Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1367–1371, February 1990
Cell Biology

Potential for two isoforms of the A1 ribonucleoprotein in *Xenopus laevis*

(heterogeneous nuclear ribonucleoprotein/RNA-binding protein)

BRIAN K. KAY*, RAVI K. SAWHNEY*, AND SAMUEL H. WILSON†

*Department of Biology, CB 3280, University of North Carolina, Chapel Hill, NC 27599-3280; and †Laboratory of Biochemistry, National Cancer Institute, Bethesda, MD 20892

Communicated by Joseph G. Gall, November 21, 1989

ABSTRACT **We have identified the *Xenopus* cognates for the RNA-binding protein A1. This protein has previously been shown to be one of the components of the heterogeneous nuclear ribonucleoprotein (hnRNP) complex in rat and human cells. We have isolated several *Xenopus* clones from oocyte, tailbud embryo, and leg muscle cDNA libraries and determined their nucleotide sequences. Potentially, two different A1 isoforms are expressed in *Xenopus*; they have been termed XA1a and XA1b. Besides insertions and deletions, the XA1a protein sequence is 92% identical to the rat protein and suggests very similar secondary structures. There are two segments in the COOH-terminal domain where deletions or insertions are apparent: the rat protein does not have a 48-residue sequence that is present in the frog protein, and the frog protein does not have a 12-residue sequence that is present in the rat protein. We have confirmed that the XA1a protein is larger than rat A1 by *in vitro* transcription, translation, and gel electrophoresis. The second isoform, XA1b, is very similar to the XA1a isoform, except it has a different COOH terminus due to the absence of a 73-nucleotide region from its cDNA clones. Transcripts representing both isoforms have been detected in various *Xenopus* RNA preparations by polymerase chain reaction experiments with A1-specific oligonucleotides. Our findings suggest that the isoforms are encoded by one or two genes and are the result of alternative splicing. We discuss the biological implications of having two forms of the A1 component of hnRNP particles.**

In the heterogeneous nuclear ribonucleoprotein (hnRNP) particles of eukaryotes, there are a number of distinct protein components. One of the best characterized is the A1 species (1). It has M_r 34,192, is conserved at the primary structure level between rat and man, and is immunologically related to the other protein components of the hnRNP complex. A1 has an ordinary globular NH_2-terminal domain that notably consists of two 90- to 93-amino acid repeats, each containing two short sequences that are highly conserved among RNPs (2–4). The COOH-terminal domain is flexible, extended, and glycine-rich (5). Experiments suggest that both domains are involved in RNA binding and make direct contact with the nucleic acid; the COOH-terminal domain is involved in cooperative protein–protein interactions as well (4).

Recently, cDNAs and genes encoding eukaryotic A1 proteins have been identified. The first cDNA, isolated from a rat brain library, corresponds to an mRNA encoding a protein of 320 amino acid residues (5). Equivalent cDNA clones have also been isolated from human sources (6). The deduced human and rat A1 proteins are identical, indicating that the primary structure of this nuclear protein has been highly conserved during evolution (7). Recently, a cDNA has been identified from *Drosophila melanogaster* (8), and the open reading frame encodes a protein with significant similarity to the rat A1 protein, especially in the NH_2-terminal domain.

In this paper we describe the isolation and characterization of several A1 cDNA clones from the frog, *Xenopus laevis*. This study was conducted for two reasons: (*i*) to learn from comparison what regions of A1 are conserved and most likely to have important functions and (*ii*) to study the dynamics of this protein during frog development. There are many precedents for differential expression of RNAs and proteins in oocytes and embryos of *Xenopus* (9). From examination of different cDNA clones we find that at least two isoforms of the A1 protein exist.‡ These different A1 species, though highly related to the characterized rat species in primary structure, differ in their COOH-terminal domains. This region of the A1 protein has previously been suggested to regulate protein–protein interactions in the hnRNP (4).

MATERIALS AND METHODS

Materials. Restriction and DNA modification enzymes were purchased from Promega and United States Biochemicals and used according to the manufacturer. The plasmid vector pGEM-7zf was obtained from Promega. For *in vitro* transcription and translation reactions, reagents were obtained from Stratagene, and [^{35}S]methionine and Amplify were purchased from Amersham. *In vitro* transcription reactions also included m7G(5′)ppp(5′)G cap analog (New England BioLabs) and RNasin (Promega), and DNA was removed with RNase-free DNase. Oligonucleotides were synthesized on an Applied Biosystems machine and purified by HPLC.

