The EMBO Journal vol.6 no.12 pp.3841-3848, 1987

cDNA cloning of the human U1 snRNA-associated A protein: extensive homology between U1 and U2 snRNP-specific proteins

Peter T.G.Sillekens, Winand J.Habets, Ria P.Beijer and Walther J.van Venrooij

Department of Biochemistry, University of Nijmegen, Geert Grooteplein N21, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Communicated by H.Bloemendal

Sera from patients with connective tissue diseases often contain antibodies against snRNA-associated proteins. Using one of these sera in an immunological screening of a human λgt11 expression vector cDNA library, two cDNA clones for the U1 snRNP-specific A protein, termed λHA-1 and λHA-2, were isolated. Monospecific antibodies, eluted from the β-galactosidase fusion protein of either clone reacted with the U1 snRNP-specific A antigen. The identity of the clones was confirmed by *in vitro* translation of hybrid selected mRNA. RNA blot analysis revealed a single polyadenylated transcript of about 1.4 kb in human cells. A cDNA of 1.2 kb, isolated from the same λgt11 expression library by cross-hybridization with a λHA-2 restriction fragment, covered the complete coding sequence of the A protein as demonstrated by *in vitro* translation of an RNA transcript synthesized from this cDNA. The deduced amino acid sequence contains one very hydrophilic region, an internal sequence duplication and a region highly homologous to the RNP consensus sequence that seems to be common to RNA binding proteins. Sequence comparison with the recently cloned U2 snRNP-specific B″ protein revealed two extremely homologous regions located in the carboxy-terminal (homology of 86%) and amino-terminal part (homology of 77%) of the proteins. This structural relationship indicates that proteins A and B″, although located in different snRNP particles, may have identical functions.

Key words: autoimmunity/cDNA cloning/RNP antigens/small nuclear ribonucleoproteins

Introduction

All eukaryotic cells contain a group of small ribonucleic acids, among which the most abundant are the U series small nuclear RNAs (U snRNAs). They reside in the nucleus of the cell as RNA-protein complexes (RNPs: for reviews see Reddy and Busch, 1983; Brunel *et al.*, 1985). Since the discovery that Sm and RNP serological specificities seen in systemic lupus erythematosus (SLE) and other disorders (Tan, 1982) recognize peptides of these U snRNPs (Lerner and Steitz, 1979), a number of different classes of these particles have been distinguished (Hinterberger *et al.*, 1983; Kinlaw *et al.*, 1983) on the basis of the U snRNA species they contain. The exact number of distinct snRNPs in mammalian cells is still unknown, but until now the very abundant U1 and U2 snRNPs, the slightly less abundant snRNPs containing U4/U6 snRNAs and U5 snRNA, and several low abundance snRNPs containing other U snRNAs have been described (Strub *et al.*, 1984; Reddy *et al.*, 1985). The U2, U4/U6 and U5 snRNA-containing particles are recognized by anti-Sm autoimmune sera and contain at least seven common peptides, called B′, B, D, D′, E, F and G (Hinterberger *et al.*, 1983; Kinlaw *et al.*, 1983; Billings and Hoch, 1984; Pettersson *et al.*, 1984; Habets *et al.*, 1985a; Bringmann and Lührmann, 1986). The U1 snRNA-containing particles are composed of the same seven proteins and therefore are also immunoprecipitated by anti-Sm sera. In addition, however, they are recognized by anti-RNP sera, due to the presence of at least three additional proteins, called 70K, A and C, which are unique to the U1 snRNP species (Billings and Hoch, 1984; Pettersson *et al.*, 1984; Habets *et al.*, 1985a). U2 snRNP particles also contain at least two unique proteins called A′ and B″, which are recognized by antibodies present in anti-(U1,U2)RNP sera. Anti-B″ antibodies in these sera were shown to crossreact with the U1 snRNP-specific A protein, suggesting a structural relationship between these proteins (Habets *et al.*, 1985b).

Although the exact function(s) of the various snRNP particles in the nucleus is still largely unknown, there is mounting evidence that most of them are essential cofactors for mRNA processing reactions (for a review see Maniatis and Reed, 1987). A critical role for snRNPs in the pre-mRNA splicing process was first proposed for the U1 snRNP particle (Lerner and Steitz, 1979; Rogers and Wall, 1980) and was based on complementarity in nucleotide sequence between the 5′ end of the U1 snRNA and the 5′ splice junctions of pre-mRNA. Experimental evidence for this idea was found in the inhibition of RNA splicing either by antibodies directed against U1 snRNPs (Yang *et al.*, 1981; Padgett *et al.*, 1983; Krämer *et al.*, 1984) or by removal of the 5′ end of the U1 snRNA (Krämer *et al.*, 1984; Krainer and Maniatis, 1985; Black *et al.*, 1985). The latter observation suggests that the RNA moiety is essential for an interaction between the U1 snRNP-particle and the 5′ splice site. The ability of a mutant U1 snRNA gene to suppress a 5′ splice site mutation (Zhuang and Weiner, 1986) also demonstrates the involvement of the RNA component of U1 snRNP in the association of the particle with the 5′ splice site. The U1 snRNP peptide components also appear to be required for binding (Mount *et al.*, 1983) but to what extent they contribute to this or any of the following events that take place at the 5′ splice junction during the splicing reaction is presently unknown. A first step to a better understanding of the biological role of these proteins, which based on immunological criteria show strong evolutionary conservation, is the elucidation of their primary structure.

Recently the molecular cloning of several autoimmune antigens has been reported (Chambers and Keene, 1985; Wieben *et al.*, 1985; Theissen *et al.*, 1986; Habets *et al.*, 1987). Here we describe the isolation, characterization and sequencing of four cDNA clones containing coding capacity for the human U1 snRNP-associated A protein. The longest of these cDNA clones contains the complete coding sequence for the A protein.

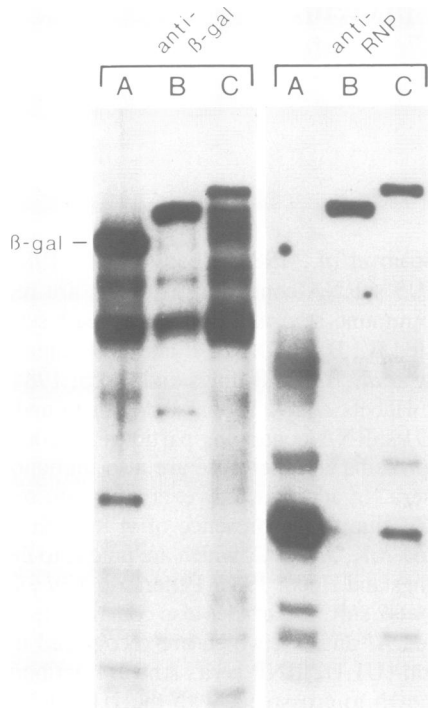

Fig. 1. Immunoblot analysis of the fusion proteins produced by the lysogens of λHA-1 and λHA-2. Purified fusion protein was resolved on a 7.5% SDS–polyacrylamide gel along with total lysate from λgt11-infected cells and transferred to nitrocellulose. Identical immunoblots were probed with either the patient anti-(U1)RNP serum with which the clones were isolated from the λgt11 library or a monoclonal antibody specific for β-galactosidase. Bound antibodies were detected with ^{125}I-labelled sheep anti-human Ig or sheep anti-mouse Ig respectively. **Lanes:** (**A**) total lysate from λgt11-infected cells; (**B**) purified fusion protein from clone λHA-1; (**C**) purified fusion protein from clone λHA-2.

