., Henning, D. and Busch, H. (1979) J.Biol.Chem. 254, 11097-11105.
104. Yang, V.W., Lerner, M.R., Steitz, J.A. and Flint, S.J.(1981) Proc.Natl.Sci.USA 78, 1371-1375.
105. Calvet, J.P. and Pederson, T. (1981) Cell 26, 363-370.
106. Calvet, J.P., Meyer, L.M. and Pederson, T. (1982) Science 217, 456-458.
107. Rogers, J. and Wall, R. (1980) Proc.Natl.Acad.Sci.USA 77, 1877-1879.
108. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A. (1980) Nature 283, 220-224.
109. Liautard, J-P., Sri-Widada, J., Brunel, C. and Jeanteur, P. (1982) J.Mol.Biol. 162, 623-643.
110. Oshima, Y., Itoh, M., Okada, N. and Miyata, T. (1981) Proc. Natl.Acad.Sci.USA 78, 4471-4474.
111. Ohno, S. (1980) Differentiation 17, 1-15.
112. Steitz, J.A. and Kamen, R. (1981) Mol.Cell.Biol. 1, 21-34.
113. Stevenin, J., Gattoni, R., Keohavong, P. and Jacob, M. (1982) J.Mol.Biol. 155, 185-205.
114. van Eekelen, C., Ohlsson, R., Philipson, L., Mariman, E., van Beek, R. and van Venrooij, W. (1982) Nucleic Acids Res. 10, 7115-7131.
115. Klessig, D.F. and Chow, L.T. (1980) J.Mol.Biol. 139, 221-242.
116. Manley, J.L., Sharp, P.A. and Gefter, M.L. (1979) J.Mol. Biol. 135, 171-197.
117. Kinniburgh, A.J. and Ross, J. (1979) Cell 17, 915-921.
118. Grosveld, G.C. and Flavell, R.A. (1981) Cell 23, 573-584.
119. Donaldson, D.S., McNab, A.R., Rovera, G. and Curtis, P.J. (1982) J.Biol.Chem. 257, 8655-8660.
120. Avvedimento, V.E., Vogeli, G., Yamada, Y., Maizel, J.V., Pastan, I. and de-Combrugghe, B. (1980) Cell 21, 689-696.
121. Kamijo, S., Zielinski, R. and Lizardi, P.M. (1982) J.Cell Biol. 95, 477a (abstract).
122. Carlock, L. and Jones, N.C. (1981) Nature 294, 572-574.
123. Goldenberg, C.J. and Raskas, H.J. (1979) Cell 16, 131-138.
124. Kitchingman, G.R. and Westphal. H. (1980) J.Mol.Biol. 137, 23-48.
125. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stevenin, J. (1982) Nucleic Acids Res. 10, 1215-1229.
126. Ruffel, G.U., Wyler, T., Muellener, D.B. and Weber, R. (1980) Cell 19, 53-61.
127. Gruss, P., Lai, C-J, Dahr, R. and Khoury, G. (1979) Proc. Natl.Acad.Sci.USA 76, 4317-4321.
128. Lai, C-J. and Khoury, G. (1979) Proc.Natl.Acad.Sci.USA 76, 71-75.
129. Hamer, D.H. and Leder, P. (1979) Cell 18, 1299-1302.
130. Campos, R., Jovanovich, S. and Villareal, L.P. (1981) Nature 291, 344-346.
131. Chu, G. and Sharp, P.A. (1981) Nature 289, 378-382.
132. Roop, D., Nordstrom, J., Tsai, S., Tsai, M. and O'Malley, B. (1978) Cell 15, 671-685.
133. Kantor, J.A., Turner, P.H. and Nienhuis, A.W. (1980) Cell 21, 149-157.

134. Suares, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033-1044.
135. Lawn, R.M., Adelman, J., Dull, T.J., Gross, M., Goeddel, D. and Ullrich, A. (1981) Science 212, 1159-1162.
136. Tavernier, J., Derynck, R. and Fiers, W. (1981) Nucleic Acids Res. 9, 461-
137. Smith, M., Leung, D.W., Gillam, S., Astell, C.R., Montgomery, D.L. and Hall, B.D. (1979) Cell 16, 753-761.
138. Alestrom, P., Akusjärvi, G., Perricaudet, M., Mathews, M., Klessig, D.F. and Pettersson, U. (1980) Cell 19, 671-681.
139. Adesnik, M. and Darnell, J.E. (1972) J. Mol.Biol. 67, 397-406.
140. Lenk, R.P., Maizel, J.V. and Crouch, R.J. (1982) Eur.J. Biochem. 121, 475-482.
141. Gerlinger, P., Krust, A., LeMeur, M., Perrin, F., Cochet, M., Gannon, F., Dupret, D. and Chambon, P. (1982) J.Mol. Biol. 162, 345-364.
142. Flavell, R.A. (1981) Nature 290, 541-542.
143. Young, R.A., Hagenbüchle, O. and Schibler, U. (1981) Cell 23, 451-458.
144. Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) Cell 21, 681-687.
145. Crabtree, G.R. and Kant, J.A. (1982) Cell 31, 159-166.
146. Denoto, F.M., Moore, D.D. and Goodman, H.M. (1981) Nucleic Acids Res. 9, 3719-3737.
147. Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M. Raschke, W. and Tonegawa, S. (1981) Cell 24, 353-365.
148. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20, 313-319.
149. Rosenfeld, M.G., Lin, C.R., Amara, S.G., Stolarsky, L., Roos, B.A., Ong, E.S. and Evans, R.M. (1982) Proc.Natl. Acad.Sci.USA. 79, 1717-1721.
150. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature 298, 240-244.
151. Beggs, J.D., van den Berg, J., van Ooyen, A. and Weissmann, C. (1980) Nature 283, 835-840.
152. Cheng, H.L., Blattner, F.R., Fitzmaurice, L., Mushinski, J.F. and Tucker, P.W. (1982) Nature 296, 410-415.
153. Benyajati, C., Place, A.R., Wang, N., Pentz, E. and Sofer, W. (1982) Nucleic Acids Res. 10, 7261-7272.
154. Zeevi, M., Nevins, J.R. and Darnell, J.E. (1981) Cell 26, 39-46.
155. Treisman, R., Orkin, S.H. and Maniatis, T. (1983) Nature 302, 591-596.
156. Langford, C., Nellen, W., Niessing, J. and Gallwitz, D. (1983) Proc.Natl.Acad.Sci.USA 80, 1496-1500.
157. Heinrich, P.C., Gross, V., Northeman, W. and Scheurlen, M. (1978) Rev.Physiol.Biochem.Pharmacol. 81, 102-134.
158. van Venrooij, W.J. and Jansen, D.B. (1978) Mol.Biol.Rep. 4, 3-8.
159. Perry, R.P. (1981) J.Cell Biol. 91, 285-385.
160. Samarina, G.P., Lukanidin, E.M., Molnar, J. and Georgiev, G.P. (1968) J.Mol.Biol. 33, 251-263.
161. Stevenin, J., Gattoni, R., Gallinaro-Matringe, H. and Jacob, M. (1978) Eur.J.Biochem. 84, 541-549.
162. Brunel, C. and Lelay, M.N. (1979) Eur.J.Biochem. 99, 273-283.

163. Maundrell, K. and Scherrer, K. (1979) Eur.J.Biochem. 99, 225-238.
164. Karn, J., Vidali, G., Boffra, L.C. and Allfrey, V.G. (1977) J.Biol.Chem. 252, 7307-7322.
165. Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Cell 11, 127-138.
166. Walker, B.W., Lothstein, L., Baker, C.L. and LeStourgeon, W.M. (1980) Nucleic Acids Res. 8, 3639-3657.
167. LeStourgeon, W.M., Lothstein, L., Walker, B.W. and Beyer, A.L. (1982) in: The cell nucleus (editor H.Busch) part IX, pp 48-81.
168. Augenlicht, L. (1979) Biochemistry 18, 3780-3786.
169. Munroe, S.H. and Pederson, T. (1981) J.Mol.Biol. 147, 437-449.
170. Wahrman, M.Z. and Augenlicht, L.H. (1979) Biochem.Biophys. Res.Commun. 87, 395-402.
171. Ohlsson, R.I., van Eekelen, C. and Philipson, L. (1982) Nucleic Acids Res. 10, 3053-3068.
172. Pederson, T. and Davis, N.G. (1980) J.Cell Biol. 87, 47-54.
173. van Eekelen, C. and van Venrooij, W. (1981) J.Cell Biol. 88, 554-563.
174. van Eekelen, C., Mariman, E., Reinders, R. and van Venrooij, W. (1981) Eur.J.Biochem. 119, 461-467.
175. van Eekelen, C., Buijtels, H., Linné, T., Ohlsson, R., Philipson, L. and van Venrooij, W.J. (1982) Nucleic Acids Res. 10, 3039-3052.
176. Beyer, A.L., Bouton, A.H. and Miller, O.L. (1981) Cell 26, 155-165.
177. Beyer, A.L., Miller, O.L. and McKnight, S.L. (1980) Cell 20, 75-84.
178. Calvet, J.P. and Pederson, T. (1979) Nucleic Acids Res. 6, 1993-2001.
179. Calvet, J.P. and Pederson, T. (1978) J.Mol.Biol. 122, 361-378.
180. Kish, V.M. and Pederson, T. (1977) Proc.Natl.Acad.Sci.USA 74, 1426-1430.
181. Berezney, R. and Coffey, D.S. (1974) Biochem.Biophys.Res. Commun. 60, 1410-1417.
182. Berezney, R. (1979) in: The cell nucleus (H.Busch, ed.) Vol.7, pp 413, Acad.Press, N.Y.
183. Agutter, P.S. and Richardson, J.C.W. (1980) J.Cell Sci. 44, 395-435.
184. Kaufmann, S.H., Coffey, D.S. and Shaper, J.H. (1981) Exp. Cell.Res. 132, 105-123.
185. van Eekelen, C.A.G., (1983) Thesis: Structural Organization of heterogeneous nuclear RNA in eukaryotic cells.
186. van Eekelen, C.A.G. and van Venrooij, W.J. (1981) J.Cell Biol. 88, 554-563.
187. Capco, D.G., Wan, K.M. and Penman, S. (1982) Cell 29, 847-858.
188. Fischer, P.A. and Blobel, G. (1982) J.Cell Biol. 92, 674-686.
189. van Venrooij, W.J., van Eekelen, C.A.G., Mariman, E.C.M. and Reinders, R. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 235-246, Alan Liss, inc., N.Y.
190. Long, B.H., Huang, C-Y. and Pogo, A.O. (1979) Cell 18, 1079-1090.

191. Lebkowski, J.S. and Laemli, U.K. (1982) J.Mol.Biol. 156, 325-344.
192. Lebkowski, J.S. and Laemli, U.K. (1982) J.Mol.Biol. 156, 300-324.
193. van Eekelen, C.A.G., Salden, M.H.L., Habets, W.J.A., van de Putte, L.B.A. and van Venrooij, W.J. (1982) Exp.Cell Res. 141, 181-191.
194. Brasch, K. (1981) Exp.Cell Res. 140, 161-171.
195. Vogelstein, B., Nelkin, B., Pardoll, D. and Hunt, B.F. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 169-181, Alan Liss inc., N.Y.
196. Berezney, R., Basler, J., Buchholtz, L.A., Smith, H.C. and Siegel, A.J. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 183-198, Alan Liss inc.,N.Y.
197. Wanka, F., Pieck, A.C.M., Bekers, A.G.M. and Mullenders, L.H.F. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 199-212, Alan Liss inc., N.Y.
198. Vogelstein, B., Pardoll, D.M. and Coffey, D.S. (1980) Cell 22, 79-85.
199. McCready, S.J., Godwin, J., Mason, D.W., Brazell, I.A. and Cook, P.R. (1980) J.Cell Sci. 46, 365-386.
200. Buckler, J., White, A.J., Humphrey, G.W. and Pigiet, V. (1980) Cell 22, 37-46.
201. Barrack, E.R. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 247-258, Alan Liss inc., N.Y.
202. Clark, J.M. and Markaverich, E.A. (1982) in: The nuclear envelope and the nuclear matrix (G.Maul, ed.) pp 259-270, Alan Liss inc., N.Y.
203. Barrack, E.R. and Coffey, D.S. (1979) J.Biol.Chem. 255, 7265-7271.
204. Jackson, D.A., McCready, S.J. and Cook, P.R. (1981) Nature 292, 552-555.
205. Jackson, D.A., Caton, A.J., McCready, S.J. and Cook, P.R. (1982) Nature 296, 366-368.
206. Faiferman, I. and Pogo, A.O. (1975) Biochemistry 14, 3808-3816.
207. Herman, R., Weymouth, L. and Penman, S. (1978) J.Cell Biol. 78, 663-674.
208. Miller, T.E., Huang, C-Y and Pogo, A.O. (1979) J.Cell Biol. 76, 675-691.
209. Herlan, G., Eckert, W.A., Kaffenberger, W and Wunderlich, F. (1979) Biochemistry 18, 1782-1787.
210. Shine, J. and Dalgarno, L. (1974) Proc.Natl.Acad.Sci.USA 71, 1342-1346.
211. Shinagawa , M. and Padmanabhan, R. (1980) Proc.Natl.Acad. Sci.USA 77, 3831-3835.
212. Desiderio, S.V. and Kelley, T.J. (1981) J.Mol.Biol. 145, 319-337.
213. Lichy, J.H., Field, J., Horwitz, M.S. and Hurwitz, J. (1982) Proc.Natl.Acad.Sci.USA 79, 5225-5229.
214. Kitchingman, G.R. (1982) Gene 20, 205-210.
215. Philipson, L. and Petters son, U.(1980) Trends in Biol.Sci. 5, 135-138.
216. Nevins, J.R., Ginsberg, H.S., Blanchard, J-M., Wilson, M.C. and Darnell, J.E. (1979) J.Virol. 32, 727-733.
217. Jones, N. and Shenk, T. (1979) Proc.Natl.Acad.Sci.USA 76, 3665-3669.

218. Berk, A.J.,Lee, F., Harrison, T., Williams, J. and Sharp, P.A. (1979) Cell 17, 935-944.
219. Nevins, J.R., Jensen-Winkler, J. (1980) Proc.Natl. Acad.Sci.USA 77, 1893-1897.
220. Foster, D.A., Spangler, R. and Zubay, G. (1981) J.Virol. 37, 493-495.
221. Esche, H., Mathews, M.B. and Lewis, J.B. (1980) J.Mol.Biol. 142, 399-417.
222. van der Eb, A.J., van Ormondt, H., Schrier, P.I., Lupker, J.H., Jochemsen, H., van de Elsen, P.J., Deleys, R.J., Maat, J., van Beveren, C.P., Dykema, R. and de Waard, A. (1980) Cold Spring Harbor Symp. Quant.Biol. XLIV, 383-399.
223. Shiroki, K., Shimojo, H., Sawada, Y., Uemizu, Y. and Fujinaga, K. (1979), Virology 95, 127-136.
224. Klein, H., Maltzman, W. and Levine, A.J. (1979) J.Biol. Chem. 254, 11051-11060.
225. Vennström,B., Persson, H., Pettersson, U. and Philipson, L. (1979) Nucleic Acids Res. 7, 1405-1418.
226. Beltz, G.A. and Flint, S.J. (1979) J.Mol.Biol. 131, 353-373.
227. Goldenberg, C.J., Rosenthal, R., Bhaduri, S. and Raskas, H. (1981) J.Virol. 38, 932-939.
228. Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475.
229. Fraser, N.W., Nevins, J.R., Ziff, E. and Darnell, J.E. (1979) J.Mol.Biol. 129, 643-656.
230. Evans, R., Weber, J., Ziff, E. and Darnell, J.E. (1979) Nature 278, 367-370.
231. Fraser, N.W., Seghal, P.B. and Darnell, J.E. (1979) Proc. Natl.Acad.Sci.USA 76, 2571-2575.
232. Shaw, A.R. and Ziff, E.B. (1980) Cell 22, 905-916.
233. Nevins, J.R. (1979) J.Mol.Biol. 130, 493-506.
234. Akusjärvi, G. and Persson, H. (1981) Nature 292, 420-426.
235. Nevins, J.R. and Wilson, M.C. (1981) Nature 290, 113-118.
236. Chow, L.T., Broker, T.R. and Lewis, J.B. (1979) J.Mol.Biol. 134, 265-303.
237. Virtanen, A., Aleström, P., Persson, H., Katze, M.G. and Pettersson, U. (1982) Nucleic Acids Res. 10, 2539-2548.
238. Chow, L.T. and Broker, T.R. (1978) Cell 15, 497-510.
239. Akusjärvi, G. Mathews, M.B., Andersson, P., Vennström, B. and Pettersson, U. (1980) Proc.Natl.Acad.Sci.USA 77, 2424-2428.
240. Monstein, H-J. and Philipson, L. (1981) Nucleic Acids Res. 9, 4239-4250.
241. Baum, S.G. and Fox, R.I. (1974) Cold Spring Harbor Symp. Quant. Biol. XXXIX, 567-573.
242. Zehner, Z.E. and Paterson, B.M. (1983) Proc.Natl.Acad. Sci.USA 80, 911-915.
243. Shine, J. and Dalgarno, L. (1974) Proc.Natl.Acad.Sci.USA 71, 1342-1346.

CHAPTER II

ALTERNATIVE SPLICING PATHWAYS EXIST IN THE FORMATION OF ADENOVIRAL LATE MESSENGER RNAs

Alternative splicing pathways exist in the formation of adenoviral late messengers RNAs
E.C.M. Mariman, R.J. van Beek-Reinders and W.J van Venrooij (1983) J.Mol.Biol. 163, 239-256.

J Mol Biol (1983) **163**, 239–256

Alternative Splicing Pathways Exist in the Formation of Adenoviral Late Messenger RNAs

EDWIN C M MARIMAN, RITA J VAN BEEK REINDERS
AND
WALTHER J VAN VENROOIJ

Department of Biochemistry
University of Nijmegen, Geert Grooteplein Noord 21
6525 EZ Nijmegen, The Netherlands

(Received 26 July 1982, and in revised form 21 September 1982)

Total rapidly labeled RNA was extracted from nuclei of HeLa cells late after infection with adenovirus type 2 Most of this nuclear RNA is transcribed from the major late transcription unit (16 2 to 100·0 map units (m u)) To study the cleavage reactions that are involved in RNA splicing we used the S_1 nuclease mapping technique with *Hin*dIII B (16 8 to 31 5 m u) and *Xho*I F (15 5 to 22 4 m u) restriction fragments as DNA probes The S_1 mapping data showed that both intron 1 (16 3 to 19 1 m u) and intron 2 (19 4 to 26 2 m u) can be excised in more than one step Kinetic labeling and pulse–chase experiments demonstrated that certain cleavage sites in the RNA are used within three minutes after the start of transcription, while other cleavages occur only after a significant time lag The use of 5,6-dichloro 1 β D ribofuranosylbenzimidazole enabled us to label the nuclear RNA exclusively from the 5′ end Such an approach showed that the first cleavages are introduced in the nascent RNA before the RNA polymerase has passed more than 2000 nucleotides beyond the cleavage site The chronology of the appearance of processing intermediates that results from cleavage of the RNA indicates that preferentially intron 1 is removed before intron 2 This is in agreement with the finding that leader 1 is ligated to leader 2 before ligation of leader 2 to leader 3 takes place However, we have found that alternative splicing pathways exist in the excision of introns 1 and 2, demonstrating that cleavage in intron 2 may occur before intron 1 is attacked by the splicing enzymes

1. Introduction

Eukaryotic messenger RNAs are usually derived from large primary transcripts Several post transcriptional modifications are needed to transform the primary transcript into a mature messenger One of these modifications is RNA splicing (Sharp, 1981, Nevins, 1982) Internal nucleotide sequences are removed from the transcript in a process that involves cleavage of the RNA chain at the 5′ and 3′ ends of an intron, followed by ligation of the external pieces of RNA In this way an RNA molecule is formed that is non colinear with the DNA Not much is known

0022-2836/83/020239-18 $03 00/0

about the way in which the intervening sequences are removed. Although it is assumed that introns can be removed in one piece, it has been reported that removal of an intron can be a stepwise process (Berget & Sharp, 1979: Avvedimento *et al.*, 1980: Grosveld *et al.*, 1981). In this case the intervening sequences are removed in parts. When a transcript contains more than one intron, the introns may be excised in an orderly fashion (Goldenberg & Raskas, 1979: Tsai *et al.*, 1980: Grosveld *et al.*, 1981; Keohavong *et al.*, 1982).

Late after infection of human cells with adenovirus type 2, mRNAs are produced that mainly code for viral coat proteins. Most of these messengers are derived from the major late transcription unit, which is located between 16·2 and 100·0 map units on the viral genome (the genome is divided in 100 m.u.†). All messengers that are transcribed from the major late transcription unit have the same three leaders spliced to the 5′ end of their body sequences (Fig. 1(a); Gelinas & Roberts, 1977; Chow *et al.*, 1977). Using heteroduplex and S_1 nuclease mapping, Berget & Sharp (1979) have suggested that splicing between the leaders occurs in an orderly fashion. The intron between leader 1 and leader 2 is removed before the intron that separates leader 2 from leader 3 is excised. Recently this was confirmed by Keohavong *et al.* (1982), who found that ligation of leader 1 to leader 2 occurs before ligation of leader 2 to leader 3.

We have used the S_1 nuclease mapping procedure to study the cleavage reactions that are involved in the splicing of the late leader sequences. Our S_1 mapping data indicate that there are alternative splicing pathways for the excision of the first and second intron from RNAs that are derived from the major late transcription unit. Although the kinetic analysis of the cleavage reactions is in agreement with the chronology of the ligation of the leaders (Keohavong *et al.*, 1982), we found that cleavage can occur in the second intron before the first intron is attacked.

2. Materials and Methods

(a) *Infection and labeling of cells*

HeLa S3 cells were infected with adenovirus type 2 (2000 particles per cell) at a density of 5×10^6 cells/ml in Eagles minimal essential medium (MEM) containing 1 mM-arginine. After adsorption of the virus to the cells during 1 h at 37°C, the cells were diluted to $0{\cdot}3 \times 10^6$ cells/ml with MEM containing 5% newborn calf serum. Cells were harvested at 18 h post-infection. To obtain rapidly labeled nuclear RNA 18 h after infection, cells were concentrated to 20×10^6 cells/ml and incubated with 20 μCi of [5-^{3}H]cytidine/ml (31 Ci/mmol; Amersham, England), 20 μCi of [5′-^{3}H]guanosine/ml (22 Ci/mmol; Amersham, England) and 20 μCi of [5,6-^{3}H]uridine/ml (45 Ci/mmol; Amersham, England). The labeling times were as indicated in the text. The pulse–chase experiment was performed without guanosine in the labeling mixture.

(b) *Preparation of nuclear RNA*

At 18 h after infection cells were harvested on frozen Na/K/Mg solution (0·13 M-NaCl, 0·05 M-KCl, 1·5 mM-$MgCl_2$) and centrifuged for 5 min at 800 *g*. The cells were washed once with Na/K/Mg solution and after pelleting they were resuspended in reticulocyte suspension

† Abbreviations used: m.u., map units; DRB, 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole; Ad2, adenovirus type 2.

buffer (RSB 10 mM NaCl 10 mM Tris pH 7.4 1.5 mM $MgCl_2$) After addition of 0.1 vol of a sodium deoxycholate/Nonidet P40 mixture (5% each) the cells were disrupted by intensive vortexing The nuclei were pelleted washed once with RSB and resuspended in the same buffer Nuclear RNA was then isolated by extraction with phenol at 55°C as described by Long *et al* (1979)

(c) *Isolation of cloned restriction fragments*

The *Aho*I F fragment of Ad2 DNA was ligated in the *Sal*I site of pBR322 pBR322 containing the *Hin*dIII B restriction fragment in the *Hin*dIII site was a generous gift from R Ohlsson Plasmids were grown in *Escherichia coli* strain JA221 Extraction and purification of plasmid DNA were performed as described by Alestrom *et al* (1980) After isolation the purified plasmids were linearized by digestion with *Eco*RI

(d) *Restriction endonuclease digestions*

All restriction enzymes were purchased from Bethesda Research Laboratories Inc Digestions were performed at 37°C for 4 h using 1 unit of enzyme per μg of DNA Standard conditions for *Eco*RI and *Hin*dIII were 50 mM Tris (pH 7.4), 7 mM $MgCl_2$, 50 mM NaCl, 3 mM dithiothreitol Digestion conditions for *Sst*I and *Bam*HI were 20 mM Tris (pH 7.4), 6 mM $MgCl_2$, 90 mM NaCl 6 mM dithiothreitol *Kpn*I digestions were performed under low salt conditions (6 mM Tris pH 7.4 6 mM $MgCl_2$, 6 mM NaCl, 6 mM dithiothreitol) Bovine serum albumin was added to the digestion mixtures to a final concentration of 1 mg/ml Restriction fragments were separated by agarose gel electrophoresis 0.8% agarose gels were run for 20 h at 80 V in Tris/Na/EDTA buffer (40 mM Tris, pH 8.0, 8 mM sodium acetate 2 mM EDTA) The DNA was recovered from agarose gel slices by electrophoretic elution

(e) *S_1 nuclease mapping and analysis of S_1 resistant hybrids*

RNA was hybridized to DNA under conditions in which DNA DNA hybridization was minimized using the procedure described by Manley *et al* (1979) A sample (1 μg) of nuclear RNA was hybridized to 2.5 μg equivalent of a restriction fragment of Ad2 DNA Hybridization concentration was 0.35 μg RNA/μl After S_1 nuclease digestion the hybrids were precipitated with ethanol, the pellet was solubilized in Tris/Na/EDTA buffer and applied to a 1.2% agarose gel Electrophoresis was performed under the same conditions as were used for the separation of DNA restriction fragments Gels were subsequently treated with 5% trichloroacetic acid methanol, methanol containing 10% PPO, and water After drying they were exposed to Kodak XR 5 film for an appropriate period of time (3 to 7 days)

3. Results

(a) *S_1 nuclease mapping using the* Hind*III B restriction fragment*

The *Hin*dIII B restriction fragment of Ad2 DNA (16.8 to 31.5 m.u.) covers the second and third leader sequences as well as part of the first, the second and (part of) the third intervening sequences (Fig 1(a)) Also the 5′ ends of the larger L1 messengers are located in this region (Akusjärvi & Persson, 1981) Therefore, we considered the *Hin*dIII B fragment to be a useful DNA probe to study RNA splicing between the late leaders via the S_1 mapping procedure. Rapidly labeled nuclear RNA was extracted from HeLa cells late after infection with Ad2 and hybridized to the *Hin*dIII B fragment When the S_1-resistant hybrids were resolved on agarose gel as described in Materials and Methods, a complex pattern of

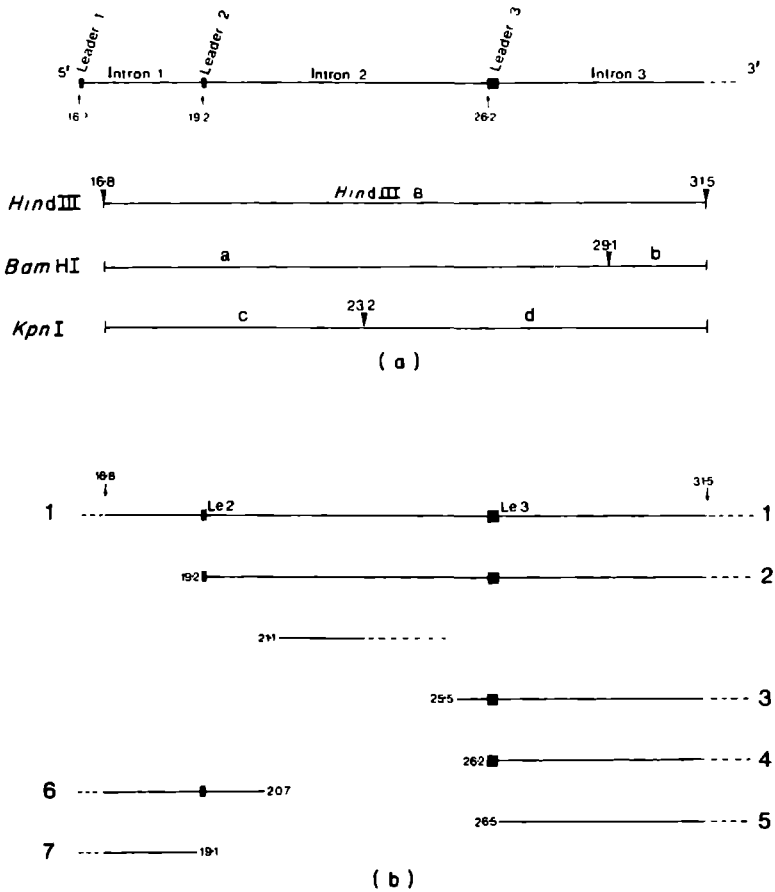


FIG. 1 (a) Position of the 3 leader sequences and intervening sequences of major late adenoviral RNAs on the viral genome. The genome is divided into 100 m.u. The exact positions of the leaders, as indicated, are deduced from the nucleotide sequence of the *Hin*dIII B restriction fragment (Peter Alestrom, personal communication). Also shown are the restriction fragments that were used as DNA probes in S_1 mapping experiments. (b) Summary of the S_1 mapping data that were obtained via S_1 analysis of rapidly labeled nuclear RNA using the restriction fragments depicted in (a) as DNA probes. The RNAs have been numbered (1 to 7) for reference. (▬▬) Leader (Le) sequence, (———) intervening sequence.

hybrids was obtained. This was to be expected, according to the findings of Berget & Sharp (1979), who have detected a complex population of RNA molecules that map with their 5′ end in this region of the genome and have their 3′ end to the right of co-ordinate 31·5. These RNA species were assumed to be processing intermediates in the formation of adenoviral late messengers. We have used subfragments of *Hin*dIII B (Fig. 1(a)) to identify the most prominent labeled RNA species and map their 5′ end in this region. Figure 2 shows the result of such an analysis in which 15-minute-labeled late nuclear RNA was hybridized to *Hin*dIII B (lane 1) and to the subfragments (lanes 2 to 5). Hybrids in lane 1 are

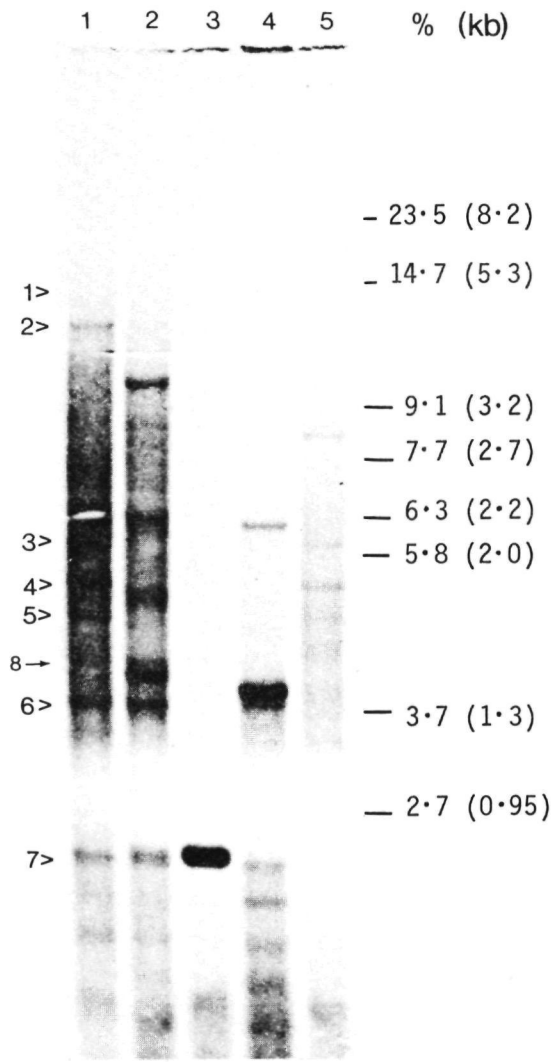


FIG. 2. S_1 nuclease analysis of total nuclear RNA in the region of *Hin*dIII B. HeLa cells 18 h after infection were labeled for 15 min with ^{3}H-labeled nucleosides. After labeling total nuclear RNA was prepared as described in Materials and Methods. The RNA (2×10^6 cts/min) was hybridized to various DNA fragments (Fig. 1(a)) and analyzed according to the S_1 mapping procedure (Materials and Methods). The Figure shows the S_1-resistant hybrids of nuclear RNA with: lane 1, *Hin*dIII B; lane 2, subfragment a; lane 3, subfragment b; lane 4, subfragment c; lane 5, subfragment d. *Eco*RI and *Hin*dIII digests of Ad2 DNA were used as molecular weight markers. The positions of the *Hin*dIII markers are indicated. The hybrid bands in lane 1 have been numbered for reference. kb, 10^3 base-pairs.

numbered for reference. Band number 1 (14·6%; 1% equals 360 base-pairs) was visible only after prolonged exposure times. This band represents the primary transcript molecules that extend beyond both ends of *Hin*dIII B and, therefore, protect the DNA fragment totally against degradation by the S_1 nuclease. A 12·3% hybrid band (number 2) is formed from RNA species that have their 5′ ends localized at 19·2 m.u. In agreement, hybrids of 9·9% or 4·0% were found when subfragments a or c were used (lanes 2 and 4). These hybrids represent processing intermediates from which intron 1 has been removed (Berget & Sharp, 1979; Manley *et al.*, 1979). Hybrid bands 3, 4 and 5, with lengths of 6·0, 5·3 and 5·0%, respectively, were detected with *Hin*dIII B (lane 1) as well as with fragment d (lane 5) but not with any of the other fragments. These colinear pieces of RNA, therefore, have their 5′ ends at co-ordinates 25·5, 26·2 and 26·5, respectively, and extend to the right of 31·5 m.u. They were described by Berget & Sharp (1979), who identified six presumptive processing intermediates mapping in the region between 24·8 and 26·9 m.u. With subfragment c a band of 2·1% was found (lane 4), which could not be detected with any of the other fragments. The RNA in this hybrid maps with its 5′ end at co-ordinate 21·1 and extends beyond 23·2 m.u.

