Volume 16 Number 9 1988 Nucleic Acids Research

cDNA cloning of human hnRNP protein A1 reveals the existence of multiple mRNA isoforms

M.Buvoli, G.Biamonti, P.Tsoulfas, M.T.Bassi, A.Ghetti, S.Riva* and C.Morandi[1]

Istituto di Genetica Biochimica ed Evoluzionistica, CNR, Via Abbiategrasso, 207-27100 Pavia and [1]Istituto di Scienze Biologiche, Università degli Studi di Verona, Italy

Received February 2, 1988; Revised and Accepted March 24, 1988 Accession no.X06747

ABSTRACT

Protein A1 is one of the major component of mammalian ribonucleoprotein particles (hnRNP). Human protein A1 cDNA cloning and sequencing revealed the existence of at least two protein isoforms. Among the cDNAs examined, sequence differences were found both in the structural portion, leading to aminoacid changes (Tyr to Phe or Arg to Lys) and in the non translated 3'-region where two T-stretches of different length were observed. Interestingly one of the aminoacid substitutions falls into a consensus sequence common to many RNA binding proteins. Northern blot analysis of poly A^+ RNAs from five human tissues revealed two mRNA forms of 1500 and 1900 n due to alternative polyadenylation. Analysis of genomic DNA showed at least 30 A1-specific sequences,some of which correspond to processed pseudogenes. These results suggest that protein A1 is encoded by a multigene family.

INTRODUCTION

In the nucleus of eukaryotic cells the primary transcripts of RNA polymerase II (hnRNAs) are associated with a specific set of proteins to form ribonucleoprotein (hnRNP) particles (1). Considerable experimental evidence points to a role of hnRNP complexes in the post transcriptional processing of hnRNA and particularly in the splicing reaction (1-4). The hnRNP complex is one of the most abundant structures in the nucleus accounting for at least one third of the proteins of the nucleoplasm (1). The hnRNA in the nucleoplasm of mechanically disrupted nuclei sediments in sucrose gradient as a heterodispersed material between 30 and 250 S (5,6). All of the heterodispersed fast sedimenting hnRNA is converted by mild RNAse digestion into relatively homodispersed particles that sediment at 30-40 S. The proteins of the 30 S particles generally referred to as "core proteins" consist of six main polypeptides in one-dimensional polyacrylamide gel

electrophoresis. Their HeLa cell nomenclature is the following: A1 = 34 Kd; A2 = 36 Kd; B1 = 37 Kd; B2 = 38 Kd; C1 = 41 Kd; C2 = 43 Kd. Cross linking experiments in vivo and in vitro confirmed that hnRNA is actually associated with this unique set of proteins (7). However immunoprecipitation experiments with specific monoclonal antibodies suggest that at least two other polypeptides of 68 and 120 Kd are associated to the hnRNP complex (7). Proteins A and B are members of two related families of basic proteins that share common antigenic determinants (8) and extensive homologies in the primary structure (9). Proteins A1 and A2 (together with protein C1) are the most abundant components of the hnRNP complex. The ratio of A1 to A2 is about 1:1 in the particles from rapidly growing cells (1) but much less A1 is found in stationary cells (10). Thus, in spite of their striking similarities, these two proteins seem to be differently expressed. In particular A1 can be classified as a proliferation sensitive protein (10) which further stimulates the interest for its more detailed characterization. This can be obtained by cDNA cloning and sequencing, and by the study of homologous mRNAs and of genomic DNA sequences.

In a previous paper (11) we described the use of cDNA cloning to demonstrate the existence of a precursor-product relationship between hnRNP A1 and the mammalian single stranded DNA binding protein UP1. The analysis of A1 cDNA from human cells revealed the peculiar two-domains structure of this protein that was also confirmed by peptide sequencing on the purified protein (9,11). These studies also evidenced a striking evolutionary conservation of protein A1 between human and rat (12). Considering that a similar evolutionary conservation was recently reported for protein C1 (13) it is reasonable to assume that the same could be true for other hnRNP core proteins. In this paper we extend our studies and demonstrate that protein A1 has different isoforms. The analysis of A1 mRNAs also suggests a differential expression of the various isoforms in different tissues. Genomic DNA analysis clearly indicates that the A1-specific DNA sequences are members of a conserved gene family containing several pseudogenes as it was shown to be the case for protein C1 (13).

MATERIALS AND METHODS

Screening of a human genomic DNA library

A library of human liver DNA, partially digested with EcoRI and cloned in λ Ch 4A (34) was screened by plaque hybridization (35) using pRP15 cDNA probe ^{32}P labeled by the random priming method (36).

Oligonucleotide synthesis, library screening and analysis of cDNA clones

Oligonucleotide probes were synthesized on a Beckman System 1 Plus DNA Synthesizer, purified by polyacrylamide gel electrophoresis and ^{32}P end-labeled with T4 polynucleotide kinase. The screening of an SV40 transformed human fibroblast cDNA library in λ gt11 with a 20-Mer oligonucleotide, was performed according to Mason and Williams (37). Hybridization and washings were performed at 46°C.

cDNA clones were analyzed by standard procedures (38). Bluescribe M13$^+$ vector was purchased from Vector Cloning System (San Diego).

DNA sequencing and sequence analysis

DNA sequencing was performed by the chemical modification method of Maxam and Gilbert (39) except that diphenylamine formate was used instead of piperidine formate in the (A+G) reaction. Nucleotide and deduced aminoacid sequences were analyzed by the Beckman "Microgenie" computer program.

