Molecular cloning of the cDNA for the human U2 snRNA-specific A' protein

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

Department of Biochemistry, University of Nijmegen, St Adelbertusplein 1, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Received December 20, 1988; Revised and Accepted February 8, 1989 EMBL/Genbank accession no. X13482

ABSTRACT

The A' polypeptide is one of the protein constituents of the U2 snRNP particle. A potentially full-length cDNA clone containing the complete coding sequence for this U2 snRNP-specific protein was isolated by screening of a human λ gt11 expression vector library with an autoimmune anti-(U1,U2)RNP serum. Monospecific antibodies, eluted from the 140-150 kD fusion protein of this cDNA recombinant, specifically recognized the A' protein on immunoblots and immunoprecipitated U2 snRNP particles from nuclear extracts. The identity of the clone was confirmed by *in vitro* translation of hybrid-selected mRNA or an RNA transcript synthesized from the cDNA insert. RNA blot analysis showed strong hybridization to a single polyadenylated transcript of 1.3 kb in human cells. The nucleotide sequence of the 1054 bp cDNA contains an open reading frame of 756 bp encoding a polypeptide of 255 amino acids with a predicted molecular weight of 28,444 D. The coding sequence is preceded by a 49 bp 5'-untranslated region and followed by a 226 bp 3'-untranslated region containing a single polyadenylation signal. Most striking feature of the deduced primary structure for the A' protein is a leucine-rich region in the amino-terminal half of the polypeptide. In contrast to the other U2 snRNP-specific protein B'', the A' protein does not contain segments homologous to the RNP consensus sequences RNP1 and RNP2, common amino acid motifs found in several RNA-binding proteins. In the A' protein, however, the extremely hydrophilic carboxy terminus may constitute an RNA-binding moiety.

INTRODUCTION

U2 snRNA belongs to the family of abundant small nuclear RNAs of the U snRNA class which are found in the nucleoplasm of eukaryotic cells (1). In the nucleus, however, it does not function as a naked RNA molecule, but is organized in a ribonucleoprotein particle (U2 snRNP). In human cells at least nine polypeptides are associated with purified U2 snRNP, termed A', B, B', B'', D, D', E, F and G (2). While the proteins B, B', D, D', E, F and G are also present in the other abundant nucleoplasmic snRNPs U1 and U5 and the dimeric snRNP particle U4/U6, the A' and B'' polypeptides are unique to the U2 snRNP species (2-5). All these snRNPs appear to participate in the splicing of nuclear messenger RNA precursors (pre-mRNAs), which takes place in a large multicomponent complex termed spliceosome (for a review see 6). In this spli-

cing complex U2 snRNP interacts with the branch point (7). This central area in the intron where lariat formation takes place is primarily selected by its distance from the 3' splice site. In contrast to U1 snRNP whose binding to the 5' splice site is largely determined by sequence complementarity between the 5' end of the U1 snRNA and the 5' splice junction of the pre-mRNA (8,9), the requirements for U2 snRNP binding are more complex since only a weak and limited consensus seems to exist for mammalian branch points. Still base-pairing between U2 snRNA and the degenerate sequences surrounding the branch point appear to play a role in the selection of the branch-site nucleotide in metazoan splicing (10,11). Furthermore, U2 snRNP binding is ATP-dependent and a purified U2 snRNP alone cannot bind to its cognate pre-mRNA binding site. At least two protein factors, involved in the complex formation at the 3' splice site are also necessary for a productive association of U2 snRNP with the branch point region (12). One of these protein factors might be identical to the intron binding protein (IBP) that was shown to specifically interact with pre-mRNA substrates at the 3' end of the intron (13,14), whereas the other component coelutes through several purification steps with the auxiliary protein splicing factor, U2AF, that binds to the polypyrimidine stretch at the 3' splice site and is necessary for the U2 snRNP branch point interaction (12,15). To what extent the protein components of the U2 snRNP particle itself contribute to this or any of the following events that take place at the branch point or the nearby 3' splice site remains unclear. It is thought, however, that the unique polypeptides A' and B'' are responsible for the function of U2 snRNP in pre-mRNA splicing and make up essential elements for the specific interactions between U2 snRNP and pre-mRNA or other components of the splicing machinery.