Isolation of *Xenopus* cDNA Clones. A nucleic acid hybridization probe was prepared from a rat A1 cDNA clone (5). The rat cDNA clone was digested with *Eco*RI to release a 1397-base-pair (bp) fragment with the entire A1 coding region; the fragment was labeled by nick-translation (10) and used to screen several different *Xenopus* cDNA libraries under moderate stringency conditions (1 M NaCl, 55°C); three clones were isolated from an oocyte library (11), two from an embryonic stage 24 library (gift of K. Richter and I. Dawid, National Institute of Child Health and Human Development, Bethesda, MD), and one from an adult leg muscle library (gift of R. Harland, University of California, Berkeley). The recombinant λ phages were propagated and the *Eco*RI inserts were inserted at the *Eco*RI site of the plasmid pGEM-7zf. The oocyte clone inserts were 1313, 1070, and 1168 bp; the stage 24 clone inserts were 1263 and 1037 bp; and the leg muscle insert was 628 bp in length.

Sequence Analysis. Nucleotide sequences were determined by the dideoxy method (12) on double-stranded DNA (13).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: RNP, ribonucleoprotein; hnRNP, heterogeneous nuclear RNP; PCR, polymerase chain reaction; nt, nucleotide(s).

‡The sequences reported in this paper have been deposited in the GenBank data base (accession no. M31041 for XA1a and for XA1b).

The insert termini were sequenced with primers corresponding to the flanking T7 and SP6 RNA polymerase promoters, and internal regions were sequenced for the largest oocyte cDNA clone by a progressive deletion method (14). The second strand of the inserts was sequenced with oligonucleotide primers synthesized according to the nucleotide sequence of the first strand; these 17-mers were located at nucleotide positions 135, 298, 440, 608, 765, 927, 1080, and 1247 of the complete cDNA sequence. The largest oocyte clone was sequenced on both strands, and the other clones were completely sequenced only on one strand. The data were analyzed with the Wisconsin Genetics Computer Group programs (15).

***In Vitro* Transcription and Translations.** RNA transcripts were prepared from the recombinants by *in vitro* transcription and translated into proteins for SDS/polyacrylamide gel electrophoresis. The *Xenopus* subclones in pGEM-7zf were linearized with *Hin*dIII, and the DNA was recovered after phenol/chloroform extraction by isopropyl alcohol precipitation. Sense RNA transcripts were then generated *in vitro* with T7 RNA polymerase using an *in vitro* transcription system according to the manufacturer, except that RNasin and 5′ cap analog were included in the reaction mixture (16). Recovered transcripts were translated *in vitro* with a commercial rabbit reticulocyte lysate and [^{35}S]methionine. To analyze the reaction products, the samples were denatured by boiling in sample loading buffer and resolved by SDS/polyacrylamide gel electrophoresis. In parallel, a small amount of the rat A1 protein, generated by expression of a full-length cDNA clone in bacteria (17), was loaded on the gels. Proteins were then localized in the gel by staining with Coomassie blue and x-ray film autoradiography after treatment with Amplify.

Polymerase Chain Reaction (PCR) Experiments. Total cellular RNA was prepared from accessory cell-free oocytes and adult gastrocnemius leg muscles by standard procedures (18). cDNA was primed from 5 μg of total cellular RNA with reverse transcription (Promega) and oligo(dT)$_{15}$ (Boehringer Mannheim) and then used directly for PCR according to a published procedure (19). One nanogram of cDNA plasmid DNA, representing either A1 isoform, was also separately used for PCR. The PCR primers were 32 nucleotides (nt) in length and corresponded to positions 992 and 1222 of the XA1a sequence. One-tenth of each of the PCR products was analyzed by electrophoresis in 1.4% agarose gels and specific sequences were detected on nitrocellulose blots by hybridization with ^{32}P-labeled oligonucleotides and x-ray film autoradiography. The oligonucleotide probes were 45 nt, had equal G+C % content, and corresponded to positions 1054 and 1166, which represent "a"-specific and common ("a + b") regions, respectively.

RESULTS

Identification of a *Xenopus* A1 cDNA Sequence. We screened a *Xenopus* oocyte cDNA library, using the rat A1 cDNA as a probe, and isolated three clones that cross-hybridized. The inserts from these clones were subcloned into plasmids and their inserts were sequenced. Fig. 1 shows the nucleotide sequence and the hypothetical coding region of the long open reading frame from our longest cDNA. The other two clones were found to be colinear at their 3′ termini and identical in sequence for the regions of overlap, differing only in length at their 5′ termini. The longest open reading frame is 1095 nt residues and can encode a polypeptide of 365 amino acids with M_r of 38,296 and pI of 9.5. The putative protein encoded by this sequence is named XA1a.