Except for RNA molecules that extend beyond 31·5 m.u. in this experiment, we identified two species that seemed to extend to the left of 16·8 m.u. Bands 6 and 7, 3·9 and 2·3% in length, were found when *Hin*dIII B or subfragments a or c were used to hybridize (lanes 1, 2 and 4). These RNA species extend from the left of 16·8 m.u. to co-ordinates 20·7 and 19·1, respectively. In further experiments we identified a third species that extends beyond 16·8 m.u. It is indicated by an arrow and numbered 8 for reference. From the data presented in this section it cannot be concluded that these three RNAs are derived from the major late promoter. Their identity will be discussed further, below.

Figure 1(b) summarizes the S_1 mapping data and shows the RNA species that were identified in the rapidly labeled total nuclear RNA. The positions of the leaders, as indicated, are deduced from the nucleotide sequence of the *Hin*dIII B restriction fragment (Peter Alestrom, personal communication).

(b) *Kinetics of the cleavage reactions as part of RNA splicing*

To study the formation of the RNA species a kinetic labeling experiment was performed in the following manner. Four portions of 18-hour-infected HeLa cells (100×10^6 cells each) were labeled with ^{3}H-labeled nucleosides for 3, 7, 15 and 30 minutes, respectively (see Materials and Methods). After labeling, nuclear RNA was prepared and analyzed via the S_1 nuclease procedure. From each portion of cells 2×10^6 cts/min of nuclear RNA were hybridized to *Hin*dIII B. Thus, the pattern of S_1-resistant hybrids (Fig. 3) shows the relative accumulation of label in the individual RNA species. After three minutes of labeling RNAs 1 and 2 were already labeled, while no label could be detected in RNA 4 (Fig. 3, lane 1). Only after 15 minutes of labeling did radioactivity begin to accumulate in RNA 4 (Fig. 3, lane 3), which became the most predominant band after a labeling time of 30 minutes (Fig. 3, lane 4) or longer. The method of analysis that we have used can detect only where the ends of a colinear RNA chain are located. Our results show

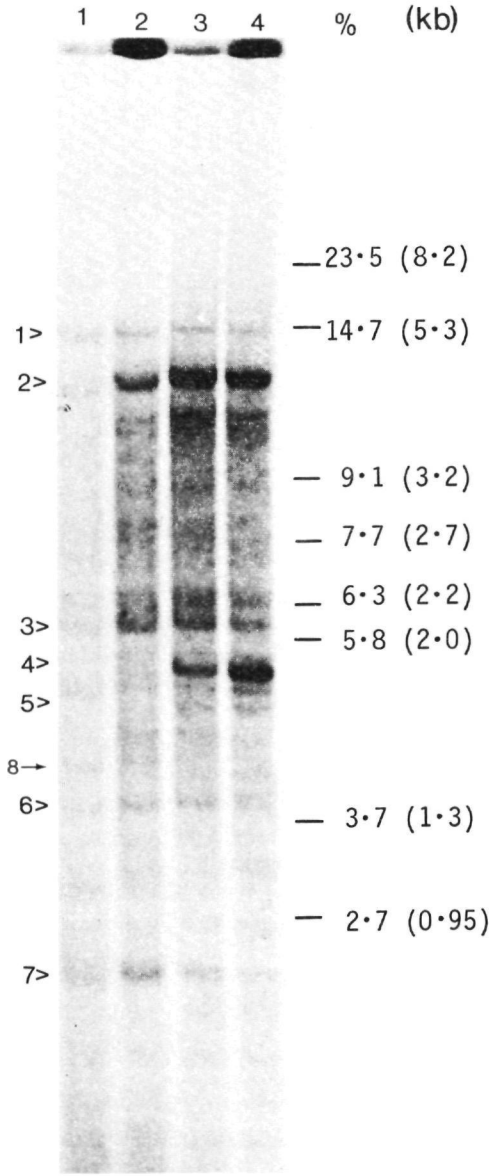


FIG. 3. S_1 nuclease analysis of nuclear RNA labeled for various periods of time. HeLa cells 18 h after infection were labeled for 3, 7, 15 or 30 min with ^{3}H-labeled nucleosides. After labeling total nuclear RNA was prepared as described in Materials and Methods. The RNA (2×10^6 cts/min) was hybridized to the *Hin*dIII B restriction fragment and analyzed according to the S_1 mapping procedure (Materials and Methods). The Figure shows the S_1-resistant hybrids. *Eco*RI and *Hin*dIII digests of Ad2 DNA were used as molecular weight markers. The positions of the *Hin*dIII markers are indicated. Lane 1, 3 min labeling; lane 2, 7 min labeling; lane 3, 15 min labeling; lane 4, 30 min labeling. The hybrid bands have been numbered for reference.

that within three minutes after the start of transcription the RNA is cleaved at the acceptor site between intron 1 and leader 2 (19·2 m.u.), while it takes almost 15 minutes before cleavage occurs at the acceptor site between intron 2 and leader 3 (26·2 m.u.). In all labeling experiments that we have performed RNA 3 appeared also to accumulate label within three minutes (Fig. 3, lane 1). From this we concluded that intron 2 may be rapidly cleaved at 25·5 m.u. Therefore, it seems likely that intron 2 can be removed in at least two steps, in which RNA 3 is created as an early intermediate of this splicing event.

(c) *Pulse–chase analysis of nuclear RNA*

We were interested to know if RNA 4 is derived from the primary transcript or from a rapidly formed processing intermediate. Therefore, pulse-chase type experiments were performed. HeLa cells 18 hours after infection were pulse-labeled (5 min) with ^{3}H-labeled nucleosides (see Materials and Methods). After labeling, the cells were rapidly cooled to 0 °C, pelleted and suspended in medium containing 10 mM-glucosamine and 0·1 mg of cytidine and uridine/ml (Scholtissek, 1971). This procedure is known to reduce the uridine pool in the nucleus dramatically. After a chase of 5, 10, 20 or 30 minutes at 37°C, samples containing 100×10^6 HeLa cells were collected, cooled rapidly to 0°C and nuclear RNA was prepared as described in Materials and Methods. By this procedure accumulation of label was stopped in less than ten minutes of chasing. Nuclear RNA (2×10^6 cts/min) was hybridized to *Hin*dIII B. The S_1-resistant hybrids, as shown in Figure 4, demonstrated that after a five-minute chase RNAs 1, 2 and 3 were labeled (lane 1). The primary transcript RNA 1 contained only a very low amount of label while RNA 2 was the most predominant band. This agrees with our findings that the primary transcript is very rapidly processed by cleavage at 19·2 m.u. to form RNA 2. Therefore, RNA 2 accumulates rather than RNA 1. RNA 4 began to accumulate label after a 10-minute chase and became the most predominant band after a 30-minute chasing period. On the score of the relative intensities of the hybrid bands we concluded that RNA 4 is a processing product of RNA 2. The primary transcript is cleaved at 19·2 and 26·2 m.u. in the 5′ to 3′ direction. These findings are in full agreement with the results of Keohavong *et al.* (1982), who found that ligation of leader 1 to leader 2 precedes ligation of leader 2 to leader 3. Removal of intron 1 seems to occur before removal of intron 2

(d) *S_1 nuclease mapping using the* Xho*I F restriction fragment*

As can be seen from Figure 3 (lane 1) RNAs 6, 7 and 8 are labeled within three minutes. To find out if these three species were derived from the major late transcription unit, we have performed S_1 mapping using the *Xho*I F restriction fragment (15·5 to 22·4 m.u.) as a probe for hybridization. However, for accurate mapping 20-minute-labeled late nuclear RNA was not only hybridized to full-length *Xho*I F (Fig. 5, lane 1), but also to the DNA fragment after its digestion with *Sst*I (Fig. 5, lane 2) or with *Hin*dIII (Fig. 5, lane 3 for restriction sites see Fig. 6(a)). To identify RNA species 6, 7 and 8 the RNA was hybridized to *Hin*dIII B as well and scored for S_1-resistant hybrids (Fig. 5, lane 4).

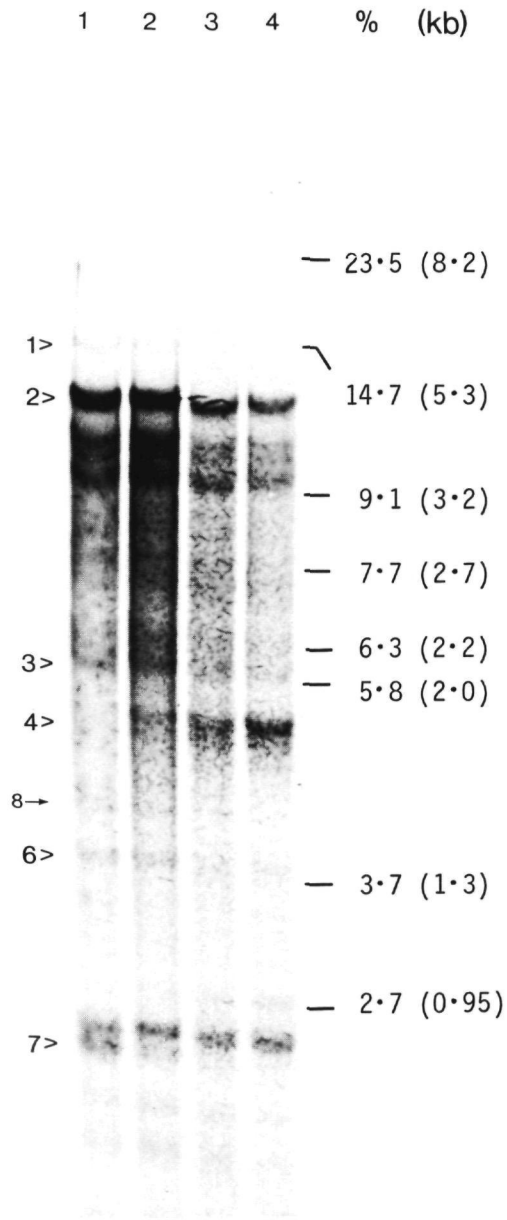


FIG. 4. Pulse–chase analysis of nuclear RNA. HeLa cells 18 h after infection were labeled for 5 min with ^{3}H-labeled nucleosides (Materials and Methods). After labeling the cells were rapidly cooled to 0°C, pelleted and resuspended in chase medium as described in Results, section (c). Chasing was performed for 5, 10, 20 and 30 min, respectively. After chasing total nuclear RNA was extracted and analyzed according to the S_1 mapping procedure (Materials and Methods) using *Hin*dIII B as DNA probe. The Figure shows the S_1-resistant hybrids. *Eco*RI and *Hin*dIII digests of Ad2 DNA were used as molecular weight markers. The positions of the *Hin*dIII markers are indicated. Lane 1, 5 min chasing; lane 2, 10 min chasing; lane 3, 20 min chasing; lane 4, 30 min chasing. The hybrids have been numbered for reference.

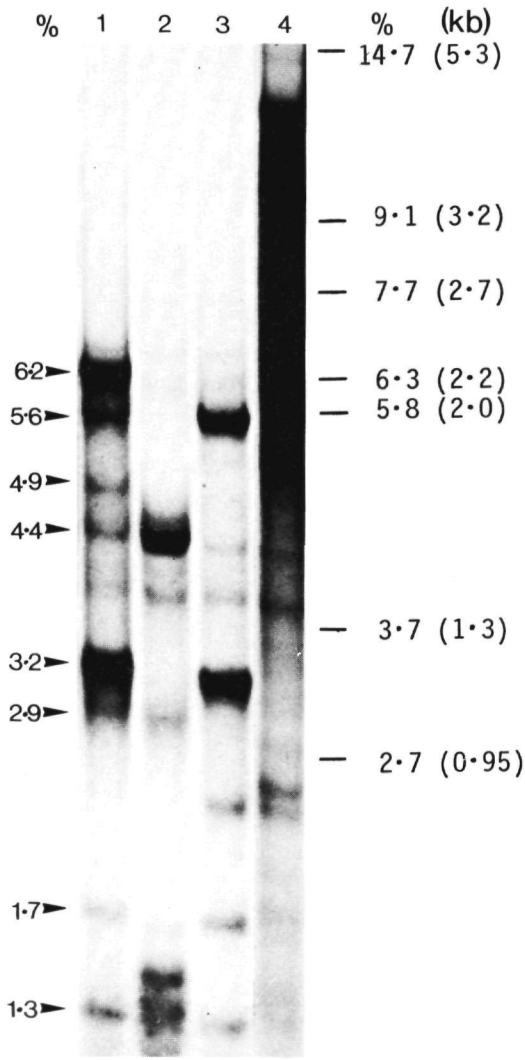


FIG. 5. S_1 nuclease analysis of total nuclear RNA in the region of *Xho*I F. HeLa cells 18 h after infection were labeled for 20 min with ^{3}H-labeled nucleosides. After labeling total nuclear RNA was prepared as described in Materials and Methods. The RNA (2×10^6 cts/min) was hybridized to *Xho*I F and to the same restriction fragment after its digestion with *Sst*I or *Hin*dIII (Fig. 6(a)). Further analysis was performed according to the S_1 mapping procedure (Materials and Methods). The Figure shows the S_1-resistant hybrids of nuclear RNA with: lane 1, *Xho*I F; lane 2, *Sst*I-digested *Xho*I F; lane 3, *Hin*dIII-digested *Xho*I F; lane 4, *Hin*dIII B. *Eco*RI and *Hin*dIII digests of Ad2 DNA were used as molecular weight markers. The positions of the *Hin*dIII markers are indicated. The lengths of hybrids in lane 1 are indicated for reference.

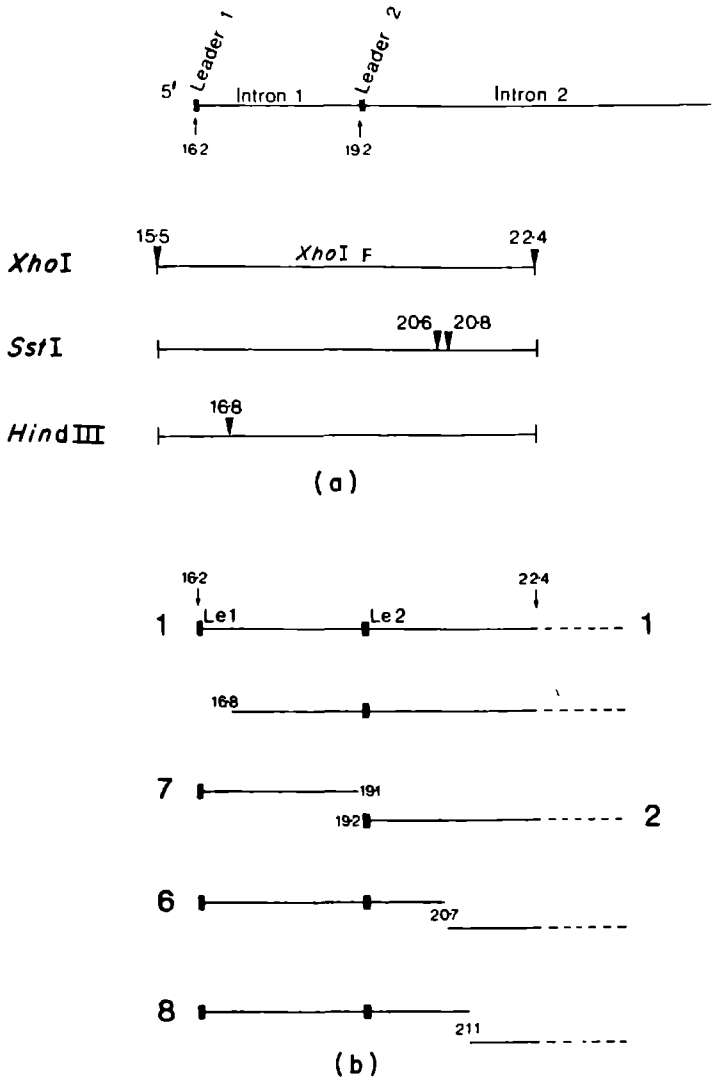


FIG. 6 (a) Position of the leader sequences and intervening sequences of major late adenoviral RNAs on the viral genome. Also shown are the position of the *Xho*I F restriction fragment and the location of the cleavage sites in this fragment for the restriction enzymes that were used in the S_1 mapping experiments. (b) Summary of the S_1 mapping data that were obtained via S_1 analysis of rapidly labeled nuclear RNA using the *Xho*I F fragment and digested *Xho*I F DNA as probes for hybridization (a). The RNAs have been numbered for reference. Le, leader.

With *Xho*I F a 6·2% band was found (lane 1), which represents the primary transcript extending from 16·2 to beyond 22·4 m u. Similarly, a hybrid of 4·4% was found when the *Sst*I-digested fragment was used (lane 2). The most predominant band was 3·2% long (lane 1). This hybrid was also found with the *Hin*dIII-digested

fragment (lane 3), but not with the *Sst*I-digested fragment (lane 2), in which case a predominant 1·4% hybrid was detected. The 5′ end of this RNA maps at 19·2 m.u. and extends beyond 22·4 m.u. It was earlier identified as RNA 2 (Fig. 1(b)). Two smaller bands 1·7% and 1·3% in length, respectively, were detected with *Xho*I F (lane 1). Both bands were also present in the hybrid pattern obtained with the *Hin*dIII-digested fragment (lane 3). The 1·7% hybrid was not found with the *Sst*I-digested probe (lane 2). A slightly smaller band of about 1·6% was probably derived from RNA that hybridized to the small *Sst*I-*Xho*I fragment (20·8 to 22·4 m.u.). In this case, however, we cannot explain why this band was only faint instead of predominant as would be expected. The RNA in the 1·7% hybrid maps with its 5′ end at 20·7 m.u. and extends beyond 22·4 m.u. This was confirmed by the presence of a weak 2·5% long hybrid in the pattern obtained with subfragment c (Fig. 2, lane 4). When subfragment c was used as a probe for hybridization we identified an RNA species with its 5′ end at co-ordinate 21·1 (see section (a), above). This species corresponded to the 1·3% hybrid found with *Xho*I F (lane 1) as well as with the digested restriction fragments (lanes 2 and 3). A band of 5·6% was detected with *Xho*I F (lane 1), while a 3·8% long hybrid was found with the *Sst*I-digested fragment (lane 2). These hybrids were formed from RNA molecules that map with their 5′ end at 16·8 m.u. and extend beyond co-ordinate 22·4.

Except for the RNAs that have their 3′ end to the right of 22·4 m.u., we have mapped the RNA molecules that extend beyond 16·8 m.u. A 4·9% long hybrid was found with *Xho*I F (lane 1), but not with the *Sst*I-digested fragment (lane 2). It corresponds to the hybrid of 4·3%, which was detected with the *Hin*dIII-digested fragment (lane 3) and was earlier numbered 8 for reference (Fig. 3, lane 1; Fig. 4, lane 1). This RNA extends from the promotor region at 16·2 m.u. to co-ordinate 21.1.

RNA 6 gave rise to a hybrid of 3·9% with *Hin*dIII B (lane 4) and was also detected with the *Hin*dIII-digested *Xho*I F fragment (lane 3). In accordance with the supposition that this RNA extends from 16·2 to 20·7 m.u., a hybrid of length 4·4% was found with the *Xho*I F fragment (lane 1).

RNA 7, which was found with *Hin*dIII B (lane 4) and with the *Hin*dIII-digested *Xho*I F fragment (lane 3), was earlier mapped with one end at 19·1 m.u. and extending to the left of 16·8 m.u. (see section (a), above). This appeared to be correct since a band of 2·9% was found when *Xho*I F or the *Sst*I-digested fragment was used as DNA probe (lanes 1 and 2). Thus this RNA maps between 16·2 and 19·1 m.u. Since this RNA could represent the spliced-out intron 1, we have verified the presence of leader 1 in this RNA by separating the hybrids on agarose gel and transferring them to nitrocellulose. The blot was incubated with ^{32}P-labeled complementary DNA of the fiber messenger (Zain *et al.*, 1979*b*), a generous gift from M. Mathews. After incubation and washing, the 2·9% hybrid appeared as a ^{32}P-labeled band (data not shown). Thus, RNA 7 contains leader 1 sequences.

Figure 6(b) summarizes the S_1 mapping data described above. Two conclusions can be drawn. First, there are two cleavage sites in intron 1 at 16·8 and 19·1 m.u., respectively, demonstrating that excision of intron 1 may be a stepwise process. Second, RNAs 6, 7 and 8 map with their 5′ end at the start site of transcription (16·2 m.u.). Most striking, however, is the fact that their 3′ ends map at locations

where other colinear RNAs have their 5′ end. It is, therefore, reasonable to assume that pairs of these RNAs (e.g. RNAs 2 and 7) are created by the cleavage of the primary transcript at 19·1, 20·7, or 21·1 m.u., respectively. This finding also makes it unlikely that RNAs 6, 7 and 8 might be the result of premature termination of transcription, a phenomenon that has been reported to occur in the major late transcription unit (Evans *et al.*, 1979); for in that case one should not expect to detect the 3′ RNA pieces. Furthermore, we have labeled HeLa cells at the late stage of infection in the presence of DRB for 20 minutes. When the nuclear RNA was extracted and analysed according to the S_1 mapping procedure the RNAs 6, 7 and 8 were not detectable as labeled species. DRB seems to inhibit the formation of these three RNAs to the same extent as the synthesis of promoter-distal RNA sequences. Since products of premature termination are continued to be formed in the presence of DRB (Fraser *et al.*, 1978,1979) this indicates that RNAs 6, 7 and 8 are not the result of premature termination but rather of cleavage activities.

(e) *Analysis of nascent RNA using DRB-treated cells*

The kinetic labeling experiment described in section (b), above, showed that after three minutes of labeling both RNAs 2 and 7 were already labeled (Fig. 3, lane 1). Thus, the primary transcript is cleaved at 19·1 m.u. within three minutes after the start of transcription. To obtain more information about the moment of cleavage of the RNA we have performed a kinetic labeling experiment in which the RNA was labeled from the 5′ end. Therefore, HeLa cells were treated with 75 μM-DRB (Fraser *et al.*, 1978; Tamm *et al.*, 1980; Cheng-Kiang *et al.*, 1982) from 17 to 18 hours after infection. DRB allows initiation of transcription to take place, but inhibits the elongation of RNA synthesis. At 18 hours after infection the cells were rapidly cooled to 0°C, pelleted, washed in Na/K/Mg solution and resuspended in medium to label the cells as described in Materials and Methods. Cells were labeled for 5, 10, 20 and 45 minutes, respectively. The nuclear RNA was extracted and analyzed via the S_1 nuclease procedure. The RNA (2×10^6 cts/min) was hybridized to the *Hin*dIII B restriction fragment and S_1-resistant hybrids were scored as shown in Figure 7. The effect of DRB seemed not to be reversed instantly upon its removal, since after a labeling time of five minutes hardly any labeled RNA 1 or 2 could be detected. Elongation of transcription seemed to be slowed down, possibly by the presence of a small residual amount of DRB in the cells. However, RNAs 6, 7 and 8 were clearly labeled after five minutes (lane 1). This indicates that cleavage can occur in the nascent transcript, preferentially at 19·1, but also at 20·7 and 21·1 m.u. The RNA was cleaved before transcription had passed beyond co-ordinate 31·5. As soon as transcription had passed beyond this co-ordinate RNA 3 appeared as a labeled species in the hybrid pattern (lane 2). Again this suggests that the nascent RNA transcript is cleaved at 25·5 m.u. as soon as the RNA polymerase has passed beyond this co-ordinate. Only after 20 minutes of labeling did RNA 4 appear (lane 3), in agreement with the relatively slow cleavage of the RNA at 26·2 m.u. It should be noted, however, that the altered rate of transcription in DRB-treated cells may lead to the accumulation of nascent transcripts and a higher efficiency of cleavage as compared to non-treated cells.

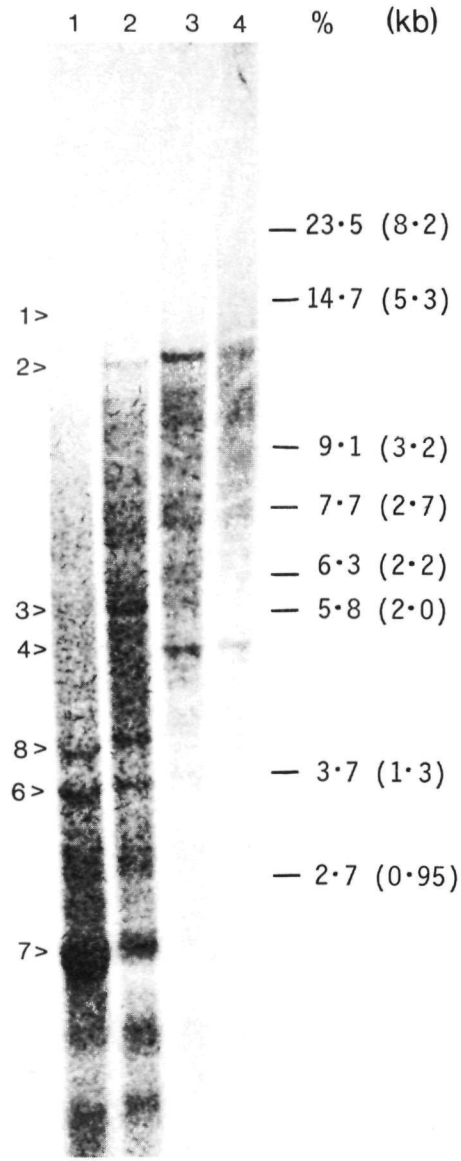


FIG. 7. S_1 analysis of nascent RNA. HeLa cells were treated with DRB (75 μM) from 17 to 18 h after infection. Thereafter the cells were rapidly cooled to 0°C, pelleted, washed with Na/K/Mg solution and resuspended in medium to label the cells with ^{3}H-labeled nucleosides for 5, 10, 20 or 45 min (see Materials and Methods). After labeling total nuclear RNA was prepared as described in Materials and Methods. The RNA (2×10^6 cts/min) was hybridized to the *Hin*dIII B restriction fragment and analyzed according to the S_1 mapping procedure (see Materials and Methods). The Figure shows the S_1-resistant hybrids. *Eco*RI and *Hin*dIII restriction digests of Ad2 DNA were used as molecular weight markers. The positions of the *Hin*dIII markers are indicated. Lane 1, 5 min labeling; lane 2, 10 min labeling; lane 3, 20 min labeling; lane 4, 45 min labeling. The hybrid bands have been numbered for reference.

4. Discussion

Splicing is one of the post-transcriptional modifications that are part of the production process of eukaryotic messengers. Although not much is known about the mechanism of splicing it evidently involves cleavage of the RNA chain at the 5′ and 3′ ends of an intervening sequence, followed by the ligation of the external RNA pieces. Using the S_1 mapping procedure (Manley *et al.*, 1979) we have studied the cleavage reactions that occur during splicing of the tripartite leader sequences in RNAs that are derived from the major late transcription unit of adenovirus type 2 (Gelinas & Roberts, 1977; Chow *et al.*, 1977). We have analyzed pulse-labeled nuclear RNA rather than steady-state RNA. This has enabled us to draw the following conclusions.

(1) Both intron 1 and intron 2 can be spliced out in more than one step.
(2) Certain cleavages are rapidly introduced in the primary transcript, while other cleavage sites are used only after a significant time-lag. In this case the cleavages occur in a processing intermediate rather than in the primary transcript.
(3) The rapid cleavages can occur in the nascent transcript. In such cases the first acts of splicing can occur before polyadenylation or termination of transcription.
(4) Alternative splicing pathways exist in the excision of intron 1 as well as intron 2.

