PROTEINS: Structure, Function, and Genetics 1:195–210 (1986)

Review Article

Set of Novel, Conserved Proteins Fold Pre-Messenger RNA Into Ribonucleosomes

Su Yun Chung[1] and John Wooley[2]
[1]*Lab of Experimental Carcinogenesis, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892;* [2]*Molecular Biosciences Division, National Science Foundation, Washington, DC 20550*

Key words: protein domains, oligomeric assembly, RNA splicing, multiple binding sites, RNP core proteins

INTRODUCTION

All nuclear RNA molecules are associated with protein, but different classes of RNA are associated with different polypeptide species, forming ribonucleoprotein complexes (RNP). Recent studies on the processing of messenger precursors (pre-mRNA) and on RNP proteins have pointed to the importance of understanding the structure and function of primary transcripts (hnRNA) in their ribonucleoprotein form (hnRNP). Concomitant with transcription, a set of structural proteins organize hnRNA into a chain of globular nucleoprotein particles or ribonucleosomes. There are exciting suggestions that hnRNP architecture is gene or transcript specific and that the structural integrity of ribonucleosomes is required for efficient and accurate processing of pre-mRNA. Interest in hnRNPs and ribonucleosomes has been further stimulated by recent direct evidence for an obligatory role of RNP proteins in splicing.

This article is a brief overview of the current understanding of the major structural proteins associated with hnRNA and the organization of the globular hnRNP subunit or ribonucleosome. Comprehensive reviews of studies on hnRNPs have appeared.[1–6] The current understanding of the details of RNA processing and role of small nuclear RNAs in nucleoprotein form (snRNP) in splicing are described elsewhere.[7–10]

A consensus view has emerged for the protein components of ribonucleosomes. Six major polypeptides, termed the core proteins, that are very abundant in transcriptionally active cells, conserved in evolution, and nuclear-restricted, serve to fold hnRNA into a beaded ribonucleoprotein complex or hnRNP. The individual beads, or ribonucleosomes, can be visualized in electron micrographs as 200–240-Å particles and isolated as globular complexes, sedimenting at 40S. Some details of the organization of ribonucleosomes, previously termed 40S particles or hnRNP monoparticles, are understood. The RNA is largely single stranded, compacted about 10–20 fold (RNA length/particle width), and possibly on the periphery of the complex. Within the past year, the stoichiometry of the core proteins has been determined, as have some details of their interactions with each other, and their relative spatial location. New insight has come from the realization that previously characterized eukaryotic helix-destabilizing proteins are specific proteolytic fragments of RNP core proteins. Rapid progress has also come from the molecular cloning of the genes coding for some of the core proteins. As a result, complete or partial amino acid sequences are now available. These studies have established the core proteins as a separate class of non-sequence-specific single-stranded RNA binding proteins. Individual RNP polypeptides are strongly conserved in evolution and appear to have an intriguing domain structure: the N-terminal region differs substantially from the C-terminal region in amino acid composition and secondary structure. The N-terminal region also contains one or more copies of an extremely conserved amino acid sequence element which is thought to be important for binding single-stranded RNA. Recent experiments unambiguously demonstrate that one group of the core proteins is required for splicing of pre-mRNA in vitro.

IMPLICATIONS OF RIBONUCLEOPROTEIN ARCHITECTURE

Both electron microscopic studies and biochemical studies indicate that all primary transcripts are associated with protein and folded into hnRNPs concomitant with transcription.[1,4–6,11,12] For nearly two decades investigators studying pre-messenger ribonucleoproteins have assumed that hnRNPs are the substrate, not naked RNA, for the various steps in RNA maturation. Since the discovery of splicing, it has seemed possible that the protein component serves as a matrix, scaffold, or workbench for that processing. Indeed, processing events—namely, loop formation and subsequent shortening of primary transcripts—presumably related to splicing, have been

Received October 2, 1986; accepted October 2, 1986.

Address reprint requests to J. Wooley, Molecular Biosciences Division, National Science Foundation, Washington DC, 20550.

visualized along hnRNP in situ on chromatin-associated transcripts.[4,13–15] Ribonucleosomes isolated from sucrose gradients contain both intervening sequences (introns) and exons.[1,16–18] However, only recently has biochemical evidence been obtained for a direct role of core proteins and ribonucleosomes in RNA splicing.[19]

A Multicomponent Complex for RNA Splicing

The processing of messenger RNA precursors in eukaryotes includes 5′ capping (addition of 7-methylguanosine), removal of intervening sequences (introns) by splicing, and 3′ polyadenylation (by poly [A] polymerase). A few aspects of splicing particularly relevant to ribonucleosomes are outlined here. Details of RNA splicing, beyond the scope of this article, and references have been recently provided.[10]

Splicing must occur in a precise fashion to generate correct proteins, since many protein-coding regions are interrupted by more than one intron. Accurate splicing requires sequence specific cleavage at the 5′ and 3′ splice sites and joining of the two ends of the consecutive exons. Consensus sequences at the 5′ and 3′ junctions between exons and introns have been found. The ligation step requires that the cut exon ends be held in juxtaposition, presumably by a RNA-protein complex, to prevent diffusion or incorrect joining. Studies on the mechanism of RNA splicing have exploited a cell-free extract system that can precisely splice exogenous RNA. In vitro splicing is ATP-dependent and occurs in at least two reaction steps. The first step involves cleavage at the 5′ splice site. A novel RNA form having a circular component and a tail, termed a lariat RNA, is generated in the next step. The lariat RNA is formed by the linkage of the 5′ end of the intervening sequence to an adenosine residue (at a specific branch site in the intron) through a 2′-5′ phosphodiester bond. This branch site is typically 20–50 nucleotides upstream of the 3′ splice site. The splicing intermediate consists of two RNA components, the 5′ exon and the lariat form of the intron sequences linked by a normal phosphodiester bond to the 3′ exon. The subsequent step(s), involving the cleavage of the 3′ splice site and ligation of the two adjacent exons, is less understood. The final product is the spliced mRNA; the intron is released in lariat form. (The lariat is debranched and processed in an independent reaction.)

In vitro studies also demonstrate that splicing occurs on large multicomponent nucleoprotein complexes.[20–23] Substrate RNA is first assembled, generally following a lag period, into a 30S to 40S complex. ATP and Mg^{2+} are required for formation of this complex and for splicing. Specificity in the reaction is believed to depend in part on additional obligatory components—namely, four individual snRNP complexes containing U1, U2, U5 and U4-U6, respectively.[10,23–29] U1 snRNP interacts with the 5′ splice site, U2 snRNP with the nucleotide stretch where the branch will occur, and another snRNP, probably U5, with the 3′ splice site. However, some of the snRNPs, notably U1, may act transiently, and then dissociate from the larger complex.[28,29]

The multicomponent, active complex, termed a splicesome,[20] sediments at 50S–60S in mammalian extracts. This 60S complex forms in the extracts during the lag period, initially including intact pre-mRNA; then two RNA intermediates, the 5′ exon and the lariat-3′ exon, are observed. The intermediates are found only after formation of the splicesome (and not in the 40S complex), and are presumably held together in this complex until completion of the splicing reaction.[10,19,26,27] At least transiently, the splicesome must contain a wide diversity of snRNPs, possibly other factors and enzymes, and, as described below, at least some of the RNP core proteins.

Nonrandom, Gene-Specific hnRNP Architecture

Much of the information on the organization of long hnRNP transcripts comes from electron microscopic studies employing a nuclear spreading method.[30] Low ionic strength and centrifugal force disrupts and disperses nuclear contents, extending chromatin fibers and permitting visualization of chromatin-associated transcripts as lateral fibrils extended from the chromatin filament. Under a variety of preparatory conditions, dense chains of globular particles, or ribonucleosomes, about 200–240 Å in size, are seen along nascent transcripts.[4,13,15,31] Under conditions of moderate dispersion, a good correspondence is seen for the spacing and size of ribonucleosomes along hnRNPs and the particulate or granular complexes of pre-mRNA observed in thin sections.[32] A lucid discussion of the relationship between the many different RNA-containing structures in the nucleus and known biochemical entities has been provided elsewhere.[33]

Evidence for gene-specific hnRNP architecture has come from experiments using very low ionic strength at basic pH for the spreading conditions. Under these stringent spreading conditions, hnRNPs are extended and partially unfolded; since some particulate regions resist unfolding, it appears that ribonucleosomes differ in their chemical and physical stability.[13,15,34–40] Extended hnRNPs exhibit three morphological features: intact 200–240-Å particles which appear to be ribonucleosomes, smooth 40-Å fibrillar segments, and some smaller particulate regions, intermediate in size between the fiber segments and intact ribonucleosomes. Thus, a resistant, stable subset of particles has been found to exist under complete dispersion. Strikingly, the location of these resistant particles, or stable RNPs, is consistent and specific for a given transcriptional unit. Linear regression analysis of transcript length versus transcript position indicates that the individual transcripts in each unit form a smoothly increasing length gradient. This suggests that each transcript initiated from a common promoter and is folded equivalently. The pattern of the arrangement of stable RNPs along hnRNP fibrils dif-

fers among different transcriptional units. When sister chromatids are found adjacent to each other, the paired sister transcriptional units show the same pattern for transcript folding. The location of stable RNPs relative to the 5′ ends has been mapped.[34] The experiments demonstrate: 1) stable RNPs form soon after synthesis of a given sequence and then are observed at that site on all longer transcripts; 2) the pairs of transcripts on daughter chromatids exhibit identical patterns in terms of the localization of stable nRNPs; 3) the localization of stable particles along hnRNPs differs among all genes observed.