Apart from their cellular function, the U2 snRNP-specific polypeptides A' and B'' also play a role as antigens in autoimmune diseases, since both proteins carry antigenic determinants for autoantibodies. Like anti-Sm autoantibodies, which have been mainly assigned to the common polypeptides B, B' and D, or anti-(U1)RNP autoantibodies, which recognize the U1 snRNP-specific proteins 70K, A and C, these antinuclear antibodies with specificity for the U2 snRNP-specific proteins are found in sera from patients with systemic lupus erythematosus (SLE) and other related disorders (16). The determination of the primary structure and the identification of immunoreactive epitopes in the snRNP proteins are a prerequisite for the elucidation of the cellular function and the role these polypeptides play in autoimmune disease pathogenesis. As the first step in a study of the A' protein we report here the mo-

lecular cloning and sequencing of a human cDNA clone containing the complete coding sequence for this U2 snRNP-specific protein.

MATERIALS AND METHODS

Cell labelling and preparation of HeLa cell total protein extracts or nuclear extracts: 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. Labelling of cellular protein with [^{35}S]methionine and preparation of total protein fractions was performed as previously described (17). Labelling of the cells with [^{32}P]orthophosphate and preparation of nuclear extracts was carried out as described by Bringmann *et al.* (18).

Antibody screening, immunoblotting and immunoprecipitation: A patient's anti-(U1,U2)RNP serum, termed V26 (4), was used (diluted 1:500) to screen a human teratocarcinoma cDNA library constructed with the λ gt11 vector (a gift from J.Skowronski, NCI/NIH, Bethesda) as previously described (17). To detect specifically bound antibody, ^{125}I -labelled sheep anti-human Ig [F(ab) $_2$ fragment] was used.

For immunoblotting, bacterial lysates were prepared essentially as described by Adam *et al.* (19) from lysogens prepared in *E. coli* strain Y1089 (20). Proteins were fractionated on SDS-polyacrylamide gels and electroblotted onto nitrocellulose sheets as described (21). For the detection of antigens the protein blots were treated and processed according to the method of Habets *et al.* (22).

Immunoprecipitation of proteins from ^{35}S -labelled HeLa cell extracts or from cell-free translation mixtures was carried out as previously described (17). Precipitation of ^{32}P -labelled snRNP particles from nuclear extracts was performed as described by Bringmann *et al.* (18).

DNA sequence analysis: For sequence analysis, suitable cDNA fragments were ligated into M13 mp18 or mp19 (23), grown in JM109, and subjected to dideoxy chain termination reactions (24).

Hybrid selection, RNA transcription and in vitro translation: Plasmid pHA' was digested with EcoRI. 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.* (25).

RNA transcripts were synthesized *in vitro* from HindIII-digested pHA' DNA in the presence of the dinucleotide primer G(5')ppp(5')G and SP6 RNA polyme-

rase (26).

In vitro translations in a nuclease-treated reticulocyte lysate (Amersham, UK) were performed for 90 min at 30°C in 25 µl reaction mixtures.

RNA blot analysis: Total RNA was isolated from HeLa cells (17) and poly(A)⁺ RNA was selected by oligo(dT)-cellulose chromatography. Ten µg of poly(A)⁺ RNA was glyoxalated, fractionated on a 1.0% agarose gel and transferred to Hybond-N (Amersham, UK)(27,28). Concurrently electrophoresed 18S and 28S ribosomal RNAs from HeLa cells were used as molecular weight markers.

Hybridization of Northern blots was performed as described by Church and Gilbert (29). After hybridization, filters were washed twice with 0.5 M sodium phosphate, pH 7.0, 1% SDS, 1 mM EDTA at 65°C for 30 min and once with 0.25 M sodium phosphate, pH 7.0, 1% SDS, 1 mM EDTA at 65°C for 30 min.