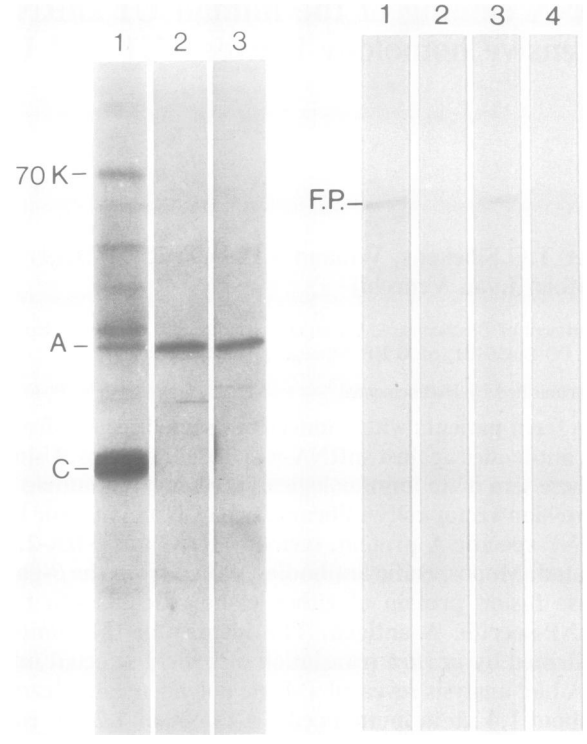

Fig. 2. Immunoblotting with antibodies affinity-purified from fusion proteins (**A**) or from U1 snRNP-specific proteins (**B**) . (**A**) Purified fusion proteins from the induced phages λHA-1 and λHA-2 were resolved on a 7.5% SDS–polyacrylamide gel and transferred to nitrocellulose. A strip of the immunoblots was processed for the detection of antigens and the area of each blot containing the fusion protein was excised. These regions were used to adsorb antibodies specific for the fusion proteins from a human anti-(U1)RNP serum containing antibodies against the 70K, A and C proteins. The antibodies were eluted (Smith and Fisher, 1984) and re-used to probe an immunoblot of HeLa total protein extract. **Lanes:** (**1**) total human anti-(U1)RNP serum; (**2**) affinity-purified antibody from the fusion protein of λHA-1; (**3**) affinity-purified antibody from the fusion protein of λHA-2. (**B**) HeLa total protein blots were incubated with the human anti-(U1)RNP serum of **lane 1** in (**A**). Reactive regions of the proteins 70K, A and C were excised and bound antibodies were eluted. These fractions were re-used to probe an immunoblot containing purified fusion protein of λHA-1. **Lanes:** (**1**) total human anti-(U1)RNP serum; (**2**) affinity-purified antibody from the 70K protein; (**3**) affinity-purified antibody from the A protein; (**4**) affinity-purified antibody from the C protein.

Results

Isolation and identification of cDNA clones for the U1 snRNP-specific A protein

An anti-(U1)RNP serum was used to screen a λgt11 expression library containing cDNAs prepared from human teratocarcinoma cells for clones encoding the U1 snRNP-specific A antigen. Initial screening of about 600 000 recombinants identified eight putative A protein clones. Upon several rounds of purification and rescreening two clones which possibly contain coding sequences for the A polypeptide could be identified. Both clones, referred to as λHA-1 and λHA-2, reacted with a panel of nine sera from different autoimmune patients, all containing antibodies against the A antigen, whereas the reaction with normal human control sera was negative (not shown).

To demonstrate that the reactivity of the phage clones with the antibodies was due to antigenic determinants encoded by the cDNA inserts, λHA-1 and λHA-2 were transferred to a lysogenic host. Lysogens were induced by temperature shift and addition of isopropyl-β-D-thiogalactopyranoside (IPTG) allowing maximal fusion protein synthesis. Immunoblots of the lysogen proteins were probed separately with a monoclonal anti-β-galactosidase antibody and the anti-(U1)RNP serum used for the initial selection. Lysogens containing phage λHA-1 produced a fusion protein with an apparent mol. wt on an SDS–polyacrylamide gel of about 126 kd, which was reactive both with antibodies to β-galactosidase and with antibodies from the patient serum (Figure 1, lane B). The fusion protein of phage λHA-2, giving the same immunological response as the λHA-1 fusion protein, had a mol. wt of about 138 kd (Figure 1, lane C).

To identify the antigenic specificity from the anti-(U1)RNP serum which is responsible for a positive immunoreaction, antibodies specifically reacting with the fusion proteins were affinity-purified from total serum and used to probe an immunoblot of HeLa total protein extract. As shown in Figure 2A, antibodies affinity-purified from both fusion proteins reacted only with the U1 snRNP-specific A protein. When the reciprocal experiment was carried out, a similar result was obtained: antibodies eluted from the region of an immunoblot containing the *in vivo* synthesized A polypeptide stained the fusion protein band of λHA-1 and λHA-2, whereas no reaction was observed with the other affinity-purified anti-(U1)RNP antibodies (anti-70K, anti-C). The results for the λHA-1 clone are shown in Figure 2B.

Hybrid selection and in vitro translation

The identity of cDNA clones λHA-1 and λHA-2 was confirmed by hybrid selection of HeLa poly(A)$^+$ RNA and subsequent *in*

