Nuclear RNA-Protein Interactions and Messenger RNA Processing

THORU PEDERSON
Cell Biology Group, Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545

ABSTRACT Eucaryotic messenger RNA precursors are processed in nuclear ribonucleoprotein particles (hnRNP). Here recent work on the structure of hnRNP is reviewed, with emphasis on function. Detailed analysis of a specific case, the altered assembly of hnRNP in heat-shocked Drosophila and mammalian cells, leads to a general hypothesis linking hnRNP structure and messenger RNA processing.

A unifying principle emerging from the modern era of biology is the realization that cellular processes can be understood in terms of chemical binding equilibria among macromolecules, such as those between nucleic acids and proteins. Here I review recent progress in the area of eucaryotic messenger RNA (mRNA)\(^1\) biosynthesis, with particular emphasis on nuclear RNA-protein interactions involved in mRNA processing.

Eucaryotic genes that code for mRNA are copied by RNA polymerase II into transcripts collectively termed heterogeneous nuclear RNA (hnRNA). The great majority of these nuclear transcripts undergo subsequent covalent modifications, through which some (but, importantly, not all) are converted into mRNA (1). The posttranscriptional modifications of hnRNA include addition of inverted guanosine nucleotides ("caps") at the 5' termini of most transcripts, specific base and ribose methylations, 3' processing followed by addition of poly(A) sequences at the 3' ends of some hnRNA molecules, and the excision of intervening DNA sequence transcripts followed by ligation of the mRNA sequences (splicing).

In addition to these covalent modifications of the RNA transcript, another important step in the maturation of mRNA is the assembly of heterogeneous nuclear RNA into ribonucleoprotein complexes, termed hnRNP particles (2). This begins while the transcript is still a nascent RNA chain (3–5). As 3'-OH poly(A) addition obviously cannot occur on nascent hnRNA chains (elongation proceeding 5' → 3'), and because poly(A) addition normally takes place before splicing (6), it follows that hnRNP assembly precedes both of these mRNA processing steps (7).

The central question is the functional significance of hnRNP particles. There are two extreme possibilities, which are not mutually exclusive. One is that hnRNP is simply a metabolically inert packaging device, involving a regular array of stable hnRNA-protein contacts analogous to the nucleoprotein organization of chromatin or viral nucleocapsids. The other possibility is that hnRNP particles reflect dynamic interactions of proteins at specific hnRNA sites related to mRNA processing, for example splicing. The former view of hnRNP has prevailed for many years, but recent evidence now points to the latter possibility. The distinction comes down to determining the extent to which the structure of hnRNP is nucleotide sequence-specific (8).

Historical Perspective

The study of hnRNP complexes has proceeded concurrently since about 1960 in the theaters of both cytology and biochemistry. Several reviews of both aspects of hnRNP research have been published recently (2, 9–12). The focus of the present article is on current and future directions in the field, for which the following synopsis is intended to serve as a background.

The idea that eucaryotic gene transcripts exist in the cell as nuclear RNP particles arose from cytological studies of meiotic prophase ("lambrush") chromosomes in amphibian oocytes. Rapidly labeled nascent hnRNA (13, 14) on the lateral loops of DNA was observed to be particulate (15), and...
hnRNPs. This is pertinent to the question of whether or not hnRNPs generate 30S RNP complexes from larger (50-300S) particles. The analysis of hnRNPs has been confirmed and further detailed through ultrastructural studies (17-19), including the analysis of nascent hnRNPs on chromatin spread from lysed nuclei by the procedures developed by Miller (see references 20-30).

The biochemical isolation of nuclear RNP particles containing hnRNA was pioneered by G. P. Georgiev and colleagues (31) in the Soviet Union. They showed that if rat liver nuclei are incubated in an isotonic buffer at pH 8.0, a major fraction of the rapidly labeled nuclear RNA is extracted in the form of 30S RNP complexes (31). The metabolic instability of this particle-associated RNA fraction and its DNA-like base composition led the Moscow group to conclude correctly that the particles contain pre-messenger RNA. The 30S RNP complexes were termed "informers" ("information bearers"), to contrast them from the cytoplasmic messenger RNA-protein complexes (mRNP) that Spirin had described several years earlier, which had been named "informosomes" ("information bodies"). The important work of the Moscow group on nuclear hnRNPs and cytoplasmic mRNPs, published mainly between 1965 and 1970, has been recently reviewed in detail (10, 32).

