Volume 15 Number 24 1987 Nucleic Acids Research

The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding

Richard A.Spritz*, Kathleen Strunk, Carol S.Surowy, Sallie O.Hoch[1], David E.Barton[2] and Uta Francke[2]

Department of Medical Genetics, 309 Laboratory of Genetics, University of Wisconsin, Madison, WI 53706, [1]The Agouron Institute, 505 Coast Blvd South, La Jolla, CA 92037 and [2]Department of Human Genetics, New Haven, Yale University School of Medicine, 333 Cedar St., CT 06510, USA

Received August 25, 1987; Revised and Accepted October 11, 1987 Accession no. Y00886

ABSTRACT

We have isolated and sequenced cDNA clones encoding the human U1-70K snRNP protein, and have mapped this locus (U1AP1) to human chromosome 19. The gene produces two size classes of RNA, a major 1.7-kb RNA and a minor 3.9-kb RNA. The 1.7-kb species appears to be the functional mRNA; the role of the 3.9-kb RNA, which extends further in the 5' direction, is unclear. The actual size of the hU1-70K protein is probably 52 kd, rather than 70 kd. The protein contains three regions similar to known nucleic acid-binding proteins, and it binds RNA in an *in vitro* assay. Comparison of the cDNA sequences indicates that there are multiple subclasses of mRNA that arise by alternative pre-mRNA splicing of at least four alternative exon segments. This suggests that multiple forms of the hU1-70K protein may exist, possibly with different functions *in vivo*.

INTRODUCTION

During processing of eukaryotic nuclear mRNA precursors (pre-mRNAs) to mRNAs, the intervening sequences (IVSs) are excised and the exons ligated. Pre-mRNA splicing takes place in association with a large nuclear ribonucleoprotein (RNP) complex, the spliceosome (reviewed in 1-4). The spliceosome contains several different small nuclear ribonucleoprotein complexes (snRNPs), each containing one or two discrete small nuclear RNAs ("U" snRNAs), several "core" proteins, and usually one or more proteins specific to the individual snRNP species. The U1 snRNP binds to 5' splice sites via base-pairing between the 5' end of U1 snRNA and the 5' splice site consensus sequence, and also interacts with U2 snRNP during splicing. Human U1 snRNP contains at least three specific proteins: the 70-kd (hU1-70K), 34-kd (hU1-A), and 22-kd (hU1-C) snRNP proteins (reviewed in 5). The hU1-70K snRNP protein is the major antigen recognized by anti-(U1)RNP sera from patients with autoimmune diseases.

We describe here the isolation and analysis of cDNA clones encoding the human U1-70K snRNP protein. Our results complement those of Theissen *et al* (6), who have independently reported the isolation and structure of cDNA clones encoding this protein. We show that the hU1-70K protein is encoded by a single gene on chromosome 19. We observe two classes of hybridizing mRNA, sized approximately 1.7 and 3.9 kb. The 1.7-kb mRNA, which is colinear with the 3' half of the 3.9-kb mRNA, appears to be the true mRNA, encoding a protein of actual molecular weight 52 kd, rather than 70 kd. The hU1-70K snRNP protein contains three regions of similarity to known nucleic acid-binding proteins, and β-galactosidase/hU1-70K fusion proteins bind RNA in an *in vitro* assay. These data suggest that the hU1-70K snRNP protein binds to U1 snRNA or pre-mRNA directly *in vivo*. Finally, characterization of different cDNA and genomic clones demonstrates that multiple mRNA subspecies result from alternative splicing of the hU1-70K pre-mRNA. This suggests that multiple forms of the hU1-70K protein may exist *in vivo*, perhaps with different properties and functions.

MATERIALS AND METHODS

Isolation of hU1-70K snRNP Protein cDNA Clones

Human λgt11 cDNA libraries were screened using a mouse monoclonal antibody against the mouse U1-70K snRNP protein (7) as described (8). EcoR I cDNA fragments from immunologically detected recombinants were tested for cross-hybridization by Southern blotting (9). Subsequently, other cDNA and genomic clones were isolated by hybridization.

Nucleotide Sequence Analyses

EcoR I cDNA fragments were subcloned in M13mp18 and DNA sequences were determined by the method of Sanger *et al.* (10).

Southern Blot Analyses

For mapping the human gene encoding the hU1-70K snRNP protein, DNA (3 μg per lane) from normal human 721 fibroblasts was first digested with various restriction enzymes and analyzed essentially according to Southern (9) using the complete 1.7-kb EcoR I cDNA fragment of clone FL1.7 as probe. For more detailed mapping, replicate lanes of Pst I digests were probed with a 309-

bp EcoR I-Xho I fragment (containing the 5' end of FL1.7) and the 527-bp EcoR I cDNA fragment of RNP3 (the mRNA 3' terminus). For chromosomal mapping, DNA (10 μg per lane) from 17 rodent x human somatic cell hybrid lines (11) and controls was digested with Hind III and hybridized to the complete FL1.7 probe.

RNA Analyses

Total and poly(A)$^+$ cytoplasmic RNA was prepared from cultured human cells [HepG2 (hepatoma; 12), FEM XII (melanoma), 880 (lymphoblastoid), L144 (lung fibroblast), and HeLa (cervical adenocarcinoma)] and analyzed by Northern blot hybridization (13) using the complete FL1.7 probe.

In Vitro RNA-Binding Analyses

λgt11 recombinant phage containing hU1-70K protein partial cDNAs were plated in top agar containing 5 mM IPTG, plaques were harvested, and the proteins were denatured in SDS and electrophoresed in 7.5% SDS-polyacrylamide gels as described (14). Proteins were electrophoretically transferred to nitrocellulose (15) and the β-galactosidase/hU1-70K fusion proteins were detected immunologically (8) using the mouse anti-U1-70K protein monoclonal antibody (7). The amount of each specific fusion protein applied to gels was normalized by prior immunodetection using an anti-β-galactosidase monoclonal antibody (Promega) and scanning densitometry. To assay RNA-binding, the nitrocellulose-bound proteins were "decorated" (16,17) with radiolabeled SP6 *in vitro* transcripts and autoradiographed. Transcripts tested included human U1 snRNA (from pSP64U1c+), capped poly(A)$^+$ human β-globin pre-mRNA (from pSP6β), capped poly(A)$^+$ human β-globin mRNA (from pSP6βc), and a transcript of prokaryotic plasmid (pSP64) sequences.

RESULTS

Isolation of human U1-70K snRNP protein cDNA clones

Several human cDNA libraries constructed in the expression vector λgt11 (8) were screened immunologically using a mouse monoclonal antibody to the mouse U1-70K snRNP protein (7). Screening of approximately 7 million independent recombinant phage yielded five independent cross-hybridizing recombinants (RNP3, liver; RNP6, placenta; RNP8, HeLa cell; RNP11 and RNP12,

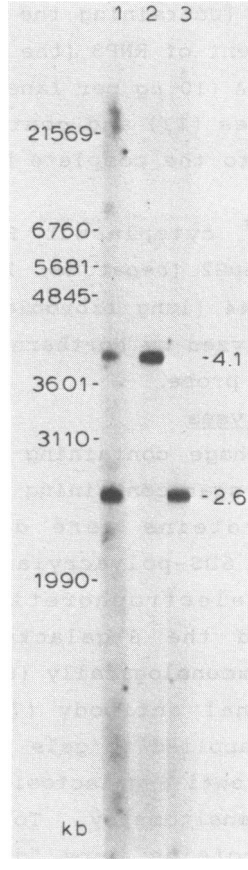


Figure 1. Southern hybridization analyses of the hU1-70K snRNP protein gene. Human DNA was digested with Pst I and hybridized to cDNA probes. Lanes: 1, 1.7-kb FL1.7 cDNA probe; 2, 309-bp EcoR I-Xho I probe (5' end of FL1.7); 3, 527-bp RNP3 cDNA probe (3' end of mRNA).

MOLT4 cell). Subsequently, FL1.7 was isolated from an HL60 cell cDNA library by hybridization to RNP6, and RNPH1, RNPH2, and RNPW1 were isolated from human liver cDNA libraries by hybridization to a 309-bp EcoR I-Xho I fragment containing the 5' end of FL1.7.

Chromosomal mapping of the human U1-70K protein gene

In Southern hybridization analyses of human 721 fibroblast DNA, the complete 1.7-kb FL1.7 probe detected two Pst I fragments, sized 4.1 and 2.6 kb (Figure 1, lane 1). Only the

TABLE I. Correlation of human FL1.7 sequences with human chromosomes in rodent x human somatic cell hybrids.

	Human Chromosomes																						
Hybridization	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X
+/+	4	1	3	5	2	5	2	5	3	1	3	4	4	7	6	5	3	6	9	4	5	7	2
-/-	7	7	2	5	5	5	5	5	6	5	5	5	5	2	4	6	8	4	8	5	4	4	1
+/-	5	8	5	3	7	4	6	3	5	8	4	5	4	2	3	4	6	3	0	5	2	1	1
-/+	1	1	6	1	2	3	2	3	2	3	3	2	3	3	4	1	0	3	0	3	4	4	2
Discordant hybrids	6	9	11	4	9	7	8	6	7	11	7	7	7	5	7	5	6	6	0	8	6	5	3
Informative hybrids	17	17	16	14	16	17	15	16	16	17	15	16	16	14	17	16	17	16	17	17	15	16	6

The numbers of hybrids showing concordant (+/+ or -/-) and discordant (+/- or -/+) segregation with FL1.7 sequences are given for each chromosome. Data on rearranged chromosomes or chromosomes present in fewer than 10% of cells were excluded.

4.1-kb Pst I fragment was detected when a 309-bp EcoR I-Xho I fragment containing only the 5' end of the FL1.7 cDNA was used as probe (lane 2). In contrast, only the 2.6-kb Pst I fragment was detected when the 527-bp RNP3 EcoR I fragment (containing only the 3' end of the mRNA) was used as probe (lane 3). Analysis of genomic clones (unpublished data) indicates that Pst I fragments corresponding to central portions of the cDNA are quite small, and are apparently not detected in this experiment because they bind inefficiently to nitrocellulose and contain only small hybridizing exon segments. These results thus indicate that the 4.1- and 2.6-kb Pst I fragments derive from a single gene. In addition, the complete 1.7-kb FL1.7 probe detected single large fragments in both BamH I and EcoR V digests of human DNA (data not shown), also suggesting that this is a single-copy gene.

To determine on which human chromosome the hU1-70K protein gene is located, we performed Southern hybridization analyses of DNA from 17 rodent x human somatic cell hybrid lines, using the complete FL1.7 probe. The results of this experiment are consistent with localization of this gene on human chromosome 19 (Table I). As shown in Figure 2, the complete FL1.7 probe detected four fragments, sized 12.5, 8.4, 7.1, and 3.6 kb, in

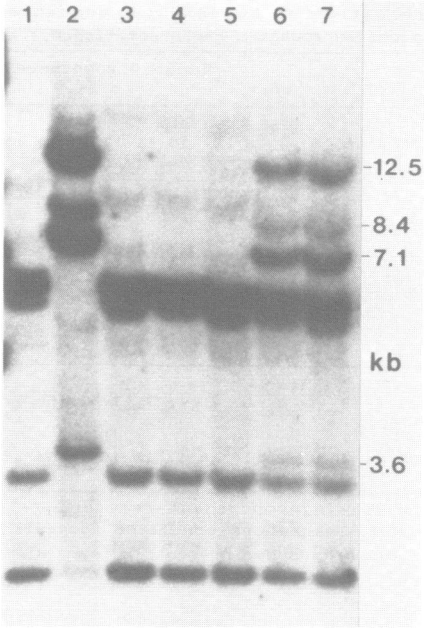


Figure 2. Southern hybridization of human x Chinese hamster hybrid cell lines. DNAs were cleaved with Hind III and hybridized to the complete FL1.7 probe. DNAs from a total of 17 hybrid cell lines were analyzed. Lanes: 1, Chinese hamster 380-6 DNA; 2, human DNA; 3-5, DNAs from human x Chinese hamster hybrid cell lines lacking human chromosome 19; 6-7, DNAs from human x Chinese hamster hybrid cell lines retaining human chromosome 19.

Hind III digests of normal human DNA (lane 2) and in DNAs of all hybrid cells that retained human chromosome 19 (see for examples Figure 2, lanes 6 and 7). These fragments were not detected in Hind III digests of DNAs from hybrid cells that lacked human chromosome 19 (see for examples Fig. 2, lanes 3-5). All other human chromosomes were excluded by at least five discordant hybrids. Four other fragments, sized 5.8, 5.4, 3.4, and 2.5 kb, were derived from the hamster genome, as they were detected in Hind III digests of Chinese hamster DNA (lane 1). Weakly hybridizing fragments in both human (lane 2) and hamster (lane 1) DNAs disappeared after more stringent washing of the filters (data not shown), and probably represent related sequences in the human and rodent genomes. These genomic mapping data (and also those described above) are consistent with restriction maps of

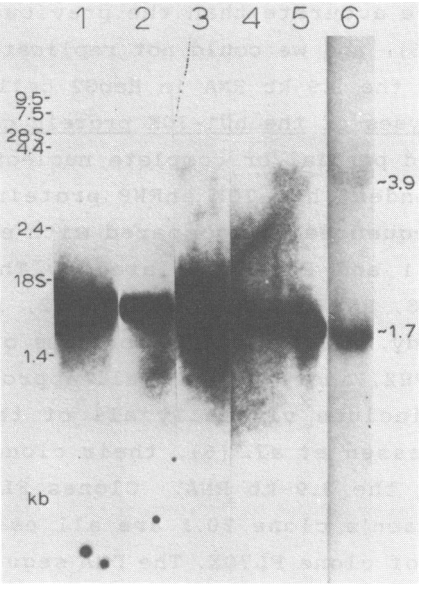


Figure 3. Blot hybridization analysis of RNAs from cultured human cells. RNAs were hybridizided to the complete FL1.7 cDNA probe. Lanes: 1, poly(A)$^+$ HepG2; 2, poly(A)$^+$ FEM XII; 3, poly(A)$^+$ 880; 4, poly(A)$^+$ L144; 5, poly(A)$^+$ HeLa; 6, total HeLa.

genomic clones of this gene (our unpublished data).

Blot hybridization analyses of hU1-70K protein RNAs

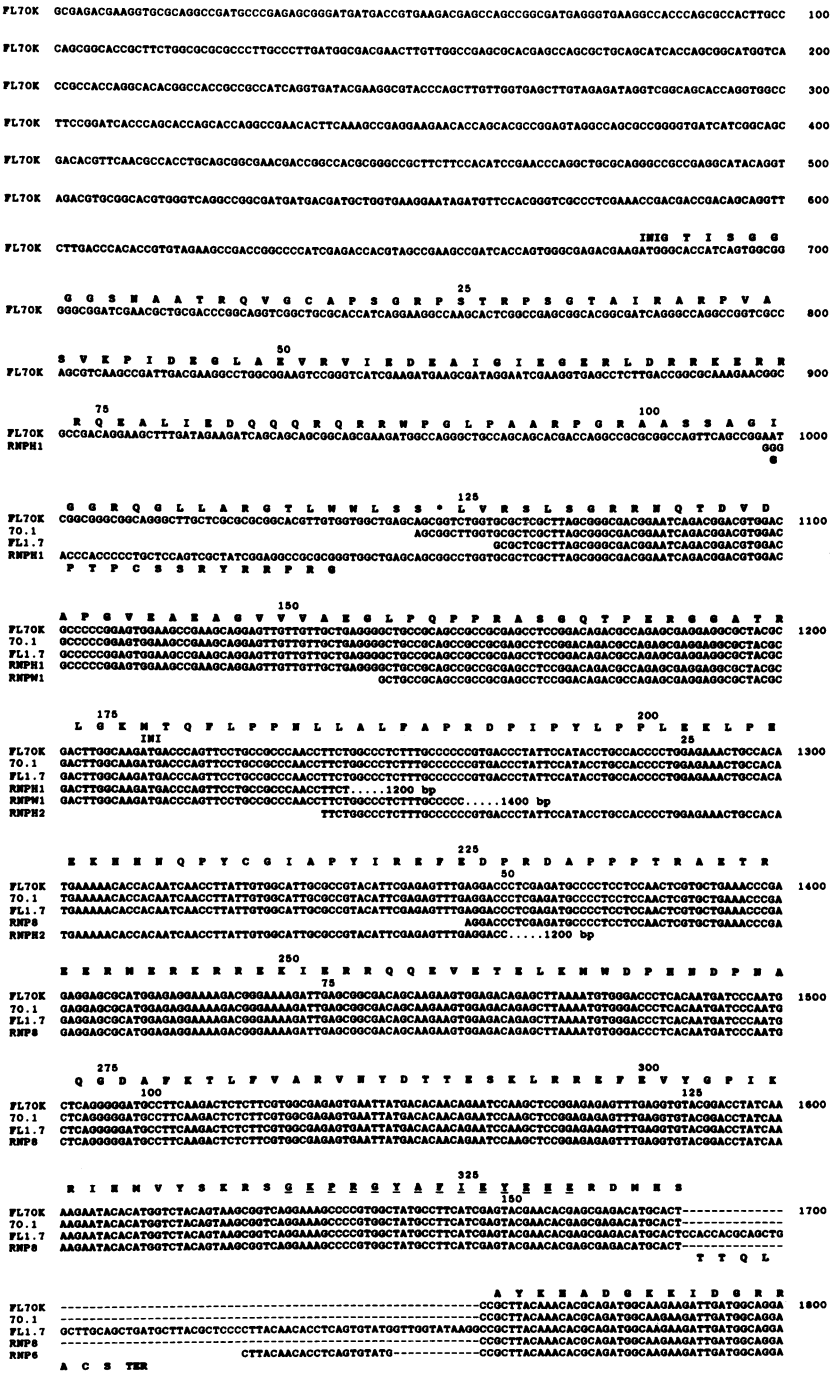
To characterize the mRNAs encoded by the human U1-70K protein gene, we conducted blot hybridization analyses of cytoplasmic poly(A)$^+$ RNAs from a number of cultured human cell lines, using the complete FL1.7 probe. As shown in Figure 3, we detected two size classes of hybridizing RNA: a major one of approximately 1.7 kb and a minor one of approximately 3.9 kb. As discussed below, the 1.7-kb RNA is apparently the true mRNA encoding the hU1-70K protein. HepG2 hepatoma (lane 1), 880 fibroblast (lane 3), L144 lymphoblastoid (lane 4), and HeLa (lane 5) cells contained both the major and minor RNA species. However, we were unable to detect the 3.9-kb RNA in FEM XII melanoma cells (lane 2). The proportions of the 1.7-kb and 3.9-kb RNAs were similar in cytoplasmic (lane 5) and total cellular (lane 6) poly(A)$^+$ RNA from HeLa cells. The present sizing of the

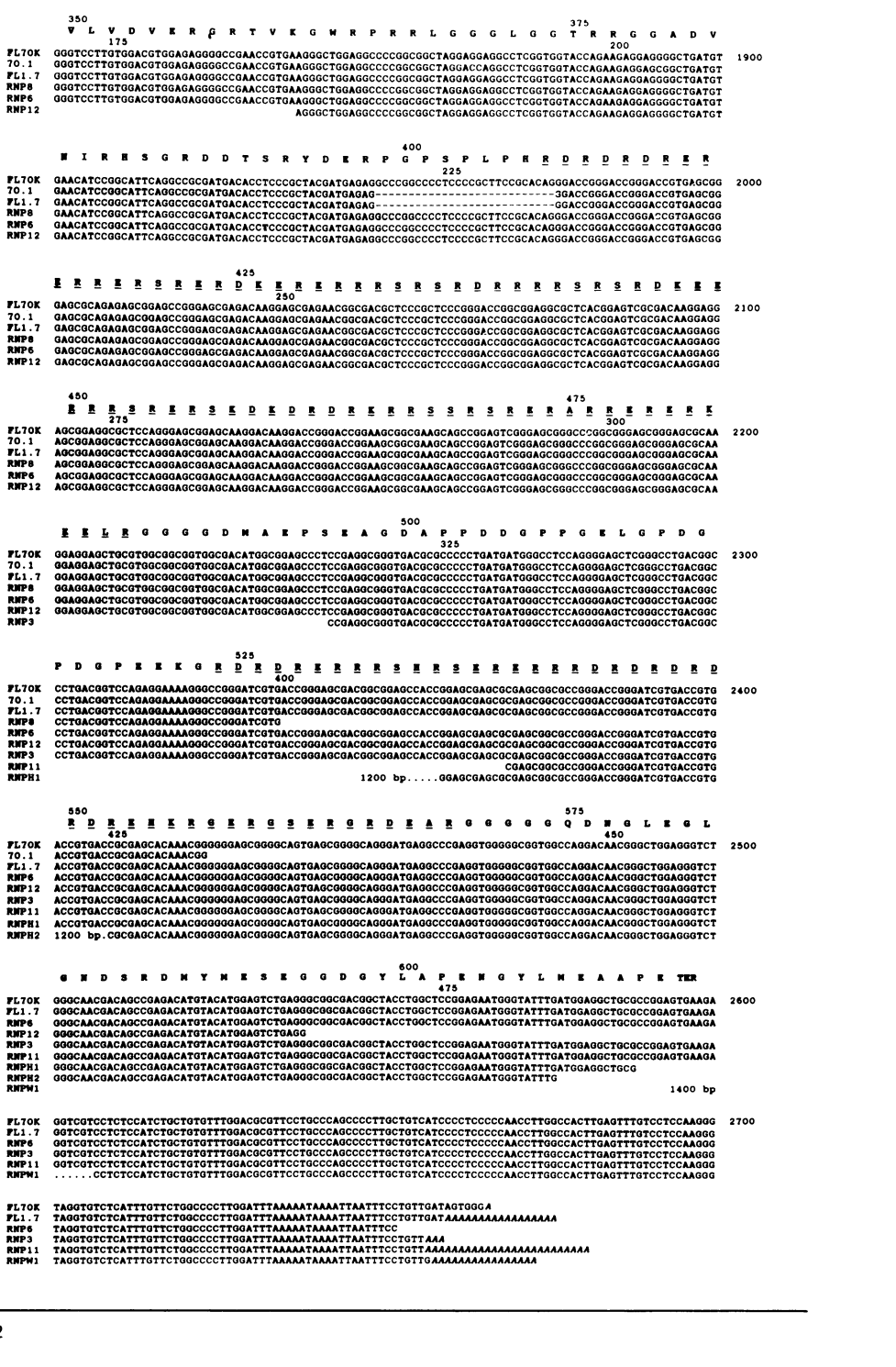
3.9-kb RNA is more accurate than the previous 3-kb estimate of Thiessen *et al.* (6), and we could not replicate their observation of virtually only the 3.9-kb RNA in HepG2 cells.

DNA sequence analyses of the hU1-70K protein cDNA clones

We determined partial or complete nucleotide sequences of a number of independent hU1-70K snRNP protein cDNA clones. In Figure 4 these sequences are compared with each other and also with clones 70.1 and FL70K isolated by Thiessen *et al.* (6). Clones RNP3, RNP6, RNP8, RNP11, and RNP12, isolated using the monoclonal antibody, all include 3' portions of the mRNA. Clones FL1.7, RNPH1, RNPH2, and RNPW1 are all approximately 1.7 kb in size, and thus include virtually all of the 1.7-kb mRNA. As discussed by Thiessen *et al.* (6), their clone FL70K apparently includes most of the 3.9-kb RNA. Clones FL1.7, RNPH1, RNPH2, RNPW1, and Thiessen's clone 70.1 are all essentially colinear with the 3' half of clone FL70K. The DNA sequences of all of the cDNA clones are identical except for differences at nt 1055-1056 and 1863-1864, which apparently represent polymorphisms.

As shown in Figure 4, the various cDNA clones also differ by the presence or absence of four different segments ofsequence. Analysis of genomic clones encoding the human U1-70K snRNP protein shows that this gene, which is greater than 30 kb in size, is interrupted by numerous IVSs (unpublished data), and that the observed differences result from alternative pre-mRNA splicing (Figure 5). First, the first 41 bases of RNPH1 are not represented in FL70K, possibly resulting from an alternative splice in this region. Second, FL1.7 and probably RNP6 differ from FL70K, 70.1, and RNP8 by the inclusion of all or part of a 60-bp alternative exon (nt 1687-1756) between codons 335 and 336. This exon contains an in-frame termination codon; therefore, mRNAs containing it cannot encode a 70-kd protein. This segment exists as a discrete exon in genomic DNA. Third, FL1.7 also contains an additional 12 bases immediately downstream from the 60-bp alternative exon. This 12-bp alternative segment (nt 1757-1768) begins with the sequence GTTGGT (similar to the 5' splice site consensus sequence), and in genomic DNA this segment occurs immediately 3' to the 60-bp alternative exon discussed above (unpublished data). Thus, two different alternative 5' splice





sites can be used when the downstream IVS is excised. Alternatively, in other transcripts the 5' splice site of the upstream IVS is ligated to the 3' splice site of the downstream intervening sequence, completely eliminating this exon. Fourth, a 27-bp alternative exon segment (nt 1949-1964; codons 399-407) is included in FL70K, RNP6, RNP8, and RNP12, but is absent from 70.1 and FL1.7. This segment consists of a pyrimidine-rich tract followed by an AG (the mammalian 3' splice site consensus sequence), and sequence analysis of genomic clones demonstrates that inclusion of this segment results from use of the upstream of two alternative 3' splice sites of an 82- versus 109-base IVS at this site.

There is also considerable heterogeneity of the sites of mRNA polyadenylylation in the various cDNA clones. As shown in Figure 4, at least four poly(A) addition sites are utilized within a 10-base region downstream from a single AAUAAA sequence near the mRNA 3' termini.

Immunologic mapping of the hU1-70K snRNP protein

Mapping studies of immunoreactive sites in the hU1-70K snRNP protein using two different human anti-(U1)RNP lupus sera were inconclusive (data not shown). However, we were able to map the epitopes that react with the monoclonal antibody to the mouse U1-70K protein. Clones RNP3, RNP6, RNP8, RNP11, and RNP12 were all isolated using the mouse monoclonal antibody to the U1-70K snRNP protein, and immunoblots of proteins produced in *E. coli* infected by these phage all demonstrated a single immunoreactive fusion protein (Figure 6B). However, as shown in Figure 6A, clones RNP8 and RNP11 do not overlap. Therefore, this monoclonal antibody must recognize at least two epitopes in the hU1-70K protein. To locate these epitopes, we compared the amino

Figure 4. Sequences of hU1-70K snRNP protein cDNA clones. DNA sequences and deduced protein sequences are indicated. 100 bases are presented per line; nucleotides are numbered at right. Sequences of FL70K and 70.1 and proposed amino acid sequence are from Thiessen *et al.* (6). Numbers above the amino acid sequence refer to the 3.9-kb RNA; those below the amino acid sequence refer to the 1.7-kb RNA. An asterisk indicates a multiply polymorphic site in the putative amino acid sequence. Absent regions are indicated by dashes. The "RNP" consensus sequence (modified from refs. 22,31) and the arginine-rich regions PT1 and PT2 are underlined. Italics indicate the poly(A) tails.

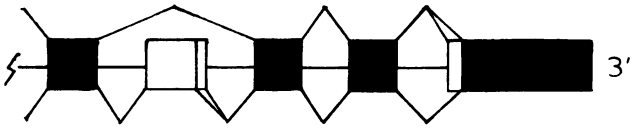


Figure 5. Alternative splicing patterns of the hU1-70K snRNP protein pre-mRNA. The three patterns of alternative splicing that we have observed involving the 3' portion of the mRNA are indicated. An additional included/excluded exon segment near the pre-mRNA 5' terminus may account for the difference between the 3.9-kb and 1.7-kb RNAs. Filled boxes, invariant exons; open boxes, alternative exon segments. Splices are indicated. Exons and IVSs are not to scale.

acid sequences of RNP8 and RNP11. The only identical segments of 6 or more amino acids encoded by these clones are tracts of alternating arginine and aspartic acid residues at amino acids 407-413 (RDRDRDR) and 542-552 (RDRDRDRDRDR). Shorter arginine-aspartic acid tracts also occur at amino acids 461-46 (DRDR) and 524-528 (RDRDR). Incubation of the monoclonal antibody with a synthetic oligopeptide containing this "RD" sequence motif (RDRDRDRDRDR) abolished binding of the antibody to the rabbit U1-70K snRNP protein (data not shown), confirming the identification of the "RD" motif as the antigenic epitope. The RD sequence motif, which is probably highly charged and on the protein surface, is presumably quite immunoreactive, as five of 13 independent mouse monoclonal antibodies against the mouse U1-70K snRNP protein recognized this epitope (data not shown).

RNA-binding by the hU1-70K snRNP protein

To determine whether the hU1-70K snRNP protein binds RNA, we first tested the ability of human snRNP proteins to bind RNAs *in vitro*. Intact snRNPs were purified from HeLa cells by affinity binding to a anti-"Sm" IgG-Sepharose column (18), and the proteins were electrophoresed in 10% SDS-polyacrylamide gels, transferred to nitrocellulose (15), and "decorated" with various radiolabeled RNAs transcribed *in vitro* (16,17). We found that the hU1-70K, hU1-A, hU-B, hU-B', hU1-C, hU-D, and hU-D' snRNP proteins (but not control proteins) bind RNA quite strongly (data not shown). However, the type of RNA bound in this *in vitro* assay was not specific; these snRNP proteins all bound U1 snRNA, β-

globin pre-mRNA, β-globin mRNA, and a transcript of plasmid sequences, although they did not bind single-stranded DNA.

To begin to localize the RNA-binding site(s) in the hU1-70K snRNP protein, we assayed RNA-binding by λgt11 β-galactosidase/hU1-70K fusion proteins produced in *E. coli*. As shown in Figure 6A, fusion proteins produced by the five original RNP recombinant phages contain different portions of the hU1-70K polypeptide. The RNP3 and RNP11 proteins contain only the carboxyl segment, including the PT2 arginine-rich region. The RNP6 and RNP12 proteins contain both the PT1 and PT2 arginine-rich regions. The RNP8 protein contains the RNP consensus (plus upstream sequences) and the PT1 region. As shown in Figure 6C, the RNP6, RNP8, and RNP12 fusion proteins bound U1 snRNA *in vitro*, but the RNP3 and RNP11 fusion proteins did not. However, the type of RNA bound by the fusion proteins was not specific *in vitro*; the RNP6, RNP8, and RNP12 proteins also bound pre-mRNA, mRNA, and the plasmid transcript (data not shown).

DISCUSSION

We have isolated cDNA clones encoding portions of the human U1-70K snRNP protein. This protein is encoded by a single gene located on human chromosome 19, with no detectable pseudogenes. We propose the designation U1AP1 for this locus. No other genes encoding nucleic acid-binding proteins or proteins involved in RNA metabolism have been previously mapped to chromosome 19 (19).

The human U1-70K snRNP protein cDNAs hybridize to two poly(A)$^+$ RNAs, sized approximately 1.7- and 3.9-kb, from cultured human cells. The 1.7-kb mRNA is at least 40-fold more abundant than the 3.9-kb species in all human cells assayed. The 1.7-kb mRNA is colinear with the 3' half of the 3.9-kb RNA. Thus, the two RNAs might be transcribed from alternative promoters, they might be derived from a common pre-mRNA by alternative splicing, the 1.7-kb RNA might be a stable degradation product of the 3.9-kb RNA. or the 3.9-kb RNA might even be a splicing intermediate that is further processed to the 1.7-kb RNA.

The hU1-70K protein migrates as a 70-kd polypeptide in standard SDS-polyacrylamide gel electrophoresis systems. A 1.7-

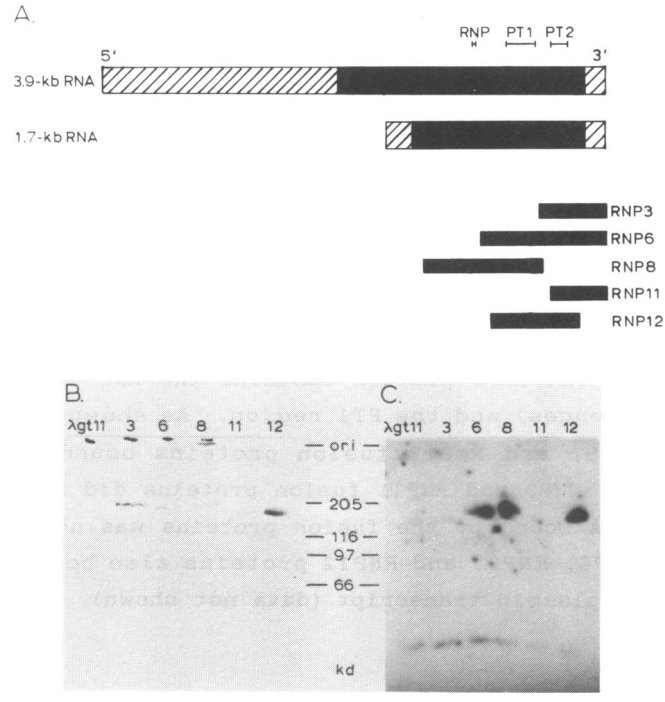


Figure 6. RNA-binding by β-galactosidase/hU1-70K snRNP fusion proteins *in vitro*. A. Structures of the hU1-70K protein RNAs and cDNAs. The 3.9-kb and 1.7-kb RNAs are illustrated. The protein coding regions are filled and the 5' and 3' untranslated regions are hatched. The RNP, PT1, and PT2 regions are indicated. Solid bars denote segments included in the various RNP cDNA clones. B. Detection of the β-galactosidase/hU1-70K fusion proteins by immunoblotting with the mouse monoclonal antibody to the mouse U1-70K snRNP protein. Lanes 3, 6, 8, 11, and 12 denote lysates of proteins produced by *E. coli* infected with the respective RNP λgt11 recombinant phage. Molecular size standards are indicated. C. Binding of synthetic U1 snRNA by the fusion proteins. Lanes are as in B.

kb mRNA cannot encode a 70-kd protein; therefore, Thiessen and coworkers (6) suggested that the 3.9-kb minor RNA (represented in part by FL70K) is the full-length hU1-70K protein mRNA, encoding a 613-amino acid 70-kd polypeptide. However, several points now suggest that the hU1-70K protein is actually encoded by the 1.7-kb RNA, and that the molecular size of the protein is only 52 kd. First, the U1-70K mRNA of *Xenopus laevis* is only 1.85 kb in size, and it is highly homologous to the human 1.7-kb hU1-70K RNA

from the translational initiation codon (Met 177 according to the numbers of ref. 6) through the carboxyl terminus; there is no homology 5' to Met "177" (M. Etzerodt, pers. comm.). This suggests that Met "177" is probably the translational initiation codon of the human U1-70K protein as well. Second, the hU1-70K snRNP protein seems likely to be ubiquitous; therefore, the low abundance of the 3.9-kb RNA compared to the 1.7-kb RNA, and its apparent absence from FEMXII cells (which contain abundant U1-70K protein), seem incongruous. Third, putative smaller protein(s) apparently encoded by the much more abundant 1.7-kb RNA have not been observed. These putative proteins would include the epitopes recognized by the anti-U1-70K monoclonal antibody; however, this antibody detects no smaller proteins on immunoblots of HeLa total, nuclear, or snRNP protein fractions (18; unpublished data). Together, these data suggest that the hU1-70K protein is encoded by the 1.7-kb mRNA, rather than by the 3.9-kb RNA, and that the correct molecular size of the hU1-70K protein is only 52 kd, rather than 70 kd.

If the human U1-70K protein is only 52 kd in actual molecular weight, its retarded migration on SDS-polyacrylamide gel electrophoresis might be the result of the unusual distribution of highly charged regions within this protein (see below; refs. 20-22). In addition, the hU1-70K protein is phosphorylated (23), which can also lead to aberrant electrophoretic migration (rev. in 24). The sequence of the putative 52-kd protein is in agreement with the limited amino acid sequence and overall amino acid composition of the hU1-70K protein reported by Thiessen *et al.* (6). However, it has not been possible to determine the (blocked) amino terminal sequence of the hU1-70K protein (unpublished data and R. Lührmann, pers. comm.) or to directly characterize the hU1-70K protein by *in vitro* translation (our unpublished data; 25,26).

It is likely that the native hU1-70K snRNP protein binds U1 snRNA or pre-mRNA *in vivo*, and the amino acid sequence of the hU1-70K polypeptide contains three regions with homology to known nucleic acid-binding proteins. Two of these regions, amino acids 407-486 (PT1; numbers according to ref. 6) and 524-569 (PT2), are arginine-rich, and have considerable similarity (>50%) to

protamines, chicken galline, and mammalian sperm histones (27), at least some of which bind both DNA and RNA (28). Both PT1 and PT2 end with the sequence RGGGG, and include the alternating arginine-aspartic acid tracts described above. Among other known protein sequences, "RD" repeats occur only in the *C. elegans* v-*abl* homologue (29) and the bovine myosin I heavy chain-like protein (30), and these are quite short. As noted by Thiessen *et al.* (6), a third region, from amino acids 266-362 (RNP), has homology to the yeast (31,32) and human (33) poly(A)-binding proteins, bovine (34) and rat (35) helix-destabilizing proteins, the human hnRNP C proteins (22), hamster nucleolin (36), the human A (W. van Venrooij, pers. comm.) and B'' (37) snRNP proteins, and the *Xenopus* La protein (S. Clarkson, pers. comm.). All of these nucleic acid-binding proteins include at least one copy of the so-called "RNP" consensus sequence $GlyLysSer^{Lys}_{Arg}GlyPhe^{Gly}_{Ala}PheValX^{Phe}_{Tyr}XX^{Glu}_{His}$ (modified from refs. 22,31). This sequence may bind RNA *in vitro* (38), but is apparently not required for function of the yeast poly(A)-binding protein (32).

The hU1-70K snRNP protein does bind RNA *in vitro* (16,17; present study), although without specificity. In preliminary studies to map the RNA-binding site(s), we found that fusion proteins produced by phages RNP6, RNP8, and RNP12 bind RNA *in vitro*, but those produced by phages RNP3 and RNP11 do not. Therefore, sequences from the amino terminus through amino acid 361 (including the RNP consensus) and from amino acid 527 through the carboxyl terminus (including the PT2 arginine-rich region) are not necessary for RNA-binding *in vitro*. Furthermore, the latter region is also not sufficient to bind RNA *in vitro*. We cannot, however, conclude that sequences from amino acids 362-526 are either necessary or sufficient for RNA-binding. It may be that sequences upstream from this region are sufficient, or that sequences either upstream or downstream are necessary. RNA-binding was not specific *in vitro*, probably because the protein is denatured in this assay system. In the hU1-70K polypeptide, the arginine-rich regions PT1 and PT1 are positively charged, whereas the segments between the RNP consensus and PT1, between PT1 and PT2, and from PT2 to the carboxyl terminus are negatively charged. The RNP consensus, PT1, and PT2 might thus be charge-

associated in the native hU1-70K snRNP protein, comprising a complex RNA-binding site specific for U1 snRNA, pre-mRNA, or mRNA *in vivo*.

Finally, we also observed heterogeneity among hU1-70K protein cDNAs (Figure 4) resulting from alternative patterns of splicing of the hU1-70K protein pre-mRNA. As shown in Figure 5, these include use of both alternative 5' and 3' splice sites, as well as at least one complete included/excluded exon (our unpublished data). The existence of these multiple hU1-70K protein mRNAs suggests that multiple forms of the hU1-70K protein, differing by small exon segments, may be produced *in vivo*. Furthermore, because the included/excluded exon contains an in-frame termination codon, mRNAs that include this exon would encode relatively short proteins that do not contain the arginine-rich segments PT1 and PT2 (and would thus not react with the monoclonal antibody to the U1-70K protein), but which do include the RNP consensus region. It may be, therefore, that these different hU1-70K-related polypeptides have different functions *in vivo*.

ACKNOWLEDGMENTS

We thank S. Woo, P. Henthorn, and S. Orkin for human cDNA libraries, R. Demars for L144 and 880 cells and 721 cell DNA, K. Huebener for HepG2 cells, V. Hearing for FEM XII cells, J. Dahlberg for U1 transcription plasmid pSP64U1c+, R. Lührmann for permission to reproduce the sequences of clones 70.1 and FL70K, and M. Etzerodt for unpublished information on the structure of *Xenopus laevis* U1-70K cDNAs and corrections of the published FL70K and 70.1 sequences. This work was supported by Research Grant DMB-8505902 from the N.S.F. (to R.A.S.), and N.I.H. Research Grants AI21083 (to S.O.H.) and GM26105 (to U.F.). This is paper no. 2930 from the Laboratory of Genetics, University of Wisconsin.

*To whom correspondence should be addressed

REFERENCES

1. Green, M.R. (1986) *Ann. Rev. Genet.* **20**, 671-708.
2. Sharp, P.A. (1987) *Science* **235**, 766-771.

3. Maniatis, T. and Reed, R. (1987) *Nature* (London) **325**, 673-678.
4. Padgett, R.A., Seiler, S.R. and Sharp, P.A. (1986) *Ann. Rev. Biochem.* **55**, 1119-1150.
5. Bringmann, P. and Lührmann, R. (1986) *EMBO J.* **5**, 3509-3516.
6. Thiessen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Philipson, L. (1986) *EMBO J.* **5**, 3209-3217.
7. Billings, P.B., Allen, R.W., Jensen, F.C. and Hoch, S.O. (1982) *J. Immunol.* **128**, 1176-1180.
8. Young, R.A. and Davis, R.W. (1985) In Setlow, J. and Hollaender, A. (eds), *Genetic Engineering*, Plenum, New York, Vol. 7, pp.29-41.
9. Southern, E.M. (1975) *J. Mol. Biol.* **98**, 503-517.
10. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
11. Yang-Feng, T.L., DeGennaro, L.J. and Francke, U. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8679-8683.
12. Knowles, B., Howe, C.C. and Aden, D.P. (1980) *Science* **209**, 497-499.
13. Thomas, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201-5205.
14. Saul, A. and Yeganeh, F. (1986) *Anal. Biochem.* **156**, 354-356.
15. Towbin, H., Staehelin, T. and Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354.
16. Tazi, J. Alibert, C., Temsemani, J. Revaillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986) *Cell* **47**, 755-766.
17. Gerke, V. and Steitz, J.A. (1986) *Cell* **47**, 973-984.
18. Billings, P.B. and Hoch, S.O. (1984) *J. Biol. Chem.* **259**, 12850-12856.
19. New Haven Gene Mapping Library (1987) In *Chromosome Plots*, No. 2, Howard Hughes Med. Inst., New Haven, CT.
20. Hope, I.A. and Struhl, K. (1986) *Cell* **46**, 885-894.
21. 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.
22. Swanson, M.S., Nakagawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) *Mol. Cell Biol.* **7**, 1731-1739.
23. Wooley, J.C., Zukerberg, L.R. and Chung, S.-Y. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5208-5212.
24. Allen, G. (1981) *Sequencing of Proteins and Peptides*, Elsevier/North-Holland, Amsterdam.
25. Wieben, E.D., Madore, S.J. and Pederson, T. (1983). *Proc. Natl. Acad. Sci. USA* **80**, 1217-1220.
26. Fisher, D.E., Conner, G.E., Reeves, N.H., Wisniewolski, R. and Blobel, G. (1985) *Cell* **42**, 751-758.
27. Dayhoff, M.O. (1972) *Atlas of Protein Sequence and Structure*, NBRF, Washington, DC.
28. Warrant, R.W. and Kim, S.H. (1978) *Nature* (London) **271**, 130-135.
29. Goddard, J.M., Weiland, J.J. and Capecchi, M.R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2172-2176.
30. Hoshimaru, M. and Nakanishi, S. (1987) *J. Biol. Chem.* **262**, 14625-14632.
31. Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6**, 2932-2943.
32. Sachs, A.B., Bond, M.W. and Kornberg, R.D. (1986) *Cell* **45**, 827-835.

33. Grange, T., de Sa, C.M., Oddos, J. and Pictet, R. (1987) *Nucleic Acids Res.* **15**, 4771-4787.
34. Williams, K.R., Stone, K.L., LoPresti, M.B., Merrill, B.M. and Planck, S.R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 5666-5670.
35. Cobianchi, F., Sen Gupta, D.N., Zmudzka, B.Z. and Wilson, S.H. (1986) *J. Biol. Chem.* **261**, 3521-3543.
36. Lapeyre, B., Bourbon, H., and Amalric, F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1472-1476.
37. 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.
38. Bugler, B., Bourbon, H., Lapeyre, B., Wallace, M.O., Chang, J.-H., Amalric, F. and Olson, M.O.J. (1987). *J. Biol. Chem.* **262**, 10922-10925.