In the case of excision of introns as a whole one should expect to detect only two processing intermediates with the S_1 mapping technique. Removal of intron 1 should lead to an intermediate with its 5′ end at co-ordinate 19·2, while removal of intron 2 should create a colinear RNA with its 5′ end localized at 26·2 m.u. Indeed, both intermediates, RNAs 2 and 4, were detected. Except for these RNAs, which extend from leader 2 or leader 3, other RNA species were found that have their 5′ ends at 16·8 m.u. In intron 2 three species were mapped with their 5′ ends at 20·7, 21·1 and 25·5 m.u., respectively. RNA molecules mapping with their 5′ ends in intron 2 were reported to be present in steady-state nuclear RNA (Berget & Sharp, 1979; Manley *et al.*, 1979) and were regarded as intermediates in a stepwise splicing process. Our results thus indicate that not only intron 2, but also intron 1, can be excised in a stepwise manner. Not all of the intermediates that we have found agree with those first described by Berget & Sharp (1979). This may be ascribed to the fact that we analyzed rapidly labeled RNA instead of steady-state RNA. Rapidly formed intermediates do not necessarily have to correspond to the most abundant species present in nuclear RNA.

As shown by a pulse–chase experiment (Fig. 4), the intermediate that extends from leader 3 (26·2 m.u.) is derived from another processing intermediate extending from leader 2 (19·2 m.u.). Cleavage at 26·2 m.u. occurs at about 15 minutes after the start of transcription. Keohavong *et al.* (1982) have reported that in ligation of leader 2 to leader 3 there is a time-lag of 12 minutes. This suggests that the time-lag is due to the slow cleavage reaction at the 3′ end of intron 2. Cleavage at 26·2 m.u. seems to be one of the last events in splicing of leader 2 to leader 3, followed by rapid

ligation of these exons. Further, they have reported that ligation of leader 1 to leader 2 occurs without detectable delay. Accordingly, we have found that intron 1 is cleaved at its 3′ end (19·1 m.u.) as soon as transcription has proceeded some distance beyond this site. These results indicate that cleavage at the 3′ end of an intron may be involved in the regulation of splicing.

Labeling from the 5′ end (Fig. 7) has shown that the primary transcript may be cleaved at certain places while transcription is still proceeding. These cleavage sites are located at 19·1, 20·7, 21·1 and 25·5 m.u., suggesting that the RNA polymerase has added not more than 2000 nucleotides to the growing RNA chain before it is cleaved. For the first three cleavage sites we were able to detect both the 5′ and 3′ pieces resulting from cleavage of the primary transcript. In case of cleavage of the primary transcript at co-ordinate 25·5 one should expect to find an RNA molecule that extends from 16·2 to 25·5 m.u. Until now we have not been able to detect such a species, indicating that the cleavage at 25·5 m.u. occurs in a more rapidly produced processing intermediate rather than in the primary transcript. Recent results obtained by Manley *et al.* (1982), who have studied transcription and processing in isolated nuclei of late-infected HeLa cells, also point to the rapid excision of intron 1 from the nascent transcript.

Our S_1 mapping data have demonstrated at least two processing intermediates arising from splicing of intron 1 (Fig. 6(b)). One intermediate has its 5′ end at 16·8 m.u. and extends beyond leader 2. The other preferentially formed intermediate extends from the 5′ end of leader 1 and has its 3′ end at 19·1 m.u. These processing intermediates cannot be derived from one another. Therefore, alternative splicing pathways seem to exist for the excision of intron 1. We have found also that the first cleavage of the primary transcript can occur at either of two locations in intron 2, at 20·7 or 21·1 m.u. Thus, alternative pathways for the excision of intron 2 also seem to exist. It is possible that one splicing pathway is used to introduce the i-leader into mature messengers (Akusjärvi & Persson, 1981).

Our data indicate that the preferred splicing pathway follows the removal of intron 1 before intron 2. However, cleavage may occur in intron 2 before intron 1 is attacked. These findings do not contradict the results of Keohavong *et al.* (1982), since in this case ligation of leader 1 to leader 2 may still precede the ligation of leader 2 to leader 3.

Although our data demonstrate that the primary transcript is preferentially cleaved at 19·1 m.u., there are a small number of molecules (band number 1, Fig. 3) that seem to escape rapid cleavage at a site within the *Hin*dIII B region. It is possible that the first event of splicing in these molecules may occur at the right of 31·5 m.u. Another explanation could be that splicing in these molecules is delayed, for instance, because they miss the correct processing proteins. It is also possible that these unattacked molecules for some reason are not determined or not fit to be processed into mRNA.

Until now we could never identify pieces of RNA that could be ascribed to spliced-out intervening sequences (Mariman *et al.*, 1982). We have demonstrated the presence of leader 1 sequences in RNA 7 (16·2 to 19·1 m.u.). However, since leader 1 is only 41 nucleotides in length (Zain *et al.*, 1979*a*) it cannot be ruled out that the hybrid bands derived from RNA 7 are actually derived from a mixture of

RNA molecules that contain leader 1 (16·2 to 19·1 m u), and spliced-out intron 1 of about the same size (16·3 to 19·1 m u) It is also possible that rapid degradation may prevent the accumulation of spliced-out intervening sequences to a detectable concentration

We have reported the existence of alternative splicing pathways in the formation of the tripartite leader sequences of late adenoviral mRNAs This may be a more general feature of RNA splicing in eukaryotic cells Further studies are directed to determine the splicing pathways in more detail, since this may help to elucidate the mechanism of the regulation of RNA splicing

We thank the collaborators of the department of Microbiology, Biomedical Center Uppsala, in particular Drs L Philipson, R Ohlsson and P Alestrom, for generously providing us with cloned *Hin*dIII B DNA and the nucleotide sequence of this restriction fragment We also thank Dr M Mathews from the Cold Spring Harbor Laboratory for generously providing us with cloned cDNA of fiber mRNA (pJAW43, Zain *et al* , 1979*b*) We are indebted to Dr C van Eekelen for help and advice and A Groeneveld for culturing the cells

This work was supported by the Netherlands Foundation for Chemical Research (S O N) and the Netherlands Organization for the Advancement of Pure Research

REFERENCES

Akusjarvi, G & Persson, H (1981) *Nature (London)* **292**, 420–426
Alestrom, P , Akusjarvi, G , Perricaudet, M , Mathews, M B , Klessig, D F & Pettersson, U (1980) *Cell*, **19**, 671–681
Avvedimento, V E , Vogeli, G , Yamada, Y , Maizel, J V , Pastan, I & deCombrugghe, B (1980) *Cell*, **21**, 689–696
Berget, S M & Sharp P A (1979) *J Mol Biol* **129**, 547–565
Chen Kiang, S , Wolgemuth, D J , Hsu M T & Darnell, J E (1982) *Cell*, **28**, 575–584
Chow, L T , Gelinas, R E , Broker, T R & Roberts, R J (1977) *Cell*, **12**, 1–8
Evans, R , Weber, J , Ziff, E & Darnell, J E (1979) *Nature (London)*, **278**, 367–370
Fraser, N W , Seghal, P B & Darnell, J E (1978) *Nature (London)*, **272**, 590–593
Fraser, N W , Seghal, P B & Darnell, J E (1979) *Proc Nat Acad Sci , U S A* **76**, 2571–2575
Gelinas, R E & Roberts, R J (1977) *Cell*, **11**, 533–544
Goldenberg, C J & Raskas, H J (1979) *Cell*, **16**, 131–138
Grosveld, C J , Koster, A & Flavell, R A (1981) *Cell*, **23**, 573–584
Keohavong, P , Gattoni, R , LeMoullec, J M , Jacob, M & Stevenin, J (1982) *Nucl Acids Res* **10**, 1215–1229
Long, B H , Huang C Y & Pogo, A O (1979) *Cell*, **18**, 1079–1090
Manley, J L , Sharp P A & Gefter, M L (1979) *J Mol Biol* **135**, 171–197
Manley, J L , Sharp, P A & Gefter, M L (1982) *J Mol Biol* **159**, 581–599
Mariman, E C M , van Eekelen C A G , Reinders, R J , Berns, A J M & van Venrooij, W J (1982) *J Mol Biol* **154**, 103–119
Nevins, J R (1982) *Cell*, **28**, 1–2
Scholtissek, C (1971) *Eur J Biochem* **24**, 358–365
Sharp, P A (1981) *Cell*, **23**, 643–646
Tamm, I , Kikuchi, T , Darnell, J E & Saldiet Georgieff, M (1980) *Biochemistry*, **19**, 2743–2748
Tsai, M J , Ting, A C , Nordstrom, J L , Zimmer, W & O'Malley, B W (1980) *Cell*, **22**, 219–230

Zain, S , Gingeras, T. R , Bullock, P , Wong, G. & Gelinas, R E. (1979*a*) *J Mol Biol* **135**, 413–433
Zain, S , Sambrook, J , Roberts, R J , Keller, W , Fried, M & Dunn, A R (1979*b*) *Cell*, **16**, 851–861

Edited by P. Chambon

CHAPTER III

A MODEL FOR THE EXCISION OF INTRONS 1 AND 2 FROM ADENOVIRAL MAJOR LATE PRE-mRNAs

A model for the excision of introns 1 & 2 from adenoviral major late pre-mRNAs.

E.C.M. Mariman, R.J. van Beek-Reinders and W.J. van Venrooy

Department of Biochemistry, University of Nijmegen,
The Netherlands.

SUMMARY

Total steady state RNA was extracted from nuclei of Hela cells late after infection with adenovirus type 2. Most of the nuclear RNA is transcribed from the major late transcription unit (16.2 to 100.0 map units (m.u.)). To study the cleavage reactions involved in the splicing of leaders 1 and 2, we have used the S1 nuclease mapping technique with restriction fragments located in the region of intron 1 as DNA probes. The S1-mapping data showed that RNA species accumulate in total nuclear RNA from which the 5' part of intron 1 has been removed but still contain the 3' part. This indicates that intron 1 can be removed in a stepwise fashion following the 5' to 3' direction. We have located the ends of these putative processing intermediates in the nucleotide sequence. It seems that they do not resemble consensus acceptor on donor sites. Reverse transcription experiments have been performed in which a primer, located with its left end in leader 2, is extended into intron 1. The results show that a major cleavage site is located about 35 nucleotides before the acceptor site. Cleavage at the acceptor site seems to be rapidly followed by ligation of leader 1 to leader 2. A model for RNA splicing is presented.

INTRODUCTION

At the late stage of infection of Hela cells with adenovirus type 2 most of the viral RNAs are transcribed from the major late transcription unit. Three intervening sequences are present in the primary transcript (1). The first two,

introns 1 and 2, are located between leaders, while the third intron is located between leader 3 and the body of the mRNA. Analysis of the steady state nuclear RNA extracted from Hela cells at the late phase of infection has revealed the existence of a set virus-specific RNAs which lack the 5' part of the second intron (2). Those RNAs may be intermediates in a stepwise excision of intron 2 in the 5' to 3' direction. Since our former results (3) indicated that the removal of intron 1 may, at least for some of the primary transcripts, also be performed in a stepwise fashion, we have tried to find further indications for such a process. Therefore, steady state nuclear RNA from Hela cells 18 h after infection was analysed with the S1-mapping procedure using DNA fragments of 813 and 555 base pairs in length, which overlap the 5' and 3' half of intron 1, respectively. Our results are consistent with splicing in which the intron is excised in the 5' to 3' direction. Reverse transcription experiments have been performed to explore the situation close to the acceptor site. A major cleavage site was found to be located about 35 nucleotides before the acceptor site. Our results suggest that cleavage at the actual acceptor site is rapidly followed by ligation of leader 1 to leader 2.

MATERIALS AND METHODS

a) Infection of cells

Hela S3 cells were infected with adenovirus type 2 (2000 particles per cell) at a density of 5×10^6 cells/ml in Eagles minimal essential medium containing 1 mM arginine (MEM). After adsorption of the virus to the cells during 1 h at 37 °C, the cells were diluted to $0,3 \times 10^6$ cells/ml with MEM containing 5% newborn calf serum. Cells were harvested 18 h after infection.

b) Isolation of nuclear RNA

Nuclear RNA was isolated as previously described (5). At 18 h after infection cells were harvested on frozen Na/K/Mg solution (0,13 M NaCl, 0,05 M KCl, 1,5 mM $MgCl_2$) and centrifuged for 5 min at 800 g. The cells were washed once with Na/K/Mg solution and after pelleting they were resuspended in reticulocyte

suspension buffer (RSB: 10 mM NaCl, 10 mM Tris,pH 7.4, 1.5 mM $MgCl_2$). After addition of 0,1 volume of a sodium deoxycholate/ Nonidet P40 mixture (5% each) the cells were disrupted by intensive vortexing. The nuclei were pelleted, washed once with RSB and resuspended in the same buffer. Nuclear RNA was then isolated by extraction with phenol at 55 °C as described by Long et al. (6).

c) Isolation of restriction fragments

PBR 322 containing the adenoviral Xho I fragment F (15.5-22.4 m.u.) in the Sal I site was propagated in E.coli strain JA221 and extracted and purified as described by Alestrom et al. (4). All the DNA fragments used in this study were obtained from this plasmid DNA by restriction digestion and separation of the digestion products. The restriction enzymes used are depicted in Figure 1. All the enzymes were purchased from Bethesda Research Laboratories, Inc. and digestions were performed under the conditions prescribed by this company. DNA fragments were separated on mixed agarose/polyacrylamide gels (0.5% /5%). After electrophoresis the DNA was isolated from gel slices by the method of electrophoretic elution. When a particular fragment was obtained from a single digestion of the plasmid DNA, as for instance Taq I fragment 813, the eluted DNA was further purified by a second gel run and elution. When a DNA fragment was obtained by subdigestion of a restriction fragment such an extra purification step was omitted.

d) Labeling of DNA fragments

DNA restriction fragments were labeled via nick translation. A quantity of the fragment equivalent to 5 µg of Ad DNA was dissolved in 25 µl NT buffer (50 mM Tris, pH 7.5, 25 mM $MgCl_2$, 10 mM DTT, 50 µg/ml BSA, dGTP, dATP, dTTP (2 µM each), 40 µCi (α-^{32}P)-dCTP). To this solution 5 units Kornberg DNA-polymerase (Bethesda Res. Lab. Inc.) were added. The mixture was incubated for 1.5 h at 15 °C and then diluted to 200 µl with water, run over G50-coarse to remove the excess ^{32}P-dCTP and extracted with one volume of phenol followed by two extractions with chloroform. After addition of 10 µg of RNA and NaCl to 0.3 M the nucleic acids were precipitated with ethanol. This method

of labeling uses the nicks already present in the DNA. Although the 3' end would be labeled more efficiently than the 5' end, adding DNAse I (1 µg/ml) to the labeling mixture did not alter the pattern of S1 hybrids. Moreover, labeling without DNAse I seemed to improve the S1 results by removing nicks from the DNA strands.

e) S1-nuclease mapping and analysis of S1-resistant hybrids

RNA was hybridized to DNA under conditions in which DNA:DNA hybridization was minimized, using the procedure described by Manley et al. (7). Nuclear RNA (10 µg) was hybridized with a quantity of a restriction fragment equivalent to 5 µg Ad2 DNA. Hybridization concentrations was 0,35 µg RNA/µl. After S1 nuclease digestion the hybrids were precipitated with ethanol, the pellet was solubilized in TE buffer (10 mM Tris, pH 7.4, 1 mM EDTA), and applied on a 1 mm thick 5% polyacrylamide gel. Electrophoresis was for 20 h at 100 V in Tris/Na/EDTA buffer (40 mM Tris, pH 8.0, 8 mM sodium acetate, 2 mM EDTA). After electrophoresis the gels were directly exposed to Kodak XR-5 film for an appropiate period of time (3 to 16 h).

f) Reverse transcription

A quantity of fragment 366 equivalent to 5 g of Ad2 DNA was hybridized to 10 µg of nuclear RNA as described above. After hybridization the nucleic acid mixture was precipitated with ethanol, pelleted and solubilized in 20 µl RT buffer (50 mM Tris, pH 8.0, 8 mM $MgCl_2$, 10 mM DTT, dATP, dCTP, dGTP, dTTP (1 mM each)). To this solution 20 units reverse transcriptase (Dr J.W. Beard, Life Science Inc., Florida) were added and incubation was for 1.5 h at 42°C. Then SDS, EDTA and NaOH were added to a final concentration of 0.5%, 10 mM and 200 mM, respectively. Incubation for 30 min at 70°C was appropriate to degrade the RNA but to leave the DNA intact. After neutralization with 10% acetic acid the cDNA was precipitated with ethanol, pelleted and dissolved in TE buffer. The solution was applied on a 5% polyacrylamide gel containing 7M urea. Electrophoresis was for 16 h/120 V, 30 h/120 V or 10 h/220 V. After gel electrophoresis the cDNA was electrophoretically transferred to Gene Screen (New England Nuclear) using a Biorad Transblot

Cell (8). Blots were baked at 80 °C for 4 h in a vacuum oven, prehybridized for 3 h at 42 °C in 50% formamide, 5x SSC, 250 µg/ml ssDNA, 5x Denhardts mix, 0.1% SDS, 10 mM phosphate buffer, pH 6.5, and hybridized for 16 h with ^{32}P-labeled Xho I fragment F under the same conditions. After extensive washing the cDNA bands were visualized by exposing the blots to Kodak XR-5 film for 3 to 16 h.

RESULTS

a) S1 nuclease mapping in the region 1 (Taq I) - 813 (Taq I).

In order to explore the mechanism involved in the removal of the first intron from RNAs which are transcribed from the adenoviral major late transcription unit, we have used the S1 nuclease mapping procedure to locate ends of major components of viral RNA in the region of intron 1. First the situation in the 5' half of intron 1 was analysed using ^{32}P-labeled DNA fragments 813 and 662 (Figure 1).

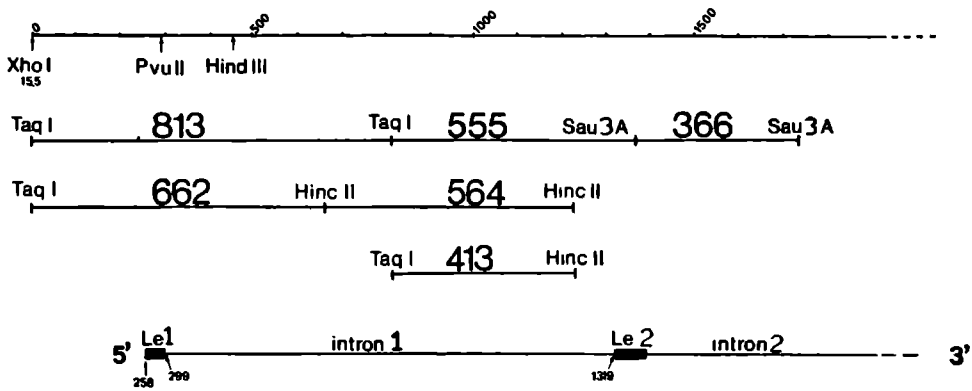


Fig. 1. DNA fragments used for S1-mapping and reverse transcription. Location of the restriction fragments used in this study relative to the left end of the Xho I fragment F (15.5 - 22.4 m.u.) from which they were derived. The length of the fragments (r-strand) is indicated in number of nucleotides. The nucleotide sequence is numbered from left to right starting with 1 at the r-strand nucleotide T of the Taq I restriction site at 15.5 m.u. in the genome (nucleotide 5782 of the Ad2 DNA (9)). The restriction sites used are: Taq I: 1, 814; Hinc II: 663, 1227; Sau3A: 1369, 1735. The position of the leaders and introns of the major late pre-mRNAs in this region are as depicted.

The labeled restriction fragments were hybridized to nuclear RNA which was isolated from Hela cells 18 h after infection (Materials & Methods). After hybridization the S1-resistant hybrids were resolved on a 5% polyacrylamide gel and visualized by direct autoradiography of the gel as described in the Material & Methods section. The deviation in the length of hybrid bands as determined in this way was estimated to be about 20 base pairs. When, as a control, labeled DNA fragments were treated in the same way as the DNA/RNA mixtures, no labeled bands could be detected showing that under these hybridization conditions no DNA:DNA hybridization occurred.

The results are shown in Figure 2A. When fragment 813 was used three major bands were observed with lengts of 540, 425 and 350 base pairs, respectively (lane 1). None of those, as most of the minor hybrid bands, could be detected when fragment 662 was used (lane 2). This demonstrates that the corresponding RNA species map with one end in the region of fragment 662 and extend to the right beyond nucleotide 662. Related to the 540 base pairs long hybrid the most prominent band obtained with fragment 662 was found to be 400 base pairs in length (lane 2). The RNA thus maps with its end near nucleotide 265. It is most likely that this RNA represents the primary transcript, starting with nucleotide 258. To match the hybrid of 425 a band of 295 base pairs was detected in the hybrid pattern obtained with fragment 662 (lane 2). These hybrids are derived from an RNA species which has its end near nucleotide 380. In agreement with the length of the 350 base pairs long hybrid a band of about 210 base pairs was detected using fragment 662 for S1-mapping. The corresponding RNA species has its end near nucleotide 460 and extends to the right beyond nucleotide 813. This RNA may well be the same species as we previously observed as the result of cleavage of viral pre-mRNA at coordinate 16.8 (7).

If the primary transcript would be cleaved at the donor site of intron 1 bands of 515 (299 - 813) and 364 (299 - 662) base pairs should be detectable with fragments 813 and 662, respectively. Indeed, bands of 505 and 365 base pairs (indicated by the arrow in lanes 1 and 2) can be seen as minor components of the hybrid patterns. This could mean that cleavage at the donor site does not preferentially occur in transcripts longer than

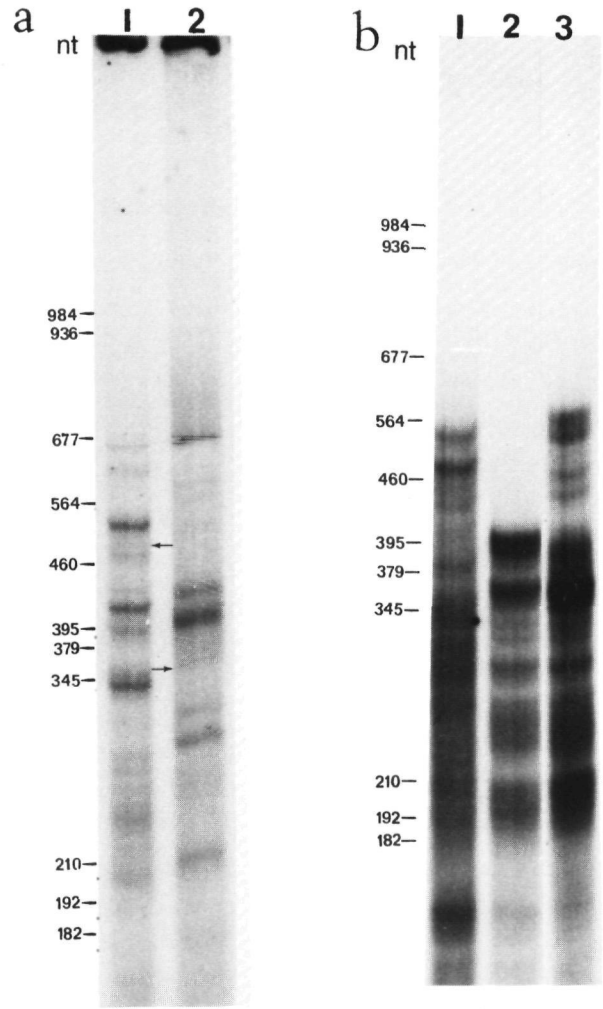


Fig. 2. S1-mapping in the region of intron 1. ^{32}P-labeled DNA fragments, as depicted in Figure 1, were hybridized with steady state nuclear RNA extracted from Hela cells 18 h after infection. S1-resistant hybrids were separated on 5% polyacrylamide gels and were visualized via autoradiography of the gels. (Materials & Methods). Taq I and Hinc II digests of the adenoviral HindIII restriction fragment B were used as molecular weight markers. The position of the HincII marker bands, as observed by ethidium bromide staining, was as indicated.
2A. Lane 1: fragment 813; lane 2: fragment 662.
2B. Lane 1: fragment 555; lane 2; fragment 413; lane 3: fragment 564.

405 nucleotides (258-662). Alternatively, as soon as RNAs are cleaved at the donor site they might directly be submitted to

further processing.

Except for the hybrid bands discussed above, minor bands can be observed some of which, on the basis of their length, seem to run over the site of initiation of transcription of the major late RNAs. We suspect that these species are transcribed from the l-strand and belong to the E2b or IVa2 transcription units (9).

b) S1 nuclease mapping in the region 814 (Taq I) - 1368 (Sau3A).

In a second series of S1-experiments we have tried to look into the situation at the 3' half of intron 1 regarding possible processing intermediates. We have used DNA fragments 564, 555 and 413 as depicted in Figure 1. Fragments 564 and 413 share the right end (nucleotide 1226), while fragments 555 and 413 share the left end (nucleotide 814).This should enable us to locate all the RNAs which map with one or both ends in the region between nucleo-tides 814 and 1226. Figure 2B shows the hybrid patterns obtained in a typical S1 experiment using ^{32}P-labeled DNA fragments and late steady state nuclear RNA (Materials & Methods). In the size range of 413 to 200 base pairs fragments 564 and 413 give a similar gel pattern (lanes 2 and 3) different from the pattern obtained with fragment 555 (lane 1). This indicates that nearly all the RNA species detected with fragment 413, map with one end in this region and extend to the right of nucleotide 1226. Lengths of such bands that can be distinguished are 365, 300, 250 and 200 base pairs. The last two of these are rather diffuse bands probably because their size is too small to yield discrete bands on 5% polyacrylamide gels. In accordance with these bands hybrids of 500, 440, 380 and 340 base pairs can be detected with fragment 555, showing that the corresponding RNAs extend beyond nucleotide 1368 and thus contain sequences of leader 2. The cleavage sites are located near nucleotides 865, 925, 980 and 1025, respectively.

With fragment 564 hybrids longer than 413 base pairs were detected. However, with the restriction fragments that we have used these species cannot be accurately mapped.

The acceptor site of intron 1 is located within the region of fragment 555. In case of cleavage at the acceptor site one would expect to find a band of 506 base pairs. (814-1319) and, indeed, a band of 500 base pairs is a major component of the

gel pattern (lane 1). However, based on the relative intensities of hybrid bands the 500 base pairs long hybrid rather seems to originate from cleavage of the RNA near nucleotide 865 than from cleavage at the acceptor site. Although cleavage of long RNAs at the acceptor site cannot be excluded, such a 3' cleavage product would be a very minor species in steady state nuclear RNA. As already mentioned a similar result was found for cleavage products generated from the donor site.

c) Reverse transcription of virus-specific steady state nuclear RNA.

In a previous study (3) we have described that nascent transcripts can be cleaved at coordinate 19.1, i.e. at or near the acceptor site (19.2 m.u.) of intron 1. Although in rapidly labeled nuclear RNA the 3' cleavage product (16.2-19.1 m.u.) can be easily identified, labeling the RNA for 45 min or longer makes it impossible to detect this species among the labeled RNA via S1-mapping using Xho I restriction fragment F (15.5-22.4 m.u.). This and the results described in section (b) of this chapter indicate that the 3' cleavage product does not accumulate in steady state nuclear RNA. We were interested to know if we could detect the 5' cleavage product resulting from cleavage at coordinate 19.1. Therefore, steady state nuclear RNA extracted from Hela cells 18 h after infection was hybridized to DNA fragment 366 (1369-1734, Figure 1), from which the left end is located in leader 2. After hybridization the nucleic acid mixture was treated with reverse transcriptase, as described in the Materials & Methods section, in order to obtain cDNAs by the extension of fragment 366 into intron 1. Then the mixture was treated with NaOH to degrade the RNA and, after neutralization and precipitation, the cDNA mixture was resolved on a 5% polyacrylamide gel containing 7M urea. The cDNAs were transferred to Gene Screen and visualized by hybridization of the blots with ^{32}P-labeled Xho I fragment F followed by autoradiography. The results of two such experiments are shown in Figure 3. Reverse transcription seemed to proceed up to the 5' end of leader 1 in the primary RNA transcript resulting in a cDNA of 1477 nucleotides in length (258-1734). The most prominent band is the 366 primer since this fragment

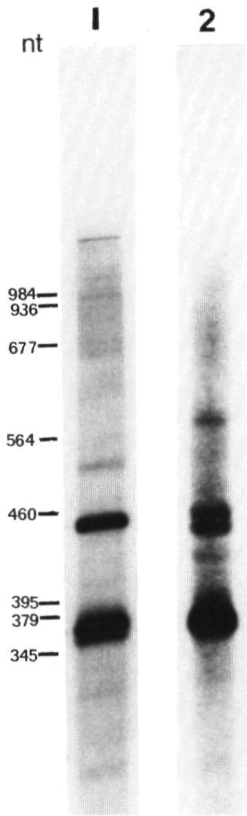


Fig. 3. Reverse transcription into intron 1.
DNA fragment 366, of which the location is depicted in Figure 1, was hybridized with steady state nuclear RNA extracted from Hela cells 18 h after infection. After hybridization the mixture was treated with reverse transcriptase. Then, RNA was degraded and the cDNAs were separated on a 5% polyacrylamide gel containing 7 M urea. The cDNA bands were electrophoretically transferred to Gene Screen and visualized by hybridization with ^{32}P-labeled Xho I restriction fragment F (Materials & Methods). The results of two experiments are shown. In Experiment 1 electrophoresis was for 16 h at 120 V, while in Experiment 2 electrophoresis was for 10 h at 220 V. Taq I and Hinc II digests of the adenoviral HindIII restriction fragment B and Taq I digest of PBR 322 were used as molecular weight markers. The position of the Hinc II marker bands was as indicated.

was added in excess over the RNA. The most prominent cDNA band observed in experiment 1 has a length of about 450 nucleotides which agrees with what would be expected for the splicing product in which leader 1 is linked to leader 2, i.e. 456 nucleotides. However, this band appeared to consist of two different

species, which could be separated by running the gel for longer times or at a higher voltage. In the latter case the heat which is generated in the gel may promote the denaturing effect. Experiment 2 in Figure 3 shows the bands separated.

