Molecular Biology Reports 18: 73–78, 1993.

The hnRNP proteins

Matthias Görlach, Christopher G. Burd, Douglas S. Portman & Gideon Dreyfuss*
*Howard Hughes Medical Institute and the Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, PA, USA (*author for correspondence)*

Received and accepted 2 June 1993

Key words: heterogeneous nuclear ribonucleoproteins, hnRNP, pre-mRNA, RNA-binding proteins, RNA processing

HnRNAs, or pre-mRNAs, are the primary transcripts of RNA polymerase II in eukaryotic cells. These transcripts, which become functional mRNAs upon capping, splicing, and polyadenylation, are bound by hnRNP proteins as they emerge from the polymerase complex [1]. In human (HeLa) cells, immunopurification of hnRNP complexes has shown that they consist of a set of approximately 20 major (and many minor) polypeptides, ranging in size from 34 kDa to 120 kDa and denoted hnRNP A1 through hnRNP U [2]. The major hnRNP proteins are extremely abundant components of HeLa nuclei, their amounts far exceeding those of the major snRNPs, and approximately the same as those of the core histones.

Analysis of the hnRNP proteins has provided strong evidence that most (if not all) of them bind directly to RNA, [2–5] that they exhibit binding preferences and that the composition of hnRNP complexes appears to be transcript-specific [6, 7]. Immunolocalization studies of these proteins have shown (with several notable exceptions) a general, uniform nucleoplasmic staining, consistent with their role in nuclear RNA metabolism [8–10]. Recently, however, several hnRNP proteins (e.g. hnRNP A1) have been found to shuttle between the nucleus and the cytoplasm [11] (see below). The emerging picture is one of a dynamic hnRNP complex whose composition changes during the course of mRNA metabolism.

HnRNP proteins have also been studied in divergent organisms. Vertebrate organisms, such as mouse and chicken, have an assortment of hnRNP proteins very similar to those observed in HeLa cells, both immunologically and structurally. Lower metazoans, in particular *Drosophila melanogaster*, contain a somewhat less diverse collection of hnRNP proteins, most of which have a similar domain structure to the vertebrate hnRNP A/B group proteins [12–15]. Studies of nascent transcripts of the polytene chromosomes of *Drosophila* have clearly demonstrated the differential association of hnRNP proteins with pre-mRNAs in a transcript-specific manner [6]. In the yeast *S. cerevisiae* and in other lower eukaryotes, hnRNP proteins remain less well-defined. Candidates for such proteins have recently begun to emerge [16, Matunis MJ, Matunis EL & Dreyfuss G, submitted], and studies in these organisms will facilitate a better understanding of the functions of hnRNP proteins.

The structure of hnRNP proteins

The cloning and sequencing of cDNAs for most of the hnRNP proteins has revealed a modular structure and allowed the identification of several different motifs. Three major structural motifs have been identified: the RNP consensus sequence, the RGG box and the KH domain [1, 17].

The RNP consensus sequence (RNP-CS)

The most prevalent motif is the 90-100 aa RNP consensus sequence (RNP-CS [18]; also referred to as RRM [19, 20] or RNP motif [1]) with its canonical RNP 1 and RNP 2 sequences [18–21]. It is found in human hnRNP A1, A2/B1, C1/C2, E, G, F/H, and *D. melanogaster* hrp36, hrp40.1, hrp40.2, and hrp48 [1, 15]. Several hnRNP proteins (I, L, M) contain multiple copies of noncanonical RNP-CS domains [10, 22, 23]. The RNP-CS and sequences immediately adjacent to it confer the specific RNA binding activity of these proteins [3, 4, 24–26]. In proteins with multiple such RNA-binding domains (RBDs), the specificity of RNA-binding can be a property of multiple RBDs (e.g. hnRNP A1 and the poly(A) binding protein) [4, 27, 28], or each RBD can function independently (e.g. the U1 snRNP A protein) [24–26].

The structure of two different RBDs, the N-terminal RBD of the U1A snRNP A protein [29, 30] and that of the hnRNP C proteins [31], has recently been solved. Both RBDs fold into a $\beta\alpha\beta\beta\alpha\beta$ structure, forming an antiparallel four-stranded β-sheet which is packed against the two α-helices. The RNP1 and RNP2 consensus sequences are juxtaposed on the two central β-strands ($\beta 3$ and $\beta 1$). Refinement of the hnRNP C RBD has recently revealed a new structural feature between $\alpha 2$ and $\beta 4$, a short two-stranded antiparallel β-sheet with a tight turn [M. Wittekind, M. Görlach, G. Dreyfuss & L. Mueller, unpublished].

