# Sequence dependent interaction of hnRNP proteins with late adenoviral transcripts

Chris van Eekelen, Rolf Ohlsson*, Lennart Philipson+, Edwin Mariman, Rita van Beek and Walther van Venrooij

Department of Biochemistry, University of Nijmegen, Geert Grooteplein Noord 21, 6525 EZ Nijmegen, The Netherlands, and Department of Microbiology, Uppsala University, Biomedical Center, Uppsala, Sweden

Received 22 September 1982; Accepted 21 October 1982

## ABSTRACT

Irradiation with ultraviolet light was used to induce covalent linkage between hnRNA and its associated proteins in intact HeLa cells, late after infection with adenovirus type 2. Covalently linked hnRNA-protein complexes, containing polyadenylated adenoviral RNA, were isolated and their protein moiety characterized. Host 42,000 Mr hnRNP proteins proved to be the major proteins crosslinked to viral hnRNA. To investigate their possible involvement in RNA processing, the localization of these cross-linked polypeptides on adenoviral late transcripts was determined. Sequences of RNA around the attachment sites of the protein were isolated. After in vitro labeling they were hybridized to Southern blots of adeno DNA fragments. The hybridization patterns revealed that the 42,000 Mr polypeptides can be linked to adenoviral transcripts over the entire length of the RNA, corresponding to 16.2-91.5 m.u. of the viral genome. Fine mapping within the Hind III B region (16.8-31.5 m.u.) established, however, that the localization of the cross-linked polypeptides was not random in all parts of the transcript. Sequences around the third leader and the 3' part of the i-leader were overrepresented, whereas the regions encoding VA I and VA II RNA and the late region 1 mRNA bodies were underrepresented in the cross-linked RNA. Using genomic DNA fragments and a cDNA clone containing the tripartite leader it appeared that leader and intervening sequences were represented about equally in cross-linked RNA fragments. Although these results do not support the notion that introns or exons are specifically interacting with one RNP protein, they demonstrate that the 42,000 hnRNP proteins are non randomly positioned on the RNA sequence.

## INTRODUCTION

The structural organization of heterogeneous nuclear RNA (hnRNA)(reviewed in 1 and 2) has been studied extensively in order to understand the mechanism of RNA processing and its regulatory role in the control of gene expression (3).All hnRNA,also those containing mRNA sequences, are associated with protein to form ribonucleoprotein particles (1,2,4). HnRNA-protein complexes, generally referred to as hnRNP particles, can be isolated from nuclei by extraction at elevated pH or by 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)





0305-1048/82/1022-7115$ 2.00/0

can be recovered, whereas after mild nuclease degradation 30-40 S monoparticles accumulate (2,5). The polypeptide composition of the isolated complexes is also dependent on the isolation procedure. Polypeptide compositions ranging from almost all non-histone nuclear proteins to a few distinct polypeptides have been reported (1,2,6-10). The protein composition of purified monoparticles, which are considered as structural building blocks of hnRNP, is relatively simple. 30 S particles from HeLa cells contain 6 major polypeptides, subdivided in three groups,each with two polypeptides,called A,B and C proteins and ranging in molecular weight from 32,000-44,000 dalton (10-12).

A higher level of structural organization of hnRNA has also been encountered. If endogenous nuclease and protease activity is prevented during isolation, hnRNA can be isolated in association with a chromatin-free structural network, the nuclear matrix (13-16). The interaction between hnRNA and the matrix is probably mediated by the hnRNP C-proteins (16). Precursor RNAs as well as intermediates and products of processing have been found associated with hnRNP C-proteins, implicating that processing takes place while hnRNA is bound (17).

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 (18-20). The function of these double stranded structures is still unresolved, but it is of interest that snRNAs appeared to be associated with hnRNP (21-23). In the case of U1 snRNA it has been proposed that this RNA is involved in splicing by base pairing to splice junctions in the hnRNA molecule (24-27).

An important experimental approach in studying hnRNP structure involves limited nuclease digestion, which has been applied to determine the RNA sequences protected by and thereby associated with proteins. Although no discrete sequence specificity was found using a total cellular hnRNP population (28), Munroe and Pederson (29) found, following this strategy, that mRNA sequences in hnRNP are complexed with proteins. No discrete lengths of protein protected RNA were obtained upon limited ribonuclease digestion distinguishing hnRNP organization from the repeating subunit structure of chromatin (30). 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 (5). 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 feasible to isolate intact pre-mRNA from hnRNP particles (31,32) and from nuclear matrix preparations, in which the hnRNA is also associated with a specific set of hnRNP proteins (16,17). Recently we have reported on the purification of photochemically cross-linked hnRNA-protein complexes by hybridization to immobilized DNA probes (31,33,34).

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 not random with respect to the RNA sequence (35,36).

Steitz and Kamen (37) have studied RNP organization in 30 S 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. (31) indicate that within a specific RNA precursor there is a non-random protection of RNA sequences by RNP proteins.

In this study photochemical RNA-protein cross-linking in vivo was used to localize specific hnRNP proteins on late adenoviral transcripts. This experimental approach was chosen to circumvent problems that can arise by non-specific binding of proteins to hnRNA or rearrangements of hnRNP during isolation. It enabled us to determine the localization of two distinct hnRNP proteins on pre-mRNA molecules.