RNA preparation and Northern blot analysis

Total RNA was isolated using the guanidinium thiocyanate solubilization method of Chirgwin et al. (40) except for human fibroblast RNA that was isolated according to Morandi et al. (20). PolyA$^+$ RNA selected by chromatography over oligo dT cellulose column was electrophoresed onto a 1.5% agarose, 6% formaldehyde gel (41), transferred to nylon membrane (Hybond N, Amersham) and cross-linked by UV irradiation following the instructions of the manufacturer. Hybridization with nick-translated probes (10^6 cpm/ml) was carried out in 4xSSC, 4xDenhardt's, 25 mM Na PPi, 0.2% SDS, 125 µg/ml yeast tRNA, 50% formamide at 42°C. Hybridization with end-labelled oligonucleotides (10^6 cpm/ml) was performed in 6xSSC, 10x Denhardt's, 25 mM NaPPi, 0.1% SDS, 125 µg/ml yeast tRNA at 32°C. High stringency washing conditions with the two types of probes were respectively: 0.2xSSC, 0.1% SDS

at 65°C for 30 min and 6xSSC at 43°C for 2 min. Removal of probes for re-use of RNA blots (when required) was obtained by washing for 2 hrs at 65°C in 5 mM Tris-HCl pH8, 2 mM EDTA, 0.1 x Denhardt's. Slot blots were performed using a Minifold II apparatus (Schleicher and Schuell) according to manufacturer's instructions. RNA was denatured with 6% formaldehyde at 68°C diluted to the proper concentration with 10xSSC and loaded onto the nitrocellulose filters prewetted with 10xSSC.

Human genome DNA analysis, Southern blot, dot blot.

10 µg of HeLa cells DNA, digested with EcoRI were separated by electrophoresis on a 0.8% agarose gel, transferred to nitrocellulose paper and baked for 2 hr at 80°C. Hybridization conditions were as follows: 5x SSC, 5x Denhardt's, 25 mM NaPPi, 200 ug/ml yeast tRNA, 50% formamide; high stringency washing conditions were: 0.2 x SSC 0.1% SDS at 68°C for 30 min. Dot blot experiments were performed according to Mariani and Schimke (42) with a Minifold II (see above), apparatus. Hybridization conditions were as described above.

RESULTS

Isolation of a cDNA encoding the human hnRNP core protein A1

In a previous paper (11) we described the isolation of a cDNA fragment of 949bp encoding the last 194 aminoacids of hnRNP core protein A1 (34 Kd). This cDNA clone (pRP10), isolated from a human liver cDNA library in plasmid expression vector pEX1 (14), was the starting material for the isolation of other A1 specific DNAs. A 20 n long oligonucleotide complementary to the 5'-end of pRP10 was synthesized and used to screen a human cDNA library in λ gt11 prepared from polyA$^+$ mRNA of SV 40 transformed human fibroblasts (kindly provided by B. Wold), as described in Materials and Methods. Plaque hybridization experiments yielded three clones (pRP12, pRP13, and pRP15) with inserts longer than pRP10 (949 bp) and their cDNAs were subcloned in a plasmid vector (Bluescribe M13$^+$) for further characterization. The properties of pRP12 and pRP13 will be described in the following section. pRP15 contained the longest insert of about 1700 bp as judged from agarose

gel electrophoresis (not shown) and its size roughly corresponded to the longer A1-specific mRNA species detected in Northern blots experiments with HeLa cell polyA$^+$ mRNA (11; see ahead).

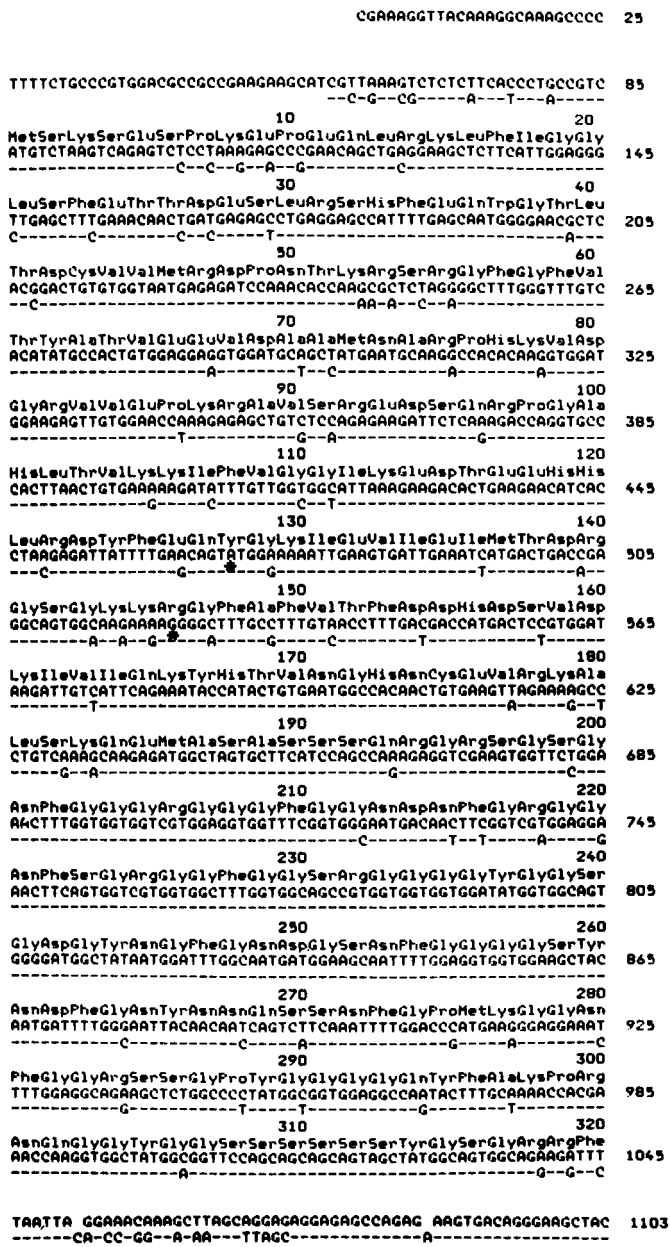
The complete sequence of pRP15, along with the restriction map and the sequencing strategy, is shown in Fig. 1. The total length of pRP15 cDNA is 1767 n, it contains an open reading frame of 960 nucleotides and two non coding regions of 85 and 722 nucleotides at the 5'-end and at the 3'-end, respectively; two possible polyadenylation sites are present.