Mutagenesis [29, 32] and NMR studies [3] have identified candidate residues involved in RNA binding to reside mainly on the antiparallel β-sheet, including conserved positions of the RNP1 and RNP2 consensus sequences. In addition, the NMR data [3] and deletional analysis [M. Görlach, C.G. Burd & G. Dreyfuss, submitted] have identified residues flanking the RNP-CS of the hnRNP C proteins to be involved in RNA binding. Interestingly, several hnRNP proteins differ from each other only by small peptide inserts [33] immediately adjacent to the RNP-CS, suggesting that those amino acids could modulate sequence-specific RNA binding. Structural studies, combined with mutagenesis and RNA-binding studies, should reveal the determinants of RNA-binding specificity.

The RGG box

The RGG box, originally described in hnRNP U [34], is also found in many RNA-binding proteins including hnRNP A1. This sequence motif contains clustered repeats of Arg-Gly-Gly tripeptides with interspersed aromatic (Phe, Tyr) residues. In hnRNP U, the RGG box is necessary and sufficient for RNA binding [34]. However, hnRNP A1 and G contain RNP-CSs in addition to an RGG box, suggesting an additional role for this domain in these proteins (see below). The RGG box, like the RNP-CS, is also found in RNA-binding proteins involved in the biogenesis of pre-rRNA [34–36]. Circular dichroism studies suggest that the RGG box can adopt a β-spiral structure and that it changes the structure of the RNA substrate by partially unstacking its bases upon binding [37].

The KH domain

A new structural motif, termed the *K homology* domain (KH domain), has recently been discovered in the hnRNP K protein [17]. This domain spans about 40 amino acids and is found in a large number of proteins from evolutionarily distant organisms [17, 38]. Its main features are a conserved spacing of the hydrophobic branched side chain amino acids (I, L, V) and a central IGxxG pentapeptide that is reminiscent of but distinct from the core sequences of mono- and dinucleotide binding sites [39]. The fact that this domain is primarily found in known or likely RNA-binding proteins [17], including archaebacterial and eubacterial ribosomal S3 proteins [40] and the yeast MER1 protein [41], and that it bears some resemblance to the mono/dinucleotide binding motif, suggests that this domain is an RNA-binding domain.

Auxiliary domains

The hnRNP proteins of the RNP-CS family also contain domains outside their RBDs, which are termed auxiliary domains [18]. With a few exceptions, the functions of these auxiliary domains are unknown. It is likely that these domains are involved in protein-protein interactions, in modulating the RNA-binding properties of these proteins, and in determining their nuclear localization. Auxiliary domains frequently contain potential sites for covalent modification, and several human hnRNP proteins have been shown to be post-translationally modified. RGG box-containing proteins, such as hnRNP A1 [42], undergo dimethylation of specific arginine residues in their carboxy-terminal regions, and hnRNP A/B, C, and U are substrates for phosphorylation *in vivo* [43–45]. The precise effects of these modifications are not yet clear. One very interesting feature of these domains is that they usually are rich in one or a few particular amino acids, reminiscent of several families of transcription factors [46].

The glycine-rich domains found in the A/B-type hnRNP proteins from divergent organisms is the most prevalent type of auxiliary domain. Since all the proteins of this group have two RBDs, they are referred to as 2xRBD-Gly proteins [15]. The glycine-rich domain confers some nonspecific RNA-binding activity to hnRNP A1 [47], though this binding is weak compared to that of the glycine-rich RGG box region of hnRNP U [34]. The glycine-rich domain also mediates the cooperative binding mode of A1 [48] and exhibits RNA annealing activity (see below). Proline-rich regions [as found in hnRNP K, L, and the poly(A) binding protein] and glycine-rich domains are similar to regions in other proteins that have been shown to mediate protein-protein interactions [49, 50]. These regions may serve a similar function in hnRNP proteins.

Studies of oligomeric forms of hnRNP proteins suggest that protein-protein interactions may be important for the structure of hnRNP complexes [51, 52].