In other words, the packaging of the RNA can be considered at one level as a "Morse code" pattern, a structural approximation of dots and dashes, that is unique to each gene. The underlying sequence of RNA dictates the overall hnRNP architecture. If real, gene-specific hnRNP packaging must have significant implications for pre-mRNA processing.

In general, a continuous fiber gradient is observed upon examination of a collection of hnRNPs along a transcriptional unit. Occasionally, two fiber gradients, consistent with concomitant RNA processing during transcription, are observed. For transcriptional units exhibiting apparent processing, two stable RNPs are sometimes seen to coalesce, forming a larger particulate structure, about 400 Å in diameter, with a loop of RNP fibril extended out. The loops are then released, leaving a shorter transcript.

To evaluate the relationship of gene-specific hnRNP architecture to RNA processing, it is necessary to examine known genes. In general, it is not technically feasible to identify genes in nuclear spreads. However, the two *Drosophila* eggshell protein (chorion) genes provide a convenient, characterized system for the correlation of the location of stable RNPs with known DNA sequence.[15] The closely linked chorion genes, s36-1 and s38-1, each have a small 5′-proximal intron, are vigorously transcribed during specific stages of oogenesis, and lie within a 16–20-fold amplified region of the X chromosome.[41,42] These details permit identification of the s36 and s38 transcriptional units in nuclear spreads from staged follicle cells. Under moderate spreading conditions, the chorion transcripts are folded into a chain of tightly packed ribonucleosomes. Under stringent spreading conditions, only two stable particles are observed, and these lie at the splice junctions. On longer transcripts, the stable RNP particles coalesce to form a single 400-Å particle. This experiment suggests a direct relationship between stable RNPs and RNA splicing. It is possible that the coalescence of two ribonucleosomal-sized particles into a larger particle visualized in the electron microscope represents the same dynamic event as the conversion of the 40S component to the 60S splicesome in cell-free splicing systems. Recently, antibodies recognizing snRNP components, as well as antibodies recognizing core proteins, have been found to bind to hnRNP transcripts in nuclear spreads.[43] A tight complex of snRNPs, hnRNPs, and possibly other factors and core proteins, may well be the most resistant or stable structures along hnRNPs. Stable RNPs are considerably larger than individual snRNPs,[44] but the size and shape of the multicomponent complexes active in splicing are not known. Although they are the same size, the precise relationship of the stable RNPs to the less-stable, bulk population of ribonucleosomes along hnRNP chains is uncertain.

Alternative or Differential Splicing

Some eukaryotic transcriptional units are spliced in more than one pathway, that is, alternate splice sites are utilized, producing more than one mature mRNA. In some cases, the different forms of the mRNA encode different polypeptides. The mechanisms for differential processing are not understood, although many examples have been characterized from yeast to mammals.[9] An extreme case of diversity is observed for the splicing of the 18 exons of rat troponin T[45]; up to 64 mature RNA species can be produced in the various muscle types or during development.

Tissue-specific factors, possibly modified snRNPs, are thought to be involved in the selection of alternate sites.[9,10,25] Such factors could also modulate hnRNP structure, at least with respect to placement of "stable particles" and thus splice site selection. Isomeric conformations of ribonucleosomes or changes in their placement along hnRNPs could serve to mask particular sites from binding of snRNPs. The rich diversity of splicing in late adenovirus infections provides a version of alternative splice selection that has been examined with the nuclear spreading method. The adenovirus late transcriptional units were found to exhibit gene-specific hnRNP architecture, with a finite number of arrangements for a given transcript.[36,46] The observations are consistent with a limited set of alternative splicing pathways, but any relationship of hnRNP architecture to alternative splicing remains speculative.

THE RNP CORE PROTEINS AND RIBONUCLEOSOMES

Biochemical investigations on non-ribosomal, nuclear RNPs were begun 20 years ago by Samarina, Georgiev, and colleagues, who hypothesized the existence of a repetitive subunit with a RNA wrapped around a protein core, much like the current models for the nucleosome and the chromatin fiber.[47,48] They observed both a putative monomer and a set of oligomers, partially separated across sucrose gradients of low salt-extracted nuclei, and demonstrated that a monomer or short oligomeric chains could be reconstituted from separated RNA and protein components. The overall implications of their work are consistent with current research, even though some specific details of their experimental observations have turned out to be incorrect, owing to technical limitations of the methods employed. The evidence

for their key point, the existence of a stable, globular, RNP complex, or ribonucleosome, released by ribonucleolytic cleavage of hnRNP, is unambiguous. Many independent observations have demonstrated that isolated ribonucleosomes do reflect a conserved, unique structural element of nuclei in vivo.[1,3,49-53] These complexes are readily obtained as the 40S peak in sucrose gradient fractions of nuclear extracts. It has proven much more difficult to isolate discrete oligomeric units, although long RNA molecules in ribonucleoprotein form have been obtained by sonicating nuclei, or by carrying out extractions in the presence of esoteric RNase inhibitors.[5,6,54-57] The proteins observed for such hnRNP preparations are more heterogeneous, but it is now clear that the structural protein components of ribonucleosomes are also the major, consistent components of hnRNP.[57-59] The minor species observed may include snRNP polypeptides and enzymes of hnRNA metabolism. In our review, we will focus on the ribonucleosome and its major protein components.

The Major Proteins: Identification and Categories

Ribonucleosomes are primarily composed of (multiple copies of) six nucleus-restricted polypeptides that are of fixed stoichiometry, as expected for integral components of a specific macromolecular assembly. These major polypeptides, or RNP core proteins, migrate in SDS-polyacrylamide gels as three groups of doublet bands: each polypeptide of a doublet band shares some common chemical and physical properties with its pair, which led to the terminology A1, A2, B1, B2, C1, and C2 for the polypeptide species (see Table I for properties based on mammalian sources). These polypeptides are the most abundant nonhistone nuclear proteins in growing cells.

The core proteins have been identified and categorized by a set of criteria. A key point, along with the nuclear abundance of these species, indicating that these are genuine structural polypeptides was the recognition of an invariant, simple stoichiometry[51] for RNP polypeptides over a range of tissues and species, and across the 40S peak under a variety of sedimentation conditions. A1, A2, and C1 are apparently equimolar based on gel scans, and are present at three times the amount of B1, B2, and C2.[60] The absolute ratios are certainly not proven, but assuming the simplest case, represent the best estimate. A1 and A2 always appear equimolar and can be cross-linked into equimolar heteropolymers within ribonucleosomes.[60] Chemical cross-linking data (see below) also strongly support the three-to-one ratio between the three most abundant polypeptide species (A1, A2, C1) and the others (B1, B2, C2).

Another essential feature comes from reconstitution studies that have established that the isolated core proteins can fold exogenous RNA into 40S ribonucleosomes in vitro.[50,52,61,62] The polypeptides appear to be present in the reconstituted complexes in the same stoichiometric ratios as in the native ribonucleosomes. In addition, monoclonal antibodies recognizing C1 and C2 precipitate A1, A2, B1, B2, C1, and C2 as a large RNA-protein complex from nuclear extracts.[59] Again, the protein ratios in this large complex are the same as those observed for total nuclear extracts and for isolated ribonucleosomes.

