Volume 16 Number 17 1988 Nucleic Acids Research

Human U1 snRNP-specific C protein: complete cDNA and protein sequence and identification of a multigene family in mammals

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 June 28, 1988; Revised and Accepted August 3, 1988 Accession no. X12517

ABSTRACT

A complementary DNA clone for the human U1 snRNP-specific C protein has been isolated. The nucleotide sequence of the 733 bp cDNA insert includes a 15 bp 5'-untranslated region, an open reading frame of 477 bp corresponding to 159 amino acids (M_r=17,373 D), and a 223 bp 3'-untranslated region. The identity of the clone was confirmed by in vitro translation of hybrid-selected mRNA or an RNA transcript synthesized from the cDNA. The in vitro synthesized C protein has a slightly greater mobility on SDS-polyacrylamide gels, indicating that the in vivo product is post-translationally modified. The deduced primary structure contains a segment of high proline and methionine content. A region homologous to the RNP consensus sequence, found in the other two U1 snRNP-specific proteins 70K and A, is absent. Analysis of genomic DNA restriction enzyme digests shows hybridizing fragments in the genome of all vertebrate classes. The results are consistent with multi-copy representation of the C protein gene in mammals, whereas in the other vertebrate classes the related protein seems to be encoded by a single-copy gene.

INTRODUCTION

Small nuclear ribonucleoproteins (snRNPs) are a class of RNA-containing particles in the nucleus of eukaryotic cells (1). They consist of an uridylate-rich small RNA (U snRNA) complexed wih several proteins. The U snRNAs are synthesized in the nucleus as precursors that are transported to the cytoplasm where they are processed to their mature sizes, assembled into snRNPs, and then transported back to the nucleus (for a review see 2). With about 10^6 copies per cell the U1 snRNP particle is the most abundant among these snRNP species. So far, ten polypeptides have been identified as constituents of the U1 snRNP particle. Seven of these polypeptides with approximate molecular weights 29 kD (B'), 28 kD (B), 16 kD (D), 15.5kD (D'), 12 kD (E), 11 kD (F) and 10 kD (G) make up the RNP core structure, which is also found in the other major monomeric snRNP species U2 and U5 and in the dimeric snRNP particle U4/U6 (3). The smaller core polypeptides D, E, F and G first assemble into an RNA-free 6S heterooligomeric complex (4), which subsequently binds to an internal single-stranded sequence of the type $A(U)_nG$ (with n>3)

© IRL Press Limited, Oxford, England.

that is conserved among snRNAs U1, U2, U4 and U5 (5). Further methylation of the 7-methylguanosine cap structure of the RNA polymerase II transcribed U snRNAs is dependent on the presence of these core proteins (5).

In addition to the common polypeptides, U1 snRNP contains at least three characteristic proteins with apparent molecular weigths 70 kD (70K), 31 kD (A) and 22 kD (C) (3,6,7). These unique polypeptides may be involved in the fine tuning of the function that U1 snRNP has in the splicing of nuclear precursors to mature messenger RNAs, a process which occurs in a large multicomponent complex termed a spliceosome (for a review see 8). U1 snRNP specifically recognizes the 5' splice sites of these pre-mRNAs. Although this interaction is mediated by RNA base pairing between the complementary nucleotide sequences of the 5' end of the U1 snRNA and the 5' splice junction of the pre-mRNA (9,10), the specificity of this association appears to be guided by the U1 snRNA protein components (11). The precise role of the individual proteins is still unknown.

The snRNP proteins are not only of interest because they participate in the processing of nuclear pre-mRNAs, but they also play a role as antigens in the autoimmune response of patients with connective tissue diseases (12). The use of these so-called anti-Sm and anti-RNP sera has greatly facilitated the study of snRNPs and their protein components. The U1 snRNP-specific proteins 70K and A, and with a lower frequency also polypeptide C, are the target antigens recognized by anti-RNP autoantibodies. Using human autoimmune sera containing these anti-RNP activities, cDNA clones encoding the human 70K protein (13,14) and A protein (15) have been isolated from libraries constructed in expression vectors. In this study we describe the molecular cloning and sequence analysis of a full-length cDNA clone encoding the human U1 snRNP-specific protein C. In vitro translation of an SP6 transcript derived from this cDNA revealed that the C protein is post-translationally modified. In Southern blot analysis the C protein shows DNA sequence conservation in the genome of all vertebrate classes. Furthermore, the results indicate that it is encoded by a member of a multigene family in mammals.

MATERIALS AND METHODS

Antibody screening, immunoblotting and immunoprecipitation: 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 (15). To detect specifically bound antibody, ^{125}I-labelled sheep anti-human Ig [F(ab)$_2$ fragment] was used.