Structural Predictions of the *Xenopus* A1 Protein. Examination of a matrix comparison of the frog protein reveals that the NH$_2$-terminal and COOH-terminal domains of the protein

```
GGAGTCACCATGCACAAGTCCGAGGCACCCAACGAGCCAGAGCAACTTCGCAAGCTGTTC     60
        M  H  K  S  E  A  P  N  E  P  E  Q  L  R  K  L  F      17

ATTGGAGGCTTGAGTTTTGAAACCACAGATGAAAGTCTCCGCGAGCACTTTGAGCAATGG    120
I  G  G  L  S  F  E  T  T  D  E  S  L  R  E  H  F  E  Q  W      37

GGCACCCTTACAGACTGTGTGGTTATGAGGGATCCAAACTCAAAACGTTCCCGTGGCTTT    180
G  T  L  T  D  C  V  V  M  R  D  P  N  S  K  R  S  R  G  F      57

GGATTTGTTACATACTTATCTACAGATGAAGTAGATGCTGCCATGACTGCTCGCCCACAT    240
G  F  V  T  Y  L  S  T  D  E  V  D  A  A  M  T  A  R  P  H      77

AAAGTGGATGGGCGAGTGGTTGAACCTAAAAGGGCTGTCTCTAGAGAGGATTCCTCTAGG    300
K  V  D  G  R  V  V  E  P  K  R  A  V  S  R  E  D  S  S  R      97

CCTGGTGCACACCTCACCGTAAAGAAAATCTTTGTAGGTGGTATCAAGGAGGACACAGAA    360
P  G  A  H  L  T  V  K  K  I  F  V  G  G  I  K  E  D  T  E     117

GAAGATCATTTACGAGAATATTTCGAGCAATATGGCAAAATTGAAGTTATAGAGATAATG    420
E  D  H  L  R  E  Y  F  E  Q  Y  G  K  I  E  V  I  E  I  M     137

ACTGACCGAGGCAGTGGCAAGAAAAGAGGCTTTGCATTTGTCACATTTGAAGATCATGAT    480
T  D  R  G  S  G  K  K  R  G  F  A  F  V  T  F  E  D  H  D     157

TCCGTTGACAAGATTGTCATCCAGAAATATCACACCGTCAACAACCACAATTCTCAAGTG    540
S  V  D  K  I  V  I  Q  K  Y  H  T  V  N  N  H  N  S  Q  V     177

CGGAAGGCACTCTCCAAACAGGAAATGGCAAGTGTTTCTGGCAGTCAGAGAGAACGTGGT    600
R  K  A  L  S  K  Q  E  M  A  S  V  S  G  S  Q  R  E  R  G     197

GGCTCTGGAAACTATGGAAGCCGTGGTGGGTTTGGTAATGATAACTTTGGTGGTCGTGGT    660
G  S  G  N  Y  G  S  R  G  G  F  G  N  D  N  F  G  G  R  G     217

GGCAACTTTGGTGGCAACAGAGGAGGAGGGGGTGGATTTGGTAATCGAGGTTATGGTGGA    720
G  N  F  G  G  N  R  G  G  G  G  G  F  G  N  R  G  Y  G  G     237

GACGGCTACAATGGTGATGGCCAATTATGGTGGCAGCCCTCCCTACTCGGGTGGAACCGA    780
D  G  Y  N  G  D  G  Q  L  W  W  Q  P  S  L  L  G  W  N  R     257

GGCTATGGTGCTGGCCAAGGAGGTGGATATGGTGCTGGCCAAGGAGGTGGATATGGTGGT    840
G  Y  G  A  G  Q  G  G  G  Y  G  A  G  Q  G  G  G  Y  G  G     277

GGTGGCCAAGGAGGTGGATATGGTGGAAATGGAGGATACGATGGTTATAATGGCGGAGGC    900
G  G  Q  G  G  G  Y  G  G  N  G  G  Y  D  G  Y  N  G  G  G     297

AGTGGCTTCAGTGGCTCTGGTGGTAACTTTGGTAGCAGTGGGGGATATAACGACTTTGGC    960
S  G  F  S  G  S  G  G  N  F  G  S  S  G  G  Y  N  D  F  G     317

AACTACAACAGTCAGTCATCATCCAACTTTGGCCCAATGAAAGGTGGAAATTATGGTGGT   1020
N  Y  N  S  Q  S  S  S  N  F  G  P  M  K  G  G  N  Y  G  G     337

GGCAGAAATTCTGGACCATATGGAGGTGGCTATGGCGGAGGGTCTGCTAGCAGCAGTAGC   1080
G  R  N  S  G  P  Y  G  G  G  Y  G  G  G  S  A  S  S  S  S     357

GGATATGGCGGTGGGAGGAGGTTTTAATTTTTCCACAGGGGAATATCCATCTACATGATT   1140
G  Y  G  G  G  R  R  F                                         365

AAACGAAACGTGGACCTTGGTTTGTGAAGCAGTGTTCAGATACAGGTTTTAAGGTTAACT   1200

GATGTGACGGATCTGACCAACTTCAAATTTCACAGCTACTGTGTAGCTCTGTTTTGTTTT   1260

TTTGTTCTTTTCCAAAATAAAGTTTTTATGATTCTCTAAAAAAAAAA                1307
```