An alternate approach to identifying hnRNP polypeptides is to cross-link protein to RNA in vivo with UV light.[63] RNA-protein complexes are then isolated under denaturing conditions. It appears likely that the levels of UV light employed are insufficient to unfold or denature ribonucleosomes. Although this may miss some genuine ribonucleosomal proteins, the polypeptides identified by this procedure are intimately associated with RNA. C1 and C2 are most readily cross-linked to RNA, but with increased exposure, A1, A2, B1, and B2 are also cross-linked to RNA.[57,58,64,65]

The three protein groups were originally categorized on the basis of their salt-extraction properties, amino acids composition, and isoelectric properties, along with their gel mobility.[51,52] The properties of chromatin and histones[66] suggest that these physical and chemical parameters might be informative in

TABLE I. Core Proteins of Ribonucleosomes*

Polypeptide species	M_r (kd)	MW (kd)	Stoichiometry	P_I	Selected amino acid content				Dissociation with NACL	Phosphorylated
					GLY	LYS & ARG	GLX & ASX	DMA		
A1	32	35	3	9.3	20	13	20	+	0.25 M	−
A2	34	ND	3	8.4	25	11	20	+	0.25	+
B1	36	ND	1	8.8	22	16.5	20.5	+	0.30	+
B2	37	ND	1	9.2	24	12	14.5	+	0.30	−
C1	42	30	3	4.8	9.2	15	29	−	0.70	+
C2	44	30	1	4.8	9.8	15	29	−	0.70	+

*Nomenclature, apparent molecular weight, M_r, for HeLa, and the molarity at which the species are released from ribonucleosomes is from Beyer et al.[50] Molecular weight for A1 is from Cobiachi et al.;[77] for C_1 and C_2 from Swanson et al.[75] Stoichiometry from Lothstein et al.[60] Apparent P_I, amino acid data from Wilk et al.[63] DMA is dimethylarginine. Phosphorylation data from Dreyfus et al.[65] and Holcomb and Friedman.[68]

establishing structural and functional distinctions among the different core proteins. Some of these distinctions can now be expanded.

1. The A group polypeptides, A1 and A2, with M_r 32 kilodaltons (kd) and 34 kd, respectively, constituting 50% of the protein mass of ribonucleosomes, have basic isoelectric points, unusually high amounts of glycine, and contain the unusual methylated amino acid, N^G,N^G-dimethylarginine. A1 and A2 simultaneously dissociate from ribonucleosomes upon exposure to increasing salt, coming off at 0.25 M NaCl, possibly as a linked pair.

2. The B group polypeptides, B1 and B2, with M_r 36 kd and 37 kd, respectively, are also basic, glycine-rich, and contain dimethylarginine. (20% of the arginines in B1 are the dimethyl form.) The amino acid composition is similar to that of the A group, but B1 and B2 dissociate simultaneously from ribonucleosomes at a slightly higher salt—namely, 0.3 M NaCl.

3. The C group polypeptides, C1 and C2, M_r 42 kd and 44 kd, respectively, are distinctly different. They are acidic, and do not fully dissociate from RNA until 0.7 M NaCl. Subsequently, it has become clear that these proteins are significantly lower in glycine content, and may lack dimethylarginine.[67] The C proteins are highly phosphorylated, and in HeLa, are the major phosphorylated RNP polypeptides[65,68]. More recently, the C proteins have been shown to be intimately associated with RNA, since they can be quantitatively cross-linked to RNA, in the virtual absence of cross-linking of other proteins, by relatively low levels of exposure to ultraviolet light[64,65] (G. Conway and W. LeStourgeon, personal communication) and are rapidly released from isolated ribonucleosomes by mild digestion with RNase.[60]

The properties described above show that the distinctions between the C proteins and the A and B proteins are quite clear. We believe that this distinction is real and reflects functional and structural differences. The A and B groups of core proteins are not so readily distinguished. This is particularly true if one considers the minor, glycine-rich RNP polypeptides (described in Table II) that are very closely related to the major A and B group proteins, based on the criteria used above.[67,69–72] Some of the polypeptides might arise by secondary modifications. They are present at less than one copy per ribonucleosome, with the possible exception of B1b.[67,69] (Owing to their low abundance and lack of characterization, these putative minor core proteins are only discussed in this context. Some less abundant, higher molecular weight species are also consistently observed.[51,56]) Recent studies employing monoclonal antibodies as probes indicate a close relatedness among all of the A and B polypeptides.[71,72] More information is required for a definitive categorization of the A and B proteins. An alternative scheme, showing closely related members of four hypothetical subgroups within the A and B group of glycine-rich core proteins, is provided (Table III). This scheme is based on amino acid difference indices[67] and is consistent with data on the effects of nuclease and protease digestion of ribonucleosomes.[60,69]

A1, A2, B1, and B2 are clearly chemically distinct from one another; that is, none arises as a secondary modification of another. In contrast, C1 and C2 have similar amino acid compositions and the same isoelectric point, and behave similarly on all physical and chemical treatments of ribonucleosomes. All characterized anti-C monoclonal antibodies react with both proteins,[73] even though these antibodies recognize different epitopes. The fixed stoichiometry makes it hard to see C2 as a precursor to C1, or as alternative conformations of the same chemical sequence. Indeed, the same hybrid-selected mRNA has been found to direct synthesis of both C1 and C2 in vitro; sequence analysis of a cDNA clone for the C proteins has confirmed that they have the same primary sequence.[74,75] C1 and C2 must differ in very quantitatively regulated, specific secondary modifications, which do not alter the charge (pI). Whatever this modification, it can occur in a regulated fashion in in vitro translation systems. On the other hand, the consistent, invariant ratio of C1 to C2, and the slight difference in protease sensitivity between C1 and C2 in ribonucleosomes, but not as isolated polypeptides, suggest distinct structural roles for C1 and C2 in the supramolecular assembly.[60]

Conservation and Relatedness of Core Proteins

Two features of the core proteins have been elucidated through comparative analysis. One, the polypeptides are evolutionary conserved, and two, the different polypeptides are closely related to one another. Since only one RNP polypeptide species has been found to be present in ribonucleosomes from

TABLE II. Less Abundant Glycine-Rich RNP Polypeptides*

	M_r (Kd)	P_I	Molar percent glycine	Presence of dimethylarginine	Phosphorylated
B1b	37	8.3	24	+	+
B1c	37	5.7	18	+	+
C3	42	5.6	19	+	−
C4	44	9.2	26	+	−

*Data derived from Wilk et al.,[65] except apparent molecular weight from Beyer et al.[50] Nomenclature is that of Celis et al.[67]

TABLE III. Possible Relationships Among A and B Group Proteins*

Subset	Polypeptides
1	A1
2	A2, B2, C4
3	Bla, Blb
4	Blc, C3

*Relatedness among polypeptides was assigned from amino acid difference indices of Wilk et al.;[66] nomenclature is that of Celis et al.[67]

lower eukaryotes (and possibly only two in higher plants), it has seemed reasonable to assume that all core proteins arose from a common ancestor and evolve by way of gene duplication and subsequent divergence.

Evidence for conservation during evolution came first from observations of immunological cross-reactivity among these polypeptides. For example, a set of polyclonal antibodies recognizing all of the A and B core proteins were observed to bind transcriptionally active nuclei of human, amphibian, and chicken brain cells, but not to transcriptionally inactive chicken erythrocytes.[76] Mouse monoclonal antibodies that react against one or more A and B proteins recognize core polypeptide species (in immunoblots) across a wide variety of vertebrates.[71,72] Similarly, a mouse monoclonal antibody (termed 4F4) that reacts with the C polypeptides identifies C1 and C2 in cells ranging from lizard to human (based on immunoblotting and immunofluorescence).[73] However, at least one region of the C polypeptides is less conserved, since a different monoclonal antibody (termed 2B12) was found to react only with proteins from higher mammals.

More direct evolutionary and structural comparisons are now possible through amino acid sequencing of peptide fragments or molecular cloning of the core protein genes. The complete amino sequence deduced from a cDNA clone for rat A1,[77] the first 195 residues of calf A1[78,79], and the last 196 residues of human A1[80] are known. The partial human sequence is identical to the rat sequence and to the overlapping region of calf A1; the partial N-terminal calf sequence is also identical to that of rat. Sequencing the N-terminus of human A1 has picked up only one change, which turns out to be conservative; rat and calf A1 have an arginine at position 31, whereas human A1 has a lysine at this position.[81] This high degree of homology in A1 indicates more significant evolutionary constraints than had been expected. Less sequence data are available for the B and C proteins. However, genomic DNA blot analysis with a cDNA clone of human C proteins as probe was found to give multiple bands at high stringency with human, mouse, chicken and frog, and at lower stringency with *Drosophila* and yeast.[74]

Studies on RNP proteins of lower eukaryotes also support sequence conservation in evolution, although lower eukaryotes have a much simpler complement of polypeptide species. Only one RNP core protein can be isolated from the slime mold, *Physarum*,[82] or the brine shrimp, *Artemia*[83], M_r of 36 kd and 40 kd, respectively. Both polypeptides are basic, glycine-rich, contain dimethyarginine, and thus look like members of the vertebrate A and B groups. Polyclonal antisera against the polypeptide from *Physarum* cross-react with mammalian core proteins.[51] Similarly, a partial cDNA clone corresponding to the RNP protein in brine shrimp, HD40, hybridizes at reduced stringency with genomic DNA from avian, rodent, and plant sources.[84] More studies are required before a precise relationship between these proteins from primitive eukaryotes and the mammalian core proteins can be established.