For immunoblotting, fusion protein was prepared essentially as described by Adam et al. (16) from lysogens prepared in E. coli strain Y1089 (17). Proteins were fractionated on SDS-polyacrylamide gels and electroblotted onto nitrocellulose sheets as described (18). For the detection of antigens the protein blots were treated and processed according to the method of Habets et al. (19).

Immunoprecipitation of proteins from ^{35}S-labelled HeLa cell extracts or from cell-free translation mixtures was carried out as previously described (15).

DNA sequence analysis: For sequence analysis, suitable cDNA fragments were ligated into M13 mp18 or mp19 (20), grown in JM101, and subjected to dideoxy chain termination reactions (21).

Hybrid selection, RNA transcription and in vitro translation: Plasmid pHC-1 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. (22).

RNA transcripts were synthesized in vitro from HindIII-digested pHC-1 DNA in the presence of the dinucleotide primer G(5')ppp(5')G and SP6 RNA polymerase (23).

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 and DNA blot analysis: Total RNA was isolated from Hela cells (15) and poly(A)$^+$ RNA was selected by oligo(dT)-cellulose chromatography. Ten µg of poly(A)$^+$ RNA was glyoxalated, fractionated on 1.0% agarose gels and transferred to Hybond-N (Amersham, UK)(24,25). HindIII-digested λ-DNA fragments were used as molecular weight markers. Denaturation of the DNA markers was performed as described for RNA.

Total genomic DNA was isolated from diverse tissues of all species described using standard methods (26). Ten-microgram amounts were digested, electrophoresed on 0.7% agarose gels and transferred onto nitrocellulose. Hybridization of Northern and Southern blots was performed as described by Church and Gilbert (27), except that for Southern blotting 100 µg of herring sperm single-stranded DNA was added in the prehybridization and hybridization mix. 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.

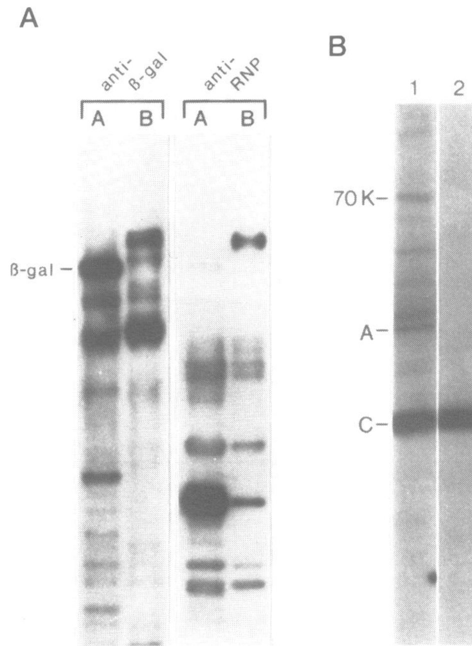

Figure 1. Immunoblot analysis of the fusion protein produced by the lysogen of λHC-1 **(A)** and immunoblotting with antibodies affinity-purified from λHC-1 fusion protein **(B)**. **(A)** Purified fusion protein was resolved on a 7.5% SDS-polyacrylamide gel along with total lysate from λgt11-infected cells and transferred to nitrocellulose. Identical immunoblots were probed either with the patient anti-(U1)RNP serum with which the clone was 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 λHC-1. **(B)** A strip of an immunoblot as described above 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 a human anti-(U1)RNP serum containing antibodies against the 70K, A, and C proteins. The antibodies were eluted 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 λHC-1.

RESULTS

Isolation and identification of a full-length cDNA clone for the U1 snRNP-specific C protein.

An anti-(U1)RNP serum containing a high titer of antibody against the U1 snRNP-specific C protein (Figure 1B, lane 1) was used to screen a human teratocarcinoma cDNA library constructed in the vector λgt11 for recombinants expressing this antigen. Initial screening of about 600,000 individual plaques

