Nucleic Acids Research, 1993, Vol. 21, No. 3 439–446

The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins

Kshama V.Datar, Gideon Dreyfuss[1] and Maurice S.Swanson*
Department of Immunology and Medical Microbiology, College of Medicine, University of Florida, Gainesville, FL 32610-0266 and [1]Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104-6148, USA

Received November 10, 1992; Revised and Accepted December 19, 1992 GenBank accession no. L03532

ABSTRACT

Recent reports indicate that proteins which directly bind to nascent RNA polymerase II transcripts, the heterogeneous nuclear ribonucleoproteins (hnRNPs), play an important role in both transcript-specific packaging and alternative splicing of pre-mRNAs. Here we describe the isolation and characterization of a group of abundant hnRNPs, the M1–M4 proteins, which appear as a cluster of four proteins of 64,000–68,000 daltons by two-dimensional electrophoresis. The M proteins are pre-mRNA binding proteins *in vivo*, and they bind avidly to poly(G) and poly(U) RNA homopolymers *in vitro*. Covalently associated polyadenylated RNA-protein complexes, generated by irradiating living HeLa cells with UV light, were purified and used to elicit antibodies in mice. The resulting antisera were then employed to isolate cDNA clones for the largest M protein, M4, by immunological screening. The deduced amino acid sequence of M4 indicates that the M proteins are members of the ribonucleoprotein consensus sequence family of RNA-binding proteins with greatest similarity to a hypothetical RNA-binding protein from *Saccharomyces cerevisiae*. The M proteins also possess an unusual hexapeptide-repeat region rich in methionine and arginine residues (MR repeat motif) that resembles a repeat in the 64,000 dalton subunit of cleavage stimulation factor, which is involved in 3′-end maturation of pre-mRNAs. Proteins immunologically related to M exist in divergent eukaryotes ranging from human to yeast.

INTRODUCTION

More than twenty abundant nuclear RNA-binding proteins associate with heterogeneous nuclear RNA (hnRNA) following the initiation of transcription by RNA polymerase II (1–3). It has been proposed that these proteins, heterogeneous nuclear ribonucleoproteins (hnRNPs), package hnRNAs in a transcript-specific manner so that the nuclear machinery required for specific pre-mRNA processing events, such as small nuclear ribonucleoprotein particles (snRNPs) (4), recognizes each transcript as a unique entity (5–9). This possibility would help explain why many of these proteins bind RNA in a sequence-specific fashion (10–13). The processes of alternative RNA processing, including splicing, polyadenylation, and nucleocytoplasmic transport of mRNA, might begin with this transcript-specific packaging by proteins of the hnRNP complex. Experimental evidence that hnRNPs may play a direct role in constitutive pre-mRNA splicing has been provided for the hnRNP C (14) and A/B proteins (15), while the relative concentrations of the hnRNP A1 protein and splicing factor SF2 appear to influence splice site choice *in vitro* (16). Recent studies have also demonstrated that several hnRNPs including the A1, D, C1/C2, and I proteins (also termed the polypyrimidine tract-binding protein or PTB) (11–13) bind to pre-spliceosomes *in vitro* with sequence preference and in a transcript-dependent manner (9). Due to the apparent important role that hnRNPs play in nuclear RNA metabolism, it is therefore essential to isolate and characterize individual members of this family to begin to determine their various nuclear functions.

During the last several years, we have been studying human hnRNP complexes immunopurified from nucleoplasm using monoclonal antibodies to previously authenticated hnRNPs (17, 18). Prominent among the proteins associated with hnRNA isolated using this technique is a cluster which migrates at 68,000 daltons by SDS–PAGE, the L and M proteins (18). The L protein is composed of four ~80 amino acid domains related to the ribonucleoprotein consensus sequence RNA-binding domain (CS-RBD) (6, 13, 19), and a considerable amount of L is detectable outside of hnRNP complexes (7). Monoclonal antibodies to the L proteins not only detect discrete non-nucleolar structures within the nucleus of human cells, but L appears to be able to discriminate among different RNA polymerase II transcripts since it binds at high concentration to the 'giant loops' of amphibian lampbrush chromosomes (7). Here we characterize the hnRNP complex M proteins, and identify an unusual protein structural motif in heterogeneous nuclear ribonucleoproteins.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Tissue culture, subcellular fractionation, and immunopurifications

HeLa (JW36 and S3) and mouse (3T3) cells were grown to subconfluent densities in Dulbecco's modified Eagle's medium (DMEM) supplemented with 1% penicillin/streptomycin and 10% calf serum. Rabbit kidney cells (RK13) were grown in Eagle's MEM with Earles's salts, and chicken embryo fibroblasts in Medium 199 with Earle's salts, supplemented with 1% penicillin/streptomycin and 5% fetal bovine serum. *Schizosaccharomyces pombe* cells were grown to $O.D._{600} = 1.0$ in YES media (20). HeLa cells were labeled with either ^{35}S-methionine or TRAN^{35}S at 20 μCi/ml for 12 hr in the presence of 10% of the normal methionine level and 5% calf serum. Nucleoplasm was prepared by sonication of isolated nuclei as described previously (17). Immunopurifications of both HeLa S3 cell hnRNP complexes, and *in vitro* translated M proteins were performed as described previously (7, 17).

Preparation of polyclonal antisera, monoclonal antibodies, and isolation of cDNA clones

UV crosslinked polyadenylated RNA-protein complexes were prepared and used to raise antibodies in Balb/c mice as described previously (21, 22). Polyclonal antisera were used to screen both total HeLa S3 proteins and immunopurified hnRNP complexes by immunoblotting. Antisera from one mouse recognized the M proteins, and was used to screen a λgt11 HeLa cell cDNA library using a 1:200 dilution of the antisera. Positive plaques were isolated, purified and subcloned into pGEM1 (Promega). One of these expression clones, pHCMex1 (encoding amino acids 286–531, see Figure 6), was sequenced and subsequently employed as a hybridization probe to isolate a full-length clone, pHCM4, from a human osteosarcoma λZAP cDNA library (23). To identify cDNA clones encoding the M1/M2 proteins, the λgt11 expression library was screened with both 1D8 and 2A6 and those clones reacting with 1D8, but not 2A6, were purified and subcloned as described above. One of these clones, pHCM1/2, contained a precise deletion of nucleotides 486–603 or amino acid residues 159–197 (see Figure 6).