Evidence for the closely related nature of the A, B, and C core proteins has come from antibody studies and recently from a few amino acid sequences. Both polyclonal and monoclonal antibodies to core proteins have been observed to react with several or all six of the core polypeptides.[71,76] Only a few hybridomas expressed immunoglobulins recognizing an epitope present on only one RNP polypeptide.[71,72] Recent sequence data on A1, A2, and their proteolytic fragments confirm a high degree of sequence relatedness.[77–81,85] Tryptic peptide maps and partial sequencing data suggest that human A1 and A2 share as much as 80% sequence homology.[81] Difference indices of amino acid composition and partial proteolytic data suggest all A and B polypeptides are closely related; for example, B2 may be even more closely related to A2 than is A1.[67,72]

Implications of Amino Acid Sequence Information

A class of putative eukaryotic single-stranded DNA binding proteins, namely calf UP1, UP1b, and U2 (UP for unwinding protein), rat helix-destabilizing protein (HDP), and mouse HDP-1, have recently been shown to be specific proteolytic degradation products of the RNP polypeptides, A1 and A2, and possibly of other RNP polypeptides as well.[80,86,87] As a consequence a considerable amount of sequence and chemical information on fragments of A1 and A2, as well as a complete cDNA clone for rat A1, are now available.

Several lines of evidence indicate that the protein A1 has several distinct structural domains. Limited proteolytic cleavage in vitro of A1 yields discrete polypeptide fragments (including UP1), indicating the existence of a protease sensitive region in the "border" between two different structural domains.[77,81] This cleavage may be directed by the N^G,N^G-dimethylarginine at position 193, out of the 195 amino acids of UP1.[81] Overall, A1 can be seen as composed of a globular, organized N-terminal region and a very glycine rich C-terminal region.[77,79,81,85] Based on the deduced amino acid sequence from a complete cDNA clone of a rat helix-destabilizing protein that turned out to be A1, a sharp distinction has been clearly

demonstrated between the aminoterminal 190 residues and the carboxyterminal residues from 195 to 320.[77,85] Plots of secondary structure all make an abrupt transition in predicted conformation between residues 190 and 195. The aminoterminal regions are predicted to be highly structured with four alpha helices, plus six or seven beta-sheets with alternating hydrophobic and hydrophilic regions.[77,79] In contrast, the glycine-rich C-terminal is slightly hydrophilic and predicted to form beta-pleated sheet or random coil. Preliminary analysis of A2 and its proteolytic fragment UP1b indicates that A2 is also divided into distinctly different domains.[81]

Putative domains for protein-RNA interactions

The N-terminal segment of A1, corresponding to UP1, is itself divided into two repetitive regions, termed the A and B domains.[79] The globular A and B domains have a high degree of internal sequence homology. When residues 4 to 94 and 95 to 185 are aligned, 32% of the amino acids are identical, and 39% of the changes can be accomplished by single nucleotide substitution.[79,81] Studies involving covalent oligonucleotide-protein cross-linking and limited proteolysis suggest that A and B represent two domains that can independently bind nucleic acids; also, the binding is not sequence specific.[81,88] UP1 was originally identified by its single-stranded DNA binding activity and therefore helix-destabilizing activity; UP1 also depresses the melting temperature of double-stranded nucleic acid polymers.[89] UP1 can virtually coat single-stranded nucleic acids by way of noncooperative binding and the binding site spans about seven nucleotides.[90]

The entire A and B domains of A1 are rich in basic and aromatic amino acids, which is a common feature in the nucleic acid binding sites of prokaryotic single-stranded DNA binding proteins. For these ssDNA binding proteins, the aromatic residues are believed to interact with the bases through intercalation and the positively charged amino acids are presumably involved in ionic interactions with the backbone phosphates.[91] When the A and B domains are aligned, 62% of the basic residues and 80% of the aromatic residues occur in precisely the same relative positions; in contrast, only 26% exact homology is observed for non-basic and non-aromatic amino acids.[79] Based on this observation, two sequence elements or motifs in A1 (UP1) spanning residues 52–62 and 143–153, chosen for their high density of aromatic and basic amino acids, have been hypothesized[79] to constitute two independent single-stranded nucleic acid binding sites (Fig. 1). The presence of multiple nucleic acid binding domains may be a general feature of eukaryotic single-stranded nucleic acid binding proteins.[81,91]

A sequence similar to the putative single-stranded nucleic acid binding site has been found within another RNP polypeptide, based on the amino acid sequence predicted from a partial human cDNA clone, termed DL4.[92] Overall, the DL4 clone contains an open reading frame for 225 amino acids that share homology with A1, A2, and notably, with calf UP2. In particular, a segment homologous to the putative binding site of A1 is present (Fig. 1). The RNP polypeptide to which DL4 corresponds remains to be established. DL4 does not correspond to the C proteins, as originally suggested.[75,92] In contrast, the predicted amino acid sequence from DL4 shows a perfect match with several tryptic peptides of calf UP2,[75,88] a 40-kd polypeptide that is glycine-rich, contains dimethylarginine, and binds ssRNA.[78,81] It is likely that DL4 corresponds to a less abundant member of the A-B group of core proteins. More sequence data are required before precise relationships can be established.

Studies on another type of RNA-binding proteins, those associated with poly(A), have strengthened and extended our understanding of the significance of the

A1	^{53}arg	SER	ARG	GLY	PHE	GLY	PHE	VAL	thr	tyr^{62}			
A1	^{143}GLY	LYS	lys	ARG	GLY	PHE	ala	PHE	VAL	thr	PHE^{153}		
DL4		LYS	arg	ARG	GLY	PHE	cys	PHE	ile	thr	PHE		
PAB			^{79}SER	LYS	GLY	tyr	ala	tyr	VAL	asn	PHE^{87}		
PAB	^{161}glu	asn	GLY	LYS	SER	LYS	GLY	PHE	GLY	PHE	VAL	his	PHE^{173}
PAB	^{255}asp	GLY	LYS	leu	LYS	GLY	PHE	GLY	PHE	VAL	asn	tyr^{266}	
PAB	^{357}glu	asn	GLY	LYS	SER	LYS	GLY	PHE	GLY	PHE	VAL	cys	PHE^{369}
CONSENSUS SEQUENCE			GLY	LYS	SER	LYS/ ARG	GLY	PHE	GLY	PHE	VAL	X	PHE

Fig. 1. A putative consensus sequence or RNP motif for eukaryotic ssRNA binding proteins. The sequence of A1 is from Cobiachi et al.[77] and Williams et al.[78] The DL4 sequence is from Lahiri and Thomas.[92] PAB is poly(A) binding protein; the PAB sequence is from Sachs et al.[93] and Adams et al.[94]

putative ssRNA-binding domain. The polyadenylate-binding protein (PAB) is associated only with poly(A), not with the rest of the RNA molecule.[1] The yeast gene for the polyadenylate-binding protein has been cloned by two groups.[93,94] The complete cDNA clones were found to have an open reading frame encoding 577 amino acids with several distinct structural domains. The carboxyterminal one-third (bounded by a run of eighth consecutive alanine residues) contains a 60 residue stretch rich in proline and methionine. The N-terminal can be divided into four internally repeated, homologous domains, about 90 amino acids in length.[93,94] All four of the domains contain sequence motifs rich in positively charged and aromatic amino acids. Three of the PAB motifs, approximately 13 amino acids in length, are very homologous to the putative ssRNA-binding site of A1 (Fig. 1). This sequence element or motif has been termed an RNP consensus sequence.[94] The spacing among the motifs is also similar between A1 and PAB: in PAB, the adjacent elements are spaced by 80 to 90 amino acids, whereas in A1, the spacing is approximately 90 amino acids. A schematic diagram of the domain structure and the putative ssRNA-binding sites or motifs for rat A1 and yeast PAB is shown in Figure 2.