The functions of hnRNP proteins

It has long been expected that hnRNP proteins, being very abundant and avid RNA-binding proteins, play important roles in the metabolism of hnRNAs. Most experiments have focused primarily on the role of hnRNP proteins in pre-mRNA splicing reactions. Addition of antibodies to hnRNP C proteins [53] or to hnRNP A/B and C proteins [54] has been shown to inhibit the splicing of several different pre-mRNAs *in vitro*. More recent experiments have directly explored the role of hnRNP proteins in splicing by complementation assays. HnRNP A1 has recently been shown to affect 5′ splice site choice in concert with another protein, ASF/SF2 [55]. Interestingly, ASF/SF2 contains all the hallmarks of hnRNP proteins – namely an RNP-CS and an auxiliary (SR) domain [56, 57]. It is likely that the ratio of 2xRBD-Gly proteins to SR proteins is important for regulated pre-mRNA splicing [55]. An essential role for a *D. melanogaster* 2xRBD-Gly protein has recently been identified [58, E.L. Matunis, R. Kelley & G. Dreyfuss, submitted], and others have been shown to have important developmental functions [59]. However, a murine cell line completely deficient of hnRNP A1 has been obtained [60], indicating that A1 is not essential for basal splicing and raising the possibility that hnRNP proteins exhibit functional redundancy. The 57 kDa hnRNP I protein (also known as PTB) was purified on the basis of its ability to bind to the polypyrimidine tract of the 3′ splice site region of pre-mRNAs and to influence 3′ splice site choice [61–63]. It co-purifies with a 100 kDa protein, called PSF, in a large complex that is required to restore splicing to a depleted extract [64].

Functional significance of hnRNP proteins to RNA metabolism is also suggested by RNA-binding site mapping experiments. Preferred binding sites for the hnRNP A1, C1/C2, and D proteins have been mapped to the 3′ ends of several intron/exon borders [65] and SELEX experiments have begun to identify preferred binding sequences for many pre-mRNA-binding proteins. Thus far, the results of these experi-

ments have correlated well with previous studies and these experiments suggest that hnRNP proteins have a range of RNA-binding affinities. It will be necessary to determine the relative affinities for high-affinity and low-affinity sequences. Although the significance of high-affinity hnRNP protein binding sites is not known, it is obvious that these sites will affect the local constellation of bound proteins on a particular RNA. Since the major hnRNP proteins are very abundant, it is likely that each protein will be bound to many different binding sites. The result of such binding, in conjunction with cooperative interactions, would be to form a protein-RNA fibril in which most, if not all, of the pre-mRNA is bound by hnRNP proteins.

Understanding the biochemical consequences of the binding of hnRNP proteins to pre-mRNA is essential to understanding the function of hnRNP proteins. An important biochemical property of hnRNP proteins that has emerged from studies with hnRNP A1 is the ability to strongly stimulate nucleic acid annealing *in vitro* [66–68]. The domain primarily responsible for the RNA annealing activity of A1 is its glycine-rich auxiliary domain [68]. Moreover, phosphorylation of a serine in this domain abolishes this activity of A1 [69]. Fractionation of RNA annealing activities in HeLa cell extracts has recently revealed that many additional hnRNP proteins have RNA annealing activity [D. Portman & G. Dreyfuss, submitted]. The functional significance of RNA annealing activity may be related to processes such as pre-mRNA–snRNA interaction [70], snRNA–snRNA interaction, or the modulation of pre-mRNA secondary structure. However, RNA annealing activity may also be thought of as a manifestation of a more general effect caused by certain types of RNA-protein interaction, which may act to facilitate interactions in *trans*. Such activity could be due to protein-protein interaction and/or the modulation of RNA conformation and accessibility.

In thinking about their functions in the cell, the hnRNP proteins have traditionally been considered to be strictly nuclear proteins. However, it has recently been found that some hnRNP proteins shuttle between the nucleus and the cytoplasm and it has been demonstrated that hnRNP A1 (one of the shuttling proteins) is bound to polyadenylated RNA in both cellular compartments [11]. It is possible that the mRNA transport machinery is, actually, a protein transport machinery in which the RNA may be the cargo. Furthermore, since the mRNA in the cytoplasm is bound by hnRNP proteins, it invites the thought that hnRNP proteins may modulate gene expression by participating in a variety of cytoplasmic aspects of mRNA metabolism including translational regulation, mRNA degradation and subcellular mRNA localization.

Conclusions

The isolation of hnRNP complexes has identified many new proteins and their characterization has led to the identification of several motifs that are important for RNA binding. These motifs are present in a wide variety of proteins including splicing factors, ribosomal proteins, and several proteins of unknown function. These findings have blurred the lines of demarcation between proteins previously thought of as RNA "packaging" proteins and RNA processing factors. Recent findings on hnRNP proteins have suggested a plethora of possible functions along the pathway of mRNA metabolism. It can be expected that the next few years will see the unraveling of the detailed functions of hnRNP proteins.

References

1. Dreyfuss G, Matunis MJ, Piñol-Roma S & Burd C (1993) Annu Rev. Biochem. 62: 289–321
2. Piñol-Roma S, Choi YD, Matunis MJ & Dreyfuss G (1988) Genes & Dev. 2: 215–227
3. Görlach M, Wittekind M, Beckman RA, Mueller L & Dreyfuss G (1990) EMBO J. 11: 3289–3295
4. Buvoli M, Cobianchi F, Biamonti G & Riva S (1990) Nucl. Acids Res. 18: 6595–6600
5. Merrill BM, Stone KL, Cobianchi F, Wilson SH & Williams (1988) J. Biol. Chem. 263: 3307–3313

6. Matunis EL, Matunis MJ & Dreyfuss G (1993) J. Cell Biol. 121: 219–228
7. Bennett M, Piñol-Roma S, Staknis D, Dreyfuss G & Reed R (1992) Mol. Cell. Biol. 12: 3165–3175
8. Choi YD & Dreyfuss G (1984) J. Cell Biol. 99: 1997–2004
9. Leser GP, Escara-Wilke J & Martin TE (1984) J. Biol. Chem. 259: 1827–1833
10. Piñol-Roma S, Swanson MS, Gall JG & Dreyfuss G (1989) J. Cell Biol. 109: 2575–2587
11. Piñol-Roma S & Dreyfuss G (1992) Nature 355: 730–732
12. Matunis MJ, Matunis EL & Dreyfuss G (1992) J. Cell Biol. 116: 245–255
13. Haynes SR, Raychaudhuri G & Beyer AL (1990) Mol. Cell. Biol. 10: 316–323
14. Haynes SR, Johnson D, Raychaudhuri G & Beyer AL (1990) Nucl. Acids Res. 19: 25–31
15. Matunis EL, Matunis MJ & Dreyfuss G (1992) J. Cell Biol. 116: 257–269
16. Anderson JT, Wilson SM, Datar KV & Swanson MS (1993) Mol. Cell. Biol. 13: 2730–2741
17. Siomi H, Matunis MJ, Michael WM & Dreyfuss, G (1993) Nucl. Acids Res. 21: 1193–1198
18. Bandziulis RJ, Swanson MS & Dreyfuss G (1989) Genes Dev. 3: 431–437
19. Query CC, Bentley RC & Keene JD (1989) Cell 57: 89–101
20. Kenan DJ, Query CC & Keene JD (1991) Trends Biochem. Sci. 16: 214–220
21. Mattaj IW (1989) Cell 57: 1–3
22. Ghetti A, Piñol-Roma S, Michael WM, Morandi C & Dreyfuss G (1992) Nucl. Acids Res. 20: 3671–3678
23. Datar KV, Dreyfuss G & Swanson MS (1993) Nucl. Acids Res. 21: 439–446
24. Scherly D, Boelens W, Dathan NA, Venrooij WJ & Mattaj IW (1990) Nature 345: 502–506
25. Lutz-Freyermuth, C, Query CC & Keene JD (1990) Proc. Natl. Acad. Sci. USA 87: 6393–6397
26. Nagai K (1992) Curr. Opin. Struct. Biol. 2: 131–137
27. Burd C, Matunis EL & Dreyfuss G (1991) Mol. Cell. Biol. 7: 3419–3424
28. Nietfeld W, Mentzel H & Pieler T (1990) EMBO J. 9: 3699–3705
29. Nagai K, Oubridge C, Jessen TH, Li J & Evans PR (1990) Nature 346: 515–520
30. Hoffman DW, Query CC, Golden BW, White SW & Kenne JD (1991) Proc. Natl. Acad. Sci. USA 88: 2495–2499
31. Wittekind M, Görlach M, Friedrichs M, Dreyfuss G & Mueller L (1992) Biochemistry 31: 6254–6265
32. Jessen T-H, Oubridge C, Teo CH, Pritchard C & Nagai K (1991) EMBO J. 10: 3447–3456
33. Burd CG, Swanson MS & Dreyfuss G (1989) Proc. Natl. Acad. Sci. USA 86: 9788–9792
34. Kiledjian M & Dreyfuss G (1992) EMBO J. 11: 2655–2664
35. Fournier MJ & Maxwell ES (1993) Trends Biochem. Sci. 18: 131–135
36. Tollervey D, Lehtonen H, Jansen R, Kern H & Hurt EC (1993) Cell 72: 443–457
37. Ghisolfi L, Joseph G, Amalric F & Erard M (1992) J. Biol. Chem. 267: 2955–2959
38. Gibson TJ, Thompson JD & Heninger J (1993) FEBS Lett. (in press)
39. Saraste M, Sibbald PR & Wittinghofer A (1990) Trends Biochem. Sci. 15: 430–434
40. Rinke-Appel J, Jünke N, Stade K, Brimacombe R (1991) EMBO J. 10: 2195–2202
41. Nandabalan K, Price L & Roeder GS (1993) Cell 73: 407–415
42. Beyer AL, Christensen ME, Walker BW & LeStourgeon WM (1977) Cell 11: 127–138
43. Wilk HE, Werr H, Friedrick D, Kiltz HH & Schäfer KP (1985) Eur. J. Biochem. 46: 71–81
44. Holcomb ER & Friedman DL (1984) J. Biol. Chem. 259: 31–40
45. Dreyfuss G, Choi YD & Adam SA (1984) Mol. Cell. Biol. 4: 1104–1114
46. Mitchell PJ & Tjian R (1989) Science 245: 371–378
47. Kumar A, Casas-Finet JR, Luneau CJ, Karpel RL, Merrill BM, Williams KR & Wilson SH (1990) J. Biol. Chem. 265: 17094–17100
48. Cobianchi F, Karpel RL, Williams KR, Notario V & Wilson SH (1988) J. Biol. Chem. 263: 1063–1071
49. Ren R, Mayer BJ, Cicchetti P & Baltimore D (1993) Science 259: 1157–1161.
50. Parry DAD & Steiner PA (1992) Curr. Opin. Cell Biol. 4: 94–98
51. LeStourgeon WM, Barnett SF & Northington (1990) In: Struss PR & Wilson SH (Ed) The Eukaryotic Nucleus: Molecular Biochemistry & Macromolecular Assemblies (pp 477–502) Caldwell, NJ, Telford
52. Barnett SF, Friedman DL & LeStourgeon WM (1989) Mol. Cell. Biol. 9: 492–498
53. Choi YD, Grabowski PJ, Sharp PA & Dreyfuss G (1986) Science 231: 1534–1539
54. Sierakowska H, Szer HW, Furdon PJ & Kole R (1986) Nucl. Acids Res. 14: 5241–5254
55. Mayeda A & Krainer AR (1992) Cell 68: 365–375
56. Ge H, Zuo P & Manley JL (1991) Cell 66: 373–382
57. Krainer AR, Mayeda A, Kozak D & Binns G (1991) Cell 66: 383–394
58. Kelley R (1993) Genes Dev. (in press)
59. Karsch-Mizrachi I & Haynes SR (1993) Nucl. Acids Res. 21: 2229–2235
60. Ben-David Y, Bani MR, Chabot B, De Koven A & Bernstein A (1992) Mol. Cell. Biol. 12: 4449–4455
61. Gil A, Sharp PA, Jamison SF, García-Blanco M & (1991) Genes Dev. 5: 1224–1236

62. Patton JG, Mayer SA, Tempst P & Nadal-Ginard B (1991) Genes Dev. 5: 1237–1251
63. Brunel F, Alzari PM, Ferrara P, Zakin MM (1991) Nucl. Acids Res. 19: 5237–5245
64. Patton JG, Porro EB, Galceran J, Tempst P & Nadal-Ginard B (1993) Genes Dev. 7: 393–406
65. Swanson MS & Dreyfuss G (1988) EMBO J. 11: 3519–3529
66. Pontius BW & Berg P (1990) Proc. Natl. Acad. Sci. USA 87: 8403–8407
67. Kumar A & Wilson SH (1990) Biochemistry 29: 10717–10722
68. Munroe SH & Dong XF (1992) Proc. Natl. Acad. Sci. USA 89: 895–899
69. Cobianchi F, Calvio C, Stoppini M, Buvoli M & Riva S (1993) Nucl. Acids Res. 21: 949–955
70. Buvoli M, Cobianchi F & Riva S (1992) Nucl. Acids Res. 20: 5017–5025