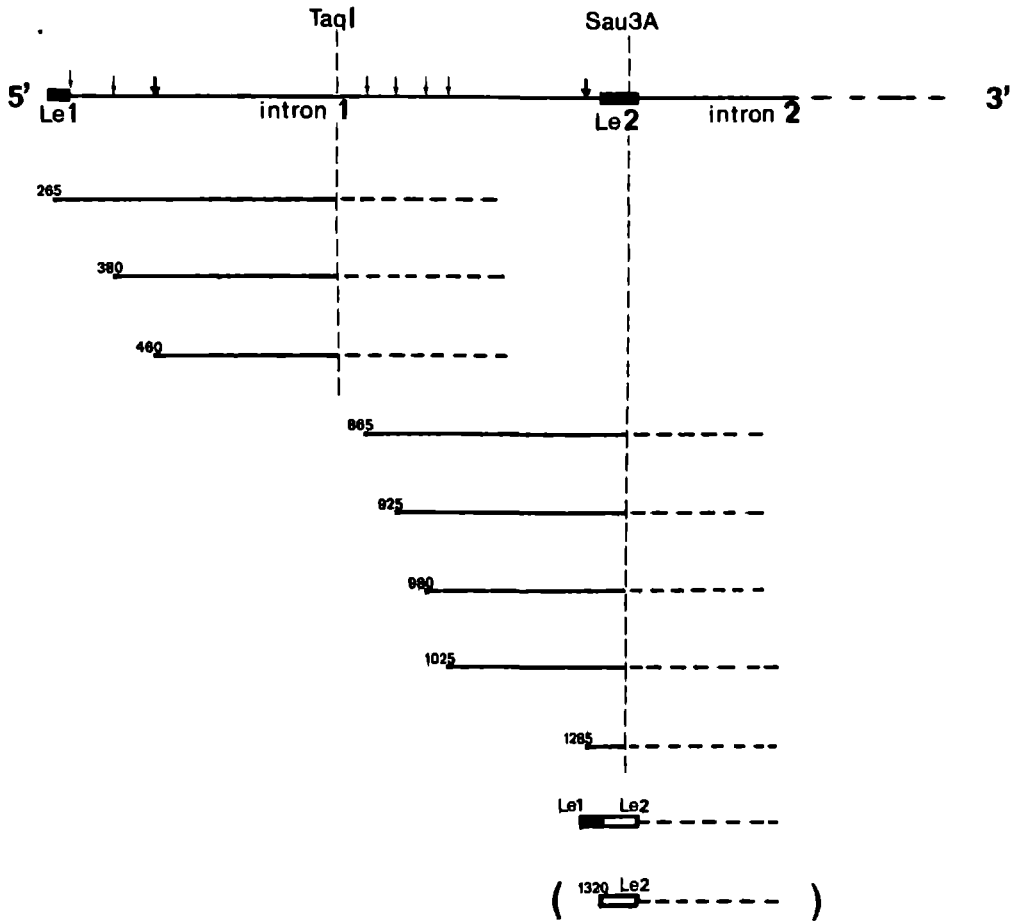


Fig. 4. Cleavage sites in intron 1.

Summary of the data obtained in the present study. The ends of the putative processing intermediates are located near the indicated nucleotide. The preferred (heavy arrow) and non-preferred (thin arrow) cleavage sites in the primary transcript are shown. Heavy bar: exon sequences, thin bar: intron sequences. The RNA placed in parentheses is the species which would result from cleavage at the acceptor site but which cannot be detected in steady state RNA.

They differ only 5 to 10 nucleotides in length. The upper band could be identified as the splicing product, because with a specific probe the precence of leader 1 sequences could be demonstrated in this RNA species (data not shown). The lower band represents an RNA species which seems to contain the 3' ultimate 30 - 35 nucleotides of intron 1. Since this species exists in steady state RNA in about the same concentration as the splicing product, it suggests that there is a major cleavage site some 35 nucleotides before the acceptor site in intron 1. Cleavage at the acceptor site would result in a reverse transcription product of 415 nucleotides (1320-1734). At this location in the gel only a faint band can be detected showing that in total RNA the 5' cleavage product resulting from cleavage at the actual acceptor site does not accumulate. Cleavage at this site seems to be followed rapidly by ligation of leader 1 to leader 2. These data together suggest that rapid cleavage of the nascent transcript occurs near nucleotide 1285 in intron 1 (19.1 m.u.) rather than at the actual acceptor site (19.2 m.u.).

Next to the bands described above, some additional minor cDNA bands can be observed. Some of them could be derived from reverse transcription of processing intermediates, such as those described in sections (a) and (b) of this paragraph, but the low intensity of these bands does not allow accurate identification.

Figure 4 summarizes the data obtained in the present study. As earlier proposed for intron 2, our results suggest that removal of intron 1 is a stepwise process as well and proceeds in the 5' to 3' direction.

DISCUSSION

Earlier studies on steady state nuclear RNA isolated from Hela cells at the late phase of infection with adenovirus 2 have revealed the existence of RNAs which lack part of intron 2 (2). These RNAs can be regarded as intermediates of a splicing process in which the intron is removed in a stepwise fashion following the 5' to 3' direction. In the present study we have analysed steady state nuclear RNA via S1-mapping and

reverse transcription in the region of intron 1. Our data show that RNA species which have lost the 5' part of intron 1 but still contain the 3' part accumulate in the nucleus. Thus, a similar mechanism seems to be operative in both introns 1 & 2, i.e. stepwise excision in the 5' to 3' direction.

Previously we have identified two major cleavage sites in intron 1, at 16.8 and 19.1 m.u., respectively (3). In case of cleavage at 19.1 both the 5' and 3' cleavage products could be detected in rapidly labeled RNA. In steady state RNA, however, only the 5' cleavage products seem to be present. One way to explain this observation is that the removal of an intron follows discrete steps in the 5' to 3' direction, while the intron sequences of the 3' cleavage product are gradually degraded. In this way no discrete 3' cleavage products can be detected in steady state RNA. Stepwise splicing could also explain why large spliced-out intron sequences have not been found yet. When the hybrid patterns obtained with fragments 413, 555 and 564 are compared (Figure 2B) there seems to be no common hybrid band longer than 200 base pairs. Below this size some similarity in the patterns can be seen which might indicate that spliced-out intron-pieces are about 150 nucleotides or shorter. It should be noted that we have not detected RNAs which could be the result of premature termination of transcription. Since these RNAs seem to be rather heterogeneous in length they may contribute to the background of the gel patterns instead of appearing as discrete major bands. (10).

Rapid cleavage of the primary transcript can occur at coordinates 16.8 and 19.1, respectively (3). According to cleavage at 16.8 m.u. an RNA species is present in steady state nuclear RNA with its 5' end near nucleotide 460. In case of cleavage at 19.1 m.u. S1-mapping using the Xho I fragment F did not allow to distinguish between cleavage close to or at the acceptor site (19.2 m.u.). Reverse transcription experiments indicate, however, that a major component of steady state RNA results from cleavage of the RNA some 35 nucleotides before the acceptor site. Therefore, it seems likely that the site for rapid cleavage is located about 35 nucleotides before the actual acceptor site in intron 1 near nucleotide 1285. All the rapid cleavages which are introduced in the (nascent) primary transcript seem

to be located at an internal site in the intron. These results argue against a simple RNA scanning model for splicing (11), in which the excision starts at one of the ends of the intron. Rather the excision starts somewhere within the intron sequence. Other arguments against a simple intron scanning model have been reported by Kühne et al. (12). Combining the former and present results, our data suggest that in intron 1 preferred and non-preferred cleavage sites can be distinguished (Figure 4).

With reverse transcription hardly any RNA species can be found which has its 5' end at nucleotide 1320, the acceptor site. This indicates that cleavage at the acceptor site is rapidly followed or coupled with ligation of leader 1 to leader 2. Kinetic labeling experiments have shown that a similar situation might exist for splicing of leader 2 to leader 3 (3,13). Cleavage at the acceptor site seems to be the last step in the excision of the introns and is rapidly followed by the ligation of the leaders. This may be a more general phenomenon of RNA splicing.

The S1-mapping experiments described above enabled us to compare the nucleotide sequences around the 5' end of the putative processing intermediates (14). As far as we can decide, the internal cleavage sites do not resemble the consensus donor or acceptor sequences (15) suggesting that cleavage in an intron is not determined by the primary nucleotide sequence but rather by other factors such as RNP-structure or secondary structure.

The following model can be proposed to fit all our results (Figure 5). The primary transcript is rapidly cleaved at a preferred site within the intron. The intron sequences of the 3' cleavage product are gradually degraded up to the 3' end of the first exon. The 5' cleavage product is further processed by cleavage of the RNA at non-preferred sites removing the intron via discrete steps in the 5' to 3' direction. The last cleavage event takes place at the acceptor site and is rapidly followed by ligation of the exons. The model requires a higher ordered structure to hold the 5' and 3' cleavage products together. The RNP structure, or the nuclear matrix as an intranuclear ultrastructure (16), may meet such a requirement (5). We have analysed rapidly labeled nuclear matrix RNA via the S1-mapping procedure using the Xho I restriction fragment F. Both

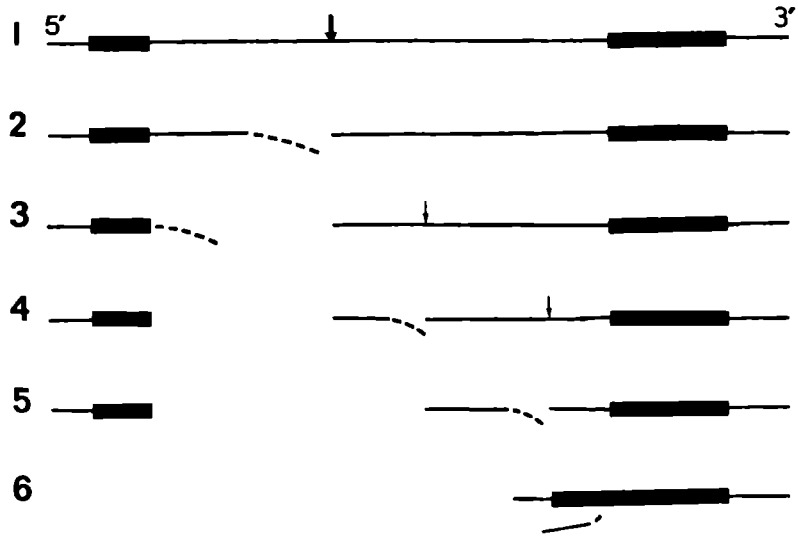


Fig. 5. A model for RNA splicing.
The model as depicted in this figure fits all the presently and previously obtained data on splicing involved in the formation of the tripartite leader. A description of the model is given in the text of the Discussion section. Heavy bar: exon sequences, thin bar: intron sequences. The arrows indicate preferred (heavy) and non-preferred (thin) cleavage sites.

the 5' and 3' cleavage product generated by cleavage of the RNA at coordinate 19.1, 20.7 and 21.1, respectively, could be detected showing that the cleavage products are indeed associated with the nuclear matrix (data not shown). Also the results of van Eekelen et al. (8) who showed that both intron and exon sequences are associated via the C-proteins to the nuclear matrix support the hypothesis that higher ordered structures might be an essential element in the process of RNA splicing.

The procedure of S1-mapping as we have used it, can only locate the ends of RNA molecules which are colinear to the DNA and longer than 150 base pairs. Therefore, it is possible that ligation of leader 1 occurs to the end of each 5' cleavage product. However, the shortest intermediate was readily detected by the reverse transcription and contains only about

35 nucleotides linked to leader 2, not derived from leader 1. Further, the nucleotide sequence at the 5' end of the cleavage products does not resemble the consensus donor sequence. Although these findings argue against such a process of repeated ligation, it is one of the many aspects of RNA processing which are subject of future studies.

ACKNOWLEDGEMENTS

We are indebted to Drs C. van Eekelen and P. Sillekens for help and advice and A. Groeneveld for culturing the cells.

This work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

1. Broker, T.R. and Chow, L.T. (1980) Trends in Biol. Sci. 7, 174-178.
2. Berget, S.M. and Sharp, P.A. (1979) J.Mol.Biol. 129, 547-565.
3. Mariman, E.C.M., van Beek-Reinders, R.J. and van Venrooij, W.J. (1983) J.Mol.Biol. 163, 239-256.
4. Alestrom, P., Akusjärvi, G., Perricaudet, M., Mathews, M., Klessig, D. and PEttersson, U. (1980) Cell 19, 671-681.
5. Mariman, E.C.M., van Eekelen, C.A.G., Reinders, R.J., Berns, A.J. and van Venrooij, W.J. (1982) J.Mol.Biol. 154, 103-119.
6. Long, B.H., Huang, C-Y. and Pogo, A.O. (1979) Cell 18, 1070-1090.
7. Manley, J.L., sharp, P.A. and Gefter, M.L. (1979) J.Mol Biol. 135, 171-197.
8. van Eekelen, C., Ohlsson, R., Philipson, L., Mariman, E., van Beek, R. and van Venrooij, W. (1982) Nucleic Acids Res. 10, 7115-7131.
9. Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J. (1982) J.Biol.Chem. 257, 13475-13491.
10. Fraser, N.W., Seghal, P.B. and Darnell, J.E. (1979) Proc. Natl.Acad.Sci.USA 76, 2571-2575.
11. Sharp, P.A. (1981) Cell 23, 643-646.
12. Kühne, T., Wieringa, B., Reiser, J. and Weissman, C. (1983) EMBO Journal 2, 727-733.
13. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stevenin, J. (1982) Nucleic Acids Res. 10, 1215-12229.
14. Alestrom, P., Akusjärvi, G., Pettersson, M. and Pettersson, U. (1982) J.Biol.Chem. 257, 13492-13498.

15. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-471.
16. van Eekelen, C.A.G. and van Venrooij, W.J. (1981) J.Cell Biol. 88, 554-563.

CHAPTER IV

SOME ASPECTS OF THE EXCISION OF INTRON 3 FROM ADENOVIRAL MAJOR LATE PRE-mRNAs

Some aspects of the excision of intron 3 from adenoviral major late pre-mRNAs.

E.C.M. Mariman, R.J. van Beek-Reinders and W.J. van Venrooij

Department of Biochemistry, University of Nijmegen, The Netherlands.

SUMMARY

We have used different methods of rapid RNA labeling to study cleavage reactions, which are involved in the removal of intron 3 from adenoviral pre-mRNAs produced from the major late transcription unit. The labeled RNA was analysed via the S1-mapping procedure. Our results show that cleavage as part of RNA splicing can occur at the acceptor site of intron 3 before the tripartite leader is completed. Cleavage at the donor site of intron 3 cannot occur before intron 2 is excised, suggesting that ligation of the four exons which constitute the major late mRNAs follows the 5' to 3' direction.

INTRODUCTION

At the late stage of infection of human cells with adenovirus type 2 most of the viral RNA is transcribed from the major late transcription unit (16.2 - 100.0 m.u.; the genome is divided in 100 map units (m.u.), 1% of the genome equals 360 nucleotides). Via polyadenylation and splicing the primary transcript is processed to one out of several possible polyadenylated mRNAs, which are divided into five 3' coterminal families L1 to L5 (1). Polyadenylation is assumed to be an early step in the processing of pre-mRNA and can occur in about one minute after the RNA-polymerase has passed the polyadenylation site (2). Splicing involves the linkage of three leader sequences to form the tripartite leader, and linkage of the last leader to the protein coding sequence. Keohavong et al. (3) have shown that completion of the tripartite leader takes about

15 min. Since splicing is relatively slow and rapid polyadenylation seems to occur on non-spliced molecules (2), one would expect long polyadenylated pre-mRNAs to accumulate in the hnRNA of infected cells, the more, while late after infection 50% of the nuclear RNA appears to be virus-specific (4). In uninfected Hela cells about 15 - 20% of the hnRNA is polyadenylated (5). In spite of the expectation, however, in late infected Hela cells not more than 6% of the hnRNA can be selected via its poly(A)-tail (6). We have tried to find an explanation for this relatively low amount of polyadenylated pre-mRNA. Earlier we have found that rapid cleavages can be introduced in the primary transcript (7). The sites for cleavage are located at coordinates 16.8, 19.1, in intron 1, and 20.7, 21.1, 25.5, in intron 2, respectively. Since these sites are relatively close to the site for initiation of transcription, this cannot by itself be the cause for the low percentage of poly(A)-containing RNA. Therefore, we hypothesized that rapid cleavage as part of RNA splicing might be introduced in the promotor distal part of the pre-mRNA as well, i.e. in intron 3 between the third leader and the protein coding sequence. We have used the S1-mapping procedure with restriction fragments HindIII B (16.8 - 31.5 m.u.; Fig 1A) and EcoRI B (58.5 -70.7 m.u.; Fig 1B) to analyse nuclear RNA which was labeled in different ways. Our results show that rapid cleavage can occur at an acceptor site in intron 3. Cleavage at the donor site of intron 3 can occur only if cleavage at the acceptor site of intron 2 has taken place, suggesting that ligation of the four exons which constitute the mRNA follows the 5' to 3' direction.

MATERIALS AND METHODS

a) Infection and labeling of cells

Hela S3 cells were infected with adenovirus type 2 (2000 particles per cell) at a density of 5×10^6 cells/ml in Eagles minimal essential medium (MEM) containing 1 mM arginine. After adsorption of the virus to the cells during 1 h at 37°C, the cells were diluted to 0.3×10^6 cells/ml with MEM containing 5% newborn calf serum. Cells were harvested at 18 h after

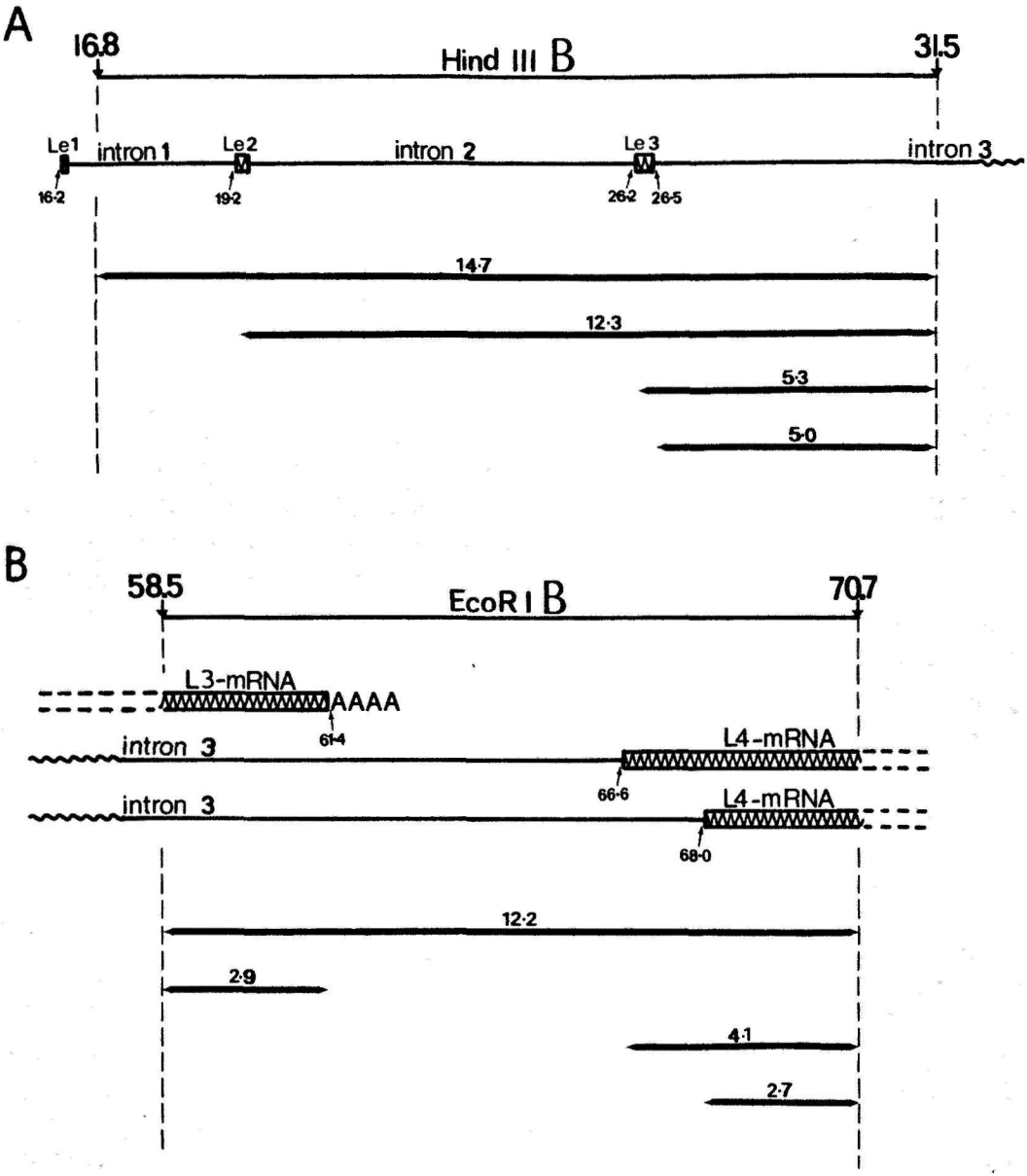


Fig. 1. Location of the RNA species relative to the HindIII and EcoRI restriction fragments B.
A. HindIII restriction fragment B is colinear to the 3' part of intron 1, leader 2, intron 2, leader 3 and the 5' part of intron 3. The heavy double-pointed bars represent the DNA:RNA hybrids obtained in S1 experiments as discussed in the text.

The length of each hybrid is indicated in percentages of the genome. 14.7%: primary transcript; 12.3%: splicing intermediate from which intron 1 is excised; 5.3%: intermediate from which intron 2 is excised; 5.0%: intermediate resulting from cleavage at the donor site of intron 3.
B. EcoRI restriction fragment B is colinear to the 3' part of L3-RNAs, precursors for L4- and L5-RNAs and the 5' part of L4-mRNA body sequences. The region E2a RNAs are not shown. The length of each hybrid is indicated in percentages of the genome. 12.2%: precursors for L4- and L5-specific RNAs; 2.9%: the 3' part of L3-specific RNAs which are cleaved at the L3-polyadenylation site (61.4 m.u.); 4.1%: body sequence of the longest L4-mRNA (starting at 66.6 m.u.); 2.7%: body sequence of the shorter L4-mRNA in this region (starting at 68.0 m.u.).

infection. To obtain rapidly labeled nuclear RNA 18 h after infection, cells were concentrated to 20 x 10^6 cells/ml and incubated with 20 µci/ml (5-^{3}H)-cytidine (31 Ci/mmol, Amersham, England),20 µCi/ml (5'-^{3}H)-guanosine (22 Ci/mmol, Amersham, England) and 20 µCi/ml (5,6-^{3}H)-uridine (45 Ci/mmol, Amersham, England). The labeling times were as indicated in the text.

b) Preparation of nuclear RNA

18 h after infection cells were harvested on frozen NKM solution (0,13 M NaCl, 0.05 M KCl, 1.5 mM $MgCl_2$) and centrifuged for 5 min at 800 g. The cells were washed once with NKM and after pelleting they were resuspended in RSB (10 mM NaCl, 10 mM Tris, pH 7.4, 1.5 mM $MgCl_2$). After addition of 0.1 vol. of a sodium deoxycholate/Nonidet P40 mixture (5% each) the cells were disrupted by intensive vortexing. The nuclei were pelleted, washed once with RSB and resuspended in the same buffer. Nuclear RNA was then isolated by phenol extraction at 55°C as described by Long et al. (8).

c) Isolation of restriction fragments

pBR322 containing either the HindIII A or B restriction fragments in the HindIII site (generous giftsof R. Ohlsson and M. Mathews) were propagated in E.coli strain JA221. Extraction and purification of plasmid DNA were performed as described by Aleström et al (9). Fragment HindIII B was isolated by digestion of the cloned DNA with HindIII. Fragment EcoRI B was obtained from the cloned HindIII A fragment. Restriction digestions with HindIII and EcoRI were performed at 37°C for 4 h using 1 unit of enzyme per µg of DNA. Standard conditions were

50 mM Tris, pH 7.4, 7 mM $MgCl_2$, 50 mM NaCl, 3 mM dithiothreitol. BSA was added to the digestion mixtures to a final concentration of 1 mg/ml. Restriction fragments were separated by agarose gel electrophoresis; 0,8% agarose gels were run for 20 h at 80 V in TNE buffer (40 mM Tris, pH 8.0, 8 mM sodium acetate, 2 mM EDTA). The DNA was recovered from agarose gel slices by electrophoretic elution.

d) S1-nuclease mapping and analysis of S1-resistant hybrids

RNA was hybridized to DNA under conditions in which DNA:DNA hybridization was minimized, using the procedure described by Manley et al(10). 1 µg of nuclear RNA was hybridized to 2.5 µg equivalent of a restriction fragment of Ad2 DNA. Hybridization concentration was 0.35 µg RNA/µl. After S1 nuclease digestion the hybrids were precipitated with ethanol, the pellet was solubilized in TNE buffer and applied on a 1.2% agarose gel. Electrophoresis was performed under the same conditions as were used for the separation of DNA restriction fragments. Gels were subsequently treated with 5% TCA, methanol, methanol containing 10% PPO and water. After drying they were exposed to Kodak XR-5 film for an appropriate period of time (3 to 7 days).

RESULTS

a) S1 nuclease analysis of rapidly labeled nuclear RNA.

Previously we have analysed nuclear RNA extracted from Hela cells at the late stage of infection (17). Using restriction fragment EcoRI B and steady state labeled polyA(+) nuclear RNA two S1 resistant hybrids, 4.1% and 2.7% in length (Fig 1B), could be identified which appeared to represent the 5' part of the body of two L4-mRNAs. These species were also found with 20 min labeled nuclear matrix RNA. In a first attempt to find out how fast cleavage at the acceptor site of intron 3 could occur, we analysed 8 min labeled total nuclear RNA via S1-mapping. As can be seen in Figure 2, a 2.9% long hybrid representing the 3' end of L3-specific sequences which end at the L3-polyadenylation site (61.4 m.u.) was clearly visible as should be expected in case of a rapid polyadenylation process (2).

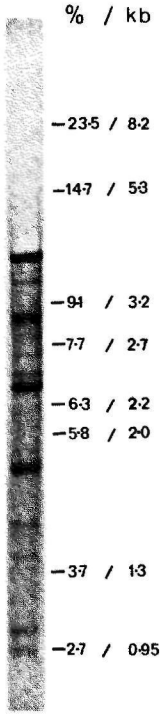


Fig. 2. S1 analysis of 8 min labeled total nuclear RNA. Hela cells, 18 h after infection, were labeled for 8 min with ^{3}H-nucleosides and total nuclear RNA was isolated as described in the Materials & Methods section. The RNA was analysed via the S1-mapping procedure using HindIII B as a probe. S1 resistant hybrids were resolved on agarose gels and visualized via fluorography (Materials & Methods). EcoRI and HindIII digests of Ad2 DNA were used as molecular weight markers. The positions of the HindIII markers are indicated.

As minor components of the labeled RNA, the 5' part of both L4 body sequences located in the region of EcoRI b could be detected (Fig. 2). This demonstrated that 8 min after the start of the RNA labeling cleavage at the acceptor site of intron 3 already had occurred. The leader sequences are not colinear to the EcoRI B restriction fragment (58.5 - 70.7 m.u.). Therefore, we cannot distinguish wether these hybrids are formed from the splicing product, in which leader 3 is ligated to the body

sequences, or from a splicing intermediate, derived from cleavage of the RNA at the acceptor site of intron 3, at 66.6 m.u. or 68.0 m.u., respectively.

b) S1 analysis of labeled nuclear RNA isolated from DRB-treated cells.

It has been demonstrated that cleavage at the acceptor site of intron 2 and ligation of leader 2 to leader 3 occurs about 12 - 15 min after the start of transcription (3,7). We were interested to know if cleavage at the acceptor sites of the L4-mRNA bodies in intron 3 occurred before or after completion of the tripartite leader. The site for initiation of transcription (16.2 m.u.) and the acceptor site of the longest L4-mRNA body (66.6 m.u.) are separated by more than 50% of the genome. Taking the speed of the RNA polymerase to be 50 -100 nucleotides per second (11-13) it may last about 6 min before the enzyme reaches the L4-specific region. Therefore, the 8 min labeled RNA contains L4-species which are being synthesized and processed for at least 6 and maximally 14 min. Since in this time period the leaders could have been spliced, analysis of the 8 min labeled RNA cannot give an answer to the question asked. Shorter labeling times are not adequate because then the RNA in this region does not accumulate enough label to enable accurate S1 analysis. Therefore, we performed a kinetic labeling experiment in which the RNA is labeled exclusively from the 5' end. Hela cells 18 h after infection were treated for 1 h with 70 µM DRB (5,6-dichloro-1-βD-ribofuranosylbenzimidazole). Then the cells were cooled to $0^{\circ}C$, pelleted and washed, to remove the drug, and resuspended in culture medium containing ^{3}H-uridine, -cytidine and -guanosine in order to label the RNA at $37^{\circ}C$. DRB synchronizes the RNA-polymerases close to the promotor site (7,14,15). It seems to inhibit the elongation of transcription, maybe by promoting premature termination (16), but not the initiation of transcription. Removal of the drug restores the elongation activity so that accumulation of label in the RNA follows the 5' to 3' direction. When labeling in this way is performed for 8 min the labeled RNA contains species which have been synthesized and processed for not longer than 8 min. Since previous experiments suggested that the elongation

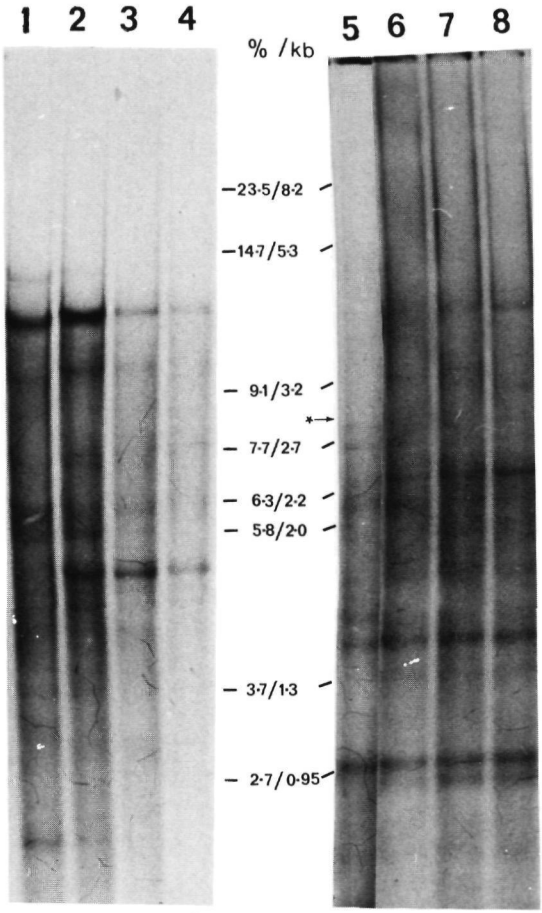


Fig. 3. S1 analysis of nuclear RNA extracted from DRB treated cells. From 17 to 18 h after infection Hela cells were treated with DRB by adding the drug to the culture medium to a final concentration of 70 µM. At 18 h after infection the drug was removed and the cells were labeled with ^{3}H-nucleosides in culture medium at 37°C for 10, 20, 30 and 45 min, respectively. Then, total nuclear RNA was isolated and analysed via the S1-mapping procedure using HindIII B (lanes 1 to 4) and EcoRI B (lanes 5 to 8) fragments as probes. S1 resistant hybrids were resolved on agarose gels and visualized by fluorography (Materials & Methods). Lanes 1 & 5: 10 min labeling; lanes 2 & 6: 20 min labeling; lanes 3 & 7: 30 min labeling; lanes 4 & 8: 45 min labeling. EcoRI and HindIII digests of Ad2 DNA were used as molecular weight markers. The position of the HindIII markers is indicated.

of transcription after DRB treatment is somewhat slower, we choose a minimum labeling time of 10 min. The RNA was labeled

for 10, 20, 30 and 45 min, respectively, and analysed via S1-mapping using HindIII and EcoRI restriction fragments B as DNA probes (see Materials & Methods). The results are shown in Figure 3. With the HindIII B fragment as a probe, cleavage at the acceptor site of intron 2 will yield an RNA:DNA hybrid of 5.3% in length (26.2 - 31.5 m.u.) (Fig 1A). In the pattern obtained with 10 min labeled RNA this hybrid cannot be detected (Fig 3, lane 1). The hybrid band only appears after 20 min of labeling (Fig 3, lane 2). This demonstrates again (see chapter II) that 10 min after the start of transcription cleavage at the acceptor site of intron 2 has not yet occurred. Analysis of the RNA with fragment EcoRI B (58.5 - 70.7 m.u.; Fig 1B) showed that, when after DRB treatment the RNA was labeled for 20 min or longer, all the RNA species known to be present in steady state polyA(+) RNA or in 8 min labeled total RNA could be detected (Fig 3, lanes 6 to 8). The pattern of hybrids obtained with RNA labeled for 10 min, after the DRB treatment, however, differs significantly (Fig 3, lane 5). The most prominent band is derived from cleavage of the RNA at the polyadenylation site of the L3-coterminal family (61.4 m.u.). Polyadenylation can be a very rapid processing step. Looking for the hybrids of 4.1% and 2.7% in length, representing the L4-mRNA bodies in this region, we can detect the longer species, but not the shorter one. So, cleavage at the acceptor site of intron 3 to give the longest L4-mRNA (66.6 m.u.) can occur rapidly, that is before cleavage at the acceptor site of intron 2 and before ligation of leaders 2 and 3. Cleavage at the other acceptor site (68.0 m.u.) to produce the shorter L4-mRNA seems to be a later event.