**hnRNP Structure**

The 30S nuclear RNP "informers" contain ~80% protein and 20% RNA as estimated from their buoyant density of 1.39 g/cm³ in CsCl. The proteins were initially reported to consist of a single protein species of ~40,000 mol wt (33), but subsequent studies using higher resolution gel electrophoresis systems have revealed the presence of a major sextet of hnRNP proteins, also known as "core" proteins, with molecular weights between ~32,000 and 42,000 (5, 8, 34-36). They are a closely related family of proteins as evidenced by their biochemical (34-36) and immunological interrelatedness (37; S. L. George and T. Pederson, manuscript submitted for publication).

A particularly important observation in Georgiev's group's original study (31) was that incubation of rat liver nuclei in the pH 8.0 isotonic buffer plus a cytoplasmic extract known to contain a potent ribonuclease inhibitor resulted in the rapid extraction of ~80% protein/20% RNA mass ratio. This indicated that the 30S particles are produced, in the absence of added ribonuclease inhibitors, by the action of endogenous nuclease(s) on large native RNP structures, presumably cutting at the exposed RNA sites linking adjacent 30S particles. This interpretation was supported by the conversion of the larger structures to 30S particles by deliberate ribonuclease treatment (31), and has been confirmed in several subsequent studies. However, it is to be noted that the generation of 30S RNP complexes from larger (50-300S) hnRNPs does not define the intersubunit spacing pattern, because the linking segments of RNA are destroyed in the process. We will later return to this important point, which is pertinent to the question of whether or not hnRNP structure is RNA sequence-specific.

**The Search for a Specific hnRNP**

With the development of methods for isolating large, native hnRNPs (2, 5, 38-41), the analysis of hnRNP organization was accelerated and a number of new features were revealed (e.g., 42-47), which have been reviewed elsewhere (2). However, around 1980 it became clear that hnRNP structure might be investigated more effectively using a specific pre-mRNA rather than the highly complex totality of hnRNA (48, 49). Our first encounter was with β-globin pre-mRNA in mouse erythroleukemia cells (50), but these sequences were found to comprise only 0.01% of the hnRNA (51), and therefore this was not a very propitious setting for purifying β-globin hnRNP. Virus-specific hnRNPs was investigated by others, but the mRNA processing pathways in the first examples studied proved too complex to be easily related to major differences in RNP structure with the techniques employed (52-54).

At this point in our studies, we turned to heat shock in *Drosophila* (55, 56), hoping that the transcripts of heat shock genes might be such prevalent components of the nuclear RNA that their isolation as hnRNP would be straightforward. But despite their prominence as preferentially translated mRNA in the cytoplasm (57-61), the heat shock gene transcripts turned out to be only 1-2% of the hnRNA (62). Once again, this was not judged to be a very hospitable situation for our objective of isolating a specific hnRNP. Despite this disappointment, we nevertheless decided to push on and characterize the total hnRNP particles of heat-shocked *Drosophila* cells. This is when we had a surprise.

**Heat Shock Blocks hnRNP Assembly**

In cultured *Drosophila* cells, we found that transcription of high molecular weight, nonribosomal, heterogeneous (10-40S) nuclear RNA continues at near normal rates after heat shock (63), confirming results reported previously by others (64). However, to our complete surprise, we found that the hnRNA synthesized after heat shock is not properly assembled into hnRNP. Instead of residing in the usual particles having an 80% protein/20% RNA mass ratio, the hnRNA synthesized during heat shock was found to reside in protein-deficient structures, with a composition of ~10% protein/90% RNA, as determined by CsSO₄ isopycnic banding (63). That this reflects a true alteration of hnRNP assembly in vivo and not an increased lability of particles during cell fractionation was established (63) by RNA-protein cross-linking experiments conducted in intact cells (65).