More recently, the same RNP motif has been found for the human C proteins, and brine shrimp HD40.[75,88] (The C proteins have only one copy of the RNP motif but appear to have an additional region involved in binding RNA.[75]) Thus, this motif is unusually conserved and may represent a consensus sequence involved in the binding of single-stranded nucleic acids. Although conservation for the motif is consistent with a common ancestral gene for the core proteins, it will be necessary to determine how widespread this motif is among different, and potentially unrelated, nucleic acid binding proteins before the implications of its discovery are clarified. In addition, the sequences flanking the motif are also important and may play a role in modulating binding specificity and affinity.[77,81,92–95]

A Putative domain for protein-protein interactions

The carboxyterminal segment, or C domain, excised by protease from A1, is strikingly different than the N-terminal nucleic acid binding regions. Starting at the site of cleavage, residue 196, the amino acid composition of the C domain is very simple, containing 40% glycine and 40% four other amino acids: serine, asparagine, tyrosine, and phenylalanine.[77,80,81] Furthermore, two of the three dimethylarginines are in this region, the third being in the boundary between the A-B or UP1 domains and the C domain.[78,81] The C domain has no homology with the A and B domains. Along with the slightly hydrophilic nature, the C terminus is predicted to be conformationally flexible, with a high tendency toward random coil or beta-sheet.[75,85]

The C-terminal in A1 contains 16 inexact repeating oligopeptide units, beginning with residue 202.[77] A consensus sequence which itself occurs once is phenylalanine-(glycine)$_4$-serine. This glycine- and serine-rich region exhibits significant homology to the N- and C-termini of keratin. In fact, sequence comparison reveals that the C-terminal region of the rat clone for A1 is very similar to a sequence in a cDNA clone for mouse epidermal keratin I; 92 out of 128 nucleotide residues match.[77] Similarly, a comparison of a human cDNA clone for A1 revealed a partial analogy of the C-terminal region to human keratin type II.[80]

Recently, a nuclear envelope structural protein, nuclear lamin A, which shares structural analogy (for the alpha-helical rod domain) with the intermediate filament proteins, has been found also to have a glycine- and serine-rich stretch in its C-terminus.[96,97] The significance of this unusual amino acid sequence is unclear. The glycine-serine–rich sequence seems unique to keratin proteins[98] and may be specific for the assembly of the filament. It is known that the C-terminal domain is essential for assembly of intermediate filaments; removal of the C-terminus abolishes intermediate filament formation.[99] In the absence of RNA, similar interactions may occur among core proteins, which have been observed to form regular filaments in vitro.[60] Therefore, it is conceivable that the C-terminal domains of the core proteins could be involved in nucleating protein-protein interactions in ribonucleosome assembly. The implication of C-terminal interaction among filamentous polypeptides raises the interesting possibility of RNP polypeptide-lamin interactions. Such interactions may

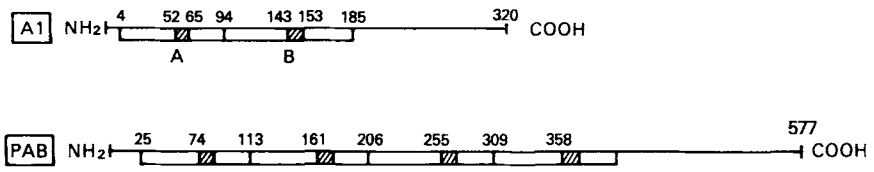

Fig. 2. Putative domain structure for the ssRNA binding proteins associated with pre-mRNA or with poly(A). The RNP motif is shown hatched, and the internally-repeated N-terminal domains are indicated. The A1 data is from Merrill et al.[79] and Cobiachi et al.[77] PAB is poly(A) binding protein; the PAB sequence is from Sachs et al.[93] and Adams et al.[94]

be the basis for the observations on the structural interrelatedness of hnRNPs, nuclear lamins, and the nuclear matrix.[100,101]

The presence of the C-terminal domain alters the properties of A1 compared to UP1.[81,95] A1 binds somewhat better to RNA whereas UP1 has somewhat more affinity for single-stranded DNA. UP1 is a helix-destabilizing protein, whereas pure A1 appears to have no effect on the conformation of polynucleotides. Analogous functional domains for single-stranded DNA binding proteins—namely, HMG-1, T_4 gene 32 and *Escherichia coli* SSB behave similarly.[81,91,102] That is, proteolytic removal of a C-terminal fragment increases the unwinding activity of the truncated protein; the C-terminal domain for these DNA binding proteins has more clearly been shown to be involved in protein-protein interactions. Thus, the amino acid composition and its effects on the properties of the "UP1 region" strongly suggest the C domain is involved in protein-protein interactions. However, it is important to note that the properties of isolated A1 may differ significantly from A1 associated with other core proteins. In ribonucleosomes, A1 interacts with A2, and codissociates with A2 around 0.25 M NaCl.[51,60] In contrast, purified A1 sticks much tighter to nucleic acids.[52] When the C domain is involved in protein-protein interactions, the A and B domains might still act to remove secondary structure from nucleic acids, consistent with helix-destablizing properties observed from bulk core proteins.[103]

Recently, A1 protein has been obtained in large quantities and pure form by expressing the cloned gene in *E. coli*.[101] Proteolytic fragments of A1 containing increasing amounts of the C domain were found to elute at higher salt from single-stranded DNA cellulose; intact A1 eluted last over the range from 0.6 to 0.8 M NaCl. The C-terminal domain of A1 appears to permit tighter binding of the protein to nucleic acids.[101] The difference between these binding studies and the release of A1 from ribonucleosomes in 0.25 M NaCl may arise from the absence of the other core proteins or that RNA, rather than RNP, is the substrate. Since naturally occurring A1, modified in vivo, also requires a higher salt for elution,[52] this effect is not a consequence of producing unmodified A1 protein in *E. coli*.

Additional evidence for the importance of C-terminal interactions in RNP assembly has come from the study of the periodic organization of the complex of cytoplasmic poly(A) binding protein with Poly(A). Nuclease cleaves the cytoplasmic poly(A)-PAB protein complex, but not the nuclear complex, in a regular fashion, every 27 nucleotides.[104] The periodicity of cleavage might be related in part to the multiple binding domains. The nuclear protein is derived from the cytoplasmic form by cleavage of the C-terminal region.[93,105,106] The nuclear form retains the four regions of internal homology, and the capacity to bind poly(A). Cleaving the C-terminus may abolish the protein-protein interactions along the poly(A) strand necessary for establishing the specific placement of the poly(A)-binding protein.[93]

HD40: A model system for ribonucleosomes

The brine shrimp, *Artemia salina*, major hnRNP protein, HD40 (helix-destabilizing protein, molecular weight 40,000) is the best-characterized core protein in terms of its interactions with polyribonucleotides.[83,103,107,108] *Artemia* ribonucleosomes contain predominantly three basic, isoelectric variants of HD40, all post-translational modifications of a simple precursor. As described above, HD40 shares many of the characteristic features of the A and B group of core proteins from higher eukaryotes. HD40 binds preferentially to single-stranded RNA, compared to single-stranded DNA, disrupting any secondary structure. Binding of HD40 occurs in two stages.[108] The first step extends the nucleic acid, while disrupting the secondary structure. At one HD40 protein per 12 to 15 nucleotides, the RNA in the extended nucleoprotein filament is maximally unwound. No further change in secondary structure occurs in the second stage, that is, in the formation of the beaded complex at the saturating limit of one HD40 per eight nucleotides. The first stage of binding appears to involve noncontiguous association. At higher protein concentrations, a slight cooperativity is observed. MS2 phage RNA, hnRNA, ribohomopolymers, and ϕX174 phage DNA are folded into beaded structures. Each bead is estimated to contain about 300 nucleotides. In reconstitution experiments with poly(A), twenty HD40 molecules bind to a poly(A) chain of 160 nucleotides to form a disk about 30 Å high by 180 Å in diameter. Smaller poly(A) chains associate with HD40 to form sectors of this disk: poly$(A)_{40}$-HD40 forms a quarter-disk; poly$(A)_{80}$-HD40 forms a half-disk. Longer poly(A) chains complex with HD40 either to form a second disk or an extended helical cylindrical 180 Å in diameter. Both continuous helices and discrete beads on a string structures are seen. Lateral protein-protein interactions are believed to produce the disk and a marginally-stable helical stacking of disks. RNA is on the periphery of the nucleoprotein structure; thus, 300 nucleotides would make about 1.8 turns of RNA around the bead observed for longer RNA molecules (assuming a simple helix).

Ribonucleosome Architecture

The structural organization of ribonucleosomes is known only in very general terms. We outline the information available on the chemical composition, size and shape, molecular weight, protein interactions, and relative spatial location of the constituents. From a low resolution perspective, it is interesting to compare the properties of ribonucleosomes to two other well characterized nucleoprotein complexes, the

nucleosome and the ribosome. The 10S nucleosome has a copy number of two for each of the four histones, which together form an octamer in the center of the complex. The ribosome has a much larger number of unique polypeptides. The ribonucleosome is comparable in size as well as sedimentation coefficient to the 40S ribosomal subunit, which has a mass of 1.5 million. Thus, the ribonucleosome having a high copy number of a simple protein complement (six major structural polypeptides) can be viewed as having an intermediate structural complexity between these nucleoprotein paradigms.