When rapid cleavages introduced in introns 1 and 2 were studied both the 5' and 3' cleavage products could be identified in rapidly labeled RNA (see chapters II and III). Using the EcoRI B fragment (58.5 -70.7 m.u.) a hybrid of 8.1% in length (58.5 - 66.6 m.u.) would be expected from cleavage of the RNA at coordinate 66.6. Indeed, such a band can be detected with 10 min labeled RNA (Fig 3, lane 5) as indicated by the arrow. Although some additional bands can be observed, the identity of these RNA species remains to be elucidated.

c) Kinetics of the cleavage reactions around leader 3.

We have seen that cleavage at the acceptor site of intron 3 can occur before splicing of the tripartite leader is completed. Then we asked the same question for cleavage at the donor site of intron 3: Does that cleavage occur before completion of the tripartite leader. In a previous study (7) we identified an RNA species that yielded a hybrid of 5.0% in S1-experiments with Hind III B as a probe. Using subfragments of HindIII B the RNA was accurately mapped. It has its 5' end at coordinate 26.5 and extends to the right beyond 31.5 m.u. It is produced by cleavage of pre-mRNA at the donor site of intron 3 (26.5 m.u.; Fig 1A). We have performed kinetic labeling experiments to learn how fast this species is produced. Hela cells 18 h after infection were labeled for 3, 7, 15, 30 or 5, 15, 25, 35 min, respectively. After labeling total nuclear RNA was isolated (Materials & Methods). The RNA was analysed via the S1-mapping procedure using HindIII B (16.8 - 31.5 m.u.) as described in the Materials & Methods section. Earlier we have depicted the results obtained with RNA labeled for 3 to 30 min (7). Here we show the results of an experiment in which the RNA was labeled for 5 to 35 min (Figure 4). In both types of labeling experiments the appearance of the 5.0% long hybrid follows that of the 5.3% long band. This suggests that cleavage at the acceptor site of intron 2 (26.2 m.u.), which itself is a late event in the splicing of intron 2, precedes the cleavage at the donor site of intron 3. Cleavages around leader 3 thus seem to occur in an ordered fashion. Despite the rapid acceptor site cleavage of intron 3, cleavage at the donor site of intron 3, and thus also ligation of leader 3 to the mRNA body, has to wait for the cleavage at the acceptor site of intron 2 and is therefore, a late event in the splicing process.

DISCUSSION

The adenoviral major late mRNAs can be divided into five 3' coterminal families, referred to as L1 to L5 (1). The members of one family differ in the length of the body sequences and thus in the location of the 5' end of the mRNA-body.

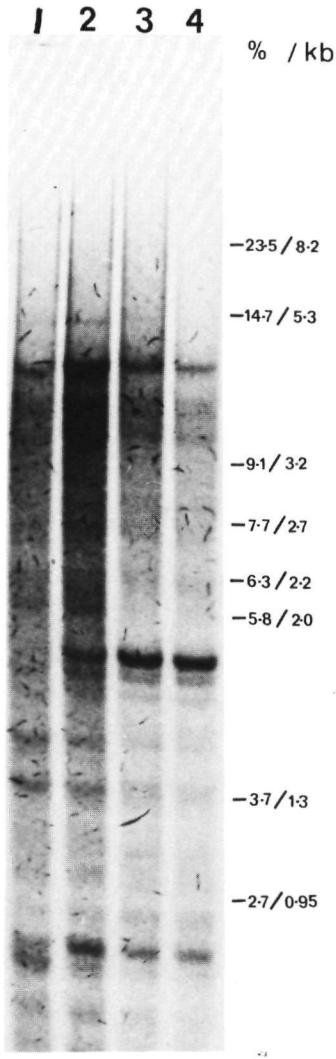


Fig. 4. S1-nuclease analysis of nuclear RNA labeled for various periods of time. Hela cells 18 h after infection were labeled with ^{3}H-nucleosides for 5, 15, 25 and 35 min, respectively. After labeling total nuclear RNA was isolated and analysed via the S1-mapping procedure using HindIII B as a probe. S1 resistant hybrids were resolved on agarose gels and visualized by fluorography (Materials & Methods).
EcoRI and HindIII digests of Ad2 DNA were used as molecular weight markers. The position of the HindIII markers is indicated.
Lane 1: 5 min labeling; lane 2: 15 min labeling; lane 3: 25 min labeling; lane 4: 35 min labeling.

Consequently, intron 3 which is situated between leader 3 and the mRNA-body, can have different lengths. In removing intron 3 always the same donor site is used which is located at the 3' end of leader 3 (26.5 m.u.; Fig 1A). Depending on which member of which coterminal family will be produced, one out of several acceptor sites is selected. We have attempted to gain more knowledge about the excision of intron 3. Therefore, we analysed nuclear RNA extracted from Hela cells at the late stage of infection via the S1-mapping procedure. The RNA was labeled in several ways. As DNA probes the HindIII B restriction fragment, which overlaps leader 3 and the 5' part of intron 3, and fragment EcoRI B, which overlaps the 3' part of intron 3 and the 5' part of two L4-mRNA bodies, were used (Fig. 1A and 1B). Our conclusions can be summarized as follows:

- cleavage at the acceptor site of the longest L4-mRNA body (66.6 m.u.) can occur before the tripartite leader is completed.
- the fast cleavage at the donor site at the 3' end of leader 3 is in the processing pathway a late event, because it has to wait for cleavage of the acceptor site of intron 2 at the 5' end of leader 3.

Cleavage at the acceptor site of intron 3 can either be rapid (at 66.6 m.u.) or relatively late (at 68.0 m.u.). Rapid cleavages are introduced in introns 1 and 2 at the level of the nascent transcript (7). Cleavage at the acceptor site for the longest L4-mRNA body can occur before complete removal of intron 2, which takes about 12 to 15 min (3,7). This indicates that cleavage events may take place simultaneously at sites which are separated by 50% of the genome, i.e. about 18000 nucleotides. The cleaved-off RNA pieces have to be kept together in the right orientation in order to allow further processing and ligation of the exons. The nuclear matrix as a higher ordered intranuclear structure meets this requirement since all cleavage products are found to be associated with the matrix (17, chapter III and V). It is, therefore, possible that the nuclear matrix indeed plays an essential role in RNA splicing.

It should be noted that, although rapid cleavages in the leader region occur in the majority of the primary transcripts

(7), there are some transcripts which seem to be exempted from rapid cleavage in this region. Therefore, the possibility remains that rapid cleavage at an acceptor site of intron 3 occurs only in primary transcripts which are not yet cleaved in the leader region. In this case, however, there has to exist a mechanism which can discriminate between early and late cleavage at a particular acceptor site of intron 3 .

The cleavage reactions around the third leader occur in an ordered fashion. The acceptor site of intron 2 is cleaved before the donor site of intron 3. Together with the data of Keohavong et al (3) our earlier results indicated that cleavage at the acceptor site of intron 2, 12 to 15 min after the start of transcription, is the last step in the excision of this intron and is rapidly followed by the ligation of leaders 2 and 3. Accordingly, our present results suggest that intron 2 is removed before intron 3 and that ligation of leaders 2 and 3 precedes ligation of leader 3 to the body of the mRNA. Since ligation of leaders 1 and 2 occurs without detectable delay after the start of transcription, the ligation reactions involved in the production of the major late mRNAs seem to follow the 5' to 3' direction.

When the excision of intron 3 is considered, acceptor site cleavage can precede cleavage of the donor site. Rapid cleavage at the acceptor site and possible other rapid cleavages occuring in intron 3, provides an explanation for the relatively low amount of poly(A)-containing RNA in rapidly labeled nuclear RNA isolated from late infected cells. Further it shows that removal of intron 3 may start at the 3' end of the intron. However, when splicing patterns of Ad2 RNA in monkey cells were studied, L5-specific RNAs were found to accumulate, which still contained the L4-specific sequences between 66 and 78 m.u. (18). Although such RNA species were not observed in human cells, they may be regarded as intermediates of splicing. Therefore, the site located at 66.6 may either be used as an acceptor site for L4-mRNAs or as an internal cleavage site for the production of L5-mRNAs and rapid cleavage at 66.6 m.u. may be related to one of these functions.

ACKNOWLEDGEMENTS

We thank A. Groeneveld for culturing the cells.

This work was supported by the Netherlands Foudation for Chemical Research (S.O.N.) and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

1. Nevins, J.R. (1979) J.Mol.Biol. 130, 493-506.
2. Nevins, J.R. and Darnell, J.E. (1978) Cell 15, 1477-1493.
3. Keohavong, O., Gattoni, R., LeMoullec, J.M., Jacob, M.and Stevenin, J. (1982) Nucleic Acids Res. 10, 1215-1229.
4. van Eekelen, C.A.G., Mariman, E.C.M., Reinders, R.J. and van Venrooij, W.J. (1981) Eur.J.Biochem. 119, 461-467.
5. van Eekelen, C.A.G. and van Venrooij, W.J. (1981) J.Cell Biol. 88, 554-563.
6. Mariman, E.C.M., Hagebols, A-M. and van Venrooij, W.J. (1982) Nucleic Acids Res. 10, 6131-6145.
7. Mariman, E.C.M., van Beek-Reinders, R.J. and van Venrooij, W.J. (1983) J.Mol.Biol. 163, 239-256.
8. Long, B.H., Huang, C-Y.and Pogo, A.O. (1979) Cell 18, 1079-1090.
9. Alestrom, P., Akusjärvi, G., Perricaudet, M., Mathews, M., Klessig, D.F. and Pettersson, U. (1980) Cell 19, 671-681.
10. Manley, J.L., Sharp, P.A. and Gefter, M.L. (1979) J.Mol. Biol. 135, 171-197.
11. Darnell, J.E., Girard, M., Baltimore, D., Summers, D.F. and Maizel, J.V. (1967) The Molecular Biology of Viruses, J.Colter, ed. (New York:Acad. Press), 375-401.
12. Seghal, P., Derman, E., Molloy, G.R., Tamm, I and Darnell, J.E. (1976) Science 194, 431-433.
13. Greenberg, H. and Penman S. (1966) J.Mol.Biol. 21, 527-535.
14. Tamm, I., Kikuchi, T., Darnell, J.E.and Salditt-Georgieff, M. (1980) Biochemistry 19, 2743-2748.
15. Chen-Kiang, S., Wolgemuth, D.J., Hsu, M-T. and Darnell, J.E. (1982) Cell 12, 1-8.
16. Fraser, N.W., Seghal, P.B. and Darnell, J.E. (1979) Proc. Natl.Acad.Sci.USA 76, 2571-2575.
17. Mariman, E.C.M., van Eekelen, C.A.G., Reinders, R.J., Berns, A.J.M. and van Venrooy, W.J. (1982) J.Mol.Biol. 154, 103-119.
18. Klessig, D.F. and Chow, L.T. (1980) J.Mol.Biol. 139, 221-242.

CHAPTER V

ADENOVIRAL HETEROGENEOUS NUCLEAR RNA IS ASSOCIATED WITH THE NUCLEAR MATRIX DURING SPLICING

Adenoviral heterogeneous nuclear RNA is associated with the host nuclear matrix during splicing
E.C.M. Mariman, C.A.G. van Eekelen, R.J. Reinders, A.J.M. Berns and W.J. van Venrooij (1982) J.Mol.Biol. 154, 103-119.

J Mol Biol (1982) **154**, 103–119

Adenoviral Heterogeneous Nuclear RNA is Associated with the Host Nuclear Matrix During Splicing

EDWIN C M MARIMAN, CHRIS A G VAN EEKELEN, RITA J REINDERS
ANTON J M BERNS AND WALTHER J VAN VENROOIJ

Department of Biochemistry, University of Nijmegen
Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands

(Received 17 July 1981, and in revised form 11 September 1981)

After infection of HeLa cells with adenovirus type 2, virus specific heterogeneous nuclear RNA is quantitatively associated with a higher ordered structure, the nuclear matrix Analysis of this matrix associated RNA by S_1 nuclease mapping showed that precursors as well as processed messenger RNAs from the late region L4 were present By irradiation of intact cells with ultraviolet light, proteins tightly associated with heterogeneous nuclear RNA can be induced to cross link with the RNA Characterization of the cross linked RNA–protein complexes showed that all viral polyadenylated RNAs (precursors products and processing intermediates) could be cross linked to two host proteins, earlier found to be involved in the association of host specific heterogeneous nuclear RNA to the nuclear matrix (van Eekelen & van Venrooij, 1981) Our results thus further support the concept that the nuclear matrix may function in the localization and the structural organization of (viral) heterogeneous nuclear RNA during its processing

1. Introduction

Adenovirus infected cells have proven to be an efficient tool in the study of hnRNA† processing The virus specific RNA is capped, methylated, polyadenylated and spliced in much the same way as the hnRNA of eukaryotic cells (Darnell, 1979, Ziff, 1980) By methods of hybridization with specific parts of the viral genomic DNA, individual mRNA precursors can be identified and pursued during processing In this way the various steps of processing of pre-mRNA to mature mRNA can be studied independently (Blanchard *et al*, 1978, Babich *et al*, 1980, Nevins *et al*, 1980)

There is not much known about the localization and the structural organization of hnRNA in the nucleus hnRNA is associated with a characteristic subset of nuclear proteins and it can be isolated as hnRNP particles In addition, the cell nucleus seems to contain a higher ordered proteinaceous structure (Herman *et al*,

† Abbreviations used hnRNA, heterogeneous nuclear RNA, hnRNP heterogeneous nuclear ribonucleoprotein, DBP, DNA binding protein snRNA small nuclear RNA

0022-2836/82/010103-17 $02 00/0

1978; Kaufmann *et al.*, 1981; Miller *et al.*, 1978*a*; van Eekelen & van Venrooij, 1981), usually referred to as the nuclear matrix. Heterogeneous nuclear RNA is associated with the nuclear matrix (Herman *et al.*, 1978; Miller *et al.*, 1978*a*), and it has been reported that two specific hnRNP proteins are involved in this association (van Eekelen & van Venrooij, 1981). In this study we present evidence that adenoviral-specific nuclear matrix RNA contains precursors, intermediates and products of RNA processing, indicating that pre-mRNA is associated with the nuclear matrix during its processing to mRNA.

2. Materials and Methods

(a) *Infection and labelling of cells*

HeLa S3 cells were infected with adenovirus type 2 (2000 particles/cell) at a density of 5×10^6 cells/ml in MEM (Eagles minimal essential medium containing 1 mM-arginine). After adsorption of virus to the cells during 1 h at 37°C, the cells were diluted to $0{\cdot}3 \times 10^6$ cells/ml with MEM containing 5% (v/v) newborn calf serum. Cells were harvested at 18 h post infection. To obtain steady-state labelled RNA, cells were concentrated to a density of 5×10^6/ml and incubated with 10 μCi [5,6-^{3}H]uridine/ml (45 Ci/mmol; Amersham, England) from 16 to 18 h after infection. To obtain rapidly labelled hnRNA 18 h post infection, cells were concentrated to 20×10^6/ml and incubated for 10 min with 20 μCi [5,6-^{3}H]uridine/ml and 20 μCi [5-^{3}H]cytidine/ml (31 Ci/mmol; Amersham, England).

(b) *Purification and oligo(dT)-cellulose chromatography of cytoplasmic and nuclear RNA*

At 18 h after infection, cells were harvested on frozen NKM solution (0·13 M-NaCl, 0·05 M-KCl, 1·5 mM-$MgCl_2$) and centrifuged for 5 min at 800 *g*. The cells were washed once with NKM solution and, after pelleting, they were resuspended in RSB (10 mM-NaCl, 10 mM-Tris (pH 7·4), 1·5 mM-$MgCl_2$). After addition of 0·1 vol. 5% sodium deoxycholate, 5% (w/v) Nonidet P40 (Fluka, Buchs, Switzerland), the cells were disrupted by intensive vortexing. The nuclei were pelleted and, from the supernatant, cytoplasmic RNA was isolated by extraction with phenol/chloroform as described by Palmiter (1973). The nuclei were washed once in RSB and resuspended in the same buffer. Nuclear RNA was then isolated by extraction with phenol at 55°C as described by Long *et al.* (1979). To select the poly (A)-containing RNA, oligo(dT)-cellulose chromatography (Collaborative Research Inc.; grade T2) was performed as described by Aviv & Leder (1972).

(c) *Preparation of nuclear matrix RNA*

At 18 h after infection of HeLa cells with adenovirus type 2, nuclei were isolated as described above. The nuclei were resuspended in RSB containing 0·1 M-NaCl and treated with DNAase I for 30 min at 10°C (500 μg/ml). After digestion with DNAase I, the nuclei were pelleted through a 30% (w/v) sucrose layer. The pelleted material was extracted once with 0·4 M-$(NH_4)_2SO_4$. After repelleting, the nuclear matrices were washed with RSB, suspended in RSB containing 1% (w/v) sodium dodecyl sulphate and incubated at 100°C for 3 min. The boiled mixture was adjusted to 0·2 M-NaCl and extracted with saturated phenol/chloroform (1·1, v/v). After 2 additional extractions with chloroform, the RNA was precipitated with 2 vol. ethanol.

(d) *RNA–protein cross-linking* in vivo *and subsequent isolation of the cross-linked RNA*

RNA was cross-linked to its associated proteins by irradiation of the cells for 5 min with ultraviolet light (254 nm) at 0°C according to the procedure described in detail by Wagenmakers *et al.* (1980). After irradiation, the nuclear matrices were isolated as described

above. The covalently linked RNA–protein complexes were isolated using the phenol/chloroform extraction procedure as described (van Eekelen & van Venrooij, 1981; Wagenmakers *et al.*, 1980).

(e) *Characterization of the proteins cross-linked to RNA*

The covalently linked RNA–protein complexes were isolated as described above and the RNA moiety was digested with a mixture of ribonuclease A and micrococcal nuclease The proteins were analyzed by sodium dodecyl sulphate/polyacrylamide gel electrophoresis as described (van Eekelen & van Venrooij, 1981)

(f) *Adenoviral DNA isolation and restriction endonuclease digestions*

Viral DNA was isolated from the purified virions as described by Petterson & Sambrook (1973). Restriction digestions were performed at 37°C for 4 h using 1 unit of enzyme per μg DNA. Standard conditions of *Eco*RI, *Hin*dIII and *Bgl*II (Boehringer, Mannheim) digestions were 50 mM-Tris (pH 7·4), 7 mM-$MgCl_2$, 50 mM-NaCl, 3 mM-dithiothreitol. *Kpn*I digestion (Bethesda Research Laboratories Inc) was performed under low salt conditions (10 mM-Tris (pH 7·4), 7 mM-$MgCl_2$, 10 mM-KCl, 3 mM-dithiothreitol). Restriction fragments were separated by agarose gel electrophoresis; 0·8% (w/v) agarose gels were run for 20 h at 80 V in TNE buffer (40 mM-Tris (pH 8), 8 mM-sodium acetate, 2 mM-EDTA).

(g) *S_1-nuclease mapping and analysis of S_1-resistant hybrids*

RNA was hybridized to DNA under conditions in which DNA-DNA hybridization was minimized, using the procedure described by Manley *et al.* (1979). After digestion with S_1 nuclease, the hybrids were precipitated with ethanol, the pellet was solubilized in TNE buffer and applied to a 1·2% (w/v) agarose gel. Electrophoresis was performed under the same conditions as those used for the separation of DNA restriction fragments. Gels were subsequently treated with 5% (w/v) trichloroacetic acid, methanol, methanol containing 10% (w/v) PPO, and water. After drying, they were exposed to Kodak XR-5 film for an appropriate period of time (3 to 7 days).

3. Results

(a) *The nuclear matrix*

The nuclear matrix of HeLa cells is a higher ordered proteinaceous structure to which the rapidly labelled hnRNA is quantitatively bound (Herman *et al.*, 1978; Miller *et al.*, 1978*a*; van Eekelen & van Venrooij, 1981). The isolation of the nuclear matrix generally involves digestion of the DNA with purified DNAase I and removal of the digested DNA together with its associated proteins. In our procedure, this is achieved by sedimenting the digested nuclei through a 30% (w/v) sucrose layer followed by an extraction with a high salt buffer. Late after infection, the nuclei are somewhat more fragile as compared to nuclei from uninfected cells. This is probably due to the structural changes caused by the accumulation of virus particles. Although more fragile, intact nuclei and matrices can be isolated 18 hours post infection. The newly replicated viral DNA that has already been encapsidated by the viral protein coat seems to be relatively resistant to the DNAase I. Most of the virions, however, are removed by the high salt extraction, so that the resulting matrices contain only a small number of virus particles as determined by electron microscopy (van Eekelen *et al.*, 1981). Less than 1% of the host DNA is left in the

matrices. As in uninfected cells, hnRNA from infected cells remains associated with the nuclear matrix. About 95% of the viral-specific RNA is associated with the nuclear matrix. In agreement with earlier observations, we have found that the presence of inhibitors of proteolytic activities is a prerequisite for the isolation of intact matrices and undegraded hnRNA. The integrity of the nuclear matrix seems to protect the RNA against degradation.

It has been reported by Beltz & Flint (1979) that after infection of HeLa cells with adenovirus the nuclear RNA becomes only partly viral specific, because transcription of cellular sequences continues. Liquid hybridization studies have confirmed that nuclear matrix-associated hnRNA of infected cells is a mixture of viral and host RNA. It was found that, although 18 hours post infection the concentration of host poly(A)$^+$ mRNA in the cytoplasm has decreased to less than 20% of the level found in the uninfected cell, the nuclear poly(A)$^+$ hnRNA still contains about the same proportion of host mRNA sequences (van Eekelen *et al.*, 1981).

The gradient profiles of total nuclear and nuclear matrix RNA show a close resemblance (Fig. 1(a)). We have verified the presence of viral-specific sequences in nuclear matrix RNA by hybridization with adenovirus DNA fragments. Figure 1(b) shows the profile of matrix RNA containing nucleotide sequences complementary to the DNA strands of the *Eco*RI restriction fragment B (58·5 to 70·7 map units). RNA species as large as 45 S could be detected, probably representing the precursors for the region L5 messenger RNAs, extending from 16·45 map units towards the right end of the viral genome.

(b) *S_1 nuclease mapping of poly(A)$^+$ cytoplasmic and nuclear RNA with the* Eco*RI B restriction fragment*

In order to determine the complement of RNA molecules bound to the nuclear matrix, it was necessary to analyse first the complement of RNA species present in cytoplasmic and nuclear extracts. For this purpose, an S_1 nuclease mapping procedure was used, as described in Materials and Methods. When poly(A)$^+$ cytoplasmic RNA, labelled from 16 to 18 hours post infection, was hybridized to the *Hin*dIII A fragment of adenoviral DNA, two major bands were detected, which were 11·3% and 6·3% in length (Fig. 2(a), lane 1; 1% equals 350 base-pairs). This would correspond to the length of hybrids derived from L3 and L4 messengers (Berget *et al.*, 1977; Chow *et al.*, 1977; Chow & Broker, 1978; Berget & Sharp, 1979). Additional evidence was obtained when smaller DNA fragments, as depicted in Figure 2(b), were used for hybridization. The *Eco*RI B fragment gives two major bands, numbered 5 and 6 (Fig. 2(a), lane 2). The larger one, band number 5, is 4·1% in length and was also found when subfragment b or e was used (Fig. 2(a), lanes 3 and 7). This hybrid thus contains the 5′ end of the L4 messenger RNA and should therefore start at 66·6 ± 0·2 map units. The co-ordinate of the 5′ end of the messenger RNA for the DNA binding protein, which is transcribed from the opposite strand, has been determined to reside at 66·5 map units (Chow *et al.*, 1979). Obviously, the 5′ ends of the L4 and DBP messenger RNAs are very close,

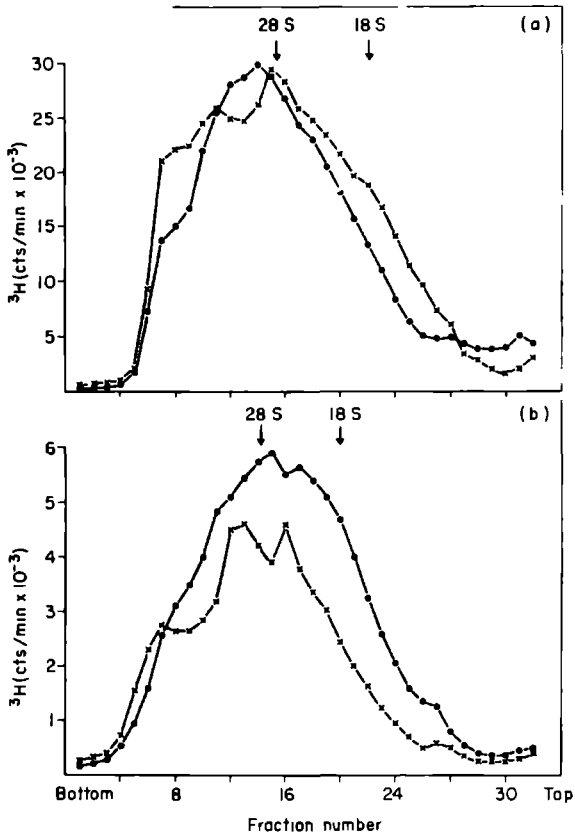


FIG 1 Sucrose gradient profiles of rapidly labelled hnRNA isolated from nuclei and nuclear matrices Nuclei were isolated from 10 min labelled HeLa cells (see Materials and Methods) 18 h after infection One portion of the nuclei was used for the isolation of hnRNA From the other batch of nuclei nuclear matrices were isolated and the RNA was purified from the matrices as described in Materials and Methods The RNA was dissolved in a buffer of low ionic strength (10 mM Tris (pH 7 4) 2 mM EDTA 0 2% (w/v) sodium dodecyl sulphate) heated to 100 C for a few minutes and after rapid cooling the solution was layered on top of a 16% to 43 5% (w v) isokinetic sucrose gradient containing the same buffer Centrifugation was for about 16 h at 20°C in an IEC SB283 rotor at 200 000 *g* The gradients were fractionated and radioactivity was counted using Pico fluor™ 15 (Packard Inst) (a) (●) Nuclear RNA, (×) nuclear matrix RNA After fractionation the nuclear matrix RNA was precipitated and redissolved in 200 μl of 3 × SSC containing 1 mM EDTA 0 1% (w/v) sodium dodecyl sulphate 50% formamide In this solution a nitrocellulose filter to which 2 μg equivalents of *Eco*RI B fragment was immobilized (Gillespie & Gillespie, 1971) was incubated for 2 min at 65 C and subsequently for 60 h at 37 C Filters were washed several times with 2 × SSC dried and the radioactivity of the hybridized RNA was counted in a toluene based scintillant (SSC is 0 15 M NaCl 0·015 M sodium citrate, pH 7) (b) (●) Nuclear matrix RNA, (×) matrix RNA hybridized to the *Eco*RI B fragment

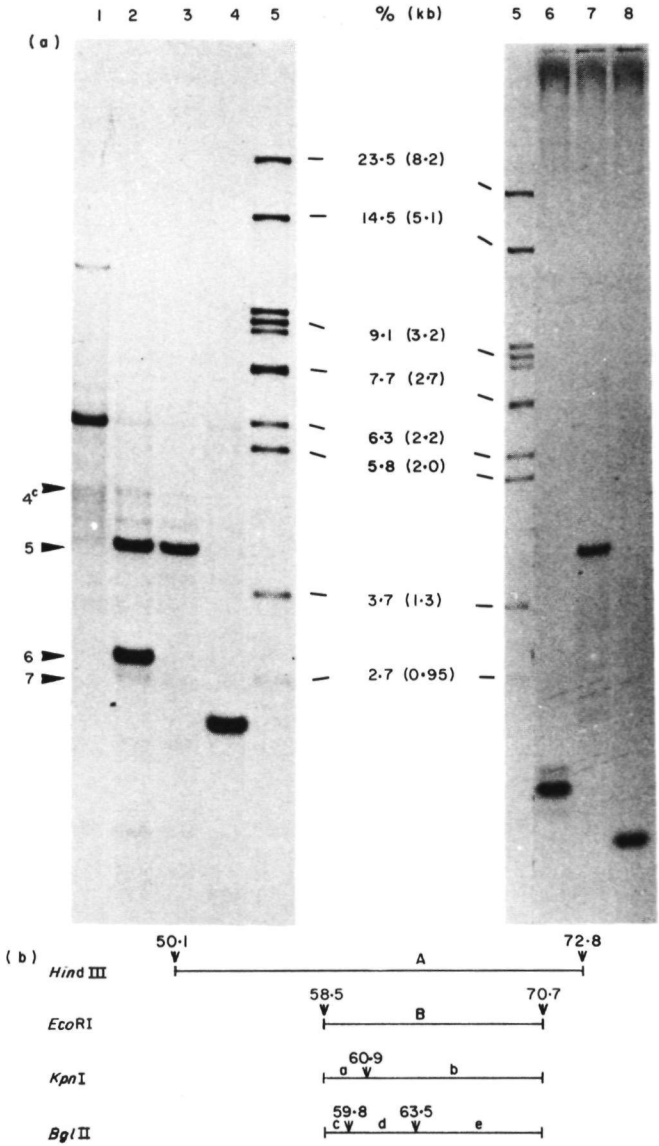


FIG. 2. S_1 nuclease analysis of poly(A)-containing cytoplasmic RNA. HeLa cells were labelled with [^{3}H]uridine from 16 to 18 h after infection. Cytoplasmic extracts were prepared as described in Materials and Methods, and cytoplasmic mRNA purified *via* oligo(dT)-cellulose chromatography (Aviv & Leder, 1972). The poly(A)$^+$ mRNA was hybridized to various DNA fragments ((b)) and analysed according to the S_1 mapping procedure as described in Materials and Methods. The Figure shows the S_1-resistant hybrids of poly(A)$^+$ mRNA with: lane 1, *Hin*dIII A; lane 2, *Eco*RI B; lane 3, subfragment b; lane 4, subfragment a; lane 6, subfragment d; lane 7, subfragment e; lane 8, subfragment c; lane 5 shows the *Hin*dIII restriction fragments of adenovirus type 2 DNA as molecular weight markers. The hybrids in lane 2 (arrows) have been numbered for reference. kb, 10^3 bases. (b) Location of restriction fragments and subfragments of *Eco*RI B on the adenovirus type 2 genome. The molecular weight of the subfragments were determined on agarose gels with *Hin*dIII and *Eco*RI restriction digests of adenovirus type 2 DNA as molecular weight markers. From these molecular weights, the restriction sites within the *Eco*RI B fragment were calculated.

indicating that the viral genome sequences are most economically used without large spacer regions between individual genes. This has recently been confirmed by sequence studies reported by Akusjarvi *et al* (1981). The second band obtained with the *Eco*RI B fragment (Fig. 2(a) lane 2, band number 6), which is 2·8% long, was not found with any of the subfragments. Therefore it probably represents the 3′ end of the region L3 messenger RNA which maps within the region of subfragment d at 61·4±0·2 map units (Fig. 2(a) lane 6).