FIG. 1. Nucleotide and deduced amino acid sequences of the *Xenopus* XA1a cDNA clone and protein. Both strands of the insert were sequenced by the dideoxy method and only the coding strand and long open reading frame are shown. The putative poly(A) signal (AATAAA) is underlined just upstream of the poly(A) 3′ terminus. The *Eco*RI sites at the ends of the cloned insert are not shown. Northern blots of oocyte RNA suggest that this sequence is near full-length (data not shown).

contain internal repeats. In the NH$_2$-terminal domain, these repeats are 91 amino residues in length, corresponding to residues 4–94 and 95–185, and have 34% identity. Each of the repeats contains the RNP consensus sequence no 1, RGF(G/A)FVT(Y/F), described by Adam *et al.* (2), and also the RNP consensus sequence no. 2, K(I/L)F(V/I)GG(L/I), described by Merrill *et al.* (4). In the COOH-terminal domain, there are five and seven copies of the sequences GN(F/Y)G(S/G) and GYG(A/G), respectively.

Optimal amino acid alignment of the frog protein with the rat A1 protein is shown in Fig. 2. The proteins are strikingly homologous, including perfect matches in four regions that appear to be in nucleic acid-binding pockets in rat A1 (4). Besides deletions and insertions, the two proteins match in 276 of 301 overlapping residues. Nine of the 25 amino acid differences are Gly-Ser differences in the COOH-terminal domain, residues 190 to the end; Gly-Ser substitutions could be functionally conservative in the A1 COOH-terminal domain, since neither residue promotes α-helix formation, thus

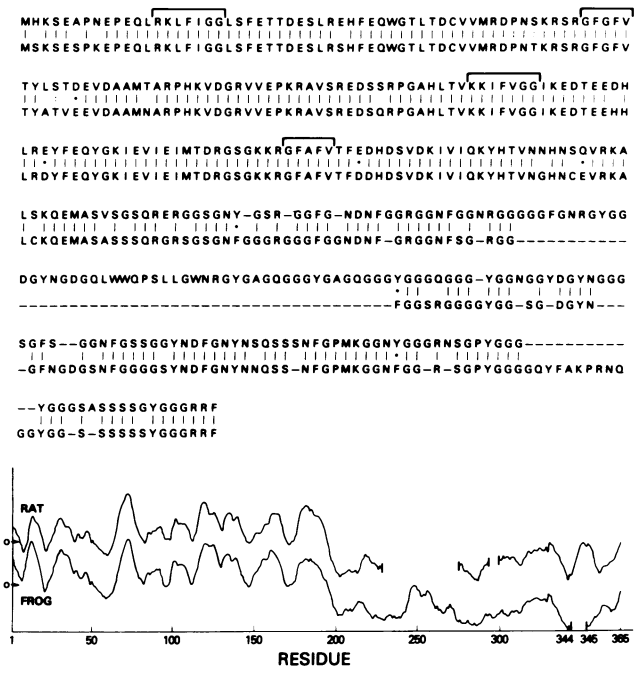

FIG. 2. (*Upper*) Optimal alignment of amino acid sequences of *Xenopus* XA1a (365 residues) and rat A1 (320 residues). The alignment spans 380 residues with 60 residues shown on each line, except the last line of 20; the *Xenopus* sequence is on the top line and the rat sequence is on the bottom. Dashes (–) represent apparent deletions or insertions. Symbols between lines mark amino acid identities (|), conservative amino acid substitutions (·), or other amino acid substitutions (:) commonly considered to be conservative (20). RNP consensus sequences (2, 4) are marked by brackets. (*Lower*) Comparison of predicted α-helix structure for *Xenopus* XA1a and rat A1 proteins illustrating locations and structure of apparent deletions or insertions. Computer-derived secondary structure predictions were as described (21, 22). Alignment is according to *Upper*. Arrows on the ordinate indicate the position of predictive score = 0; values above 0 represent greater probability of α-helix structures.

