Annu. Rev. Biochem. 1993. 62:289–321

hnRNP PROTEINS AND THE BIOGENESIS OF mRNA

Gideon Dreyfuss, Michael J. Matunis, Serafín Piñol-Roma, and Christopher G. Burd

Howard Hughes Medical Institute and Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6148

KEY WORDS: heterogeneous nuclear ribonucleoproteins, hnRNA, pre-mRNA, RNA processing, nucleus

CONTENTS

0066-4154/93/0701-0289$02.00

I. INTRODUCTION

Messenger RNAs (mRNAs) are formed in the nuclei of eukaryotic cells by extensive posttranscriptional processing of primary transcripts of protein-coding genes (1, 2). These transcripts are produced by RNA polymerase II and are termed heterogeneous nuclear RNAs (hnRNAs), a historical term that describes their size heterogeneity and cellular location. The terms hnRNA and pre-mRNA are often used interchangeably, although only a subset of hnRNAs may actually be precursors to mRNAs, while the rest, their function obscure, turn over in the nucleus. From the time hnRNAs emerge from the transcription complex, and throughout the time they are in the nucleus, they are associated with proteins. The collective term for the proteins that bind hnRNAs, and that are not stable components of other classes of ribonucleoprotein (RNP) complexes such as small nuclear RNPs (snRNPs) (for reviews of these complexes see 3–5), is hnRNP proteins (6). The full range of functions and mechanism of action of hnRNP proteins is not yet known. It can be anticipated, however, that as hnRNA-binding proteins, hnRNP proteins influence the structure of hnRNAs and facilitate or hinder the interaction of hnRNA sequences with other components that are needed for processing of pre-mRNAs, thus affecting the fate of hnRNAs. hnRNP proteins may also play important roles in the interaction of hnRNA with other nuclear structures, in nucleocytoplasmic transport of mRNA, and in other cellular processes. Together, the hnRNP proteins are as abundant in growing vertebrate cells as histones, and hnRNA-hnRNP protein complexes (hnRNP complexes) are thus also of interest because they are major nuclear structures. In addition, what has been learned from the study of hnRNP proteins turned out to be extremely instructive for other RNA-binding proteins, including those that control developmentally important pathways (7), snRNP proteins, and mRNA-binding (mRNP) proteins. Once formed, mRNAs are transported to the cytoplasm where mature mRNAs associate with a different set of proteins, the mRNP proteins, which are likely to be involved in the regulation of the translation and stability of mRNAs and in their cellular location (8).

Much progress has been made in the understanding of hnRNP proteins and hnRNP complexes over the past several years, but many questions of fundamental importance still need to be answered. There are several central questions on hnRNP proteins: What are their characteristics (e.g. structure, RNA-binding, protein-protein interaction, localization, posttranslational modifications, amount)? What is their arrangement on RNAs? What are their functions? Obviously, these questions are intimately related, and they are divided in this way here primarily to facilitate thinking about and reviewing these issues. In the following sections we briefly summarize the currently available information relating to these issues, and outline what we consider

important to further their understanding. The earlier work in this field has been reviewed previously (8–13), so this review emphasizes recent developments.

II. THE hnRNP PROTEINS

Definition and Experimental Criteria

Nascent, chromatin-associated hnRNAs associate with proteins and snRNP particles. This can most vividly be seen by microscopy on the amphibian oocyte lampbrush chromosomes, and it has been recognized for decades (14–17). RNP complexes are, with very few exceptions, multiprotein complexes (see 18 for review), and the complexes that assemble on hnRNAs are composed of a particularly large number of proteins (19). One of the major tasks in studying hnRNP complexes has been to identify their composition definitively. Nascent hnRNA-hnRNP-snRNP complexes contain many additional proteins that are involved in transcription and RNA processing. However, because these complexes are insoluble, their biochemical analysis is difficult and their complete composition is not known. Several methods have been developed for the isolation and characterization of soluble nucleoplasmic hnRNP complexes, or complexes released from nuclei after limited RNase digestion.

hnRNP complexes are labile (e.g. to RNases), and due to shortcomings of the earlier methods, unambiguous identification of authentic hnRNP proteins was difficult. The first method used to isolate hnRNP complexes relied on cosedimentation of proteins and hnRNA through sucrose density gradients (20–27). This method is lengthy, subjects the complexes to deleterious conditions (RNases, proteases, and centrifugal drag), and it can lead to both loss of hnRNP proteins and to nonspecific binding of proteins. Moreover, hnRNP complexes cannot be resolved from other cellular structures that have similar sedimentation properties. Despite the limitations of sucrose gradients for obtaining pure and intact complexes, data from these studies led gradually to the consensus that hnRNAs in vertebrate cells are associated with a group of proteins in the 30–43-kDa range, which include the hnRNP A, B, and C groups. Proteins in direct contact with hnRNA in vivo have been subsequently identified by UV-induced RNA-protein crosslinking (28–35). After UV irradiation of intact cells, covalent protein-RNA complexes are purified from nuclei by oligo(dT)-chromatography under protein-denaturing conditions. This procedure overcomes the problems of specificity associated with isolating complexes by sucrose gradient sedimentation. The major limitations of this method are the dependence on the photoreactivity of the particular proteins and the RNA sequences, that the proteins are denatured during isolation, and

that only proteins crosslinked to poly(A)$^+$ RNA are identified. This method definitively identified the 30–43-kDa proteins as in vivo hnRNA-binding proteins and identified additional proteins of 120, 68, and 53 kDa. The most recent and specific method for the isolation of hnRNP complexes is immunopurification with monoclonal antibodies, which were initially raised against authentic hnRNA-contacting proteins purified by UV crosslinking in vivo (19, 36). The immunopurification procedure is specific and rapid, and it yields pure, intact hnRNP complexes. This procedure has been particularly useful for identifying the proteins associated with hnRNA more definitively than was possible with previously used methods, and it led to the discovery of more than 20 proteins that are components of hnRNP complexes in human cells.

Although the UV-crosslinking and immunopurification methods have inherent shortcomings, in concert they provide a powerful set of experimental criteria for the identification of hnRNP proteins. A general theme that has emerged is that most, if not all, hnRNP proteins that have been identified by these methods are RNA-binding proteins (19, 37). It is therefore possible to think of hnRNP proteins as all of the proteins that bind hnRNAs and that are not stable components of other classes of RNP complexes such as snRNPs. This definition provides an important unifying theme, as it does not make a distinction between proteins previously thought of as "hnRNP proteins" (e.g. A1, A2, C1, C2, I, etc) and "RNA processing factors" (e.g. U2AF, ASF/SF2, CStF, etc; for reviews, see 38, 39). Recent information on the structure, RNA-binding activities, and functions of these proteins makes it difficult to distinguish meaningfully between them. Proteins that interact with hnRNP proteins and with the hnRNA-hnRNP-snRNP complexes solely by protein-protein interaction have not been detected so far, although we expect that such associated proteins exist. Another important general theme is that there is a large number of hnRNP proteins. The characterization of hnRNP proteins has therefore turned out to be a considerable undertaking, but progress has been rewarding in what has already been learned. This review focuses on the more abundant group of hnRNP proteins, and emphasizes the human hnRNP proteins, as these are the best characterized.

Human

The overall protein composition of hnRNP complexes immunopurified from nucleoplasm of growing HeLa cells (the post-chromatin, post-nucleolar fraction prepared at 100 mM NaCl) is shown in Figure 1 (19). As it includes all soluble hnRNP complexes, the protein composition of individual hnRNPs cannot be determined by such immunopurification from total nucleoplasm. About 20 major proteins, or groups of proteins, are resolved by two-dimensional gel electrophoresis; these are designated A1 (34 kDa) to U (120 kDa)

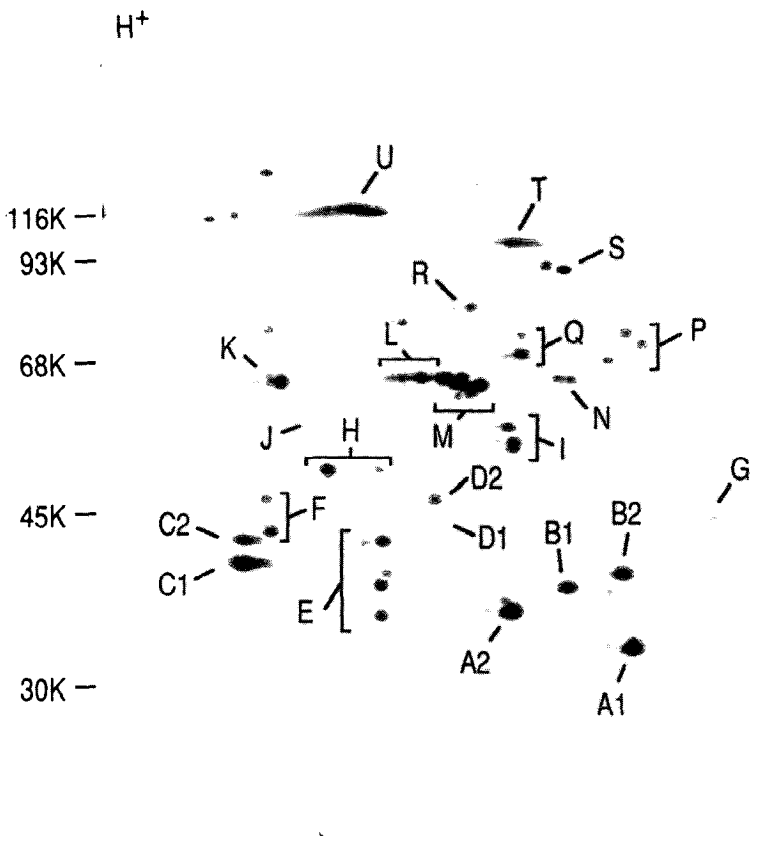

Figure 1 Protein composition of hnRNP complexes immunopurified with a monoclonal antibody, 4F4, to the C proteins. The hnRNP complexes were immunopurified from the nucleoplasm of [^{35}S]methionine-labeled HeLa cells (19, 36). The proteins were separated by non-equilibrium pH gradient gel electrophoresis (NEPHGE) in the first dimension and by SDS-PAGE in the second dimension, and visualized by fluorography.

(19). In addition, immunopurified hnRNP complexes contain RNAs larger than 10,000 nucleotides (36). The hnRNP proteins are among the most abundant proteins in the nucleus (8). The hnRNP proteins A1 and C1, for example, are much more abundant than U1 snRNP (D. S. Portman, G. Dreyfuss, unpublished observations). While the A, B, and C proteins (25), initially referred to as the "core" hnRNP proteins, are abundant components of immunopurified hnRNP complexes, many other proteins of similar

abundance are also apparent. In addition, numerous less abundant proteins, including snRNP proteins, can be visualized with longer fluorographic exposures. hnRNP complexes of similar composition are isolated with monoclonal antibodies to various hnRNP proteins, including A1 (19), C (36, 40), D (S. Piñol-Roma, G. Dreyfuss, unpublished observations), K (41), L (42), and U (36), indicating that all of these proteins are common constituents of the same supramolecular complexes. Table 1 lists the hnRNP proteins and details some of their key features. In the following section, these characteristics are described for the proteins that have been published. The RNA-binding activity and the functions of the proteins, where known, are discussed in the relevant subsequent sections.