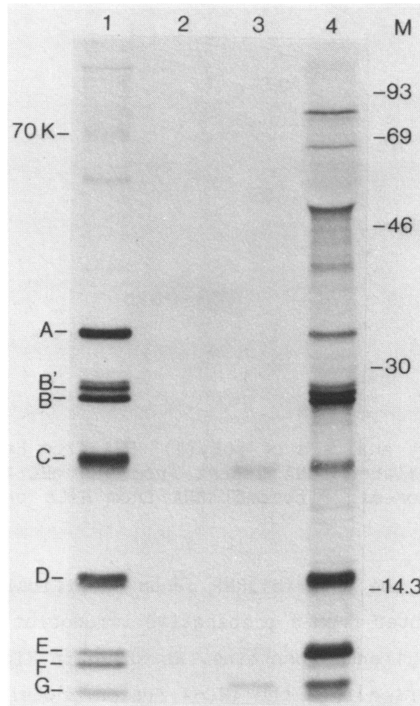

Figure 2. Hybrid-selection translation of HeLa mRNA with pHC-1. Immunoprecipitations were performed with a human anti-RNP/Sm serum recognizing on an immunoblot of HeLa total protein extract the snRNA-associated proteins 70K, A, B', B, C, and D. **Lanes: (1)** Immunoprecipitation of U snRNP proteins from in vivo ^{35}S-methionine labelled HeLa protein extract; **(2)** immunoprecipitation of translation products from a control translation (no mRNA added); **(3)** immunoprecipitation of translation products from pHC-1 hybrid-selected mRNA; **(4)** immunoprecipitation of translation products from total HeLa poly(A)$^{+}$RNA. Molecular weights (in kD) of protein standards are indicated on the right **(M)**.

yielded a single clone, designated λHC-1, which possibly contained coding sequences for the C polypeptide.

To confirm reactivity with antigenic determinants encoded by the cDNA insert, λHC-1 and wild-type λgt11 were used to infect lysogenic strain Y1089. Bacterial lysates were prepared after induction of the phage β-galactosidase gene by isopropyl-β-D-thiogalactopyranoside (IPTG) and analyzed by immunoblotting (Figure 1A). The 116 kD β-galactosidase present in the wild-type culture was replaced by a 126 kD fusion protein in the λHC-1 culture, which was reactive both with antibodies to β-galactosidase and with antibodies from the anti-(U1)RNP serum. No reaction was observed with serum from normal individuals.

Figure 3. Northern blot analysis of poly(A)$^+$ RNA from HeLa cells probed with ^{32}P-labelled nick-translated cDNA insert from the pHC-1 clone. The size of concurrently electrophoresed ribosomal RNA from HeLa cells is indicated on the left.

The antibodies from the anti-(U1)RNP serum specifically reacting with the fusion protein were eluted from a preparative immunoblot (28) and assayed for their reactivity with U1 snRNP proteins. As shown in Figure 1B (lane 2), antibodies affinity purified from the λHC-1 fusion protein exclusively recognized the U1 snRNP-specific C protein.

The cDNA insert, isolated from the λHC-1 phage by digestion with EcoRI, was recloned into the SP6 bacteriophage promoter plasmid, pSP65, to yield pHC-1. To confirm the identity of the cDNA clone, mRNA specifically hybridizing to the pHC-1 cDNA insert was selected from total HeLa poly(A)$^+$ RNA and subsequently translated _in vitro_ using a rabbit reticulocyte lysate. The hybrid-selected mRNA translated _in vitro_ two major polypeptides with apparent molecular weights of about 21 kD and 11 kD, respectively, and a minor product of approximately 12.5 kD, which all were immunoprecipitable with the anti-(U1)RNP serum that was initially used for the isolation of the λHC-1 phage (Figure 2, lane 3). The 21 kD protein comigrated with the _in vitro_ synthesized C protein (Figure 2, lane 4), which on an SDS-polyacrylamide gel had a slightly greater mobility than the _in vivo_ labelled C protein (Figure 2, lane 1). This shift in mobility was estimated to correspond to approximately 1.0 kD. The fact that the 12.5 and 11 kD polypeptides were also specifically immunoprecipitated from the translation reaction directed by the hybrid-selected mRNA, implies that they originate from other in-phase AUG codons erroneously used for initiation of translation (see below) or, alternatively, represent degradation products of the C protein. In fact, they also appeared

in the immunoprecipitation of snRNP proteins synthesized in vitro by translating total poly(A)$^+$ RNA (Figure 2, lane 4). Since Northern blots containing total HeLa poly(A)$^+$ RNA revealed only one mRNA species when probed with the nick-translated cDNA insert of pHC-1 (Figure 3), the possibility that these additional translation products originate from a contaminating mRNA could be ruled out.

Restriction mapping of the cDNA insert of pHC-1 showed that it was approximately 750 bp in length. Assuming an average length of about 150 residues for the poly(A) tail of a eukaryotic mRNA, this cDNA matches well with the size of the homologous mRNA which was determined by Northern blot analysis to be about 0.9 kb (Figure 3). To determine whether the cDNA insert is indeed close to full-length, a primer extension experiment was carried out. An end-labelled 24-residue oligonucleotide primer complementary to nucleotides 64-87 as presented in Figure 4B, was hybridized to HeLa poly(A)$^+$ RNA and used as a primer for reverse transcription. The primer was extended about 60-70 nucleotides (data not shown), meaning that the cDNA insert of λHC-1 lacked at most a few bp, if any, of the 5'-untranslated end of the C protein mRNA. Also rescreening of the λgt11 expression library with a ^{32}P-labelled nick-translated 5'-terminal fragment of the λHC-1 insert did not yield recombinants with inserts longer than that of the λHC-1 phage clone.