Chemical composition

Protein composes about 80%–85% of the mass of ribonucleosomes[48–52]; under some conditions the core proteins in turn compose 90% more of the protein mass.[3,60] The length of the RNA chain in a particle is not certain. A very broad and continuous length distribution has always been observed for RNA isolated from ribonucleosomes. Under conditions of limited nuclease digestion, the ribonucleosomes isolated from mammalian sources predominantly contain RNA molecules ranging from 600 to 1,200 nucleotides in length.[49–52] About 900–1,200 nucleotides are packaged into each ribonucleosome that has been reconstituted along defined, long nucleic acid molecules (intact phage ϕX174, MS2, and TMV RNAs).[61,62] Although further nuclease digestion sufficient to reduce the average length below 400 nucleotides can cause the complex to dissociate,[60] under other conditions, ribonucleosomes are still observed to sediment at 40S despite extensive nicking of the RNA. Ribonucleosomes that contain RNA molecules only 100–150 nucleotides in length have been found to be structurally intact[2,59] and ribonucleosomes containing even shorter RNA chains may be stable under some conditions[17,18] (W. LeStourgeon, personal communication). It is not clear if multiple copies of the shorter RNA molecules are required to hold the complex together.

In sum, the length of nucleic acid per particle, which was reliably ascertained for nucleosomes, has been a difficult parameter to obtain for ribonucleosomes. This is presumably related to the less monotonic nature of the structural units along hnRNPs. The requirements for processing RNA molecules varying widely in length, and the heterogeneity observed for ribonucleosomes, suggest that there may not be a canonical length of RNA per ribonucleosome in vivo. Instead, the reconstitution experiments may indicate the upper limit to the amount of RNA that can bind to core proteins to form an individual ribonucleosome.

The mass distribution of ribonucleosomes isolated following limited nuclease digestion is very broad, ranging from 1 to 2 million.[60,109] To obtain a clearer picture, reconstituted ribonucleosome monomers and dimers have been prepared from in vivo transcribed RNA of defined length and sequence (G. Conway and W. LeStourgeon, unpublished). After correcting for the known mass of the RNA, all monomer and dimer reconstituted ribonucleosomes yield a consistent mass of 1.2 to 1.3 million daltons for the protein component of an individual particle (J. Wooley and J. Wall, unpublished.) From the polypeptide stoichiometry and their apparent mass, this implies that there are on the order of 36 polypeptides in the reconstituted ribonucleosomes.

Size and shape

Ribonucleosomes in thin sections and along nuclear spreads are of the order of 200 Å in size, as noted above. In principle, more detailed information about the morphology should be obtainable from isolated ribonucleosomes. It can be difficult to determine the width accurately for small objects such as nucleoprotein complexes, owing to inherent problems with the methods required for obtaining contrast for electron microscopy. This problem is enhanced for ribonucleosomes because of their lability. Unfixed ribonucleosomes are unfolded, amorphous structures.[50] After fixation with aldehydes and negative staining, ribonucleosomes appear more compact, roughly ellipsoidal in profile, about 180 to 240 Å in width.[49,50,109] The morphology is variable and the profiles are not consistent with different views of a single object. The number of different profiles is limited and suggest a finite number of morphologies.[50]

A supramolecular protein complex

Several observations suggested early on that the major proteins could be separated from the RNA in the form of a discrete, supramolecular 30S complex,[47,110] hence the traditional designation "core proteins." However, this RNA-free, protein complex is a result of cross-linking (by free radicals) during the iodination procedure.[50] The iodinated 30S protein complex was found to be extensively cross-linked. Yet, it is quite interesting that RNA could be removed after such extensive cross-linking, and that a RNA-protein complex could again be created simply by adding back RNA. The same RNA dissociation and reassociation can be accomplished on ribonucleosomes exhaustively cross-linked with imidate reagents.[111] This supports the original assertion that RNA is wrapped around a protein core, although the core is not stable in low salt and in the absence of RNA.[50,112] In addition, the observation that a zero-length cross-linking reagent is capable of cross-linking the polypeptides up to a 30S complex indicates the extent of oligometric protein-protein interactions in ribonucleosomes.[50]

Direct evidence for intimate polypeptide associations and oligometric interactions among the core proteins has come from reversible cross-linking studies.[60] Analysis of cross-linking products produced by dithiobispropionimidate (DTBP), which can cross-link primary amino amines within 11 Å of each other,

unambiguously identifies a set of homotrimers of A1, A2, and C2. Trace amounts of hexamers of these major polypeptides are also observed, as are equimolar A1-A2 heterotypic complexes. The zero-length reagent (ortho-phenanthroline)$_2$ CU(II), or CuP, catalyzes the formation of disulfide bridges between existing reduced cysteine residues in the protein. Following CuP treatment, homotrimers of A2 and of C1 can be identified, and equimolar A1-A2 complexes. CuP also cross-links polymers of A2 with B1, up to the heterotypic tetramer $(A2)_3B1$. (The cross-linking data is the best evidence for the polypeptide stoichiometry given in Table I). Thus, it seems likely that A1, A2, and C1 exist as homotypic trimers in ribonucleosomes. Interactions between trimers of A1 and of A2 and between A2 trimers and B1 also occur. Interaction between A1 and A2 had been hypothesized based on their co-elution from ribonucleosomes at low salt.[3,51,52] In addition, a stable tetrameric complex of $(A2)_3$ B1 has been isolated free in solution.[60] In summary, the cross-linking studies reveal a complex pattern of oligomeric polypeptide species. The oligomeric polypeptide interactions may underlie the folding and packaging of nascent hnRNA.

Some information on the topology of ribonucleosomes has been provided by in vitro nuclease and protease digestion studies.[60] Digestion of nRNPs with micrococcal nuclease or pancreatic RNase rapidly releases much of the RNA. A1, C1, and C2 are lost concomitantly. Proteolytic treatment of ribonucleosomes with trypsin or chymotrypsin causes rapid degradation of C1 and B2, followed by A1 and C2. These observations are consistent with the UV cross-linking data indicating intimate contact of RNA and the C proteins, and with the accessibility of C protein within ribonucleosomes to antibody binding.[59,64,65] Thus, it seems likely that the C proteins and RNA, and possibly A1, are located on the outside or periphery of ribonucleosomes.

Secondary structure of RNA in ribonucleosomes

It is well established that purified hnRNA is capable of forming extensive intramolecular or intermolecular duplex regions. However, isolated core proteins possess helix-destabilizing activity.[103,107,108] After excluding the contribution from small nuclear RNAs, 0.4% or less of the RNA of ribonucleosomes has been found to be double-stranded.[50] Reconstitution experiments indicated that MS2 phage RNA, which has extensive intramolecular base pairing, is folded into ribonucleosomes indistinguishable from those formed by single-stranded homopolymers.[61] It would seem likely that the RNA in ribonucleosomes contains little secondary structure. If the core proteins behave like prokaryotic single-stranded nucleic acid binding proteins, such as *E. coli* SSB and T_4 phage gp32,[91] they unfold all secondary structure in partial duplexes except for longer inverted repeats, or inverted repeats with a loop too small for interactions and melting to occur. Consistent with this expectation, core proteins have low affinity for double-stranded polynucleotides;[61] ribonucleosomes are not formed in reconstitution experiments with fully base-paired duplexes;[61] nuclease digestion experiments have suggested that intramolecular duplexes are excluded from ribonucleosomes and are largely protein free.[113]

The core proteins may serve to maintain the sequences of hnRNA accessible to enzymes and factors for processing by destabilizing RNA secondary structure and organizing RNA on the periphery of the complex. Exposed ssRNA may be a requirement for base pairing to small nuclear RNAs and recognition by other sequence-specific factors.[10] The inverted repeats excluded from ribonucleosomes might be recognition sites for eukaryotic enzymes, similar to the role of duplex segments in prokaryotic RNA processing by RNase III.[114] Although there would seem to be little stable base pairing within ribonucleosomes, transient or dynamic roles for RNA secondary structure have been suggested.[115] Helical regions formed immediately after transcription could well influence the placement of the polypeptides on the nascent transcripts, but then be unwound upon particle formation. The effect of intron secondary structure in splice site selection and accuracy of processing is still uncertain.[10]

Transcript-specific interactions within ribonucleosomes

The electron microscopic evidence for gene-specific hnRNP architecture suggests specific placement of proteins along primary transcripts, or, if there is a unitary protein assembly within ribonucleosomes, the phasing or unique placement of RNA along the core protein suprastructure. There is little biochemical evidence to support such specificity.[17,18] Only minimal suggests of specificity has come from nuclease digestion of hnRNP coupled with Southern blot hybridization.[116,117] A higher-resolution approach for RNA mapping has been used to look for specific protection of beta-globin sequences in ribonucleosomes from chicken reticulocytes.[17,18] Discrete nuclease-resistant regions of 20 to 50 nucleotides in length were mapped in both exons and introns. Notably, both introns have a resistant region immediately upstream of the lariat branch site. In general, these RNA mapping experiments do not distinguish among contributions from the different proteins and nucleoproteins, such as the components of the splicing machinery, and the core proteins, that might contribute to nuclease protection (or sensitivity). Analysis of reconstituted ribonucleosomes and of transcripts for other genes are needed to clarify this picture.