The effect of heat shock on hnRNP assembly is extremely rapid. For example, if *Drosophila* cells are raised from their normal culture temperature (25°C) to 37°C for 5 min and then labeled for an additional 5 min with [H]uridine, the newly synthesized hnRNA is not assembled into hnRNPs (63). As this is well before the onset of heat shock protein synthesis (66), these proteins are apparently not required for the block of hnRNP assembly. When heat-shocked cells are returned to 25°C, the capacity for normal hnRNP assembly is restored (63), in parallel with the gradual return of normal gene transcription and protein synthesis (62, 66). In addition, hnRNPs assembled at 25°C are found to promptly disassemble when cells are heat-shocked (63). This effect is not obtained when nuclei or hnRNPs from 25°C cells are incubated at 37°C in vitro, ruling out temperature per se as the important factor. Taken together, these results suggest that the inability of hnRNA made at 37°C to assemble into hnRNP is a very early component of the cell's physiological response to heat shock. A similar block of hnRNP assembly was observed in murine and human cells exposed to temper-
atures of 39–43°C (63), indicating that this is part of the evolutionary conserved stress response of eucaryotic cells.

The block of hnRNP assembly is not absolute, however, because in all cases the hnRNA made during heat shock is associated with a small amount of protein, rather than none at all. This is shown by the fact that the hnRNA from heat-shocked cells bands in Cs2SO4 at 1.58–1.63 g/cm³, even after RNA-protein cross-linking in situ (63), whereas RNA completely free of protein bands at 1.66 g/cm³ in these gradients. The conclusion that hnRNA made during heat shock is associated with some protein is also supported by direct analysis of the proteins by electrophoresis. While hnRNP from normal Drosophila cells contains the familiar group of 30,000–42,000-mol wt core proteins, the hnRNA from heat-shocked cells is associated mainly with a single component of ~35,000 mol wt (S. L. George and T. Pederson, manuscript submitted for publication).

The mechanism underlying this block of hnRNP assembly is not understood. It could involve rapid changes in the hnRNP proteins, leading to greatly reduced affinity for hnRNA. For example, hnRNP proteins are known to carry posttranscriptional modifications such as phosphoserine, phosphothreonine, and dimethyl-arginine residues (35, 36); and it is possible that these are altered during heat shock. It is also interesting that rapid dephosphorylations of histones have been reported after heat shock in both Drosophila and Tetrahymena (67, 68). Another general category of possibilities is the intranuclear solvent environment. Heat-shock induced changes in water content, pH or other ion activities could lead to altered equilibrium binding constants between hnRNP proteins and hnRNA. If such solvent changes do occur, they apparently do not affect all nucleoprotein structures in the nucleus to the same extent as hnRNP assembly. For example, chromatin retains normal nucleosome structure after heat shock, and the RNP structure of U1 small nuclear RNA is unaffected by the criterion of its reactivity with autoantibodies (69).

The heat shock-induced block on hnRNP assembly is the only known biological situation in which this process is under the investigator’s control, through the vehicle of imposed culture temperature. This leads us to believe that, irrespective of its utility for studying mRNA processing (vide infra), this is a very attractive system for learning about the assembly of hnRNP in vivo.

**hnRNP Structure and mRNA Processing—A Hypothesis**

The great majority of hnRNA made after heat shock in Drosophila cultured cells is not processed into mRNA (e.g., 70, 71), and this may be related to the block of hnRNP assembly (63). How then are the transcripts of the heat shock genes processed into mRNA under this condition of blocked hnRNP assembly? One possibility is that these transcripts, which comprise only 1–2% of the total hnRNA (62), do assemble into complete hnRNP particles and that it is the other 98–99% of the hnRNA whose assembly into hnRNP is blocked. This intuitively unlikely possibility is contradicted by hybridization of RNA from Cs2SO4 gradients, which suggests that heat shock gene transcripts are in the same protein-deficient structures as total hnRNA.