RESULTS AND DISCUSSION

Isolation and identification of a full-length cDNA clone for the U2 snRNP-specific A' protein.

Serum V26, an anti-(U1,U2)RNP serum containing antibodies against the U2 snRNP-specific proteins A' and B" (ref. 4 and Figure 1A, lane 1) was used to screen a human teratocarcinoma cDNA library constructed in the phage vector λgt11 for the selection of recombinants expressing the A' antigen. Initial screening of approximately 600,000 independent plaques yielded several antigen-producing clones. Among these, clones containing coding sequences for the U1 snRNP-specific proteins 70K (unpublished results) and A (17) and the U2 snRNP-specific B" protein (30) were identified. In addition, one putative A' clone was selected. After plaque purification this positive clone, designated λHA', was first characterized by immunoblot analysis. Therefore, λHA' phage was transferred to a lysogenic host and bacterial lysates, prepared after induction of the phage β-galactosidase gene by isopropyl-β-D-thiogalactopyranoside (IPTG), were fractionated on a 7.5 % SDS-polyacrylamide gel and blotted onto a nitrocellulose filter. A single polypeptide of approximately 140-150 kD in the bacterial lysate of λHA' was specifically recognized by antibodies from the anti-(U1,U2)RNP serum. This hybrid protein was absent from the lysate of wild-type λgt11-infected cells (data not shown). The region of a preparative immunoblot containing the expressed fusion protein or a non-reactive control region were used to immunoselect antibody from the polyvalent autoantiserum. The eluted antibodies were assayed for their reactivity with intact U snRNPs or snRNP proteins. As shown in Figure 1, antibody affinity-purified from the fusion protein precipitated U2 snRNPs (Figure 1B, lane 3) and on an

Figure 1. Immunoblotting (A) and immunoprecipitation of snRNPs (B) with antibodies affinity-purified from λ HA' fusion protein. (A) A strip of an immunoblot containing total lysate of λ HA'-infected cells was processed for the detection of antigens and the area of the blot containing the fusion protein was excised. This region was used to adsorb antibodies specific for the fusion protein from the human anti-(U1,U2)RNP serum V26 containing antibodies against the U1 snRNP-specific proteins 70K and A and the U2 snRNP-specific proteins A' and B" (4). The antibodies were eluted and re-used to probe an immunoblot of HeLa total protein extract. **Lanes:** (1) total human anti-(U1,U2) RNP serum; (2) affinity-purified antibody from the fusion protein of λ HA'. (B) Antibodies affinity-purified from the region containing the fusion protein or from a non-reactive control region of an immunoblot as described above were re-used for immunoprecipitation of snRNPs from 32 P-labelled HeLa cell nuclear extracts. snRNAs present in the immune precipitates were fractionated on polyacrylamide gels. **Lanes:** (1) total human anti-(U1,U2)RNP serum; (2) affinity-purified antibody from the control region; (3) antibody affinity-purified from the fusion protein.

immunoblot of HeLa total protein extract specifically recognized the A' protein (Figure 1A, lane 2). Background levels of U1 and U2 snRNPs, as seen in the immunoprecipitation with antibodies eluted from the control region of the immunoblot (Figure 1B, lane 2), probably are due to non-specific binding of antibodies to nitrocellulose.