Except for these major components, a few minor RNA species were found to hybridize to the *Eco*RI B fragment. A hybrid numbered 7, of 2·7%, can be identified with subfragments b and e (Fig. 2(a) lanes 3 and 7) but not with *Hin*dIII A (Fig. 2(a) lane 1). Therefore this hybrid contains the 5′ end of the region L4 messenger RNA which has been mapped at 68·0 map units (Chow *et al* 1977; Chow & Broker 1978). Using our procedure the 5′ end of this RNA was found to be located at the same co-ordinate 68·0±0·2 map units. Band number 4^c (Fig. 2(a) lane 2), representing a hybrid of 4·9%, is probably derived from one of the described DBP mRNAs (61·6 to 66·5 map units; Chow *et al* 1979), but the concentration of this RNA in the cytoplasm late after infection is too low for accurate mapping with our procedure. This hybrid has the same length as a hybrid formed from a typical nuclear RNA species (band 4^n, see below) but can be distinguished from the nuclear species as will be described below. When long exposure times were used, some additional hybrids of relatively low molecular weight were detected. These probably contain RNA species present in the cytoplasm in very low quantities. The existence of one such messenger RNA has been deduced from sequence data of the *Eco*RI B fragment of adenovirus type 2 (Akusjarvi *et al* 1981). The body of this L3 messenger maps between 59·9 and 61·8 map units.

We then undertook the analysis of total nuclear RNA. As expected, a more complex pattern of hybrids was obtained, although the species present in the cytoplasmic RNA (band numbers 4, 5, 6 and 7) could also be detected in steady-state nuclear RNA (compare Figs. 2(a) and 3). It should be mentioned here that we never found any difference between the pattern of total nuclear RNA and that of poly(A)-selected RNA. Most experiments, however, were performed with poly(A)-containing nuclear RNA. The long precursors of the region L4 and L5 mRNAs protect the *Eco*RI B fragment totally against degradation by the S_1 nuclease (12·2%, Fig. 3, lane 2, band number 1). In the case of the DBP precursor RNA one expects to find a hybrid that is 9·1% (61·6 to 70·7 map units). Indeed, a band of 8·9±0·2% was detected with the *Eco*RI B fragment (Fig. 3, lane 2, band number 2) or subfragment b as probe (Fig. 3, lane 3).

Besides hybrids that can be ascribed to known pre-mRNAs (bands 1 and 2) and mRNA (bands 4^c, 5, 6 and 7), a few more hybrid bands (numbered 3A+B and 4^n) were detected. These were not the result of contamination of *Eco*RI B with other *Eco*RI restriction fragments, since these bands were not found when other *Eco*RI fragments were hybridized to nuclear RNA (Fig. 4, lanes 2 to 5). Furthermore the *Eco*RI B fragment isolated from cloned *Hin*dIII A, a generous gift from Dr M. Matthews, gives an identical pattern of hybrids as the fragment isolated from agarose gels (data not shown). This indicates that the three bands (numbers 4^n

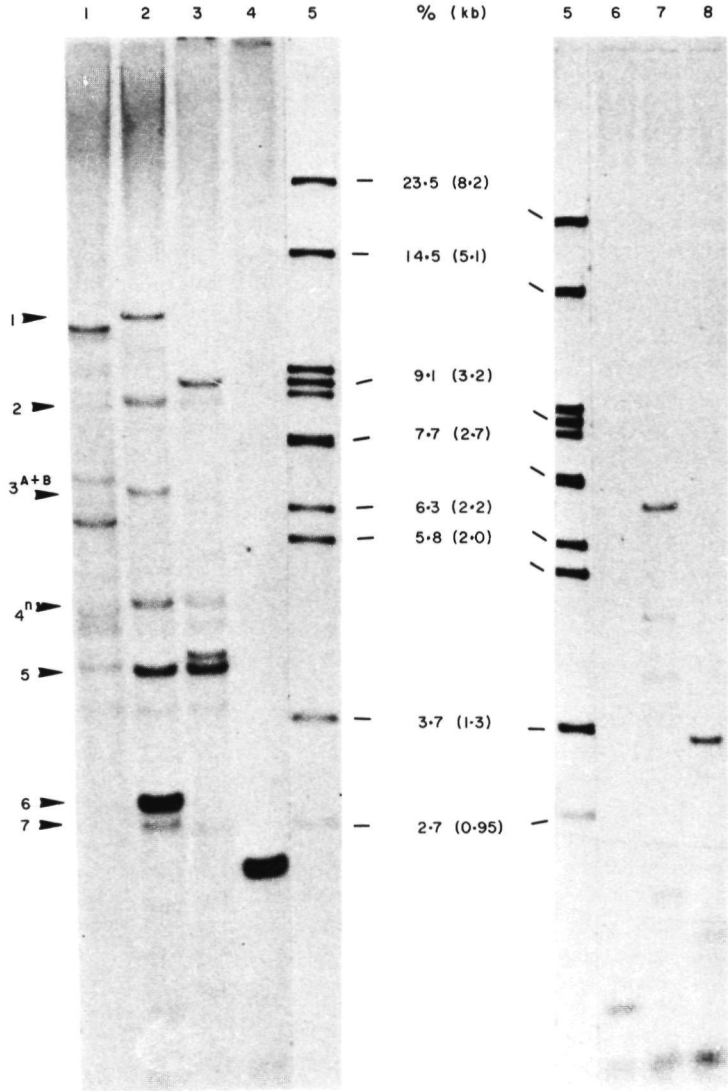


FIG. 3. S_1 nuclease analysis of poly(A)-containing nuclear RNA. HeLa cells were labelled with [^{3}H]uridine from 16 to 18 h after infection. Nuclei were isolated as described in Materials and Methods. Nuclear RNA was prepared by extraction with hot phenol (Long *et al.*, 1979) and purified *via* oligo(dT)-cellulose chromatography (Aviv & Leder, 1972). The poly(A)$^+$ nuclear RNA was hybridized to various DNA fragments (Fig. 2(b)) and analysed according to the S_1 mapping procedure as described in Materials and Methods. The Figure shows the S_1-resistant hybrids of poly(A)$^+$ nuclear RNA with: lane 1, *Hin*dIII A; lane 2, *Eco*RI B; lane 3, subfragment b; lane 4, subfragment a; lane 6, subfragment c; lane 7, subfragment e; lane 8, subfragment d; lane 5 shows the *Hin*dIII molecular weight markers. The hybrids in lane 2 (arrows) have been numbered for reference. kb, 10^3 bases.

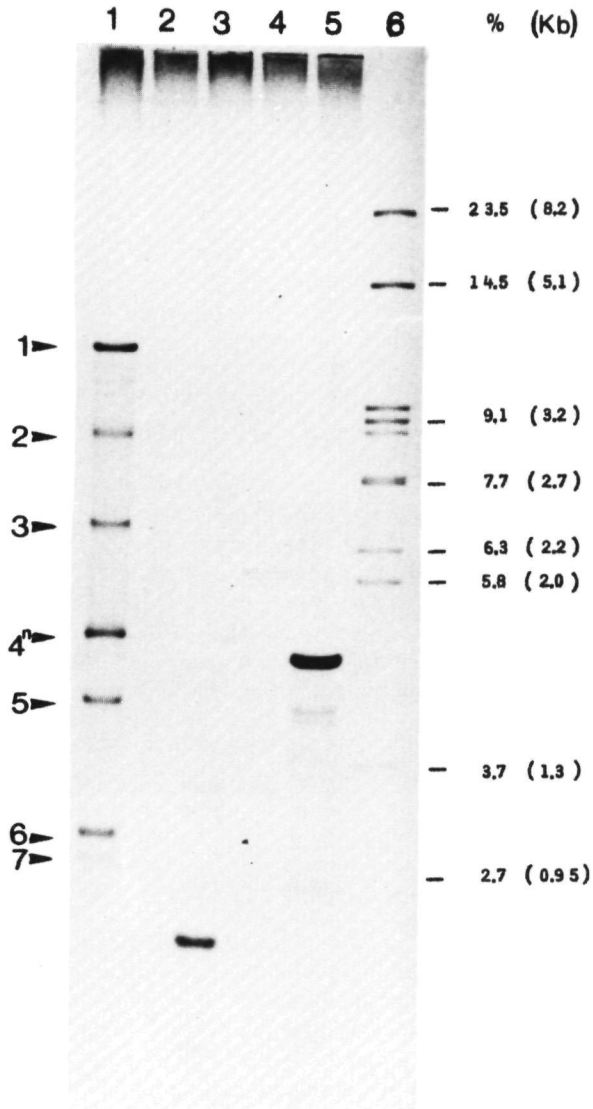


FIG. 4. S_1 nuclease analysis of poly(A)-containing nuclear RNA with various *Eco*RI restriction fragments. Infected HeLa cells were labelled and poly(A)$^+$ nuclear RNA was isolated as indicated in the legend to Fig. 3. The poly(A)$^+$ nuclear RNA was hybridized to various *Eco*RI restriction fragments of adenovirus type 2 DNA and analysed according to the S_1 mapping procedure as described in Materials and Methods. The Figure shows the S_1-resistant hybrids of poly(A)$^+$ nuclear RNA with: lane 1, *Eco*RI B; lane 2, *Eco*RI C; lane 3, *Eco*RI D; lane 4, *Eco*RI E; lane 5, *Eco*RI F; lane 6 shows the *Hin*dIII molecular weight markers. The hybrids in lane 1 (arrows) have been numbered for reference. kb, 10^3 bases.

3A+B) represent RNA molecules from which at least one end maps within the *Eco*RI B fragment. Hybridization of nuclear RNA to the *Eco*RI B subfragments (Fig. 2(b)) provides a better localization of these species. The band of 6·8% was found to consist of two components, 3A and 3B. The RNA in hybrid 3A yielded with subfragment b a band of 4·3%, which is not found with *Eco*RI B (Fig. 3, lanes 2 and 3). This indicates that the RNA extends from left of 58·5 to 65·2±0·2 map units. This is further substantiated by hybridization with subfragment e, which yielded the expected fragment of 1·8% (Fig. 3, lane 7). It is possible that this species represents a partly spliced poly(A)$^+$ RNA molecule from which an intron extending between 65·2 and a co-ordinate to the right of 70·7 has been removed. An alternative possibility is the existence of an additional polyadenylation site at 65·2 map units. Hybrid band 3B was also found with subfragment b (Fig. 3, lane 3). Therefore, this hybrid most likely represents a splicing intermediate which is generated by the excision of the 5′ intron of the DBP precursor RNA (61·6 to 68·6 map units; Blanchard *et al.* 1978; Goldenberg & Raskas, 1979). However, accurate mapping was not possible because of the relatively low concentration of this intermediate.

The 4·9% long hybrid band number 4^n (Fig. 3, lane 2) is probably not due to accumulation of processed DBP RNA in the nucleus because no 3·0% band could be detected with subfragment e (Fig. 3, lane 7). Furthermore, band number 4^n is also found with subfragments b and e but not with *Hin*dIII A (Fig. 3, lanes 1, 3 and 7). Thus, the RNA extends from 65·8±0·2 map units to beyond 70·7 and could therefore represent processing intermediates from region L5 RNAs. They resemble molecules that constitute a large part of region L5 messenger RNAs in abortive infection (Klessig & Chow, 1980). Finally, the hybrid numbered 6 (2·8%, Fig. 3, lane 2) is probably generated both from the 3′ end of the region L3 messengers, as discussed above, as well as from their polyadenylated precursors which extend between 16·45 and 61·5 map units. The results of our analyses are summarized in Figure 5.

(c) *S_1 nuclease analysis of nuclear matrix RNA using the* Eco*RI B fragment as a probe*

The data summarized in Figure 5 show that precursors (bands number 1 and 2) and processing intermediates (bands number 3A, 3B and 4^n) and mRNAs (bands number 4^c, 5, 6 and 7) are present in nuclear RNA. It would be interesting to know which of these RNA species are associated with the nuclear matrix. For this reason rapidly labelled nuclear matrix RNA was isolated and hybridized to the various DNA fragments shown in Figure 2(b). As determined for RNA sequences that are transcribed from the genome between co-ordinates 50·1 and 72·8, the S_1 nuclease mapping results indicate (compare Figs. 3 and 6) that the matrix RNA most likely has the same composition as steady-state poly(A)$^+$ nuclear RNA. Therefore, all the RNA species present in the poly(A)$^+$ nuclear RNA, i.e. precursors, intermediates and products of RNA processing, are associated in some way with the nuclear matrix. To exclude the possibility that the products of processing that were found in the matrix RNA were the result of cytoplasmic contamination, the poly(A)$^+$

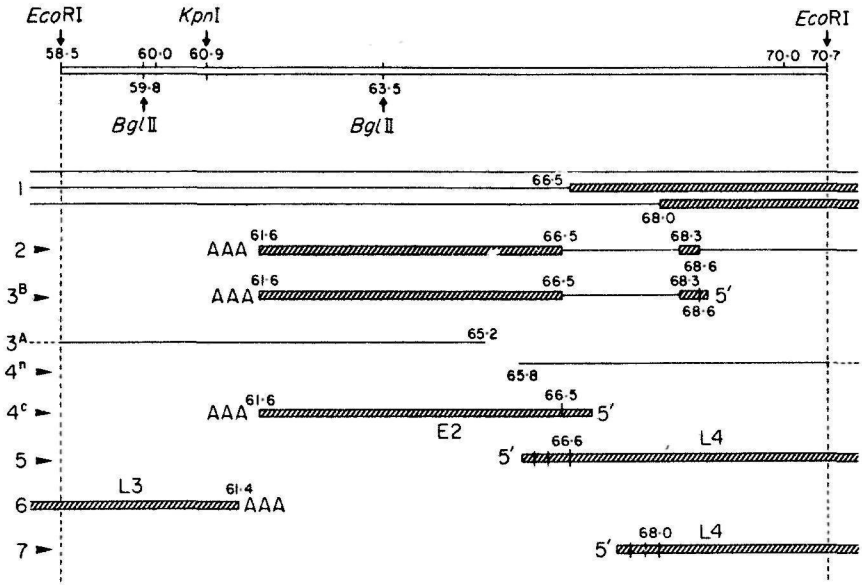


FIG. 5. Summary of the data obtained *via* S_1 mapping experiments. RNA species complementary to the *Eco*RI B fragment found in the nucleus of HeLa cells at the late stage of infection with adenovirus type 2. Detection was performed by means of the S_1 nuclease mapping procedure as described in Materials and Methods. The hatched bars represent the mRNA sequence, the lines indicate the intervening sequence.

cytoplasmic RNA from the same batch of cells was analysed as well. No labelled S_1-resistant hybrids could be detected (Fig. 6, lane 8). Even after labelling the cells for 20 minutes, only very low levels of labelled mRNAs were present in the cytoplasm. This excludes the possibility of contamination and strongly suggests that pre-mRNA is associated with the nuclear matrix during its processing.

(d) *Viral RNA is associated with host proteins during processing*

Recent studies have indicated that hnRNA is bound to the nuclear matrix *via* two proteins of M_r 41,500 and 43,000. The evidence underlying this conclusion was based on the findings that these proteins were tightly bound to rapidly labelled hnRNA *in vivo*, and that they were not released from the matrices during exhaustive treatment with ribonuclease, which does remove other hnRNA-associated proteins (van Eekelen & van Venrooij, 1981). In order to show that adenovirus-specific sequences were bound to the nuclear matrix in a similar way as host hnRNA, the ultraviolet light cross-linking method (Wagenmakers *et al.*, 1980) was applied to infected cells. The covalently linked RNA–protein complexes were separated from non-cross-linked complexes by boiling the matrices in 1% (w/v) sodium dodecyl sulphate followed by extraction with phenol/chloroform. Free RNA remains in the aqueous phase, while covalent RNA–protein complexes move

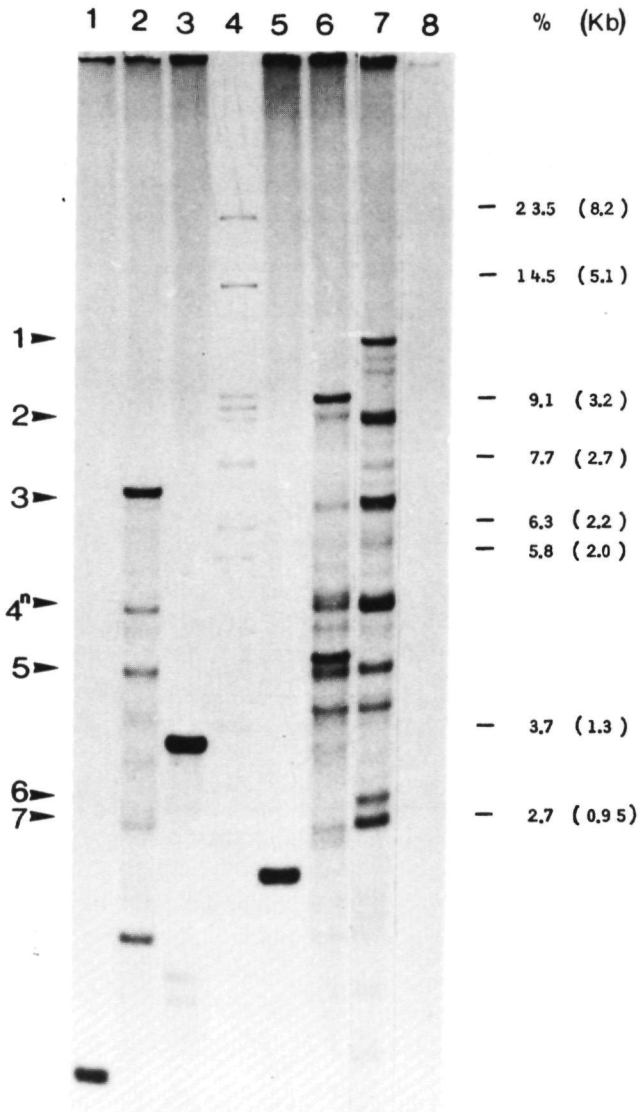


FIG. 6. S_1 nuclease analysis of nuclear matrix RNA. HeLa cells were labelled for 10 min with [^{3}H]uridine and [^{3}H]cytidine 18 h after infection. Nuclear matrices were prepared as described in Materials and Methods. The associated RNA was isolated by boiling the matrices in 1% (w/v) sodium dodecyl sulphate for a few minutes, followed by extraction with phenol/chloroform and 2 extractions with chloroform. The RNA was precipitated with ethanol. pelleted and hybridized to various DNA fragments (Fig. 2(b)). Analysis was according to the S_1 mapping procedure as described in Materials and Methods. The Figure shows the S_1-resistant hybrids of total nuclear matrix RNA with: lane 1, subfragment. c; lane 2, subfragment e; lane 3, subfragment d; lane 5, subfragment a; lane 6, subfragment b; lane 7, *Eco*RI B; lane 8, S_1-resistant hybrids of poly(A)$^+$ cytoplasmic RNA from 10 min labelled infected HeLa cells. hybridized to *Eco*RI B. Lane 4 shows the *Hin*dIII molecular weight markers. The hybrids in lane 7 have been numbered (arrows) for reference. kb, 10^3 bases.

into the interphase. After irradiating the cells for five minutes, 55 to 80% of the nuclear matrix RNA was found in the interphase fraction. In control experiments in which the cells were not irradiated, only 3 to 7% of the RNA was found in the interphase. To determine which RNA species were cross-linked, the RNA–protein complexes in the interphase and organic layer were precipitated with ethanol and deproteinized by extensive treatment with proteinase K. The RNA was then analysed by the S_1 nuclease mapping procedure using *Eco*RI B as a probe. Our results showed that all the RNA species that were originally present in the nuclear matrix RNA preparation (Fig. 6) were also present in the cross-linked RNA fraction (Fig. 7(a)).

We also analysed the proteins that were cross-linked to the matrix-associated RNA of infected cells. For this purpose, at 18 hours after infection cells were labelled for ten minutes with [^{3}H]uridine and [^{3}H]cytidine, harvested and irradiated. The nuclear matrices were dissolved in sodium dodecyl sulphate and the covalent hnRNA–protein complexes were isolated by oligo(dT)-cellulose chromatography and digested extensively with nucleases (van Eekelen & van Venrooij, 1981). By virtue of the few remaining ^{3}H-labelled nucleotides, the proteins that were cross-linked to poly(A)$^+$ hnRNA were identified by gel electrophoresis and fluorography (Wagenmakers *et al.*, 1980). The result of such an analysis is depicted in Figure 7(b), which shows that two proteins (M_r 41,500 and 43,000) were cross-linked to rapidly labelled hnRNA from infected cells. Since the cross-linked RNA probably is a mixture of viral and host-specific sequences (Beltz & Flint, 1979), and the cross-linked proteins comigrate exactly with the proteins cross-linked to hnRNA (van Eekelen & van Venrooij, 1981), we also have selected the viral-specific complexes by hybridizing the RNA–protein complexes to viral DNA immobilized on Sepharose. The analysis of these viral-specific RNA–protein complexes gave the same result as is shown in Figure 7(b).

In conclusion, we may state that all viral RNA species (precursors, intermediates and processed mRNAs) in the nucleus are bound to the nuclear matrix. They can be cross-linked *in vivo* to host proteins, which are well-described components of the 30 S hnRNP core particles (Beyer *et al.*, 1977) and which cannot be removed from the matrix by digestion with ribonuclease, in contrast to other core proteins (van Eekelen & van Venrooij, 1981). This suggests that they are implicated in the binding of viral and host RNA to the nuclear matrix.

4. Discussion

The nucleus of eukaryotic cells contains a structure, generally referred to as nuclear matrix, which can be isolated after removal of the DNA and DNA-associated proteins by treatment with DNAase I and various extraction procedures. It has been described by several workers (Herman *et al.*, 1978; Miller *et al.*, 1978*a*; Long *et al.*, 1979; van Eekelen & van Venrooij, 1981) that heterogeneous nuclear RNA is tightly associated with this nuclear framework. However, no attempt has been made so far to analyse the RNA moiety that is associated with the matrix. Such an analysis might provide important information

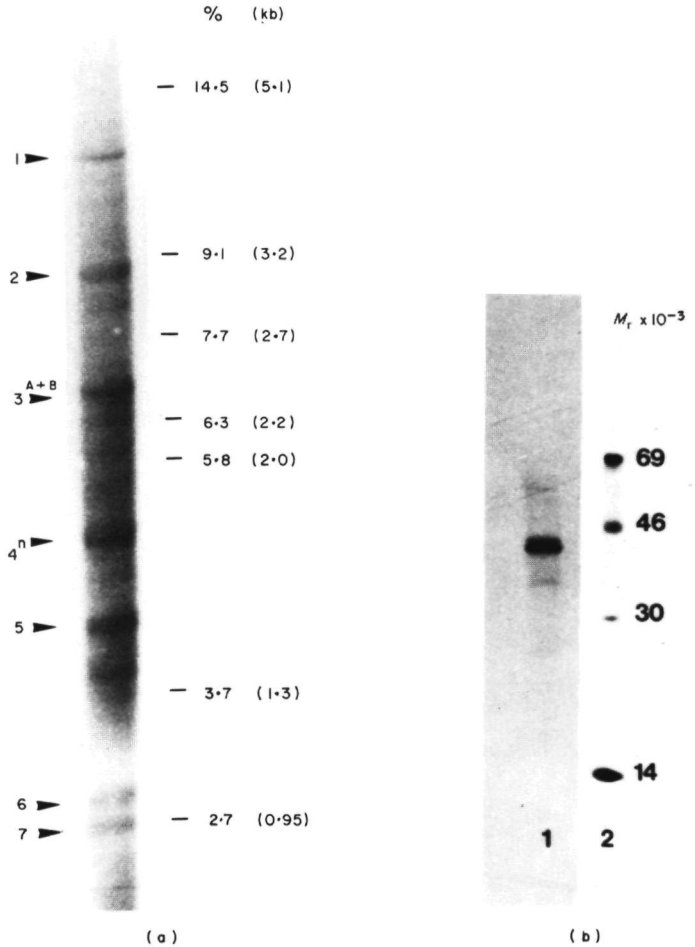


FIG. 7. Cross-linking of hnRNA to proteins *in vivo* by irradiation with ultraviolet light. Infected HeLa cells were labelled for 10 min with [^{3}H]uridine and [^{3}H]cytidine and irradiated with ultraviolet light in order to cross-link RNA to its associated proteins. Covalently linked RNA–protein complexes were prepared from the isolated nuclear matrices by extraction with phenol/chloroform (see Materials and Methods). (a) The RNA was isolated from the complexes by removal of the covalently linked proteins with proteinase K, hybridized to *Eco*RI B and analysed with the S_1 nuclease mapping procedure. The Figure shows the S_1-resistant hybrids. The positions of the *Hin*dIII molecular weight markers are as indicated. kb, 10^3 bases. (b) The proteins were freed from covalently linked RNA by extensive digestion with nucleases and analysed by sodium dodecyl sulphate/polyacrylamide gel electrophoresis as indicated in Materials and Methods.

on the topology of RNA processing in the nucleus. Using adenovirus-infected cells, we have tried to answer some of these questions and our main conclusions can be summarized as follows.

(1) Viral-specific hnRNA is associated quantitatively with the nuclear matrix. In the viral-specific nuclear matrix RNA, precursors, processing intermediates and processed mRNAs can be distinguished.

(2) Precursors, intermediates and products of RNA processing can be cross-linked *in vivo*, by means of irradiation with ultraviolet light, to host proteins of M_r 41,500 and 43,000.

The results underlying the first conclusion depend on the integrity of the nuclear matrix. The structural integrity seems to protect the associated RNA against degradation. Using an S_1 nuclease mapping procedure, we found that viral-specific nuclear matrix RNA contains precursors and products of RNA splicing. This result strongly suggests that the splicing process occurs while the RNA is associated with the nuclear matrix. It could be envisaged that all components necessary for the splicing process are assembled in a splicing complex localized on the nuclear matrix structure. Since it has been proposed that certain snRNAs may function as "splicer RNA"; i.e. to hold neighbouring exons together during splicing (Lerner *et al.*, 1980), it is interesting to note that some snRNA species have also been found in association with the matrix (Miller *et al.*, 1978*b*).

The fate of the spliced-out intervening sequences is unclear. Our S_1 nuclease analyses never showed differences between hybrid patterns of poly(A)-containing RNA and total RNA, not selected for poly(A), from nuclei or matrices, indicating that hybrids that could be ascribed to spliced-out intervening sequences were not present in detectable quantities. It is possible that the intervening sequences, after being spliced out, are degraded rapidly, so that appreciable accumulation does not occur. Alternatively, the intervening sequences could be excised in a stepwise manner by removal of oligonucleotides too small to be detected. In accordance with the latter model, it has been reported that splicing of the globin pre-mRNA (Kinniburgh & Ross, 1979; Reymond *et al.*, 1980) seems to take place in a stepwise mode.

S_1 nuclease mapping further shows the presence of RNA species that seem to be intermediates of RNA processing. Probably, the viral-specific hnRNA becomes attached to the nuclear matrix shortly after the start of transcription and remains associated with the matrix during all steps of processing. It seems reasonable to expect that this is a general phenomenon of hnRNA of eukaryotic cells, and that the adenovirus makes use of the host processing machinery.