The deduced aminoacid sequence of pRP15 cDNA corresponds exactly to that reported for both human and rodent protein A1 (9,11,12). The comparison with the nucleotide sequence of a rodent A1 cDNA (see Fig. 1) (12) evidences a remarkable conservation. This point will be further addressed to in the Discussion.

Starting from nucleotide 458 the sequence is identical to that of pRP10 cDNA (11) except that the latter ends at the first polyadenylation signal and except for a small difference in the length of the T-streches at positions 1306 and 1376 (see next sections). The 5'-end of pRP15 cDNA is probably truncated 100-200 n from the start point transcription as indicated by the fact that the corresponding mRNA is about 1900 n long (see next section). Further experiments are being carried out to determine the complete sequence of the 5'-end non coding region.

mRNA variants of hnRNP A1

The results reported in the previous sections already indicated the existence of at least two mRNAs encoding for the protein A1 but differing for the length of the 3'-non translated region. The sequence analysis of the other two cDNA clones isolated (pRP12 and pRP13) revealed variations also in the coding region. Both cDNAs appeared truncated, the first starting at nucleotide 209 (numbering as in Fig. 1) and the second at nucleotide 285. pRP12, like pRP15 had a 3'-end non coding region of 722 n, while in pRP13, like in pRP10, the non coding region was only 307 n long since it ended at the first polyadenylation site. Also in the case of these two cDNAs the T-stretches had different lengths (Fig. 1).



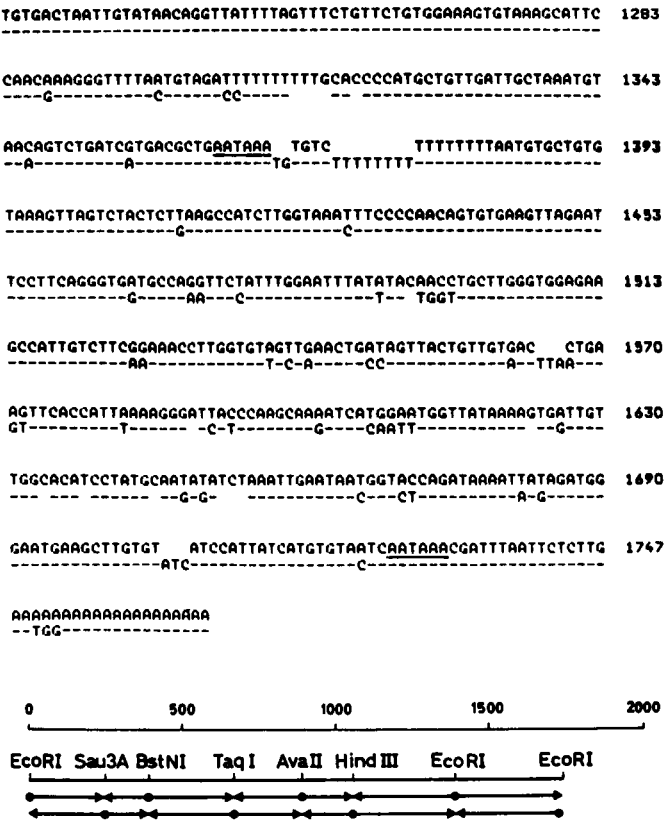


FIG. 1 - Restriction map, sequencing strategy and nucleotide sequence of pRP15 cDNA. The sequence of a rodent A1-specific cDNA clone (12) is shown underneath for the sake of comparison; dotted line: identical sequence, black bars: polyadenylation sites, asterisks: site of nucleotide change in type ß cDNA (pRP12, see Figure 2).

Surprisingly however pRP12 showed also two nucleotide substitutions at positions (468) and (522) causing two amino acid substitutions (Tyr --> Phe and Arg --> Lys) in the protein at positions 128 and 146, respectively. Thus at least two protein A1 isoforms (α = pRP15 and ß = pRP12) exist and the possibility that other variations might occur in the part of the molecule not covered by pRP12 and pRP13 cannot be ruled out. The results of these experiments are summarized in Fig. 2.