## MATERIALS AND METHODS

### Isolation of photochemically cross-linked hnRNA-protein complexes

HeLa S3 cells were cultured and infected with adenovirus type 2 as described (16,17,33). Cells 18 h post infection, were pulse labeled (15 min) with (10 μCi/ml) 5,6-$^{3}$H uridine and (10 μCi/ml) 5-$^{3}$H cytidine (Amersham International, specific activities 45 and 31 Ci/mmol, respectively). The cells were then concentrated and irradiated for 5 min with ultraviolet light (254 nm) as described (16,38) to induce RNA-protein cross-linking. Subsequently nuclear matrices containing more than 80% of the pulse labeled RNA were prepared following the procedure described by van Eekelen et al.(16). These matrices were dissolved by heating them for 2 min at 90$^{o}$C in 1% SDS and the solubilized material was subsequently applied to an oligo dT-cellulose column to select

for polyadenylated hnRNA and hnRNA-protein complexes (39). The selection procedure was performed twice. SnRNA or other small molecular weight RNA, possibly base paired to hnRNA, was lost during this procedure by melting the samples before application to the oligo dT-cellulose column. 15-20% of the incorporated $^{3}H$ label applied to the column was present in the polyadenylated material.

Preparation of the RNA probes

Covalently linked polyadenylated hnRNA-protein complexes were treated with 100 U/ml Micrococcal nuclease ( P&L Biochemicals Inc.) for 30 min at 37°C to generate RNA fragments of lengths less than 100 nucleotides. After this digestion two classes of RNA fragments can be distinguished: free RNA fragments and RNA fragments cross-linked to protein. These classes were carefully separated by phenol/chloroform extraction as indicated in figure 1. The cross-linked RNA fragments were subsequently deproteinized (see fig. 1). Both RNA preparations, the cross-linked fragments (Q) and the unlinked fragments (W), were stored dissolved in aqueous buffer, so that, when needed, portions could be taken for 5' end labeling and hybridization. Hybridization was performed under identical conditions for both RNA preparations and with exactly the same quantities of RNA ( calculated from the $^{3}H$ label incorporated in the RNA). After 5' end labeling (40) the reaction mixture was phenol/chloroform extracted and to the water phase tRNA (25 µg/ml), NaCl (200 mM) and 3 volumes of ethanol were added. After 1 h at -30°C the precipitated RNA was centrifuged (10 min, 10,000xg), then dissolved and precipitated again. This procedure minimizes contamination of the probes with free $\gamma$-$^{32}P$-ATP .

Preparation and hybridization of blots

Adenoviral genomic DNA was isolated as described by Petterson and Sambrook (41). Plasmids were propagated in E.coli strain JA 221 and extracted and purified as described by Aleström et al. (42). All restriction enzymes were purchased from Bethesda Research Lab.Inc.. Restriction digests containing DNA fragments larger than 1 kb were analyzed on agarose gels and the DNA was subsequently transferred to nitrocellulose as described by Southern (43). Smaller DNA fragments (0.14-1.5 kb) were separated on mixed agarose/acrylamide gels (0.4% / 4%). After staining and photography, the gels were treated with 0.2 M NaOH and 0.6 M NaCl for 30 min and then washed extensively with 0.05 M phosphate buffer pH 6.8 (4x30 min).The DNA was then electrophoretically transferred (16 h, 1 A) in this low salt buffer to Gene Screen (New England Nuclear) using a Biorad Transblot Cell. All blots were baked at 80°C for 4 h in a vacuum oven. Gene Screen proved to bind small DNA fragments (140-1500 bp)