The photo-induced cross-linking method, mentioned in conclusion (2), has a very high degree of specificity, because a firm association between RNA and protein must be present *in vivo* at the time of irradiation of the cells in order to achieve covalent cross-linkage (Wagenmakers *et al.*, 1980). In previous studies, it was shown that only two proteins can be efficiently cross-linked to host nuclear matrix RNA (van Eekelen & van Venrooij, 1981). It was proposed that these proteins could be involved in the binding of hnRNA to the matrix, since they were, in contrast to other hnRNA-associated core proteins, not released from the matrix by

extensive treatment with RNAase After infection, viral-specific nuclear matrix RNA, including precursors, intermediates and products of RNA processing, can be efficiently cross linked to the same host proteins, suggesting that the association of these RNA species with these proteins is not significantly influenced by the processing of pre-mRNA to mRNA Although our method of analysis does not allow a conclusion to be reached as to whether both proteins are randomly associated with all identified RNA species, or if there is specificity between some RNA species and one of the proteins, we believe that our data support the hypothesis that, during its processing, viral hnRNA is associated with nuclear matrix host proteins

We thank Dr L Philipson of the Biomedical Center, Uppsala, for critically commenting on the S_1 nuclease mapping data We also thank Dr M Matthews from the Cold Spring Harbor Laboratory for generously providing us with cloned *Hin*dIII A fragment, and A Groeneveld for culturing the cells

This work was supported by the Netherlands Foundation for Chemical Research (S O N) and the Netherlands Organization for the Advancement of Pure Research (Z W O)

REFERENCES

Akusjarvi, G , Zabielski, J , Perricaudet, M & Petterson, U (1981) *Nucl Acids Res* **9**, 1–17

Aviv H & Leder, P (1972) *Proc Nat Acad Sci U S A* **69** 1408–1412

Babich, A , Nevins, J R & Darnell, J E (1980) *Nature (London)*, **287**, 246–248

Beltz, G A & Flint, S J (1979) *J Mol Biol* **131**, 353–373

Berget, S M & Sharp, P A (1979) *J Mol Biol* **129**, 547–565

Berget, S M , Moore, C & Sharp, P A (1977) *Proc Nat Acad Sci , U S A* **74**, 3171–3175

Beyer, A L , Christensen, M E , Walker, B W & LeStourgeon, W M (1977) *Cell*, **11**, 127–138

Blanchard, J M , Weber, J , Jelinek, W & Darnell J E (1978) *Proc Nat Acad Sci , U S A* **75**, 5344–5348

Chow, L T & Broker, T R (1978) *Cell*, **15**, 497–510

Chow, L T , Roberts, J M , Lewis, J B & Broker, T R (1977) *Cell*, **11**, 819–836

Chow, L T , Broker, T R & Lewis, J B (1979) *J Mol Biol* **134** 265–303

Darnell, J E (1979) *Progr Nucl Acid Res Mol Biol* **22**, 327–353

Gillespie, S & Gillespie, D (1971) *Biochem J* **125**, 481–487

Goldenberg, C J & Raskas, H J (1979) *Cell*, **16**, 131–138

Herman, R , Weymouth, L & Penman, S (1978) *J Cell Biol* **78**, 663–674

Kaufmann, S H , Coffey, D S & Shaper, J H (1981) *Exp Cell Res* **132**, 105–123

Kinniburgh A J & Ross, J (1979) *Cell*, **17** 915–922

Klessig, D F & Chow, L T (1980) *J Mol Biol* **139**, 221–242

Lerner, M R , Boyle, J A , Mount, S M , Wolin, S L & Steitz, J A (1980) *Nature (London)*, **283**, 220–224

Long, B H , Huang, C Y & Pogo, A O (1979) *Cell*, **18**, 1079–1090

Manley, J L , Sharp, P A & Gefter, M L (1979) *J Mol Biol* **135**, 171–197

Miller, T E , Huang, C Y & Pogo, A O (1978*a*) *J Cell Biol* **76**, 675–691

Miller, T E , Huang C Y & Pogo, A O (1978*b*) *J Cell Biol* **76**, 692–704

Nevins, J R , Blanchard, J M & Darnell, J E (1980) *J Mol Biol* **144**, 377–386

Palmiter, R D (1973) *J Biol Chem* **248**, 2095–2106

Petterson, L & Sambrook, J (1973) *J Mol Biol* **73** 125–130

Reymond, C , Anghiano, A & Spohr, G (1980) *Proc Nat Acad Sci , U S A* **77**, 4683–4687

van Eekelen, C A G & van Venrooij, W J (1981) *J Cell Biol* **88**, 554–563
van Eekelen, C A G , Mariman, E C M , Reinders, R J & van Venrooij, W J (1981) *Eur J Biochem* **119**, 461–467
Wagenmakers, A J M , Reinders, R J & van Venrooij, W J (1980) *Eur J Biochem* **112**, 323–330
Ziff, E B (1980) *Nature (London)*, **287**, 491–499

Edited by P Chambon

CHAPTER VI

ON THE LOCALIZATION AND TRANSPORT OF SPECIFIC ADENOVIRAL mRNA-SEQUENCES IN THE LATE INFECTED HELA CELL

On the localization and transport of specific adenoviral mRNA-sequences in the late infected Hela cell
E. Mariman, A-M. Hagebols and W. van Venrooij (1982)
Nucleic Acids Res. 10, 6131-6145.

On the localization and transport of specific adenoviral mRNA-sequences in the late infected HeLa cell

Edwin Mariman, Anne-Marie Hagebols and Walther van Venrooij

Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

Received 23 July 1982; Revised and Accepted 14 September 1982

ABSTRACT

In the nucleus of HeLa cells late after infection with adenovirus type 2 mRNA-sequences which are processed via RNA splicing are attached to the nuclear matrix (Mariman et al., 1982). Although the mRNA, which codes for polypeptide IX, is not formed via splicing, about 70% of the non-polyadenylated pre-mRNA and the polyadenylated pIX mRNA are bound to the matrix structure, indicating that polyadenylation is performed while the RNA is associated with the matrix. Binding to the nuclear matrix seems to be a common property of all mRNA-sequences in the nucleus. At the late stage of infection most of the newly synthesized mRNAs which appear in the cytoplasm are viral specific (Beltz & Flint, 1979). Kinetic analysis of the newly synthesized poly(A)-containing mRNA on sucrose gradients reveals that 7-12 S messengers appear more rapidly in the cytoplasm than messengers larger than 13 S. More specifically, the nuclear exit time of the pIX-mRNA, which is the major 9 S adenoviral messenger late after infection, was determined to be about 4 min, while messengers transcribed from the late region 3 need more than 16 min to arrive in the cytoplasm. In the cytoplasm about 70% of the mRNA is bound to the cytoskeletal framework, while 30% remains as free mRNP. Analysis of the mRNA in both fractions reveals that L3-, E1B- and pIX-specific polyadenylated mRNA preferably exist as cytoskeleton-bound mRNA. However, significant differences occur in the partition of specific messengers over free and cytoskeletal RNA fractions.

INTRODUCTION

It has been reported that hnRNA is quantitatively associated with a higher ordered proteinaceous structure, generally referred to as the nuclear matrix (1-4). At the late stage of infection of HeLa cells with adenovirus type 2 the viral specific hnRNA is also quantitatively attached to the nuclear matrix (4). It has been suggested that the nuclear matrix may function in localizing and structurally organizing hnRNA during its processing, since both mRNA-precursors and spliced products can be detected in the matrix associated RNA (5). After completion of its processing the matured messenger RNA is transported to the cytoplasm where it is translated.

mRNA-sequences in the nucleus as well as in the cytoplasm are complexed with several proteins. However, a different set of proteins seems to be associated

0305-1048/82/1019-6131$ 2.00/0

with mRNA-sequences in the nucleus as compared to cytoplasmic mRNA, indicating that during nucleo-cytoplasmic transport one set of proteins is exchanged for another (6). Finally, in the cytoplasm mRNA can exist either as free mRNP or be bound to the cytoskeleton, a cytoplasmic framework composed of structural proteins, which can be isolated by treatment of cells with a non-ionic detergent (7, 8).

In this study we pursued specific adenoviral mRNA-sequences from transcription to their destination in the cytoplasm. Although all adenoviral mRNA-sequences roughly seem to follow the same pathway through the infected HeLa cell, significant differences between specific mRNAs occur in the rate of their nucleo-cytoplasmic transport and their localization in the cytoplasm.

MATERIALS AND METHODS

Infection and labeling of cells

HeLa S3 cells were infected with adenovirus type 2 (2000 particles per cell) at a density of $5x10^6$ cells/ml in Eagles minimal essential medium containing 1 mM arginine (MEM). After adsorption of the virus to the cells during 1 h at 37 ^oC, the cells were diluted to $0.3x10^6$ cells/ml with MEM containing 5% newborn calf serum. Cells were harvested at 18 h after infection. To obtain a time course of labeled mRNA 18 h after infection, cells were concentrated to $20x10^6$ cells/ml and incubated for the indicated periods of time with 50 µCi/ml (5,6-^{3}H) uridine (45 Ci/mmol, Radiochemical Centre Amersham, England), 25 µCi/ml (5'-^{3}H) guanosine (20 Ci/mmol, Amersham) and 25 µCi/ml (5-^{3}H) cytidine (31 Ci/mmol, Amersham). To obtain steady state labeled mRNA the cells were labeled from 16 to 18 h after infection with 5 µCi/ ml (5,6-^{3}H) uridine at a density of $4x10^6$ cells/ml.

Purification and oligo(dT)-cellulose chromatography of cytoplasmic and nuclear RNA

The incorporation of labeled precursors in the RNA was terminated by rapidly mixing the cells with an equal volume of crushed, frozen NKM solution (0.13 M NaCl, 0.05 M KCl, 1.5 mM $MgCl_2$). After centrifugation for 5 min at 800g the cells were washed once with NKM, pelleted and resuspended in RSB (10 mM NaCl, 10 mM Tris, pH 7.4, 1.5 mM $MgCl_2$). After the addition of 0.1 vol of a sodium deoxycholate-Nonidet P40 mixture (5% each) the cells were disrupted by intensive vortexing. The nuclei were pelleted and from the supernatant cytoplasmic RNA was isolated by phenol/chloroform extraction (9). The nuclei were washed once in RSB and resuspended in the same buffer. Nuclear RNA was isolated by phenol extraction at 55^oC as described by Long et al.

(10). To select the poly(A)-containing RNA oligo(dT)-cellulose chromatography was performed as described by Aviv and Leder (11) using the T2 grade of Collaborative Research Inc.

Fractionation of cytoplasmic RNA into free and cytoskeleton-bound mRNA

Cytoskeletons from HeLa cells were prepared essentially as described by van Venrooij et al. (12). Cells were harvested by centrifugation and washed once with isotonic NKM solution and once with hypertonic buffer (0.3 M sucrose 10 mM KCl, 1.5 mM Mg-acetate, 10 mM Tris-acetate, pH 7.4). The cells were resuspended in hypertonic buffer ($40x10^6$ cells/ml) and 1 vol. of 1% Triton X-100 (BDH Chem. Ltd., Poole, Dorset) in hypertonic buffer was added. The suspension was gently swirled in ice for 30 sec and centrifuged (5 min, 800g) to sediment the cytoskeletons. The supernatant containing the free mRNA was pipetted off and the pellet was washed with hypertonic buffer containing 0.5% Triton X-100. After centrifugation the pellet was resuspended in RSB. To solubilize the cytoskeletal fraction 0.1 vol. of a sodium deoxycholate-Nonidet P40 mixture (5% each) was added to the suspended cytoskeletons and the suspension was intensively vortexed. The nuclei were removed from the cytoskeletal fraction by centrifugation at 800g for 5 min. Immediately after cell fractionation the various fractions were made 1% in SDS and deproteinized by phenol/chloroform extraction (9). The RNA was then passed over oligo(dT)-cellulose (T2-grade, Collaborative Research Inc.) to select the poly(A)-containing mRNA (11).

Preparation of nuclear matrix RNA

Nuclear matrices of HeLa cells 18 h after infection were isolated as previously described (4). Nuclei were isolated from infected HeLa cells as described above. The nuclei were resuspended in RSB containing 0.1 M NaCl and treated with DNAse I for 30 min at 10^oC (500 µg/ml). After DNAse I digestion the nuclei were pelleted through a 30% sucrose layer by centrifugation at low speed. The pelleted material was extracted once with 0.4 M $(NH_4)_2SO_4$. After repelleting the nuclear matrices were washed with RSB, suspended in RSB buffer containing 1% SDS and incubated at 100^oC for 3 min. The boiled mixture was adjusted to 0.2 M NaCl and extracted with phenol/chloroform (1:1). After two additional extractions with chloroform the RNA was precipitated with two volumes of ethanol.

Adenoviral DNA isolation and restriction endonuclease digestions

Viral DNA was isolated from the purified virions as described by Petterson and Sambrook (13). EcoRI and HindIII (Boehringer, Mannheim) restriction digestions were performed at 37^oC for 4 h using 1 unit of enzyme per µg DNA.

Standard conditions were 50 mM Tris, pH 7.4, 7 mM $MgCl_2$, 50 mM NaCl, 3 mM DTT. SstI (Bethesda Research Laboratories Inc.) digestions were performed at 37°C for 12 h (1 unit/µg DNA) in Tris buffer containing 10 mM Tris, pH 7.4, 7 mM $MgCl_2$, 10 mM KCl, 20 mM NaCl, 3 mM DTT. In case of SmaI (Boehringer, Mannheim) the DNA was digested at 25°C for 12 h (1 unit/µg DNA) in a buffer containing 15 mM Tris, pH 8.6, 6 mM $MgCl_2$, 15 mM KCl, 3 mM DTT. Restriction fragments were separated by agarose gel electrophoresis; 0.8% agarose gels were run for 20 h at 80 V in buffer TNE (40 mM Tris, pH 8.0, 8 mM NaAc, 2 mM EDTA). The DNA was recovered from agarose gel slices by the method of electrophoretic elution.

RNA-DNA hybridization

The DNA restriction fragments were immobilized to nitrocellulose filters as described by Gillespie and Gillespie (14). RNA was dissolved in 200 µl 3x SSC (SSC, 0.15 M sodium chloride-0.015 M sodium citrate), 1 mM EDTA, 0.1% SDS, 50% formamide. To this solution a filter containing 2 µg equivalent of DNA of a particular restriction fragment was added. The RNA was denatured by incubating the mixture at 65°C for 5 min. Hybridization was then carried out at 38°C for an appropriate period of time. After hybridization the filters were washed several times in 2x SSC and subsequently with 0.1 N NaOH at 45°C for 10 min. The solution containing the eluted RNA was then neutralized with HCl and diluted with one volume of water. The radioactivity in the hybridized RNA was determined by means of liquid scintillation counting using pico-fluor™15 (Packard Inst.) as a scintillant.

RESULTS

Association of mRNA-sequences with the nuclear matrix

It has been shown that the nucleus of HeLa cells, as well as of other cell types, contains a higher ordered proteinaceous structure, usually referred to as the nuclear matrix, to which the hnRNA is quantitatively attached (1-4). Recently it was shown that the process of RNA splicing is performed while the RNA is attached to the matrix structure (5). With this finding the possibility arose that only mRNA-sequences which have to be spliced are bound, for example via intron sequences, to the matrix structure. Release of processed mRNA would then occur simultaneous with the final splicing step. Unspliced mRNAs, in that case, would not be found associated with the matrix. The messenger which codes for polypeptide IX (15), a structural component of the virion, is transcribed from the viral r-strand between coordinates 9.8 and 11.2 (16; the viral genome is divided into 100 map units). Recently it was shown that the nucleotide sequences of this mRNA is colinear with the DNA (17) and until

now this messenger is the only known adenoviral mRNA which is unspliced.

To test our hypothesis HeLa cells 18 h after infection with Ad2 were labeled with ^{3}H-nucleosides for 30 min. After labeling the nuclei were isolated as indicated in Materials and Methods and divided in two. From one part nuclear RNA was isolated by means of the hot phenol extraction procedure (see Materials and Methods). From the other part of the nuclei nuclear matrices were prepared. The nuclear matrix RNA was then isolated as described in the Materials and Methods section. Both nuclear and nuclear matrix RNA were fractionated in polyadenylated (6% of the labeled RNA) and non-polyadenylated RNA (94% of the labeled RNA) via oligo(dT)-cellulose chromatography. This demonstrates that not only polyA(+) but also polyA(-) hnRNA is bound to the nuclear matrix. The amount of RNA containing nucleotide sequences specific for L3, E1B or pIX in each fraction was then quantitated by the method of filter hybridization as described in the Materials and Methods section. The nucleotide sequences of the pIX-specific RNA reside in the region E1B (18; 4.6-11.2 m.u.), which is active in transcription at the early stage of infection. However, transcription of this region seems to continue in the late phase of infection. Two early messengers are transcribed from the region E1B which both contain the nucleotide sequences that make up the mRNA for pIX (18; Fig. 1A). Nuclear and nuclear matrix RNA were hybridized to a HindIII/SmaI restriction fragment of the Ad2 DNA (7.9-10.8 m.u.; Fig. 1A), which binds the E1B-specific RNA as well as the pIX-specific RNA-sequences. The RNA was also hybridized to a SstI/HindIII fragment (5.0-7.9 m.u.; Fig. 1A), which only detects the E1B-specific RNA. The percentage of hybridization due to the pIX RNA was calculated by substracting the percentage found with the Sst/Hind fragment from the percentage that was found when the Hind/Sma fragment was used. To detect the L3-specific RNA a Hind/EcoRI restriction fragment (19; 50.1-58.5 m.u.; Fig. 1B) was used. The results, listed in table I, show that 70% or more of the nuclear RNA, which contains L3-, E1B- or pIX-specific sequences is attached to the nuclear matrix. There is no significant difference in the binding to the nuclear matrix between RNA which contains sequences for the unspliced pIX mRNA and the RNA sequences which form the L3 or E1B messengers via splicing. Therefore, binding to the nuclear matrix seems to be a common property of mRNA sequences, that have to be transported to the cytoplasm and is not limited to the messengers which are generated via splicing.

Sucrose gradient analysis of newly accumulated mRNA in the cytoplasm

At the late stage of infection of HeLa cells with Ad2 most of the newly synthesized mRNAs, which appear in the cytoplasm are viral specific (20).

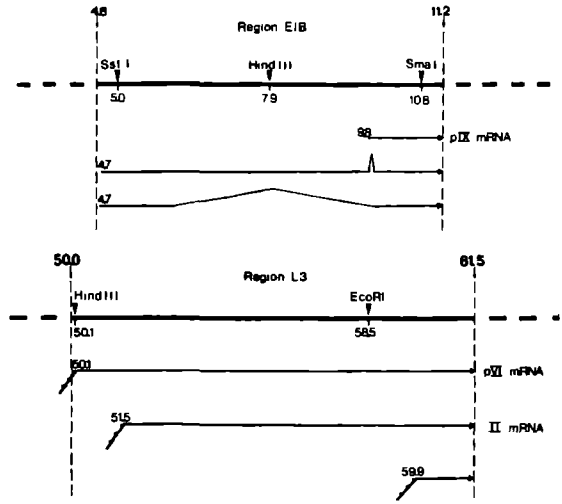


Fig. 1. A. Map coordinates of the early region 1B. The mRNAs which are transcribed from the region E1B are represented by arrows, arrowheads point at the 3' end of the mRNAs. Caret symbols indicate the intervening sequences which are absent from the spliced messengers. The cleavage sites of the endonuclease restriction enzymes which have been used to obtain Ad2 DNA fragments, are as indicated. B. Map coordinates of the late region 3. The mRNAs which are transcribed from the region L3 are represented by arrows, arrowheads point at the 3' end of the mRNAs. Jointed stalks represent the spliced tripartite leader sequences. The cleavage sites of the endonuclease restriction enzymes which have been used to obtain Ad2 DNA fragments, are as indicated.

Figure 2 shows the sedimentation profile of steady state mRNA late after infection as determined by sucrose gradient centrifugation. Three size classes can be distinguished: 18-24 S (class I), 14-17 S (class II) and 7-12 S (class III). The first type of experiment we undertook was to see if there were appreciable differences in the time of appearance in the cytoplasm between mRNAs of different size classes. Therefore, HeLa cells 18 h after infection were labeled with ^{3}H-nucleosides (see Materials and Methods) for 5, 15, 30 or 45 min, respectively. The poly(A)-containing cytoplasmic RNA was isolated, denatured and layered on top of a 16-43.5% (w/v) isokinetic sucrose gradient. After centrifugation the gradients were fractionated and the radioactivity in each fraction was determined. As can be seen from the gradient profiles in Figure 3 class III mRNA appears quite rapidly in the cytoplasm (Fig. 3A). After labeling for 30 min or longer the gradient profile (Fig. 3C and 3D) already closely resembles that for steady state mRNA (Fig. 2)

Table I. Determination of the percentages of L3-, E1B- and pIX-specific nuclear RNA sequences which are bound to the nuclear matrix.

		Hybridization of nuclear RNA (%)	Hybridization of matrix RNA (%)	Nuclear RNA bound to the matrix (%)	Poly A(+)/ poly A(-)
L3-specific	pA^+	7.1	6.1	86	0.8
	pA^-	9.8	7.6	78	
E1B-specific	pA^+	7.4	5.3	72	0.9
	pA^-	6.7	5.7	85	
pIX-specific	pA^+	1.0	0.7	70	0.2
	pA^-	3.8	3.2	84	

Ad2 infected HeLa cells 18 h after infection were labeled with ^{3}H-nucleosides for 30 min. After labeling nuclear and nuclear matrix RNA were prepared and fractionated in polyA(+) and polyA(-) RNA by oligo(dT)-cellulose chromatography (Materials and Methods). In each fraction of RNA the amounts of L3-, E1B- and pIX specific sequences were determined by filter hybridization as described in the Materials and Methods section. The restriction fragments of Ad2 DNA which were used, are as indicated in the text. The percentages of hybridization were corrected for non-specific adsorption by incubation of RNA with a blank filter under the same conditions. The third column shows the percentages of specific RNA sequences which are bound to the nuclear matrix. These data were calculated from the amounts of certain specific RNA sequences in nuclear and nuclear matrix RNA as shown in the first and second column. The fourth column shows the ratio between the amounts of certain specific RNA sequences in polyA(+) and polyA(-) nuclear matrix RNA. Each percentage is the mean value of the results obtained in 5 separate hybridizations. The deviation of the values in the first and second column is within 1/10 of the indicated values.

while at the intermediate labeling time of 15 min the size distribution also was found to be intermediate (Fig. 3B). The ratios between the amounts of newly synthesized class I and class III mRNAs and between class II and class III mRNAs rise between 5 and 45 min labeling illustrating the fact that there is a significant difference between the time of appearance in the cytoplasm of 7-12 S poly(A)-containing messengers on one hand and messengers larger than 13 S on the other.

Nuclear exit times of different adenoviral messengers

The gradient profiles of Figure 3 showed that 7-12 S messengers appear more rapidly in the cytoplasm than larger mRNAs. We were interested to see if this result could be verified for specific mRNAs by determining the nuclear exit

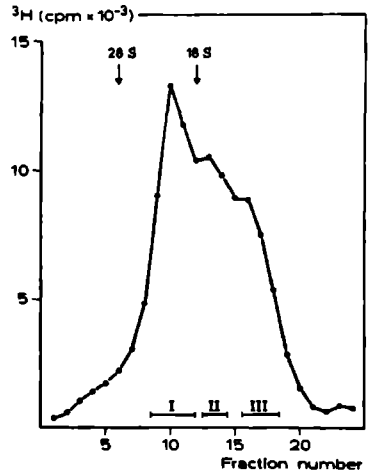


Fig. 2. Sucrose gradient profile of steady state labeled poly(A)-containing cytoplasmic RNA. HeLa cells were infected with Ad2 and from 16 to 18 h after infection the cells were labeled with 5 µCi/ml (5,6-^{3}H) uridine. The cytoplasmic RNA was isolated as described in the Materials and Methods section. To isolate the poly(A)-containing fraction of the cytoplasmic RNA, oligo(dT)-cellulose chromatography was performed as described by Aviv and Leder (11). The RNA was dissolved in a buffer of low ionic strength containing 10 mM Tris, pH 7.4, 2 mM EDTA, 0.2% SDS, heated to 100°C for a few minutes, and after rapid cooling the solution was layered on top of a 16-43.5% (w/v) isokinetic sucrose gradient containing the same buffer. Centrifugation was for about 16 h at 20°C in a SW41 rotor at 150,000g. The gradient was fractionated and the radioactivity counted with pico-fluor™15 (Packard Inst.). The horizontal bars indicate the different size classes of RNA which can be distinguished in the gradient profile.

time of the messenger for pIX, which is the major 9 S adenoviral messenger at the late stage of infection (15), and that of the messengers which are transcribed from the late region 3 (Fig. 1B) and which sediment at a higher density than 18 S (15). HeLa cells 18 h after infection were labeled for 4, 8, 12, 16 min and 5, 15, 30, 45 min, respectively. Poly(A)-containing cytoplasmic RNA was isolated and hybridized to appropriate restriction fragments of the Ad2 DNA, immobilized to nitrocellulose filters, as described in the first section of this paragraph (Fig. 1; Materials and Methods). The same quantities of RNA were, under the same conditions, also hybridized to filters containing 2 µg of total Ad2 DNA. The percentages of RNA hybridized to restriction fragments were then divided by the percentages of RNA hybridized to total Ad2 DNA.

The results are depicted in Figure 4 and in agreement with our expectation they show that the first pIX-specific sequences appear in the cytoplasm already after 4 min labeling, while labeled L3 messengers begin to accumulate after 16 min (Fig. 4).

Partition of messengers over free and cytoskeleton-bound mRNA fractions

Morphological studies including immunofluorescence and electron microscopy have revealed the existence of a cytoskeletal framework composed

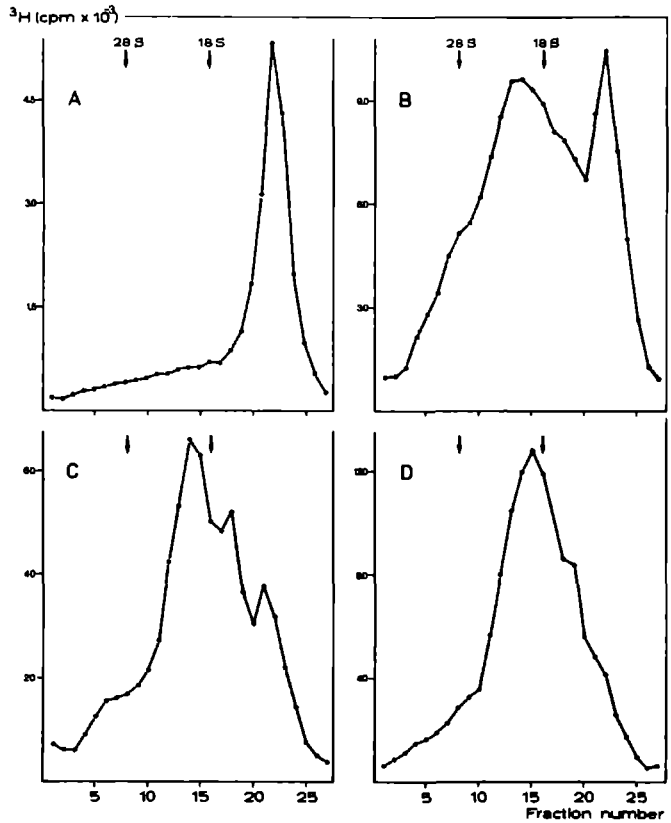


Fig. 3. Sucrose gradient analysis of newly accumulated poly(A)-containing mRNA in the cytoplasm. HeLa cells were infected with Ad2. After labeling the cells with ^{3}H-nucleosides (see Materials and Methods) for a period of time as indicated in the text, 18 h after infection, the poly(A)-containing cytoplasmic RNA was isolated and analysed on an isokinetic sucrose gradient as described in the legend of Figure 2. Labeling time: 5 min (3A); 15 min (3B); 30 min (3C); 45 min (3D).

of structural proteins in the cytoplasm of eukaryotic cells, which can be isolated by gently extracting the cells with a non-ionic detergent such as TX-100 (7, 8). Some workers have found that cytoplasmic mRNA is in part bound to the cytoskeleton but that it can exist also as free mRNP. Further results have indicated some specificity in the distribution of mRNA-sequences over the free and cytoskeleton-bound mRNA fractions (12, 21, 22).

To determine if adenoviral mRNA preferably exists as free mRNP or as

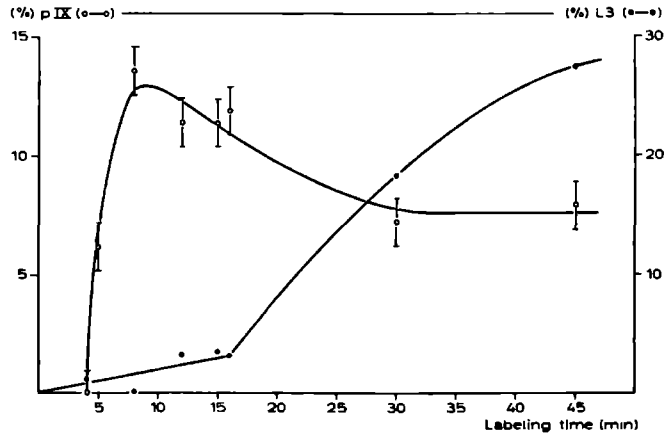


Fig. 4. Determination of nuclear exit times by filter hybridization. Ad2 infected HeLa cells 18 h after infection were labeled with ^{3}H-nucleosides for 4, 8, 12, 16 min and 5, 15, 30, 45 min respectively (Materials and Methods). Poly(A)-containing cytoplasmic RNA was prepared, dissolved in 200 µl 3x SSC, 1 mM EDTA, 0.1% SDS, 50% formamide and hybridized to DNA fragments immobilized to nitrocellulose filters. After hybridization the percentage of RNA which was hybridized to the DNA was determined as described in the Materials and Methods section. To detect L3-specific sequences the RNA was hybridized for 70 h to filters containing 2 µg equivalent of the HindIII/EcoRI fragment indicated in Figure 1B. The percentages were corrected for non-specific adsorption by incubation of a blank filter with RNA under the same conditions. To detect pIX-specific sequences the RNA was hybridized for 150 h to 2 µg equivalent of an SstI/HindIII as well as a HindIII/SmaI fragment (Fig. 1A). The percentages of hybridization due to pIX-specific sequences were calculated by substracting the percentages found with the first fragment from those which were found with the second fragment. Using identical conditions the same quantity of RNA was hybridized to filters containing 2 µg Ad2 DNA. The percentage of the RNA hybridized to a DNA frgment was divided by the percentage of the RNA which hybridized to total Ad2 DNA under these conditions. Each point in the curves is the mean value of 5 separate hybridizations.

cytoskeleton-bound mRNA, HeLa cells were labeled from 16 to 18 h after infection to obtain steady state labeled mRNA (Materials and Methods). After labeling poly(A)-containing mRNA was prepared from both the free and cytoskeletal RNA fraction as described in the Materials and Methods section. In accordance with earlier results (12) it was found that only 30% of the steady state labeled mRNA exists as free mRNP while as much as 70% is bound to the cytoskeleton. The sucrose gradient profiles as depicted in Figure 5 show that there is only a slight difference in the size distribution and composition between free and cytoskeletal mRNA. By the method of filter hybridization as described in the first section of this paragraph we

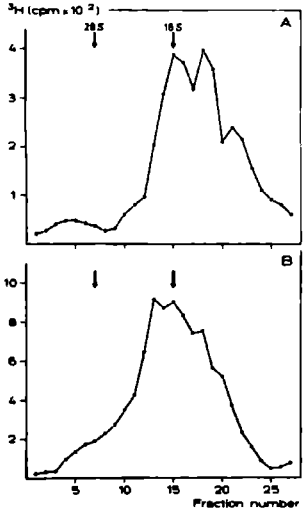


Fig. 5. Sucrose gradient analysis of steady state labeled polyA(+) mRNA isolated from the free and cytoskeletal RNA fractions. HeLa cells were infected with Ad2. After labeling the cells with (5,6-^{3}H) uridine (5 µCi/ml) from 16 to 18 h after infection, the free and cytoskeleton RNA fractions were isolated as described in the Materials and Methods section. Oligo(dT)-cellulose chromatography was performed (11) to isolate the poly(A)-containing RNA. The RNA was then dissolved in a buffer of low ionic strength containing 10 mM Tris, pH 7.4, 2 mM EDTA, 0.2% SDS, heated to 100°C for a few minutes, and after rapid cooling the solution was layered on top of a 16-43.5% (w/v) isokinetic sucrose gradient containing the same buffer. Centrifugation was for about 16 h at 20°C in a SW41 rotor at 150,000g. A. Free polyA(+) mRNA. B. Cytoskeleton-bound polyA(+) mRNA.

quantitated the amounts of L3-, E1B- or pIX-specific mRNA-sequences in both the free and the cytoskeletal mRNA fraction. The results (Table II) demonstrate that each of the individual mRNAs preferably is bound to the cytoskeletal framework. However, significant differences can be detected in the partition of specific messengers over both cytoplasmic fractions. This possibly reflects the affinity of individual messengers for the cytoskeletal framework but, as was suggested earlier (12), can also point to a regulation mechanism of the translation of specific (adenoviral) mRNAs.