Nucleotide sequence and complete deduced amino acid sequence of the C protein.

The nucleotide sequence of the C protein cDNA contained in clone λHC-1 was determined by the dideoxy method (21) according to the strategy outlined in Figure 4A. The sequence is shown in Figure 4B and covers 733 nucleotides. At the 3' end it contains the consensus polyadenylation signal AATAAA (29) 25 nucleotides upstream of the terminating poly(A) tract. The reading frame, which is in register with the junctional β-galactosidase sequence of λgt11 and therefore is expressed in λHC-1, extends from the first nucleotide to nucleotide 493, where a TAA stop codon is encountered. The first putative ATG initiation codon is found at position 16 and is located within a sequence segment which fits the initiation codon rules of Kozak (30). Experimental evidence that the designated ATG codon indeed encodes the initiation methionine was obtained from the in vitro translation of an SP6 RNA transcript of the pHC-1 insert. The plasmid pHC-1, containing the SP6 promoter immediately upstream of the cDNA insert, was transcribed in vitro after linearization of the circular DNA by HindIII digestion. In vitro translation of the resulting RNA yielded three major products (Figure 5, lane 2). The longest polypeptide

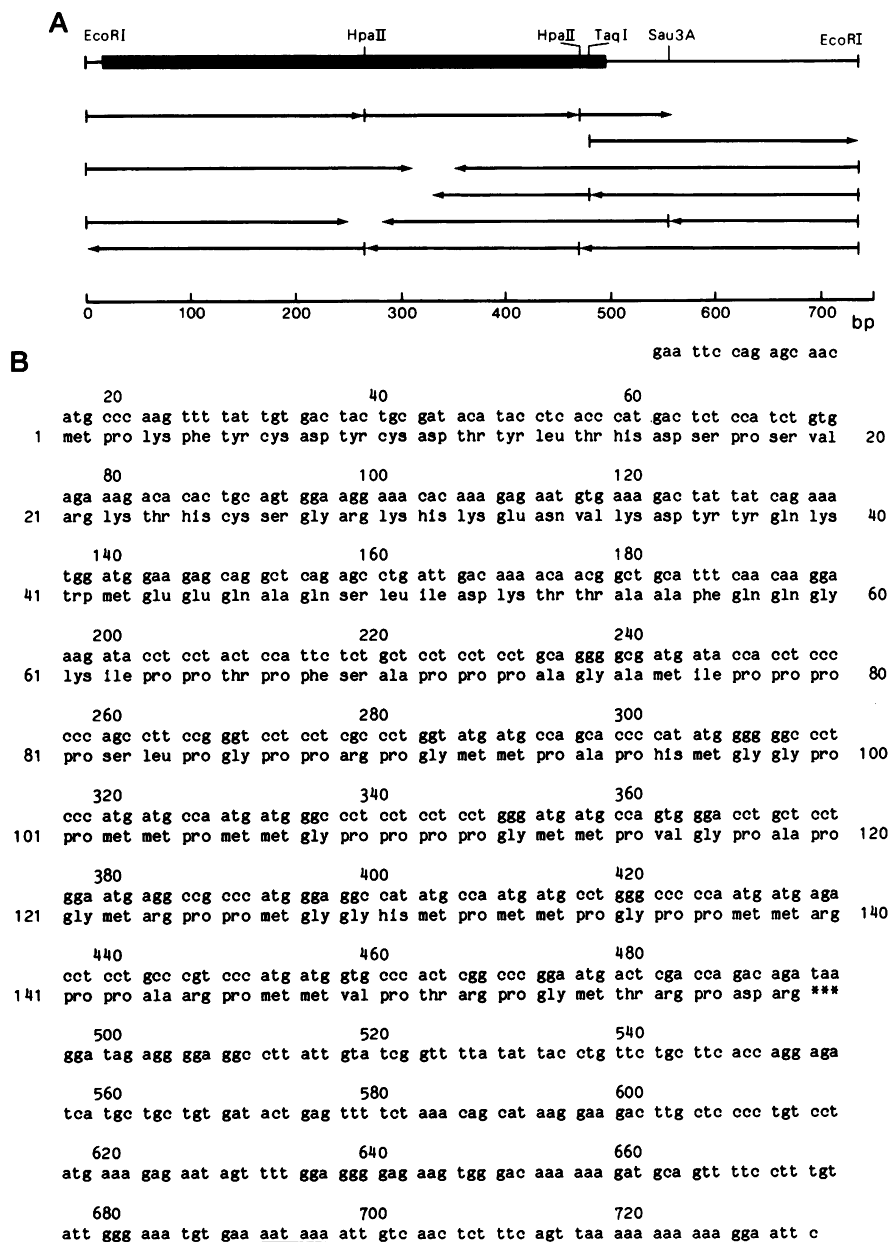

Figure 4. (A) Partial restriction map and sequencing strategy for the full-length cDNA clone pHC-1. The thick bar represents the protein coding region. **(B)** Nucleotide sequence and deduced amino acid sequence of the encoded C protein. The predicted amino acids are shown below their triplets. The polyadenylation signal is underlined and the stop codon is marked by asterisks.