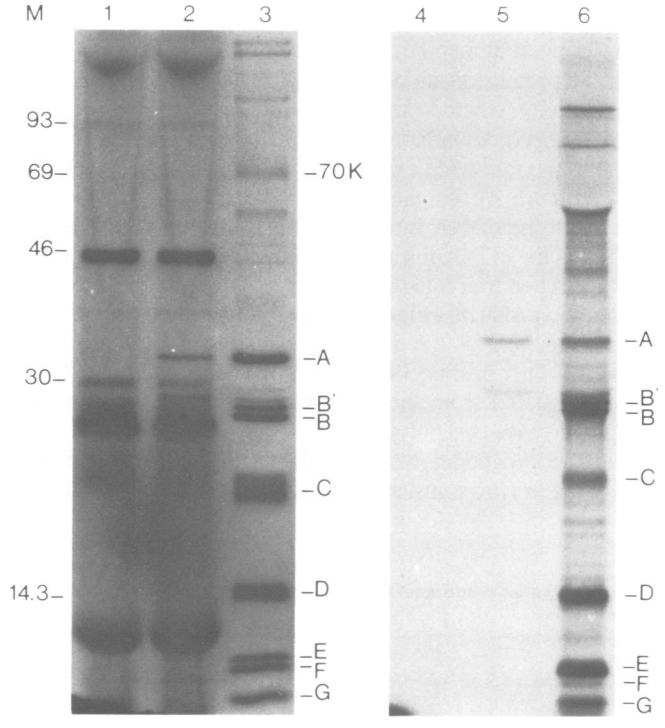

Fig. 3. Hybrid-selection translation of HeLa mRNA with pHA-2. HeLa poly(A)$^+$ RNA was hybrid selected and translated in a reticulocyte lysate as described in Materials and methods. Immunoprecipitation of the translation products of mRNA selected on the pHA-2 cDNA fragment was carried out with a human anti-RNP/Sm antiserum recognizing on an immunoblot of HeLa total protein extract the snRNA-associated proteins 70K, A, B/B', C and D. The same serum was used for an immunoprecipitation of U snRNPs from *in vivo* [^{35}S]methionine labelled HeLa total protein extract. ^{35}S-labelled translation products were analysed on 15% SDS–polyacrylamide gels and visualized by fluorography. **Lanes:** (**1**) background translation without addition of exogenous RNA; (**2**) translation products of mRNA hybrid selected on pHA-2; (**3**) immunoprecipitation of U snRNP proteins from *in vivo* ^{35}S-labelled HeLa protein extract; (**4**) immunoprecipitation of translation products from the background translation; (**5**) immunoprecipitation of translation products from pHA-2 hybrid selected mRNA. (**6**) immunoprecipitation of translation products from total HeLa poly(A)$^+$ RNA. Mol. wt (in kd) of protein standards are indicated to the left (**M**).

vitro translation. The cDNA inserts of λHA-1 and λHA-2 were recloned into the vector pSP65, resulting in the plasmid clones pHA-1 and pHA-2. cDNA inserts were isolated by digestion with *Eco*RI and subsequently spotted onto nitrocellulose. After hybridization with HeLa poly(A)$^+$ RNA, filter-bound mRNA was eluted and translated *in vitro* using a rabbit reticulocyte lysate. In Figure 3 the results for pHA-2 are shown. A 32 000 dalton protein was synthesized by the hybrid selected mRNA (Figure 3, lane 2), which on an SDS–polyacrylamide gel comigrates with the *in vivo* labelled U1 specific A protein immunoprecipitated by a human serum containing both Sm and RNP serological specificities (Figure 3, lane 3). When the same serum was used in an immunoprecipitation of an *in vitro* translation directed by total HeLa poly(A)$^+$ RNA, every U1 snRNP protein could be shown to be synthesized in an immunoprecipitable form *in vitro*, except for the 70K protein, which *in vivo* as well as *in vitro* seems to be hardly labelled by [^{35}S]methionine. The 32 000 dalton translation product of the pHA-2 selected mRNA was also found to be immunoreactive with this serum. The immunoprecipitated protein (Figure 3, lane 5) comigrates with the *in vitro* labelled A polypeptide (Figure 3, lane 6). In addition, two lower mol. wt polypeptides were also specifically immunoprecipitated (Figure 3, lane 5). Since Northern blots revealed only one mRNA species (see below), these extra protein bands are more likely to be degradation products from the A protein, rather than translation products from additional mRNAs hybrid selected by the cDNA insert of the A clones.

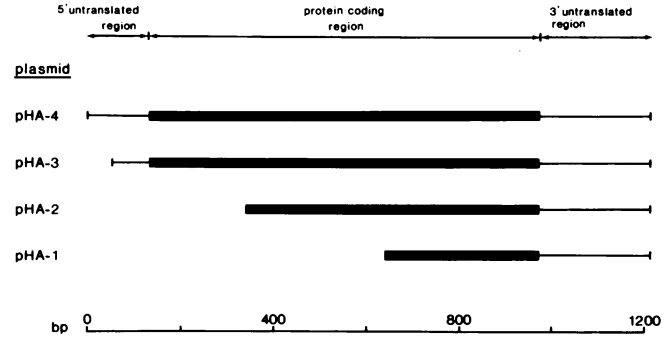

Fig. 4. Schematic representation of the cDNA clones pHA-1, pHA-2, pHA-3 and pHA-4. The relative positions of the cDNA inserts of the A protein clones are shown. Regions coding for the A protein are indicated by solid bars. The 5' and 3' noncoding regions are indicated by thin lines.

In conclusion, these results together with the immunological data presented demonstrate that the λHA-1 and λHA-2 clones contain coding capacity for the U1 snRNP-specific A protein.

Isolation of a full-length cDNA clone of the A protein

Determination of the insert size revealed a 0.6 kb cDNA insert for the λHA-1 clone, whereas the λHA-2 clone contained an insert of 0.9 kb. Partial restriction analysis and DNA sequencing of both cDNA inserts established that we had isolated two overlapping clones, showing in addition that the sequence of the insert of λHA-1 is fully contained in the λHA-2 cDNA insert. The relative position of the two cDNAs is shown in Figure 4.

The size of the homologous mRNA was determined by RNA blot analysis. Hybridization of a blot containing poly(A)$^+$ RNA from HeLa cells to the nick-translated 0.9 kb cDNA insert of λHA-2 revealed a single band at about 1.4 kb (not shown). Taken together, these data show that the cDNA insert of the λHA-2 clone comprises only about 60% of the total message for the A polypeptide. Therefore, attempts were made to isolate a full-length cDNA clone. The λgt11 expression library, originally used for screening with an anti-(U1)RNP serum, was now rescreened with a nick-translated *Eco*RI/*Sph*I fragment covering the 5'-terminal 216 nucleotides of the λHA-2 insert (nucleotides 342–557 of the cDNA, as presented in Figure 5). Out of 300 000 plaques, 30 hybridized to this probe. Size analysis of the inserts revealed two phages with inserts longer than the 0.9 kb insert of the λHA-2 clone, the longest insert having a size of about 1200 bp. Assuming an average length of about 150 residues for the poly(A) tail of a eukaryotic mRNA, this cDNA insert matches well with the size of the homologous mRNA. The other insert was about 1150 bp in size. The cDNA inserts of both phage clones isolated by cross-hybridization were subcloned in the vector pSP65. The relative position of these cDNAs (pHA-3, pHA-4), established by Southern blotting and restriction enzyme maps, is shown in Figure 4, together with the two cDNAs isolated via immunological screening with the anti-(U1)RNP serum.

Nucleotide sequence and deduced amino acid sequence of the A protein

The nucleotide sequence of the cDNA inserts of all four clones was determined by the dideoxy method (Sanger *et al.*, 1977). The combined nucleotide sequence is shown in Figure 5 and

covers 1209 nucleotides. It should be mentioned that no differences between corresponding parts of the four A protein clones were found. There is only one large uninterrupted open reading frame (covering nucleotides 81−971) encoding 297 amino acids (Figure 5). The designated ATG translation initiation codon is located at position 126 and is the first ATG codon occurring. It exists within a sequence which only diverges in position −3 (3 nucleotides upstream of the ATG codon) from the optimal context for initiation, ACCATGG (Kozak, 1986). The TAG translational stop signal at position 972 is followed by a non-coding region of 223 nucleotides. At the 3′ end a stretch of eight adenines is found, which is preceded by the hexamer ATTAAA. This sequence is known as a functional polyadenylation signal (Leff *et al.*, 1986), though it is slightly altered from the conserved AATAAA hexamer normally found in eukaryotic mRNAs (Proudfoot and Brownlee, 1976).

The predicted amino acid sequence for the A protein (Figure 5) indicates that the mol. wt of the polypeptide is 31 243 daltons. This is in good agreement with the observed mol. wt of 32 kd for the A protein on an SDS−polyacrylamide gel. To verify experimentally whether the nucleotide sequence presented in Figure 5 indeed contains the complete coding sequence for the A protein, the following experiment was conducted. The plasmid pHA-4 (Figure 4), containing the SP6 promoter immediately upstream of the cDNA insert, was linearized by digestion with *Hin*dIII and the cDNA insert was transcribed *in vitro*. The resulting RNA, when translated in a rabbit reticulocyte lysate, yielded three major products (Figure 6, lane 2), the longest of which had the same size as the *in vivo* synthesized A protein (Figure 6, lane 3). This primary translation product could also be immunoprecipitated from the translation mixture (Figure 6, lane 5), together with two polypeptides of lower mol. wt. The same additional polypeptides were also found in the immunoprecipitation of the *in vitro* translation reaction programmed with hybrid

```
                                                                    20
                                                  ga att cct gac ttc ctt ttc gga gga aga tcc ttg

     40                        60                        80                       100                       120
agc agc cga cgt tgg gac aaa gga ttt gga gaa acc cag ggc taa agt cac gtt ttt cct cct tta aga ctt acc tca aca ctt cac tcc

                  140                       160                       180                       200
    atg gca gtt ccc gag acc cgc cct aac cac act att tat atc aac aac ctc aat gag aag atc aag aag gat gag cta aaa aag tcc ctg
1   Met Ala Val Pro Glu Thr Arg Pro Asn His Thr Ile Tyr Ile Asn Asn Leu Asn Glu Lys Ile Lys Lys Asp Glu Leu Lys Lys Ser Leu   30

     220                       240                       260                       280                       300
    tac gcc atc ttc tcc cag ttt ggc cag atc ctg gat atc ctg gta tca cgg agc ctg aag atg agg ggc cag gcc ttt gtc atc ttc aag
31  Tyr Ala Ile Phe Ser Gln Phe Gly Gln Ile Leu Asp Ile Leu Val Ser Arg Ser Leu Lys Met Arg Gly Gln Ala Phe Val Ile Phe Lys   60

                  320                       340                       360                       380
    gag gtc agc agc gcc acc aac gcc ctg cgc tcc atg cag ggt ttc cct ttc tat gac aaa cct atg cgt atc cag tat gcc aag acc gac
61  Glu Val Ser Ser Ala Thr Asn Ala Leu Arg Ser Met Gln Gly Phe Pro Phe Tyr Asp Lys Pro Met Arg Ile Gln Tyr Ala Lys Thr Asp   90

     400                       420                       440                       460                       480
    tca gat atc att gcc aag atg aaa ggc acc ttc gtg gag cgg gac cgc aag cgg gag aag agg aag ccc aag agc cag gag acc ccg gcc
91  Ser Asp Ile Ile Ala Lys Met Lys Gly Thr Phe Val Glu Arg Asp Arg Lys Arg Glu Lys Arg Lys Pro Lys Ser Gln Glu Thr Pro Ala  120

                  500                       520                       540                       560
    acc aag aag gct gtg caa ggc ggg gga gcc acc ccc gtg gtg ggg gct gtc cag ggg cct gtc ccg ggc atg ccg ccg atg act cag gcg
121 Thr Lys Lys Ala Val Gln Gly Gly Gly Ala Thr Pro Val Val Gly Ala Val Gln Gly Pro Val Pro Gly Met Pro Pro Met Thr Gln Ala  150

     580                       600                       620                       640                       660
    ccc cgc att atg cac cac atg ccg ggc cag ccg ccc tac atg ccg ccc cct ggt atg atc ccc ccg cca ggc ctt gca cct ggc cag atc
151 Pro Arg Ile Met His His Met Pro Gly Gln Pro Pro Tyr Met Pro Pro Pro Gly Met Ile Pro Pro Pro Gly Leu Ala Pro Gly Gln Ile  180

                  680                       700                       720                       740
    cca cca ggg gcc atg ccc ccg cag cag ctt atg cca gga cag atg ccc cct gcc cag cct ctt tct gag aat cca ccg aat cac atc ttg
181 Pro Pro Gly Ala Met Pro Pro Gln Gln Leu Met Pro Gly Gln Met Pro Pro Ala Gln Pro Leu Ser Glu Asn Pro Pro Asn His Ile Leu  210

     760                       780                       800                       820                       840
    ttc ctc acc aac ctg cca gag gag acc aac gag ctc atg ctg tcc atg ctt ttc aat cag ttc cct ggc ttc aag gag gtc cgt ctg gta
211 Phe Leu Thr Asn Leu Pro Glu Glu Thr Asn Glu Leu Met Leu Ser Met Leu Phe Asn Gln Phe Pro Gly Phe Lys Glu Val Arg Leu Val  240

                  860                       880                       900                       920
    ccc ggg cgg cat gac atc gcc ttc gtg gag ttt gac aat gag gta cag gca ggg gca gct cgc gat gcc ctg cag ggc ttt aag atc acg
241 Pro Gly Arg His Asp Ile Ala Phe Val Glu Phe Asp Asn Glu Val Gln Ala Gly Ala Ala Arg Asp Ala Leu Gln Gly Phe Lys Ile Thr  270

     940                       960                       980                      1000                      1020
    cag aac aac gcc atg aag atc tcc ttt gcc aag aag tag cac ctt ttc ccc cca tgc ctg ccc ctt ccc ctg ttc tgg ggc cac ccc ttt
271 Gln Asn Asn Ala Met Lys Ile Ser Phe Ala Lys Lys

                 1040                      1060                      1080                      1100
    ccc cct tgg ctc agc ccc ctg aag gta agt ccc ccc ttg ggg gcc ttc ttg gag ccg tgt gtg agt gag tgg tcg cca cac agc att gta

      1120                      1140                      1160                      1180                      1200
    ccc aga gtc tgt ccc cag aca ttg cac ctg gcg ctg tta ggc cgg aat taa agt ggc ttt ttg agg ttt ggt ttt tca caa aaa aaa gga

    att c
```

Fig. 5. Combined nucleotide sequence of pHA-1, pHA-2, pHA-3 and pHA-4, and deduced amino acid sequence of the encoded A protein. The predicted amino acids are shown below their triplets. The polyadenylation signal and the RNP consensus sequence are underlined.

selected A protein mRNA (Figure 3, lane 5) and may either be degradation products of the A protein or alternatively represent incomplete translation products originating from the same reading frame.

Discussion

Using autoantibodies from a patient with mixed connective tissue disease (MCTD), we have isolated two human cDNA clones for the U1 snRNA-associated A protein from a λgt11 expression library. Screening of the same library with restriction fragments of these clones revealed two additional recombinants with longer cDNA inserts. The identity of the cDNAs was established by hybrid selection and subsequent translation of the selected mRNA (Figure 3) and by *in vitro* translation of an RNA transcript synthesized from the longest cDNA insert (Figure 6). Several immunological experiments also confirm the identity of the isolated cDNAs: the β-galactosidase fusion protein of the λgt11 recombinants was recognized by anti-A antibodies from several patients' sera (not shown) and antibodies affinity-purified from the fusion proteins exclusively recognized the A antigen on an immunoblot of HeLa total protein extract (Figure 2A).

The longest cDNA identified covered the complete coding sequence and 3′ noncoding region of the A protein mRNA. In addition, it contained 125 nucleotides upstream of the putative initiation codon. In most eukaryotic mRNAs no extraneous ATG triplets are found upstream of the functional initiation codon for protein synthesis. Indeed, the first ATG codon occurring in the long open reading frame of this clone is at the same time the 5′-proximal ATG of the entire nucleotide sequence. Furthermore, this designated initiation codon exists in a fairly strong consensus sequence (Kozak, 1986). Only the highly conserved purine in position −3 is replaced by a uridine residue, possibly resulting in a somewhat weaker initiation signal. The primary translation product consists of 282 amino acids, resulting in a protein with a predicted mol. wt of 31 243 daltons. This is in good agreement with the apparent mol. wt of 32 kd for the A protein, as estimated from SDS−polyacrylamide gels.

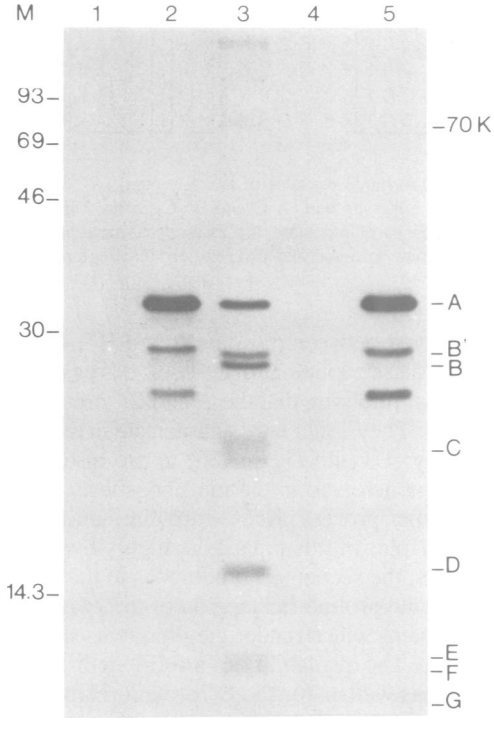

Fig. 6. *In vitro* translation of pHA-4 encoded RNA. Plasmid pHA-4 was linearized by digestion with *Hin*dIII and transcribed *in vitro* with SP6 RNA polymerase. The RNA was gel purified and translated in a reticulocyte lysate. Immunoprecipitation of the translation products was performed with the same anti-RNP/Sm antiserum that was used in the hybrid selection experiment (Figure 3). **Lanes:** (1) background translation without addition of exogenous RNA; (2) translation products of pHA-4 encoded RNA. (3) immunoprecipitation of U snRNP proteins from *in vivo* ^{35}S-labelled HeLa protein extract. (4) immunoprecipitation of translation products from the background translation; (5) immunoprecipitation of translation products encoded by RNA from pHA-4. Mol. wt (in kd) of protein standards are indicated to the left (**M**).

The amino acid sequence of the A protein deduced from the nucleotide sequence of the cDNA inserts (Figure 5) has an interesting distribution of charged and aromatic amino acids. They are organized in two distinct domains, separated by a segment of very high proline content (amino acids 140−206 in Figure 5). Most of the charged and aromatic amino acids are located in the amino-terminal part of the protein. In this domain a cluster of 11 charged amino acids represents the most hydrophilic region of the entire protein (amino acids 103−112 in Figure 5). Due to the absence of negatively charged amino acids, the middle part of the protein, containing the proline-rich region, also has an overall positive charge. Though not common to all RNA binding proteins, similar regions of very high proline content and an overall positive charge have been identified in several single-stranded nucleic acid-binding proteins (Garoff *et al.*, 1980; Kruijer *et al.*, 1981; Adam *et al.*, 1986; Theissen *et al.*, 1986).

The U1 snRNP-specific A polypeptide has at least one epitope in common with the U2 snRNP specific B″ protein, as was demonstrated by a crossreactivity of monospecific anti-B″ antibodies with the A antigen (Habets *et al.*, 1985b). Monoclonal antibodies that recognize an epitope shared by these two proteins could also be produced (Reuter and Lührmann, 1986; Reuter *et al.*, 1986). Protease digestion of proteins A and B″ suggests that the two proteins share a similar structural domain which is even larger than just an epitope (Reuter *et al.*, 1987). Therefore, the amino acid sequence of the A protein was compared with that of the earlier reported B″ protein (Habets *et al.*, 1987) for sequence homology. Two regions of extensive homology were found. In Figure 7 the A protein sequence is shown in the upper line and the aligned sequence of the B″ protein in the bottom line. Homology is highest in the carboxy-part: out of the 78 carboxy-terminal residues of both proteins, 67 are identical, resulting in a homology of 86%. The second region of homology is found near the amino-terminal end. Here, in the region that spans amino acids 7−101 of the A protein, only 22 residues deviate from amino acids 4−98 in the B″ protein. Though somewhat less extensive than in the carboxy-part, the degree of homology in this region is still 77%. Recently a segment, LYS/ARG-GLY-PHE-GLY/ALA-PHE-VAL-X-PHE/TYR, has been described as a common context found in the amino acid sequence of several RNA binding proteins (Swanson *et al.*, 1987), including the U1 snRNA-associated 70K protein (Theissen *et al.*, 1986). The amino-terminal domain of the A protein also contains a fragment which is highly homologous to this RNP consensus sequence. The segment ARG-GLY-GLN-ALA-PHE-VAL-ILE-PHE (amino acids 52−59 in Figure 5) diverges only in the third amino acid from the RNP consensus sequence. Considerable sequence homology with other RNA binding proteins extends also beyond the RNP consensus sequence, over a domain of 80−100 amino acids (Sachs *et al.*, 1986). Since the RNP consensus sequence is located in a region of the A protein which is homologous with the B″ protein, a similar sequence was ex-

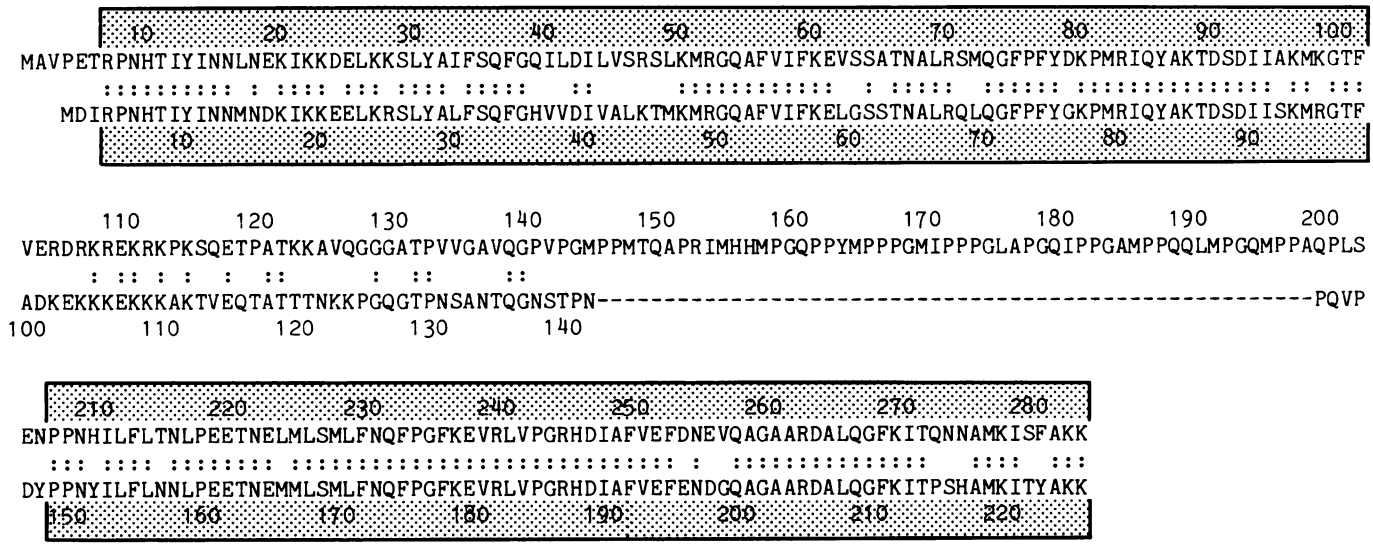

Fig. 7. Comparison of deduced amino acid sequences of proteins A and B″. The predicted amino acid sequence of the A protein is displayed in the upper line, and that of the aligned B″ protein (Habets *et al.*, 1987) in the bottom line. Regions of homology are indicated by the dotted areas. Common residues are marked by colons.

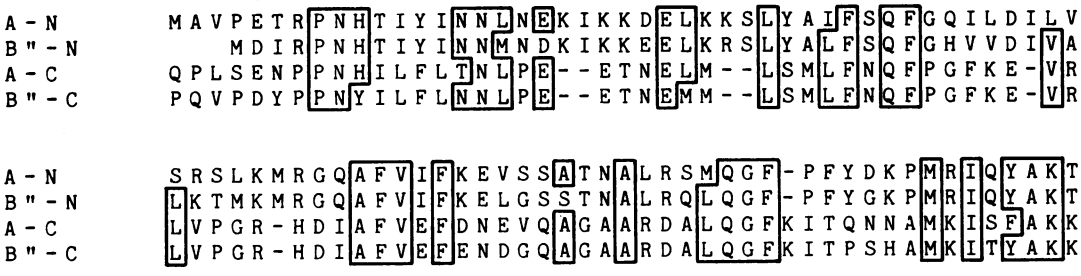

Fig. 8. Comparison of amino-terminal and carboxy-terminal ends of proteins A and B″. The amino-terminal residues of the A protein (A-N) and B″ protein (B″-N) (Habets *et al.*, 1987) are displayed in the upper lines, and the carboxy-terminal residues of either protein (A-C and B″-C) in the bottom lines. Alignment was achieved by the program of Wilbur and Lipman (1983) and as few as possible deletions or insertions are assumed. Amino acids shared by both regions of the proteins are enclosed in boxes. For the A protein the homology between the amino-terminal and carboxy-terminal region is 31%. For the B″ protein this homology is 35%.

pected to be present in this protein. Indeed, this putative RNA binding domain is completely conserved in the B″ protein (amino acids 49–56 in Figure 7). Even the amino acid in position 3, which in the A protein deviates from the RNP consensus sequence of Swanson *et al.* (1987), and the undefined amino acid at position 7 are identical in both proteins.

Another salient feature of the derived amino acid sequence of the A protein is a significant sequence homology between two parts of the protein itself. Alignment of the amino-terminal and carboxy-terminal region of the protein shows a homology of 31%. Because of the extensive homology between both proteins the corresponding regions of the B″ protein were also compared, resulting in a homology of 35%. Alignment of the amino-terminal regions from the A and B″ proteins with the carboxy-parts of both proteins is shown in Figure 8. From the RNP consensus sequence found in the amino-terminal part of both proteins, only four amino acids are conserved in the carboxy-terminal parts.

From this extensive structural relationship between the polypeptides A and B″ it is quite obvious that the genes encoding these proteins have emerged from a common ancestral gene. The internal sequence homology suggests that prior to this gene duplication event a sequence duplication within the precursor gene itself probably took place. During the evolution of the separate genes the sequence of these duplicated gene elements has been highly conserved between proteins A and B″. It is tempting to speculate that these evolutionarily conserved regions reflect functional domains, implying that the A and B″ proteins fulfill identical functions. They could either participate in the biological role of respectively, U1 and U2 snRNP in pre-mRNA splicing, or alternatively be involved in the morphogenesis of these snRNP species. In either process, RNA–protein interactions probably play a crucial role. In this regard the highly hydrophilic stretch of amino acids, the sequence homologous to the RNP consensus sequence and the proline-rich region of the A protein represent structural domains which render these proteins suitable to associate with RNA. The availability of a full-length cDNA clone for the A protein as well as for the B″ protein (Habets *et al.*, 1987) opens the way for a more detailed study of the interaction between these proteins and their respective U snRNA or defined RNA sequences.