permitting flexibility in the protein backbone. There are two segments in the COOH-terminal domain where deletions or insertions are apparent. The rat protein does not have a 48-residue sequence in the frog protein, residues 230–277 in the alignment (Fig. 2 *Upper*), and the frog protein does not have a 12-residue sequence in the rat protein, residues 351–362. The two proteins are similar at the level of secondary structure prediction as well (Fig. 2 *Lower*), with patterns for the first 200 residues that are virtually identical. However, similarities in the COOH-terminal domain are less noticeable. The 48-residue "insertion" in the frog A1 proteins has higher prediction of α-helix than other regions of the COOH-terminal domain, and the 12-residue sequence novel to the rat protein creates a difference in the predicted structure of the extreme COOH-terminal residues of the two proteins.

A Second A1 Isoform Is Expressed in *Xenopus*. To examine the possibility that different isoforms of the A1 protein are expressed in *Xenopus*, we have isolated several additional cDNA clones. Two clones were obtained from an embryonic (stage 24) library, and one clone was isolated from a leg muscle library. The coding strand of each of these clones has been sequenced by the dideoxy nucleotide method using a set of eight evenly spaced oligonucleotide primers. Examination of the nucleotide sequences for the three clones revealed that they are identical in the regions of overlap, except that they lack a 73-nt region corresponding to the COOH terminus of XA1a (Fig. 3). This second cDNA isoform theoretically encodes a second form of A1 protein, which we have termed XA1b. The two isoforms differ in their COOH termini, where 15 amino acids (GYGGGSASSSSGYGGGRRF) at the COOH terminus of XA1a are replaced by a short sequence of five residues (GYPST) in XA1b.

To confirm the sizes of the proteins predicted from the nucleotide sequences, we prepared mRNA *in vitro* and translated the transcripts with a rabbit reticulolysate supplemented with [^{35}S]methionine. The protein products were then resolved by SDS/polyacrylamide gel electrophoresis, followed by autoradiography. As seen in Fig. 4, rat A1 (lane 1) and A1 proteins encoded by *in vitro* transcripts generated from XA1b (lane 2) and XA1a (lane 3) cDNAs were 34, 38, and 39 kDa, respectively. The sizes of these proteins match the sizes predicted from the open reading frames. It is of interest to note that the XA1a and XA1b proteins differed by 2 kDa in this experiment; such a difference theoretically represents a difference of 20 amino acids, which agrees well with the expected difference of 14 amino acids based on the nucleotide sequences (Fig. 3). Based on this *in vitro* translatability we expect that both protein isoforms exist *in vivo*.

Two important questions are raised by these findings. (*i*) Can we verify that the isolated cDNA clones represent authentic RNA isoforms? (*ii*) What is the relative proportion of the RNAs encoding the two isoforms in different tissues? To answer these questions we have examined RNA preparations by PCR. Oligonucleotide primers, flanking the region of the A1 RNA sequence that differed between XA1a and XA1b forms, were used to amplify the A1 cDNAs from total RNA of oocyte and adult leg muscle. Amplified DNA segments were then resolved by agarose gel electrophoresis, and the different isoforms were then detected on blots with another set of oligonucleotides that were either specific to XA1a or common to both types of isoform cDNAs (Fig. 5). By this analysis, both isoforms were detected in oocyte (lanes 3) and leg muscle (lanes 4) RNAs. From densitometric measurements of autoradiograms, it appears that the XA1b isoform is 3-fold more abundant than the XA1a isoform in