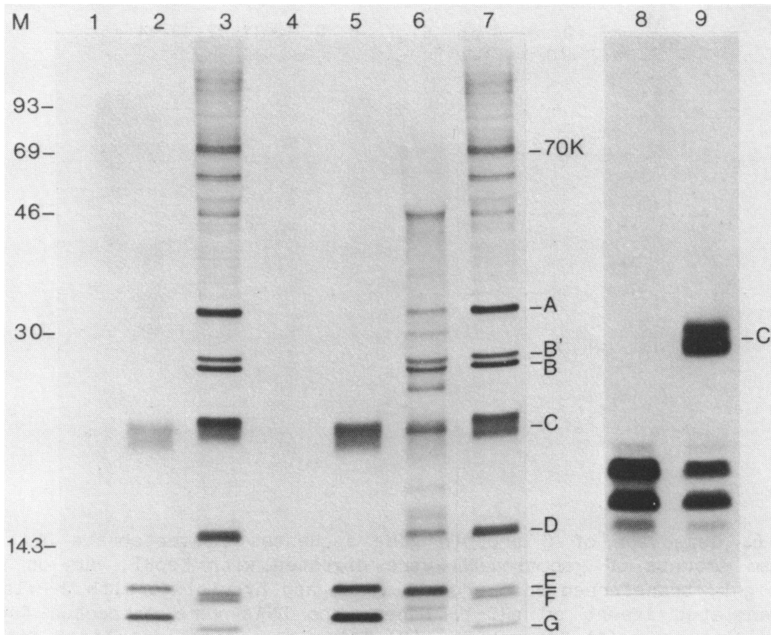

Figure 5. In vitro translation of pHC-1 encoded RNA. Immunoprecipitation of the translation products was performed with the same anti-RNP/Sm serum that was used in the hybrid selection experiment (Figure 2). **Lanes: (1)** control translation without addition of exogenous RNA; **(2)** and **(9)** translation products of pHC-1 encoded RNA; **(3)** and **(7)** immunoprecipitation of U snRNP proteins from in vivo ^{35}S-methionine labelled HeLa protein extract; **(4)** immunoprecipitation of translation products from the control translation; **(5)** immunoprecipitation of translation products encoded by RNA from pHC-1; **(6)** immunoprecipitation of translation products from total HeLa poly(A)$^+$ RNA; **(8)** translation products of RNA derived from pHC-1 after deletion of the first 87 nucleotides of the cDNA by partial PstI digestion. Molecular weights (in kD) of protein standards are indicated on the left **(M)**.

had exactly the same size as the C protein synthesized in vitro by the bona fide C protein mRNA (Figure 5, lane 6), demonstrating that the cDNA sequence presented in Figure 4B contains the entire protein coding region. As already observed in the hybrid selection experiment the in vitro translation product had a slightly smaller molecular weight as the in vivo synthesized C protein. The primary translation product could also be immunoprecipitated from the translation mixture (Figure 5, lane 5), together with the two polypeptides of lower molecular weight, indicating that these additional products originate from the same reading frame. The longest of the incomplete translation products results from incorrect initiation of translation at the ATG codon at

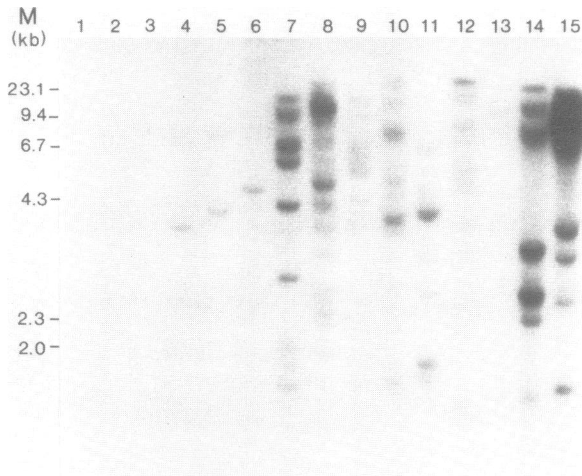

Figure 6. Detection of C protein gene sequences in vertebrate DNAs. Ten-microgram amounts of genomic DNA were digested with EcoRI, run on a 0.7% agarose gel, transferred to nitrocellulose, and hybridized with ^{32}P-labelled nick-translated insert of pHC-1. **Lanes:** The DNAs were extracted from the following species: **(1)** Drosophila; **(2)** Tilapia, a cichlid teleost fish; **(3)** Xenopus; **(4)** Varan (lizard); **(5)** chicken; **(6)** duck; **(7)** rabbit; **(8)** rat; **(9)** mouse; **(10)** hamster; **(11)** mole; **(12)** mole rat; **(13)** calf; **(14)** monkey; **(15)** man. As markers, HindIII digested λ-DNA fragments were run in parallel **(M)**.