An attractive possibility raised by these results is that some gene transcripts may not form an hnRNP particle at all during their nuclear processing and maturation. Instead of being assembled into the usual hnRNP structures (80% protein/20% RNA), these transcripts might only bind a small number of proteins needed for a streamlined route of processing. If such transcripts were a minority, then their different RNP organization could easily have been missed in previous studies on total hnRNP. As no more than a working hypothesis and a basis for further work, I raise the possibility that the transcripts of genes that lack intervening sequences (introns) may represent such a case. As shown schematically in Fig. 1, two pathways of mRNA processing can be envisioned from the standpoint of RNP. One of these relates to genes with introns and is termed “major” because most of the hnRNA mass in mammalian cells comes from intron-containing genes (72). The removal of intron transcripts and splicing of mRNA sequences takes place in hnRNP particles (Fig. 1, major pathway), as is evidenced by their content of specific pre-mRNA sequences; e.g., see reference 50. RNP containing small nuclear RNA such as U1 may also be involved in splicing (73–76), adding to the overall complexity of RNP involved. (That the splicing process is complex is also suggested by the fact that it appears to comprise a considerably larger part of the nuclear residence time of pre-mRNA than the other posttranscriptional processing events; see reference 1.) In contrast to this major pathway, there may be a second, “minor” pathway (Fig. 1). This involves only the transcripts of intron-lacking genes and is a fast pathway that consists merely of capping and polyadenylation of the primary transcript, producing a messenger RNA ready for export to the cytoplasm. As indicated (Fig. 1), this pathway might involve only a small number of RNA-binding proteins, e.g., those associated with poly(A) or caps. In the context of this hypothesis (Fig. 1), heat shock is envisioned as shutting down the major pathway, by blocking hnRNP assembly. This does not exclude the possibility that the major pathway may also be blocked to some extent at the level of transcription rate of some pre-RNA (see Discussion, reference 63), but the fact that total hnRNA transcription occurs at nearly normal rates after heat shock (63) argues that the major factor in blocking this pathway is posttranscriptional. In contrast, heat shock would have no effect whatsoever on the operation of the minor pathway. This would mean that any intron-lacking gene could continue to produce functional mRNA during heat shock (as long as its transcription were not shut off).

Now it is possible to view this speculation in full perspective. All but one of the Drosophila heat shock mRNA-coding genes are known to lack intervening sequences (77, 78). (The one heat shock gene that does contain an intron [79] is expressed

---

**FIGURE 1** Two pathways of nuclear mRNA processing and ribonucleoprotein assembly. This is a schematic representation of the hypothesis developed in the text.
in normal cells [66, 80] and may therefore be subject to translational control.) In addition, newly transcribed histone mRNA continues to enter the cytoplasm during heat shock in Drosophila cultured cells [71] and these genes also lack introns [81]. The rapidity with which histone mRNA is exported to the cytoplasm [82] is also compatible with the hypothesis. Moreover, the genes for human α- and β-interferons (but not γ) lack introns, and these must produce functional mRNA during viral infection which, interestingly, is usually accompanied by fever. In fact, interferon production follows the course of fever [83] and our studies with cultured human cells show that hnRNP assembly is blocked at such temperatures (102–104°F) [63]. In addition, as predicted by the hypothesis (Fig. 1), interferon production is not inhibited by heat shock [84]. The more general possibility emerging from this hypothesis is that evolution has strongly selected against interruption by introns of those (few) genes that code for proteins essential for cell survival during environmental stresses such as heat shock, where intron-containing transcripts would abort due to blocked hnRNP assembly. The ability of intron-lacking “stress” genes to function in the face of blocked hnRNP assembly could also apply to short-term thermal acclimation (e.g., 85).

The hypothesis makes predictions. One is that intron-lacking gene transcripts circumvent hnRNP assembly even in normal cells, i.e., the minor pathway in Fig. 1 operates for intron-lacking gene transcripts in the absence of heat shock. We are testing this by examining the RNP structure of histone and interferon gene transcripts in non-heat-shocked mammalian cells. (With the advantage of hindsight, we now realize that we have always observed a small amount of hnRNA banding at 1.58–1.63 g/cm² in Cs₂SO₄ gradients of hnRNP from non-heat-shocked cells; e.g., see references 46 and 63.) Another prediction relates to the main question posed at the outset of this article: is the structure of hnRNP sequence-specific? Clearly, if it were the case that only intron-containing gene transcripts form hnRNP, then hnRNP assembly would have to be based on sequences restricted to these RNA, the most obvious possibilities being consensus sequences at exon–intron borders or splice signals within introns (e.g., 86). There is now independent evidence that this may be the case.