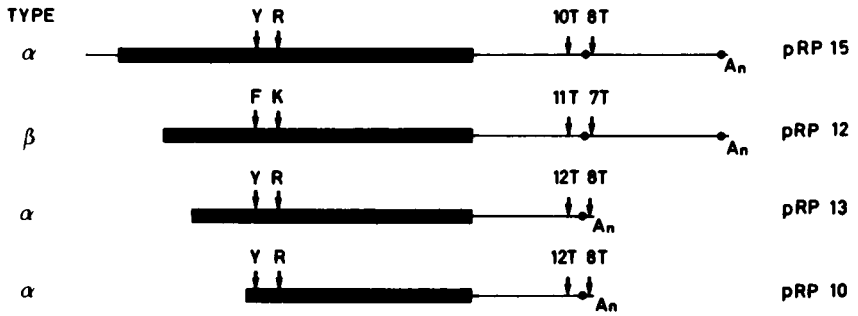


FIG. 2 - Different A1-specific cDNAs isolated from human fibroblasts (pRP12, pRP13, pRP15) and from human liver (pRP10, 11). Aminoacid substitutions corresponding to type α and ß isoforms (see text) are indicated along with the length of T-stretches in the 3'-end non coding region. The black dots represent polyadenylation sites.

The existence of A1-specific mRNAs with different polyadenylation sites was confirmed by a Northern type hybridization on polyA$^+$ RNA from HeLa cells with pRP15 cDNA as probe. As shown in Fig. 3(A) two mRNA species of about 1.5 and 1.9 Kb specifically hybridized to the probe, the smaller being more abundant. The densitometric scanning of the autoradiogram (not shown) gave a 2:1 ratio in the intensity of the short and long RNA signals.

To show that the two mRNA species actually differed in the 3'-end non coding region as observed in the cDNAs (see Figure 2), the same blot was hybridized with a probe lying between the two polyadenylation sites of pRP15; as expected this probe hybridized only to the 1.9 Kb long mRNA (Fig. 3(A), lane B).

Overall abundance of A1 specific RNA

The abundance of A1 specific mRNA was measured by a slot-blot hybridization experiment with HeLa polyA$^+$ RNA by comparing the hybridization signal of the pRP15 probe (see previous section) with that of a human ß-actin cDNA probe. The abundance of ß-actin mRNA is relatively constant in the different cell lines and has been estimated to be of the order of 500 molecules/cell (15). The result of the experiment is shown in Fig. 3 (B); total A1 mRNA is approximately as abundant as that of ß-actin and constitutes about 0.1% of the total polyA$^+$ RNA. This figure is in agreement

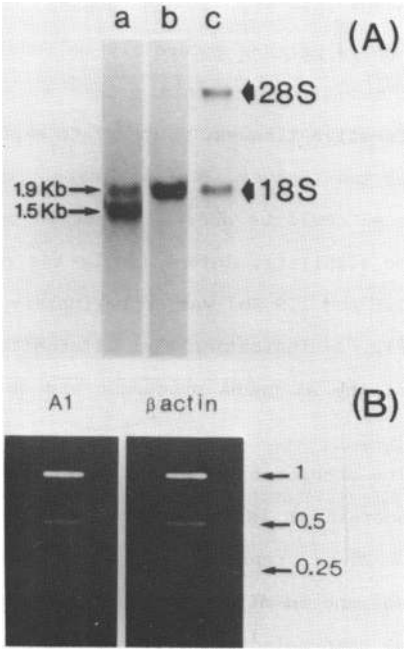


FIG. 3 - A: Northern type hybridization on poly A^+mRNA from HeLa cells (see Materials and Methods). Probe: (Lane a) complete pRP15 cDNA; (Lane b) EcoRI-EcoRI fragment of pRP15 of 309 n, at the 3'-end between the two polyadenylation sites (see Fig. 1). (Lane c) rRNA markers.
B: slot blot hybridization experiment for determining the overall abundance of A1-specific mRNAs relative to that of human ß-actin mRNA. The amounts of $polyA^+$ RNA from HeLa cells (µg) per slot are indicated alongside.

with the fact that hnRNP proteins are the most abundant non histone nuclear proteins in mammalian cells and that protein A1 is a major constituent of the complex (1).

A1 specific mRNAs in different human tissues

hnRNP core protein A1 has been reported to be the only protein of the hnRNP 40s complex whose intracellular level varies significantly with the proliferative state of the cell (10) being more abundant in actively growing cells. We wanted to test whether also the A1 mRNA expression was proliferation sensitive. Northern blot analysis using pRP15 cDNA probe was performed on $polyA^+$ RNAs from five human tissues: fibroblasts, HeLa cells,

kidney, placenta and thymus (see Fig. 4) and compared to ß-actin mRNA. Since the amount of loaded mRNA (4 µg) was accurately calibrated in all lanes, it is evident that kidney and placenta contain less A1 mRNA compared to the other three more proliferative tissues. Contrary to expectation also ß-actin mRNA is reduced in these two tissues. Thus the effect of cell proliferation on the level of protein A1 could be due, at least in part, to an effect on A1 mRNA synthesis and/or stability. Interestingly the relative abundance of the two mRNA species (1.5 and 1.9 Kb) was approximately the same in all the examined tissues (see Fig. 4) indicating that alternative polyadenylation is an intrinsic property of the A1 hnRNA precursor and does not depend on the cellular environment.