A Functional Role for Ribonucleosomal Core Proteins

Direct evidence for a functional role for RNP core proteins, besides compacting RNA and protecting it

from nonspecific nuclease degradation, comes from recent observations with an in vitro splicing system and monoclonal antibodies directly against C1 and C2.[19] The addition of antibodies does not disrupt preexisting ribonucleosomes or prevent the formation of the 60S spliceosome. One of the anti-C-protein antibodies, termed 4F4, inhibits splicing of pre-mRNA at an early step: endonucleolytic cleavage at the 5′ splice site and the formation of the intron lariat. Following the addition of 4F4 to extracts in the absence of ATP, sedimentation studies show that the pre-mRNA is incorporated into a 40S complex. Upon addition of ATP, the complex is converted to the 60S spliceosome but the pre-mRNA remains intact. Nonimmune sera and boiled 4F4 have no effect on splicing. Also, another monoclonal antibody, termed 2B12, that recognizes different epitopes on the C proteins[73] does not inhibit splicing. Splicing of exogenous pre-mRNA is also inhibited in extracts from which C protein has been previously depleted. Quantitative analysis of the splicing products suggests that the removal of C proteins severely effects the second cleavage step and ligation as well as inhibiting the first endonucleolytic cleavage of the 5′ splice site. The 60S complex or spliceosome does not form in C-depleted extracts, although the RNA is converted into 30S material, presumably some form of a ribonucleoprotein complex. This study clearly demonstrates the essential role of the C proteins in splicing. It is not yet known if the C proteins participate directly in splicing or whether an intact ribonucleosome is required, perhaps as a precursor for the formation of the splicesome.

Additional evidence for the involvement of core proteins in the splicing process comes from the observation that polyclonal anti-RNP antibodies block the in vitro splicing of a truncated beta-globin pre-mRNA.[119] The polyclonal antibodies employed recognize common or conserved regions among the core polypeptides, reacting strongly with A1 and A2 but also to a lesser extent with B1, B2, C1, and C2.[92] Addition of low levels of anti-RNP to the extract prior to incubation was shown to block the second step, cleavage at the 3′ splice site and ligation, but higher levels of antibody were required for inhibition of the first step. Since addition of antibodies differs in effect on the two steps in the reaction, it seems likely that the inhibition directly involves interference in the splicing process, not simply removal of the transcript from the reaction by precipitation. It is not clear if the reaction is inhibited largely as a result of antibody reaction with C1 and C2, or if the effect is also due to antibody binding to other core proteins. This study further emphasizes the distinction between the two known steps in splicing; presumably, the ligation step is more dependent on the integrity of ribonucleosomes, which would maintain the spatial proximity of the ends of the cleaved RNA intermediates.

Higher-Order Structure of hnRNPs

The unique nature of hnRNPs extends to the higher-order structure of primary transcripts.[120,121] One case has been analyzed in detail: the giant RNA transcription sites or puffs in the insect *Chironomus* known as the Balbiani rings (BR). The products of the BR gene are exceptionally large salivary proteins on the order of 10^6 daltons. The 75S RNA molecules are transcribed in the BR genes and packaged into a globular 500-Å hnRNP.[121–124] The folded hnRNP structure for the BR transcript changes during translocation across the nuclear envelope into an extended, rod-like structure seen on the cytoplasmic side of the nuclear pores.[125] The growing hnRNP looks like a mushroom with a stalk and a dense cap; the size of the "cap" increases and reaches a roughly 500 Å spherical shape when fully mature; in two dimensional profiles the particles look like an almost complete donut. A detailed (85–90 Å resolution) reconstructed map of the mature hnRNP using electron microscopic tomography shows that this ring-shaped configuration is asymmetric with four domains.[121] Individual BR hnRNPs are quite similar to each other, and to the average reconstructed structure. On the other hand, there is no obvious relationship between this structure and ribonucleosomes or other nuclear RNA structures visualized in thin sections or in nuclear spreads. The BR hnRNP structure highlights the unique supramolecular nature of nascent RNP in vivo.

CONCLUSIONS

Rapid progress has been made in the last year in our understanding of the ribonucleosomal protein components and their involvement in RNA splicing. The core proteins are highly conserved in mammals. The degree of conservation now established was unexpected and suggests that the proteins are involved in multiple component interactions. In addition, the core proteins have distinct internal domains; the N terminal section of A1 contains two nucleic acid binding domains. It is likely that multiple nucleic binding domains are a general feature of the A and B core proteins. An RNP consensus sequence or motif has been identified within this domain; this motif is conserved among different types of RNA-binding proteins from yeast to man. The glycine-rich C-terminal segment, which is very different in predicted secondary structure and shares extensive sequence analogy with the keratins, may be essential for the interactions of core proteins.

The core proteins are organized in an oligomeric assembly in ribonucleosomes. C1 and C2, previously shown to be intimately associated with RNA, appear to be peripherally located. All of the polypeptides are likely to be closely associated with each other. The set of protein interactions within ribonucleosomes includes homotrimers of A1, A2, and C1. A2 must also

interact with A1 and with B1. The heterotypic assembly of an A2 trimer with one molecule of B1, detected in ribonucleosomes, can be isolated as a heterotypic tetramer free in solution. The demonstration of these homotypic and heterotypic interactions strongly supports the stoichiometry proposed for the core proteins: 3:3:1:1:3:1 for A1,A2,B1,B2,C1,C2, respectively.

The C proteins have been shown to be an obligatory for splicing and to be a component of the spliceosome. C proteins are required for formation of the spliceosome, so their role is not limited to holding the excised intermediates in a nucleoprotein matrix. They may, however, bring the ends together, that is, participate in a dynamic or intramolecular reassembly of the spliceosome necessary for the ligation step.

Many different experiments suggest a significant amount of diversity for ribonucleosomes and gene-specific hnRNP organization. At the least, different stretches of RNA, or the same stretch during different steps in processing, must lie in ribonucleosomes of alternate conformation. Conformational flexibility in polypeptide assembly, like that for spherical viruses,[126] is likely to play a role. It is also possible that there is more than one class of particulate units, differing in specific polypeptide conformation. The underlying basis for this structural heterogeneity is unknown. Introns and exons in different genes vary greatly in length and this may well be a major constraint on RNP architecture. A number of observations suggest that discontinuities in RNP architecture are introduced at exon-intron junctions by snRNPs. Specific proteins, regions of residual RNA secondary structure or pauses in transcription, changing the details of assembly, could also lead to diversity in RNP architecture.

In contrast, some aspects of hnRNPs are consistent with a regular structure, like the nucleosomal paradigm. The term ribonucleosome has been introduced both for clarity and to reflect the evidence for a regular structural element. The simple, fixed protein stoichiometry, the high copy number of the individual polypeptides, their similar sequences and structural properties, and the contiguous uniform arrays of particles on nascent transcripts observed under mild spreading conditions are all consistent with an oligomeric protein unit, around which RNA is wrapped. Although the RNA is likely to be coiled or folded in a continuous fashion around the outside of the particle, the data as a whole suggest to us that ribonucleosomes have a large number of discrete RNA binding sites. This organization would facilitate rearrangement of the RNA and protein components during splicing. The distribution of sites might be related to a lower limit or minimum length for introns and exons.[10] The need to accommodate the diversity of RNA exons and introns may in fact underlie the high polypeptide copy number and thus the large number of RNA binding sites, only a few of which may be occupied in a given ribonucleosome. A broad range for the length of RNA per particle would be expected. There may also be sites for the potential association of snRNPs and other factors.

There is a dichotomy between the evidence for the diverse architecture of hnRNPs and the simple set of highly conserved core proteins. Further insight into the features of hnRNPs will undoubtedly come soon from studies on clones corresponding to other core proteins as well as on the role of the A and B group core proteins in splicing. It is also essential to relate the electron microscopic observations more precisely to the biochemical details of ribonucleosomes. Nonetheless, after 20 years of research, studies on hnRNPs are now beginning to uncover details about their structure and function. Ribonucleosomes probably contain varying length of RNA, and must be capable of dynamic, intraparticle processing, making them quite different from nucleosomes and ribosomes. Internally, the RNP core proteins are structurally distinct, with different domains. The supramolecular complex they form with RNA is an enigma. Elucidating the role of the core proteins, the structure of ribonucleosomes, and their relationship to splicesomes promises to be quite exciting.