To exclude the possibility that we had cloned a cross-reacting polypeptide sharing at least one epitope with the A' protein, the identity of clone λ HA' was confirmed by hybrid selection of HeLa poly(A)⁺ RNA and subsequent *in vitro* translation. cDNA insert of λ HA' was isolated by digestion with EcoRI and spotted onto nitrocellulose. After hybridization with HeLa poly(A)⁺ RNA, filter-bound mRNA was eluted and translated *in vitro* in the presence of [35 S]me-

Figure 2. Hybrid-selection translation of HeLa mRNA with pHA'. 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' cDNA fragment was performed with the anti-(U1,U2)-RNP serum described in Figure 1. Immunoprecipitation of U snRNP proteins from in vivo [³⁵S]methionine labelled HeLa total protein extract or from a translation reaction directed by HeLa total poly(A)⁺ RNA were carried out with a mixture of the anti-(U1,U2)RNP serum and a human anti-RNP/Sm serum recognizing on an immunoblot the snRNA-associated proteins 70K, A, B', B, C, and D. **Lanes:** (1) Immunoprecipitation of U snRNP proteins from in vivo [³⁵S]methionine labelled HeLa protein extract; (2) immunoprecipitation of translation products from a control translation (no mRNA added); (3) immunoprecipitation of translation products from pHA' hybrid-selected mRNA; (4) immunoprecipitation of translation products from total HeLa poly(A)⁺ RNA. The position of the A' protein in lanes (1) and (4) was determined from the autoradiogram after a longer exposure time. Molecular weights (in kD) of protein standards are indicated on the right (M).

thionine using a rabbit reticulocyte lysate. The hybrid-arrested mRNA specifically translated in vitro two polypeptides of approximately 28 kD and 13 kD, which both were immunoprecipitable with the anti-(U1,U2)RNP serum V26 initially used for the isolation of the pHA' recombinant (Figure 2, lane 3). The 28 kD protein comigrates with the U2 snRNP-specific A' protein derived from in vivo [³⁵S]methionine labelled HeLa protein extract (Figure 2, lane 1) or from an in vitro translation reaction directed by total HeLa poly(A)⁺ RNA

Figure 3. Northern blot analysis of poly(A)⁺ RNA from HeLa cells probed with ³²P-labelled nick-translated cDNA insert from the pHA' clone. The size of concurrently electrophoresed ribosomal RNAs from HeLa cells is indicated on the left.

(Figure 2, lane 4), respectively. Since Northern blot analysis revealed only a single mRNA species when probed with the cDNA insert of λ HA' (Figure 3), the extra polypeptide band of 13 kD, which was also specifically immunoprecipitated from the translation reaction directed by the hybrid-selected mRNA (Figure 2, lane 3), presumably results from anomalous translation initiation at some site in the same reading frame or, alternatively, represents a degradation product of the A' protein.

The λ HA' clone contains an insert of about 1.1 kb as determined by restriction mapping and sizing of the DNA fragments on agarose gels. The size of the homologous mRNA was determined by RNA blot analysis. Probing of a blot containing total HeLa poly(A)⁺ RNA with the nick-translated 1.1 kb cDNA insert of λ HA' resulted in cross-hybridization to a single mRNA species of 1.3 kb (Figure 3). Assuming an average length of approximately 150 residues for the poly(A) tail of a eukaryotic mRNA, the λ HA' cDNA insert is close to full-length and might well contain the entire coding region for the 28 kD A' polypeptide. To experimentally verify this assumption, the cDNA insert was isolated from λ HA' by digestion with EcoRI and recloned into the SP6 bacteriophage promoter plasmid, pSP65, to yield pHA'. Plasmid DNAs, containing the insert in either orientation, were linearized by digestion with HindIII and used as templates in *in vitro* transcription reactions with SP6 RNA polymerase. The resulting RNAs, representing sense and anti-sense RNA copies of the clone, were translated in a rabbit reticulocyte translation system employing [³⁵S]-

Figure 4. In vitro translation of pHA' encoded RNA. Immunoprecipitation of the translation products was performed with the same anti-(U1,U2)RNP serum and anti-RNP/Sm serum that were used in the hybrid selection experiment (Figure 2). **Lanes:** (1) translation products of pHA' encoded anti-sense RNA; (2) translation products of pHA' encoded sense RNA; (3) and (6) immunoprecipitation of U snRNP proteins from in vivo [35 S]methionine labelled HeLa protein extract with a mixture of the anti-(U1,U2)RNP serum and the anti-RNP/Sm serum; (4) immunoprecipitation of translation products encoded by anti-sense RNA from pHA' with anti-(U1,U2)RNP serum; (5) immunoprecipitation of translation products encoded by sense RNA from pHA' with anti-(U1,U2)RNP serum. The position of the cellular A' antigen in lanes (3) and (6) was determined from the autoradiogram after a longer exposure time. Molecular weights (in kD) of protein standards are indicated on the left (M).

methionine and subsequently subjected to SDS-polyacrylamide gel analysis. Two prominent labelled species were observed in the translation reaction directed by the sense RNA (Figure 4, lane 2). The longest polypeptide synthesized had the same mobility as the cellular A' protein (Figure 4, lanes 3 and 6). This primary translation product could also be immunoprecipitated from the translation mixture (Figure 4, lane 5), together with the same additional polypeptide of lower molecular weight that was already observed in the hybrid selection experiment. Taken together, these results demonstrate that RNA transcribed from the pHA' cDNA insert encodes for the same polypeptides as the authentic A' protein mRNA.

Nucleotide sequence and deduced amino acid sequence of the A' protein.

The complete nucleotide sequence of the human A' protein cDNA contained in pHA' was determined according to the strategy outlined in Figure 5. The se-

Figure 5. Partial restriction map and sequencing strategy for the full-length cDNA clone pHA'. The thick bar represents the protein coding region.

quence so obtained is shown in Figure 6 and covers 1054 nucleotides (EcoRI-linkers included). The only large uninterrupted reading frame starts at the beginning of the sequence and terminates at position 822. Remarkably, this reading frame is not in register with the β -galactosidase coding sequence. Still detectable quantities of antigen are produced as a 140-150 kD fusion protein. Therefore, internal initiation of protein synthesis at some site(s) within the insert, as has been observed in other systems (31,32), could be ruled out. To explain this phenomenon it was necessary to assume the presence of three EcoRI-linker sequences instead of one. Incubation of an immunoblot with anti- β -galactosidase revealed only a relatively low quantity of the 140-150 kD fusion protein, whereas a hybrid protein with approximately the size of wild-type β -galactosidase was detected in large amounts (data not shown). The latter protein does not react with antibodies from the anti-(U1,U2)RNP serum. These findings could be explained by the fact that the putative stretch of EcoRI-linkers takes the configuration of a stem-and-loop structure causing termination of transcription. Only occasionally the RNA polymerase can pass through this multi-linker-region and reach the cDNA insert sequence for transcription.

