Nucleic Acids Research, 1993, Vol. 21, No. 4 ***999–1006***

Three new members of the RNP protein family in *Xenopus*

Peter J.Good, Martha L.Rebbert and Igor B.Dawid
Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA

Received September 28, 1992; Revised and Accepted January 7, 1993 GenBank accession nos L02593–L02597 (incl.)

ABSTRACT

Many RNP proteins contain one or more copies of the RNA recognition motif (RRM) and are thought to be involved in cellular RNA metabolism. We have previously characterized in *Xenopus* a nervous system specific gene, *nrp1*, that is more similar to the hnRNP A/B proteins than to other known proteins (K.Richter, P.J.Good, and I.B.Dawid (1990), New Biol. 2, 556–565). PCR amplification with degenerate primers was used to identify additional cDNAs encoding two RRMs in *Xenopus*. Three previously uncharacterized genes were identified. Two genes encode hnRNP A/B proteins with two RRMs and a glycine-rich domain. One of these is the *Xenopus* homolog of the human A2/B1 gene; the other, named hnRNP A3, is similar to both the A1 and A2 hnRNP genes. The *Xenopus* hnRNP A1, A2 and A3 genes are expressed throughout development and in all adult tissues. Multiple protein isoforms for the hnRNP A2 gene are predicted that differ by the insertion of short peptide sequences in the glycine-rich domain. The third newly isolated gene, named *xrp1*, encodes a protein that is related by sequence to the nrp1 protein but is expressed ubiquitously. Despite the similarity to nuclear RNP proteins, both the nrp1 and xrp1 proteins are localized to the cytoplasm in the *Xenopus* oocyte. The *xrp1* gene may have a function in all cells that is similar to that executed by *nrp1* specifically within the nervous system.

INTRODUCTION

RNA binding proteins are involved in various aspects of RNA metabolism in the cell including RNA stability, RNA processing, and translation (1). A subset of the RNA binding proteins form a family that is identified by one or more copies of an 80–90 amino acid motif named the RNP motif or RNA recognition motif (RRM); the RRM has been shown to mediate RNA binding by these proteins (for review see ref. 2). This motif contains two conserved regions, an octapeptide RNP1 and a hexapeptide RNP2 (3,4). Many of the genes encoding RRMs are expressed in all or most cell types, but some appear to have a tissue specific or developmentally regulated function (3,5–10).

The heterogeneous nuclear ribonucleoproteins (hnRNPs) are defined as proteins that associate with hnRNA (pre-mRNA) in the nucleus. A family of core hnRNP proteins, the A, B and C type proteins, have been described as forming a complex with hnRNA (11); cDNAs encoding such proteins have been isolated from mammals (12–15), frogs (16,17), and insects (18–23). These proteins contain either 1 (type C) or 2 (type A/B) copies of the RRM in addition to auxiliary domains. The sequence of the two RRMs of hnRNP A/B proteins is conserved such that the first RRM is more closely related to the first RRM of other proteins than to the second RRM of the same protein (4,19,21). This conservation implies that the two RRMs have maintained distinct protein structures, presumably for specific functions such as preferential binding to distinct classes of RNA sequence (24–27). The original view of the role of core hnRNPs held that they are involved in packaging of newly transcribed RNA (11,28), but recent experiments suggested that these proteins also have a role in RNA processing (29–34).

Previously, we identified a gene, *nrp1*, that encodes a protein that is more similar in sequence to the hnRNP A/B proteins than to other previously described proteins (6). The nrp1 protein has a similar domain structure as the hnRNP A/B proteins with two RRMs in the amino terminal half. The *nrp1* gene is specifically expressed in the nervous system of *Xenopus*, except for a maternal RNA component accumulated in the egg. During development, *nrp1* is expressed in the neurectoderm after neural induction, and later is localized to the ventricular zone of the tadpole brain, a region where cell proliferation takes place (6,35).

We report here the use of PCR amplification with degenerate primers to isolate cDNAs related to both *nrp1* and the type A/B hnRNPs from *Xenopus*. This screen identified three previously unreported genes. One is the *Xenopus* homolog of the mammalian A2/B1 gene, and another is related to both the A1 and A2/B1 genes. The third gene, named *xrp1*, is similar to *nrp1*, yet is expressed in all cells and tissues of *Xenopus* that were tested. In *Xenopus* oocytes injected with either protein or the corresponding mRNA, the nrp1 and xrp1 proteins are localized to the cytoplasm.

MATERIALS AND METHODS

Isolation of PCR fragments from cDNAs encoding two RRMs

PCR amplification with degenerate primers was performed on cDNA template from either an embryonic stage 24/25 library or an adult brain library as described previously (36). Fully

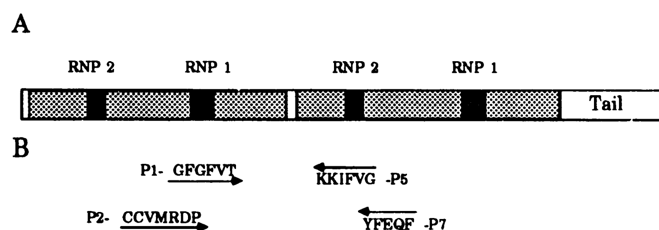

Figure 1. Domain structure of a generic RNP protein (3) with two RNA recognition motifs (RRMs). **(A)** The box represents the entire protein with the stippled boxes being the RRMs; the relative locations of RNP1 and RNP2 consensus sequences (2,4) are indicated. This figure is not drawn to scale. **(B)** Location and sequence of conserved amino acids used to design degenerate PCR primers.

degenerate primers corresponding to conserved amino acid sequences (Fig. 1) were synthesized that included an EcoRI site at their 5′ ends. The relevant amplified bands were identified by size, gel purified, and inserted into the EcoRI site of pBluescript KS(+). About 1000 of the resulting colonies were screened as described in Results.

Isolation and sequencing of cDNA inserts

Specific oligonucleotides were synthesized for the A2, A3 and *xrp1* genes and used to screen cDNA libraries for full-length cDNAs. Initially, a λZAP (Stratagene) cDNA library of RNA isolated from *Xenopus* XTC cells was screened and the inserts subcloned by helper phage-induced recombination. For *xrp1*, one insert, xrp1-3, was sequenced and is 1774 bp long, similar in size to the xrp1 mRNA detected on northern blots, and ends in a stretch of polyA (Accession #L02953). For the hnRNP A2, four cDNAs were sequenced and two classes of cDNAs (A2-7/A2-9 and A2-10/A2-11) were identified, presumably representing both copies of the hnRNP A2 gene present in the pseudotetraploid *Xenopus* laevis. Inserts A2-7 and A2-11 were sequenced on both strands with specific primers as described above. Inserts A2-9 and A2-10 were sequenced on one strand except where the sequence differed from inserts A2-7 and A2-11, respectively. In these areas, the opposite strand was also sequenced. For the hnRNP A3 gene, a single cDNA was sequenced and appeared to be a partial cDNA; a stage 24/25 embryonic cDNA library (35) was screened for full-length inserts. Three inserts of about 1.6 kb (A3-13, A3-16, and A3-17) were subcloned and sequenced on both strands and encode the entire presumed A3 protein. Two inserts are identical in sequence (A3-13 and A3-17) while the other insert (A3-16) differs slightly; again these two classes of cDNAs probably represent the two copies of the hnRNP A3 gene. All sequences were analyzed with the GCG package of computer programs (37).

Northern blot analyses

Northern blots of tissue and embryonic RNAs were performed as described previously (36,38). For the xrp1, nrp1, and hnRNP A1 probes, the entire cDNA was used as a probe and the blots were washed under stringent conditions (0.1×SSPE, 0.1% SDS; 65°C 2×15 min). For the hnRNP A2 and A3, specific probes were made from restriction fragments of the 3′ untranslated region. Blots were stripped of probe by washing twice with 10 mM Tris−Cl (pH 8.0), 1 mM EDTA, and 0.1% SDS at 95°C for 10 min followed by washing with the same solution at room temperature. The blots were exposed to film before reprobing to ensure sufficient probe removal.

In vitro translations

The appropriate inserts were subcloned into the EcoRI site of the plasmid pSD3, a vector that allows the synthesis of RNA with a polyA tail by SP6 RNA polymerase (39). The hnRNP A1 insert is from plasmid pXlA1a (17). Capped RNA was synthesized by including a 5-fold excess of 7^{me}GpppG in the transcription reactions (Promega). Following DNase treatment, RNA was translated according to the manufacturer's instructions in a rabbit reticulocyte lysate (Promega). Proteins were labeled by incorporation of ^{35}S-methionine, electrophoresed on 10% SDS/polyacrylamide gels, and detected by autoradiography.

Epitope-tagged version of nrp1

To make an epitope-tagged version of nrp1, a primer was synthesized that encodes 12 aa of a myc epitope (40) at its 5′ end and 18 nt of nrp1 sequence beginning with the first AUG codon at its 3′ end. This primer was used with an antisense primer from the 3′ UTR of nrp1 to amplify a modified nrp1 insert, and this DNA was inserted into pSD3.

Xenopus oocyte injections

In vitro translation products labeled with ^{35}S-methionine were purified and equilibrated with injection buffer by ultrafiltration (Amicon Centricon 10). Samples of either capped mRNA (30 nl of 0.5 μg/μl) or in vitro translated proteins (30 nl) were injected into the cytoplasm of stage VI oocytes (41). Protein-injected oocytes were harvested after 6h at 18°C; RNA-injected oocytes were labeled by incubation with ^{35}S-methionine (1 mCi/ml in OR2) beginning 4h after injection and then harvested after 24h at 18°C. Oocytes were manually dissected into cytoplasm and nuclear (GV) fractions, the proteins extracted, and analysed on SDS/polyacrylamide gels (42).

Primers for PCR and library screening

P1	GGAATTCCGGNTT[C,T]GGNTT[C,T]GTNAC
P2	GGAATTCTG[C,T]NTNGTNATG[G,T]GNGA[C,T]CC
P5	GGAATTCCNCCNAC[G,A]AANA[A,G,T][C,T]TT[C,T]TT
P7	GGAATTCAC[G,A]AA[C,T]TG[C,T]TC[G,A]AA[G,A]TA
nrp1	GGACCAAGCGGGAGTGGACAAAGT
Xl A1	TTCCTCTAGGCCTGGTGCACACC
xrp1	AGATCCTGCAAGTCTAGAT
Xl A2	CATGAATGAAGTTGACGCAG
Xl A3	CGCACAAAGTTGATGGCCG

RESULTS

PCR amplification of genes encoding proteins with two RRMs

Since the spacing of RRMs is invariant in hnRNP proteins with two RRMs (Fig. 1), we expected to amplify DNA of a discrete size from RNAs that encode at least two RRMs by using primers that span an RRM boundary. We aligned the available vertebrate and Drosophila hnRNP A/B protein sequences with the nrp1 protein and designed degenerate PCR primers based on conserved regions within the first and second RRMs (Fig. 1B). All combinations of these primers were used to amplify DNA from either embryonic or brain cDNA, and the products of the correct size were inserted into a plasmid vector. Initially, 48 randomly picked plasmids were sequenced to determine the identity and distribution of different sequences. Most inserts were either hnRNP A1 or nrp1 sequences. Next, an ordered array of 960

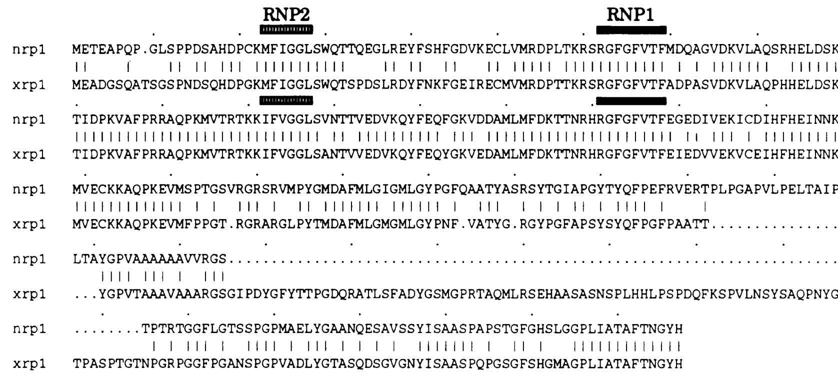

Figure 2. Comparison of the nrp1 and xrp1 protein sequences. The sequences were aligned with the GCG Gap program and the alignment optimized by hand. The lines between the sequences indicate identity, the dots represent gaps. The locations of RNP1 and RNP2 are indicated.

plasmids was screened by hybridization with the PCR fragment to identify plasmids with inserts, and counterscreened with oligonucleotide probes specific for hnRNP A1 and nrp1 to eliminate plasmids containing these abundant cDNAs. The remaining 145 inserts were sequenced. A total of five genes were identified. Two of these, hnRNP A1 and nrp1 were known; the three remaining genes are described below.

A sequence related to nrp1

One set of PCR fragments were similar, but not identical, in sequence to nrp1. We named the corresponding gene *xrp1* for *Xenopus* **R**NP **p**rotein. A full length cDNA (see below and Materials and Methods) derived from the *xrp1* gene was isolated and sequenced. The predicted xrp1 protein shares extensive sequence similarity with the nrp1 protein in the RRMs and also, to a lesser extent, in the carboxy terminal tail (Fig. 2 and Table 1). The major difference between the two proteins are insertions/deletions, resulting in a length of 406 aa for xrp1 as compared to 347 aa for nrp1. Whereas the tail of the core hnRNP A/B proteins is highly enriched for glycine, the tail of xrp1, as that of nrp1, shows no such enrichment although the frequency of glycine, alanine, tyrosine, and proline is about twice that of an average *Xenopus* protein. While there is little sequence similarity, this composition is somewhat reminiscent of that in the tail domain of the *Drosophila* hnRNP hrp48.1 (21).

Sequences related to the human hnRNP A2/B1 proteins

Two different sets of PCR fragments were closely related to the human hnRNP A/B sequences. Four apparently full-length cDNAs were isolated corresponding to one set of PCR products and the predicted protein was most similar to the human hnRNP A2/B1 proteins (Fig. 3, Table 1). The human A2/B1 proteins are most likely products of alternative splicing of one transcript (13); we refer to the predicted frog protein as *Xenopus* hnRNP A2 because the sequence of the amino terminus, where the human A2 and B1 proteins differ, is identical in all of the four predicted proteins to the A2 protein. To search for the expected B1-like RNA in *Xenopus*, we carried out PCR amplification with primers flanking the site where A2 and B1 isoforms differ. DNA from cDNA libraries of st I & II oocytes, stage 24 embryos, and brain was used as template. Only the product corresponding to the A2 isoform was observed (data not shown), suggesting that the B1

Table 1. Percent identity between regions of different proteins containing two RRMs.

	nrp1	xrp1	Xl A2	Hu A2	Xl A1	Hu A1	Xl A3
nrp1		81	41	41	40	39	39
xrp1	56		40	40	39	38	40
Xl A2	12	18		93	78	78	78
Hu A2	13	21	84		77	78	78
Xl A1	09	14	51	50		90	81
Hu A1	13	18	47	50	67		84
Xl A3	13	17	49	49	55	58	

Comparisons prepared by the Pileup and Distance programs of the GCG package are presented. Numbers above and below the line are comparisons of the RRM-containing amino terminus and the carboxy terminus, respectively. The protein sequences are described in Figures 2 and 3; Hu A1 is from ref.15

isoform may not exist in *Xenopus*. *Xenopus laevis* is a pseudotetraploid organism with two non-allelic copies of most genes; these copies are very similar but not identical, and usually both are expressed (43). On the basis of nucleic acid sequence we conclude that two of the hnRNP A2 cDNAs (A2-7 and A2-9; accession #L02954) represent one copy, while the other two cDNAs (A2-10 and A2-11; accession #L02955) represent the other copy. These cDNAs are 92% identical at the nucleic acid level while the predicted proteins are 95% identical.

Protein isoforms of the *Xenopus* hnRNP A2 protein

The four hnRNP A2 cDNAs each encode different protein isoforms. Two of these different isoforms are due to the presence of two copies of the gene in the pseudotetraploid *Xenopus laevis* genome, as described above. A more substantial difference appears to result from alternative splicing. Within the glycine-rich carboxy terminus, the A2-7 and A2-11 proteins line up with

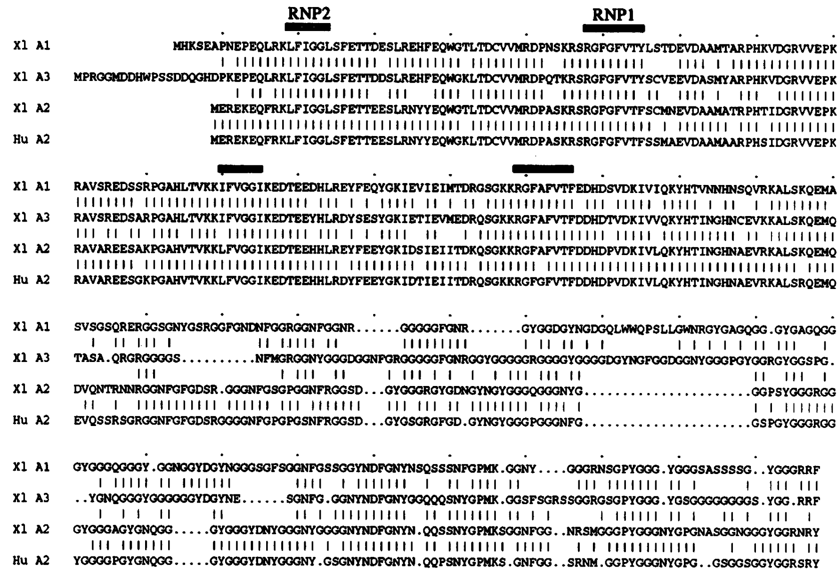

Figure 3. Comparison of the *Xenopus* and human hnRNP A/B protein sequences. The sequences were aligned and displayed as described in the legend to Figure 2. Human A2 (Hu A2; ref. 13), *Xenopus* A1 (Xl A1; ref. 17). The protein sequences for the *Xenopus* A2 and A3 proteins are translated from cDNAs A2-7 and A3-16, respectively.

the human hnRNP A2 protein (with the exception of an occasional single amino acid deletion or substitution) (Fig. 4A). However, the A2-9 and the A2-10 proteins have a 34 aa deletion and a 12 aa insertion, respectively, at the identical location relative to the human hnRNP A2 protein. In vitro translation of mRNA transcribed from these cDNA clones demonstrates that the A2-9 and A2-10 cDNAs encode isoforms with different molecular masses from the A2-7 and A2-11 cDNAs, consistent with their sequence (Fig. 4B). This result also shows that the A2 isoforms are translated as efficiently as the *Xenopus* hnRNP A1 and A3 proteins. Multiple isoforms of a similar nature, differing within the carboxy terminal domain, are seen in the *Xenopus* nrp1 (6), *Xenopus* hnRNP A1a/A1b (17), human hnRNP A1/A1^{B} (44), *Drosophila* hrp40.1/hrp40.2 (21), and human hnRNP C1/C2 proteins (13).

A novel hnRNP A/B protein

The sequence of the second set of PCR fragments was closely related to but distinct from both the *Xenopus* hnRNP A1 and A2 sequences; we named this gene the hnRNP A3 gene. Three nearly full-length cDNAs were isolated; two inserts, A3-17 and A3-13, represent one copy of the A3 gene (accession #L02957). While both these cDNAs encode the same protein and have 3′ terminal polyA stretches, the A3-13 insert ends 76 nt upstream of the A3-17 insert; both polyA termini are preceded by a canonical AATAAA polyadenylation site. Presumably, these cDNAs represent mRNAs with alternative polyadenylation sites. The third plasmid, A3-16 represents the other 'tetraploid' copy in the *Xenopus laevis* genome (accession #L02956). These two copies of the hnRNP A3 cDNA are 94% identical at the nucleic acid level, while the encoded proteins share 96% identity.

Comparisons between the hnRNP A/B class proteins (Fig. 3, Table 1) show that sequence identities are higher in the RRMs than in the tail domains, yet these proteins also share substantial homology in the tail. Both hnRNP A2 and A3 have glycine-rich carboxy terminal domains characteristic of the type A/B proteins. As is common among members of most protein families,

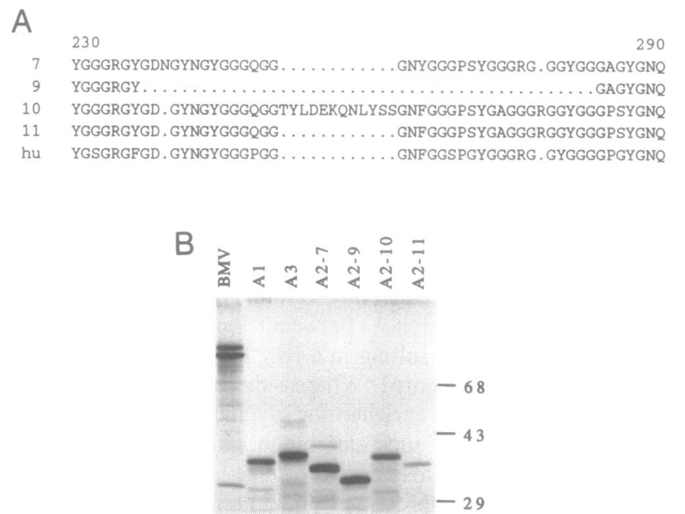

Figure 4. Multiple protein isoforms encoded by the hnRNP A2 gene. **(A)** Sequence in the relevant region of the predicted protein from different A2 cDNAs. The first four sequences are proteins encoded by cDNAs described in the text with #7 and #9 corresponding to one copy and #10 and #11 corresponding to the other copy of the A2 gene in the pseudotetraploid *X. laevis* genome. The sequence labeled hu corresponds to human A2. Dots indicate gaps. The numbering refers to the sequence of the A2-7 predicted protein. **(B)** Synthetic RNAs encoding the *Xenopus* hnRNP A1a (17), A3 (from the A3-16 cDNA), and four A2 cDNAs listed under (A) were translated in vitro as described in Materials and Methods. The protein products were electrophoresed on 10% SDS/polyacrylamide gels; BMV is Brome Mosaic Virus RNA used as control.

homologs in different species (e.g., frog and human A2 proteins) are more similar to each other than even closely related family members within the same species (e.g., frog A1, A2, and A3). The basis for classifying the *Xenopus* A3 protein as distinct is also apparent from Table 1; A3 is no more similar to *Xenopus* A1 and A2 than it is to human A1 and A2. Therefore it is possible that a mammalian A3 homolog remains to be discovered.

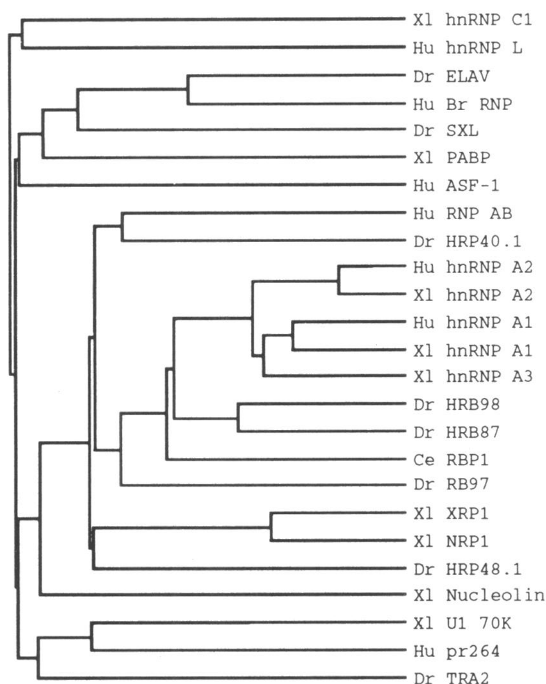

Figure 5. Dendrogram of a multiple sequence comparison of various RRM-containing proteins. Sequences derived from selected RRM-containing proteins were compared using the PILEUP program of the GCG package. The branch lengths are proportional to the similarity between sequences. The references or accession numbers are: Xl (*X.laevis*) hnRNP C1 (16), Hu (human) hnRNP L (57), Dr (*Drosophila*) ELAV (9), Hu Br RNP (7), Dr SXL (10), Xl PABP (M27072), Hu ASF-1 (M72709), Dr HRP40.1 (21), Hu RNP AB (46), Hu hnRNP A2 (13), Xl hnRNP A2 (this paper), Hu hnRNP A1 (15), Xl hnRNP A1 (17), Xl hnRNP A3 (this paper), Dr HRB98 (18), Dr HRB87 (19), Ce (*Caenorhabditis elegans*) RBP1 (45), Dr RB97 (54), Xl XRP1 (this paper), Xl NRP1 (6), Dr HRP48.1 (21), and Xl Nucleolin (X63091), Xl U1 70K (X12430), Hu pr264 (X62447), Dr TRA2 (M30939).

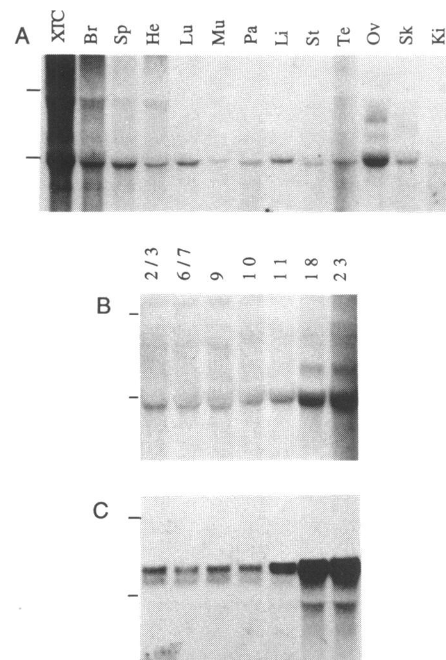

Figure 6. Expression of the *xrp1* gene. **(A)** Northern blot with 10 μg of RNA from different adult tissues probed with an xrp1 cDNA probe. XTC, XTC cell line; Br, brain; Sp, spleen; He, heart; Lu, lung; Mu, skeletal muscle; Pa, pancreas; Li, liver; St, stomach; Te, testis; Ov, ovary; Sk, skin; Ki, kidney. **(B,C)** Total RNA (10 μg) from staged embryos was analyzed by Northern blots probed with either the xrp1 (B) or nrp1 (C) probes. The lanes are labeled with the embryo stage. Migration of 28S and 18S RNA is indicated at left.

The relationship of these newly-identified proteins to other RRM-containing proteins is demonstrated by a pairwise comparison presented in Figure 5. The hnRNP A2 and A3 together with hnRNP A1 form a subgroup of hnRNPs distinct from but related to *Drosophila* hnRNPs (19,21), the *Caenorhabditis elegans* rbp1 protein (45), and the human RNP AB (46). Both nrp1 and xrp1 are most similar to a *Drosophila* hnRNP protein hrp48.1 (21), with all three proteins being most similar to the hnRNP A/B class. Other RRM-containing proteins, including the nervous system-specific *Drosophila* elav (9) and human brain RNP (7) proteins, are more distantly related to both the hnRNP A/B and the nrp/xrp proteins.

Conserved 3′ untranslated region of the hnRNP A2 mRNA

A comparison of the 3′ untranslated region (UTR) from the human and *Xenopus* A2 hnRNP mRNAs indicates a surprising degree of similarity (83% identity) between these sequences over a length of approximately 500 nt (not shown). Comparing the 3′ UTR of the hnRNP A1 and C1 mRNAs from human and *Xenopus* reveal no similarity above random, although the 3′ UTRs from rat and human hnRNP A1 are highly conserved. In addition, a comparison of the 3′ UTR of the human and *Xenopus* poly(A) binding protein also indicates a strong sequence similarity (75% identity over 400 nt of 3′ UTR comparing accession #Y00345 to #M27072). This conservation of sequence within the UTR implies a functional role in regulation of mRNA stability or translation.

Expression pattern of the *xrp1* gene

To determine if *xrp1* is expressed in a tissue-specific manner, a northern blot containing RNA from different adult tissues was probed with an xrp1 cDNA probe. An approximately 1.8 kb RNA is detected that is present in all tissues tested, including XTC cells (Fig. 6A). This contrasts with the nervous system specificity of its close relative *nrp1* (35). Northern blots with RNA from different developmental stages indicated that *xrp1* is expressed throughout development, increasing in levels after neurulation (Fig. 6B). This increase paralleled the increase in RNA levels seen for *nrp1* (Fig. 6C). The abundance of the xrp1 RNA is at least 5-fold less than that of the nrp1 RNA as seen from comparative northern blots (data not shown).

Coordinate expression of the hnRNP A/B genes

RNA blotting showed that the hnRNP A1, A2, and A3 genes are expressed maternally and throughout embryogenesis (data not shown). All three genes are expressed in all adult tissues that were tested, although some quantitative variations in expression level are seen (Fig. 7). For all three genes, the ovary and late embryonic stages contained the highest level of RNA, suggesting a requirement for hnRNP proteins during rapid cell proliferation (embryogenesis), or in preparation for proliferation (oogenesis). This result is consistent with expression of hnRNP genes in *Drosophila* (18,19) and with the observation that hnRNP protein levels are high in proliferating cells (47). Direct comparisons

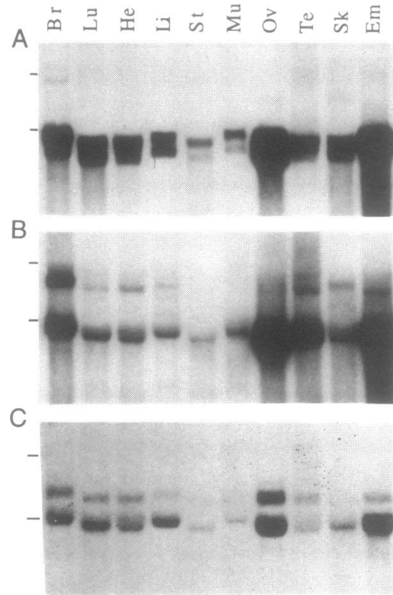

Figure 7. Comparison of *Xenopus* hnRNP A/B gene expression in different adult tissues. Northern blots containing 10 μg of total RNA were probed with **(A)** the complete hnRNP A1 cDNA, **(B)** the 3′ UTR of the hnRNP A2 cDNA, and **(C)** the 3′ UTR of the hnRNP A3 cDNA. Tissues are indicated as in Figure 6; Em is RNA from stage 16/17 embryos.

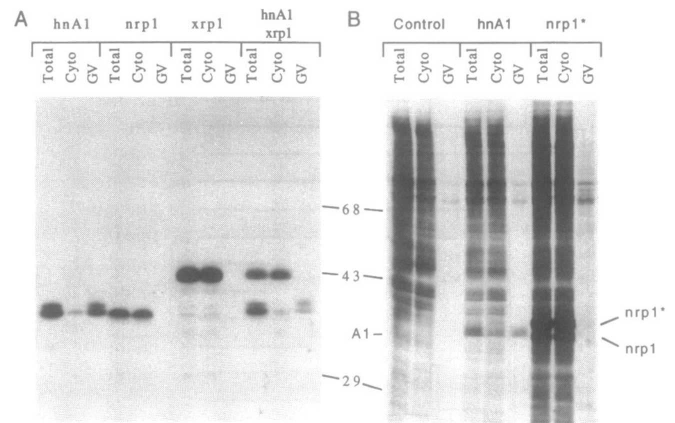

Figure 8. Localization of nrp1 and xrp1 proteins in *Xenopus* oocytes. Autoradiogram of proteins from injected stage VI oocytes, extracted from whole oocytes (total) or manually dissected nuclear (GV) and cytoplasmic (Cyto) fractions. **(A)** Oocytes were injected with the indicated labeled proteins and analysed as described in Materials and Methods. After incubation for 6 hr, the proteins were separated by gel electrophoresis. **(B)** Oocytes were injected with mRNA synthesized in vitro that encode the indicated proteins. After labeling the oocytes with ^{35}S-methionine, proteins were separated by electrophoresis. The nrp1* mRNA is translated into both the epitope-nrp1 fusion protein (nrp1*) and the natural nrp1 protein, yielding two closely-spaced bands. The migration of the hnRNP A1 and the nrp1 proteins is indicated.

between these RNAs indicate that the hnRNP A3 gene is expressed at approximately 10-fold lower levels than the hnRNP A1 or A2 genes, which are expressed at similar levels (data not shown). Thus, since the mRNAs are translated at equivalent efficiencies (Fig. 4), the hnRNP A3 protein is a minor component of hnRNP complexes.

All A/B class hnRNP gene probes detected multiple RNA species in all adult tissues. Two A1 RNA species are seen, approximately 1.4 and 1.7 kb. The predominant A2 RNA species is approximately 1.9 kb, with a minor RNA of about 3.2 kb. Two RNAs of 1.7 and 2.4 kb are detected with the A3 gene-specific probe. Small variations in mobility of RNA species between lanes is a gel artefact since the same filters probed for *Xenopus* elongation factor-1α demonstrate parallel variations (data not shown). While the nature of these additional RNA species, whose ratio varies with tissue type, is not known, these differences may be the result of alternative polyadenylation as seen in human A1 (15) and *Drosophila* Hrb98DE (18) and Hrb87F (19) mRNAs.

Cytoplasmic localization of the nrp1 and xrp1 proteins

The similarity in sequence of both the nrp1 and xrp1 proteins to the core hnRNP A/B proteins suggested that these proteins may be hnRNP proteins involved in nuclear RNA processing. To test this hypothesis, we examined the intracellular localization of the nrp1 and xrp1 proteins in *Xenopus* oocytes by two separate experiments (Fig. 8). First, labeled proteins synthesized in vitro were injected into the oocyte cytoplasm, and after an incubation of 6h the oocytes were fractionated into nucleus (germinal vesicle or GV) and cytoplasm. As expected, the majority of the injected hnRNP A1 protein migrated into the nucleus. However, neither nrp1 or xrp1 protein was found in the nucleus, even though the proteins appeared quite stable in the oocyte (Fig. 8A). When A1 and xrp1 proteins were coinjected, each behaved as it did in singly injected ooytes. In the second approach, synthetic mRNAs were injected into oocytes, the proteins labeled by incubation with ^{35}S-methionine, and the oocytes dissected into nuclear and cytoplasmic fractions; the products were then analyzed by gel electrophoresis. We used an epitope tagged version of nrp1 for this experiment (see Materials and Methods) because this RNA was translated more efficiently than unmodified nrp1 or xrp1 mRNA. Again, hnRNP A1 was efficiently transported to the nucleus as expected; in contrast, the efficiently translated nrp1 protein remained in the cytoplasm and could not be detected in the nuclear fraction (Fig. 8B). Similar results were obtained for xrp1 translated from its mRNA after injection in the oocyte (data not shown). We conclude that, at least in the *Xenopus* ooctye, both nrp1 and xrp1 are cytoplasmic proteins.

DISCUSSION

Isolation of cDNAs for proteins containing two RRMs from *Xenopus*

The use of PCR amplification with degenerate primers enables one to identify and isolate genes that share various degrees of similarity. With this approach, we have isolated cDNAs corresponding to three previously unreported genes encoding putative RNA binding proteins in *Xenopus*. All three predicted proteins contain two RRMs, as expected from the design of the PCR primers. We do not feel that this was an exhaustive search for cDNAs encoding RNA binding proteins with two RRMs in *Xenopus*, since the approach used is subject to several variables including the match of the primers to the actual sequence, and the abundance of the target cDNA. A different PCR screen for RRM sequences from *Drosophila* and plants yielded several distinct sequences (23,48).

The three genes identified here include the *Xenopus* hnRNP A2 gene, a gene encoding a putative core hnRNP protein named A3, and the *xrp1* gene that is similar to the previously studied

nrp1 gene. All three predicted proteins have the same domain structure characteristic of hnRNP A/B proteins. The predicted hnRNP A2 and A3 proteins also share extensive sequence similarity, not only within the conserved RNP1 and RNP2 sequences but throughout the RRM. In addition both proteins share glycine-rich tails with other hnRNP A/B proteins. Given its similarity to other core hnRNP proteins and its ubiquitous tissue distribution, A3 most likely is a core hnRNP protein. In contrast, the predicted xrp1 protein, along with nrp1, is similar to the hnRNP A/B proteins mainly in the conserved RNP1 and RNP2 sequences and in a few other residues, while the carboxy terminal domain does not contain a single predominant amino acid.

Cytoplasmic RNP proteins expressed in *Xenopus*

The *xrp1* gene is closely related to the nervous system-specific *nrp1* gene yet is expressed in all tissues and developmental stages tested. Both xrp1 and nrp1 are most similar to nuclear RNA binding proteins (hnRNPs; Fig. 5), and because of this sequence similarity nrp1 was classified with the hnRNPs in our original report (6) as well as in a recent compilation (23). On the basis of the results shown in Figure 8 this conclusion appears to be incorrect, casting doubt on the functional classification of RRM proteins solely on the basis of sequence relationships. The nrp1 and xrp1 proteins may regulate events in RNA metabolism that occur in the cytoplasm such as localization of specific mRNAs within the cytoplasm, transport of RNAs after their exit from the nucleus, cytoplasmic polyadenylation, and regulation of translation. Alternatively, these proteins may after all function in the nucleus, but nuclear localization may occur only at certain developmental stages or in certain tissues. A number of proteins contain nuclear localization signals that are regulated in different tissues and developmental stages (49,50). The most dramatic example of regulated localization is provided by the product of the *dorsal* gene, which is nuclear in ventral cells but cytoplasmic in dorsal cells of *Drosophila* embryos (49,50). In addition, the localization of a protein may depend on the synthesis of its target RNA, as is the case with some snRNP proteins whose localization depends on the expression of their corresponding snRNA (51). In the case of nrp1 and xrp1 proteins, however, preliminary experiments (unpublished observations) have suggested a cytoplasmic localization in *Xenopus* embryos, similar to the situation in oocytes.

Are the functions of *nrp1* and *xrp1* similar? A possible hypothesis in favor of such an interpretation holds that the xrp1 protein has a ubiquitous function in RNA metabolism, while the nrp1 protein carries out the analogous function in the nervous system. If the nrp1 and xrp1 proteins have in fact similar functions, it remains unclear why the brain should express both; however, it is possible that the expression of *nrp1* and *xrp1* is cell type-specific within the nervous system, leading to non-overlapping patterns for the two genes. Alternatively, *nrp1* may have evolved a more specific function that is needed for a specialized RNA processing event that occurs only in the nervous system. In either case, identification of other proteins and/or RNA species that interact with the nrp1 or xrp1 protein should provide clues to their functions.

Diversity of core hnRNP proteins in *Xenopus*

We propose that the *Xenopus* hnRNP A2 gene is the homolog of the human A2/B1 gene. The *Xenopus* A2 gene encodes proteins with alternative peptide inserts within the glycine-rich domain; similar alternative products are seen in the human A1 gene (17). However, the difference between the mammalian B1 and the A2 protein is a 12 aa insert near the amino terminus (13); as described above, the corresponding alternative product appears to be absent from *Xenopus*. With the identification of the A3 gene it appears that *Xenopus* contains at least three closely related hnRNP genes whereas only two have been identified in mammals within this subgroup (the A1 and A2/B1 genes). It is possible that the frog A3 protein may perform the function served by the B1 isoform in mammals. Alternatively, the A3 gene may have evolved separately in a lineage leading to *Xenopus*, or the corresponding mammalian gene may exist. For example, the A3 protein may be the homolog of the human B2 protein for which no cDNA has been identified.

Monoclonal antibodies against hnRNP proteins detected multiple reactive proteins in *Xenopus* cell lines (21,52). We do not know whether the *Xenopus* cDNAs isolated so far (i.e., A1, A2, A3, and C1/C2) account for this diversity. A minor core hnRNP protein with two RRMs, the human hnRNP AB protein (earlier referred to as the hnRNP C3 protein (46,53)), is only 37% identical to the *Xenopus* hnRNP A3 protein. The human hnRNP AB gene has a different domain structure in that the amino terminal sequence upstream of the RRMs is longer than in the hnRNP A/B proteins, and it lacks a long glycine-rich carboxy terminal domain (46,53). It remains to be seen if a *Xenopus* homolog of this gene exists. In *Drosophila*, a family of at least five genes encodes the major hnRNP core proteins, with no clear indication from sequence similarity to suggest which gene may be the homolog of which mammalian gene (18,19,21,54). Using an experimental definition of an hnRNP as the association of a protein with nuclear pre-mRNA, a large number of hnRNP proteins have been described in humans and flys (55,56), including some that are encoded by the hnRNP A/B family, others with RRMs diverged from the hnRNP A/B type (57,58), and still others that do not contain RRMs (59,60).

ACKNOWLEDGEMENTS

We thank Brian Kay for the *Xenopus* hnRNP A1 plasmid, Michael Lai for help with the A3 cDNA characterizations, and Sue Haynes and Gopa Raychaudhuri for discussions and comments on the manuscript.

REFERENCES

1. Dreyfuss, G., Philipson, L. and Mattaj, I.W. (1988) J. Cell Biol. 106, 1419–1425.
2. Kenan, D.J., Query, C.C. and Keene, J.D. (1991) Trends Biochem. Sci. 19, 214–220.
3. Bandziulis, R.J., Swanson, M.S. and Dreyfuss, G. (1989) Genes and Dev. 3, 431–437.
4. Dreyfuss, G., Swanson, M.S. and Piñol-Roma, S. (1988) Trends Biochem. Sci. 13, 86–91.
5. Ayane, M., Preuss, U., Köhler, G. and Nielsen, P.J. (1991) Nucl. Acids Res. 19, 1273–1278.
6. Richter, K., Good, P.J. and Dawid, I.B. (1990) New Biol. 2, 556–565.
7. Szabo, A., Dalmau, J., Manley, G., Rosenfeld, M., Wong, E., Henson, J., Posner, J.B. and Furneaux, H.M. (1991) Cell 67, 325–333.
8. Lantz, V., Ambrosio, L. and Schedl, P. (1992) Development 115, 75–88.
9. Robinow, S., Campos, A.R., Yao, K.-M. and White, K. (1988) Science 242, 1570–1572.
10. Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988) Cell 55, 1037–1046.
11. Dreyfuss, G. (1986) Annu. Rev. Cell Biol. 2, 459–498.
12. Cobianchi, F., SenGupta, D.N., Zmudzka, B.Z. and Wilson, S.H. (1986) J. Biol. Chem. 261, 3536–3543.

13. Burd, C.G., Swanson, M.S., Görlach, M. and Dreyfuss, G. (1989) Proc. Natl. Acad. Sci. 86, 9788–9792.
14. Nakagawa, T.Y., Swanson, M.S., Wold, B.J. and Dreyfuss, G. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 2007–2011.
15. Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M.T., Ghetti, A., Riva, S. and Morandi, C. (1988) Nucl. Acids Res. 16, 3751–3770.
16. Preugschat, F. and Wold, B. (1988) Proc. Natl. Acad. Sci. 85, 9669–9673.
17. Kay, B.K., Sawhney, R.K. and Wilson, S.H. (1990) Proc. Natl. Acad. Sci. 87, 1367–1371.
18. Haynes, S.R., Raychaudhuri, G. and Beyer, A.L. (1990) Mol. Cell. Biol. 10, 316–323.
19. Haynes, S.R., Johnson, D., Raychaudhuri, G. and Beyer, A.L. (1991) Nucl. Acids Res. 19, 25–31.
20. Ball, E.E., Rehm, E.J. and Goodman, C.S. (1991) Nucleic Acids. Res. 19, 397–397.
21. Matunis, E.L., Matunis, M.J. and Dreyfuss, G. (1992) J. Cell Biol. 116, 257–269.
22. Hovemann, B.T., Dessen, E., Mechler, H. and Mack, E. (1991) Nucl. Acids Res. 19, 4909–4914.
23. Kim, Y.-J. and Baker, B.S. (1993) Mol. Cell. Biol. 13, 174–183.
24. Swanson, M.S. and Dreyfuss, G. (1988) Mol. Cell Biol. 8, 2237–2241.
25. Minoo, P., Martin, T.E. and Riehl, R.M. (1991) Biochem . Biophy. Res. Comm. 176, 747–755.
26. Nietfeld, W., Mentzel, H. and Pieler, T. (1990) EMBO J. 9, 3699–3705.
27. Burd, C.G., Matunis, E.L. and Dreyfuss, G. (1991) Mol. Cell Biol. 11, 3419–3424.
28. LeStourgeon, W.M., Barnett, S.F. and Northington, S.J. (1990) in The eukaryotic nucleus: Molecular biochemistry and macromolecular assemblies, Strauss, P.R. and Wilson, S.H. Eds. pp. 477–502, Telford Press, Caldwell,NJ.
29. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G. and Reed, R. (1992) Mol. Cell Biol. 12, 3165–3175.
30. Choi, Y.D., Grabowski, P.J., Sharp, P.A. and Dreyfuss, G. (1986) Science 231, 1534–1539.
31. Mayeda, A. and Krainer, A.R. (1992) Cell 68, 365–375.
32. Piñol-Roma, S. and Dreyfuss, G. (1992) Nature 355, 730–732.
33. Piñol-Roma, S. and Dreyfuss, G. (1991) Science 253, 312–314.
34. Wilusz, J. and Shenk, T. (1990) Mol. Cell Biol. 10, 6397–6407.
35. Richter, K., Grunz, H. and Dawid, I.B. (1988) Proc. Natl. Acad. Sci. USA 85, 8086–8090.
36. Taira, M., Jamrich, M., Good, P.J. and Dawid, I.B. (1992) Genes and Dev. 6, 356–366.
37. Devereux, J., Haeberli, P. and Smithies, O. (1984) Nulceic Acids Res. 12, 387–395.
38. Good, P.J., Richter, K. and Dawid, I.B. (1990) Proc. Natl. Acad. Sci. 87, 9088–9092.
39. Good, P.J., Welch, R.C., Barkan, A., Somasekhar, M.B. and Mertz, J.E. (1988) J. Virol. 62, 944–953.
40. Kolodziej, P.A. and Young, R.A. (1991) Methods Enzymol. 194, 508–519.
41. Kay, B.K. (1991) Methods Cell Biol. 36, 663–669.
42. Evans, J.P. and Kay, B.K. (1991) Methods Cell Biol. 36, 133–148.
43. Graf, J.-D. and Kobel, H.R. (1991) Methods Cell Biol. 36, 19–34.
44. Buvoli, M., Cobianchi, F., Bestagno, M.G., Mangiarotti, A., Bassi, M.T., Biamonti, G. and Riva, S. (1990) EMBO J. 9, 1229–1235.
45. Iwasaki, M., Okumura, K., Kondo, Y., Tanaka, T. and Igarashi, H. (1992) Nucl. Acids Res. 20, 4001–4007.
46. Khan, F.A., Jaiswal, A.K. and Szer, W. (1991) FEBS Lett. 290, 159–161.
47. Minoo, P., Sullivan, W., Solomon, L.R., Martin, T.E., Toft, D.O. and Scott, R.E. (1989) J. Cell Biol. 109, 1937–1946.
48. Mieszczak, M., Klahre, U., Levy, J.H., Goodall, G.J. and Filipowicz, W. (1992) Mol. Gen. Genet. 234, 390–400.
49. Richter, J.D. and Standiford, D. (1992) in Nuclear trafficking, Feldherr, C.M. Ed. pp. 89–120, Academic Press, San Diego, CA.
50. Yamasaki, L. and Lanford, R.E. (1992) in Nuclear trafficking, Feldherr, C.M. Ed. pp. 121–174, Academic Press, San Diego, CA.
51. Zeller, R., Nyffenegger, T. and DeRobertis, E.M. (1983) Cell 32, 425–434.
52. Leser, G.P., Escara-Wilke, J. and Martin, T.E. (1984) J. Biol. Chem. 259, 1827–1833.
53. Kumar, A., Sierakowska, H. and Szer, W. (1987) J. Biol. Chem. 262, 17126–17137.
54. Karsch-Mizrachi, I. and Haynes, S.R. (1993) Nucl. Acids Res. (manuscript submitted)
55. Matunis, M.J., Matunis, E.L. and Dreyfuss, G. (1992) J. Cell Biol. 116, 245–255.
56. Piñol-Roma, S., Choi, Y.D., Matunis, M.J. and Dreyfuss, G. (1988) Genes Dev. 2, 215–227.
57. Piñol-Roma, S., Swanson, M.S., Gall, J.G. and Dreyfuss, G. (1989) J. Cell Biol. 109, 2575–2587.
58. Ghetti, A., Piñol-Roma, S., Michael, W.M., Morandi, C. and Dreyfuss, G. (1992) Nucl. Acids Res. 20, 3671–3678.
59. Matunis, M.J., Michael, W.M. and Dreyfuss, G. (1992) Mol. Cell. Biol. 12, 164–171.
60. Kiledjian, M. and Dreyfuss, G. (1992) EMBO J. 11, 2655–2664.
61. Biamonti, G., Buvoli, M., Bassi, M.T., Morandi, C., Cobianchi, F. and Riva, S. (1989) J. Mol. Biol. 207, 491–503.