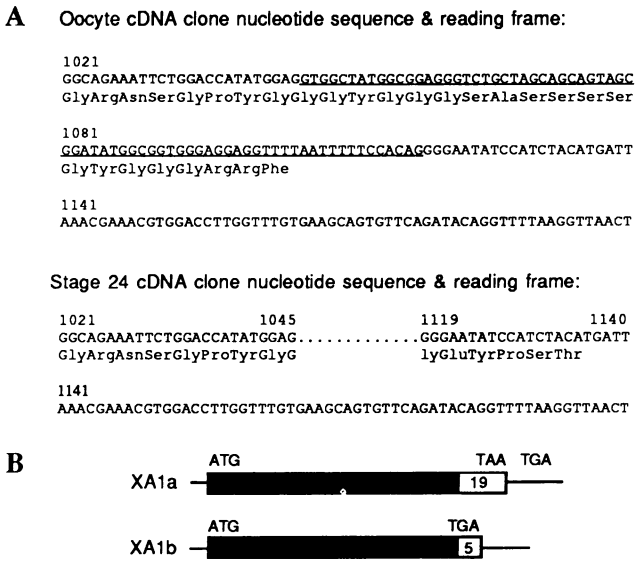

FIG. 3. (*A*) Sequence comparison of XA1a and XA1b cDNA clones and predicted proteins. The nucleotide sequences and coding frames are shown for the region that differs between the two major types of cDNA clones. The nucleotides are numbered according to the oocyte cDNA clone. The region in the XA1a cDNA that is absent in the XA1b cDNA is underlined; because there are strings of guanine residues flanking both sides of this segment, the termini of the missing region are not precisely defined. (*B*) Diagram of the two cDNA and protein isoforms in *Xenopus*. The boxes correspond to portions of the coding region, with the number of amino acid residues listed within the two different COOH-terminal segments. The sites of the initiation (ATG) and termination (TAA, TGA) codons are noted.

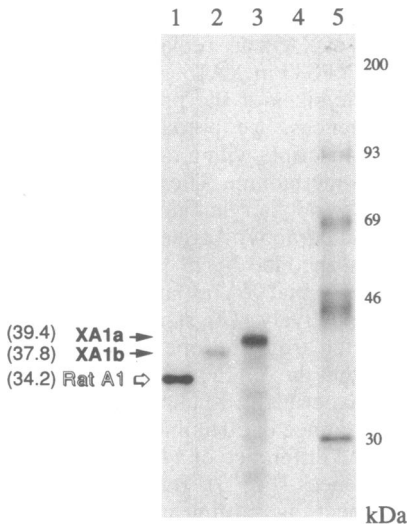

FIG. 4. Gel analysis of the rat and *Xenopus* XA1 proteins. Lane 1, rat A1 protein; lane 2, XA1b *in vitro* product; lane 3, XA1a *in vitro* product; lane 4, no RNA added to the translation mixture; lane 5, radioactive molecular mass markers. Synthetic mRNAs were prepared from the XA1a and XA1b cDNA clones and translated in a rabbit reticulolysate system in the presence of [^{35}S]methionine. The reaction products were then resolved by SDS/polyacrylamide gel electrophoresis, along with radioactive size markers and purified rat A1 protein. After gel electrophoresis, these various proteins were detected by Coomassie blue staining and x-ray film autoradiography. The mobilities of the major protein species and molecular mass standards (kDa) are indicated on the left and right, respectively.

either cell type. Several additional points should be noted. (*i*) These results are template and primer dependent. (*ii*) These amplified segments are derived from RNA and not from contaminating genomic DNA, as separate PCRs of frog DNA generate different-sized products (data not shown). (*iii*) Due to the repetitive nature of the A1 COOH-terminal amino acid sequence, one of the PCR primers hybridizes to a second upstream site to yield additional minor bands from the XA1a and XA1b RNA species. (*iv*) Since our three initial isolates

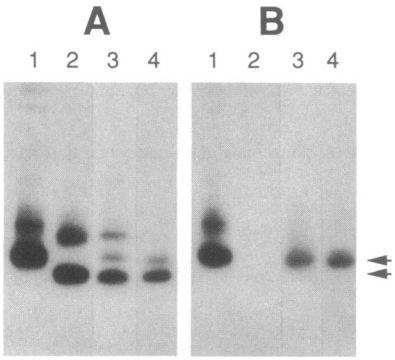

FIG. 5. PCR amplification of cDNAs representing both isoforms from various RNA preparations. A1 sequences were amplified by PCR with two oligonucleotide primers flanking the region that differs between XA1a and XA1b sequences. The PCR products were then resolved by gel electrophoresis in 1.4% agarose, blotted to nitrocellulose, and hybridized with ^{32}P-labeled oligonucleotide probes corresponding to both A1 forms (*A*) or XA1a alone (*B*). The autoradiograms show the hybridizing PCR products of the XA1b plasmid (lanes 1), XA1a plasmid (lanes 2), oocyte RNA (lanes 3), and leg muscle RNA (lanes 4). The upper and lower arrows on the right denote the positions of the major products for the XA1a and XA1b species, respectively. Secondary bands in the autoradiograms correspond to those PCR products that are generated from one oligonucleotide (992–1023) hybridizing to a second upstream region (815–845) in both A1 cDNA species.

from an oocyte cDNA λ library only corresponded to the XA1a isoform, we have rescreened the library and have recently identified several oocyte XA1b clones.