We next turned to the problem of the mRNA variants (α and ß, see Fig. 2) that can produce two protein A1 isoforms (Tyr $\rightarrow$ Phe and Arg $\rightarrow$ Lys) and asked the question of their representation among the two mRNA length variants (1.5 and 1.9 Kb) and in different tissues. To discriminate between the two mRNA isoforms we used selective oligonucleotide hybridization on the same filter used for hybridization with pRP15 cDNA. Two 15 n long oligodeoxynucleotides (I and II) were synthesized (see Materials and Methods) complementary to the same coding region of the two mRNAs (nucleotides 461-475) and thus differing only for one central nucleotide (A$\rightarrow$T). Then in a preliminary slot blot hybridization experiment (not shown) we determined the conditions under which each oligonucleotide hybridized only to the cognate RNA made in vitro with no cross hybridization to the other (16). As shown in Fig. 4 (upper panel), the 15-mer oligo I probe (form α) hybridized strongly to the HeLa cell RNA (both 1.5 and 1.9 Kb) and to a much lower degree to the RNAs from the other four tissues. Surprisingly however, by the same procedure, we were unable to detect significant hybridization with the 15-mer oligo II (form ß). Hybridization with oligo I sharply contrasts with that of cDNA probe only in the case of thymus (t) where the relative amount of type α mRNA seems to be lower than in the other tissues. Since type α protein (Tyr at residue 128) is predominant in all tissues examined so far it is possible that in human thymus other type α mRNA variants exist but escape detection with oligo I because other

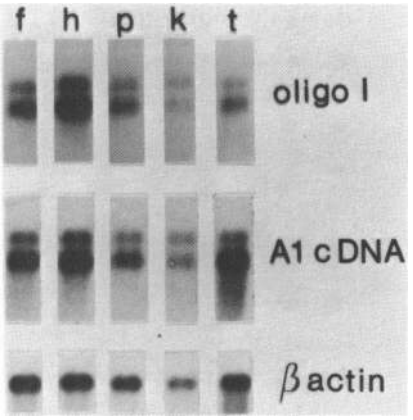


FIG. 4 - Northern type hybridization experiment to determine the relative abundance of A1-specific mRNAs in five human tissues. (f): SV40 transformed human fibroblasts; (h): HeLa cells; (p): placenta, (k): kidney; (t): thymus. 4 μg of polyA$^+$ RNA from the five tissues were loaded onto the gel. The nitrocellulose filter was hybridized first (upper panel) with a 15-mer oligonucleotide (Oligo I) (nucleotides 461-475 of pRP15) corresponding to the type α isoform (see Fig. 2). The same filter was then washed and rehybridized to the complete pRP15 cDNA probe (middle panel) and to a ß-actin cDNA probe (lower panel) as a control.

nucleotide changes can occur within the short oligo sequence. This fact could also account far the failure to detect hybridization with oligo II and explain at least in part the discrepancy with other data. These results are in line with our finding of different cDNA species within the libraries analyzed and with the results of Southern blot analysis of the human genome which showed at least 30 hybridizing bands (see next section). We therefore postulate that hnRNP protein A1 is most likely encoded by multiple mRNAs some of which direct the synthesis of the two protein isoforms hitherto found (α and ß) while other isoforms are still to be discovered.

Genomic complexity of A1 genes

To understand the origin of the protein isoforms described in the previous sections we undertook the analysis of the structure of A1 specific DNA sequences in the human genome. Dot blot hybridization to human DNA carried out at high stringency demonstrated the presence of about 30 sequences homologous to pRP15 cDNA per haploid genome (see Fig. 5(A)).

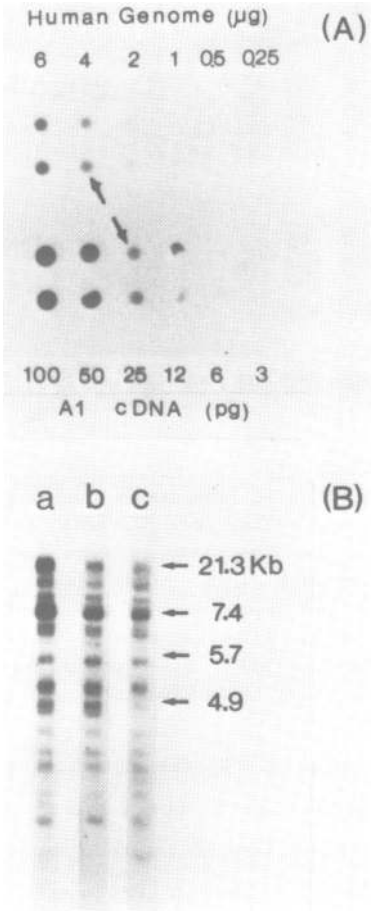


FIG. 5 - A: Dot blot hybridization experiment to determine the number of A1-specific sequences in human DNA genome under stringent conditions (see Materials and Methods). The indicated amounts of HeLa cell DNA (upper panel) and of pRP15 cDNA cloned in Bluescribe M13$^+$ (lower panel) were applied onto the blots in duplicate. The filter was probed with a complete pRP15 cDNA. The number of chromosomal sequences was determined by comparison of signal intensities (arrows).
B: Southern type hybridization of pRP15 cDNA to HeLa cell DNA digested with EcoRI. Lane a): complete pRP15 cDNA probe. Lane b): BstNI-Hind-III fragment of pRP15 cDNA in the coding region. Lane c): Hind-III-EcoRI fragment of pRP15 cDNA in the 3'-end non coding region. Molecular weight markers are shown alongside.