DISCUSSION

The mRNAs of eukaryotic cells are usually derived from larger precursor molecules. These primary products of transcription are modified by a number of processing steps, i.e. capping, polyadenylation, methylation and splicing. After processing, which is performed in the cell nucleus, the newly formed mRNA is transported out to the cytoplasm. At some time during mRNA transport, the proteins which are associated with the mRNA in the nucleus are replaced by a cytoplasm-specific set of mRNA-associated proteins (6). In the cytoplasm messengers can occur either as free mRNP or be bound to the cytoskeleton, a cytoplasmic framework (12, 21, 22). In the present study we pursued some adenoviral mRNA-sequences from transcription to their destination in the cytoplasm. Our main results can be summarized as follows:

Table II. Partition of L3-, E1B- and pIX-specific poly(A)-containing mRNA over free and cytoskeletal RNA fractions.

	free RNA fraction	cytoskeletal RNA fraction
L3-specific mRNA	11.5%	88.5%
E1B-specific mRNA	2.2%	97.8%
pIX-specific mRNA	28.5%	71.5%

Ad2 infected HeLa cells were labeled from 16 to 18 h after infection with ^{3}H-nucleosides. After labeling the polyA(+) mRNA was isolated from both the free and cytoskeleton-bound RNA fractions (Materials and Methods). In each fraction of cytoplasmic mRNA the amounts of L3-, E1B- and pIX-specific mRNA were determined by filter hybridization as described in the Materials and Methods section. The restriction fragments of the Ad2 DNA, which were used, are indicated in the first section of this paragraph. The percentages of hybridization were corrected for non-specific adsorption by incubation of RNA with a blank filter under the same conditions. From these data the partition of specific mRNA sequences over the free and cytoskeletal RNA fractions were calculated. Each percentage is the mean value of the results of 5 separate hybridizations. The absolute deviation in the indicated percentages is less than 2.5%. Recently we have found the same results using 1 h labeled RNA and cloned DNA fragments.

1. In the nucleus pre-mRNA sequences, polyadenylated or non-polyadenylated, are associated with the nuclear matrix regardless of the fact that their processing pathway includes the process of RNA splicing or not.
2. Generally, small-sized messengers ($\pm$ 9 S) appear more rapidly in the cytoplasm than larger mRNAs. When specific mRNAs are compared, the nuclear exit time differs significantly.
3. In the cytoplasm messengers are preferably bound to the cytoskeletal framework although significant differences occur in the steady state distribution of specific mRNAs over free and cytoskeleton-bound RNA fractions.

HnRNA is quantitatively attached to a nuclear matrix (1-4). Adenovirus region L4 RNAs are attached to the nuclear matrix both as precursors and as products of splicing, indicating that splicing is performed while the RNA is associated with the matrix structure (5). Accordingly, using the method of filter hybridization for a more quantitative approach, we have found that at least 70% of the L3- and E1B-specific nuclear RNAs, which both are processed via RNA splicing, are attached to the nuclear matrix. Similarly it was shown that the polyA(-) nuclear matrix RNA contains at least 70% of the non-

polyadenylated pIX-specific sequences present in the nucleus. From this we concluded that pre-mRNA for the pIX messenger is attached to the nuclear matrix, although the mRNA is not formed via splicing. At least 70% of the nuclear poly(A)-containing pIX RNA is also attached to the nuclear matrix, indicating that not only splicing but also polyadenylation is performed while the RNA is bound to the matrix structure. From Table I it can be seen that polyA(+) and polyA(-) nuclear matrix RNA roughly contain the same amount of L3- or E1B-specific sequences. However, the polyA(+) fraction contains about 4 times less pIX-specific sequences than the polyA(-) fraction. This suggests that polyadenylation is the last step in the processing of pIX RNA and indicates that the mature mRNA, after polyadenylation, is immediately transported out to the cytoplasm. The measurement of the nuclear exit time of these specific sequences confirms this suggestion. As seems to be the case for most adenoviral mRNAs (23) processing and transport of mRNA-sequences of the late region 3 last more than 16 min. However, the processing and transport of pIX mRNA is performed in only about 4 min (Fig. 4). The relatively rapid appearance of pIX mRNA in the cytoplasm could be explained by the fact that the processing pathway does not include RNA splicing, which could be a time-consuming processing step. An alternative explanation would be that the rapid appearance of pIX mRNA could be the result of an enhanced rate of nucleo-cytoplasmic transportation. Our findings do not support the existence of different transport mechanisms for pIX and L3 messengers. However, rapid transport could simply be the result of the smaller size of the pIX messenger. The gradient profiles of Figure 3 show that small-sized messengers ($\pm$ 9 S) are the first to appear in the cytoplasm. This seems to be a more general feature since other workers have obtained similar results using uninfected HeLa cells and cultured L-cells (24, 25). They reported that 9 S mRNA accumulated within 10 min in the small polyribosomes of these cells. In addition it was reported that the β-globin mRNA, a 10 S mRNA which is formed via RNA splicing, reached the cytoplasm within 5 min after termination of transcription (26, 27). Studies on the accumulation of E1A and E1B mRNAs have been performed by Wilson and Darnell (28).

In the cytoplasm about 70% of the adenoviral poly(A)-containing mRNA is bound to the cytoskeletal framework. Binding to the cytoskeleton might be obligatory for the mRNA to be translated (12, 22). Sucrose gradient analysis of free and cytoskeletal polyA(+) mRNA late after infection reveals that both fractions of RNA sediment with roughly the same velocity (Fig. 5). However, translation experiments have indicated that among adenovirus-specific

messengers there may be species which exist preferably as free or as cytoskeletal mRNA (12). Using filter hybridization we found that L3-, E1B- and pIX-mRNA species are preferably bound to the cytoskeletal framework (Table II). However, while the E1B mRNA was almost quantitatively bound to the cytoskeletal framework, about 30% of the pIX mRNA was present as free mRNP. This demonstrates that among specific mRNA species there are differences in the binding to the cytoskeletal framework supporting the hypothesis that cytoskeleton-binding may be a regulatory step in the synthesis of (viral) proteins (12, 21, 22).

Summarizing, our results show that various adenoviral mRNA-sequences seem to follow a similar pathway through the infected HeLa cell, notwithstanding large differences in their nuclear processing, their exit time and their partition over free and cytoskeleton-bound mRNA fractions.

ACKNOWLEDGEMENTS

We thank A. Groeneveld for culturing the cells. This work was supported by the Netherlands Foundation for Chemical Research (S.O.N.) and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

1. Herman, K., Weymouth, L. and Penman, S. (1978) J. Cell Biol. 78, 663-674.
2. Miller, T.E., Huang, C.-Y. and Pogo, A.O. (1978) J. Cell Biol. 76, 675-691.
3. van Eekelen, C.A.G. and van Venrooij, W.J. (1981) J. Cell Biol. 88, 554-563.
4. van Eekelen, C.A.G., Mariman, E.C.M., Reinders, R.J. and van Venrooij, W.J. (1981) Eur. J. Biochem. 119, 461-467.
5. Mariman, E.C.M., van Eekelen, C.A.G., Reinders, R.J., Berns, A.J.M. and van Venrooij, W.J. (1982) J. Mol. Biol. 154, 103-119.
6. van Eekelen, C.A.G., Riemen, T. and van Venrooij, W.J. (1981) Febs Letters 130, 223-226.
7. Brown, S., Levinson, W. and Spudich, J.S. (1976) J. Supramol. Struc. 5, 119-130.
8. Lenk, R., Ransom, L., Kaufman, Y. and Penman, S. (1977) Cell 10, 67-78.
9. Palmiter, R.D. (1973) J. Biol. Chem. 248, 2095-2106.
10. Long, B.H., Huang, C.-Y. and Pogo, A.O. (1978) Cell 18, 1079-1090.
11. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci., USA 69, 1408-1412.
12. van Venrooij, W.J., Sillekens, P.T.G., van Eekelen, C.A.G. and Reinders, R.J. (1981) Exp. Cell Res. 135, 79-91.
13. Petterson, L. and Sambrook, J. (1973) J. Mol. Biol. 73, 125-130.
14. Gillespie, S. and Gillespie, D. (1971) Biochem. J. 125, 481-487.
15. Chow, L.T., Roberts, J.M., Lewis, J.B. and Broker, T.R. (1977) Cell 11, 819-836.
16. Ziff, E.B. (1980) Nature 287, 491-499.
17. Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D.F. and Petterson, U. (1980) Cell 19, 671-681.

18. Chow, L.T., Broker, T.R. and Lewis, J.B. (1979) J. Mol. Biol. 134, 265-303.
19. Broker, T.R. and Chow, L.T. (1980) Trends in Biol. Sci. 5, 174-178.
20. Beltz, G.A. and Flint, S.J. (1979) J. Mol. Biol. 131, 353-373.
21. Benze'ev, A., Horowitz, M., Solnik, H., Abulafia, R., Laub, O. and Aloni, Y. (1981) Virology 111, 475-487.
22. Cervera, M., Dreyfuss, G. and Penman, S. (1981) Cell 23, 113-120.
23. Nevins, J.R. and Darnell, J.E. (1978) Cell 15, 1477-1493.
24. Adesnik, M. and Darnell, J.E. (1972) J. Mol. Biol. 67, 397-406.
25. Schochetman, G. and Perry, R.P. (1972) J. Mol. Biol. 63, 591-596.
26. Bastos, R.N. and Aviv, H. (1977) Cell 11, 641-650.
27. Kinniburgh, A.J. and Ross, J. (1979) Cell 17, 915-922.
28. Wilson, M.C. and Darnell, J.E. (1981) J. Mol. Biol. 148, 231-251.

SUMMARY

In nuclei of eukaryotic cells transcripts are produced by RNA polymerase II which, in order to be able to function as messenger RNA in the cytoplasm, have to undergo modification processes such as capping, methylation, polyadenylation and splicing. We have focussed our attention on RNA splicing. In particular the excision of intron sequences from the adenoviral major late pre-mRNAs was studied. The major late mRNAs are composed of four exons: leader 1, which contains the cap, leader 2, leader 3 and a protein coding sequence, to which the poly(A)-tail is added. When the leaders have been spliced they are referred to as the tripartite leader sequence. Three introns, numbered 1, 2 and 3, separate the exons in the pre-mRNAs: intron 1 is located between leaders 1 and 2, intron 2 between leaders 2 and 3, and intron 3 between leader 3 and the protein coding sequence.

Some aspects of the splicing mechanism.

Under steady state conditions putative processing intermediates, missing part of an intron, can be found in the nucleus indicating that removal of introns 1 and 2 is a stepwise process and occurs in the 5' to 3' direction (Chapter III). Within an intron preferred and non-preferred cleavage sites can be distinguished. Rapid cleavage at preferred sites can occur already in the nascent transcript(Chapter II) showing that splicing can precede polyadenylation. Because an intron can have more than one preferred cleavage site, the first cleavage event may occur at different locations in the intron. Alternative splicing pathways seem to be operative in the removal of introns 1 and 2 (Chapter II). The last cleavage is performed at the intron-exon boundary (acceptor site) and is rapidly followed by the ligation of the exons (Chapters II and III). A model for splicing based on our findings is presented in Chapter III.

Splicing occurs in the 5' to 3' direction.

When splicing of the tripartite leader sequence is studied, early and late cleavage events can be distinguished. The last cleavage occurs at the acceptor site of intron 2 and is intro-

duced in a splicing intermediate from which intron 1 is already excised (Chapter II). Since ligation directly follows acceptor site cleavage, this indicates that ligation of leader 1 to leader 2 precedes ligation of leaders 2 and 3. These results are in agreement with the findings of Keohavong et al. (Nucleic Acids Res. 10, 1215-1229 (1982)). Although the ligation of the tripartite leader sequences thus follows the 5' to 3' direction, cleavages can be introduced in intron 2 before intron 1 is attacked (Chapter II). Around leader 3 the cleavage events occur in an ordered fashion as well. Cleavage at the 3' end of this leader (donor site of intron 3) can only be performed when the 5' end is already cleaved (acceptor site of intron 2) (Chapter IV). Because cleavage at the acceptor site of intron 2 is rapidly followed by ligation of leaders 2 and 3 (Chapter II), ligation of the leader sequences thus precedes ligation of leader 3 to the protein coding sequence. This indicates that ligation of the four exons, which constitue major late mRNAs, follows the 5' to 3' direction. Although cleavage at the donor site of intron 3 may be fast, it is a late event in the splicing pathway because it has to wait for the late cleavage at the acceptor site of intron 2 (Chapter IV).

Possible involvement of the nuclear matrix in the splicing and transport of mRNA.

Introns 1 and 2 can be cleaved internally without concomitant ligation of the bordering exons. This creates the necessity to keep or bring the cleaved-off RNA pieces together in the right orientation in order to make correct ligation of the exons possible after complete removal of the intervening sequences. The requirement for such a factor is emphasized by the finding that rapid cleavage in intron 3 can occur before completion of the tripartite leader (Chapter IV). This indicates that simultaneous cleavages can occur in a pre-mRNA at locations which are separated from each other by more than 18000 nucleotides. The nuclear matrix, a predominantly proteinaceous intranuclear substructure, may provide such a framework, because both cellular and adenoviral hnRNA are quantitatively associated with the matrix (van Eekelen and van Venrooij, J.Cell Biol. 88, 554-563 (1981); van Eekelen et al., Eur.J.Biochem. 119, 461-467 (1981); Chapters V and VI).Analysis of nuclear matrix associated

adenoviral RNA revealed the presence of precursors, intermediates and products of RNA splicing demonstrating that splicing occurs while the RNA is attached to the matrix structure (Chapter V). All the nuclear mRNA sequences which have some relation to the splicing process, including the products of rapid cleavage in either of the three introns, can be found in association with the matrix (Chapter III and V) showing that indeed the nuclear matrix may be involved in fixing the position of the RNA during splicing. Furthermore, all these RNAs can be cross-linked, by in vivo UV-irradiation of intact cells, to the host C-proteins, which are probably identical to the C-type core proteins (Chapter V). The C-proteins are thought to be involved in the binding of the RNA to the nuclear matrix (van Eekelen and van Venrooy, J.Cell Biol. 88, 554-563 (1981)).

Studying the nuclear localization of specific viral mRNA sequences it appears that polyadenylated as well as non-polyadenylated RNAs are almost quantitatively associated with the nuclear matrix (Chapter VI). This demonstrates that not only splicing but also polyadenylation takes place while the RNA is associated with the matrix structure. The same was found for pIX-specific RNA, of which the processing pathway does not include splicing. This indicates that "to be spliced" is obviously not a requirement for association of the RNA with the nuclear matrix. As soon as processing is completed the RNA is rapidly released from the nuclear matrix into the cytoplasm, where it to a certain extent becomes associated with the cytoskeletal framework (Chapter VI). The pIX-specific mRNA sequences have a nuclear dwell time of 4 min, while the major late mRNAs appear in the cytoplasm only about 20 min after transcription (Chapter VI). It is not unlikely that the late cleavage event at the acceptor site of intron 2 is responsible for the long stay of most late mRNA sequences in the nucleus. However, the size of the mRNA may also have its influence on the speed of nucleocytoplasmic transportation (Chapter VI).

SAMENVATTING

In de kern van eukaryotische cellen worden door het RNA polymerase II transcripten gevormd die, om als boodschapper RNA (mRNA) in het cytoplasma te kunnen functioneren, moeten worden gemodificeerd (geprocessed) door o.a "capping", methylering, polyadenylering en "splicing". Wij hebben onze aandacht gericht op RNA splicing. In het bijzonder werd gekeken naar het losknippen van introns uit pre-mRNAs, die overgeschreven zijn van het "major late" transcriptie gebied van het adenovirus genoom. De mRNAs die in dit gebied worden gecodeerd zijn samengesteld uit vier exons: leader 1, waar de 5' cap aan vast zit, leader 2, leader 3 en het gedeelte dat voor een eiwit codeert en dat ook de poly(A)-staart draagt. Eenmaal met elkaar verbonden worden de drie leaders samen de "tripartite leader" genoemd. Drie introns, genummerd 1, 2 en 3, scheiden de exons van elkaar in de pre-mRNAs: intron 1 ligt tussen leader 1 en 2 in, intron 2 tussen leader 2 en 3, en intron 3 ligt tussen leader 3 en het exon dat voor een eiwit codeert.

Enkele aspecten van het splicing mechanisme.

In het kern RNA zijn intermediairen van de processing aanwezig, die een deel van een intron missen, wat er op wijst dat het uitknippen van zowel intron 1 als intron 2 in stappen verloopt in de 5' - 3' richting (hoofdstuk III). De plaatsen waar in het intron geknipt kan worden, kunnen onderscheiden worden in voorkeursplaatsen en niet-voorkeursplaatsen. Het knippen op voorkeursplaatsen kan al optreden in een RNA molecuul dat nog bezig is gesynthetiseerd te worden (hoofdstuk II). Kennelijk kan splicing aan polyadenylering vooraf gaan. Omdat een intron meerdere voorkeurs-knipplaatsen kan bevatten, kan de eerste knip op verschillende plaatsen in het intron worden aangebracht. Dergelijke alternatieve wegen voor splicing treden op bij het verwijderen van introns 1 en 2 (hoofdstuk II). Op de grens van intron en exon (acceptor site) vindt de laatste knip plaats, die snel gevolgd wordt door het ligeren van de exons (hoofdstukken II en III). Op grond van onze bevindingen wordt in hoofdstuk III een model voor splicing gepresenteerd.

Splicing geschiedt in de 5'-3' richting.

Bij de vorming van de tripartite leader via splicing kan er onderscheid worden gemaakt tussen vroege en late knippen. De laatste knip wordt aangebracht op de acceptor site van intron 2. Dit gebeurt in een RNA-molecuul waar intron 1 al uit verwijderd is (hoofdstuk II). Aangezien het knippen van een acceptor site direct gevolgd wordt door ligering van de exons, gaat het ligeren van leader 1 met leader 2 kennelijk vooraf aan het ligeren van leaders 2 en 3. Dit resultaat is ook verkregen door Keohavong en medewerkers (Nucl. Acids Res. 10, 1215-1229 (1982)). Hoewel het ligeren van de leaders dus in 5'-3' richting verloopt, kan er toch in intron 2 geknipt worden voordat intron 1 wordt aangetast (hoofdstuk II). Ook rondom leader 3 wordt in een bepaalde volgorde geknipt. Op het 3' eind van deze leader (donor site van intron 3) kan slechts worden geknipt, indien het knippen op het 5' eind (acceptor site van intron 2) al heeft plaatsgehad (hoofdstuk IV). Omdat knippen op de acceptor site van intron 2 snel gevolgd wordt door ligering van leader 2 met leader 3 (hoofdstuk II), gaat het ligeren van de leaders onderling dus vooraf aan het ligeren van leader 3 met het eiwitcoderende exon. Ligeren van de vier exons, waaruit de mRNAs van het major late gebied zijn opgebouwd, gebeurt geheel in de 5'-3' richting. Hoewel het knippen op de donor site van intron 3 dus snel gebeurt, is het toch een late stap in het splicing proces omdat er gewacht moet worden op de late knip op de acceptor site van intron 2 (hoofdstuk IV).

De mogelijke betrokkenheid van de kernmatrix bij splicing en transport van mRNA.

In introns 1 en 2 kan worden geknipt zonder dat dit gepaard gaat met ligering van de exons. Daardoor ontstaat de behoefte om de losgeknipte stukken RNA op de een of andere wijze bij elkaar te houden of bij elkaar te brengen, en wel in een zodanige oriëntatie dat de exons, nadat het intron in zijn geheel verwijderd is, op de juiste manier worden geligeerd. Daarbij komt dat ook in intron 3 al kan worden geknipt voordat de tripartite leader klaar is (hoofdstuk IV). De knippen in het pre-mRNA kunnen in feite gelijktijdig worden aangebracht op plaatsen die meer dan 18000 nucleotiden van elkaar verwijderd zijn. De kern-

matrix, een structuur die voornamelijk is opgebouwd uit eiwit, zou in een dergelijke behoefte kunnen voorzien, omdat zowel cellulair als adenovirus heterogeen nucleair RNA, waar het pre-mRNA een onderdeel van is, kwantitatief met de matrix geassocieerd zijn (van Eekelen en van Venrooij, J.Cell Biol. 88, 554-563; van Eekelen et al., Eur.J.Biochem. 119, 461-467 (1981); hoofdstukken V en VI). Bij analyse van adenovirus-specifiek matrix RNA werden precursors, intermediairen en producten van splicing aangetroffen, waaruit kan worden geconcludeerd dat splicing optreedt terwijl het RNA aan de matrix gebonden is (hoofdstuk V). Alle mRNA-sequenties in de kern, die enige relatie hebben met het splicing proces, daarbij inbegrepen de producten die ontstaan door snel knippen in elk van de drie introns, zijn geassocieerd met de matrix (hoofstukken III en V). Hieruit volgt dat de kernmatrix inderdaad betrokken kan zijn bij de oriëntatie van het RNA gedurende het splicing-gebeuren. Bovendien kunnen al deze RNAs, door middel van UV-bestraling van intacte cellen, covalent worden gebonden aan gastheer eiwitten, die waarschijnlijk identiek zijn met de C-type hnRNP core eiwitten (hoofdstuk V). Deze eiwitten spelen vermoedelijk een rol bij de binding van het RNA aan de kernmatrix (van Eekelen en van Venrooij, J.Cell Biol. 88, 554-563 (1981)).

Bij het bestuderen van de lokalisatie van specifieke virale mRNA-sequenties in de kern bleek dat zowel gepolyadenyleerde als niet-gepolyadenyleerde RNAs nagenoeg kwantitatief geassocieerd zijn met de kernmatrix (hoofdstuk VI). Hieruit volgt dat niet alleen splicing, maar ook polyadenylering optreedt terwijl het RNA geassocieerd is met de matrix-structuur. Eenzelfde resultaat werd gevonden voor pIX-specifiek RNA, waarvan het mRNA niet gevormd wordt via splicing. "Splicing ondergaan" is klaarblijkelijk niet vereist voor het binden van RNA aan de matrix. Meteen na het voltooien van de processing wordt het RNA van de kernmatrix losgemaakt en gaat het naar het cytoplasma, waar het voor een bepaald percentage geassocieerd wordt met het cytoskelet (hoofdstuk VI). Het pIX-specifieke RNA heeft een verblijftijd van slechts 4 minuten in de kern, terwijl mRNAs afkomstig van het major late gebied pas zo'n 20 minuten na transcriptie in het cytoplasma verschijnen. Het is niet onwaarschijnlijk dat het vertraagde knippen op de acceptor site van

intron 2 de oorzaak is van het lange verblijf in de kern van de meeste late mRNA-sequenties. Ook de grootte van het mRNA zou echter van invloed kunnen zijn op de transportsnelheid (hoofdstuk VI).

CURRICULUM VITAE

Edwin C.M. Mariman werd geboren op 24 februari 1955 te Clinge. Na het behalen van het diploma Gymnasium B (Jansenius Lyceum, Hulst) begon hij in september 1973 zijn studie Chemie aan de Katholieke Universiteit van Nijmegen. Het kandidaatsexamen Chemie (S2) werd afgelegd in juni 1976 met het predikaat "cum laude". Het doctoraal examen Chemie, met als hoofdvakken Biochemie (Prof. dr. H. Bloemendal) en Organische Chemie (Prof. dr. R.J.F. Nivard) en met als uitbreiding Quantum Chemie (Prof. dr. ir. A. van de Avoird), werd afgelegd in maart 1979 met het predikaat "cum laude". Tijdens de studie werd in november 1978 aan hem de Unilever Chemie Prijs voor Nijmegen verleend. Van april 1979 tot en met maart 1983 was hij werkzaam als wetenschappelijk medewerker op het laboratoriom voor Biochemie van de Katholieke Universiteit. In deze periode werd onder auspiciën van de Stichting Scheikundig Onderzoek in Nederland (SON) en met financiële steun van de Nederlandse Organisatie voor Zuiver Wetenschappelijk Onderzoek (ZWO) het onderzoek verricht dat in dit proefschrift beschreven staat.

In november 1982 is hij gehuwd met Juanita R. Quiño te Danao City, Cebu City (Filippijnen). Op 27 september 1983 werd hun dochter Rayna geboren.

Vanaf april 1983 is hij werkzaam als wetenschappelijk medewerker op het laboratorium voor Biochemie van de Katholieke Universiteit, waar hij onderzoek verricht naar de mogelijke effecten van antilichamen op de RNA/eiwit synthese bij patiënten die lijden aan bepaalde reumatische aandoeningen.

Edwin C.M. Mariman was born February 24, 1955 at Clinge, The Netherlands. In 1973 he graduated from the "Jansenius Lyceum" at Hulst, The Netherlands. In September of the same year he started his chemistry studies at the Catholic University of Nijmegen, Nijmegen, The Netherlands. He obtained his bachelor's degree with distinction in June 1976. In March 1979 he obtained his master's degree with distinction majoring in biochemistry and organic chemistry with theoretical chemistry as minor subject. During his study in November 1978 he obtained the Unilever Chemistry Award, Nijmegen. In april 1979 he joined the Department of Biochemistry of the Universitey of Nijmegen, Nijmegen,

The Netherlands, to study RNA processing in eukaryotic cells. The investigations, as described in this thesis, were carried out under auspices of the Netherlands Foundation for Chemical Research (SON) and with financial aid of the Netherlands Organization for the Advancement of Pure Research (ZWO).

In November 1982 he married Juanita R. Quiño at Danao City, Cebu City (Philippines). On September 27, 1983 their daughter Rayna was born.

In April 1983 he started to investigate the influence of antibodies from patients, suffering from certain rheumatic affections, on RNA/protein synthesis (supported by the Netherlands League against Rheumatism). This work is carried out at the Department of Biochemistry of the University of Nijmegen, Nijmegen, The Netherlands.

STELLINGEN

1

De bewering van Lenk et al. dat de verlaagde aanmaak van bepaalde virale eiwitten een gevolg is van de introductie van anti-U1 RNP antilichamen in geinfecteerde cellen wordt in twijfel getrokken door de aanwezigheid van anti-virale antilichamen in humaan serum.

Lenk, R.P., Maizel, J.V. and Crouch, R.J. (1982) Eur. J. Biochem. 121, 475-482.
Habets, W.J., den Brok, J.H. Boerbooms, A.M.T., van de Putte, L.B.A. and van Venrooij, W.J. (1983) EMBO Journal 2, 1625-1631.

2

Door enkel gebruik te maken van het ampholine bereik 3.5-10 bij hun twee-dimensionale gels, verkrijgen Garber en Gold een zodanig slechte resolutie dat een goede karakterisering van de water-oplosbare eiwitfractie van ooglenzen van rund en muis door hen niet kan worden gedaan.

Garber, A.T. and Gold, R.J.M. (1982) Exp. Eye Res. 35, 585-596.

3

De opvallende overeenkomst van de EBERs met de VA-RNA's wat betreft functie, genetische organisatie en expressie, duidt op een convergente evolutie van de herpes-virussen en de adenovirussen.

Rosa, M.D., Gottlieb, E., Lerner, R. and Steitz, J.A. (1981) Molec. Cell. Biol. 1, 785-796.
Bath, R.A. and Thimmappaya, B. (1983) Proc. Natl. Acad. Sci.USA 80, 4789-4793.

4

Volgens de konventionele r- en l-streng benaming, die gebruikt wordt bij viraal dubbel strengs DNA, is bij de presentatie van de nucleotide-sequentie van een deel van het adenovirus genoom door Gingeras et al. de afgebeelde streng foutief benoemd.

Gingeras, T.R., Sciaky, D., Gelinas, R.E., Bing-Dong, J., Yen, C.E., Kelly, M.M., Bullock, P.A., Parsons, B.L., O'Neill, K.E. and Roberts, R.J. (1982) J. Biol. Chem. 22, 13475-13491.

5

De veronderstelling van Wallace en Edmonds, dat gepolyadenyleerde kern RNA's welke vertakt zijn via een 2'-5' fosfodiester binding intermediairen van pre-mRNA splicing zijn, is onwaarschijnlijk.

Wallace, J.C. and Edmonds, M. (1983) Proc. Natl. Acad. Sci.USA 80, 950-954.

6

Het is onjuist dat de intrinsieke klaring door Keiding en Andreasen wordt beschouwd als een factor die afhankelijk is van de bloedstroom.

Keiding, S. and Andreasen, P.B. (1979) Pharmacology 19, 105-110.

7

Het feit dat door Lerner en Steitz met anti-RNP en anti-Sm sera zowel snRNA's als een zevental eiwitten geprecipiteerd kunnen worden is onvoldoende om te stellen dat het componenten zijn van één RNP complex.

Lerner, M.R. and Steitz, J.A. (1979) Proc. Natl. Acad. Sci.USA 76, 5495-5499.

8

Niet alle RNA moleculen welke door Fraser et al. worden toegeschreven aan premature terminatie van transcriptie zijn het gevolg van "attenuation".

Fraser, N.W., Seghal, P.B. and Darnell, J.E. (1979) Proc. Natl. Acad. Sci.USA 76, 2571-2575.

9

De toenemende belangstelling voor een natuurlijke leefwijze zal in de komende jaren het aantal gevallen van ingewandskanker doen teruglopen.

10

Om verwarring te voorkomen tussen het Engelse "to splice" en het Nederlandse "splitsen" zou als vertaling een van de werkwoorden "splassen", als combinatie van splitsen en lassen, of "spligeren", als combinatie van splitsen en ligeren, kunnen worden gebruikt.

11

De stimulering van het onderzoek op het gebied van reumatische aandoeningen dient niet voorbij te gaan aan de konstatering dat bij 41 procent van de reuma-patiënten verbetering van de klachten optrad nadat ze gedurende drie maanden een stuk malachiet op de huid boven het hart hadden gedragen.

Reuma Bulletin 2 (1983).

12

Omdat bij een vergelijking tussen Nederland en de Filippijnen meteen opvalt dat ons systeem van openbaar vervoer uiterst beperkt is, zou moeten worden nagegaan in hoeverre het Filippijnse systeem hier ingevoerd kan worden.

Nijmegen, 14 december 1983

E.C.M. Mariman.