DISCUSSION

In this paper, we identify cDNA clones encoding the A1 hnRNP protein from the frog *Xenopus laevis*. Because of the great overall similarity between the predicted frog and mammalian species, we believe that the frog protein described in this paper is A1 and not one of the other hnRNP components, such as A2 (23). However, the *Xenopus* and mammalian A1 proteins do differ significantly in their COOH-terminal region, where a 48-residue sequence is absent from the mammalian proteins and at the same time the frog species lacks a 12-residue sequence that is present in the mammalian proteins. This finding is in agreement with previous immunoblot experiments that showed that a monoclonal antibody (iD2) reacts with hnRNP core proteins in *Xenopus* that are larger than those detected in mammals (24). The 48-residue region is unique to *Xenopus* so far, as it is also absent from the putative *Drosophila* A1 species (8). We also report in this paper the potential for two A1 isoforms in *Xenopus*, which differ only in their COOH-terminal polypeptide sequences.

From previous studies of the A1 protein, it appears to have at least two different functional domains. The NH_2-terminal half of the protein consists of two 91-amino acid repeats that form a nucleic acid-binding pocket. This region acts in a noncooperative manner to bind a stretch of 7 single-stranded nt *in vitro* (17). Since the NH_2-terminal domains of the frog and mammalian proteins are nearly identical, we can assume that this region in the frog protein will have the same nucleic acid-binding activity. This 91-amino acid repeat has been observed in a number of other types of RNA-binding proteins, including human and *Xenopus* C1 (3, 25), hamster nucleolin (26), and yeast poly(A)-binding protein (27).

The most notable difference between the frog (XA1a and XA1b) and mammalian A1 species occurs in the COOH-terminal domain, which has previously been shown to be involved in RNA-binding and protein–protein interactions (17). Because the COOH-terminal domains of the frog and mammalian A1 species are glycine and serine rich, and have similar secondary structures, we anticipate that the overall interactions among the protein components of the hnRNP complex are probably similar in frogs and mammals. However, if the primary sequence differences noted between the frog and mammalian A1 species are significant, we speculate that they may be compensated by changes in the other frog hnRNP proteins. This hypothesis will be testable in the future as other components of the frog hnRNP are identified and sequenced. Recently, the C component of the hnRNP has been identified in *Xenopus* (25) and man (3); the two proteins are 92% similar, with all amino acid substitutions being functionally conservative.

What may be the biological significance of two isoforms for the A1 species in *Xenopus*? Considering the fundamental nature of the hnRNP to RNA metabolism in eukaryotes (1), it may be surprising that any variation occurs at all. However, in mammals it now appears that hnRNPs contain 9–15 different proteins in the 30- to 45-kDa range and that protein heterogeneity may be the result of expression from multiple genes and posttranscriptional or posttranslational mechanisms (28, 29). In humans, for example, single amino acid variants have been detected among cDNA clones for A1 (30), and two different A1 species have been detected by two-dimensional gel electrophoresis (28). Moreover, transcripts of the human A1 gene may be spliced alternatively (31) and utilize either one of two poly(A) sites in the 3′ untranslated region (30). It is possible that these various A1 protein isoforms are developmentally regulated and/or that there is

selective incorporation of the isoforms into hnRNPs to alter RNA stability or splicing efficiency.

From our data, we speculate that alternative splicing is responsible for the generation of the two isoforms in *Xenopus*. We favor this hypothesis based on three findings. (*i*) The cDNA clones encoding the two isoforms are identical in nucleotide sequence in the coding and noncoding regions, except for a simple deletion in the embryonic cDNA clones. (*ii*) Both types of cDNA clones show the same polymorphism at position 1170, where the nucleotide at this 3′ untranslated region is either adenine or guanine. (*iii*) Preliminary genomic blots indicate that the A1 gene exists in low copy number (data not shown). Therefore, the simplest conclusion is that these proteins are the products of alternatively spliced transcripts from allelic genes, such that the exon encoding the COOH terminus of XA1a is retained during processing of some primary transcripts and is removed from some other transcripts to yield XA1b. This pattern of alternative splicing has been documented for a number of eukaryotic genes (32, 33). It is also formally possible, as suggested by the similarity of the sequences at the ends of the deleted 73-nt domain (Fig. 3*A*) to intron/exon junctions (34), that the 73-nt segment is a retained intron in XA1a.