The designated ATG initiation codon in the open reading frame is located at position 57, resulting in a translational unit which encodes 255 amino acids (Figure 6). In most eukaryotic mRNAs no extraneous ATG triplets are found upstream of the functional initiation codon for protein synthesis. Indeed, the designated methionine-encoding initiation site of this protein coding region also is the 5'-proximal ATG codon occurring in the sequence. Furthermore, with an A residue at position -3 and a G residue at position +4 it exists within a sequence which fits the optimal context for translation initiation at the most crucial positions (33). Starting with a TGA translational stop codon at position 822 the sequence terminates with a 3' non-coding re-

```

                                20                                40
                    ga att ccg cgg gag gcc acg ggc ttt cca cag cgc ggg gga acg gga ggc tgc agg

                                60                                80                                100
1   atg gtc aag ctg acg gcg gag ctg atc gag cag gcg gcg cag tac acc aac gcg gtg cgc   20
   met val lys leu thr ala glu leu ile glu gln ala ala gln tyr thr asn ala val arg

                                120                                140                                160
21  gac cgg gag ctg gac ctc cgg ggg tat aaa att ccc gtc att gaa aat cta ggt gct acg   40
   asp arg glu leu asp leu arg gly tyr lys ile pro val ile glu asn leu gly ala thr

                                180                                200                                220
41  tta gac cag ttt gat gct att gat ttt tct gac aat gag atc agg aaa ctg gat ggt ttt   60
   leu asp gln phe asp ala ile asp phe ser asp asn glu ile arg lys leu asp gly phe

                                240                                260                                280
61  cct ttg ttg aga aga ctg aaa aca ttg tta gtg aac aac aac aga ata tgc ogt ata ggt   80
   pro leu leu arg arg leu lys thr leu leu val asn asn asn arg ile cys arg ile gly

                                300                                320                                340
81  gag gga ctt gat cag gct ctg ccc tgt ctg aca gaa ctc att ctc acc aat aat agt ctc   100
   glu gly leu asp gln ala leu pro cys leu thr glu leu ile leu thr asn asn ser leu

                                360                                380                                400
101 gtg gaa ctg ggt gat ctg gac cct ctg gca tct ctc aaa tog ctg act tac cta agt atc   120
   val glu leu gly asp leu asp pro leu ala ser leu lys ser leu thr tyr leu ser ile

                                420                                440                                460
121 cta aga aat ccg gta acc aat aag aag cat tac aga ttg tat gtg att tat aaa gtt ccg   140
   leu arg asn pro val thr asn lys lys his tyr arg leu tyr val ile tyr lys val pro

                                480                                500                                520
141 caa gtc aga gta ctg gat ttc cag aaa gtg aaa cta aaa gag cgt cag gaa gca gag aaa   160
   gln val arg val leu asp phe gln lys val lys leu lys glu arg gln glu ala glu lys

                                540                                560                                580
161 atg ttc aag ggc aaa cgg ggt gca cag ctt gca aag gat att gcc agg aga agc aaa act   180
   met phe lys gly lys arg gly ala gln leu ala lys asp ile ala arg arg ser lys thr

                                600                                620                                640
181 ttt aat cca ggt gct ggt ttg cca act gac aaa aag aga ggt ggg cca tct cca ggg gat   200
   phe asn pro gly ala gly leu pro thr asp lys lys arg gly gly pro ser pro gly asp

                                660                                680                                700
201 gta gaa gca atc aag aat gcc ata gca aat gct tca act ctg gct gaa gtg gag agg ctg   220
   val glu ala ile lys asn ala ile ala asn ala ser thr leu ala glu val glu arg leu

                                720                                740                                760
221 aag ggg ttg ctg cag tct ggt cag atc cct ggc aga gaa cgc aga tca ggg ccc act gat   240
   lys gly leu leu gln ser gly gln ile pro gly arg glu arg arg ser gly pro thr asp

                                780                                800                                820
241 gat ggt gaa gaa gag atg gaa gaa gac aca gtc aca aac ggg tcc tga gca gtg agg cag
   asp gly glu glu glu met glu glu asp thr val thr asn gly ser ***

                                840                                860                                880
   atg tat aat aat agg ccc tct tgg aac aag tct tgc ttt tog aac atg gta taa tag cct

                                900                                920                                940
   tgt ttg tgt tag caa agt gga atc tat cag cat tgt tga aat gct taa gac tgc tgc tga

                                960                                980                                1000
   taa ttt tgt aat ata agt ttt gaa atc taa atg tca att ttc tac aaa tta taa aaa taa

                                1020                                1040
   act cca ctc tct atg cta aaa aaa aaa aaa agg aat tc

```

gion of 212 nucleotides. At the 3' end a stretch of 14 adenines is found, which is preceded by the consensus polyadenylation signal AATAAA (34) located 22 nucleotides upstream from the start of the polyadenylate tract.

The deduced amino acid sequence for the A' protein is shown in Figure 6. The molecular weight calculated from this primary structure is 28,444 D, which is in good agreement with the observed molecular weight of 28 kD for the A' protein on SDS-polyacrylamide gels. The polypeptide contains a total of only three methionine residues, located at positions 1, 161 and 246 in the amino acid sequence (Figure 6). This rather low content of methionine is consistent with predictions from *in vivo* studies in which the A' protein in contrast to most other known snRNP proteins is hardly labelled by [³⁵S]methionine (Figure 3, lane 4 and ref. 35). Most striking feature of the polypeptide is the high content of leucine residues (13 % of all amino acids). Most of the leucines are clustered in a region spanning amino acids 57-122 (Figure 6). Recently, the periodic array of leucines was noted in segments of several DNA binding proteins with a predicted α -helical conformation. This hypothetical structure was referred to as the "leucine zipper" and may represent a characteristic property of a certain category of DNA binding proteins (36). The secondary structure of the A' protein, however, as predicted by the Chou-Fasman algorithm (37) did not unambiguously reveal α -helices in any region of the protein nor does the leucine-rich region show a periodic repetition of this amino acid.

It was previously shown that the so-called Sm binding site in U2 snRNA is a prerequisite for binding of the U2 snRNP-specific proteins. In addition, RNA structure 3' to this segment is required (38). The A' protein therefore appears to interact both with the so-called Sm core protein complex (proteins D, E, F and G) and directly with RNA in the 3' half of the U2 snRNA molecule. This direct association of the A' protein to the RNA moiety of the U2 snRNP particle was experimentally demonstrated in VSV-infected cells, where A' protein is bound to precursor U2 snRNA transcripts at times when binding of the core protein complex is blocked (39). In this respect computer-derived hydrophathy plots of the amino acid sequence (data not shown) revealed a strongly hydrophilic carboxy terminus which might reflect an RNA binding region of the A' protein. No similarities to the recently described RNP consensus sequences

Figure 6. Nucleotide sequence of pH^{A'} and deduced amino acid sequence of the encoded A' protein. The predicted amino acids are shown below their triplets. The polyadenylation signal is underlined and the stop codon is marked by asterisks.

RNP1 and RNP2, that are stretches of 8 and 6 amino acids, respectively, and represent conserved parts of an RNA-binding domain (40,41), were found in the deduced amino acid sequence of the A' protein. In the primary structure of the snRNP core proteins D (42) and E (43) and the U1 snRNP-specific protein C (44,45) the RNP consensus motifs are also absent. Still all these polypeptides may interact with RNA because the RNP consensus is not a general property of all RNA-binding proteins (38). It can be speculated that the interaction of these proteins with RNA is mere structural, whereas the RNP consensus segments might be involved in more specific RNA-protein interactions, which in U1 snRNP would then be reserved for the 70K and A protein and in U2 snRNP for the B" protein.

The availability of a full-length cDNA clone for the A' polypeptide opens the way to a detailed study of the function and expression of this snRNP protein and its assembly in U2 snRNP particles. Furthermore, it provides a means to investigate its antigenic structure, especially the epitopes that are the molecular targets for antibodies raised against this protein in autoimmune disorders.

ACKNOWLEDGEMENTS

The expert technical assistance of Bea Hageman is gratefully acknowledged. We would also like to thank Jack Leunissen for his help with computer analyses. 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 Scientific Research (NWO), Organon Teknika nv, the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands League against Rheumatism.

*To whom reprint requests should be addressed

REFERENCES

1. Brunel, C., Sri-Widada, J. and Jeanteur, P. (1985) In Hahn, F.E. (ed), *Progress in Molecular and Subcellular Biology*. Springer-Verlag, Vol. 9, pp. 1-52.
2. Bringmann, P. and Lührmann, R. (1986) *EMBO J.* 5, 3509-3516.
3. Mimori, T., Hinterberger, M., Pettersson, I. and Steitz, J.A. (1984) *J. Biol. Chem.* 259, 560-565.
4. Habets, W., Hoet, M., Bringmann, P., Lührmann, R. and van Venrooij, W.J. (1985) *EMBO J.* 4, 1545-1550.
5. Reuter, R. and Lührmann, R. (1986) *Proc. Natl. Acad. Sci. USA* 83, 8689-8693.
6. Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Krämer, A., Frendewey, D. and Keller, W. (1988) In Birnstiel, M.L. (ed), *Structure and Function of Major and Minor small nuclear RNAs*. Springer-Verlag, pp. 115-154.
7. Black, D.L., Chabot, G. and Steitz, J.A. (1985) *Cell* 42, 737-750.

8. Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) *Cell* **38**, 299-307.
9. Zhuang, Y. and Weiner, A. (1986) *Cell* **46**, 827-835.
10. Hartmuth, K. and Barta, A. (1988) *Mol. Cell. Biol.* **8**, 2011-2020.
11. Reed, R. and Maniatis, T. (1988) *Genes Dev.* **2**, 1268-1276.
12. Krämer, A. (1988) *Genes Dev.* **2**, 1155-1167.
13. Gerke V. and Steitz J.A. (1986) *Cell* **47**, 973-984.
14. Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986) *Cell* **47**, 755-766.
15. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) *Cell* **52**, 207-219.
16. Tan, E.M. (1982) *Adv. Immunol.* **33**, 167-240.
17. Sillekens, P.T.G., Habets, W.J., Beijer, R.P. and van Venrooij, W.J. (1987) *EMBO J.* **6**, 3841-3848.
18. Bringmann, P., Reuter, R., Rinke, J., Appel, B., Bald, R. and Lührmann, R. (1983) *J. Biol. Chem.* **258**, 2745-2747.
19. Adam, S.A., Nakagawa, T., Swanson, M., Woodruff, T.K. and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6**, 2932-2943.
20. 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.
21. 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.
22. Habets, W.J., de Rooij, D.J., Hoet, M.H., van de Putte, L.B. and van Venrooij, W.J. (1985) *Clin. Exp. Immunol.* **59**, 457-466.
23. Messing, J. (1983) *Methods Enzymol.* **101**, 20-78.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
25. Quax-Jeuken, Y., Janssen, C., Quax, W., van den Heuvel, R. and Bloemendal, H. (1984) *J. Mol. Biol.* **180**, 457-472.
26. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) *Nature (London)* **313**, 552-557.
27. Carmichael, G.G. and McMaster, G.K. (1980) *Methods Enzymol.* **65**, 380-391.
28. Thomas, P.S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205.
29. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991-1995.
30. Habets, W.J., Sillekens, P.T.G., Hoet, M.H., Schalken, J.A., Roebroek, A.J.M., Leunissen, J.A.M., van de Ven, W.J.M. and van Venrooij, W.J. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2421-2425.
31. Young, R.A. and Davis, R.W. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1194-1198.
32. Kemp, D.J. and Cowman, A.F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4520-4524.
33. Kozak, M. (1986) *Cell* **44**, 283-292.
34. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* **263**, 211-214.
35. Kinlaw, C.S., Schwartz, S.K. and Berget, S.M. (1982) *Mol. Cell. Biol.* **2**, 1159-1166.
36. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) *Science* **240**, 1759-1764.
37. Chou, P.Y. and Fasman, G.D. (1978) *Adv. Enzymol.* **47**, 45-148.
38. Mattaj, I.W., Habets, W.J. and van Venrooij, W.J. (1986) *EMBO J.* **5**, 997-1002.
39. Fresco, L.D., Kurilla, M.G. and Keene, J.D. (1987) *Mol. Cell. Biol.* **7**, 1148-1155.
40. Swanson, M.S., Nakagawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7**, 1731-1739.
41. Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) *TIBS* **13**, 86-91.

- 42. Rokeach, L.A., Haselby, J.A. and Hoch, S.O. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4832-4836.
- 43. Wieben, E.D., Rohleder, A.M., Nenninger, J.M. and Pederson, T. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 7914-7918.
- 44. Sillekens, P.T.G., Beijer, R.P., Habets, W.J. and van Venrooij, W.J. (1988) *Nucl. Acids Res.* **16**, 8307-8321.
- 45. Yamamoto, K., Miura, H., Moroi, Y., Yoshinoya, S., Goto, M. Nishioka, K. and Miyamoto, T. (1988) *J. Immunol.* **140**, 311-317.