For the preparation of monoclonal antibodies, the entire M4 cDNA clone was excised using XbaI from pGEM 1 (Promega) and recloned in-frame into the T7 gene 10 fusion expression vector pGEMX (Promega). The resulting fusion protein consists of the entire M4 protein sequence and six 'bridge' amino acids positioned between the gene 10 and M4 proteins (CTA/LEU, GAG/GLU, GAC/ASP, GCG/ALA, GAG/GLU, AAA/LYS). IPTG-induced fusion protein was purified by SDS–PAGE, electroeluted, and dialyzed against PBS prior to injection into Balb/c mice using RIBI's (MPL/TDM) adjuvant (RIBI Immunochemical Research). Following three injections of 50 μg/injection, hybridomas were produced using SP2/O myeloma cells, and hybridoma supernatants screened by ELISA, immunoblotting, and cellular immunofluorescence (22). The majority of hybridomas (480), represented by the 1D8 monoclonal, were reactive against all four M proteins, M1–M4, while only two hybridomas, represented by 2A6, reacted against only M3 and M4. Monoclonal antibody heavy chain subclasses are IgG1 (1D8) and IgG2b (2A6).

Gel electrophoresis and immunoblotting

Proteins were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), using 12.5% (final acrylamide concentration) separation gels (24). For visualization of ^{35}S-labeled proteins by fluorography, Coomassie Blue-stained gels were impregnated with 2,5-diphenyloxazole. Two-dimensional NEPHGE was performed according to O'Farrell (25) with separation in the first dimension accomplished using a pH 3–10 ampholine gradient for 4 hr at 400 V and the second dimension by SDS–PAGE. Electroblotting was done using a semi-dry apparatus (Hoefer) for 1 hr at 100 mA. Immunoblots were probed with either 1D8 or 2A6 at 1:1000 dilution and a goat anti-mouse HRP-conjugated secondary antibody at 1:5000 dilution. Reactive antigens were visualized by enhanced chemoluminescence (Amersham) with film exposures of 2–45 sec.

RNA blot analysis, hybrid selection, and *in vitro* transcription and translation

Polyadenylated RNAs were purified from HeLa S3 cells using two cycles of selection on oligo(dT) agarose, and fractionated on 1.4% formaldehyde agarose gels as described (26). Following transfer to nitrocellulose, blots were hybridized with the isolated XbaI insert from pHCM4. Hybrid selection of polyadenylated RNAs using the pHCMex1 insert and *in vitro* translations in rabbit reticulocyte lysates were done as described (26). *In vitro* transcriptions/translations of pHCM4 were performed using T7 polymerase according to the manufacturer's instructions (Promega).

RNA homopolymer binding

RNA homopolymer binding experiments using HeLa cell nucleoplasmic proteins, or the *in vitro* translation product of pHCM4, to poly(G), poly(A), poly(U), and poly(C) were performed as previously described (10) with the exception that the cells were not labeled with ^{35}S-methionine prior to subcellular fractionation. Proteins bound to the RNA homopolymers at 0.1–1.0 M NaCl were resolved by SDS–PAGE, and immunoblotting done as described above using 1D8 and 2A6, to detect the M proteins. To study the RNA binding characteristics of the *in vitro* synthesized M4 protein, 1 μl of an ^{35}S-methionine labeled reticulocyte translation mixture primed with pHCM4 mRNA was used for each salt concentration.

DNA sequencing and protein analysis

DNA sequences were determined from both strands using restriction fragments cloned into M13mp18 and M13mp19, as described (27, 28). Sequence information was analyzed using the University of Wisconsin Genetics Computer Group programs, and database searches were done using the BLAST network service at the National Center for Biotechnology Information (29). The nucleotide sequence of the pHCM4 cDNA has been listed in the EMBL/Genbank/DDBJ Nucleotide Sequence Databases under the accession number L03532.

RESULTS

Isolation of cDNA clones and monoclonal antibodies for the hnRNP M proteins

The M proteins were originally characterized as abundant members of immunopurified human hnRNP complexes (17, 18). These proteins migrate by two-dimensional gel electrophoresis as a cluster of four proteins (M1, M2, M3, M4) of 64–68 kD with pI values (7.8–8.2) slightly more basic than the L proteins (68 kD, pI = 7.4–7.7) (18). Similar to the hnRNP A, B, and C proteins, the M proteins appear to be tightly associated with

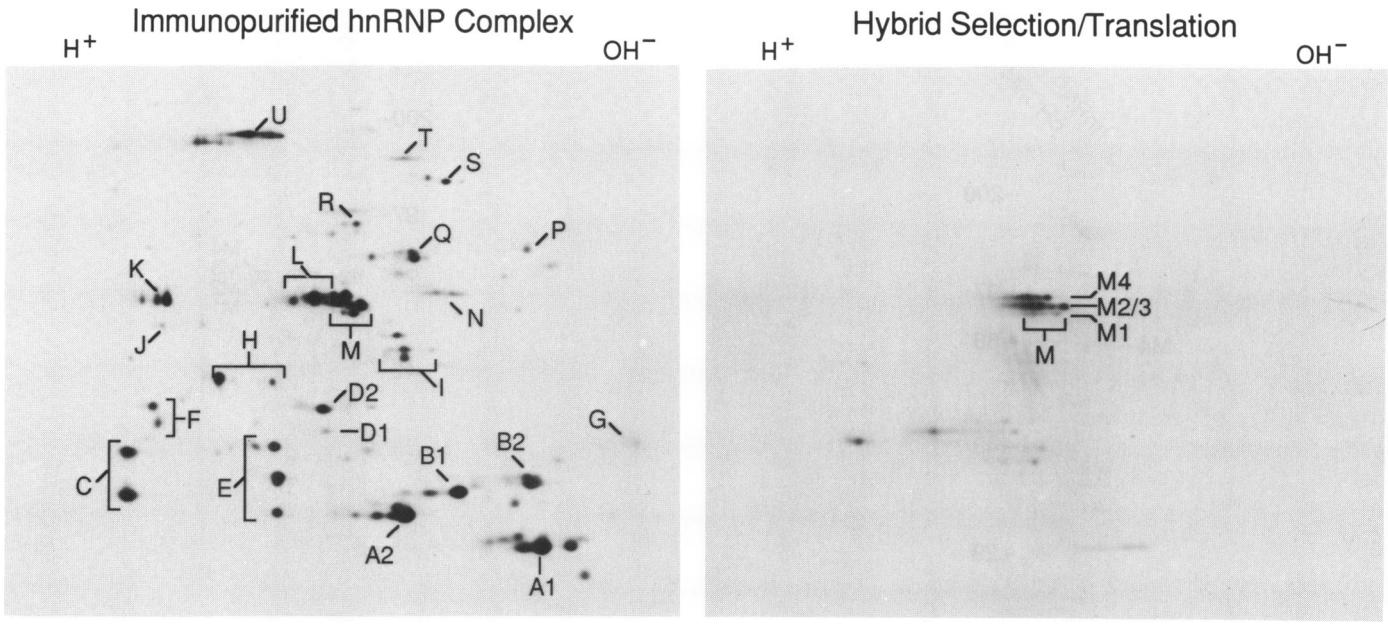


Figure 1. Identification of a cDNA clone for the M proteins. (Left Panel) HeLa cells were labeled with ^{35}S-methionine, hnRNP complexes immunopurified with a monoclonal antibody specific for the hnRNP C proteins (17, 22), proteins resolved by NEPHGE (25), and detected by fluorography. The L proteins, visible at ~68,000 daltons, have been previously characterized (7), and are separate from the indicated M proteins. (Right Panel) Hybrid selection was performed using the pHCMex1 cDNA clone and HeLa poly(A)$^+$ RNA, and the specifically bound mRNA translated *in vitro* in a rabbit reticulocyte lysate in the presence of ^{35}S-methionine. The expected positions of the four M proteins are indicated. Although M2 and M3 are resolved by prolonged electrophoresis, they co-migrate under the conditions employed here. All other labeled polypeptides are produced in the reticulocyte lysate without exogenously added RNA (data not shown).

hnRNAs since immunopurification of hnRNP complexes in the presence of heparin eliminates many of the other high molecular weight hnRNPs from the complex (18). Furthermore, the M proteins can be crosslinked to nuclear polyadenylated RNAs in living HeLa cells (17, 21), and specifically bind *in vitro* to the RNA homopolymers poly(G) and poly(U) in a salt-resistant manner (10, see below). These results demonstrate that the M proteins are tightly associated with pre-mRNAs *in vivo*, and suggest that they may play an important role(s) in the pre-mRNA processing pathway.

A detailed characterization of the hnRNP complex M proteins was initiated by producing polyclonal antibodies in Balb/c mice using crosslinked polyadenylated RNA-protein complexes. These complexes were generated by irradiating HeLa cells with UV-light and purified by oligo(dT) chromatography under protein denaturing conditions (21). Immunoblotting of both total HeLa cell proteins and immunopurified hnRNP complexes using antisera from one of these mice showed strong reactivity against several proteins in the 64−68 kd range. Isolation of cDNA clones encoding this group of proteins was accomplished by screening a λgt11 expression library as described in Materials and Methods, and several positive clones were selected, plaque purified and the cDNA inserts isolated. These cDNAs were then used to hybrid select RNA from a total polyadenylated RNA pool, and the mRNAs were subsequently translated into protein. Figure 1 demonstrates that one of these clones, pHCMex1, hybrid selected mRNA which translated into a set of 64−68 kD proteins with pI values of 7.3−8.2 which co-migrated with the hnRNP complex M proteins. The observation that pHCMex1 hybrid-selects polyadenylated RNA which translates into all four M proteins indicates that all of the M proteins are highly related to each other. Another group of proteins whose migration by two-dimensional gel electrophoresis partially overlaps the M proteins is the L proteins. However, the proteins encoded by pHCMex1 were not immunopurified by a monoclonal antibody specific for L, 4D11 (7), and L protein cDNA clones did not cross-hybridize with pHCMex1 (data not shown).

Three different experiments were performed to confirm that the pHCMex1 clone encodes an authentic M protein. First, full-length cDNA clones were obtained by hybridization screening of a λZAP human cDNA library. Figure 2 shows that *in vitro* transcription/translation of one of these clones, pHCM4, produced a protein which co-migrates by one-dimensional SDS−PAGE with the largest M protein, M4. Second, two different monoclonal antibodies were prepared from a mouse immunized with a T7 gene 10-M4 cDNA fusion protein. These monoclonals recognized proteins which co-migrated with either all four M proteins (Figure 3, panel 1D8) or only the two largest M proteins (Figure 3, panel 2A6). Third, hnRNP complexes were immunopurified from unlabeled HeLa cells with a monoclonal antibody (4F4) against another group of hnRNPs, the C1/C2 proteins (17). As shown above in Figure 1 (panel labeled 'immunopurified complexes') this technique results in the purification of all of the major hnRNP complex proteins, including the M proteins. Figure 4 demonstrates that 1D8 recognizes four proteins, and 2A6 the largest two proteins, in both total HeLa cell extracts (Figure 4, lanes total) and in immunopurified hnRNP complexes (Figure 4, lanes complex). In addition, a two-dimensional gel analysis of proteins immunopurified with 1D8 from TRAN^{35}S-labeled HeLa cell nucleoplasm, indicates that this monoclonal recognizes proteins which co-migrate with the M proteins (data not shown). Finally, as recently shown by Carmo-Fonseca *et al.* (30), using both 1D8 and 2A6, the M proteins are localized predominantly in the

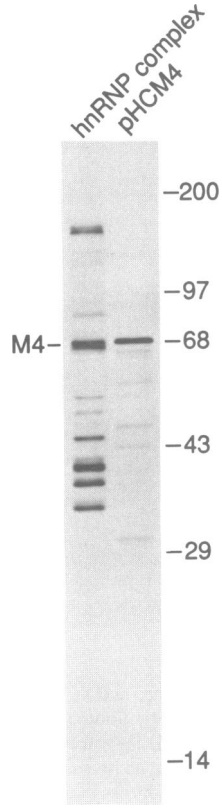


Figure 2. pHCM4 encodes the largest M protein. hnRNP complexes were immunopurified from ^{35}S-methionine-labeled HeLa cells, and the proteins detected by fluorogaphy following fractionation by SDS−PAGE (lane, hnRNP complex). To identify the M protein encoded by pHCM4, BamHI-truncated plasmid was transcribed using T7 Polymerase, the RNA translated *in vitro*, and proteins loaded into the lane (lane, pHCM4) adjacent to the immunopurified hnRNP complex. The predominant polypeptide at ~68,000 daltons comigrates with the largest M protein, M4. Polypeptides migrating below M4 are probably the result of the use of internal methionines as initiation codons since the 5′-untranslated region of pHCM4 is only 11 nt in length.

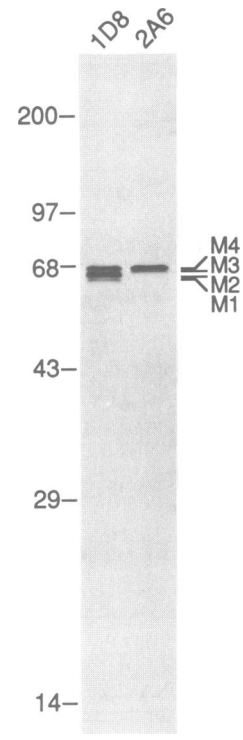


Figure 3. Monoclonal antibodies which recognize different M proteins. Nucleoplasm was prepared from TRAN^{35}S-labeled HeLa cells, and the M proteins immunopurified with either the 1D8 (lane 1D8) or 2A6 (lane 2A6) monoclonal antibodies in the presence of ionic detergent Empigen BB, which dissociates protein−protein and protein−RNA interactions without perturbing antigen−antibody interactions (17, 22). The 1D8 monoclonal immunopurifies all four M proteins while 2A6 recognizes only M3 and M4. Labeled proteins were detected by fluorography.

nucleus with a general nucleoplasmic staining pattern similar to that seen with monoclonals to other hnRNP complex proteins including A1 (18), C1/C2 (22), K/J (31) and U (21) proteins.

The M proteins possess multiple RNP-CS RNA binding domains and a methionine-rich repeat motif

RNA blot analysis of polyadenylated HeLa cell RNA, using the pHCM4 cDNA clone as a hybridization probe, detects only a single-size class of mRNA of ~2.7 kb (Figure 5). Since the pHCM4 cDNA clone hybrid selects mRNA which encodes all four M proteins (Figure 1), and because the M1−M4 mRNAs are very similar in size, the most likely interpretation, is that alternative pre-mRNA splicing generates all four proteins. In agreement with this idea, alternative splicing has previously been shown to generate multiple forms of the hnRNP complex A, B, C and I proteins which only differ by small peptide insertions (13, 23, 32).

The DNA sequence determined for the cDNA insert of pHCM4 of 2516 bp is shown in Figure 6. The cDNA sequence consists of an initiation codon in an excellent context for translational initiation, AAAATGG, 11 nt of 5′ end untranslated sequence and 257 nt of 3′ end untranslated sequence followed by a poly(A) tail of 35 nt. The predicted primary structure of M4 is a protein composed of 729 amino acids with a molecular mass of 77,606. Even though the M4 protein migrates at 68 kD by SDS−PAGE, anomalous migration of hnRNP complex C1/C2 (26), L (7), and U (33) proteins by SDS−PAGE has also been observed. Although we have been unsuccessful at isolation of M protein cDNAs with more 5′ untranslated sequence, the comigration of the *in vitro* transcription/translation product from the pHCM4 clone with the largest M protein, M4, indicates that pHCM4 encodes a full-length M4 protein. To investigate the possibility that multiple mRNAs are produced by alternative splicing of the M protein pre-mRNA, 2A6 and 1D8 were used to differentially screen a cDNA expression library, and only the cDNAs which reacted with 1D8, but not with 2A6, were further analyzed. One clone (pHCM1/2) was isolated which contained an in-frame deletion of amino acid residues 159−197 located between CS-RBD I and II (see below) (Figure 6, shaded box). This would result in a protein of 690 residues, with a calculated molecular mass of ~4 kD less than M4, which should comigate with M1.

Computer analysis of the M4 protein revealed that it possesses three RNP-CS RNA-binding domains of ~90 amino acids each. Database searches using the BLAST network service (29) revealed that M4 is most highly related to a hypothetical 48.7 kD protein determined from sequence analysis of Chromosome

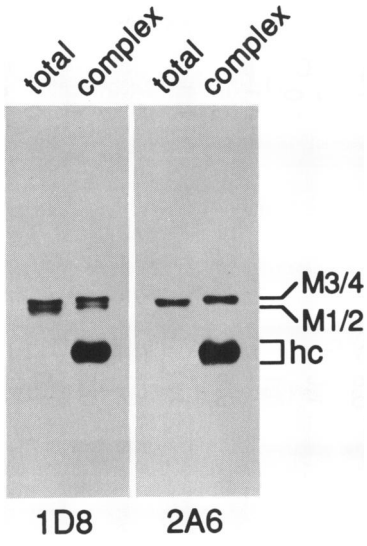


Figure 4. All four M proteins are authentic components of hnRNP complexes. Following immunopurification of the hnRNP complex with a monoclonal antibody to the C1/C2 proteins, 4F4 (17, 21), the proteins were resolved by SDS−PAGE, blotted onto nitrocellulose, and immunoblotted with either 1D8 (panel 1D8, lane complex) or 2A6 (panel 2A6, lane complex). To determine if the majority of the M proteins recognized by 1D8 and 2A6 in nucleoplasm are present in hnRNP complexes, an equivalent amount of total nucleoplasm (lanes total) was run in lanes adjacent to the immunopurified complex. Electroblotting of M proteins does not produce a clear separation of M1 from M2 (M1/2) and M3 from M4 (M3/4). The large band appearing below the M proteins in the lanes marked 'complex' is the IgG heavy chain (hc) from the 4F4 monoclonal antibody used to immunopurify the hnRNP complex.

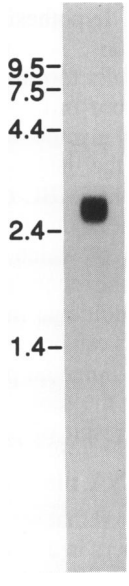


Figure 5. The M proteins are all produced from similarly-sized mRNAs. HeLa cell polyadenylated RNAs were resolved on a 1.4% formaldehyde agarose gel, blotted onto nitrocellulose, and probed with pHCM4. The major hybridizing band of ~2.7 kb must contain the mRNAs for all four M proteins since hybrid selection/translation using the pHCMex1 clone resulted in the production of M1−M4 (see Figure 1).

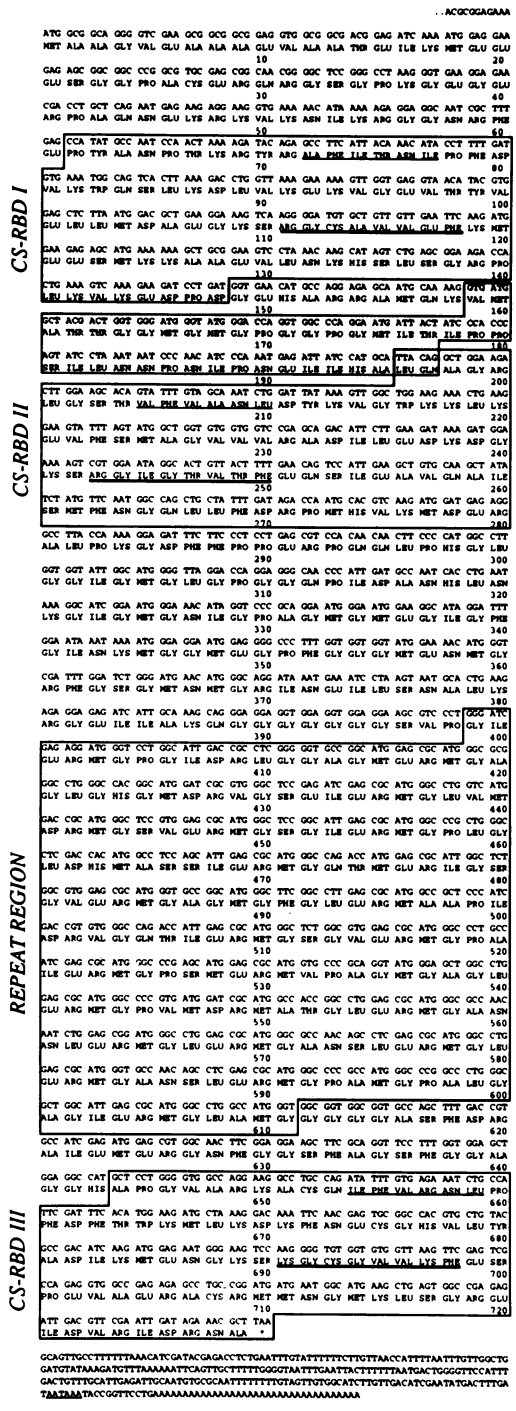
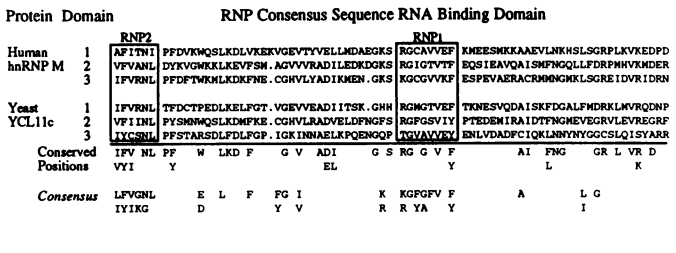


Figure 6. DNA and deduced amino acid sequences of pHCM4. Both DNA strands were sequenced from overlapping restriction fragments. The pHCM4 cDNA contains 11 nt of 5′, and 295 nt of 3′, untranslated sequence including a polyadenylation signal (underlined AATAAA) and a poly(A) tail 35 nt in length. The ribonucleoprotein consensus sequence RNA-binding domains (CS-RBDs, open boxes) were identified by the BLAST Network Service as being related to known CS-RBDs of other ribonucleoproteins. The most highly conserved motifs within the CS-RBDs (2, 6), RNP1 (also referred to as RNP-CS, thick lines) and RNP2 (thin lines), are indicated. Ihe shaded box designates the region missing in the pHCM1/2 clone. Situated between CS-RBDs II and III lies a region rich in glycine and methionine residues (between boxes marked CS-RBD II and Repeat Region) and the MR repeat motif (box marked Repeat Region).

A.



B.

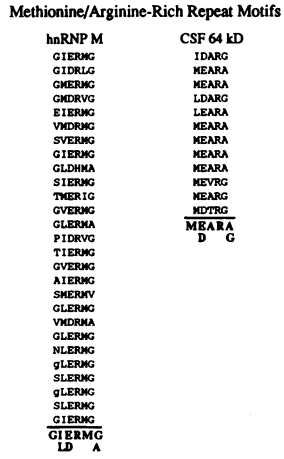
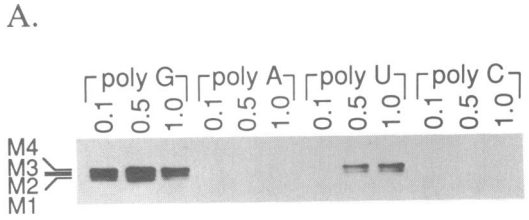


Figure 7. Distinguishing structural features of the M proteins. (**A**) The RNP consensus sequence RNA-binding domains of the human hnRNP M proteins and the yeast 48.7 kD putative RNA-binding protein are compared. Both proteins possess three CS-RBDs which are ~40% identical to one another. Positions in which one or two amino acids are repeated in 4 out of the 6 domains are indicated below the line as conserved positions. The original CS-RBD consensus (6) is also shown. (**B**) The methionine/arginine-rich repeat motifs of the hnRNP complex M proteins (27 repeats) and the 64 kD subunit of cleavage stimulation factor (CSF, 12 repeats) are compared. Glycines indicated by lower case (g) overlap with immediately upstream repeats (N^{561}LERMG/gLERMG^{571}).

III of *Saccharomyces cerevisiae* (34) (Figure 7A) and to the *Xenopus laevis* polyadenylate-tail binding protein (35) (alignment is not shown). The yeast putative RNA-binding protein also possesses three RNP-CS RNA binding domains which are all about 40% identical to the three RNA-binding domains of the M proteins. The yeast and human proteins are not significantly structurally related outside of the RNA-binding domains. In contrast to other hnRNPs whose primary sequence has been detected to date, the M4 protein contains an unusual motif rich in methionine and arginine residues (MR repeat motif), with a consensus of G(I/L)(E/D)RM(G/A), which is repeated 27 times between residues 399−611 (Figure 6, open box marked Repeat Region, and Figure 7B). Amino terminal to this repeat motif is another region (Figure 6, residues 281−398) rich in glycine and methionine with several GMG repeats. Recently, Takagaki *et al.* (36) have characterized a cleavage stimulation factor (CSF) subunit protein of 64,000 daltons, which is required for correct 3′-end cleavage and polyadenylation of mammalian pre-mRNAs. This CSF factor not only possesses a RNP-CS RNA-binding domain (CS-RBD), but a methionine/arginine-rich pentapeptide motif, M(E/D)AR(A/G), which is repeated 12 times within a 328 amino acid region rich in glycine and proline residues (Figure

A.



B.

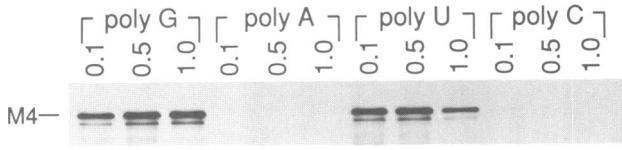


Figure 8. The M proteins bind to poly(G) and poly(U) homopolymers. (**A**) Immunoblot analysis using the 1D8 monoclonal antibody to detect the M proteins in HeLa cell nucleoplasm which specifically bind to poly(G), poly(A), poly(U) and poly(C) RNA homopolymers at the indicated salt concentrations. Longer film exposures demonstrate that all four M proteins bind to poly(U) at 0.5 M and 1.0 M NaCl. (**B**) Same as (A) except that the M4 protein encoded by pHCM4 was produced by *in vitro* transcription/translation in the presence of ^{35}S-methionine, and directly visualized by fluorogaphy.

7B). Although the primary structures of the repeat motifs of the CSF factor and the M proteins share only a limited degee of homology, they are both composed primarily of methionine, arginine, glycine/alanine, and glutamic/aspartic residues, and secondary structure predictions indicate that they are both rich in potential α-helical structure. The repeat region in the CSF factor, which has been hypothesized to be important for protein−protein interactions, is bordered by regions rich in glycine and proline potentially forming multiple β-turns. The 212 amino acid MR repeat motif in the M proteins begins with 7, and ends with 5, sequential glycine residues (Figure 6). No other proteins which contain the hexapeptide repeat motif were discovered in the Genbank/EMBL databanks. In addition, the M proteins contain nine potential casein kinase type II phosphorylation sites (six of which are within the MR repeat motif), and one tyrosine phosphorylation site within CS-RBD I (K^{93}VGEVTY^{99}). Since additional multiple discrete isoforms of each of the four M proteins can be seen in Figure 1, it is possible that, given the number of potential phosphorylation sites in the M proteins, that many of the more acidic forms seen on two-dimensional gels are the result of phosphorylation.

The M proteins have RNA binding preferences

Previous work has suggested that several hnRNPs not only bind specific RNA homopolymers in a salt-resistant manner, but also associate preferentially with different regions of pre-mRNAs, and with different types of pre-mRNAs, *in vitro* (5, 6, 9, 10). For instance, the C proteins are poly(U)-binding proteins (10) and preferentially associate with U-rich regions of pre-mRNAs within introns and downstream of the polyadenylation signal, AAUAAA (37). To characterize the RNA binding properties of the M proteins *in vitro*, HeLa cell nucleoplasm was isolated, and incubated with RNA homopolymers at increasing salt concentrations. The specifically bound proteins were then fractionated by SDS−PAGE, and finally immunoblotted using

the 1D8 monoclonal against all four M proteins. In agreement with previous work (10), Figure 8A shows that all four M proteins bind to both poly(G) and poly(U) up to at least 1 M NaCl although they do not appear to bind well to poly(U) at 100 mM NaCl. Figure 8B demonstrates that the *in vitro* transcription/translation product of the pHCM4 cDNA clone also binds to poly(G) and poly(U) in a nearly identical manner with the exception that M4 alone appears to bind to poly(U) at 100 mM NaCl. This last result suggests that at this low salt concentration other proteins in nucleoplasm may be competing with the M proteins for poly(U) binding.

Proteins immunologically-related to M exist in divergent eukaryotes

Several hnRNPs appear to have been remarkably conserved among a wide array of eukaryotic species (7, 38). The M proteins also appear to be highly conserved with immunologically cross-reacting proteins detectable in rabbit, mouse, chicken and yeast (Figure 9). Remarkably, the M proteins migrate as multiple closely-spaced polypeptides of ~68 kD in human, rabbit, and chicken whereas the mouse and *Schizosaccharomyces pombe* proteins appear as single polypeptides of 68 kD and 40 kD, respectively. The strong reactivity of 1D8 with a protein(s) from yeast is particularly surprising since none of the monoclonal antibodies raised against other human hnRNPs cross-react with yeast proteins although its authenticity as an M protein homologue has yet to be established.

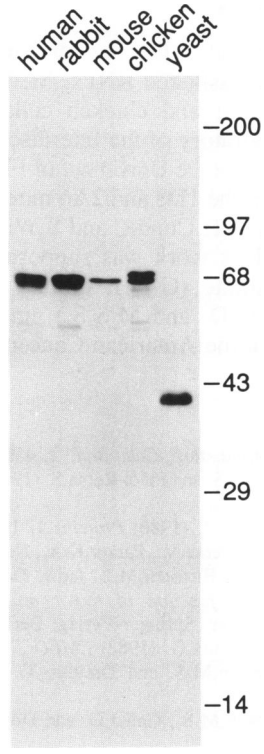


Figure 9. Proteins immunologically-related to hnRNP M exist in eukaryotes from human to yeast. Total proteins (~40 μg/lane) from human (HeLa), rabbit (RK13), mouse (3T3), chicken (CEF), and yeast (*S.pombe*) cells were fractionated by SDS−PAGE, blotted to nitrocellulose, and immunoblotted with 1D8.

DISCUSSION

When human cells are labeled with ^{35}S-methionine, and hnRNP complexes isolated either by immunopurification or by sucrose-gradient centrifugation, among the most prominent detectable hnRNPs are the L and M proteins, which migrate at ~68,000 daltons by SDS−PAGE (1, 2, 18). We demonstrate here that the M proteins possess a highly repeated methionine/arginine-rich (MR) hexapeptide motif related to an MR repeat motif in a CSF subunit factor. In contrast to this unusual structural feature, many other characteristics of the M proteins have been previously identified in proteins of the hnRNP complex. M is a member of the RNP consensus sequence family of RNA-binding proteins which includes the hnRNPs A1 (32, 39), A2/B1 (23), C1/C2 (26), I (13), and L (7). Other distinguishing features which the M proteins have in common with previously characterized hnRNPs are their ability to bind to RNA homopolymers in a salt-resistant manner (10), the existence of multiple protein isoforms probably generated by both the addition of small peptide insertions and post-translational modifications (23, 32), and their apparent evolutionary conservation.

In addition to the M1/M2/M3/M4 proteins, the hnRNP complex A1, A2/B1, C1/C2, I, L, K/J, and U proteins (the slash indicates that more than one protein appears to be produced from the same gene) have been isolated and characterized. One unexpected observation deduced from the primary stuctures of these proteins is that they not only possess different types of RNA-binding domains, which include both the CS-RBD (6, 19) and RGG box (33) motifs, but also an assortment of unusual auxiliary domains (6). In conjunction with sequence-specific RNA-binding domains, auxiliary domains have been suggested to facilitate specific protein−protein interactions, which may allow overall hnRNP complex composition to be tailored to individual hnRNAs (8, 9).

Why do both the M proteins and the CSF factor contain MR repeat motifs within their auxiliary domains? One of the distinguishing features of the auxiliary domains of the A/B group of hnRNPs is that they possess an unusual post-translational modification, dimethylarginine. This type of novel post-translational modification may provide an essential mechanism to modulate both protein−protein and protein−RNA interactions important within the RNP family, and that may be why arginine, and not lysine, is used exclusively as a basic amino acid within the MR repeat. Methionine-rich regions also appear to play an important role in ribonucleoproteins. A methionine-rich motif (M domain) has been described for a subunit of the signal recognition particle (SRP54). This region encompasses the COOH-terminal 207 amino acids of SRP54, and includes both a SRP7S RNA-binding domain and a signal sequence interaction site (40−42). Although this region of SRP54 does not possess an obvious primary structure repeat, secondary structure predictions have led to the suggestion that the M domain forms a series of amphipathic α-helices in which the polar faces interact with signal sequences and the non-polar faces with SRP7S RNA (42, 43). Other RNPs that possess regions high in methionine residues include the polyadenylate-binding protein (amino acids 433−472, carboxy terminal to CS-RBD IV) (6) and two U1 snRNA-binding proteins, A (amino acids 144−195, between CS-RBDs I and II) (44) and C (amino acids, 90−159) (45). Although we do not yet know whether the MR repeat motif can directly bind to other ribonucleoproteins and/or RNA, information on the function(s) of this particular auxiliary domain may provide insight into how hnRNP complexes form on hnRNA, and the dynamics of hnRNP complex assembly and disassembly.

ACKNOWLEDGEMENTS

We thank S.A.Adam and Y.D.Choi for the mouse polyclonal antisera against UV-crosslinked RNPs, M.Duke and R.Moyer for providing the rabbit and chicken cells, L.Green of the Hybridoma Core Laboratory of the Interdisciplinary Center for Biotechnology Research at the University of Florida for assistance with the preparation of the 1D8 and 2A6 monoclonal antibodies, and J.Anderson, J.Aris, R.Condit, and S.Wilson for comments on the manuscript. This work was supported by the Howard Hughes Medical Institute (G.D.), grants from the National Institutes of Health (G.D. and M.S.S.), and a Junior Faculty Research Award from the American Cancer Society (M.S.S.).

REFERENCES

1. Dreyfuss,G. (1986) *Annu. Rev. Cell. Biol.* **2**, 459–498.
2. Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1988) *Trends Biochem. Sci.* **13**, 86–91.
3. Chung, S.Y. and Wooley,J. (1986) *Proteins* **1**, 195–210.
4. Steitz,J.A., Black,D.L., Gerke,V., Parker,K.A., Kramer,A., Frendewey,D. and Keller,W. (1988) In, Birnsteil,M.L. (ed.), *Functions of the abundant snRNPs. Structure and function of major and minor small nuclear ribonucleoprotein particles.* Springer-Verlag, Berlin, pp. 115–154.
5. Swanson,M.S. and Dreyfuss,G. (1988) *EMBO J.* **7**, 3519–3529.
6. Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) *Genes Dev.* **3**, 431–437.
7. Piñol-Roma,S., Swanson,M.S., Gall,J.G. and Dreyfuss,G. (1989) *J. Cell Biol.* **109**, 2575–2587.
8. Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1990) In Strauss,P.R. and Wilson,S.H. (eds), *The Eukaryotic Nucleus: Structure and Function.* Telford Press, Caldwell,NJ, Vol. 2, pp. 503–517.
9. Bennett,M., Piñol-Roma,S., Staknis,D., Dreyfuss,G. and Reed,R. (1992) *Mol. Cell. Biol.* **12**, 3165–3175.
10. Swanson,M.S. and Dreyfuss,G. (1988) *Mol. Cell. Biol.* **8**, 2237–2241.
11. Gil,A., Sharp,P.A., Jamison,S.F. and Garcia-Blanco,M. (1991) *Genes Dev.* **5**, 1224–1236.
12. Patton,J.G., Mayer,S.A., Tempst,P. and Nadal-Ginard,B. (1991) *Genes Dev.* **5**, 1237–1251.
13. Ghetti,A., Piñol-Roma,S., Michael,W.M., Morandi,C. and Dreyfuss,G. (1992) *Nucleic Acids Res.* **20**, 3671–3678.
14. Choi,Y.D., Grabowski,P.J., Sharp,P.A. and Dreyfuss,G. (1986) *Science* **231**, 1534–1539.
15. Sierakowska,H., Szer,W., Furdon,P.J. and Kole,R. (1986) *Nucleic Acids Res.* **14**, 5241–5254.
16. Mayeda,A. and Kramer,A.R. (1992) *Cell* **68**, 365–375.
17. Choi,Y.D. and Dreyfuss,G. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7471–7475.
18. Piñol-Roma,S., Choi,Y.D., Matunis,M.J. and Dreyfuss,G. (1988) *Genes Dev.* **2**, 215–227.
19. Kenan,D.J., Query,C.C. and Kenan,J.D. (1991) *Trends Biochem. Sci.* **16**, 214–220.
20. Moreno,S., Klar,A. and Nurse,P. (1991) *Methods Enzymol.* **194**, 795–823.
21. Dreyfuss,G., Choi,Y.D. and Adam,S.A. (1984) *Mol. Cell. Biol.* **4**, 1104–1114.
22. Choi,Y.D. and Dreyfuss,G. (1984) *J. Cell Biol.* **99**, 1997–2004.
23. Burd,C.G., Swanson,M.S. and Dreyfuss,G. (1989) *Proc. Natl. Acad. Sci USA* **86**, 9788–9792.
24. Dreyfuss,G., Adam,S.A. and Choi,Y.D. (1984) *Mol. Cell. Biol.* **4**, 415–423.
25. O'Farrell,P.Z., Goodman,H.M. and O'Farrell,P.H. (1977) *Cell* **12**, 1133–1142.
26. Swanson,M.S., Nakagawa,T.Y., LeVan,K. and Dreyfuss,G. (1987) *Mol. Cell. Biol.* **7**, 1731–1739.
27. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
28. Messing,J. (1983) *Methods Enzymol.* **101**, 20–78.
29. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
30. Carmo-Fonseca,M., Pepperkok,R., Sproat,B.S., Ansorge,W., Swanson,M.S. and Lamond,A.I. (1991) *EMBO J.* **10**, 1863–1873.
31. Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1992) *Mol. Cell. Biol.* **12**, 164–171.
32. Buvoli,M., Cobianchi,F., Bestagno,M.G., Mangiarotti,A., Bassi,M.T., Biamonti,G. and Riva,S. (1990) *EMBO J.* **9**, 1229–1235.
33. Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.* **11**, 2655–2664.
34. Oliver,S.G., van der Aart,Q.J.M., Agostoni-Carbone,M.L., Aigle,M., Alberghina,L. *et al.* (1992) *Nature* **357**, 38–46.
35. Zelus,B.D., Giebelhaus,D.H., Eib,D.W., Kenner,K. and Moon,R.T. (1989) *Mol. Cell. Biol.* **9**, 2756–2760.
36. Takagaki,Y., MacDonald,C.C., Shenk,T. and Manley,J.L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1403–1407.
37. Wilusz,J., Feig,D.I. and Shenk,T. (1988) *Mol. Cell. Biol.* **8**, 4477–4483.
38. Matunis,E.L., Matunis,M.J. and Dreyfuss,G. (1992) *J. Cell Biol.* **116**, 257–269.
39. Kumar,A., Casas-Finet,J.R., Luneau,C.J., Karpel,R.L., Merrill,B.M., Williams,K.R. and Wilson,S.H. (1990) *J. Biol. Chem.* **265**, 17094–17100.
40. Romisch,K., Webb,J., Lingelbach,K., Gausepohl,H., and Dobberstein,B. (1990) *J. Cell Biol.* **111**, 1793–1802.
41. Zopf,D., Bernstein,H.D., Johnson,A.E. and Walter,P. (1990) *EMBO J.* **9**, 4511–4517.
42. High,S. and Dobberstein,B. (1991) *J. Cell Biol.* **113**, 229–233.
43. Bernstein,H.D., Poritz,M.A., Strub,K., Hoben,P.J., Brenner,S. and Walter,P. (1989) *Nature* **340**, 482–486.
44. Sillekens,P.T.G., Habets,W.J., Beijer,R.P. and van Venrooij,W.J. (1987) *EMBO J.* **6**, 3841–3848.
45. Sillekens,P.T.G., Beijer,R.P., Habets,W.J. and van Venrooij, W.J. (1988) *Nucleic Acids Res.* **16**, 8307–8321.