Accordingly Southern blot hybridization to HeLa cell DNA digested with EcoRI using pRP15 probe at high stringency revealed about 30 different bands (see Fig. 5(B)). The pattern of Southern hybridization did not change significantly and in any case was not simplified when the coding or the non coding regions of pRP15 were used as probes (see Fig. 5(B)). In an attempt to reduce the complexity of the hybridization pattern we used as probe a short oligonucleotide complementary to the 5'-end of pRP15. Also in this case the pattern remained quite complex with many bands recognized by the oligonucleotide (not shown). Quantitatively similar results were obtained on Southern blots of DNAs extracted from lymphocytes of five different individuals digested with a number of (restriction enzymes: Dra I, Msp I, Hind III, Cla I, Xba I (not shown)). Furthermore in these experiments no restriction site polymorphism was observed. Taken all together these results indicate the existence of a multigene family with both multiple functional genes and possibly multiple pseudogenes (see ahead) as described for several housekeeping genes (17) and for other genes like those of hnRNP C1 protein (13), snRNP proteins (17), HMG proteins (18). Given this type of results and in order to better analyze the A1 gene family, we screened a human genomic DNA library in λ Ch4A by plaque hybridization with pRP15 cDNA probe (see Materials and Methods). 22 independent positive clones with EcoRI inserts ranging from 3100 to 16000 bp were obtained. The 15 inserts giving the strongest hybridization signal were subcloned in Bluescribe M13$^+$ vector, submitted to detailed restriction analysis and the fragments so obtained were tested for hybridization to different regions of pRP15 cDNA (coding and non coding). On the basis of this analysis it can be concluded that the majority of genomic clones behaved as processed pseudogenes since, although colinear with the cDNA, they partially diverged from it. In fact some lacked restriction sites (present in the cDNA) while others had restriction fragments that did not hybridize or hybridized only weakly to the corresponding ones on the cDNA (not shown). This conclusion was further strengthened by the determination sof the nucleotide sequence of the genomic fragment that showed more similarities to the cDNA which although largely

homologous to the latter, contained many stop codons in all reading frames and a central highly divergent region of 200 n. In conclusions it appears that the majority of the genomic clones selected by plaque hybridization with pRP15 cDNA consist of non expressed pseudogenes of the processed type. In order to fish out an active gene, we have recently devised a method that favours the selection of the expressed genes against the non expressed ones. By this method we have already isolated an A1 specific gene sequence that appears to contain introns (to be published).

DISCUSSION

Identification of A1 transcripts

By screening two human cDNA libraries prepared from human liver and human fibroblasts we isolated four A1 specific cDNAs: pRP10, pRP12, pRP13 and pRP15. Of these, pRP15 contained the whole coding region, the complete 3'-end non coding region, and 85 n of the 5'-end non coding region (see Fig. 1). The other three cDNAs lacked more or less extended regions at the 5'-end probably due to premature termination of reverse transcription from the mRNAs. All four transcripts contained the same open reading frame and are extremely conserved except for the following differences (see Fig. 2):

1) the 3'-end non coding region can be either 307 or 722 nucleotide long with one or two polyadenylation sites (AATAAA) respectively.
2) the two longest T-stretches in the 3'-end non coding region have variable length.
3) one cDNA (pRP12) shows two nucleotide substitutions in the coding region that cause two aminoacid changes in the protein.

Since three of the four cDNAs were truncated, we cannot rule out the possibility that other variations could be found at the 5'-end of the cDNAs.

Properties of the non coding regions

As already mentioned in the Results, the 5'-end non coding sequence of pRP15 is probably truncated. Furthermore nucleotides 3-31 appear to be repeated in the inverted orientation at nucleotides 516-544. In spite of the possibility of a reverse transcription artifact at the cDNA tail it is worth underlining that the 85 nucleotides upstream of the first AUG are all in

frame with the following coding sequence. It is difficult however to attribute a biological significance to this observation since all the available data on the aminoacid composition and sequence of protein A1 indicate a polypeptide of 320 aminoacids (9). Furthermore, in pRP15 the AUG is immediately preceded by the CCGTC sequence that fits the consensus sequence (CCANC) for initiation of translation often found in the transcripts of vertebrates (19).

As shown in Figure 2, the 3'-end non coding region of A1 specific cDNA can consist either of 307 n (pRP10 and pRP13) or of 722 n (pRP12 and pRP15). Accordingly, Northern blot experiments with polyA$^+$ mRNA from different human tissues (see Fig. 4) evidenced two corresponding mRNA forms of 1.5 and 1.9 Kb. Given the presence of two polyadenylation sites in the longest transcripts we attribute the mRNA length heterogeneity to alternative polyadenylation of the same precursor with a slight preference for the first site. Since the relative abundance of the long vs. short mRNA is about the same in all tissues (see Fig. 4) the choice of the polyadenylation signal seems to be dependent solely on the neighbouring sequences in the precursor.

Variants in the coding region

The deduced aminoacid sequence of the protein encoded by pRP15 cDNA matches exactly the known primary structure of human protein A1. On the other hand, from the same cDNA library we could isolate a second cDNA that showed two aminoacid substitutions (Tyr --> Phe and Arg --> Lys) at residues 128 and 146 respectively producing two protein A1 isoforms. Isoforms of protein A1 were already surmised by our group (11) and by other authors (9) after the finding that certain A1 peptides obtained by protease digestion were a mixture of two peptides with one aminoacid sustitution. In addition, a certain degree of physical heterogeneity of protein A1 has been reported by several authors (1).

In a previous paper (11) we described a human A1 peptide with a substitution at residue 127 (Tyr --> Phe) that corresponds exactly to that predicted by pRP12 cDNA; another A1 peptide showed a substitution at residue 190 (Ser --> AsN) for which no corresponding cDNA has yet been found. Other authors (9) have reported an aminoacid substitution (Arg --> Lys) at residue

30 in human protein A1. Thus our isolation of A1 mRNA isoforms further supports the existence of several protein A1 isoforms. Moreover the results of Northen blot hybridization with cDNA and specific oligonucleotide probes (see Fig. 4) indicate that other isoforms besides the ones isolated so far must exist at least at the mRNA level.

It should be considered however that only conservative substitutions were observed between chemically similar aminoacids for which it is difficult to envisage a biological significance. It is possible in fact that a duplication of an ancestral A1 gene allowed the occurrence of neutral mutations subsequently fixed by random genetic drift. On the other hand, it should be observed that the aminoacid substitution at position 146 (Arg --→ Lys) falls inside a consensus for RNA recognition and/or binding (see below).

Abundance of A1 specific mRNA

The slot blot experiment reported in Fig. 3 (B) allowed an estimation of the relative abundance of A1 mRNA in HeLa cells in the order of 0.1% of the total polyA$^+$ mRNA. This figure, obtained by using as reference the mRNA for ß-actin (15) is about 10 fold lower than a previous estimation based on the abundance of human DHFR mRNA (20) but probably more accurate. In either case A1 mRNA is abundant in the nucleus in accordance with the fact that protein A1 is the major component of the hnRNP complex that in turn accounts for a large fraction of non-histone nuclear protein. Northern blot experiment on five human tissues (see Figure 4) showed that, as it was shown for protein A1, (10) also A1 mRNA is much more abundant in actively proliferating cells (HeLa), and this is the case for both the 1.5 and the 1.9 Kb mRNA forms.

This could indicate a direct involvement of A1 in some key steps of nucleic acid metabolism such as splicing as it has been shown for another protein (C1) of the hnRNP complex (2).

A1 domain structure

The domain structure of protein A1 has been previously discussed by several authors on the basis of protein sequencing studies (9,11). Secondary structure predictions suggest two distinct domains: the NH_2-terminal two thirds (residues 1-195) has a high α-helical probability while the

COOH-terminal portion (residues 197-320) has a low α-helical probability and an extremely high (40%) content of glycine. In the NH_2-terminal domain, the first 184 aminoacids, consist of two adjacent 91 residue partially homologous stretches that are believed to contain two independent binding domains for single-stranded nucleic acids also on the basis of photochemical cross-linking studies (21). Within each repeat we have located (at residues 56-63 and 146-153) an 8 aa sequence ($\begin{matrix}R\\K\end{matrix}$ G F $\begin{matrix}G\\A\end{matrix}$ F V T $\begin{matrix}Y\\F\end{matrix}$) that fits a consensus found in several eukaryotic nuclear proteins known to interact with RNA (9,11,12,22-29). As already mentioned, the two protein A1 isoforms reported here differ for two aminoacids one of which falls in the consensus. Therefore they could have a different mode of binding to RNA and a different function.

As to the glycine-rich COOH-terminal domain of protein A1, recent evidence suggests that also this part of the protein interfaces directly with nucleic acid and produces positive cooperativity of binding (30). It is worth noticing that a remarkable primary structure homology can be seen between the COOH-terminal portion of A1 and that of nucleolin (C23 fragment), (27) to indicate a similar role of the latter in the processing of ribosomal RNA.

Genomic complexity of A1 genes

Southern blot analysis on human DNAs from various sources showed multiple bands. The stringency of hybridization conditions made it unlikely that genes of other related components of the hnRNP complex could be detected. In the light of our results on the mRNA variants it is likely that some of the bands correspond to multiple functional genes. However the screening of a human DNA genomic library demonstrated many processed pseudogenes as is the case for a number of mammalian multigene families (17). Interestingly however the A1 specific sequence pattern, although very complex, was exactly the same in different individuals indicating a selective pressure in the evolution of the A1 gene family. Additional data on the possibility of a multigene family encoding protein A1 have recently been reported (31).

Conservation of protein A1 in mammals

The comparison of protein A1 from different species indicated a remarkable conservation in the primary structure. Human protein A1 (isoform α, see Fig. 2) is 100% homologous to the proteins A1 from both rat (12) and calf thymus (32). The comparison of the nucleotide sequence of our human cDNAs with that of rat cDNA (12) also gives a surprising result. As shown in Fig. 1, the two sequences are highly homologous both in the coding region (92.3%) and in the non coding regions (85.6% and 73% for the 3' and the 5'-end respectively). Only 72 point mutations are present in the coding region, 67 of which at the third nucleotide, and 5 at the first. If one considers that the two species are separated by 80 million years of evolution it should be concluded that protein A1 has a very low rate of evolution in the order of that of some histones, α-actin and somatostatin-28 (33) or lower. This again points to an essential role of this protein in some key process (splicing?) and to a strong selective pressure toward the maintenance of the structure, probably finalized to ensure the accuracy of interaction with both RNA and other proteins in the hnRNA complex. These considerations seem to be applicable also to the non coding regions that, could therefore play an important role in post-transcriptional regulation.

AKNOWLEDGMENTS

Work supported by the Progetto Finalizzato "Oncologia" and by the Progetto Finalizzato "Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie" CNR, Roma, and by an Italian Ministry of Education fund. The authors are grateful to A. Falaschi, F. Cobianchi, M. Mottes, K. Williams and G. Dreyfuss for helpful discussions and for providing unpublished information. The technical assistance of D. Arena is also aknowledged. M.B. was supported by a fellowship of the Consiglio Nazionale delle Ricerche, M.T.B. acknowledges the support of the "Fondazione Adriano Buzzati-Traverso".

*To whom correspondence should be addressed

REFERENCES

1) Dreyfuss, G. (1986) Ann. Rev. Cell. Biol. 2: 459-498.
2) Choi, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. (1986) Science 231: 1534-1539.
3) Brody, E. and Abelson, J. (1985) Science 228: 963-967.

4) Grabowski, P.J., Seiler, S.R. and Sharp, P.A. (1985) Cell. 42: 345-353.
5) Samarina, O.P., Lukanidin, E.M., Molnar, J. and Georgiev, G.P. (1968) J. Mol. Biol. 33: 251-263.
6) Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. (1977) Cell 11: 127-138.
7) Choi, Y.D. and Dreyfuss, G. (1984) Proc. Natl. Acad. Sci. USA 81: 7471-7475.
8) Leser, G.P., Escara-Wilke, J. and Martin, T.E. (1984) J. Biol. Chem. 259: 1827-1833.
9) Kumar, A., Williams, K.R. and Szer, W. (1986) J. Biol. Chem. 261: 11266-11273.
10) Celis, J.E., Bravo, R., Arenstorf, H.P. and LeStourgeon, W.M. (1986) FEBS Lett. 194: 101-109.
11) Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K.R., Multhaup, G., Beyreuther, K., Werr, H., Henrich, B. and Schäfer, K.P. (1986) The EMBO Journal 5: 2267-2273.
12) Cobianchi, F., SenGupta, D.N., Zmudzka, B.Z. and Wilson, S.H. (1986) J. Biol. Chem. 261: 3536-3543.
13) Nakagawa, T.Y., Swanson, M.S., Wold, B. and Dreyfuss, J. (1986) Proc. Natl. Acad. Sci. USA 83: 2007-11.
14) Stanley, K.K. and Luzio, J.P. (1984) EMBO J. 3: 1424-1434.
15) Ponte, C., Gunning, P., Balan, M., and Kedes, L. (1983) Mol. Cell. Biol. 3: 1783-1791.
16) Buvoli, M., Biamonti, G., Riva, S. and Morandi, C. (1987) Nucleic Acids Res.15: 9091.
17) Wagner, M. and Perry, R.P. (1985) Mol. and Cell. Biology 5: 3560-3576.
18) Landsman, D., N. Soares, Gonzalez, F.J. and Bustin, M. (1986) J. Biol. Chem. 261: 7479-7484.
19) Kozak, M. (1984) Nucleic Acid Res. 12: 857-875.
20) Morandi, C., Masters, J., Mottes, M. and Attardi, G.(1982) J. Mol. Biol. 156: 583-607.
21) Merrill, B.M., LoPresti, M.B., Stone, K.L. and Williams, K.R. (1986) J. Biol. Chem. 261: 878-883.
22) Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. 6: 2932-2943.
23) Sachs, A.B., Bond, M.W. and Kornberg, R.G. (1986) Cell 45: 827-835.
24) Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Philipson, L. (1986) The EMBO J. 5: 3209-3217.
25) Lahiri, D.K. and Thomas, J.O. (1986) Nucleic Acids Res. 14: 4077-4094.
26) 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. (1987) Proc. Natl. Acad. Sci. Usa 84: 2421-2425.
27) Lapeyre, B., Bourbon, H. and Amalric, F., (1987) Proc. Natl. Acad. Sci. USA 84: 1472-1476.
28) Swanson, M.S., Nakagawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) Mol. and Cell. Biology, p. 1731-1739.
29) Haynes, S., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.B. (1987) Proc. Natl. Acad. Sci. USA 84: 1819-1823.

30) Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1988) J. Biol. Chem. 263: 1063-1071.
31) Heesen, J.T., Melchers, K. and Schafer, K.P. (1987) Mol. Biol. Reports 12: 176.
32) Williams, K.R., Stone, K., LoPresti, B.M., Merrill, M.B. and Planck, S.R.L. (1985) Proc. Natl. Acad. Sci. USA 82: 5666-5670.
33) Li, W.H., Wu, C.I. and Lou, C.C. (1985) Mol. Biol. Evol. 2: 150-174.
34) Mottes, M., Tsai Lai, S.A., Montoya, J. and Attardi, G. (1984) Gene 27: 109-113.
35) Benton, W.D. and Davis, R.W. (1977) Science 196: 180.
36) Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132: 6-13.
37) Mason, P.J. and Williams, T.G. (1985) In: Nucleic Acid Hybridization, IRL Press, p. 113-137.
38) Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
39) Maxman, A.M. and Gilbert, W. (1980) Methods in enzymology 65: 499-560.
40) Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 24: 5294-5299.
41) Rave, N., Crkvenjakov, R. and Boedtker, H. (1979) Nucleic Acids Res. 11: 3559-3567.
42) Mariani, B.D. and Schimke, R.T. (1984) J. Biol. Chem. 259: 1901-1910.