**hnRNP Proteins May Bind Specific RNA Sequences**

The analysis of nascent RNP fibrils on chromatin spread for electron microscopy, mentioned earlier in the section Historical Perspective, has been a major force in advancing our understanding of hnRNP structure and assembly [3, 20–30]. One of the most informative applications of this approach is that developed by Beyer et al. [26], in which nascent hnRNP on sibbling chromatin fiber axes is analyzed soon after DNA replication, where RNP fibrils can be seen emerging from loci in parallel register, i.e., from the same transcription unit. These studies have provided evidence that the location of proteins, or multi-protein complexes, on nascent hnRNA is sequence-dependent [26]. Note that this does not fundamentally contradict the original view of hnRNP structure [31], described earlier under Historical Perspective, but simply adds the constraint that the hnRNP subunits are restricted to particular sites along the hnRNA molecule. Recent biochemical experiments by Ohlsson et al. [87] strongly support this conclusion. Nuclease digestion of hnRNP and hybridization of protected RNA reveals two major nuclease-resistant sites in the nuclear pre-mRNA for an adenovirus early protein. One of these two hyperresistant sites maps in the middle of an intron and other near the end of an intron [87].

Is there any evidence against the postulate that hnRNP proteins bind specific sequences? hnRNP proteins can resemble with hnRNA or synthetic polyribonucleotides in vitro, when subjected to dialysis against low salt buffer from an initial protein–RNA mixture prepared at high ionic strength (88). Similar results have been obtained with a crustacean protein that is homologous to vertebrate hnRNP core proteins (89). In vitro experiments provide useful information on the solution properties of these proteins (see also reference 7) but do not address the issue of whether or not hnRNP proteins bind specific hnRNA sequences in the cell. It is well known that proteins which form stable contacts with specific nucleotide sequences also have a nonspecific affinity for nucleic acid, e.g., the cro repressor [90]. This nonspecific affinity is high (91) and it creates a diffusion-driven search process for the specific target sequence. This often neglected principle has been presented with particular clarity by von Hippel and colleagues [91–93]. Therefore, in vitro binding to nonspecific nucleic acids does not rule out the possibility that a protein binds only specific nucleotide sequences in vivo. Indeed, in vitro binding to total DNA is the phenomenon that has led to the isolation of proteins such as Escherichia coli DNA and RNA polymerases (94), which recognize specific nucleotide sequences in vivo: origins of replication and promoters, respectively.

**Summary and Prospectus**

I began this article by reviewing research on hnRNP structure and function, and emphasized that the central question in this field is whether hnRNP structure is RNA sequence-specific. The unexpected finding that heat shock blocks hnRNP assembly in Drosophila and mammalian cells [63] then led us to a hypothesis that only intron-containing gene transcripts are assembled into hnRNP. The hypothesis makes specific predictions that are now being tested.

Sometimes the analysis of a particular case can be expanded to reveal general principles [95]. The avenues of research that will most likely lead to unifying ideas on nuclear RNP and mRNA processing are: (a) ones in which hnRNP assembly is dramatically altered in vivo (reminiscent of the use of conditional lethal mutants to study bacteriophage assembly), (b) those in which one can map hnRNP protein binding sites on a defined pre-mRNA, and (c) experiments that test the direct involvement of hnRNP proteins in mRNA splicing.

Our work on hnRNP has been supported since 1972 generously and without interruption by the National Institutes of Health (CA-13400, GM-21595, GM-28274). I also thank the Leukemia Society of America for a Scholar Award (1972–1977). This is a most appropriate juncture at which to acknowledge the students, research assistants, and postdoctoral fellows who have worked with me on hnRNP over the past eight years: James Calvet (now at University of Kansas Medical Center), Nicholas Davis (now at the Rockefeller University), Ioannis Economidis (now at the International Institute of Cellular and Molecular Pathology, Brussels), Eileen Falvey (now at Yale University), Craig Findly (now at Yale University), John Greene (now at Harvard University), Christie Holland (now at the Massachusetts Institute of Technology), Valerie Kish (now at Hobart and William Smith Colleges), Sandra Mayrand (whose work on hnRNP assembly during heat shock is the focus of this article), Leslie Meyer
76. Mount, S. M., J. Petterson, M. Hinterberger, A. Kaeman, and J. A. Stelzl. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. Cell. 33:509-578.
77. Holmgren, R., K. Livak, R. Morimoto, R. Freund, and M. Meselson. 1979. Studies of cloned sequences from four Drosophila heat shock loci. Cell. 18:1359-1370.
78. Corces, V., R. Holmgren, R. Freund, R. Morimoto, and M. Meselson. 1980. Four heat shock proteins of Drosophila melanogaster coded within a 12-kilobase region in chromosome subdivision 67B. Proc. Natl. Acad. Sci. USA. 77:5390-5393.
79. Holmgren, R., V. Corces, R. Morimoto, R. Blackman, and M. Meselson. 1981. Sequence homologies in the 5' regions of four Drosophila heat shock genes. Proc. Natl. Acad. Sci. USA. 78:3775-3778.
80. Chonym, A., and H. K. Mitchell. 1982. Synthesis of the 84K heat shock protein in normal and heat-shocked D. melanogaster cells as determined by a specific antibody. Insect Biochem. 12:105-114.
81. Lifton, R. P., M. L. Goldberg, R. W. Karp, and D. S. Hogness. 1978. The organization of the histone genes in D. melanogaster: functional and evolutionary implications. Cold Spring Harbor Symp. Quant. Biol. 43:1047-1051.
82. Adenikin, M., and J. E. Darnell. 1972. Biogenesis and characterization of histone mRNA in HeLa cells. J. Mol. Biol. 67:397-406.
83. Petrulli, J. K., T. C. Merigan, and J. R. Wilbur. 1965. Circulating interferon after measles vaccination. N. Engl. J. Med. 273:198-201.
84. Ruiz-Gomez, J., and A. Isaacs. 1963. Interferon production by different viruses. Virology. 19:8-12.
85. Levine, R. 1969. Thermal acclimation and heat resistance in Drosophila species. Am. Nat. 103:483-499.
86. Mount, S. M. 1983. Sequences that signal where to splice. Nature (Lond.) 304:309-310.
87. Ohlsson, R. I., C. van Eekelen, and L. Philipson. 1982. Non-random localization of ribonucleoprotein structures within an adenovirus mRNA precursor. Nucleic Acids Res. 10:3053-3068.
88. Pullman, J. M., and T. E. Martin. 1983. Reconstitution of nucleoprotein complexes with mammalian heterogeneous nuclear ribonucleoprotein particles. J. Mol. Biol. 77:99-111.
89. Nowak, L., D. K. Marvil, J. O. Thomas, M. Boublik, and W. Szer. 1980. A single-stranded nucleic acid-binding protein from Artemia salina. II. Interaction with nucleic acids. J. Biol. Chem. 255:6473-6478.
90. Ohlendorf, D. H., W. F. Anderson, R. G. Fisher, Y. Takeda, and B. W. Matthews. 1982. The molecular basis of DNA-protein recognition inferred from the structure of cro repressor. Nature (Lond.) 298:718-723.
91. Berg, O. G., R. B. Winter, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. I. Model and theory. Biochemistry. 20:6929-6948.
92. Winter, R. B., and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translation on nucleic acids. 2. The Escherichia coli repressor-operator interaction: equilibrium measurements. Biochemistry. 20:6948-6960.
93. Winter, R. B., O. G. Berg, and P. H. von Hippel. 1981. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The Escherichia coli lac repressor-operator interaction: kinetic measurements and conclusions. Biochemistry. 20:6961-6970.
94. Alberts, B., and G. Herrick. 1971. DNA-cellulose chromatography. Adv. Enzymol. 21(3):198-217.
95. Frazer, J. G. 1890. The Golden Bough. MacMillan, London.