THE A/B PROTEINS (M_r = 34,000–40,000 BY SDS-PAGE, pI = 8.4–9.0) By immunofluorescence microscopy the A/B proteins appear to be confined to

Table 1 The major human hnRNP proteins

Protein	M_r(kDa)/pI[a]	Structural motifs	Comments	Refs.
A1	34/9.0–9.1	2xRBD-Gly	contains DMA, may be phosphorylated	(25, 27, 46–50)
A2/B1	36 & 38/8.4–8.8	2xRBD-Gly	contains DMA B1 identical to A2 except for an 11-aa insert	(25, 27, 45)
B2	39/9.0	—		
C1/C2	41 & 43/5.9	RBD-AspGlu	phosphorylated, avid binding to poly(U), nuclear localization signal, C2 identical to C1 except for a 13-aa insert	(32, 37, 45, 66, 68, 98)
D	44–48/7.7–7.8	—		
E	36–43/7.3	RBDs; incomplete[c]	avid binding to poly(G)	(37)
G	43/9.5	RBDs; incomplete[c]		
F/H	53 & 56/6.1–7.1	RBDs; incomplete[c]	avid binding to poly(G)	(37)
I	59/8.5	4xRBD[b]	identical to the PTB	(69–72)
K/J	62 & 68/6.1–6.7	KH motif	avid binding to poly(C)	(37, 41, 73b)
L	68/7.4–7.7	4xRBD[b]		(42, 69)
M	68/7.8–8.2	4xRBD[b]		
N	70/8.7–8.9	—		
P	72/9.0	—	avid binding to poly(A)	(37)
Q	76–77/8.3	—		
R	82/8.0	—		
S	105/8.8	—		
T	113/8.4	—		
U	120/6.6–7.2	RGG box	phosphorylated, nuclear localization signal	(32, 75)

[a] M_r estimated from SDS-polyacrylamide gel electrophoresis and pI estimated from isoelectrofocusing gels
[b] Noncanonical RNP motif
[c] Unpublished observations

the nucleoplasm in interphase cells (42, 43), but they also shuttle between the nucleus and the cytoplasm (44). All of the A/B proteins that have been sequenced (A1, A2, B1) have a similar general structure: they contain two RNP-motif RNA-binding domains (RBDs) and a glycine-rich auxiliary domain at the carboxyl terminus (referred to as 2×RBD-Gly) (45–50). A2 and B1 cDNAs are identical except for 36 in-frame nucleotides in B1, probably derived by alternative splicing, that add 12 amino acids near the amino terminus of B1 (45). The RBDs of A2 and B1 have approximately 80% amino acid identity with those of A1, while the glycine-rich auxiliary domain is considerably more divergent (less than 30% identity) (45). A1 is the only hnRNP protein whose gene has been sequenced and whose gene promoter elements have been studied in detail (49). Several variants of A1 have been characterized; one contains a 50-amino-acid insert in the glycine-rich domain (48), and variants with specific amino acid substitutions have also been reported (47, 51). Diversity among the A/B proteins is also generated by posttranslational modifications, including methylation of arginines and phosphorylation (discussed below). The amino acid sequence of A1 is highly conserved among vertebrates—100% between human and rat and 92% between human and *Xenopus laevis* (52, 53), and A/B proteins are immunologically related (43, 54). Autoantibodies specific for A1 have been reported (55, 56), and autoantibodies to A2 have recently been found in about 33% of patients with rheumatoid arthritis (57). hnRNP A/B-like proteins (30% overall identity to human) have also been characterized from invertebrates, including *Drosophila melanogaster* and grasshopper (54, 58–61). Considerable evidence has accumulated to suggest that the A/B proteins have important functions in pre-mRNA processing, and these are discussed below. The amount of A1 appears to change during the cell cycle and with the state of cell proliferation (62–64). The significance of this is not clear.

THE C1/C2 PROTEINS (M_r = 41,000 AND 43,000 BY SDS-PAGE, pI = 5.9) Immunofluorescence microscopy with monoclonal antibodies shows that the C1 and C2 hnRNP proteins are confined to the nucleus of interphase cells (40, 44). The sequence of the human C2 cDNA is identical to that of C1 except for an extra 39 in-frame nucleotides, probably derived by alternative splicing, that add 13 amino acids near the middle of the C2 protein (45, 65–67). The C proteins contain two distinct parts: an amino terminal RNP-motif RBD and a carboxyl terminal negatively charged segment that contains a putative NTP-binding site and potential phosphorylation sites for casein kinase II (66). The C proteins are phosphorylated in vivo (see below), and they are highly conserved among vertebrates (40, 68).

THE I PROTEIN (M_r = 58,000 BY SDS-PAGE, pI = 8.5) Immunofluorescence microscopy with monoclonal antibodies to hnRNP I localize it to the

nucleoplasm of interphase cells, and it is also concentrated in a unique, unidentified perinucleolar structure (69). Sequences of cDNA clones for hnRNP I predict several isoforms that are likely to be derived by alternative splicing (69–72). The predicted structure of hnRNP I is highly related to that of hnRNP L; each contains four RNP motifs, but lacks the canonical consensus sequences RNP1 and RNP2. hnRNP I is released from hnRNP complexes by nuclease digestion more readily than are most other proteins, suggesting that it has a unique association with the complex and may be bound to hnRNA structures that are particularly exposed (69). Interestingly, hnRNP I is the same protein as the recently described polypyrimidine-binding protein (PTB) that binds preferentially to the polypyrimidine tract near the 3′-end of introns (71, 72).

THE K/J PROTEINS (M_r = 66,000 AND 64,000 BY SDS-PAGE, pI = 6.1–6.4) Immunofluorescence microscopy with monoclonal antibodies to the K and J proteins shows a general nucleoplasmic staining in human cells (41). K and J are immunologically related. The predicted sequence of K reveals a novel type of hnRNA-binding protein as it does not contain RNP motifs and shows no extensive similarity to any known proteins (41). hnRNP K does, however, contain two internal repeats as well as Gly-Arg-Gly-Gly and Gly-Arg-Gly-Gly-Phe sequences, which occur frequently in many RNA-binding proteins (73a, 73b). hnRNP K and J can be detected immunologically in a number of vertebrate organisms. *X. laevis* hnRNP K is a 47-kDa protein that is 90% identical to its human 66-kDa counterpart (73b). hnRNP K and J bind tenaciously to poly(C), and are the major oligo/poly(C)-binding proteins in human HeLa cells (41).

THE L PROTEIN (M_r = 64,000–68,000 BY SDS-PAGE, pI = 7.4–7.7) Monoclonal antibodies to L show strong staining of discrete non-nucleolar structures in addition to a general nucleoplasmic staining (42). hnRNP L contains glycine- and proline-rich domains and four, approximately 80-amino-acid segments that are distantly related to the RNP motif (42, 74). Sequence comparison reveals that hnRNP L is most similar in structure to hnRNP I (69). Interestingly, the L protein is associated with the majority of non-nucleolar nascent transcripts on lampbrush chromosomes from the newt, *Nothophthalmus viridescens,* but it is preferentially concentrated on the transcripts of the landmark giant loops (42).

THE U PROTEIN (M_r = 120,000 BY SDS-PAGE, pI = 6.6–7.2) Immunofluorescence microscopy with monoclonal antibodies for U show that it is confined to the nucleoplasm (32). hnRNP U is an abundant phosphoprotein (32). It

contains no extensive sequence homology to any known proteins. It has an acidic amino terminus, a glycine-rich carboxyl-terminus, a putative NTP-binding site, a putative nuclear localization signal, multiple potential phosphorylation sites, and an RGG box (75).

OTHER hnRNP PROTEINS Little is known about the remaining abundant hnRNP proteins. The characterization, cloning, and sequencing of several of them, including D, E, F, H, and M, is in progress. Proteins bound to the 5′-cap of hnRNAs remain to be identified and characterized. Candidate nuclear cap-binding proteins have been reported (76–78).

Other Organisms

The composition of hnRNP complexes isolated from cells of several other vertebrates, including rodents, avians, and amphibians, is very similar to that of human hnRNP complexes (25–27, 36, 43, 79). Notably, most of the major hnRNP proteins (A through U) appear to be highly conserved among vertebrates, both immunologically and structurally (36, 40–43, 52, 68). A comprehensive survey of the hnRNP protein composition of different cell types in the same organism has not been reported and would be very informative.

Considerable information has also become available about invertebrate hnRNP proteins, particularly from the fruitfly *Drosophila melanogaster*. hnRNP complexes immunopurified from *D. melanogaster* contain more than 10 abundant proteins with apparent molecular weights between 36,000 and 75,000 (54). Monoclonal antibodies to many of these proteins have been generated, and their sequences and genomic localization have been determined (54, 58–60). Two-dimensional gel electrophoresis and immunoblotting reveal that many of the proteins, like their human counterparts, are present as groups of immunologically related isoforms. Many of these are generated by alternative pre-mRNA processing of common primary transcripts (58–60). All of the major *D. melanogaster* hnRNP proteins cloned and sequenced thus far have a predicted structure similar to that of the human A/B proteins—2×RBD-Gly (58–60). Genetic analysis has also identified several *D. melanogaster* loci encoding for proteins with RNP motifs (80–85). These include the hnRNA-binding proteins of the *sex-lethal* gene (86, 87) and the *tra-2* gene (88), which are involved in the sex determination pathway (89, 90).

Information about an increasing number of candidate hnRNP proteins from other divergent organisms, including plants and fungi, is also accumulating, and in most cases these have been identified as proteins having sequence homologies to known human or *D. melanogaster* proteins and with other characteristics common to hnRNP proteins (91–97). However, there is still little definitive information about hnRNP complexes in these organisms.

Studies in these organisms will facilitate the application of genetic and cytological approaches to investigate the function of hnRNP proteins.

Posttranslational Modifications

In addition to alternative pre-mRNA processing, which generates a considerable diversity through the formation of isoforms, the complexity of hnRNP proteins is further increased by posttranslational modifications. The two modifications described so far are phosphorylation of serines and threonines, and methylation of arginines. The A/B proteins (26, 27), the C proteins (32, 98), and the hnRNP U (32) protein are all phosphorylated in vivo. Both A1 and A2 are methylated on arginine residues in the glycine-rich carboxyl domain (25, 27, 99, 100). Other hnRNP proteins that contain potential sites for arginine methylation (arginine residues flanked by glycines and in the proximity of phenylalanine) include the hnRNP U and K proteins (41, 75). The functions of these modifications have not been determined, but they are likely to modulate the specific interactions of the proteins with other proteins and with RNA. Extensive modifications, most of which have not been characterized, are also detected for many of the hnRNP proteins during mitosis, and these may also have regulatory roles in the localization of these proteins (63, 101).

III. THE RNA-BINDING ACTIVITY OF hnRNP PROTEINS

The sequencing of cDNA clones for many different hnRNP proteins has revealed that nearly all hnRNP proteins possess RNA-binding motifs, and experiments with the individual proteins demonstrated their RNA-binding activity. A number of studies have shown that hnRNP proteins can bind in vitro to many different single-stranded ribo- and deoxyribo-polynucleotides; this was taken to indicate that hnRNP proteins bind to hnRNA without regard to sequence (102–106). Consistent with this, most of them can be purified by affinity chromatography on single-stranded DNA (ssDNA)-agarose, to which they bind in a heparin- and moderate- or high-salt-resistant manner (19, 107, 108). Some of the hnRNP proteins (a subset of the E proteins, H, and F), however, do not bind single-stranded DNA, although they bind tenaciously to RNA (19, 37).

Subsequent more stringent in vitro assays demonstrated that hnRNP proteins have different preferences for specific sequences. The binding of hnRNP proteins to immobilized ribohomopolymers at various salt concentrations was studied (37). At 2 M NaCl the hnRNP F, P, H, M, and a subset of the E proteins bind poly(G), hnRNP P binds poly(A), the hnRNP C and M proteins bind poly(U), and the K and J proteins bind poly(C). The binding under these conditions demonstrates the striking avidity of the hnRNP proteins for their

preferred RNAs. These results indicated that different hnRNP proteins discriminate among different RNAs, and these properties provide a useful aid in the classification and the purification of hnRNP proteins. They also allow certain predictions as to where on pre-mRNAs these proteins are likely to bind avidly (with the potential functional implications that such binding specificity may have).

Sequence-specific RNA-binding by several hnRNP proteins has also been demonstrated by photochemical crosslinking and by RNA co-immunoprecipitation experiments. In crosslinking experiments a binding site for the hnRNP C proteins that consists of a stretch of five uridines was mapped on pre-mRNA polyadenylation substrates (109, 110). A similar approach using several pre-mRNA splicing substrates identified hnRNP I/PTB as a sequence-discriminating protein that crosslinks to the uridine-rich polypyrimidine stretch found at the 3′ end of most introns (71, 72). RNase T1 digestion and immunoprecipitations demonstrated that a subset of hnRNP proteins (A1, C, and D) bind preferentially to sequences found in introns at or near the 3′ splice site (111). The binding of hnRNP A1 was particularly sensitive to mutations in the highly conserved 3′ splice site AG. Studies by Riva and colleagues (112) confirmed these findings using purified recombinant A1 and synthetic deoxyoligonucleotides.

The studies mentioned above have identified preferred binding sites for several hnRNP proteins on a very limited array of RNA sequences. Ultimately it is necessary to know the intrinsic RNA-binding preference of each hnRNP protein. Recently, selection amplification from pools of random sequence RNAs (113, 114) was used to determine the preferred binding sites of several hnRNP proteins. hnRNP C1 selected (from a randomized pool of 20-mers) RNA molecules containing oligouridine stretches (U_6 stretches were the most prevalent), and the majority of RNA molecules selected by A1 contained sequences that bear resemblance to 5′ and 3′ splice sites (C. G. Burd, G. Dreyfuss, unpublished observations). Each of these proteins selected RNA molecules containing identical stretches of six contiguous bases unique for each protein, suggesting that this is the minimal length of RNA that they specifically recognize. These studies confirm that C1 and A1, and probably all hnRNP proteins, have RNA sequence binding specificity; similar experiments will allow determination of preferred binding sites for all the hnRNP proteins. Clearly though, these proteins have a spectrum of binding affinities; some sequences constitute higher-affinity binding sites, while other sequences are lower-affinity, relatively nonspecific binding sites. Thus, the term specificity as used here indicates binding preference; it does not mean exclusivity. The dissociation constants for higher-affinity sequences and for random sequences will be important parameters to determine. As the major hnRNP proteins are so abundant, it is almost certain that they are in vast

excess over the number of higher-affinity binding sites. Cooperative interactions could affect extensive contiguous binding of hnRNP proteins such that the hnRNAs may form a fibril that is completely coated with hnRNP proteins. So far, cooperative binding interactions have been reported for hnRNP A1 (107, 115, 116), but for other hnRNP proteins the influence of protein-protein interactions on binding to RNA has not been explored in detail. It is interesting and significant that the preferred binding sites found so far are sequences that are important for pre-mRNA processing; this suggests functional relevance. High-affinity sites may ensure that specialized complexes will form at sites on the hnRNA where hnRNP proteins, or the proteins they may recruit, perform essential functions in the processing pathways of hnRNAs. The relatively non-sequence-specific binding may facilitate the search for high-affinity sites by reducing the dimensionality of space through which the protein must diffuse. Finally, the RNA-binding specificity indicates that the proteins are specialized, and it partly explains why there are so many hnRNP proteins.

IV. THE STRUCTURE OF hnRNP PROTEINS

Detailed knowledge of the structure of hnRNP proteins is essential for understanding their function. The amino acid sequences of many hnRNP proteins, along with mutagenesis and binding experiments, have led to the identification of several different RNA-binding motifs. A common theme that has emerged from cDNA sequence studies is that hnRNP proteins, in fact most RNA-binding proteins, have a modular structure. That is, they possess one or more RNA-binding modules and at least one other domain, an auxiliary domain, that probably mediates protein-protein interactions.

The RNP Motif

The most common RNA-binding motif in hnRNP proteins is the RNP consensus RNA-binding domain (CS-RBD or RNP motif) (6). The RNP motif has also been called the RNA Recognition Motif (RRM) (117) and RNP 80 (118). This type of domain has been found in many RNA-binding proteins of the nucleus, cytoplasm, and cytoplasmic organelles, including hnRNA-, mRNA-, snRNA-, and pre-rRNA-binding proteins in animal, plant, and fungal cells (6, 7, 74). The hallmarks of the RNP motif are two consensus sequences, RNP1 and RNP2, located about 30 amino acids apart in this domain of approximately 90 amino acids (6). The RNP1 octapeptide, Lys/Arg-Gly-Phe/Tyr-Gly/Ala-Phe-Val-X-Phe/Tyr, is the most highly conserved segment of the RNP motif; it was noticed on the basis of primary sequence similarity between the hnRNP A1 protein and the mRNA poly(A)-binding protein (119). RNP2 is a less well conserved hexapeptide sequence that is rich in aromatic and aliphatic amino acids (6). In addition, several isolated positions throughout

the RNP motif are highly conserved (6, 7). Experimental evidence that the RNP motif is indeed an RNA-binding domain has been provided for several snRNP proteins (117, 118, 120) and hnRNP proteins (121). In addition, peptide binding studies (122) and photochemical crosslinking of phenylalanines within RNP1 and RNP2 have directly implicated these RNP consensus sequences in RNA binding (123).

The three-dimensional structure of one of the two RBDs of the U1 snRNP A protein (amino acids 1–102) at a resolution of 2.8 Å was determined using X-ray crystallographic methods by Nagai et al (124), and the global fold of the same domain was deduced by Hoffman et al using nuclear magnetic resonance (NMR) techniques (125). These structural studies showed that the RBD has a β1-α1-β2-β3-α2-β4 (α = α helix and β = β sheet) structure. The four β strands form an antiparallel β sheet that packs against the two α helices. The RNP1 and RNP2 consensus sequences are juxtaposed on the adjacent central antiparallel strands (β-strands 3 and 1, respectively). The structure of the hnRNP C RBD (amino acids 2–94) in solution was recently determined by multidimensional NMR techniques (126). The overall solution structure of this RBD (Figure 2) is very similar to that of the U1 snRNP A RBD, but there are important differences. Most notable is the complete absence, in the RBD of the C proteins, of the loop region connecting β strands 2 and 3. The corresponding region in the U1 A RBD has an insertion of five amino acids in this loop, which confers at least some of the specificity of this domain toward U1 snRNA. This loop is the region where different RBDs exhibit the greatest variability in length and in residue type, and it may be a key determinant of specificity of the RBD. The issue of specificity determinants of RBDs has been reviewed recently (74, 127). Although the structure of the RBD is of fundamental importance, the structure did not explain how the RBD functions in RNA binding.

There are two key questions to understanding the function of hnRNP proteins that center around the activity of the RBD: (*a*) What amino acids of the RBD are involved in the interaction with the RNA? and (*b*) What are the consequences to the RNA from the binding of the RBD? The role of specific amino acids in the binding of the U1 A RBD to U1 snRNA was studied by mutagenesis of many potential hydrogen-bonding residues and basic residues on the β sheet surface of the RBD (124, 128). In these studies, however, only a few important putative RNA:RBD contacts were identified (128). The U1 A RBD specifically and stably binds a unique stem-loop structure (118, 129–131), which retains its structure upon binding (128). The RNA substrates for hnRNP proteins, however, appear to be single-stranded, and thus the binding of the U1 A RBD to its stem-loop substrate may differ in important ways from that of hnRNP proteins to their substrates.

The interaction of the hnRNP C RBD with a preferred RNA substrate (U_8)

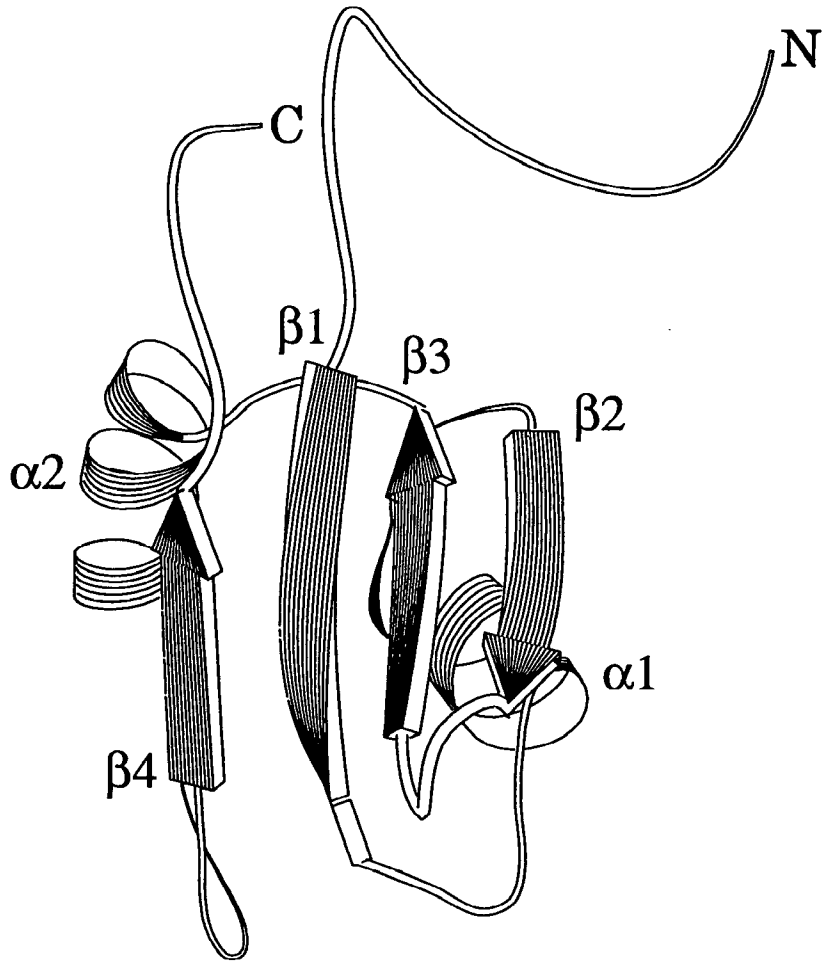

Figure 2 Schematic ribbon drawing of the hnRNP C RBD (residues 2–94). The arrows represent the four antiparallel β strands, and the curled ribbons represent the two α helices. The labels β1–β4 indicate the four strands of the β sheet; the labels α1 and α2 indicate each α helix. The RNP1 and RNP2 consensus sequences are juxtaposed on the adjacent central antiparallel strands (β strands 3 and 1, respectively). N and C denote the amino and carboxy termini of the domain, respectively (121, 126).

was recently studied by NMR methods (121). These experiments revealed that the overall structure of the RBD did not change upon binding to RNA and identified candidate amino acids that are involved in the binding to RNA. Perturbations of the chemical shifts of amide atoms of the RBD upon addition

of U_8 occurred in a large number of residues. Significant changes in the chemical shift most likely result from either direct contact with the RNA or the close proximity of amino acid residues to the RNA. Almost all the affected residues were located in the β sheet and especially in the contiguous amino- and carboxy-terminal regions. In contrast, the residues of the α helices were relatively unperturbed. These results suggest that most of the amino acids that participate in RNA binding are localized to the β sheet surface and to the contiguous termini of the RBD. These structural elements of the RBD, therefore, appear to provide an exposed surface that can serve as a platform to which the RNA binds. An important consequence of this mode of RNA binding is that the RNA, when bound, remains exposed (as opposed to buried in a binding pocket) and thus accessible to other pre-mRNA processing factors. An important role for the terminal regions of the RBD in RNA binding is supported by other findings from deletional analysis (C. G. Burd, M. Görlach, G. Dreyfuss, unpublished observations; 132) and, as these regions of the RBDs are among the most variable among this family of proteins (6, 7), they may also be important determinants of specificity of the proteins.

Very little is known about the structure of RNA as it is bound to an RBD. Circular dichroism measurements have demonstrated that both A1 and a proteolytic fragment of A1 that contains its two RBDs (UP1) can partially unstack the bases of both single-stranded and double-stranded RNA molecules (107, 115, 133). High-resolution structural studies of RNA-RBD complexes—which are essential for understanding the function of this family of RNA-binding proteins—are under way.

Many RNP motif proteins contain multiple RBDs, and it is conceivable that they can bind to more than one RNA segment simultaneously. This possibility was suggested by comparisons of the sequences of rat hnRNP A1 and a *D. melanogaster* 2×RBD-Gly protein, and also of the human and yeast PABPs (6). Each RBD appears to be evolutionarily conserved independently, and thus it is likely that each has a different function. In fact, a unique consensus for each of the individual RBDs of the 2×RBD-Gly proteins was derived (60), further indicating the functional diversity of each domain. Therefore, it is likely that in proteins containing multiple RBDs, each RBD could have a unique RNA-binding specificity. A recent biophysical study of A1 demonstrated that the linkage between the two RBDs of A1 is flexible, such that both could function independently (134).

The RGG Box

The hnRNP U protein does not contain an RNP motif. Deletional mapping of the RNA-binding activity of this protein localized the RNA-binding activity of the protein to a 26-amino-acid peptide containing four Arg-Gly-Gly (RGG) repeats with several interspersed aromatic residues (75). Several other known

RNA-binding proteins contain RGG repeats interspersed with aromatic amino acids at a characteristic spacing similar to that found in hnRNP U. This RG-rich region has been termed the RGG box, and it may represent a minimal RNA-binding domain. The number of RGG (and GRG or RRG) repeats that are required for RNA-binding activity is presently uncertain. Comparison of RGG repeats from many RNA-binding proteins suggested a consensus RGG box: G R G G N/S F X G R G G X X R G G X R G G F/Y G R/G R/G G G. It is striking that RGG boxes have a strong positive charge (+3 to +9) but there are no lysines present, suggesting that arginine is essential for the RNA-binding activity of this motif. The arginines of the RGG box may bind RNA in a similar fashion to that of the HIV tat protein to the TAR element (135–137). In that case an arginine side chain makes critical RNA contacts with RNA that lysine cannot make (135, 136). A recent structural analysis of an RGG box–containing peptide from nucleolin indicated that RGGF makes a β turn; a cluster of such repeats can form a spiral structure that can unstack RNA (138). It is also interesting to note that many of the arginines within the RGG box are potential sites for dimethylation—a known modification of several hnRNP proteins (73) that could serve to regulate the RNA-binding activity of these proteins.

Other Types of RNA-Binding Domains

Proteins bind RNA by a wide variety of different motifs. Predicted noncanonical RBDs have been found in the hnRNP I and L proteins (42, 69, 71, 72, 74). Examples of other types of RNA-binding domains include zinc fingers such as found in the 5S RNA-binding protein TFIIIA (139, 140), the arginine cluster of the HIV tat protein (141, 142), a methionine-rich domain in the SRP 53-kDa protein (143), and several unique RNA-binding domains found in RNA viruses and ribosomal proteins. Surprisingly, no significant similarity to any of these proteins has been found so far in any hnRNP proteins. A possible new RNA-binding motif was recently found in hnRNP K (41). Close to the termini of K are located 45-amino-acid repeats that are almost completely conserved between frogs and humans (73b). In addition, the protein contains three RGG peptides located between the repeats. The 45-amino-acid repeats (KH motifs) show significant homology to several known nucleic acid–binding proteins, including the archaebacterial ribosomal protein S3 and the yeast protein MER1. The KH motif may therefore be involved in RNA binding (73b).

Auxiliary Domains

RNP proteins in general have a modular structure—that is, they contain one or more RBDs and one or more other domains that are termed auxiliary

domains (7). The functional significance of auxiliary domains is a relatively unexplored area, but it is likely that these regions mediate protein-protein interactions and they may also act to localize the proteins within the cell. The most frequently found type of auxiliary domain is the glycine-rich type found in the 2×RBD-Gly proteins, such as the hnRNP A/B proteins. In A1, the glycine-rich carboxyl domain confers cooperative RNA-binding and therefore probably mediates A1-A1 interactions (107). This domain, which contains an RGG box, has also been shown to bind RNA (116), but not as avidly as other RGG box–containing proteins (e.g. hnRNP U) (75). It is likely that some auxiliary domains mediate not only homotypic interactions but also heterotypic interactions such that the binding of one hnRNP protein could significantly affect the binding of other proteins.

The auxiliary domain of the hnRNP C proteins, found at the carboxyl half of the protein, is very rich in acidic amino acids, and it also contains a putative NTP-binding site and a nuclear localization signal (66, 68). The U protein also contains a putative NTP-binding site and a putative nuclear localization signal (75), but it is not known if either the C proteins or U actually bind nucleotide triphosphates. The auxiliary domains of several of the hnRNP proteins bear resemblance to eukaryotic transcription factors in that they possess clusters rich in a few particular amino acids. For example, hnRNP U has a stretch of 50 amino acids that is composed of 28% glutamine and a region rich in acidic amino acids, and K and L contain clusters of prolines that resemble CCAAT transcription factors (CTF) (41, 42, 144).

V. LOCALIZATION, TRANSPORT, AND SHUTTLING OF hnRNP PROTEINS

Nuclear Location of hnRNP Proteins

Immunofluorescence microscopy with most of the antibodies to hnRNP proteins shows general nucleoplasmic localization of these proteins with little or no staining in the nucleoli and in the cytoplasm (32, 40–43, 101, 145). A considerable proportion of the nuclear signal probably represents hnRNP proteins bound to nascent RNA polymerase II transcripts (see following section), and the rest may result from hnRNP proteins bound to fully processed RNAs that are not yet transported to the cytoplasm, or to RNAs that are at various stages of processing. As discussed below, there is presently no evidence for free (i.e. not RNA-bound) hnRNP proteins in the nucleus. Immunoelectron microscopy studies have localized hnRNP proteins mostly to perichromatin fibrils (146, 147), which had been previously identified as the sites of formation and/or greatest accumulation of hnRNA (148, 149).

The distribution of hnRNP proteins is quite different from that of snRNPs,

which are also abundant in the nucleus. Although snRNPs are also found throughout the nucleoplasm, they concentrate in multiple discrete loci referred to as "speckles" as well as in "foci" (150–156), whose function is unclear. Several splicing factors, such as SC-35 and U2AF, localize similarly in the nucleus (145, 157–160). hnRNP proteins do not appear to be excluded from the speckles, but they are not preferentially concentrated in them. Immunocytochemical studies have shown both hnRNP proteins and snRNPs on nascent transcripts (17, 147, 161–163; E. L. Matunis, M. J. Matunis, and G. Dreyfuss, submitted), but the greatest concentration of snRNPs appears to be in interchromatin granules (which correspond to speckles) as well as in coiled bodies ("foci"; 146, 164). The nature of interchromatin granules is not understood, but pulse label studies with ^{3}H-uridine have failed to detect hnRNA within them (165, 166). Indeed, on nascent transcripts (perichromatin fibrils) where hnRNPs and snRNPs colocalize, pre-mRNA splicing has been observed (167).

In addition to the general nucleoplasmic localization, antibodies to hnRNP L stain intensely two to five discrete non-nucleolar structures in vertebrate cells (42). Similar staining is observed with antibodies to hnRNP K and J on *Xenopus laevis* cells, and the bright loci decorated by antibodies to L and to K and J overlap (73b, M. J. Matunis and G. Dreyfuss, unpublished observations) but do not colocalize with snRNP-enriched "speckles" or "foci" (145). The same antibodies stain the majority of the nascent transcripts on the loops of lampbrush chromosomes in the newt *Notophthalmus viridescens* (42), but the most intense staining is localized to the landmark giant loops. It is likely that the bright spots observed in somatic nuclei correspond to the lampbrush chromosome giant loops, and thus likely represent concentrations of L protein (and therefore specific hnRNP complexes) still associated with specific chromosome loci. Discrete brightly stained non-nucleolar structures, in addition to nucleoplasmic staining, have also been observed with antibodies to hnRNP I/PTB (69). In this case, however, there is only one (occasionally two) spot per nucleus, which is always closely apposed to, but not within, one of the multiple nucleoli of HeLa cells. By analogy to the observations with hnRNP L, these regions with higher concentrations of hnRNP I likely represent sites of transcription and/or processing of specific RNA species (69).

Shuttling of hnRNP Proteins Between the Nucleus and Cytoplasm

The nuclear staining was initially interpreted to indicate that hnRNP proteins are restricted to the nucleus, with the necessary conclusion that the functions of hnRNP proteins concern strictly nuclear processes. However, recent work has shown that this is not always the case, and that some of the hnRNP proteins, such as those in the A and B groups, shuttle between the nucleus

and the cytoplasm (44). This phenomenon, which was most clearly observed by following the migration of these proteins between nuclei in interspecies heterokaryons, is similar to that observed for some nucleolar proteins (168). In contrast, other hnRNP proteins, such as C and U, are confined to the nucleus. Significant amounts of the shuttling hnRNP proteins were not previously observed in the cytoplasm of actively growing interphase cells, probably because their presence in the cytoplasm is transient, and they rapidly reaccumulate in the nucleus. The shuttling of some of the hnRNP proteins has important implications: it suggests that these proteins may also have functions in the cytoplasm, it underscores the dynamic nature of hnRNP complexes, and it merits considering a role for these proteins in nucleocytoplasmic transport of mRNA. These issues are addressed further in the following sections.

Signals and Mechanisms for Localization of hnRNP Proteins

All of the hnRNP proteins must be imported into the nucleus, and some of them, the shuttling proteins, must also be exported to the cytoplasm. What are the signals in the hnRNP proteins that mediate these processes? The amino acid sequences of hnRNP proteins revealed the presence of SV40 T antigen- and nucleoplasmin-type nuclear localization signals in some of these proteins, such as C and U (66, 68, 75). At least for the C proteins, two different clusters of basic amino acids constitute a bona fide nuclear localization signal (H. Siomi, G. Dreyfuss, in preparation). Interestingly, however, other hnRNP proteins, e.g. A1 and A2, have no such recognizable putative nuclear localization signals, and the sequences that mediate their nuclear localization are not yet known.

An interesting role for RNA polymerase II (pol II) transcription in the nuclear localization of hnRNP proteins has recently emerged, initially from studies on mitotic cells (101). In animal cells, the nuclear envelope disassembles as they enter M-phase, and hnRNP complexes disperse throughout the cell (40, 63, 101, 169). The hnRNP proteins remain cytoplasmic and are excluded from the region of condensed chromatin until mitosis is completed and the nuclear envelopes of the daughter cells reform (63, 101, 169, 170). At this stage, the hnRNP complexes disassemble and different hnRNP proteins return to the nucleus separately (101). The separate return to the nucleus appears to reflect the existence of two modes of nuclear localization of hnRNP proteins: a transcription-independent process (e.g. C and U proteins) and a novel, transcription-dependent process (e.g. A and B proteins). Inhibition of RNA polymerase II transcription results in the cytoplasmic accumulation of the latter proteins (101). This dependence on transcription for the nuclear localization of hnRNP proteins operates also in interphase, as the shuttling

hnRNP proteins arrest in the cytoplasm in the presence of pol II transcriptional inhibitors (44).

The signals and mechanisms that mediate the export of the shuttling hnRNP proteins to the cytoplasm are not known. While it is possible that the export of the shuttling hnRNP proteins occurs as a result of their passive "piggy-backing" on the RNA as it is being transported to the cytoplasm, we consider it likely that, in addition to their nuclear import signals, shuttling hnRNP proteins also possess specific nuclear export signals.

VI. THE ARRANGEMENT OF PROTEINS AND RNA IN hnRNP COMPLEXES

hnRNP complexes differ from other RNP complexes, such as snRNPs and ribosomes, in that they contain RNAs with a wide range of lengths, different sequences, and possibly with various fates. The arrangement of the hnRNA and of the hnRNP proteins in hnRNP complexes, and in particular whether there is a sequence-specific arrangement of proteins on hnRNA, is a central question, since this will influence the structure and accessibility of the hnRNA and its interaction with other nuclear components.

The earlier morphological and biochemical studies of hnRNP complexes had led to a model invoking a basic, uniform, and repeating structure of hnRNP proteins (monoparticles or ribonucleosomes) that has a fixed composition, is RNA-sequence independent, and is common to all hnRNAs. This scenario is akin to chromatin, and envisions a more passive, packaging role for hnRNP proteins (9, 171). The more recent observations described below support an alternative, dynamic RNA sequence-dependent model, which suggests a more active and direct role for hnRNP proteins in the regulation of the fate of the hnRNAs. Although this issue has not yet been definitively resolved, of these two possibilities the evidence presently available more strongly supports a unique, sequence-dependent arrangement of hnRNP proteins on each transcript. A combination of the two models is also possible, whereby some of the proteins are positioned in a sequence-specific manner and the rest (that are in excess of the specific binding sites) are organized into some form of a repeating basic particle whose composition is more or less fixed. This possibility has been expanded on recently (172).

Studies of hnRNP Complexes Isolated from Nuclei

Electron microscopic observations of actively transcribing chromatin had demonstrated the association of proteins with nascent transcripts, and some of these proteins appeared as an array of particles or "beads" connected by the RNase-sensitive hnRNA (173–177). A more recent re-evaluation of these morphological observations, however, argues that the particles observed in

chromatin spreads are unlikely to be hnRNP "monoparticles" and that the arrangement of hnRNP proteins on the nascent transcripts takes on more the form of an hnRNA coated with hnRNP proteins throughout its length, or an RNP fibril (178). The main difficulties in interpreting the electron microscope observations are that the conditions required for the preparation of the samples disrupt the native organization of the complexes, such that severe loss and rearrangement of their components can occur, and that, most importantly, the specific composition of the particles observed was not known.

The chromatin-associated hnRNA-hnRNP-snRNP complexes are insoluble and they are also difficult to analyze biochemically as their composition is extremely complex. The question of the structure of these assemblies thus remains unsolved. Biochemical analyses have therefore focused either on particles released from nuclei by digestion with RNases (endogenous or exogenous) or on nucleoplasmic hnRNP complexes from disrupted nuclei. With little or no RNase digestion, the overall population of soluble hnRNP complexes sediment in sucrose gradients heterodispersely at 60 to >200S (20–27). Partial degradation of the RNA converts these complexes to more monodisperse particles sedimenting at about 30–40S, which consist of many of the major hnRNP proteins and hnRNA fragments of 500–800 nucleotides. Interestingly, recent studies have shown that complexes containing specific pre-mRNAs can be recovered as discrete particles sedimenting at 200S, regardless of the length of the pre-mRNA (179, 180). But the composition of the large and heterodisperse hnRNP complexes and the 30–40S particles varies in different preparations, and therefore it persisted as a matter of controversy. Nevertheless, the sedimentation data were taken together with the "beads on a string" appearance of nascent transcripts in spread chromatin preparations, to suggest a model where the 30–40S particles correspond to the observed "beads," with the hnRNA being visualized as the "string." The RNA in the 30–40S particles can be completely digested with nuclease, and it was thus inferred that it was exposed on the particles. It should be noted that while 30–40S particles have been observed repeatedly in vitro, there is no evidence for their existence in vivo. A comprehensive discussion of the possible artifactual nature of a regular array of (30–40S) monoparticles as the packaging element of hnRNAs was recently provided by Beyer & Osheim (178).

hnRNP complexes isolated from nucleoplasm by rapid immunopurification with monoclonal antibodies contain hnRNAs of heterogeneous lengths that range from a few hundred nucleotides to greater than 10,000 nucleotides, and the protein composition shown in Figure 1 (19, 36). Limited RNase digestion results in the selective loss of hnRNP I, P, and S proteins, suggesting that they are positioned on particularly exposed sites in the hnRNP complexes (69). Further digestion of the hnRNA leads to dissociation of the remaining

proteins. Some of the proteins have been shown to have the propensity to form oligomeric complexes without RNA, for example A2 and B1, and C1 and C2 (181, 182). In addition, the A1 protein binds single-stranded polynucleotides cooperatively (107, 115, 116). This reflects the capacity of hnRNP proteins to interact with each other, and may be relevant to the way they interact in the native complexes. Thus, protein-protein interactions are probably an important force in the overall structure of the complexes, but for most of the hnRNP proteins they are not sufficient to hold them together in the absence of RNA.

In vitro Assembly of hnRNP Complexes

Attempts at reconstitution of hnRNP complexes in vitro have been limited by the lack of information on native complexes. Nevertheless, some information of parameters that may be important for the organization of hnRNP complexes has been gained from in vitro assembly studies. Particles of similar general morphology and sedimentation properties to 30–40S particles can be formed in vitro, using sucrose gradient-enriched fractions of hnRNP proteins and a variety of RNAs (102, 104, 105). Essentially any ssRNA or ssDNA of greater than 700 nucleotides, regardless of its sequence, resulted in the assembly of particles with similar sedimentation properties. These results were combined with the observations on hnRNP complexes isolated from nuclei described above and interpreted to demonstrate that hnRNP complexes consist of a repeating array of regular particles of fixed stoichiometry of hnRNP proteins that package pre-mRNA in a length-dependent (ca. 700 nucleotides) but sequence-independent manner (105). While these observations may reflect some important characteristics of the hnRNP proteins that were included in the assembly experiments, their significance to the structure of hnRNP complexes in vivo is not clear.

Other approaches have also been used to investigate the assembly of proteins on defined RNAs in vitro, which led to very different conclusions about the arrangement of hnRNP proteins on hnRNA. Mapping of the binding of hnRNP proteins on specific pre-mRNAs in nuclear extracts (111) demonstrated that the binding of hnRNP A1, C, and D proteins is not random with respect to RNA sequence. The composition of complexes assembled in nuclear extracts on RNAs of defined sequence was also determined by selecting specific biotinylated RNAs with immobilized streptavidin (183). This demonstrated that most of the proteins bound to these RNAs in nuclear extracts are known hnRNP proteins. More importantly, it showed that each different RNA associated with a unique combination of hnRNP proteins. These differences, both quantitative and qualitative, are even more pronounced with shorter test RNAs, as they can accommodate fewer hnRNP proteins (183). Thus, under conditions of competition for binding sites, which more closely resembles the

situation in the nucleus, the array of hnRNP proteins bound to a given hnRNA is determined by the sequence of the RNA.

Differential Association of hnRNP Proteins with Nascent hnRNAs

Direct evidence for differential association of hnRNP proteins with hnRNAs in vivo has been obtained from immunocytochemical studies, which allow the detection of specific hnRNP proteins, as well as snRNPs, as they associate with nascent transcripts on lampbrush chromosomes of amphibian oocytes. By immunofluorescence microscopy, antibodies to both hnRNP and snRNP components stain the majority of the actively transcribing loops (17, 42, 169, 184). There are, however, notable exceptions such as the landmark giant loops and sequentially labeled loops, to which so far only the hnRNP K and L proteins have been observed to bind [see (42) and references therein]. Recently, simultaneous detection of several different hnRNP proteins on *D. melanogaster* polytene chromosomes showed that although most loci contain all of the abundant hnRNP proteins, the relative amounts of particular hnRNP proteins vary on different loci (184b). A comparison of the distribution of hnRNP proteins with that of snRNPs also showed that both hnRNPs and snRNPs colocalize on most loci, but the relative amounts of these components also vary considerably from transcript to transcript. These results provide direct evidence that the relative amounts of different individual hnRNP proteins on nascent transcripts are not fixed, and that hnRNP proteins, as well as snRNPs, associate differentially (and probably independently) with nascent hnRNAs.

hnRNP Complexes are Dynamic Structures

Recent studies on mitotic cells further argue against a fixed particle composition for hnRNP proteins in living cells. hnRNP proteins remain associated in hnRNP complexes in mitosis (63, 101, 170), but the hnRNP complexes disassemble at the end of mitosis and the proteins return to the nucleus separately at different times (101). Once in the nucleus, the pre-existing hnRNP proteins are presumably reincorporated into hnRNP complexes. Thus hnRNP proteins can exist separately in the cell and therefore the composition of the complexes is not fixed, providing an important illustration of the dynamic character of hnRNP complexes.

hnRNP complexes also undergo dramatic rearrangements during transport of mRNA from the nucleus to the cytoplasm. The recent finding that some of the hnRNP proteins shuttle between the nucleus and the cytoplasm, whereas others are restricted to the nucleus, indicates that hnRNP proteins are not part of a fixed structure and that complex rearrangements of hnRNP components occur upon mRNA transport (44). This is also illustrated by morphological

studies of the transport of Balbiani ring RNPs in *Chironomus tentans* as these RNPs undergo a remarkable morphological change during transport (185).

The General Structure of hnRNP Complexes

The evidence discussed in the preceding sections indicates that the hnRNP complexes on different hnRNAs have a different composition and unique arrangement of hnRNP proteins, and that these can change with the processing of the hnRNA. As hnRNP proteins can bind independently and differentially to RNAs, in the nucleus individual hnRNP proteins will occupy first those binding sites for which they have higher affinity. However, as the major hnRNP proteins are very abundant, each protein is likely to be in vast excess over its respective higher-affinity binding sites. Cooperative interactions, in concert with the lower-affinity, sequence-nonspecific binding of the hnRNP proteins, can allow binding to additional sites on the RNA. We suggest that this would result in extensive contiguous binding of hnRNP proteins to form an RNP fibril in which most, if not all, of the hnRNA is bound with hnRNP proteins (Figure 3). It is also possible that, if their density on the hnRNA is sufficiently high, and due to likely cooperative interactions among them, they coalesce into units (or particles) of interacting proteins. Jacob and coworkers in fact proposed a model in which hnRNP complexes are organized as RNP fibrils, interspersed by particles that exhibit differential salt sensitivity (186). Importantly, many of the hnRNP proteins contain RNP motifs, and the recent determination of the overall structure of the RBD-RNA complex, shows that the RNA is exposed as it is bound by such RBDs. This indicates that the RNA occupies an exposed position on the hnRNP complex, and RNase digestion experiments indeed support the conclusion that most of the hnRNA is exposed in the complexes.

The specific structures of individual hnRNP complexes are unknown, but the important parameters that determine them are: the sequence of the hnRNA, the binding preferences and characteristics of the hnRNP proteins, cooperative interactions between the proteins, competition between the proteins for binding sites, the relative amounts of the hnRNP proteins, and the amount of the hnRNA.

VII. THE FUNCTIONS OF hnRNP PROTEINS

Functions in Pre-mRNA Processing

It has been long expected that hnRNP proteins, being very abundant and avid RNA-binding proteins, play important roles in the metabolism of hnRNAs. One of the earliest observations that hnRNP proteins can affect the fate of pre-mRNAs came from immunoinhibition experiments in which antibodies to individual hnRNP proteins were added to in vitro splicing reactions. In these

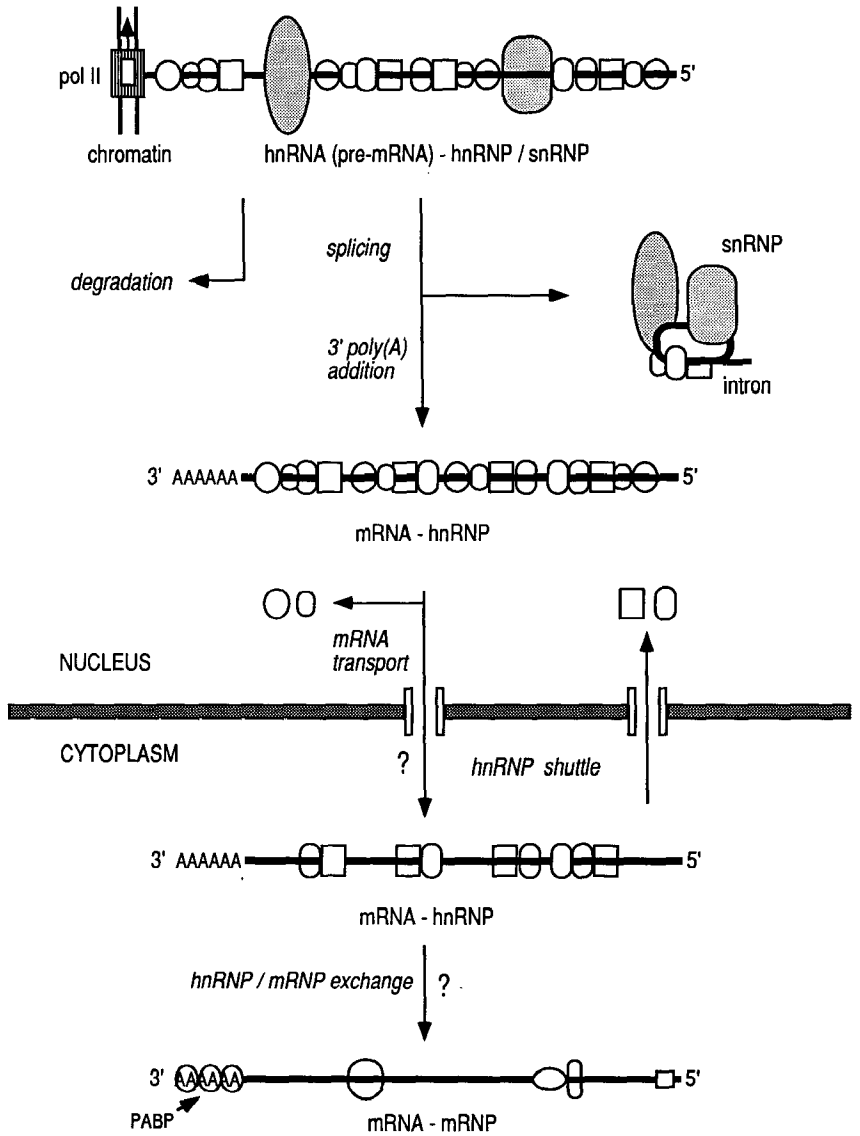

Figure 3 Schematic presentation of a generalized model for the arrangement of hnRNA-hnRNP-snRNP complexes and the pathway of mRNA biogenesis. These complexes assemble on nascent RNA polymerase II transcripts, and the model predicts a unique association of hnRNP proteins with each RNA transcript. hnRNP proteins remain associated with pre-mRNAs during and following their processing into mRNAs. A subset of the hnRNP proteins may accompany the mRNA during its transit through the nuclear pore to the cytoplasm. Once in the cytoplasm, the shuttling hnRNP proteins are exchanged for mRNP proteins. snRNPs, small nuclear ribonucleoprotein particles; PABP, poly(A)-binding protein.

studies antibodies to the hnRNP C proteins (mAb 4F4) inhibited the first cleavage event at the 5′ splice site, and immunodepleted nuclear extracts did not form spliceosomes (187). Similar experiments using an antiserum against the A, B, and C group proteins also resulted in inhibition of splicing (188). The 57-kDa hnRNP I/PTB protein has also been implicated in pre-mRNA splicing. It is UV-crosslinked specifically to the polypyrimidine tract of the 3′ splice site region (71, 72), and mutations within the polypyrimidine tract that reduce the efficiency of spliceosome formation also reduce or abolish the crosslinking of this protein (71). The specific role of hnRNP I/PTB is not known, but in the case of the splicing of a multiple intron β-tropomyosin pre-mRNA, binding of hnRNP I/PTB to the downstream polypyrimidine tract directly correlates with the selection of this 3′ splice site over the upstream splice site (189). It has not been demonstrated, however, that hnRNP I/PTB is an essential splicing factor. There is some evidence that a 100-kDa protein that copurifies with hnRNP I/PTB is essential for pre-mRNA splicing, and addition of recombinant hnRNP I/PTB alone to extracts depleted of both proteins does not restore splicing activity (72).

Recently, the hnRNP A1 protein has been shown to affect the regulation of 5′ splice site choice (190–192). In pre-mRNAs that contain multiple 5′ splice sites, A1 effects a switch in splicing from a proximal 5′ splice site to an upstream splice site. Interestingly, it is not simply the presence of A1 that causes distal splicing but rather the amount of A1 relative to another splicing factor, ASF/SF2, that activates the use of proximal 5′ splice sites (190). Therefore, the activities of A1 and ASF/SF2 may directly compete to determine 5′ splice site selection. High amounts of A1 generally favor the choice of distal 5′ splice sites, while high amounts of ASF/SF2 favor proximal splice sites (190). In addition, supplementing nuclear extracts with purified A1 may increase the general efficiency of pre-mRNA splicing (D. Stolow, S. Berget, personal communication).

How can hnRNP proteins affect RNA processing? Most hnRNP proteins contain RNP motifs and structural analyses of RNA-RBD complexes have shown that, when bound, the RNA is displayed (121, 128). We suggest that a general function of hnRNA-binding proteins is to make the hnRNA extended and accessible as it would otherwise form an inaccessible, inefficient RNA-processing substrate. In this respect, the function of hnRNP proteins would be akin to that of *Escherichia coli* SSB (193) and to the bacteriophage T4 gene 32 product (193). However, the activity of hnRNP proteins must be more sophisticated, otherwise this function could be fulfilled by a single hnRNP protein. Through cooperative interactions the abundant hnRNP proteins may bind most of the hnRNA so that almost all of it becomes exposed. This function may also explain the abundance of hnRNP proteins as much of each hnRNA must be bound by them. It may be difficult to observe this

substrate presentation function of hnRNP proteins in vitro as it may not be readily manifested in the commonly used in vitro processing reactions where very small pre-mRNA substrates are assayed over relatively long incubation periods. The in vivo situation is certainly quite different; as most of the hnRNAs are of enormous size and as processing reactions need to proceed in the cell with great fidelity at high rates, the function of hnRNP proteins is almost certainly essential. Furthermore, there is likely to be substantial functional redundancy in the requirement for this function so that specifically depleting an individual hnRNP protein from extracts (with certainty that no other components have been removed) will not necessarily cause an observable effect. Genetic approaches to reveal the function of hnRNP proteins have not yet been reported, although these will clearly be very informative.

Recent biochemical experiments illustrate how such proteins can affect the hnRNA as they demonstrated that hnRNP A1 has annealing-promoting activity towards complementary strands of RNA or DNA (194–197). Thus, hnRNP proteins may affect the structure of hnRNA and its interactions with other factors (e.g. the binding of the U1 snRNP to the pre-mRNA). As many of the hnRNP proteins have multiple RBDs, each of which can potentially act independently, they may be able to bring together different RNA sequences that would otherwise be far apart. Such activities may be important in many reactions such as annealing-promoting and trans-splicing.

Although RNP motif-containing proteins can bind RNA independently, cooperative and competitive protein-protein interactions are likely to influence the local constellation of proteins bound to a given RNA, thus facilitating or hindering the formation of specialized complexes on hnRNA. The specific arrangement of proteins on RNA, therefore, will be determined by inherent differences in binding specificities and by specific protein-protein interactions. These interactions probably result from the exclusion or recruitment of other proteins by protein-protein interactions with these hnRNA-binding proteins. Since there are several hnRNA-binding proteins that have overlapping binding specificities (e.g. hnRNP C, hnRNP I/PTB, and U2AF), there is the possibility of sequential binding to RNA, each protein with a different function.

In addition to the general role of hnRNP proteins in substrate presentation, they may also have specific functions in the regulation of pre-mRNA splicing. The functional relatedness of hnRNP proteins and splicing factors is underscored by the observation that alternative splicing can be regulated by the interplay between an abundant hnRNP protein, A1, and another RNP motif protein, ASF/SF2 (190). Information about the structure and RNA-binding properties of hnRNA-binding proteins has further blurred the distinction between "hnRNP proteins" and "RNA processing factors." With the emerging characterization of splicing factors, it has become apparent that many, if not most, of these proteins contain many of the hallmarks of hnRNP proteins.

For example, two spliceosomal proteins, SC-35 and U2AF, contain RNP motifs and auxiliary domains rich in Arg-Ser (RS) dipeptides (157, 198, 199). Two RNA-binding proteins that participate in polyadenylation, poly(A) polymerase (200–202) and CPSF (203), are also RNP motif-containing proteins. In *D. melanogaster* two loci, *transformer-2* and *sex-lethal,* encode sequence-specific RNA-binding proteins that act as negative regulators of non-sex-specific splice sites (80, 81, 86, 88, 204). As it is becoming apparent that hnRNP proteins have functions in RNA processing, a categorical division between hnRNP proteins and RNA processing factors may not be meaningful as they are a single family of hnRNA-binding proteins.

Other Functions of hnRNP Proteins

The recent finding that some of the hnRNP proteins shuttle between the nucleus and the cytoplasm (44) raises the likely possibility that these proteins also have functions in the cytoplasm. As the shuttling hnRNP proteins are found bound to mRNA in the cytoplasm (44), they could participate in a variety of cytoplasmic aspects of mRNA metabolism, including regulation of mRNA translation and stability and mRNA localization. The shuttling phenomenon also raises the possibility that these proteins could serve as carriers of (m)RNAs to the cytoplasm. Electron microscopic observations of Balbiani ring mRNAs in *Chironomus tentans* have shown that these are transported to the cytoplasm as a ribonucleoprotein particle (185, 205). However, the biochemical properties of this transport RNP are not known. It is possible that the export of hnRNP proteins from the nucleus to the cytoplasm occurs independently from mRNA export, but it is more likely that the shuttling hnRNP proteins exit the nucleus bound to mRNA during its nucleocytoplasmic transport and are components of the transported RNP.

It has been shown recently that nucleus-restricted sequences such as introns can serve as nuclear retention signals for incompletely processed pre-mRNAs (206, 207). hnRNP proteins, such as hnRNP C, which are nucleus-restricted, may participate in this retention of incompletely processed pre-mRNAs. In this regard, it is interesting to note that high-affinity binding sites for the C proteins have been mapped to such nucleus-restricted sequences, namely the 3′ end of introns (111) and sequences downstream of 3′ end cleavage and polyadenylation sites (109, 110). Finally, as many of the hnRNP proteins bind ssDNA, they may also have roles in transcription, DNA replication, and recombination.

ACKNOWLEDGMENTS

We are grateful to members of our laboratory for comments on the manuscript. Work in our laboratory was supported by the Howard Hughes Medical Institute and by grants from the National Institutes of Health.

Literature Cited

1. Darnell, J. E. 1982. *Nature* 297:365–71
2. Nevins, J. R. 1983. *Annu. Rev. Biochem.* 52:441–66
3. Steitz, J. A., Black, D. L., Gerke, V., Parker, K. A., Dramer, A., et al. 1988. In *Structure and Function of Major and Minor Small Nuclear Ribonucleoproteins,* ed. M. L. Birnstiel, pp. 115–54. Berlin/New York: Springer Verlag
4. Lührmann, R., Kastner, B., Bach, M. 1990. *Biochim. Biophys. Acta* 1087: 265–92
5. Zieve, G. W., Sauterer, R. A. 1990. *Crit. Rev. Biochem. Mol. Biol.* 25:1–46
6. Dreyfuss, G., Swanson, M. S., Piñol-Roma, S. 1988. *Trends Biochem. Sci.* 13:86–91
7. Bandziulis, R. J., Swanson, M. S., Dreyfuss, G. 1989. *Genes Dev.* 3:431–37
8. Dreyfuss, G. 1986. *Annu. Rev. Cell Biol.* 2:459–98
9. Chung, S. Y., Wooley, J. 1986. *Proteins: Struct. Funct. Genet.* 1:195–210
10. Knowler, J. T. 1983. *Int. Rev. Cytol.* 84:103–53
11. Martin, T. E., Pullman, J. M., McMullen, M. E. 1980. In *Cell Biology: A Comprehensive Treatise,* ed. D. M. Prescott, L. Goldstein, 4:137–74. New York: Academic
12. Pederson, T. 1983. *J. Cell Biol.* 97: 1321–26
13. Samarina, O. P., Krichevskaya, A. A. 1981. In *The Cell Nucleus,* ed. H. Busch, 9:1–48. New York: Academic
14. Gall, J. G., Callan, H. G. 1962. *Proc. Natl. Acad. Sci. USA* 48:562–70
15. Malcolm, D. B., Sommerville, J. 1974. *Chromosoma* 48:137–58
16. Miller, O. L., Bakken, A. H. 1972. *Karolinska Symp. Res. Methods Reprod. Endocrinol.* 5:155–67
17. Wu, Z. G., Murphy, C., Callan, H. G., Gall, J. G. 1991. *J. Cell Biol.* 113:465–83
18. Dreyfuss, G., Philipson, L., Mattaj, I. 1988. *J. Cell. Biol.* 106:1419–25
19. Piñol-Roma, S., Choi, Y. D., Matunis, M. J., Dreyfuss, G. 1988. *Genes Dev.* 2:215–27
20. Samarina, O. P., Lukanidin, E. M., Molman, J., Georgiev, G. P. 1968. *J. Mol. Biol.* 33:251–63
21. Martin, T. E., Billings, P. B., Levey, A., Ozarsian, S., Quinlan, I., et al. 1974. *Cold Spring Harbor Symp. Quant. Biol.* 38:921–32
22. Martin, T. E., Billings, P. B., Pullman, J. M., Stevens, B. J., Kinniburgh, A. J. 1978. *Cold Spring Harbor Symp. Quant. Biol.* 42:899–909
23. Pederson, T. 1974. *J. Mol. Biol.* 83: 163–83
24. Billings, P. B., Martin, T. E. 1978. *Methods Cell. Biol.* 17:349–76
25. Beyer, A. L., Christensen, M. E., Walker, B. W., LeStourgeon, W. M. 1977. *Cell* 11:127–38
26. Karn, J., Vidali, G., Boffa, L. C., Allfrey, V. G. 1977. *J. Biol. Chem.* 252:7307–22
27. Wilk, H. E., Werr, H., Friedrich, D., Kiltz, H. H., Schäfer, K. P. 1985. *Eur. J. Biochem.* 46:71–81
28. Mayrand, S., Setyono, B., Greenberg, J. R., Pederson, T. 1981. *J. Cell Biol.* 90:380–84
29. Mayrand, S., Pederson, T. 1981. *Proc. Natl. Acad. Sci. USA* 78:2208–12
30. Van Eekelen, C. A., Riemen, T., van Venrooij, W. J. 1981. *FEBS Lett.* 130:223–26
31. Van Eekelen, C. A. G., Mariman, E. C. M., Reinders, R. J., van Venrooij, W. J. 1981. *Eur. J. Biochem.* 119:461–67
32. Dreyfuss, G., Choi, Y. D., Adam, S. A. 1984. *Mol. Cell. Biol.* 4:1104–14
33. Dreyfuss, G., Adam, S. A., Choi, Y. D. 1984. *Mol. Cell. Biol.* 4:415–23
34. Bag, J. 1984. *Eur. J. Biochem.* 141: 247–54
35. Greenberg, J. R., Carroll, E. III. 1985. *Mol. Cell. Biol.* 5:342–51
36. Choi, Y. D., Dreyfuss, G. 1984. *Proc. Natl. Acad. Sci. USA* 81:7471–75
37. Swanson, M. S., Dreyfuss, G. 1988. *Mol. Cell. Biol.* 8:2237–41
38. Green, M. R. 1991. *Annu. Rev. Cell Biol.* 7:559–99
39. Wahle, E., Keller, W. 1992. *Annu. Rev. Biochem.* 61:419–40
40. Choi, Y. D., Dreyfuss, G. 1984. *J. Cell Biol.* 99:1997–2004
41. Matunis, M. J., Michael, W. M., Dreyfuss, G. 1992. *Mol. Cell. Biol.* 12:164–71
42. Piñol-Roma, S., Swanson, M. S., Gall, J. G., Dreyfuss, G. 1989. *J. Cell Biol.* 109:2575–87
43. Leser, G. P., Escara-Wilke, J., Martin, T. E. 1984. *J. Biol. Chem.* 259:1827–33
44. Piñol-Roma, S., Dreyfuss, G. 1992. *Nature* 355:730–32
45. Burd, C. G., Swanson, M. S., Görlach, M., Dreyfuss, G. 1989. *Proc. Natl. Acad. Sci. USA* 86:9788–92

46. Cobianchi, F., SenGupta, D. N., Zmudzka, B. Z., Wilson, S. H. 1986. *J. Biol. Chem.* 261:3536–43
47. Buvoli, M., Biamonti, G., Tsoulfas, P., Bassi, M. T., Ghetti, A., et al. 1988. *Nucleic Acids Res.* 16:3751–70
48. Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., et al. 1990. *EMBO J.* 9:1229–35
49. Biamonti, G., Buvoli, M., Bassi, M. T., Morandi, C., Cobianchi, F., et al. 1989. *J. Mol. Biol.* 207:491–503
50. Kumar, A., Williams, K. R., Szer, W. 1986. *J. Biol. Chem.* 261:11266–73
51. Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., et al. 1986. *EMBO J.* 5:2267–74
52. Kay, B. K., Sawhney, R. K., Wilson, S. H. 1990. *Proc. Natl. Acad. Sci. USA* 87:1367–71
53. Cobianchi, S. 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
54. Matunis, M. J., Matunis, E. L., Dreyfuss, G. 1992. *J. Cell Biol.* 116:245–55
55. Astaldi-Ricotti, G. C. B., Bestagno, M., Cerino, A., Negri, C., Caporali, R., et al. 1989. *J. Cell. Biochem.* 40:43–47
56. Montecucco, C., Caporali, R., Negri, C., deGennaro, F., Cerino, A., et al. 1990. *Arthritis Rheum.* 33:180–86
57. Steiner, G., Hartmuth, K., Skriner, K., Maurer-Fogy, I., Sinski, A., et al. 1992. *J. Clin. Invest.* 90:1061–66
58. Haynes, S. R., Johnson, D., Raychaudhuri, G., Beyer, A. L. 1990. *Nucleic Acids Res.* 19:25–31
59. Haynes, S. R., Raychaudhuri, G., Beyer, A. L. 1990. *Mol. Cell. Biol.* 10:316–23
60. Matunis, E. L., Matunis, M. J., Dreyfuss, G. 1992. *J. Cell Biol.* 116:257–69
61. Ball, E. E., Rehm, E. J., Goodman, C. S. 1991. *Nucleic Acids Res.* 19:397
62. Celis, J. E., Bravo, R., Arenstorf, H. P., LeStourgeon, W. M. 1986. *FEBS Lett.* 194:101–9
63. Leser, G. P., Martin, T. E. 1987. *J. Cell Biol.* 105:2083–94
64. Minoo, P., Sullivan, W., Solomon, L. R., Martin, T. E., Toft, D. O. 1989. *J. Cell Biol.* 109:1937–46
65. Nakagawa, T., Swanson, M. S., Wold, B., Dreyfuss, G. 1986. *Proc. Natl. Acad. Sci. USA* 83:2007–11
66. Swanson, M. S., Nakagawa, T. Y., LeVan, K., Dreyfuss, G. 1987. *Mol. Cell. Biol.* 7:1731–39
67. Merrill, B. M., Barnett, S. F., LeStourgeon, W. M., Williams, K. R. 1989. *Nucleic Acids Res.* 17:8441–49
68. Preugschat, F., Wold, B. 1988. *Proc. Natl. Acad. Sci. USA* 85:9669–73
69. Ghetti, A., Piñol-Roma, S., Michael, W. M., Morandi, C., Dreyfuss, G. 1992. *Nucleic Acids Res.* 20:3671–78
70. Brunel, F., Alzari, P. M., Ferrara, P., Zakin, M. M. 1991. *Nucleic Acids Res.* 19:5237–45
71. Gil, A., Sharp, P. A., Jamison, S. F., García-Blanco, M. 1991. *Genes Dev.* 5:1224–36
72. Patton, J. G., Mayer, S. A., Tempst, P., Nadal-Ginard, B. 1991. *Genes Dev.* 5:1237–51
73a. Christensen, M. E., Fuxa, K. P. 1988. *Biochem. Biophys. Res. Commun.* 155: 1278–83
73b. Siomi, H., Matunis, M. J., Michael, W. M., Dreyfuss, G. 1993. *Nucleic Acids Res.* In press
74. Kenan, D. J., Query, C. C., Keene, J. D. 1991. *Trends Biochem. Sci.* 16:214–20
75. Kiledjian, M., Dreyfuss, G. 1992. *EMBO J.* 11:2655–64
76. Patzelt, E., Blaas, D., Kuechler, E. 1983. *Nucleic Acids Res.* 11:5821–35
77. Rozen, F., Sonenberg, N. 1987. *Nucleic Acids Res.* 15:6489–500
78. Ohno, M., Kataoka, N., Shimura, Y. 1990. *Nucleic Acids Res.* 18:6989–95
79. Brunel, C., Lelay, M.-N. 1979. *Eur. J. Biochem.* 99:273–83
80. Bell, L. R., Maine, E. M., Schedl, P., Cline, T. W. 1988. *Cell* 55:1037–46
81. Amrein, H., Gorman, M., Nöthiger, R. 1988. *Cell* 55:1025–35
82. Robinow, S., Campos, A. R., Yao, K-M., White, K. 1988. *Science* 242: 1570–72
83. Voelker, R. A., Gibson, W., Graves, J. P., Sterling, J. F., Eisenberg, M. T. 1991. *Mol. Cell. Biol.* 11:894–905
84. Roth, M. B., Zahler, A. M., Stolk, J. A. 1991. *J. Cell. Biol.* 115:587–96
85. Von Besser, H., Schnabel, P., Wieland, C., Fritz, E., Stanewsky, R., et al. 1990. *Chromasoma* 100:37–47
86. Inoue, K., Hoshijima, K., Sakamoto, H., Shimura, Y. 1990. *Nature* 344: 461–63
87. Bell, L. R., Horabin, J. I., Schedl, P., Cline, T. W. 1991. *Cell* 65:229–39
88. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H., Shimura, Y. 1991. *Science* 252:833–36
89. McKeown, M. 1992. *Curr. Biol.* 2: 299–303
90. Baker, B. S. 1989. *Nature* 340:521–24
91. Cruz-Alvarez, M., Pellicer, A. 1987. *J. Biol. Chem.* 262:13377–80

92. Cruz-Alvarez, M., Szer, W., Pellicer, A. 1985. *Nucleic Acids Res.* 13:3917–30
93. Schuster, G., Gruissem, W. 1991. *EMBO J.* 10:1493–502
94. Cook, W. B., Walker, J. C. 1992. *Nucleic Acids Res.* 20:359–64
95. Li, Y. Q., Sugiura, M. 1990. *EMBO J.* 9:3059–66
96. Lee, W.-C., Xue, Z., Mélese, T. 1991. *J. Cell. Biol.* 113:1–12
97. Oliver, S. G., van der Aart, Q. J. M., Agostoni-Carbone, M. L., Aigle, M., Alberghina, L., et al. 1992. *Nature* 357:38–46
98. Holcomb, E. R., Friedman, D. L. 1984. *J. Biol. Chem.* 259:31–40
99. Williams, K. R., Stone, K. L., LoPresti, M. B., Merrill, B. M., Planck, S. R. 1985. *Proc. Natl. Acad. Sci. USA* 82:5666–70
100. Merrill, B. M., LoPresti, M. B., Stone, K. L., Williams, K. R. 1986. *J. Biol. Chem.* 261:878–83
101. Piñol-Roma, S., Dreyfuss, G. 1991. *Science* 253:312–14
102. Pullman, J. M., Martin, T. E. 1983. *J. Cell Biol.* 97:99–111
103. Thomas, J. O., Glowacka, S. K., Szer, W. 1983. *J. Mol. Biol.* 171:439–55
104. Wilk, H. E., Angeli, G., Schaefer, K. P. 1983. *Biochemistry* 22:4592–600
105. Conway, G., Wooley, J., Bibring, T., LeStourgeon, W. M. 1988. *Mol. Cell. Biol.* 8:2884–95
106. Schenkel, J., Sekeris, C. E., Alonso, A., Bautz, E. K. F. 1988. *Eur. J. Biochem.* 171:565–69
107. Cobianchi, F., Karpel, R. L., Williams, K. R., Notario, V., Wilson, S. H. 1988. *J. Biol. Chem.* 263:1063–71
108. Pandolfo, M., Valentini, O., Biamonti, G., Rossi, P., Riva, S. 1987. *Eur. J. Biochem.* 162:213–20
109. Wilusz, J., Shenk, T. 1990. *Mol. Cell. Biol.* 10:6397–407
110. Moore, C. L., Chen, J., Whoriskey, J. 1988. *EMBO J.* 7:3159–69
111. Swanson, M. S., Dreyfuss, G. 1988. *EMBO J.* 7:3519–29
112. Buvoli, M., Cobianchi, F., Biamonti, G., Riva, S. 1990. *Nucleic Acids Res.* 18:6595–600
113. Tuerk, C., Gold, L. 1990. *Science* 249:505–10
114. Szostak, J. W. 1992. *Trends Biochem. Sci.* 17:89–93
115. Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., et al. 1991. *Biochemistry* 30:2968–75
116. Kumar, A., Wilson, S. H. 1990. *Biochemistry* 29:10717–22
117. Query, C. C., Bentley, R. C., Keene, J. D. 1989. *Cell* 57:89–101
118. Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., Mattaj, I. W. 1989. *EMBO J.* 8:4163–70
119. Adam, S. A., Nakagawa, T. Y., Swanson, M. S., Woodruff, T., Dreyfuss, G. 1986. *Mol. Cell. Biol.* 6:2932–43
120. Lutz-Freyermuth, C., Query, C. C., Keene, J. D. 1990. *Proc. Natl. Acad. Sci. USA* 87:6393–97
121. Görlach, M., Wittekind, M., Beckman, R. A., Mueller, L., Dreyfuss, G. 1992. *EMBO J.* 11:3289–95
122. Schwemmle, M., Görlach, M., Bader, M., Sarre, T. F., Hilse, K. 1989. *FEBS Lett.* 251:117–20
123. Merrill, B. M., Stone, K. L., Cobianchi, F., Wilson, S. H., Williams, K. R. 1988. *J. Biol. Chem.* 263:3307–13
124. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., Evans, P. R. 1990. *Nature* 346:515–20
125. Hoffman, D. W., Query, C. C., Golden, B. L., White, S. W., Keene, J. D. 1991. *Proc. Natl. Acad. Sci. USA* 88:2495–99
126. Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G., Mueller, L. 1992. *Biochemistry* 31:6254–65
127. Nagai, K. 1992. *Curr. Opin. Struct. Biol.* 2:131–37
128. Jessen, T. H., Oubridge, C., Teo, C. H., Pritchard, C., Nagai, K. 1991. *EMBO J.* 10:3447–56
129. Lutz-Freyermuth, C., Keene, J. D. 1989. *Mol. Cell. Biol.* 9:2975–82
130. Patton, J. R., Habets, W., van Venrooij, W. J., Pederson, T. 1989. *Mol. Cell. Biol.* 9:3360–68
131. Surowy, C. S., van Santen, V. L., Scheib-Wixted, S. M., Spritz, R. A. 1989. *Mol. Cell. Biol.* 9:4179–86
132. Burd, C. G., Matunis, E. L., Dreyfuss, G. 1991. *Mol. Cell. Biol.* 7:3419–24
133. Karpel, R. L., Burchard, A. C. 1980. *Biochemistry* 19:4674–82
134. Casas-Finet, J. R., Karpel, R. L., Maki, A. H., Kumar, A., Wilson, S. H. 1991. *J. Mol. Biol.* 221:693–709
135. Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., Williamson, J. R. 1992. *Science* 257:76–80
136. Tao, J., Frankel, A. D. 1992. *Proc. Natl. Acad. Sci. USA* 89:2723–26
137. Calnan, B. J., Biancalana, S., Hudson, D., Frankel, A. D. 1991. *Genes Dev.* 5:201–10
138. Ghisolfi, L., Joseph, G., Amalric, F., Erard, M. 1992. *J. Biol. Chem.* 267: 2955–59

139. Christensen, J. H., Hansen, P. K., Lillelund, O., Thogersen, H. C. 1991. *FEBS Lett.* 281:181–84
140. Liao, X., Clemens, K. R., Tennant, L., Wright, P. E., Gottesfeld, J. M. 1992. *J. Mol. Biol.* 223:857–71
141. Weeks, K. M., Ampe, C., Schultz, S. C., Steitz, T. A., Crothers, D. M. 1990. *Science* 249:1281–85
142. Cordingley, M. G., LaFemina, R. L., Callahan, P. L., Condra, J. H., Sardana, V. V., et al. 1990. *Proc. Natl. Acad. Sci. USA* 87:8985–89
143. Römisch, K., Webb, J., Lingelbach, K., Gausepohl, H., Dobberstein, B. 1990. *J. Cell Biol.* 111:1793–802
144. Mitchell, P. J., Tjian, R. 1989. *Science* 245:371–78
145. Carmo-Fonseca, M., Pepperkok, R., Sproat, B. S., Ansorge, W., Swanson, M. S., Lamond, A. I. 1991. *EMBO J.* 10:1863–73
146. Fakan, S., Leser, G., Martin, T. E. 1984. *J. Cell Biol.* 98:358–63
147. Fakan, S., Leser, G., Martin, T. E. 1986. *J. Cell Biol.* 103:1153–57
148. Fakan, S., Puvion, E., Spohr, G. 1976. *Exp. Cell Res.* 99:155–64
149. Fakan, S. 1978. See Ref. 13, 5:3–53
150. Spector, D. L., Schrier, W. H., Busch, H. 1983. *Biol. Cell.* 49:1–10
151. Reuter, R., Appel, B., Bringmann, P., Rinke, J., Luhrmann, R. E. 1984. *Exp. Cell Res.* 154:548–60
152. Verheijen, R., Kuijpers, H., Vooijs, P., van Venrooij, W., Ramaekers, F. 1986. *J. Cell. Sci.* 86:173–90
153. Nyman, U., Hallman, H., Hadlaczky, G., Pettersson, I., Sharp, G., et al. 1986. *J. Cell Biol.* 102:137–44
154. Habets, W. J., Hoet, M. H., DeJong, B. A. W., Van der Kemp, A., Van Venrooij, W. J. 1989. *J. Immunol.* 143:2560–66
155. Huang, S., Spector, D. L. 1992. *Proc. Natl. Acad. Sci. USA* 89:305–8
156. Huang, S., Spector, D. L. 1992. *Curr. Biol.* 2:188–90
157. Fu, X.-D., Maniatis, T. 1990. *Nature* 343:437–41
158. Carmo-Fonseca, M., Tollervey, D., Pepperkok, R., Barabino, S. M. L., Merdes, A., et al. 1991. *EMBO J.* 10:195–206
159. Zamore, P. D., Green, M. R. 1991. *EMBO J.* 10:207–14
160. Spector, D. L., Fu, X.-D., Maniatis, T. 1991. *EMBO J.* 10:3467–81
161. Gall, J. G., Callan, H. G. 1989. *Proc. Natl. Acad. Sci. USA* 86:6635–39
162. Sass, H., Pederson, T. 1984. *J. Mol. Biol.* 180:911–26
163. Vazquez-Nin, G. H., Echeverria, O. M., Fakan, S., Leser, G., Martin, T. E. 1990. *Chromosoma* 99:44–51
164. Puvion, E., Viron, A., Assens, C., Leduc, E. H., Jeanteur, P. 1984. *J. Ultrastruct. Res.* 87:180–89
165. Fakan, S., Bernhard, W. 1971. *Exp. Cell Res.* 67:129–41
166. Fakan, S., Nobis, P. 1978. *Exp. Cell Res.* 113:327–37
167. Beyer, A. L., Osheim, Y. N. 1988. *Genes Dev.* 2:754–65
168. Borer, R. A., Lehner, C. F., Eppenberger, H. M., Nigg, E. A. 1989. *Cell* 56:379–90
169. Martin, T. E., Okamura, C. S. 1981. See Ref. 13, 9:119–44
170. Lahiri, D. K., Thomas, J. O. 1985. *J. Biol. Chem.* 260:598–603
171. LeStourgeon, W. M., Barnett, S. F., Northington, S. J. 1990. See Ref. 53, p. 477
172. Dreyfuss, G., Swanson, M. S., Piñol-Roma, S. 1990. *The Eukaryotic Nucleus: Molecular Biochemistry and Macromolecular Assemblies*, ed. P. R. Strauss, S. H. Wilson, 1:503. Caldwell, NJ: Telford
173. Foe, V. E., Wilkinson, L. E., Laird, C. D. 1976. *Cell* 9:131–46
174. Lamb, M. M., Daneholt, B. 1979. *Cell* 17:835–48
175. Malcolm, D. B., Sommerville, J. 1977. *J. Cell Sci.* 24:143–65
176. McKnight, S. L., Miller, O. L. 1979. *Cell* 17:551–63
177. Sommerville, J. 1981. See Ref. 13, 8:1–57
178. Beyer, A. L., Osheim, Y. N. 1990. See Ref. 53, p. 431
179. Spann, P., Feinerman, M., Sperling, J., Sperling, R. 1989. *Proc. Natl. Acad. Sci. USA* 86:466–70
180. Sperling, R., Sperling, J. 1990. See Ref. 172, 2:453
181. Barnett, S., Friedman, D. L., LeStourgeon, W. M. 1989. *Mol. Cell. Biol.* 9:492–98
182. Barnett, S. F., Theiry, T. A., LeStourgeon, W. M. 1991. *Mol. Cell. Biol.* 11:864–71
183. Bennett, M., Piñol-Roma, S., Staknis, D., Dreyfuss, G., Reed, R. 1992. *Mol. Cell. Biol.* 12:3165–75
184a. Roth, M. B., Gall, J. G. 1987. *J. Cell. Biol.* 105:1047–54
184b. Matunis, E. L., Matunis, M. J., Dreyfuss, G. 1993. *J. Cell Biol.* In press
185. Mehlin, H., Daneholt, B., Skoglund, U. 1992. *Cell* 69:605–13
186. Jacob, M., Devilliers, G., Fuchs, J.-P., Gallinaro, H., Gattoni, R., et al. 1981. See Ref. 13, 8:193–246
187. Choi, Y. D., Grabowski, P. J., Sharp,

P. A., Dreyfuss, G. 1986. *Science* 231:1534–39
188. Sierakowska, H., Szer, W., Furdon, P. J., Kole, R. 1986. *Nucleic Acids Res.* 14:5241–54
189. Mullen, M. P., Smith, C. W. J., Patton, J. G., Nadal-Ginard, B. 1991. *Genes Dev.* 5:642–55
190. Mayeda, A., Krainer, A. R. 1992. *Cell* 68:365–75
191. Ge, H., Manley, J. L. 1990. *Cell* 62:25–34
192. Harper, J. E., Manley, J. L. 1992. *Gene Expr.* 2:19–29
193. Chase, J. W., Williams, K. R. 1986. *Annu. Rev. Biochem.* 55:103–36
194. Pontius, B. W., Berg, P. 1990. *Proc. Natl. Acad. Sci. USA* 87:8403–7
195. Pontius, B. W., Berg, P. 1992. *J. Biol. Chem.* 267:13815–18
196. Kumar, A., Casas-Finet, J. R., Luneau, C. J., Karpel, R. L., Merrill, B. M., et al. 1990. *J. Biol. Chem.* 265:17094–100
197. Munroe, S. H., Dong, X. F. 1992. *Proc. Natl. Acad. Sci. USA* 89:895–99
198. Fu, X.-D., Maniatis, T. 1992. *Proc. Natl. Acad. Sci. USA* 89:1725–29
199. Zamore, P. D., Green, M. R. 1992. *Nature* 355:609–14
200. Wahle, E. 1991. *Cell* 66:759–68
201. Lingner, J., Killerman, J., Keller, W. 1991. *Nature* 354:496–98
202. Raabe, T., Bollum, F. J., Manley, J. L. 1991. *Nature* 353:229–34
203. Takagaki, Y., MacDonald, C. C., Shenk, T., Manley, J. L. 1992. *Proc. Natl. Acad. Sci. USA* 89:1403–7
204. Sosnowski, B. A., Belote, J. M., McKeown, M. 1989. *Cell* 58:449–59
205. Mehlin, H., Skoglund, U., Daneholt, B. 1991. *Exp. Cell. Res.* 193:72–77
206. Chang, D. D., Sharp, P. A. 1989. *Cell* 59:789–95
207. Legrain, P., Rosbash, M. 1989. *Cell* 57:573–83

Annual Review of Biochemistry
Volume 62, 1993

CONTENTS

(*continued*)