Another interesting aspect of the A protein is the presence of at least two different epitopes that are recognized by anti-RNP autoantibodies. One of these reacts with antibodies from anti-(U1)RNP sera, whereas the second is recognized by the cross-reacting anti-B″ antibodies from anti-(U1,U2)RNP sera (Habets *et al.*, 1985b). Since both types of antisera react with the fusion protein of λHA-1 as well as λHA-2, these epitopes must be located in the overlapping carboxy-terminal part of the poly-

peptides encoded by these cDNAs (amino acids 172–282 in Figure 5). Evidence for an additional epitope emerged from an experiment with a patient serum containing anti-Sm antibodies. This antiserum only recognized the fusion protein of λHA-2, showing that an epitope recognized by antibodies in this anti-Sm antiserum is located in the region of the A protein additionally encoded by λHA-2 as compared with λHA-1 (amino acids 73–171 in Figure 5). It will be of interest to investigate the antigenic structure of the A protein in more detail, as antibodies against particular epitopes on the A protein may be related to certain clinical symptoms.

Materials and methods

Screening of a λgt11 expression library with antibody probes

A patient anti-(U1)RNP serum (diluted 1:500) was used to screen a human teratocarcinoma cDNA library constructed with the λgt11 vector (a gift from J.Skowronski, NCI/NIH, Bethesda) as previously described (Habets *et al.*, 1987) using the method of Young and Davis (1983). To detect specifically bound antibody, ^{125}I-labelled sheep anti-human Ig [F(ab)$_2$ fragment] was used.

Purification of fusion protein

Lysogens were prepared in *Escherichia coli* strain Y1089 (Huynh *et al.*, 1985). A 50 ml culture of lysogenic cells was grown to OD_{600} = 0.5 at 32°C and lytic replication was induced by a temperature shift (42°C, 15 min). Isopropyl-β-D-thiogalactopyranoside (Sigma) was added to a final concentration of 10 mM and the culture was incubated for an additional hour at 38°C prior to harvesting the cells by centrifugation. After lysis of the cells by repeated freeze–thawing, fusion protein was prepared essentially as described by Adam *et al.* (1986) using Zwittergent 3-14 (Calbiochem-Bering).

Cell labelling and preparation of HeLa cell total protein fraction

HeLa S3 cells were grown in suspension at 37°C at densities ranging from 0.2 × 10^6 to 0.5 × 10^6 cells/ml on Suspension Minimal Essential Medium (SMEM) supplemented with 10% newborn calf serum. Cellular protein was labelled by incubating the cells for 16 h with 10 μCi/ml [^{35}S]methionine (Amersham, UK; 1000 Ci/mmol). For the first 2 h, the cells were incubated in medium without unlabelled methionine. Then, 0.1 vol of complete medium was added (van Eekelen and van Venrooij, 1981).

For immunoblotting and immunoprecipitation assays total protein fractions from HeLa cells were used (Habets *et al.*, 1983). Cells were pelleted by centrifugation (10 min at 800 *g*), washed once with TBS [130 mM NaCl, 40 mM Tris–HAc (pH 7.4)] and spun down again. Subsequently, the cell pellet was resuspended in TBS (1 × 10^7 cells/ml) and the cells were sonicated three times (40 s) on ice with a microtip sonifier (model B-30; Branson Instr. Inc., Connecticut, USA). The suspension of sonicated cells was centrifuged for 30 min at 16 000 *g*. The supernatant was either used immediately for immunoprecipitation (labelled extract) or diluted with 1 vol of 2 × SDS gel sample buffer [125 mM Tris–HAc (pH 6.8), 4% SDS, 20% glycerol, 0.01% bromophenol blue, 10% 2-mercaptoethanol] and boiled for 10 min.

Gel electrophoresis, protein blotting and detection of antigens

Proteins were analysed on SDS–polyacrylamide gels using the Laemmli (1970) buffer system. Transfer of proteins onto nitrocellulose sheets was performed as described by Habets *et al.* (1983). For the detection of antigens the protein blots were treated and processed as described (Habets *et al.*, 1985a). As primary antibodies human autoimmune sera (diluted 1:100) were used. The antibody–antigen complexes were detected with ^{125}I-labelled anti-human Ig from sheep [F(ab)$_2$ fragment].

Purification of antibodies from nitrocellulose blots was performed as described by Smith and Fisher (1984).

Preparation of RNA

HeLa cells were harvested by centrifugation and washed once with isotonic NKM solution (130 mM NaCl, 50 mM KCl, 1.5l mM $MgCl_2$). After pelleting they were resuspended in RSB [10 mM NaCl, 10 mM Tris–HAc (pH 7.4), 1.5 mM $MgCl_2$] at a density of 4 × 10^7 cells/ml and 0.1 vol sodium deoxycholate/Nonidet P-40 mixture (5% each) was added. The suspension was gently swirled and left on ice for 10 min. Subsequently, the nuclei were removed by centrifugation at 800 *g* for 5 min. The supernatant was made 0.5% in SDS and total RNA was isolated by phenol–chloroform extraction. Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose chromatography (Maniatis *et al.*, 1982).

RNA blot hybridization

Ten μg of poly(A)$^+$ RNA was glyoxalated, fractionated on 1.0% agarose gels and transferred to Hybond-N (Amersham, UK) (Carmichael and McMaster, 1980; Thomas, 1980). As mol. wt markers *Hin*dIII-digested λ-DNA fragments were used. Denaturation of the DNA markers was performed as described for RNA. Hybridization of the Northern blots was carried out as described by Church and Gilbert (1984).

Hybrid selection, RNA transcription and in vitro translation

Plasmid pHA-1 or pHA-2 was digested with *Eco*RI. The isolated cDNA insert was denatured by boiling and treatment with 0.5 M NaOH, neutralized, and applied to nitrocellulose. Hybridization selection was performed essentially as described by Quax-Jeuken *et al.* (1984).

RNA transcript was synthesized *in vitro* from *Hin*dIII-digested pHA-4 DNA in the presence of SP6 RNA polymerase and the dinucleotide primer G(5′)ppp(5′)G (Konarska *et al.*, 1985).

In vitro translations in a nuclease-treated reticulocyte lysate (Amersham, UK) were performed for 90 min at 30°C in 15 μl reaction mixtures and analysed by SDS–PAGE. Gels containing ^{35}S-labelled proteins were impregnated with 2,5-diphenyloxazole (PPO) before being dried and fluorography was performed with Kodak XAR-5 X-ray film (Laskey and Mills, 1975).

Analysis of immunoprecipitated proteins

For the immunoprecipitation of proteins from a ^{35}S-labelled HeLa cell extract the procedure described by Mimori *et al.* (1984) was used. Immunoprecipitated proteins were analysed on 15% SDS–polyacrylamide gels.

Cell-free translation mixtures were adjusted to 1% (v/v) Triton X-100, 100 mM NaCl, 25 mM Tris–HAc (pH 7.4), 5 mM EDTA, 2.5 mg of ovalbumin per ml, 2.5% (v/v) Trasylol, in a total volume of 100 μl. Translation products were immunoprecipitated as described by Fisher *et al.* (1983).

DNA sequence analysis

cDNA fragments were digested with a variety of restriction enzymes having 4- or 6-bp recognition sequences. DNA fragments were ligated into the polylinker region of M13 mp18 or mp19 (Messing, 1983). Sequence analysis of the DNA fragments was performed by the dideoxy chain termination method (Sanger *et al.*, 1977).