ACKNOWLEDGEMENTS

We thank our colleagues for providing reprints, preprints, and unpublished data. The comments and suggestions of G. Dreyfus, S. Harrison, T. Henry, D. Hill, R. Kornberg, W. LeStourgeon, T. Martin, H. Nash, J. Thomas, K. Williams, S. Wilson, and E. Zapolski are gratefully acknowledged.

REFERENCES

1. Martin, T.E., Pullman, J.M., McMullen, M.D. Structure and function of nuclear and cytoplasmic ribonucleoprotein complexes. In: "Cell Biology: A Comprehensive Treatise, Vol. 4." Prescott, D.M., Goldstein, L., eds. New York: Academic Press. 1980:137–174.
2. Georgiev, G.P., Samarina, O.P. D-RNA containing ribonucleoprotein particles. Adv. Cell Biol. 2:47–110, 1971.
3. LeStourgeon, W.M., Lothstein, L., Walker, B.W., Beyer, A.L. The composition and general topology of RNA and protein in monomer 40S ribonucleoprotein particles. In: "Nuclear Particles," Part A. Busch, H., ed. New York: Academic Press. 1981:49–87.
4. Sommerville, J. Nascent RNP. In: "The Cell Nucleus, Vol X, Nuclear Particles," Part A. Busch, H., ed. New York: Academic Press, 1981.
5. Knowler, J.T. An assessment of the evidence for the role of ribonucleoprotein particles in the maturation of eukaroyte mRNA. Int. Rev. Cytol. 84:103–153, 1983.
6. Pederson, T. Nuclear RNA-protein interactions and messenger RNA processing. J. Cell Biol. 97:1321–1326, 1983.
7. Busch, H., Reddy, R., Rothblum, L., Choi Y.C. SnRNAs, SnRNPs, and RNA processing. Ann. Rev. Biochem. 51:617–654, 1982.
8. Nevins, J.R. The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem. 52:441–466, 1983.
9. Leff, S.E., Rosenfeld, M.G., Evans, R.M. Complex transcriptional units: Diversity in gene expression by alternative RNA processing. Ann. Rev. Biochem. 55:1091–1117, 1986.
10. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., Sharp, P.A. Splicing of messenger RNA precursors. Ann. Rev. Biochem. 55:1119–1150, 1986.
11. McKnight, S.L., Miller, O.L. Ultrastructural patterns of RNA synthesis during early embryogenesis of *Drosophila melanogaster*. Cell 8:305–319, 1976.
12. Martin, T.E., McCarthy, B.J. Synthesis and turnover of RNA in the 30S nuclear ribonucleoprotein complexes of

mouse ascites cells. Biochem. Biophys. Acta 277:354–367.
13. Laird, C.D., Chooi, W.Y. Morphology of transcription units in *Drosophila melanogaster*. Chromosoma 58:193–218, 1976.
14. Beyer, A.L., Bouton, A.H., Miller, O.L., Jr. Correlation of hnRNP structure and nascent transcript cleavage. Cell 26:155–165, 1981.
15. Osheim, Y., Miller, O.L., Jr., Beyer, A. RNP particles at splice junction sequences on Drosophila chorion transcripts. Cell 43:143–151, 1985.
16. Gallinaro, H., Sittler, A., Jacob, M. In vivo splicing of the premRNAs from early region 3 of adenovirus-2: Association of precursors, intermediates and products with hnRNP. Nucleic Acids Res. 14:4171–4185, 1986.
17. Patton, J., Chae, C.B. Specific regions of B-globin RNA are resistant to nuclease digestion in RNA-protein complexes in chicken reticulocyte nuclei. Mol. Cell. Biol. 5:1220–1228, 1985.
18. Patton, J.R., Chae, C.B. Specific regions of the intervening sequences of B-globin RNA resistant to nuclease in 50S heterogeneous nuclear RNA-protein complex. Proc. Natl. Acad. Sci. USA 82:8414–8418, 1985.
19. Choi, Y.D., Grabowski, Sharp, P.J., Dreyfuss, G. Heterogeneous nuclear ribonucleoproteins: Role in RNA splicing. Science 231:1534–1539.
20. Brody, E., Abelson, J. The "spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. Science 228:963–967, 1985.
21. Grabowski, P.J., Seiler, S.R., Sharp, P.A. A multicomponent complex is involved in the splicing of messenger RNA precursors. Cell 42:345–353, 1985.
22. Frendewey, D., Keller, W. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell 42:355–367, 1985.
23. Perkins, K.K., Furneaux, H.M., Hurwitz, J. RNA splicing products formed with isolated fractions from HeLa cells associated with fast-sedimenting complexes. Proc. Natl. Acad. Sci. USA 83:887–891, 1986.
24. Berget, S., Robberson, B. U1, U2 and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. Cell 46:691–696, 1986.
25. Black, D., Steitz, J. Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. Cell 46:697–704, 1986.
26. Kaltwasser, G., Spitzer, S.G., Goldenberg, C.J. Assembly in vitro of a mouse insulin messenger RNA into a 60-40S ribonucleoprotein complex. Nucleic Acids Res. 14:3687–3701, 1986.
27. Bindereif, A., Green, M. Ribonucleoprotein complex formation during pre-mRNA splicing in vitro. Mol. Cell. Biol. 6:2582–2592, 1986.
28. Grabowski, P.J., Sharp, P.A. Affinity chromatography of splicing complexes: U2, U5, and U4 and U6 small nuclear ribonucleoprotein particles in the spliceosome. Science 233:1294–1299, 1986.
29. Konarska, M.M., Sharp, P.A. Electrophoretic separation of complexes involved in the splicing of precursors to mRNA. Cell 46:845–855, 1986.
30. Miller, O.L., Jr., Bakken, A.H. Morphological studies of transcription. Acta Endocrinol. Suppl. 168:155–177, 1972.
31. Malcolm, D.B., Sommerville, J. The structure of chromosome-derived ribonucleoprotein in oocytes of *Triturus* cristatus carnifex (Laurenti). Chromosoma 48:137–158, 1974.
32. Puvion, E., Moyne, G. *In situ* localization of RNA structures. In: "The Cell Nucleus," Vol 8. Busch, H., ed. New York: Academic Press. 1981:59–155.
33. Fakan, S., Leser, G., Martin, T. Ultrastructural distribution of nuclear ribonucleoproteins as visualizedby immunocytochemistry on thin sections. J. Cell Biol. 98:358–363, 1984.
34. Beyer, A.L., Miller, O.L., McKnight, S.L. Ribonucleoprotein structure in nascent hnRNA is nonrandom and sequence-dependent. Cell 20:75–84, 1980.
35. Beyer, A.L., Bouton, A.H., Hodge, L.D., Miller, O.L., Jr. Visualization of the major late R strand transcription unit of adenovirus serotype 2. J. Mol. Biol. *147*:269–295, 1981.
36. Beyer, A. Ultrastructural analysis of the ribonucleoprotein structure of nascent hnRNA. Mol. Biol. Rep. 9:49–58, 1983.
37. Glatzer, K.H. Visualization of gene transcription in spermatocytes of *Drosophila hydei*. Chromosoma 53:371–379, 1975.
38. Glatzer, K.H. Regular substructures within homologous transcripts of spread *Drosophila* germ cells. Exp. Cell Res. 125:519–523, 1980.
39. Glatzer, K.H., Meyer, G.F. Morphological aspects of the genetic activity in primary spermatocyte nuclei of Drosophila hydei. Biol. Cell 41:165–172, 1981.
40. Foe, V.E., Wilkinson, L.E., Laird, C.D. Comparative organization of active transcription units on Oncopeltus fasciatus. Cell 9:131–146, 1976.
41. Spradling, A.C. The organization and amplification of two chromosomal domains containing *Drosophila* chorion genes. Cell 27:193–209, 1981.
42. Osheim, Y.N., Miller, O.L., Jr., Novel amplification and transcriptional activity of chorion genes in *Drosophila melanogaster* follicle cells. Cell 33:543–553, 1983.
43. Fakan, S., Leser, G., Martin, T.E. Immunoelectron microscopic visualization of nuclear ribonucleoprotein antigens within spread transcription complexes. J. Cell Biol. 103:1153–1158, 1986.
44. Spector, D.L., Schrier, W.H., Busch, H. Immunoelectron microscopic localization of snRNPs. Biol. Cell 49:1–10, 1983.
45. Breitbart, R.E., Nguyen, H.T., Medford, R.M., Destree, A.I., Mahdavi, V., Nadal-Genard, B. Intricate combinatorial pathways of exon splicing generate multiple regulated troponin T isoforms from a single gene. Cell 41:67–82, 1985.
46. Osheim, Y.N., Beyer, A.L. Nascent ribonucleoprotein structure of polymerase I, II, and III gene transcripts. In: "Nuclear Envelope Structure and RNA Maturation." UCLA Symposia on Molecular and Cellular Biology, Vol 26. Smuckler, E.A., