Nucleic Acids Research, 1994, Vol. 22, No. 6 1059–1067

The hnRNP F protein: unique primary structure, nucleic acid-binding properties, and subcellular localization

Michael J.Matunis, Jun Xing+ and Gideon Dreyfuss*
Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6148, USA

Received November 23, 1993; Revised and Accepted February 2, 1994

ABSTRACT

More than 20 different heterogeneous nuclear ribonucleoproteins (hnRNPs) are associated with pre-mRNAs in the nucleus of mammalian cells and these proteins appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport. The arrangement of hnRNP proteins on pre-mRNAs is likely to be unique for each RNA and may be determined by the different RNA-binding preferences of each of these proteins. hnRNP F (M_r = 53 kD, pI = 6.1) and hnRNP H (M_r = 56 kD, pI = 6.7–7.1) are abundant components of immunopurified hnRNP complexes and they have distinct nucleic acid binding properties. Unlike other hnRNP proteins which display a varying range of affinities for different ribonucleotidehomopolymers and ssDNA, hnRNP F and hnRNP H bind only to poly(rG) *in vitro*. hnRNP F and hnRNP H were purified from HeLa cells by poly(rG) affinity chromatography and oligonucleotides derived from peptide sequences were used to isolate a cDNA encoding hnRNP F. The predicted amino acid sequence of hnRNP F revealed a novel protein with three repeated domains related to the RNP consensus sequence RNA-binding domain. Monoclonal antibodies produced against bacterially expressed hnRNP F were specific for both hnRNP F and hnRNP H and recognized related proteins in divergent organisms, including in the yeast *Saccharomyces cerevisiae*. hnRNP F and hnRNP H are thus highly related immunologically and they share identical peptides. Interestingly, immunofluorescence microscopy revealed that hnRNP F and hnRNP H are concentrated in discrete regions of the nucleoplasm, in contrast to the general nucleoplasmic distribution of previously characterized hnRNP proteins. The unique RNA-binding properties, amino acid sequence and distinct intranuclear localization of hnRNP F and hnRNP H make them novel hnRNP proteins that are likely to be important for the processing of RNAs containing guanosine-rich sequences.

INTRODUCTION

As RNA polymerase II transcripts are synthesized, they associate with proteins to form heterogeneous nuclear ribonucleoprotein (hnRNP) complexes. It is within these hnRNP complexes that pre-mRNAs are processed to form mature mRNA transcripts before export from the nucleus (1). Immunopurified hnRNP complexes from human HeLa cells contain more than 20 individual proteins designated hnRNP A1 (34 kD) through hnRNP U (120 kD) (2), and characterization of these proteins is important because they are likely to influence mRNA metabolism. Many hnRNP proteins have been characterized at the molecular level, including hnRNP A1 (3), A2/B1 (4), C1/C2 (5), I (6–9), K (10), L (11), M (12) and U (13). A common feature of most of the hnRNP proteins cloned and sequenced to date (with the exceptions of hnRNP K and hnRNP U) is an evolutionarily conserved 80–100 amino acid RNA-binding domain (RBD), the RNP-consensus sequence (RNP-CS), also referred to as RNP motif and RNA-recognition motif (14–16). The structure of the RBD, and experimental evidence that it functions as an RNA-binding domain, has been presented for the hnRNP C1 protein (17,18) and for the U1 snRNP A protein (19–22). Significantly, RNA appears to bind to the surface of the RBD and, therefore, remains exposed and accessible for interaction with other factors (17).

Although the precise roles of hnRNP proteins in pre-mRNA processing are not yet known, considerable insights into potential functions have been gained by the availability of cDNA clones and monoclonal antibodies to these proteins. For example, immuno-inhibition experiments have implicated the C proteins in pre-mRNA splicing (23), and hnRNP A1 has been identified as a factor that determines splice site selection *in vitro* (24). Several hnRNP proteins, including hnRNP C, hnRNP U (58) and hnRNP A1 (25, 26) have RNA annealing-promoting activity, an activity that is likely to be important for their roles in pre-mRNA splicing. In addition to influencing splicing, hnRNP proteins may also have a role in nucleo-cytoplasmic mRNA transport, since it has recently been shown that many of them shuttle between the nucleus and the cytoplasm (27).

*To whom correspondence should be addressed

+Present address: Harvard Medical School, Department of Microbiology, Cambridge, MA, USA

Evidence that individual hnRNP proteins have unique functions is suggested by their different nucleic acid-binding properties. hnRNP proteins can be classified according to their ribonucleotide homopolymer binding at 2 M NaCl (28). Under these stringent binding conditions, the hnRNP C proteins bind specifically to poly(rU), hnRNP K and hnRNP J bind to poly(rC), hnRNP P binds to poly(rA) and hnRNP M, hnRNP H, hnRNP F and the two lower molecular weight hnRNP E proteins bind to poly(rG). In addition to providing a useful means of classifying hnRNP proteins, these ribonucleotide homopolymer binding properties have also been useful for purifying hnRNP proteins (10, 29) and for predicting the binding sites of hnRNP proteins on authentic pre-mRNAs (30). Individual hnRNP proteins also display unique interactions with ssDNA. While most hnRNP proteins bind *in vitro* to ssDNA in a heparin resistant manner, the binding of hnRNP U is heparin sensitive and the lower molecular weight hnRNP E proteins, hnRNP F and hnRNP H do not bind ssDNA at all (2).

Among the hnRNP proteins characterized so far, hnRNP F (M_r = 53 kD, pI = 6.1) and hnRNP H (M_r = 56 kD, pI = 6.7−7.1) have unique nucleic acid-binding properties; they are among the only hnRNP proteins which bind RNA but do not bind ssDNA, and they bind specifically and tenaciously to poly(rG). Evidence indicating that the abundant hnRNP F and hnRNP H proteins are authentic hnRNP proteins includes their immunopurification with hnRNP complexes and their RNA-binding properties (2, 28). Here, we describe the purification and characterization of hnRNP F and hnRNP H. hnRNP F and hnRNP H were found to be highly related, based on both peptide sequence analysis and immunological reactivity. A cDNA encoding hnRNP F was isolated and its DNA sequence predicted a protein with three, approximately 80 amino acid domains that are related to the well characterized RNP-CS. Monoclonal antibodies were produced against bacterially expressed hnRNP F and one antibody recognized related proteins in all vertebrate organisms tested and *Saccharomyces cerevisiae*. In contrast to the homogeneous nucleoplasmic distribution characteristic of most hnRNP proteins, immunofluorescence microscopy detected hnRNP F and hnRNP H concentrated in discrete regions of the nucleoplasm. The nucleic acid binding properties of hnRNP F and hnRNP H, as well as their interesting intracellular localization, make them unique hnRNP proteins.

MATERIALS AND METHODS

Cell culture, labeling and fractionation

Human HeLa and Madin−Darby bovine kidney cells were cultured at 37°C in Dulbecco's modified Eagle's medium (DME) supplemented with penicillin, streptomycin (P/S) and 10% fetal calf serum. *Xenopus laevis* kidney epithelial cells were grown at 27°C in 80% DME supplemented with P/S and 10% fetal calf serum. *Drosophila melanogaster* Schneider's cells (line 2) were cultured at 25°C in Schneider's *Drosophila* medium (Gibco BRL, Gaithersburg, MD) supplemented with P/S and 10% heat inactivated fetal calf serum. *S.cerevisiae* and *Schizosaccharomyces pombe* were grown at 30°C in YPD (yeastolate, bacto-peptone, dextrose). HeLa cells were labeled with 20 μCi/ml TRAN[^{35}S] (ICN, Irvine, CA) for 20 hours in DME containing one-tenth the normal methionine levels and 5% calf serum. HeLa cell fractionation and preparation of nucleoplasm were performed as previously described (31).

Protein purification and amino acid sequencing

hnRNP F and hnRNP H were purified from nucleoplasm prepared from 2×10^9 HeLa cells grown in suspension culture. Nucleoplasm was digested with 100 units/ml micrococcal nuclease (Pharmacia LKB Biotechnology, Piscataway, NJ) for 10 min at 30°C in the presence of 1 mM $CaCl_2$. Digestion was stopped on ice by the addition of EGTA to a final concentration of 5 mM. NaCl was added to the digested nucleoplasm to a final concentration of 2 M and heparin (Sigma, St Louis, MS) was added to a final concentration of 1 mg/ml. The nucleoplasm was centrifuged at 5,000×g for 10 min and the supernatant was incubated with 3 ml of poly(rG)-agarose (Pharmacia LKB Biotechnology, Piscataway, NJ) for 15 min at 4°C with continuous rocking. The poly(rG)-agarose was packed into a column, washed extensively with binding buffer (50 mM Na phosphate, pH 7.4; 2 M NaCl) and eluted with 4 M guanidine hydrochloride. Approximately 200 μg of hnRNP H and 100 μg of hnRNP F were obtained.

Purified hnRNP F and hnRNP H were resolved by SDS-PAGE and transferred to nitrocellulose membrane. The proteins were digested *in situ* with endoproteinase Glu-C or with trypsin (Sequencing grade, Boehringer Mannheim Biochemicals, Indianapolis, IN) as described by Aebersold *et al.* (32). Peptides were separated by narrow-bore (2.1 mm internal diameter) C18 reverse-phase HPLC and sequenced on a Applied Biosystems Model 473 protein sequenator.

Bacterial expression

A partial length cDNA (F13) encoding hnRNP F amino acids 1−363, was cloned into the pET-11a bacterial expression vector (Novagen, Madison, WI) and expression of the encoded protein was induced in *Escherichia coli* [strain Bl21(DE3)] using previously described procedures (33). Expressed protein was resolved by SDS-PAGE and electroeluted from the gel using an Isco model 1750 electrophoretic sample concentrator (Isco Inc., Lincoln, NE). Purified protein was dialyzed against 6 M urea to remove residual SDS and finally against water. After lyophilization, protein was resuspended in phosphate buffered saline (pH 7.4) and used for immunization.

Preparation of monoclonal antibodies

The monoclonal antibody 4F4 was prepared as previously described (31) and monoclonal antibody Y12 was a kind gift from Dr Joan Steitz (Yale University). Anti-hnRNP F and hnRNP H antibodies were prepared by immunizing BALB/c mice with bacterially expressed protein prepared as described above. Hybridoma production, screening, and ascites production were performed as previously described (34).

Immunopurification of proteins and hnRNP complexes

hnRNP complexes were immunopurified from HeLa nucleoplasm in the presence of 0.5% Triton X-100 as previously described (35). hnRNP F and hnRNP H were immunoprecipitated from cell lysate in phosphate buffered saline (pH 7.4) containing 1% Empigen BB, 1 mM EDTA and 0.1 mM DTT (35). Ascites fluid from a BALB/c mouse innoculated with the parent myeloma cell line SP2/0 was used in each experiment as a non-immune control.

Gel electrophoresis and immunoblotting

Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and fluorographed as previously

described (34). Two-dimensional nonequilibrium pH gradient polyacrylamide gel electrophoresis (NEPHGE) was performed as described by O'Farrell *et al.* (36) using an ampholine gradient of pH 3−10 separated for 4 hours at 400 volts in the first dimension. Proteins were separated by SDS-PAGE in the second dimension. Proteins were transferred to nitrocellulose membrane (Schleicher and Schuell, Keene, NH) and the immunoblots were probed with antibodies as previously described (31).

Immunofluorescence microscopy

Immunofluorescence microscopy was performed on monolayer-adapted HeLa cells (clone JW36) as previously described (31). Cells were first incubated with mouse ascites fluid diluted 1:1000, followed by incubation with fluorescein isothiocyanate-conjugated goat anti-mouse F(ab')2 antibody (Cappel Laboratories, Malvern, PA) diluted 1:500. Localization was detected using a laser scanning confocal microscope (model MRC600; Bio-Rad Laboratories, Cambridge, MA).

Isolation and characterization of cDNA clones

Degenerate oligonucleotides corresponding to three of the determined peptide sequences, (1) IVQFF (2) EAFVQFA (3) QFASQE, were synthesized in both orientations and used to perform two rounds of nested polymerase chain reaction (PCR). Oligonucleotides corresponding to peptide (1) were used in both rounds of PCR and oligos corresponding to peptides (2) and (3), which partially overlap, were used consecutively in the first and second rounds of PCR. A 101 base-pair nucleotide fragment was generated using sense-orientation oligonucleotides to peptide (1) and antisense-orientation oligonucleotides for peptides (2) and (3). This DNA fragment was 3'-end labeled and used to screen a HeLa cell λgt11 cDNA library (Clontech, Palo Alto, CA). Several positive clones were identified and the largest one, F13, was further characterized.

The *Eco*RI fragment of F13 was subcloned into the plasmid pGEM-7Z (Promega, Madison, WI) to create phF13, and was sequenced entirely on both strands. All sequencing reactions were performed on plasmid DNA with the Sequenase Sequencing Kit (United States Biochem) according to the manufacturer's instructions. Sequence analysis was performed using the University of Wisconsin's Genetics Computer Group programs and databases were searched using the BLAST service at the National Center for Biotechnology Information (37).

In vitro transcription, translation, immunoprecipitation, and ribonucleotidehomopolymer binding

phF13 was cut at appropriate restriction sites to provide templates for *in vitro* transcription using T7 polymerase (Promega, Madison, WI). RNAs were translated in nuclease treated rabbit reticulocyte lysate in the presence of [^{35}S]methionine according to the manufacturer's instructions (Promega, Madison, WI). For ribonucleotide homopolymer binding assays, 5 μl of translation reaction and 500μl of binding buffer [RSB containing 2 μg/ml leupeptin-pepstatin-A, 0.5% aprotinin (Sigma, St Louis, MI) .05% Triton-X-100 and the indicated NaCl concentration] was added to a 50 μl slurry of poly(rX)-agarose (Sigma, St Louis, MI) that was pre-washed in binding buffer. Reactions were incubated at 4°C for 10 minutes, spun down and incubated in binding buffer containing 1mg/ml heparin (Sigma, St Louis, MI) at 4°C for 5 min. Reactions were spun down and washed 4 times with 1 ml of binding buffer. 25 μl of SDS sample buffer was added directly to the agarose beads, samples were heated to 90°C for 5 minutes, and proteins were analyzed by SDS-PAGE. Immunoprecipitation of the [^{35}S]methionine-labeled proteins was performed as described above. Ascites fluid from a BALB/c mouse inoculated with the parent myeloma cell line SP2/0 was used as a non-immune control.

RESULTS

Purification and amino acid sequencing of hnRNP F and hnRNP H peptides

Many hnRNP proteins can be classified by their RNA-binding preferences (28). A subset of hnRNP proteins, consisting of the two lower molecular weight hnRNP E proteins, hnRNP F, hnRNP H and the hnRNP M proteins, bind tenaciously to poly(rG) and they can be purified from other nuclear proteins by binding to poly(rG) at 2 M NaCl (28). Taking advantage of these properties, hnRNP F and hnRNP H were purified from HeLa cell nucleoplasm (Figure 1). Nucleoplasm was adjusted to 2 M NaCl, bound to poly(rG), and bound proteins were eluted with 4 M guanidine hydrochloride and separated by SDS-PAGE. The major proteins detected had relative molecular masses of 53 and 56 kD, [Figure 1, lane *poly(rG)*]. In addition, smaller amounts of M and the lower molecular weight E proteins were also detected.

After SDS-PAGE, the purified hnRNP F and hnRNP H were transferred to nitrocellulose and subjected to *in situ* protease

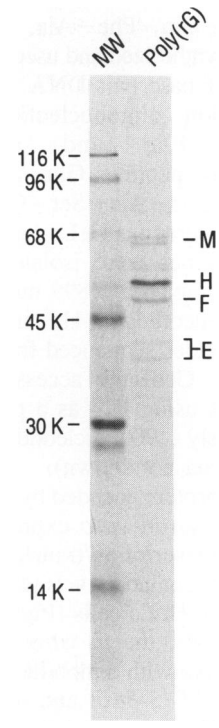

Figure 1. Purification of a subset of hnRNP proteins by poly(rG) affinity chromatography. HeLa cell nucleoplasm was adjusted to 2 M NaCl and bound to poly(rG)-agarose. Proteins retained on the column were eluted with 4 M guanidine hydrochloride and resolved by SDS-PAGE [lane *poly(rG)*]. The relative positions of the hnRNP M, H, F and E proteins are indicated on the right. Sizes of the molecular weight standards (lane *MW*) are indicated on the left in kilodaltons.

Table 1. F and H peptide sequences

Peptide	Sequence	Position
F21	IVQFFSGLE	127–135
F13	SMGHR(W/Y)	77– 82
F12	G(M/R)QSGEAFVE	51– 60
H15	IVQFFSGLE	(127–135)
H10	AFVQFASQE	(155–163)
H17	GEAVQFASQE	(153–163)

Sequences are in single-letter code. Residues in parentheses, (W/Y) and (M/R), indicate the identification of two possible amino acids at that position. The positions of peptides in the hnRNP F protein are indicated.

digestion with either endoproteinase Glu-C or with trypsin. Peptides were separated by reverse phase high performance liquid chromatography and the elution profiles for the digested hnRNP F and hnRNP H proteins were strikingly similar (data not shown). Several well resolved peptides were sequenced and the results are presented in Table 1. As suggested by the elution profiles, amino acid sequencing demonstrated that hnRNP F and hnRNP H are highly related, as several peptides obtained from the two proteins were identical (e.g. F21 and H15) or nearly identical (e.g. F12 and H17).

hnRNP F contains three RNA-binding domains

To obtain the complete amino acid sequence of hnRNP F and hnRNP H, degenerate oligonucleotides corresponding to the peptide Ile–Val–Gln–Phe–Phe, and the overlapping peptides Glu–Ala–Phe–Val–Gln–Phe–Ala, and Gln–Phe–Ala–Ser–Gln–Glu were synthesized and used to perform two rounds of nested PCR. A 101 base pair DNA fragment was obtained using sense-orientation oligonucleotides for the peptide Ile–Val–Gln–Phe–Phe and antisense-orientation oligonucleotides for the peptides Glu–Ala–Phe–Val–Gln–Phe–Ala and Gln–Phe–Ala–Ser–Gln–Glu. This DNA fragment was used to screen a HeLa λgt11 cDNA library and several hybridizing clones were isolated. The largest clone, cDNA clone F13, contained a 1,899 nucleotide insert with an open reading frame encoding a 415 amino acid protein that contains all of the peptides sequenced from both hnRNP F and hnRNP H (Figure 3; GenBank accession number L28010). Northern blot analysis using F13 as a probe detected a single mRNA of approximately 1,900 nucleotides, consistent with F13 being a full length (data not shown).

To characterize the protein encoded by F13, the lambda phage insert was cloned into an *in vitro* expression vector to create phF13. The *in vitro* transcription/translation product of phF13 migrated as a 53 kD protein, and it was identical in size to hnRNP F immunopurified from HeLa cells (Figure 2, lanes *phF13* and *HeLa/8A6*). Furthermore, the *in vitro* translated protein was immunologically reactive with antibodies specific for hnRNP F and hnRNP H (lane *phF13/8A6*) and it bound specifically to poly(rG) at 2 M NaCl as expected [lane *phF13/poly(rG)*]. In addition, monoclonal antibodies generated by injecting mice with bacterially expressed protein produced from the F13 cDNA clone were specific for hnRNP F and for hnRNP F and hnRNP H (see below). These data support the conclusion that the F13 cDNA clone encodes hnRNP hnRNP F.

The predicted amino acid sequence of the F13 cDNA encodes a protein of 45.7 kD with a pI of 5.24 which is consistent with

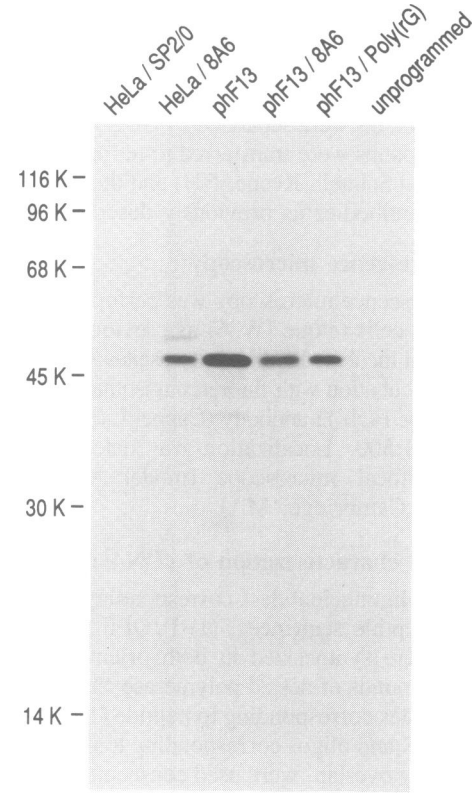

Figure 2. Immunopurification and poly(rG)-binding properties of the phF13 *in vitro* transcription–translation product. phF13 was linearized with *Bam*HI and transcribed and translated *in vitro* in the presence of [^{35}S]methionine (lane *phF13*). The translation product comigrates with hnRNP F immunopurified with from HeLa cells, is immunprecipitated with the monoclonal antibody 8A6, and binds to poly(rG) at 2 M NaCl. Lanes: *HeLa/SP2/0*, control immunoprecipitation from [^{35}S]methionine-labeled HeLa cells with SP2/0 ascites fluid; *HeLa/8A6*, immunoprecipitation of hnRNP F and hnRNP H from [^{35}S]methionine-labeled HeLa cells with monoclonal antibody 8A6; *phF13*, phF13 *in vitro* transcription–translation product; *phF13/8A6*, phF13 *in vitro* transcription–translation product immunoprecipitated with monoclonal antibody 8A6; *phF13/poly(rG)*, phF13 *in vitro* transcription-translation product bound to poly(rG)-agarose at 2 M NaCl; *unprogrammed*, product of *in vitro* transcription-translation reaction lacking RNA template. Proteins were analyzed by SDS-PAGE. Sizes of molecular weight standards are shown on the left in kilodaltons.

the properties of purified hnRNP F (Figure 3). The difference between the predicted molecular mass and the relative mass of hnRNP F, as determined by SDS-PAGE, is likely to be due to the net negative charge of the protein, which contains 13% glutamic and aspartic acid. The most notable feature of hnRNP F is the presence of three repeated domains of approximately 80 amino acids each that have been boxed and are labeled *RBDI–RBDIII* in Figure 3. Alignment of the domains revealed three highly conserved peptides; one near the beginning of the repeats (Leu–Arg–Gly–Leu–Pro), one in the center (Gly–Glu–Ala–Phe–Val) and one at the end (His–Arg–Tyr–Ile–Glu–Val–Phe–Lys–Ser) (Figure 4). Several additional positions are also conserved between all three domains, although regions outside of the three highly conserved peptides are more variable. When compared to each other, the domains have between 41% and 46% amino acid sequence identity and between 61% and 73% amino acid sequence similarity. Many hnRNP proteins characterized to date contain between one and four copies of the 80 amino acid RNP-CSs, whose hallmarks are two highly

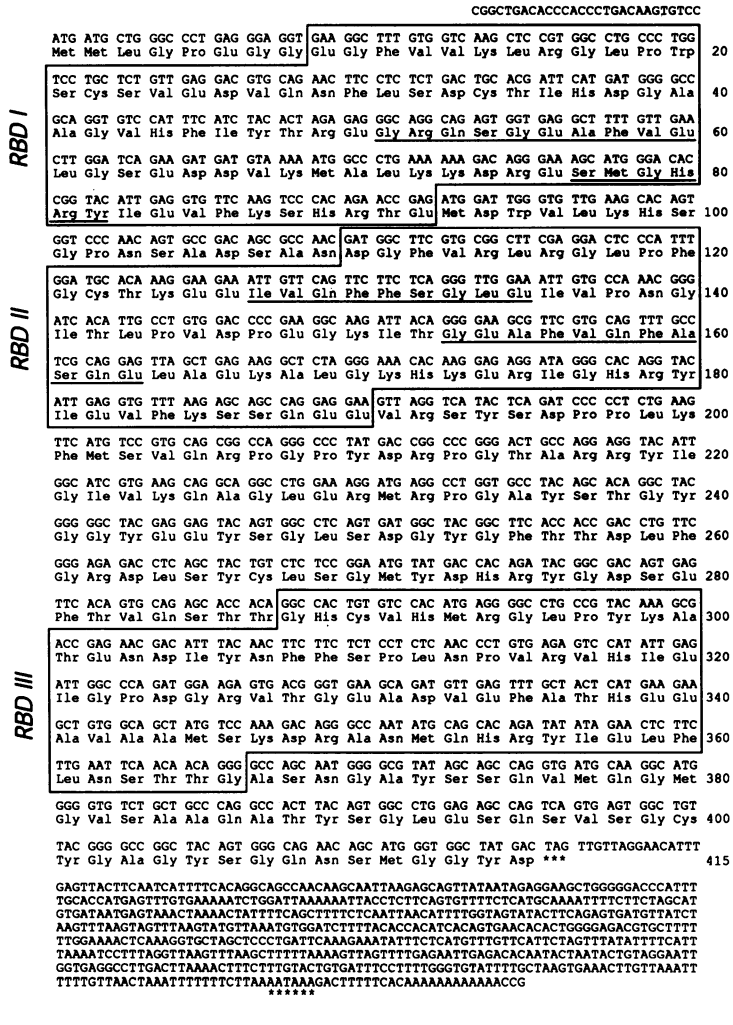

Figure 3. DNA sequence of the F13 cDNA and the predicted amino acid sequence. Boxed regions correspond to the three RBDs. Underlined amino acids correspond to sequenced peptides. A putative polyadenylation signal is noted with asterisks. The GenBank accession number for this sequence is L28010.

conserved motifs termed RNP1 and RNP2 (14–16). Comparison of the RNP-CS with the consensus sequence derived from the three domains in hnRNP F revealed that they are significantly related (Figure 4). The RBDs of hnRNP F, however, are unique from most other members of the RNP-CS family. Only five out of eight residues are conserved in a region of the hnRNP F domains corresponding to RNP1. Notable deviations include the serine or threonines at the first position of RNP1 (typically arginine or lysine), and the acidic residues at positions three and five (usually aromatic amino acids). The region corresponding to RNP2 is less well conserved. In addition to RNP1 and RNP2, there are also several more loosely conserved positions that contribute to the hydrophobic core of the RNP-CS RBD (16). Importantly, many of these residues are conserved in the domains of hnRNP F (Figure 4, positions with asterisks). One of the least conserved regions in the RNP-CS RBD is the region preceeding RNP1, referred to as loop 3. Amino acids in loop 3 are thought to be important for determining RNA-binding specificity, and this has been shown experimentally for two proteins, the U1-A and U2-B″ snRNP proteins (38). Interestingly, sequence comparison of the RBDs of hnRNP M and hnRNP F revealed a conservation of a unique tri-peptide in loop three, Glu/Asp–Gly–Arg/Lys. Because hnRNP M also binds preferentially to poly(rG) (12), it is possible that this tri-peptide contributes to the RNA-binding specificities of these proteins.

In addition to three RBDs, hnRNP F contains a 100 amino acid auxiliary domain between RBDs two and three and a 60 amino acid carboxy-terminal auxiliary domain. These domains contain several short repeated elements, including six Tyr–Ser dipeptides and four Gly–X–Tyr–(Asp/Glu) peptides. The significance of these short peptide repeats is not known, however, the interspersion of glycine, serine and tyrosine is common in the auxiliary domains of several hnRNP proteins, including the hnRNP A and hnRNP B proteins, and may be important for protein–protein interactions (3, 4).

hnRNP F and hnRNP H are immunologically related and evolutionarily conserved

To further characterize hnRNP F and hnRNP H, monoclonal antibodies were produced against a bacterially expressed portion of hnRNP F containing amino acids 1–363. Mice injected with this form of hnRNP F produced antibodies that recognized both hnRNP F and hnRNP H (data not shown), further suggesting that these proteins are highly related. The majority of the isolated monoclonal antibodies also recognized both hnRNP F and hnRNP H, although several hnRNP F-specific antibodies were also

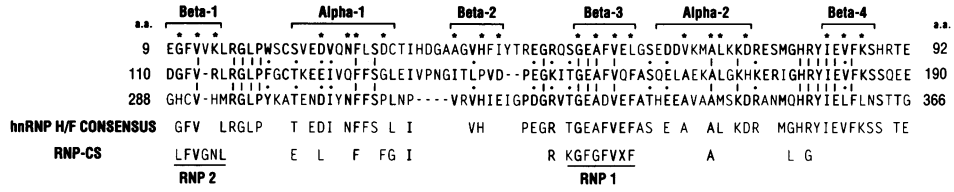

Figure 4. Sequence alignment of the three hnRNP F RBDs. Amino acid positions of the domains are indicated. Vertical lines indicate an identical amino acid at the position in all three domains. Dots indicate an identical or conserved amino acid at that position between the three domains. Bold letters indicate a conserved or identical amino acid at that position between at least two of the domains. A consensus sequence for the RBDs of hnRNP F is indicated, as well the consensus sequence for the canonical RNP-CS RBD (Dreyfuss *et al.*, 1988; Bandziulis *et al.*, 1989). Residues conserved between the consensus sequences are in bold. Asterisks indicate the positions of conserved amino acids contributing to the hydrophobic core of the RNP-CS RBD. The positioning of β-sheets and α-helixes are based on these conserved residues (see Kenan *et al.*, 1990).

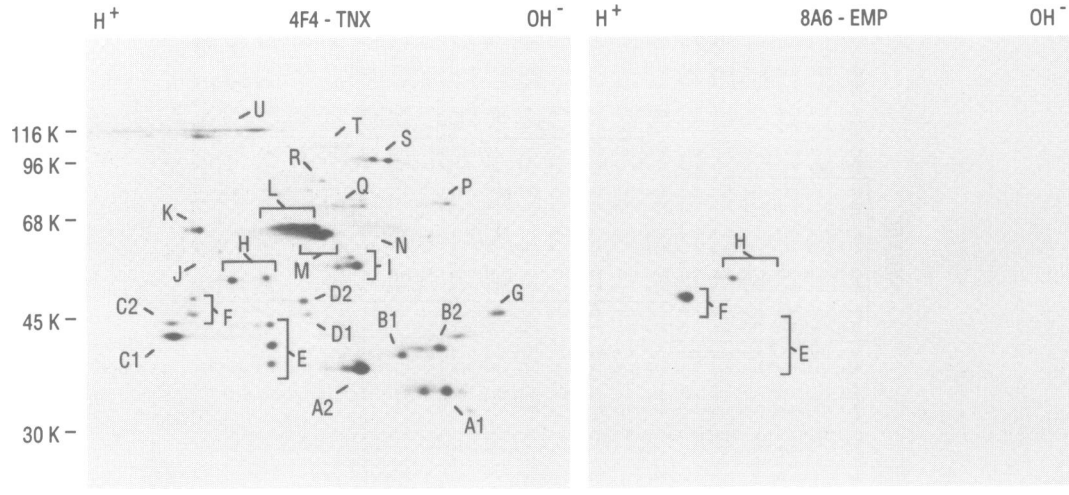

Figure 5. Monoclonal antibody 8A6 is specific for the hnRNP H, hnRNP F and smallest hnRNP E proteins. hnRNP complexes were immunopurified from [^{35}S]methionine-labeled HeLa cells with the monoclonal antibody 4F4 (left panel). hnRNP H, hnRNP F and the little hnRNP E proteins were immunopurified from [^{35}S]methionine-labeled HeLa cells in the presence of the ionic detergent Empigen BB with the monoclonal antibody 8A6 (right panel). Proteins were analyzed by two-dimensional gel electrophoresis. Sizes of molecular weight standards are indicated on the left in kilodaltons.

obtained. Interestingly, immunopurification and two-dimensional gel analysis indicated that monoclonal antibody 8A6 recognizes hnRNP F, hnRNP H and the lower molecular weight hnRNP E proteins, revealing that, in addition to their similar nucleic acid-binding properties, these proteins share a common epitope (Figure 5). Although hnRNP F was originally defined as a doublet of 44 and 53 kD (2), there is no evidence that these two proteins are related. The data presented here reveals that the 53 kD hnRNP F protein is immunologically distinct from the 44 kD protein. Furthermore, these proteins have different nucleic acid-binding properties and are, therefore, likely to be distinct hnRNP proteins (2, 28).

Immunoblot analysis of cell lines from a number of divergent organisms with 8A6 indicated that hnRNP F and hnRNP H are conserved among vertebrates (Figure 6). Whereas human hnRNP F and hnRNP H have molecular weights of 53 and 56 kD, the proteins recognized in bovine were 45 and 50 kD, and proteins recognized in frog had molecular weights of 53 and 54 kD. Surprisingly, an immunologically related protein of approximately the same molecular mass as hnRNP F was detected in the yeast *S.cerevisiae*. hnRNP proteins are commonly conserved among vertebrates, however few related proteins have been identified in yeast (1, 29). The relationship between the yeast protein and human hnRNP F and hnRNP H (which has not yet been established) is therefore of interest.

hnRNP F and hnRNP H bind specifically to poly(rG)

Although it has been demonstrated that many hnRNP proteins display RNA sequence binding preferences, all of the previously characterized proteins appear to also bind nucleic acids in an apparently sequence independent manner (2, 41, 42). One unique feature of hnRNP F and hnRNP H is that poly(rG) is the only nucleic acid that they have been found to bind to *in vitro* (2, 28). To further characterize the nucleic acid binding properties of hnRNP F and hnRNP H, their binding to ssDNA, double-stranded DNA (dsDNA), poly(rG), poly(rA), poly(rU) and poly(rC) was analyzed (Figure 7). HeLa cell nucleoplasm was bound to each of the nucleic acids at 100 mM NaCl and bound proteins were eluted with SDS sample buffer and analyzed by immunoblot analysis with monoclonal antibody 8A6. Under these conditions, hnRNP F and hnRNP H showed a strong preference for poly(rG) and only slight binding to ssDNA or poly(rU) was detected. hnRNP F and hnRNP H had no detectable binding to dsDNA, poly(rA), or poly(rC).

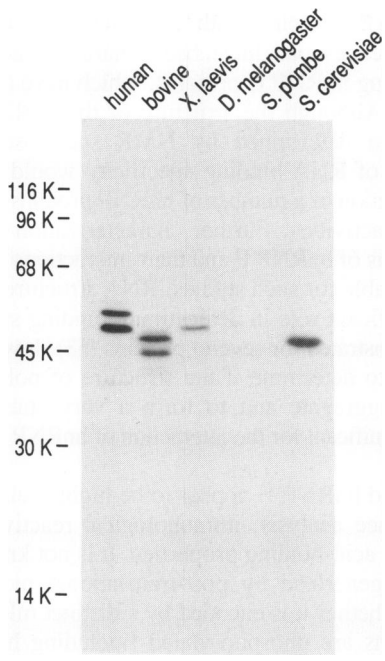

Figure 6. Proteins immunologically related to hnRNP F and hnRNP H are found in vertebrates and in the yeast *S.cerevisiae*. Total cell lysates from the indicated organisms were resolved by SDS-PAGE, transferred to nitrocellulose membrane and probed with the monoclonal antibody 8A6. Sizes of molecular weight standards are indicated on the left in kilodaltons.

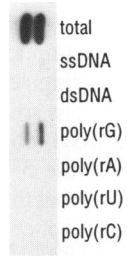

Figure 7. hnRNP F and hnRNP H bind specifically to poly(rG). HeLa nucleoplasm was incubated with single-stranded DNA–cellulose (*ssDNA*), double-stranded DNA–cellulose (dsDNA), poly(rG)-agarose, poly(rA)-agarose, poly(rU)-agarose and poly(rC)-agarose at 100 mM NaCl. Bound proteins were eluted with SDS sample buffer, resolved by SDS-PAGE and transferred to nitrocellulose membrane. hnRNP F and hnRNP H were detected using monoclonal antibody 8A6.

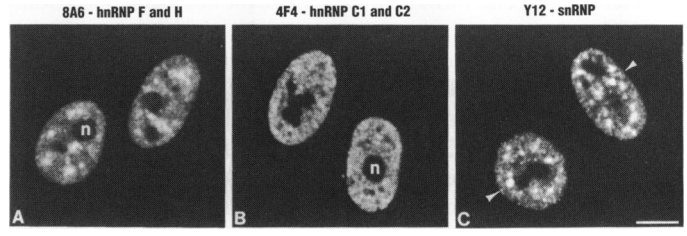

Figure 8. hnRNP F and hnRNP H are concentrated in discrete regions of the nucleoplasm. (**A**) Localization of hnRNP F and hnRNP H by indirect immunofluorescent staining of HeLa cells with the monoclonal antibody 8A6. (**B**) Localization of hnRNP C1/C2 by indirect immunofluorescent staining of HeLa cells with the monoclonal antibody 4F4. (**C**) Localization of snRNPs by indirect immunofluorescent staining of HeLa cells with the monoclonal antibody Y12. Coiled bodies stained by monoclonal antibody Y12 are indicated by arrowheads. Representative nucleoli in each cell are indicated by an n. Bar equals 10 μm. Localization was detected using a laser scanning confocal microscope.

hnRNP F and hnRNP H are concentrated in discrete regions in HeLa cell nuclei

Monoclonal antibody 8A6 was used for indirect immunolocalization of hnRNP F, hnRNP H and the lower molecular weight hnRNP E proteins (similar localization was detected with hnRNP F- and hnRNP H-specific antibodies) in cultured HeLa cells by confocal microscopy (Figure 8A). hnRNP F and hnRNP H were detected throughout the nucleoplasm, however, a low level of punctate cytoplasmic immunofluorescence was also consistently observed (not readily visible in the exposure shown in Figure 8A). The localization of hnRNP F and hnRNP H throughout the nucleoplasm was not homogeneous as seen with monoclonal antibodies to many previously characterized hnRNP proteins, including hnRNP C1/C2 (10, 11, 31). As observed by confocal microscopy, the hnRNP C proteins are completely excluded from the cytoplasm and the nucleolus and have a fine, granular, distribution throughout the nucleoplasm (Figure 8B). In contrast, hnRNP F and hnRNP H appear to be concentrated in discrete regions of the nucleoplasm, in addition to having a diffuse nucleoplasmic localization. This staining pattern is reminiscent of the 'speckled' pattern observed with snRNP-specific antibodies and with antibodies to other splicing factors such as SC35 (39). The anti-snRNP monoclonal antibody Y12 stains a large number of speckles throughout the nucleoplasm, as well as several very intense foci corresponding to coiled bodies (Figure 8C). As a general comparison, fewer speckles were observed with antibodies to hnRNP F and hnRNP H compared to staining with the anti-snRNP monoclonal antibody Y12, and the speckles observed with hnRNP F and hnRNP H antibodies were considerably larger. Furthermore, hnRNP F and hnRNP H are not present in coiled bodies, which contain snRNPs and several additional splicing factors (40, 59). The immunolocalization of snRNP antigens and hnRNP F and hnRNP H is, therefore, clearly distinct. The possibility that some overlap exists between the sites containing concentrated levels of hnRNP F and hnRNP H and sites containing splicing factors, and whether there is any interaction between these factors, will be the focus of future investigations.

DISCUSSION

hnRNP complexes isolated from human HeLa cells contain more than 20 individual proteins. Many (if not all) of these proteins have different RNA-binding preferences that are likely to be important in determining their binding sites on pre-mRNA substrates (7, 28, 43, 44). Here, we have characterized two hnRNP proteins with novel nucleic acid-binding properties, hnRNP F and hnRNP H. While the actual binding sites of hnRNP F and hnRNP H on native pre-mRNAs *in vivo* is not known, binding to synthetic nucleic acids *in vitro* demonstrated the preference of these proteins for poly(rG). By analogy, the hnRNP C proteins, which bind preferentially to poly(rU) *in vitro*, bind to the uridine-rich polypyrimidine tract at the 3'-end of introns of authentic pre-mRNAs (28, 30). It can be anticipated, therefore, that hnRNP F and hnRNP H will bind preferentially to G-rich sequences in native pre-mRNAs.

G-rich RNA elements

G-rich sequences have been implicated to be important in a number of pre-mRNA processing reactions, including splicing. For example, the 5′ end of introns are purine-rich and a prevailing element is the trinucleotide GGG (45). Although it is known that the binding of U1 snRNP to the 5′ splice site is the first step of spliceosome assembly, and that this step involves basepairing between U1 snRNA and the 5′ splice site (46), the actual process of splice site selection is not understood, but may be facilitated by these G-rich sequences and factors bound to them. G-rich elements (designated exon recognition sequences) have also been detected in the exons of a number of genes and enhance the splicing of the immediately upstream intron (47). Interestingly, enhancement of splicing appears to involve the binding of U1 snRNP to the exon recognition sequence, which also may be facilitated by factors bound to these G-rich elements. In addition to affecting splicing, G-rich elements are also important for the efficiency of 3′ end formation (48). A factor that binds specifically to a G-rich sequence downstream of the simian virus 40 late polyadenylation signal has been characterized (48), but its relationship to hnRNP F or hnRNP H is not known. Because we have found that hnRNP F and hnRNP H also bind oligo(dG) (M.J.Matunis and G.Dreyfuss, unpublished data), it is also worth noting that telomeres contain G-rich single-stranded DNA elements, and that the regulatory regions of many genes contain G-rich sequences that may adopt a single-stranded conformation (49, 50). The suggestion of a possible involvement of hnRNP proteins in both telomere function and transcription is not without precedent, as the hnRNP D and hnRNP E proteins bind specifically to DNA telomere sequences *in vitro* (44), and the hnRNP K protein binds specifically to a cytosine-rich DNA element in the promoter of the c-myc gene *in vitro* (51). The observation that hnRNP F and hnRNP H bind oligo(dG) and poly(rG), but not ssDNA, also suggests that the binding specificities of these proteins is affected by the base composition of the nucleic acid.

Unique RNP-CS RBDs

The most notable features of the hnRNP F protein sequence are the three repeated RBDs , which are significantly more similar to each other than they are to the RNP-CS RBDs of other proteins (15, 16). NMR and crystallography studies have shown that the RNP-CS RBD has a $\beta1-\alpha1-\beta2-\beta3-\alpha2-\beta4$ (β = β-sheet; α = α-helix) structure (18, 21, 22). In addition to RNP1 and RNP2, these structural studies have revealed a number of other conserved residues that contribute to the hydrophobic core of the RBD (16). Significantly, many of these residues are conserved in the RBDs of hnRNP F, and these domains are, therefore, likely to have a tertiary structure similar to the RNP-CS RBD. The determinants of RNA-binding specificity for proteins containing RNP-CS RBDs are likely to reside in regions outside of the conserved RNP1 and RNP2 peptides (16). However, because these peptides in hnRNP F are clearly distinct from the consensus, they may be as important in determining specificity as other sequences in the domain. In particular, the highly conserved aromatics at positions 3 and 5 of RNP1, which likely make direct contact with bound RNA (17, 52) are not well conserved in hnRNP F. Other determinants of specificity may be in the loop between β-sheets two and three which share homology with hnRNP M, a protein which also binds preferentially to poly(rG) (12).

Other hnRNP proteins with ribonucleotide homopolymer binding preferences that have been characterized include the poly(rU)-binding hnRNP C proteins, which have a single RNP-CS RBD (5). Although the structure of the hnRNP C protein RBD has been determined by NMR spectroscopy, further understanding of RNA-binding specificity would benefit from the characterization of a number of hnRNP proteins with different RNA-binding activities. Further characterization of the RNA-binding domains of hnRNP F and their interactions with poly(rG) should be valuable for such studies. RNA structure is also likely to play a significant role in determining binding specificity and has been demonstrated for several proteins (53). It will, therefore, be of interest to determine if the structure of poly(rG), which is known to aggregate and to form a very stable secondary structure, is significant for the interaction of hnRNP F and hnRNP H (54, 55).

hnRNP F and hnRNP H appear to be highly related based on peptide sequence analysis, immunological reactivity and their similar nucleic acid-binding properties. It is not known whether hnRNP H is generated by post-translational modification of hnRNP F or whether it is encoded by a distinct mRNA. Several hnRNP proteins are phosphorylated [including hnRNP F and hnRNP H (M.J.Matunis and G.Dreyfuss, unpublished data)], and in some cases phosphorylation of the C proteins can lead to an apparent increase in their size by SDS-PAGE (56, 57). Isoforms of hnRNP proteins are also often created by alternative splicing of a common pre-mRNA. This is the case for the human hnRNP A2/B1 proteins, the hnRNP C1/C2 proteins, and for the hnRNP M and hnRNP I proteins (4, 7, 12). Only one hnRNP F mRNA was detected by northern blot analysis, however it is possible that mRNAs encoding for hnRNP F and hnRNP H are similar in size. Reverse transcription and PCR also detected only one size of mRNA (M.J.Matunis and G.Dreyfuss, unpublished data).

Intranuclear distribution of hnRNP F and hnRNP H

Most hnRNP proteins are distributed homogeneously throughout the nucleoplasm, although several show unique localization in the nucleus, including hnRNP L and hnRNP I, which are concentrated in a limited number of bright spots in addition to diffuse nucleoplasmic staining (7, 11). Unlike previously characterized hnRNP proteins, including hnRNP L and hnRNP I, hnRNP F and hnRNP H were found to be concentrated in numerous discrete regions throughout the nucleoplasm. There is also some hnRNP F and hnRNP H signal detected in the cytoplasm. Although distinct, the observed localization of hnRNP F and hnRNP H is reminiscent of the distribution of a subset of pre-mRNA splicing factors, including snRNPs and the non-snRNP splicing factors SC-35 and SF2, which are found in regions in the nucleoplasm corresponding to perichromatin fibrils and interchromatin granules (39, 40). At present it is not certain whether these sites of concentrated splicing factors actually represent areas of active pre-mRNA processing, or are only sites for assembling or disassembling components needed for splicing elsewhere (40). It is also unclear whether a subset of hnRNP F and hnRNP H co-localize and interact with splicing factors at these sites, although this possibility is of obvious interest and will be the focus of future investigations. Immunopurification with antibodies to hnRNP F and hnRNP H identified complexes similar to those seen with antibodies to other hnRNP proteins (unpublished data, M.J.Matunis and G.Dreyfuss), however, it is possible that not all complexes containing hnRNP F and hnRNP

H are present in the soluble, nucleoplasmic fraction used for immunopurification.

In summary, the RNA-binding characteristics, amino acid sequence, and cellular localization of hnRNP F and hnRNP H make them distinct from previously characterized hnRNP proteins and suggest that they are important for the metabolism of guanosine-rich RNA sequences in the nucleus. The characterization of individual hnRNP proteins continues to demonstrate that each of these proteins has unique properties and suggests that they also have unique functions. The availability of monoclonal antibodies and cDNAs for hnRNP F and hnRNP H will make it possible to further study of the functions of these abundant hnRNP proteins.

ACKNOWLEDEGMENTS

We thank members of our laboratory, especially Erika Matunis, Megerditch Kiledjian, Christopher Burd and Haruhiko Siomi, for their comments and suggestions during the course of this work. We are also grateful to Vartkes Apkarian for synthesis of oligonucleotides, and to Erika Matunis for assistance with confocal microscopy. This research was supported by the Howard Hughes Medical Institute, by grants from the National Institutes of Health to GD, and the Medical Center Protein Chemistry Facility of the University of Pennsylvania Cancer Center Core Grant.

REFERENCES

1. Dreyfuss, G., M.J. Matunis, S. Piñol-Roma and C.G. Burd. 1993. *Annu. Review Biochem.* **62**:289–321.
2. Piñol-Roma, S., Y.D. Choi, M.J. Matunis, and G. Dreyfuss. 1988. *Genes & Dev.* **2**:215–227.
3. Cobianchi, F., Biamonti, G. Bassi, M.T., Buvoli, M., Riva, S. 1990. In *The Eukaryotic Nucleus: Structure and Function,* ed. P. Strauss, S. Wilson, 2:561. Caldwell, NJ: Telford.
4. Burd, C.G., M.S. Swanson, M. Görlach and G. Dreyfuss. 1989. *Proc. Natl. Acad. Sci. USA.* **86:** 9788–9792.
5. Swanson, M.S., T.Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. *Mol. Cell. Biol.* **7**:1731–39.
6. Brunel, F., P.M. Alzari, P. Ferrara and M.M. Zakin. 1991. *Nucleic Acids Res.* **19**:5237–5245.
7. Ghetti, A., S. Piñol-Roma, W.M. Michael, C. Morandi and G. Dreyfuss. 1992. *Nucleic Acids Res.* **14**:3671–3678.
8. Gil, A., P. A. Sharp, S. F. Jamison and M. A. Garcia-Blanco. 1991. *Genes & Dev.* **5**:1224–1236.
9. Patton, J.G., S.A. Mayer, P. Tempst and B. Nadal-Ginard. 1991. *Genes & Dev.* **5**:1237–1251.
10. Matunis, M. J., W. M. Michael and G. Dreyfuss. 1992. *Mol. Cell. Biol.* **12**:164–171.
11. Piñol-Roma, S., M.S. Swanson, J.G. Gall, and G. Dreyfuss. 1989. *J. Cell Biol.* **109**:2575–2587.
12. Datar, K.V., G. Dreyfuss and M.S. Swanson. 1993. *Nucleic Acids Res.* **21**:439–446.
13. Kiledjian, M. and G. Dreyfuss. 1992. *EMBO J.* **11**:2655–2664.
14. Bandziulis, R.J., M.S. Swanson and G. Dreyfuss. 1989. *Genes Dev.* **3**:431–437.
15. Dreyfuss, G., M.S. Swanson, and S. Piñol-Roma. 1988. *Trends Biochem. Sci.* **13**:86–91.
16. Kenan, D.J., C.C. Query and J.D. Keene. 1991. *Trends. Biochem. Sci.* **16**:214–220.
17. Görlach, M., M. Wittekind, R.A. Beckman, L. Mueller and G. Dreyfuss. 1992. *EMBO J.* **11**:3289–3295.
18. Wittekind, M., M. Gölach, M. Friedrichs, G. Dreyfuss and L. Mueller. 1992. *Biochemistry.* **31**:6254–6265.
19. Scherly D., W. Boelens, W.J. van Venrooij, N.A. Dathan, J. Hamm and I.W. Mattaj. 1989. *EMBO J.* **8**:4163–4170.
20. Lutz-Freyermuth, C., C.C. Query and J.D. Keene. 1990. *Proc. Natl. Acad. Sci. USA.* **87**:6393–6397.
21. Nagai, K., C. Oubridge, T.H. Jessen, J. Li and P.R. Evans. 1990. *Nature.* **346**:515–520.
22. Hoffman D.W., C.C. Query, B.L. Golden, S.W. White and J.D. Keene. 1991. *Proc. Natl. Acad. Sci. USA.* **88**:2495–2499.
23. Choi, Y.D., P.J. Grabowski, P.A. Sharp and G. Dreyfuss. 1986. *Science.* **231**:1534–1539.
24. Mayeda A. and A. R. Krainer. 1992. *Cell.* **68**:365–375.
25. Kumar, A. and S.H. Wilson. 1990. *Biochemistry.* **29**:10717–10722.
26. Munroe, S. H., X. Dong. 1992. *Proc. Natl. Acad. Sci. USA.* **89**:895–899.
27. Piñol-Roma, S. and G. Dreyfuss. 1992. *Nature.* **355**:730–732.
28. Swanson, M.S. and G. Dreyfuss. 1988. *Mol. Cell. Biol.* **8**:2237–2241.
29. Matunis, M.J., E.L. Matunis and G. Dreyfuss. *Mol. Cell. Biol.* **13**:6114–6123.
30. Swanson, M. S. and G. Dreyfuss. 1988. *EMBO J.* **11**:3519–3529.
31. Choi, Y.D. and G. Dreyfuss. 1984. *J. Cell Biol.* **99**:1997–2004.
32. Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood and S. B. H. Kent. 1987. *Proc. Natl. Acad. Sci. USA.* **84**:6970–6974.
33. Studier, F. W., A. H. Rosenberg, J. J. Dunn and J. W. Dubendorff. 1990. *Methods Enzymol.* **185**:62–89.
34, Choi, Y.D. and G. Dreyfuss. 1984b. *Proc. Natl. Acad. Sci. USA.* **81**:7471–7475.
35. Piñol-Roma, S., Y.D. Choi and G. Dreyfuss. 1990. *Methods Emzymol.* **181**:317–325.
36. O'Farrell, P.Z., H.M. Goodman, and P.H. O'Farrell. 1977. *Cell.* **12**:1133–1142.
37. Altschul, S.F., W. Gish, W. Miller, E.W. Myers and D.J. Lipman. 1990. *J. Mol. Biol.* **215**:403–410.
38. Scherly, D., W. Boelens, N.A. Dathan, W.J. van Venrooij and I.W. Mattaj. 1990. *Nature* **345**:502–506.
39. Spector, D.L., X.-D. Fu and T. Maniatis. 1991. *EMBO J.* **10**:3467–3481.
40. Spector, D.L. 1993. *Curr. Opin. Cell Biol.* **5**:442–448.
41. Wilk, H-E., G. Angeli, and K.P. Schafer. 1983. *Biochemistry.* **22**:4592–5000.
42. Conway, G., J. Wooley, T. Bibring and W. M. LeStourgeon. 1988. *Mol. Cell. Biol.* **8**:2884–2895.
43. Matunis, E.L., M.J. Matunis and G. Dreyfuss. 1993. *J. Cell Biol.* **121**:219–228.
44. Ishikawa, F., M.J. Matunis, G. Dreyfuss and T.R. Cech. 1993. *Mol. Cell. Biol.* **13**:4301–4310.
45. Engelbrecht, J., S. Knudsen and S. Brunak. 1992. *J. Mol. Biol.* **227**:108–113.
46. Green, M.R. 1991. *Annu. Rev. Cell Biol.* **7**:559–599.
47. Watakabe, A., K. Tanaka and Y. Shimura. 1993. *Genes and Dev.* **7**:407–418.
48. Qian, Z. and J. Wilusz. 1991. *Mol. Cell. Biol.* **11**:5312–5320.
49. Blackburn, E.H. 1991. *Nature.* **350**:569–572.
50. Clark, S.P., C.D. Lewis and G. Felsenfeld. 1990. *Nucleic Acids Res.* **18**:5119–5126.
51. Takimoto, M., E. Michelotti, M.J. Matunis, M. Avigan, H. Krutzsch, G. Dreyfuss and D. Levens. 1993. *J. Biol. Chem.* **268**:18249–18258.
52. Merrill, B.M., K.L. Stone, F. Cobianchi, S.H. Wilson and K.R. Williams. 1988. *J. Biol. Chem.* **263**:3307–3313.
53. Ellington, A.D. 1993. *Curr. Biol.* **3**:375–377.
54. Zimmerman, S.B., G.J. Cohen and D.R. Davies. 1975. *J. Mol. Biol.* **92**:181–192.
55. Kim, J., C. Cheong and P.B. Moore. 1991. *Nature.* **351**:331–332.
56. Dreyfuss, G., Y.D. Choi, and S.A. Adam. 1984. *Mol. Cell. Biol.* **4**:1104–1114.
57. Piñol-Roma, S. and G. Dreyfuss. 1993. *Mol. Cell. Biol.* **13**:5762–5770.
58. Portman, D.S. and G. Dreyfuss. 1994. *EMBO J.* **13**:213–221.
59. Lamond, A.I. and M. Carmo-Fonseca. 1993. *Trends Cell. Biol.*, **3**:198–204.