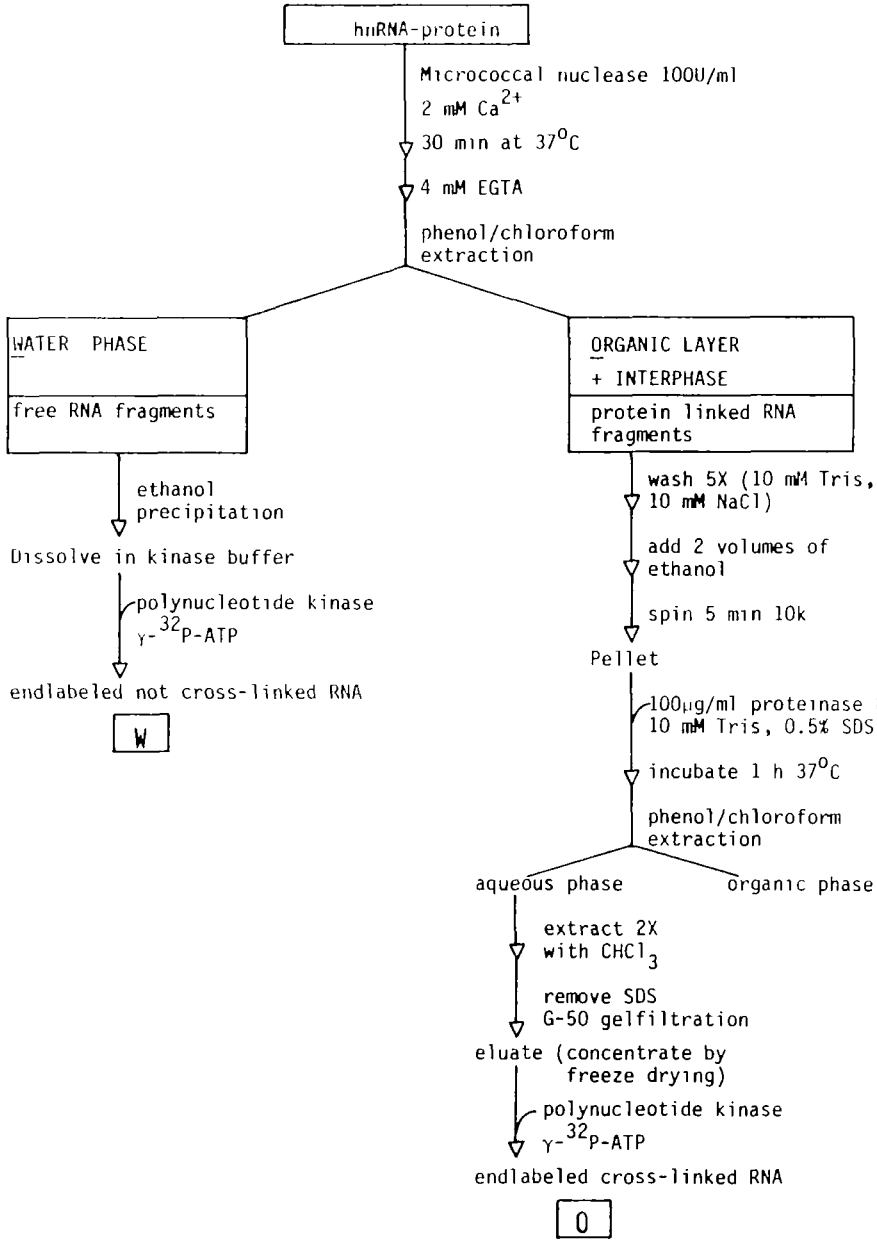


Figure 1

Scheme of isolation of hnRNA fragments cross-linked to proteins (O) and control, not protein-linked RNA fragments (W).

efficiently under these conditions. The blots were prehybridized for 4 h at 42°C in 50% formamide, 5xSSC, 250 µg/ml ssDNA, 500 µg/ml yeast RNA, 5x Denhardts mix, 0.1% SDS, 10 mM phosphate buffer pH 6.5. Hybridization of the blots with endlabeled RNA fragments was for 24 h at 42°C in a buffer of similar composition except that the concentrations of ssDNA (100 µg/ml), RNA (250 µg/ml) and Denhardts mix (2x) were reduced. The amount of DNA on the blots (5 µg/ lane) was always in large molar excess over the RNA probe (± 10 ng). After hybridization the blots were washed with hybridization buffer (2x 30 min at 42°C), 2x SSC (2x 30 min at room temperature), 0.2xSSC (2x 30 min at room temperature) then dried and exposed to X-ray film. The specificity of hybridization was etablished from the fact that plasmid DNA fragments never annealed with the RNA probes.

Quantitative analysis of blot hybridization results

To obtain a more quantitative interpretation of the extent of hybridization of the labeled probes O and W to the DNA bands on the blots, autoradiographs of these blots were scanned using a Vitatron TLD 100 densitometer. From these scans the relative intensity of each band as compared with the total intensity of all bands together, was calculated. For example, band E of the Hin II/ Eco RI digest in fig.5A represents 12% of the total intensity on the autoradiograph when probe W is used, and 35% when probe O is used. So the ratio (O/W) is 2.9, which indicates an about 3-fold overrepresentation of RNA complementary to this part of the DNA in cross-linked RNA. Autoradiographs were used only when the intensity of the bands was proportional to the exposure time. Scanning of autoradiographs, exposed for different periods of time, resulted in the same relative intensities for the individual bands, within the experimental error (< 10%).

## RESULTS

### Characterization of proteins cross-linked to hnRNA

Ultraviolet light induced RNA-protein cross-linking, performed on intact cells, is a specific way of identifying the proteins that interact with RNA. With this method copurification of proteins non-specifically associated with RNA can be avoided (38). For the detection of the cross-linked proteins either $^{35}S$ methionine labeling of cellular proteins or indirect labeling using $^{3}H$ nucleosides was used (16,17,38,44). Recently we showed that an in vitro labeling procedure, in which $\gamma$-$^{32}P$-ATP and polynucleotide kinase are used to 5' end label RNase T1 treated RNA-protein complexes, can be used to enhance the sensitivity of detecting cross-linked proteins (31,34). Figure 2 shows an

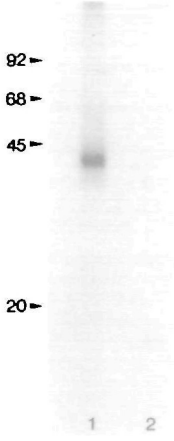


Figure 2. Analysis of polypeptides cross-linked to poly A containing hnRNA. Covalently linked poly A(+) hnRNA-protein complexes, isolated as described in the Materials and Methods section, were incubated with RNase T1 and the resulting T1-oligonucleotide-protein complexes were 5' end labeled and analyzed on 13% SDS polyacrylamide gels as described (34). The molecular weight of marker proteins ($x10^{-3}$), run on the same gel, are indicated. Lane 1: $^{32}P$ labeled, hnRNA derived, oligonucleotide-protein complex. Lane 2: as lane 1, after extensive treatment with RNase A.

example of such an experiment where proteins cross-linked to hnRNA of cells late after infection with adenovirus were identified using in vitro $^{32}P$ labeling. Polypeptides of 42,000 dalton in molecular weight are the most prominent proteins. In earlier studies using $^{3}H$ labeling we have shown that this class of proteins consists of two polypeptides with molecular weights of 41.500 and 43,000, which correspond to hnRNP C-proteins (16). Both adenoviral and host hnRNA are associated with these polypeptides (17,31,33).