position 139, since translation of an SP6 transcript of the pHC-1 insert from which the first 87 nucleotides had been deleted by partial PstI digestion, still yielded the two smaller polypeptides while the full-length C protein was no longer synthesized (Figure 5, lanes 8 and 9). Taken together, these results unambiguously identify the ATG codon at position 16 as the translational initiation codon.

Southern blot analysis of the gene encoding the C protein.

Southern blot analysis was performed to test different vertebrate genomes for the presence of sequences homologous to the cDNA of the human U1 snRNP-specific C protein. EcoRI-digested genomic DNAs of different species were hybridized using the ^{32}P-labelled nick-translated EcoRI insert of pHC-1 as probe. Representatives of all vertebrate classes, including fish, amphibia, reptiles, birds, and mammals, showed hybridizing bands under the moderate conditions of stringency used (see Materials and Methods). Whereas only two or three fragments in the genomic DNA of Tilapia, Xenopus, Varan, duck and chicken were detected in the original of Figure 6, multiple bands of widely different intensities cross-hybridized in the genome of mammals, with the

exception of calf genomic DNA which exhibits a simpler pattern. Genomic insect DNA sequences homologous to the cDNA of the C protein could not be detected, as no significant cross-hybridization with Drosophila DNA fragments (Figure 6, lane 1) was found. Southern blot analysis of human genomic DNA with other restriction enzymes confirms the data obtained with EcoRI (Figure 6, lane 15). One or two strong hybridizing bands in addition to a varying multiband pattern of lower intensity were detected in restriction digests produced with StuI, SphI, PvuII, PstI, KpnI, HpaI, HindIII, BglII, and BamHI (data not shown).

DISCUSSION

The proteins of the snRNP particles are of considerable interest because they are target antigens for the so-called anti-Sm and anti-RNP autoantibodies in sera from patients with connective tissue diseases (reviewed in 12) and because they participate in the splicing of pre-mRNA (8,11). We describe here the characterization of a cDNA clone for the human U1 snRNP-specific C protein. Northern blot analysis (Figure 3) and primer extension data indicate that the insert of 733 bp of this clone is close to full-length. It covers the complete coding and 3'-untranslated region of the C protein mRNA in addition to 15 nucleotides upstream of the designated initiation codon (Figure 4B). This rather short 5'-noncoding region might explain why ribosomes initiate at the first and second AUG codon when an SP6 transcript of the cDNA insert is translated in vitro, producing long and short forms of the encoded polypeptide (Figure 5). When the authentic mRNA encoding the C protein is translated in vitro (Figure 2 and Figure 5, lane 6), the same additional polypeptide is found, indicating that in this case the initiation codon also is located close to the cap and therefore is recognized less efficiently in vitro. Other examples of cellular mRNAs exhibiting this phenomenon have been described (for references see 31).

The derived C protein primary structure is composed of 159 amino acids (Figure 4B) and has a calculated molecular weight of 17,373 D. The deduced molecular weight does not agree with the 22 kD estimate for the C protein as determined from SDS-polyacrylamide gelelectrophoresis (Figures 2 and 5). However, anomalous migration of polypeptides on SDS-polyacrylamide gels is more often observed, the reason for this abnormal mobility being unknown (13,32). Recently experimental evidence has been provided by Fisher et al. (4) that the C protein is post-translationally modified in vivo. When cells were pulse-labelled with ^{35}S-methionine, a shift in mobility for de novo synthe-

sized C protein was observed within 1 hr of cold chase. SDS-polyacrylamide gel analysis suggested a modification corresponding to approximately 1.0 kD. In accordance with this finding a similar discrepancy in molecular weight was found between the C protein synthesized in vitro in a reticulocyte translation system and the in vivo labelled protein. Preliminary two-dimensional gel-electrophoresis data indicate that the overall charge of the polypeptide is not affected by the post-translational modification. The exact nature of the modification, however, remains to be elucidated. Specific oxidation of several of the numerous methionine residues of the C protein is a possibility because complete oxidation of either the in vitro or in vivo synthesized C protein by performic acid results in a retarded migration of the oxidized proteins on SDS-polyacrylamide gels (data not shown).

A second feature demonstrating the aberrant behaviour of the C protein on SDS-polyacrylamide gels is the finding that it sometimes separates into two faint bands which slightly differ in molecular weight. An example is shown in Figure 5. Since the in vitro synthesized C protein separates in a similar doublet, though with a slightly smaller apparent molecular weight, this effect cannot be due to the post-translational modification. Intrinsic disulfide bridge formation between cysteine residues could also be ruled out. Extensive heating of the C protein in the presence of β-mercaptoethanol did not affect the mobility of either of the doublet bands (data not shown).

When the nucleotide sequence of the pHC-1 insert was aligned with that of the recently reported pS2 DNA (33), which contains a partial sequence of the human C protein, differences were observed at seven positions. At position 88 a base change from C in pS2 to T in pHC-1 is observed and A to G changes are found at positions 94 and 295 (Figure 4B). These non-silent base changes could indicate that there is genomic polymorphism in the C protein at the DNA and amino acid level. Furthermore, a C residue around positions 301, 318 and 400, and a G residue around position 311 are absent in the pS2 nucleotide sequence. All these deletions are located within regions of the cDNA sequence, that are rich in G and C residues. In dideoxy sequencing analysis such regions are notorious for the fact that they give rise to compression artifacts. While the nucleotide differences only result in a switch to other amino acids, the deletions each time cause a shift to another amino acid reading frame, the latter giving rise to a completely different carboxy-terminal end for the C protein as the one reported by Yamamoto et al. (33). Comparison with a cDNA sequence encoding the Xenopus homologue for the human C protein revealed an identical open reading frame for the Xenopus protein as the one

presented in Figure 4B for its human counterpart. Both proteins have exactly the same length and of the 159 amino acids of either protein only a few residues are different (I. Mattaj, personal communications).

The amino acid composition of the C protein deduced from the nucleotide sequence of the cDNA (Figure 4B) is noteworthy for its extremely high content of proline residues (25% of all amino acids). Analysis of the primary sequence reveals that most of the proline residues are clustered in a region spanning amino acids 63-159 (Figure 4B). Methionine is also enriched in this carboxy-terminal part of the C protein: out of the 97 amino acids encompassing the proline-rich region, 20 are methionine residues. Due to the absence of negatively charged amino acids, the proline-rich region has an overall positive charge. Insofar, this portion of the C protein resembles the segment of high proline content in the U1 snRNP-specific A protein (15), which also has a net positive charge and an over-representation of methionine residues. For the A protein it was suggested that the proline-rich portion might constitute an RNA binding domain, since similar regions have been identified in several single-stranded nucleic acid-binding proteins (16,34,35). However, unlike the other U1 snRNA-specific proteins 70K (13) and A (15) and the U2 snRNP-specific B" protein (36), the C protein does not contain a fragment homologous to the so-called RNP consensus sequence, a common motif of eight amino acids that seems to be diagnostic of RNA binding proteins (16,37,38). This could suggest that instead of binding directly to the snRNA molecule, the C protein is incorporated into the U1 snRNP complex by virtue of its interaction with other snRNP proteins.

The genomic blot containing DNA of a number of present-day species showed hybridizing restriction fragments in all vertebrate DNAs, indicating a conservation of the sequence of the corresponding gene. This evolutionary conservation on the DNA level was also observed for other snRNA-associated proteins (ref. 39 and data not shown). For the A protein genomic fragments hybridizing with a human cDNA probe encoding this U1 snRNA-specific protein could even be detected in the invertebrate species Drosophila. Using similar hybridization conditions, however, we were not able to detect cross-hybridizing fragments in the genome of this species when the Southern blot was probed with the C protein cDNA. In concert with this finding a protein with immunological relation to the human U1 snRNP-specific C protein has not been reported for Drosophila (40,41). However, the possibility remains that the mammalian mRNA sequence can have diverged significantly during vertebrate evolution and antigenicity of the C protein may not be conserved among invertebrate species.

As far as the gene copy number is concerned the situation appears to be different for mammals as compared to the other vertebrate classes. With the C protein probe all genomic DNAs of mammals display multiple bands. Some of these bands may result from fragments of the C protein gene, but it is more likely that multiple functional genes exist or that some of the hybridizing fragments represent pseudogenes, the presence of which is a general feature in mammalian multigene families (42,43). In calf the number of hybridizing bands was found to be smaller. In this case, however, sequence divergence might be responsible, since the intensity of the hybridizing fragments also seems to be lower. Hybridization of the C protein probe with genomic DNA of the other vertebrate classes, including fish, amphibia, reptiles and birds, gives a quite different pattern. Only a few fragments cross-hybridized in the EcoRI restriction digests, suggesting that in these species the related protein is encoded by a less complex number of genes or even by a single gene. Sequence analysis of the homologous DNA fragments in the different species will give more insight in the evolution of the C protein genes.

ACKNOWLEDGEMENTS

We would like to thank Hanny Kremer (University of Nijmegen) and Iain Mattaj (EMBL, Heidelberg) for kindly providing Drosophila and Xenopus genomic DNA, respectively. 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 Scientific Research (NWO) and the Netherlands League against Rheumatism.

*To whom correspondence 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. Mattaj, I.W. (1988) In Birnstiel, M. (ed), Structure and Function of Major and Minor small nuclear RNAs. Springer-Verlag, in press.
3. Bringmann, P. and Lührmann, R. (1986) EMBO J. **5**, 3509-3516.
4. Fisher, D.E., Conner, G.E., Reeves, N.H., Wisniewolski, R. and Blobel, G. (1985) Cell **42**, 751-758.
5. Mattaj, I. (1986) Cell **46**, 905-911.
6. Pettersson, I., Hinterberger, M., Mimori, T. Gottlieb, E. and Steitz, J. A. (1984) J. Biol. Chem. **259**, 5907-5914.
7. Habets, W., Hoet, M., Bringmann, P., Lührmann, R. and van Venrooij, W. (1985) EMBO J. **4**, 1545-1550.
8. Maniatis, T. and Reed, R. (1987) Nature (London) **325**, 673-678.
9. Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) Cell **38**, 299-307.
10. Zhuang, Y. and Weiner, A. (1986) Cell **46**, 827-835.
11. Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) Cell **33**, 509-518.

12. Tan, E.M. (1982) Adv. Immunol. **33**, 167-240.
13. Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Philipson, L. (1986) EMBO J. **5**, 3209-3217.
14. Spritz, R.A., Strunk, K., Surowy, C.S., Hoch, S.O., Barton, D.E. and Francke, U. (1987) Nucl. Acids Res. **15**, 10373-10391.
15. Sillekens, P.T.G., Habets, W.J., Beijer, R.P. and van Venrooij, W.J. (1987) EMBO J. **6**, 3841-3848.
16. Adam, S.A., Nakagawa, T., Swanson, M., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. **6**, 2932-2943.
17. 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.
18. 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.
19. 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.
20. Messing, J. (1983) Methods Enzymol. **101**, 20-78.
21. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
22. Quax-Jeuken, Y., Janssen, C., Quax, W., van den Heuvel, R. and Bloemendal, H. (1984) J. Mol. Biol. **180**, 457-472.
23. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) Nature (London) **313**, 552-557.
24. Carmichael, G.G. and McMaster, G.K. (1980) Methods Enzymol. **65**, 380-391.
25. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA **77**, 5201-5205.
26. van der Putten, H., Terwindt, E., Berns, A. and Jaenisch, R. (1979) Cell **18**, 109-116.
27. Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA **81**, 1991-1995.
28. Smith, D.E. and Fisher, P.A. (1984) J. Cell Biol. **99**, 20-28.
29. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature **263**, 211-214.
30. Kozak, M. (1986) Cell **44**, 283-292.
31. Kozak, M. (1988) Nucl. Acids Res. **15**, 8125-8148.
32. Earnshaw, W.C., Sullivan, K.F., Machlin, P.S., Cooke, C.A., Kaiser, D.A., Pollard, T.D., Rothfield, N.F. and Cleveland, D.W. (1987) J. Cell Biol. **104**, 817-829.
33. Yamamoto, K., Miura, H., Moroi, Y., Yoshinoya, S., Goto, M., Nishioka, K. and Miyamoto, T. (1988) J. Immunol. **140**, 311-317.
34. Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H. and Delius, H. (1980) Proc. Natl. Acad. Sci. USA **77**, 6376-6380.
35. Kruijer, W., van Schaik, F.M.A. and Sussenbach, J.S. (1981) Nucl. Acids Res. **9**, 4439-4457.
36. Habets, W.J., Sillekens, P.T.G., Hoet, H.M., 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.
37. Swanson, M.S., Nakagawa, T.Y., Le Van, K. and Dreyfuss, G. (1987) Mol. Cell. Biol. **7**, 1730-1731.
38. Dreyfuss, G., Swanson, M.S. and Pinol-Roma, S. (1988) TIBS **13**, 86-91.
39. Wieben, E.D., Rohleder, A.M., Nenninger, J.M. and Pederson, T. (1985) Proc. Natl. Acad. Sci. USA **82**, 7914-7918.
40. Wieben, E.D. and Pederson, T. (1982) Mol. Cell. Biol. **2**, 914-920.
41. Wooly, J.C., Cone, R.D., Tartof, D. and Chung, S.-Y. (1982) Proc. Natl. Acad. Sci. USA **79**, 6762-6766.
42. Denison, R.A., Van Arsdell, S.W., Bernstein, L.B. and Weiner, A.M. (1981) Proc. Natl. Acad. Sci. USA **78**, 810-814.
43. Lemischka, I. and Sharp, P.A. (1982) Nature (London) **300**, 330-335.