We appreciate the assistance of Jeanne Alexander Harless in sequencing part of the first A1 *Xenopus* cDNA clone isolated. We acknowledge the generous sharing of recombinant cDNA libraries by I. Dawid, R. Harland, D. Melton, and K. Richter. This work was supported in part by grants from the National Institutes of Health (BRSG 2 SO7 RRO7072 and RO1 HL42250).

1. Dreyfuss, G., Swanson, M. S. & Pinol-Roma, S. (1988) *Trends Biochem. Sci.* **13,** 86–91.
2. Adam, S. A., Nakagawa, T., Swanson, M. S., Woodruff, T. K. & Dreyfuss, G. (1986) *Mol. Cell. Biol.* **6,** 2932–2943.
3. Swanson, M. S., Nakagawa, T. Y., LeVan, K. & Dreyfuss, G. (1987) *Mol. Cell. Biol.* **7,** 1731–1739.
4. Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. & Williams, K. R. (1988) *J. Biol. Chem.* **263,** 3307–3313.
5. Cobianchi, F., SenGupta, P. H., Zmudzka, B. Z. & Wilson, S. H. (1986) *J. Biol. Chem.* **261,** 3536–3543.
6. Riva, S., Moranie, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merrille, B., Williams, K. R., Multhaup, G., Beyreuther, K., Werr, B. & Schafer, K. P. (1986) *EMBO J.* **5,** 2267–2273.
7. Cobianchi, F., Biamonti, G., Bassi, M. T., Buvoli, M. & Riva, S. (1990) in *The Eukaryotic Nucleus: Molecular Structure and Macromolecular Assemblies*, eds. Strauss, P. & Wilson, S. (Telford, Caldwell, NJ), Vol. 2, pp. 131–147.
8. Haynes, S. R., Rebbert, M. L., Mozer, B. Z., Fouquiqnon, F. & Dawid, I. B. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1819–1823.
9. Davidson, E. H. (1986) *Gene Activity in Early Development* (Academic, New York), 3rd Ed.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), pp. 109–112.
11. Rebagliati, M. R., Weeks, D. L., Harvey, R. P. & Melton, D. A. (1985) *Cell* **42,** 769–777.
12. Sanger, F., Nicklen, S. & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
13. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.
14. Henikoff, S. (1984) *Gene* **4,** 351–357.
15. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387–395.
16. Krieg, P. A. & Melton, D. A. (1987) *Methods Enzymol.* **155,** 397–415.
17. Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V. & Wilson, S. H. (1988) *J. Biol. Chem.* **263,** 1063–1071.
18. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **19,** 5294–5299.
19. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.
20. Dayoff, M. O., Eck, R. V. & Park, C. M. (1972) in *Atlas of Protein Sequences and Structure*, ed. Dayoff, M. O. (Natl. Biomed. Res. Found., Silver Spring, MD), pp. 89–99.
21. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
22. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 3824–3828.
23. Merrill, B. M. & Williams, K. R. (1990) in *The Eukaryotic Nucleus: Molecular Structure and Macromolecular Assemblies*, eds. Strauss, P. & Wilson, S. (Telford, Caldwell, NJ), Vol. 2, pp. 148–175.
24. Leser, G. P., Escara-Wilke, J. & Martin, T. E. (1984) *J. Biol. Chem.* **259,** 1827–1833.
25. Preugschat, F. & Wold, B. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9669–9673.
26. Lapeyre, B., Bourbon, H. & Amalric, F. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 1472–1476.
27. Sachs, A. B., Bond, M. W. & Kornberg, R. D. (1986) *Cell* **45,** 827–835.
28. Wilk, H. E., Werr, H., Friedrich, D., Kiltz, H. H. & Schafer, K. (1985) *Eur. J. Biochem.* **146,** 71–81.
29. Pinol-Roma, S., Choi, Y. D., Matunis, M. J. & Dreyfuss, G. (1988) *Genes Dev.* **2,** 215–227.
30. Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M. T., Ghetti, A., Riva, S. & Morandi, C. (1988) *Nucleic Acids Res.* **16,** 3751–3770.
31. Biamonti, G., Buvoli, M., Bassi, M. T., Morandi, C., Cobianchi, F. & Riva, S. (1989) *J. Mol. Biol.* **207,** 491–503.
32. Crabtree, G. R. & Kant, J. A. (1982) *Cell* **31,** 159–166.
33. Andreadis, A., Gallego, M. E. & Nadal-Ginard, B. (1987) *Annu. Rev. Cell Biol.* **3,** 207–242.
34. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50,** 349–383.