*Isolation of RNA fragments linked to hnRNP proteins*

We took advantage of the simple protein composition of cross-linked hnRNP in determining the localization of the proteins on adenoviral hnRNA. To isolate the RNA fragments around the covalent linkage site between RNA and protein, the complexes were partially digested with Micrococcal nuclease (see fig.1). This digestion was monitored making use of the $^{3}H$ label, incorporated in the hnRNA prior to UV-irradiation and harvesting. The distribution of the $^{3}H$ label between the water phase on one hand and the phenol and interphase on the other after nuclease degradation reflects the extent of degradation. In the optimal case (fig.1) around 4% of the $^{3}H$ label, compared with undigested samples, moved into the organic plus interphase fraction. Since probably more than one polypeptide is linked to each RNA molecule, a cross-linked RNA fragment constitutes on the average less than 4% of the length of the original RNA molecule. As the hnRNA-protein complexes were treated with SDS during oligo dT-cellulose selection, the RNA and protein interact only at the covalent attachment site and no protection of the RNA against nucleases by linked proteins is to be expected. This is demonstrated in figure 3, in which the two endlabeled probes (O and W, Organic phase RNA fragments, containing cross-

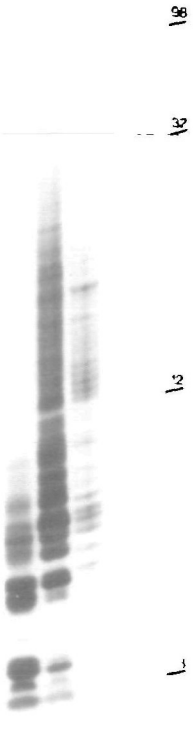


Figure 3. Determination of the length of the RNA probes. From HeLa cells 18 h post infection with adenovirus type 2, polyadenylated hnRNA-protein complexes were prepared. These complexes were fragmented by partial digestion with Micrococcal nuclease and the protein-linked (O) and unlinked fragments (W) were isolated and 5'endlabeled as depicted in figure 1. Samples of these probes were applied onto a 20% acrylamide, 7 M urea gel. The figure shows an autoradiograph of such a gel. The figures correspond to the lengths in nucleotides of marker molecules. Lane C: control, consisting of a sample prepared from a reaction mixture, to which no RNA was added. Lane O: endlabeled cross-linked RNA fragments. Lane W: endlabeled unlinked RNA fragments.

linked RNA, and Water phase RNA, unlinked RNA) were analyzed on acrylamide gels. Both probes have similar length distributions. Fig.3, lane C show the gel pattern obtained when a sample from a control 5' endlabeling reaction , in which no RNA was added, was applied. The oligonucleotide bands in this lane represent probably contaminations in commercial $\gamma$-$^{32}$P-ATP preparations or polymerization products of ATP generated during the incubation with polynucleotide kinase. The mean length of both RNA probes are therefore 10 to 30 nucleotides. Since partial degradation of the RNA fragments during 5' endlabeling occurs, the original length of the isolated RNA fragments will be slightly larger.

Localization of cross-linked RNA fragments on the adenoviral genome

To study the structural organization of specific transcripts we used cells late in adenovirus infection. A majority of the transcripts is virus specific at this time and the processing pathways of these transcripts is also known. Most transcripts are derived from the major late promoter and processing of these pre-mRNAs results in five families of 3' coterminal mRNAs. These mRNAs contain a 5' multiple leader region spliced to a body containing the coding RNA (50).

The hybridization patterns of the labeled RNA probes W and O to blots with adeno DNA restriction fragments resulting from cleavage with Xho I and Hind III are shown in figure 4. For comparison the hybridization patterns of polyadenylated cytoplasmic mRNA (fragmented with Micrococcal nuclease as for hnRNA) to the same blots are shown.

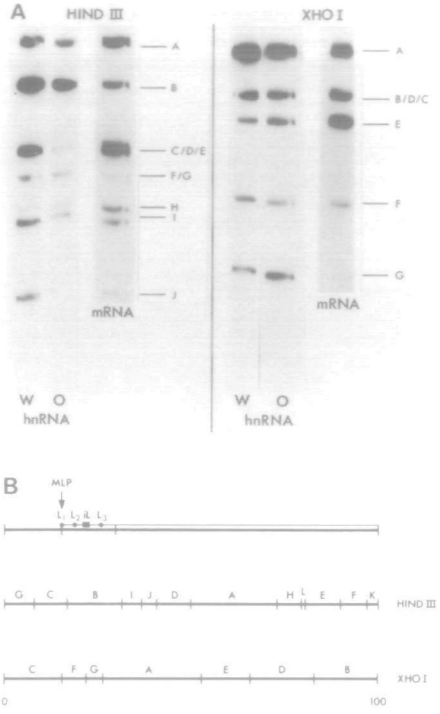


Figure 4. Localization of cross-linked hnRNA fragments on the adenoviral genomic map. A. Hybridization patterns of 5' endlabeled RNA probes O and W and fragmented poly A(+)mRNA to Southern blots of Xho I and Hind III restricted adeno DNA. B. Adenovirus DNA map showing the positions of the late promoter (MLP) and late transcripts (upper line) and the cleavage sites for Xho I and Hind III (lower lines).

First, the composition of the hnRNA probes is different from that of the mRNA probes. For example, the Hind III B fragment is recognized strongly by the hnRNA probes and only weakly by the mRNA probes, when compared with the hybridization to the Hind III A fragment. On the Xho I blot, hnRNA fragments hybridize strongly to A,F and G fragments and weakly to B/D/C and E when compared with mRNA fragments. This pattern is expected if the polyadenylated hnRNA consists predominantly of precursor RNA and intermediates of RNA processing, which both contain intervening sequences between the three leader segments or between the third leader and the body of the mRNAs. The nuclear RNA was not contaminated with cytoplasmic mRNA as revealed by RNA-protein cross-linking experiments, because other polypeptides are cross-linked to mRNA than to hnRNA (44). The main polypeptide cross-linked to poly A(+)mRNA is a 52,000 dalton protein (34,38). This polypeptide is readily detected when comparable amounts of cross-linked mRNA-protein were analyzed (34), but was not found among the hnRNA-linked proteins (fig.2).

From the hybridization patterns of figure 4 it can be deduced that cross-linked RNA fragments (O) annealed to all DNA fragments corresponding to se-

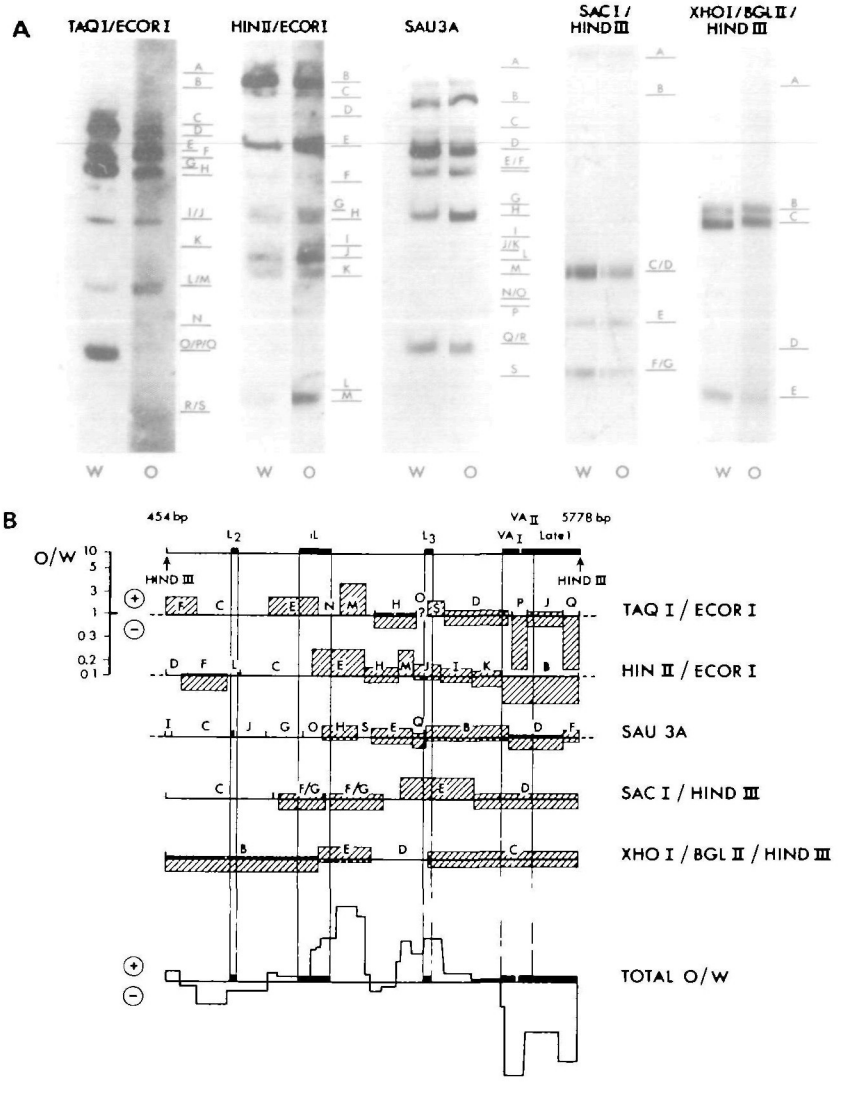


Figure 5. Fine mapping of protein linked hnRNA fragments within the Hind III B region of the adeno genome. A. Hybridization patterns of the RNA probes O and W to blots of restriction digests of pBR 322 plasmid, containing the Hind III B fragment of adeno DNA in its Hind III site. The restriction enzymes used are indicated in the figure. The letters next to the hybridization patterns correspond to the positions of ethidium bromide stained DNA bands. Blotting of DNA fragments from 1.5% agarose gels was performed onto nitrocellulose in the case of Xho/Bgl/Hind and Sac/Hind digestion. In all other cases, where smaller DNAs were separated, electrophoretic transfer of the DNA fragments from mixed 4% acrylamide/0.4% agarose gels to Gene Screen was used.

Figure 5B. Schematic representation of relative over or underrepresentation of regions within the Hind III B fragment in cross-linked RNA compared with unlinked RNA. The hatched blocks represent relative over (+) or underrepresentation (-) depending on their position to the zero level. The scale used to quantify this ratio O/W is shown on the left. The level of over or underrepresentation is calculated as described in the Materials and Methods section and forms the quantitative interpretation of the experiments shown in figure 5A. The lower line is obtained by adding up the hatched blocks for the different areas and depicts total over or underrepresentation of cross-linked hnRNA compared with unlinked RNA (total O/W).

quences that are known to be expressed and present in poly A(+)hnRNA at this time after infection (from 16.2-91.5 m.u.). This means that RNP proteins are associated with the late transcripts at several locations from the 5' to the 3' end of the RNA (see fig.4). These results were confirmed by hybridization of the same probes to Southern blots of adeno DNA fragments using other restriction enzymes (Bgl II, Eco RI)(data not shown). The main difference between the probe O and W seemed to be that the region from 30-50 m.u. was relatively underrepresented in the cross-linked RNA, but the region around the tripartite leader represented by the Hind III B fragment and the Xho G fragment were found to be cross-linked more frequently.

Localization of the protein-linked RNA fragments within the Hind III B region

In order to map the RNA-protein cross-linking sites more accurately we decided to investigate the part of the transcripts derived from the Hind III B region (16.8-31.5 m.u.) in more detail. Intron and exons have been exactly defined in this region (50). Exons are found in the leaders 2,3, the i-leader, and the region from 30.0 m.u. to 31.5 m.u. which encode late region 1 RNAs (for map see fig.4B and 5B). Other sequences represent intervening sequences. In this study we have considered 30.0 m.u. as the most extended 5' end of the late region 1 mRNA-bodies, although less abundant RNAs, whose translation products have not yet been identified, map with their 5' end of the body at 26.4 or 28.4 m.u. (45). Hybridization patterns of the probes O and W to blots of small restriction fragments of a plasmid containing the Hind III B fragment are shown in figure 5A. Plasmid DNA bands (pBR 322) show no hybridization with either RNA probe. Furthermore it is clear that not all adeno DNA containing bands, even when they are of similar size, are recognized to the same extent (see discussion). Only in two cases the level of hybridization to DNA bands can be low, due to poor recovery of the DNA on the Gene Screen, namely in the case of band S of the Taq/Eco RI digest and band S in the Sau 3A digest, because of the limited lengths of these fragments (142 and 105 bp, respectively).

Concerning the localization of the cross-linked hnRNP C-proteins on late adenoviral RNA precursors, figure 5B shows the differences in hybridization patterns of the cross-linked RNA fragments O and unlinked RNA fragments W in a schematic representation. The intensity of hybridization of probe O is expressed in blocks above (+) or below (-) the zero level depending on whether the O probe hybridized more or less than the W probe. This was quantitated as described in the Materials and Methods section. In duplicate experiments, using independently prepared RNA probes, similar results were obtained (data not shown). The lower line of the scheme summarizes the interpretation of the five hybridization experiments, by adding up the hatched blocks. This gives a general picture where the cross-linked RNA fragments are localized on the genome and therefore identifies the preferred binding sites for hnRNP C-proteins on late adenoviral transcripts.

Cross-linking efficiency of exons versus introns

Exact mapping of the RNA fragments cross-linked to hnRNP proteins with respect to the leader sequences is not possible in the experimental set up shown in fig.5 because of the small size of the leaders (41,72 and 90 nucleotides). Therefore another approach was used to determine if leader sequences are under or overrepresented in cross-linked RNA compared with the introns between them. For this analysis blots were prepared from a gel, on which the F and G fragments of Xho I restricted adeno DNA, mapping at 15.5-22.4 and 22.4-26.4 m.u. and a plasmid containing cDNA derived from fiber mRNA (a kind gift of Dr. M. Mathews, see Zain et al. (47)), digested with Sal I and Xho I , had been electrophoresed. The latter digestion gives a fragment containing PBR sequences and 169 nucleotides of cDNA sequence, consisting of leader 1,2 and the 5' part of leader 3 (band P in figure 6). This fragment contains exactly those exon sequences which are present in the genomic Xho fragments F and G, namely leader 1 and 2 in fragment F and the 5' 56 nucleotides of leader 3 in fragment G. The hybridization patterns of probes O and W to this blot (fig.6) show about equal hybridization of both probes to band P, when compared to the combined hybridization to F and G. The relative amounts of hybridization of both probes to band P on the blots were determined by densitometric scanning of two sets of autoradiographs (prepared independently). Probe W hybridized to band P for 7 $\pm$ 2%, probe O for 9 $\pm$ 2%. This result indicates that the leader sequences were cross-linked to protein in about equal proportion compared with the intervening sequences between them. Thus, the overrepresentation of RNA complementary to the Xho G fragment (see fig.4 and 6) is due to the presence of cross-linked intervening sequences and not to the presence of exon

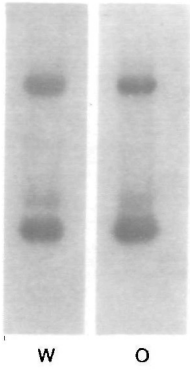


Figure 6. Relative cross-linking efficiency of introns versus exons. To a solution of purified Xho F and G fragments of adeno DNA, an equimolar amount of the fiber cDNA containing plasmid (pJAW 43) cleaved with Sal I and Xho I was added. This mixture was analyzed on 1.2% agarose gel and the part of the gel containing the Xho fragment and the smaller plasmid band (P) was blotted to Gene Screen. Band P contains pBR sequences and 169 nucleotides of cDNA (leaders 1,2 and part of leader 3). The figure shows hybridization patterns of the RNA probes to these blots.

sequences only.

Since equal amounts of leader sequences are present in band P as in the Xho F and G fragments together, and since hybridization of the hnRNA probes is predominantly to the genomic DNA fragments, the major part (>90%) of the RNA in linked and unlinked form must be derived from intervening sequences and i-leader.

## DISCUSSION

To isolate hnRNA sequences associated with RNP proteins we used ultraviolet light for cross-linking hnRNA to its associated proteins in vivo. Since the hnRNP C-proteins (Mr 42,000-44,000) are almost exclusively cross-linked to host as well as adenoviral hnRNA (16,17,33), these polypeptides can be localized on the RNA. This approach might provide more unambiguous information about RNP structure compared with a method used by other investigators, who studied the structural organization by analyzing sequences protected against RNase degradation by proteins in 30 S particles containing polyoma transcripts (37) or hnRNP complexes containing early adenovirus transcripts (31). The advantage of using cross-linking instead of RNase protection is the possibility to isolate directly the RNA fragments linked with specific polypeptides in vivo, whereas in protection experiments RNA is associated with several proteins. Moreover it is possible that RNA sequences specifically interacting with proteins are not detected in the RNase protection experiments, because the RNA might be located at the outside of the RNP complexes and is then accessible to degrading enzymes.

To facilitate the localization of the proteins on the RNA sequences we have used late adenovirus infected cells, in which the majority of the transcripts

(60%) is virus specific (48). In previous papers we have shown that the RNP organization of adenoviral transcripts is similar to those of host transcripts (17,33). Most viral RNAs isolated are derived from the major late promoter. Although the high amount of late viral transcripts is an important advantage in detecting RNA sequences specifically associated with RNP proteins, the system has also its drawbacks. The hnRNA fragments, which extend to the right of 30.0 m.u. on the genome,cannot be identified as introns or exons, because depending on the polyadenylation site and the splicing pattern of the transcripts the same sequence may be represented in both forms. We therefore decided to look in more detail in the region of the leader sequences, where more defined intron and exon sequences are available. Also in this region the population of RNA is heterogeneous. Possibly hnRNP structure changes upon processing of the RNA, so this type of study may not elucidate the RNP structure for a single transcript. The heterogeneity of the RNA can be observed in the hybridization patterns shown in figure 5A, from which it is clear that not all adeno DNA bands from the Hind III B region, even when they are of similar size, are recognized to the same extent. This would have been expected if only precursor RNA, containing the entire Hind III B sequence, was present in the isolated poly A(+)hnRNA. If only fully processed RNA would have been present, only exons would hybridize. That this is not the case is evident from the hybridization of "pure" intron regions, as for example represented in the fragments H and M in the Taq I/Eco RI digest, bands H,M and K on the Hin II/Eco RI digest and band E of the Sau 3A digest. From the total hybridization pattern it is evident that RNA complementary to the right part of the Hind III B fragment is extremely overrepresented as compared with the RNA from the left part. The explanation for this phenomenon probably is, that splicing of the intervening sequences in the left part of the RNA occurs before cleavage in more rightward located sequences. Splicing appears to occur in the 5' to 3' direction as was shown by Keohavong et al. (46) and we have found recently, using S1 nuclease mapping, that cleavage of the transcripts between leader 1 and 2 is an early event preceeding other cleavage steps (E.Mariman et al., manuscript in preparation). The major portion of the RNA used in this study has therefore probably been spliced or at least cleaved between the leaders 1 and 2, while the cleavage in the sequences between the leaders 2 and 3 and more often between leader 3 and the body of the mRNA is a slower process.

Since the length of the cross-linked and unlinked RNA probes (see figure 3) is very small the RNP proteins can theoretically be localized with high accuracy ( within less than 100 nucleotides). To attain a high resolution the

restriction fragments on the blots should also be very small. Electrophoretic blotting of DNA fragments between 1500 and 150 bp from acrylamide/agarose gels to Gene Screen was used for this purpose and proved to be efficient.

The main conclusions from our experiments are:

- hnRNP C-proteins can be cross-linked to nucleotides over the entire length of late adenoviral transcripts, not at one or a few discrete sites(fig.4).
- The distribution of these cross-linked hnRNP proteins along the viral pre-mRNAs is not random with respect to the RNA sequence (fig.5).
- Fine mapping of protein-linked RNA fragments within the Hind III B fragment of the adenoviral genome reveals regions that are overrepresented and regions strongly underrepresented as compared with unlinked RNA. Overrepresented regions are localized around the 3' end of the i-leader and around leader 3. The underrepresented regions are found in the part of the genome encoding VA I and VA II RNA and the late region 1 RNAs (fig.5B).
- Leader sequences as well as the intervening sequences between them can be cross-linked to proteins. Intron and exon sequences are equally represented in cross-linked RNA compared with unlinked RNA (fig.6).

Our results do not seem to support a simple concept of RNP structural organization, in which intron or exon sequences are specifically present in RNP structures. Protected hnRNA sequences complementary to a splice junction in an early adenoviral precursor RNA and protected RNA sequences in the middle of an intron were described by Ohlsson et al. (31). On the other hand, in protection experiments in which polyoma pre-mRNA sequences packed in 30 S particles were analyzed, mRNA sequences were preferentially protected (37).

The efficiency of cross-linking of the C-proteins to hnRNA and the fact that they are cross-linked to various parts of the viral transcripts, indicates that probably more than one of these polyeptides are associated with each molecule of RNA. This is corroborated by electron microscopy data, in which more than one RNP particle is usually found per transcript (32,49). The results from electron microscopy experiments studying nascent transcripts and their RNP structure agree with our main conclusion that RNP structure in well defined transcripts is not random and also sequence dependent. The morphology of RNP structures of late adenoviral nascent transcripts has recently been studied in more detail, and reveals non-randomly arranged RNP structures and a hairpin loop at the 5' end (leader region)(A.Beyer, personnal communication). Our results seem in good agreement with these data, although the presence of processing products and processing intermediates in our RNA does not allow detailed comparison. To study hnRNP structure it, therefore, appears advice-

ble to isolate homogeneous populations of transcripts. Nascent late adenoviral transcripts, which can be isolated as a part of transcription complexes in association with the viral DNA (51), might provide such a population.

ACKNOWLEDGEMENTS

These studies were initiated during a stay of CvE at the department of Microbiology of the University of Uppsala enabled by an E.M.B.O. short term fellowship. We are very gratefull to Drs. P.Aleström and U.Petterson for providing us with the nucleotide sequence of the Xho I/Hind III fragment of adenovirus 2 genome (15.5-31.5 m.u.) before publication. We thank Mr. A.Groeneveld for culturing the cells and we are indebted to Dr. M.Mathews from the Cold Spring Harbor Laboratory for providing us with cloned cDNA of fiber mRNA. This work was supported in part by the Netherlands Foundation for Chemical Research (S.O.N.) and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

*Present address: Department of Microbiology, University of Umeå, Sweden.

+Present address: European Molecular Biology Laboratory, Heidelberg, FRG

REFERENCES

1. Heinrich,P.C., Gross, V., Northeman. W. and Scheurlen, M. (1978) Rev. Physiol.Biochem.Pharmacol. 81 , 1o2-134.
2. van Venrooij, W.J. and Jansen, D.B. (1978) Molec.Biol.Rep. 4, 3-8.
3. Perry, R.P. (1981) J.Cell Biol. 91, 28s-38s.
4. Samarina, G.P., Lukanidin, E.M., Molnar, J. and Georgiev, G.P. (1968) J.Mol Biol. 33, 251-263.
5. Stevenin, J., Gattoni, R., Keohavong, P. and Jacob, M. (1982) J.Mol.Biol. 155, 185-205.
6. Stevenin, J., Gattoni, R., Gallinaro-Matringe, H. and Jacob, M. (1978) Eur.J.Biochem. 84, 541-549.
7. Brunel, C. and LeLay, M.N. (1979) Eur.J.Biochem. 99, 273-283.
8. Maundrell, K. and Scherrer, K. (1979) Eur.J.Biochem. 99, 225-238.
9. Karn, J., Vidali, G., Boffra, L.C. and Allfrey, V.G. (1977) J.Biol.Chem. 252, 7307-7322.
10. Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Cell 11, 127-138.
11. Walker, B.W., Lothstein, L., Baker, C,L. and LeStourgeon, W.M. (1980) Nucleic Acids Res. 8, 3639-3657.
12. 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
13. Miller, T.E., Huang, C-Y. and Pogo, A.O. (1978) J.Cell Biol. 76, 675-691.
14. Long, B.H., Huang, C-Y. and Pogo, A.O. (1979) Cell 18, 1079-1090.
15. Herman, R., Weymouth, L. and Penman, S. (1978) J.Cell Biol. 78, 663-674.
16. van Eekelen, C. and van Venrooij, W. (1981) J.Cell Biol. 88, 554-563.
17. Mariman, E., van Eekelen, C., Reinders, R., Berns, A. and van Venrooij,W. (1982) J.Mol.Biol. 154, 103-119.

18. Calvet, J.P. and Pederson, T. (1979) Nucleic Acids Res. 6, 1993-2001.
19. Calvet, J.P. and Pederson, T. (1978) J.Mol.Biol. 122, 361-378.
20. Kish, V.M. and Pederson, T. (1977) Proc.Natl.Acad.Sci.USA 74, 1426-1430.
21. Deimel, B., Louis, C. and Sekeris, C.A. (1977) FEBS lett. 73 , 80-84.
22. Seifert, H., Scheurlen, M., Northeman, W. and Heinrich, P.C. (1979) Biochim.Biophys.Acta 564, 55-66.
23. Gallinaro, H. and Jacob, M. (1981) Biochim.Biophys.Acta 652, 109-120.
24. Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L. and Steitz, J.A.(1980) Nature 283, 220-224.
25. Rogers, J. and Wall, R. (1980) Proc.Natl.Acad.Sci.USA 77, 877-879.
26. Yang, V.W., Lerner, M.R., Steitz, J.A. and Flint, S.J. (1981) Proc.Natl. Acad.Sci.USA 78, 1371-1375.
27. Calvet, J.P. and Pederson, T. (1981) Cell 26, 363-370.
28. Augenlicht, L. (1979) Biochemistry 18, 3780-3786.
29. Munroe, S.H. and Pederson, T. (1981) J.Mol.Biol. 147, 437-449.
30. Wahrman, M.Z. and Augenlicht, L.H. (1979) Biochem.Biophys.Res.Commun. 87, 395-402.
31. Ohlsson, R.I., van Eekelen, C. and Philipson, L. (1982) Nucleic Acids Res. 10, 3053-3068.
32. Pederson, T. and Davis, N.G. (1980) J.Cell Biol. 87, 47-54.
33. van Eekelen, C, Mariman, E., Reinders, R. and van Venrooij, W. (1981) Eur.J.Biochem. 119, 461-467.
34. van Eekelen, C., Buijtels, H., Linnê, T., Ohlsson, R., Philipson, L. and van Venrooij, W.J. (1982) Nucleic Acids Res. 10, 3039-3052.
35. Beyer, A.L., Bouton, A.H. and Miller, O.L. (1981) Cell 26, 155-165.
36. Beyer, A.L., Miller, O.L. and McKnight, S.L. (1980) Cell 20, 75-84.
37. Steitz, J.A. and Kamen, R. (1981) Mol.Cell.Biol. 1, 21-34
38. Wagenmakers, A., Reinders, R. and van Venrooij, W. (1980) Eur.J.Biochem. 112, 323-330.
39. Aviv, H. and Leder, P. (1972) Proc.Natl.Acad.Sci.USA 69, 1408-1412.
40. Maxam, A.M. and Gilbert, W. (1980) in: Methods in Enzymology, vol.65 pp 499-560, Acad.Press, New York.
41. Petterson, L. and Sambrook, J. (1973) J.Mol.Biol. 73, 125-130.
42. Aleström, P., Akusjärvi, G., Perricaudet, M., Matthews, M.B., Klessig, D. F. and Petterson, U. (1980) Cell 19, 671-681.
43. Southern, E.M. (1975) J.Mol.Biol. 98, 503.
44. van Eekelen, C., Riemen, T. and van Venrooij, W. (1981) FEBS lett. 130, 223-226.
45. Akusjärvi, G. and Persson, H. (1981) Nature 292, 420-426.
46. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stevenin, J. (1982) Nucleic Acids Res. 10, 1215-1229.
47. Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) Cell 16, 851-861.
48. Beltz, G.A. and Flint, S.J. (1979) J.Mol.Biol. 131, 353-373.
49. Beyer, A.L., Bouton, A.H., Hodge, L.D. and Miller, O.L. (1981) J.Mol.Biol. 147, 269-295.
50. Broker, T.R. and Chow, L.T. (1980) Trends in Biol.Sci. 5, 174-178.
51. Chen-Kiang, S., Wolgemuth, D.J., Hsu, M-T. and Darnell, J.E. (1982) Cell 28, 575-584.