The gel readings were recorded, edited and compared by the Staden programs (1979).

Acknowledgements

We would like to thank Wiljan Hendriks for advice and help in the DNA sequencing and Wilbert Boelens for the experiment described in Figure 6. The help of Jack Leunissen with computer analysis is also gratefully acknowledged. The present investigations have been carried out in part under the auspices of the Netherlands Foundation for Chemical Research (SON) and with the financial aid of the Netherlands Organization for the Advancement of Pure Research (ZWO) and the Netherlands League against Rheumatism.

References

Adam,S.A., Nakagawa,T., Swanson,M., Woodruff,T.K. and Dreyfuss,G. (1986) *Mol. Cell. Biol.*, **6**, 2932–2943.

Billings,P.B. and Hoch,S.O. (1984) *J. Biol. Chem.*, **259**, 12850–12856.

Black,D.L., Chabot,B. and Steitz,J.A. (1985) *Cell*, **42**, 737–740.

Bringmann,P. and Lührmann,R. (1986) *EMBO J.*, **5**, 3509–3516.

Brunel,C., Sri-Widada,J. and Jeanteur,P. (1985) In Hahn,F.E. (ed.), *Progress in Molecular and Subcellular Biology*. Springer-Verlag, Heidelberg, Vol. 9, pp. 1–52.

Carmichael,G.G. and McMaster,G.K. (1980) *Methods Enzymol.*, **65**, 380–391.

Chambers,J.C. and Keene,J.D. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2115–2119.

Church,G.M. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.

Fisher,D.E., Conner,G.E., Reeves,W.H., Blobel,G. and Kunkel,H.G. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6356–6360.

Garoff,H., Frischauf,A.M., Simons,K., Lehrach,H. and Delius,H. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6376–6380.

Habets,W.J., de Rooij,D.J., Salden,M.H., Verhagen,A.P., van Eekelen,C.A.G., van de Putte,L.B. and van Venrooij,W.J. (1983) *Clin. Exp. Immunol.*, **54**, 265–276.

Habets,W.J., de Rooij,D.J., Hoet,M.H., van de Putte,L.B. and van Venrooij,W.J. (1985a) *Clin. Exp. Immunol.*, **59**, 457–466.

Habets,W.J., Hoet,M.H., Bringmann,P., Lührmann,R. and van Venrooij,W.J. (1985b) *EMBO J.*, **4**, 1545–1550.

Habets,W.J., Sillekens,P.T.G., Hoet,M.H., Schalken,J.A., Roebroek,A.J.M., Leunissen,J.A.M., van den Ven,W.J.M. and van Venrooij,W.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2421–2425.

Hinterberger,M., Pettersson,I. and Steitz,J.A. (1983) *J. Biol. Chem.*, **258**, 2604–2613.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. I, pp. 49–78.

Kinlaw,C.S., Robberson,B.L. and Berget,S.M. (1983) *J. Biol. Chem.*, **258**, 7181–7189.
Konarska,M.M., Grabowski,P.J., Padgett,R.A. and Sharp,P.A. (1985) *Nature*, **313**, 552–557.
Kozak,M. (1986) *Cell*, **44**, 283–292.
Krainer,A.R. and Maniatis,T. (1985) *Cell*, **42**, 725–736.
Krämer,A., Keller,W., Appel,B. and Lührmann,R. (1984) *Cell*, **38**, 299–307.
Kruijer,W., van Schaik,F.M.A. and Sussenbach,J.S. (1981) *Nucleic Acids Res.*, **9**, 4439–4457.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Laskey,R.A. and Mills,A.D. (1975) *Eur. J. Biochem.*, **56**, 335–341.
Leff,S.E., Rosenfeld,M.G. and Evans,R.M. (1986) *Annu. Rev. Biochem.*, **55**, 1091–1117.
Lerner,M.R. and Steitz,J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495–5499.
Maniatis,T. and Reed,R. (1987) *Nature*, **325**, 673–678.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Mimori,T., Hinterberger,M., Pettersson,I. and Steitz,J.A. (1984) *J. Biol. Chem.*, **259**, 560–565.
Mount,S.M., Pettersson,I., Hinterberger,M., Karmas,A. and Steitz,J.A. (1983) *Cell*, **33**, 509–518.
Padgett,R.A., Mount,S.M., Steitz,J.A. and Sharp,P.A. (1983) *Cell*, **35**, 101–107.
Pettersson,I., Hinterberger,M., Mimori,T., Gottlieb,E. and Steitz,J.A. (1984) *J. Biol. Chem.*, **259**, 5907–5914.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211–214.
Quax-Jeuken,Y., Janssen,C., Quax,W., van den Heuvel,R. and Bloemendal,H. (1984) *J. Mol. Biol.*, **180**, 457–472.
Reddy,R. and Busch,H. (1983) *Prog. Nucleic Acids Res. Mol. Biol.*, **30**, 127–162.
Reddy,R., Henning,D. and Busch,H. (1985) *J. Biol. Chem.*, **260**, 10930–10935.
Reuter,R. and Lührmann,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8689–8693.
Reuter,R., Lehner,C.F., Nigg,E.A. and Lührmann,R. (1986) *FEBS Lett.*, **201**, 25–30.
Reuter,R., Rothe,S. and Lührmann,R. (1987) *Nucleic Acids Res.*, **15**, 4021–4034.
Rogers,J. and Wall,R. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1877–1879.
Sachs,A.B., Bond,W.M. and Kornberg,R.D. (1986) *Cell*, **45**, 827–835.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Smith,D.E. and Fisher,P.A. (1984) *J. Cell Biol.*, **99**, 20–28.
Staden,R. (1979) *Nucleic Acids Res.*, **6**, 2601–2610.
Strub,K., Galli,G., Busslinger,M. and Birnstiel,M.L. (1984) *EMBO J.*, **3**, 2801–2807.
Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.*, **7**, 1731–1739.
Tan,E.M. (1982) *Adv. Immunol.*, **33**, 167–240.
Theissen,H., Etzerodt,M., Reuter,R., Schneider,C., Lottspeich,F., Argos,P., Lührmann,R. and Philipson,L. (1986) *EMBO J.*, **5**, 3209–3217.
Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.
Van Eekelen,C.A.G. and van Venrooij,W.J. (1981) *J. Cell Biol.*, **88**, 554–563.
Wieben,E.D., Rohleder,A.M., Nenninger,J.M. and Pederson,T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7914–7918.
Wilbur,W.J. and Lipman,D.J. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 726–730.
Yang,V.W., Lerner,M.R., Steitz,J.A. and Flint,S.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 1371–1378.
Young,R.Y. and Davis,R.D. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1194–1198.
Zhuang,Y. and Weiner,A. (1986) *Cell*, **46**, 827–835.

Received on July 22, 1987; revised on September 3, 1987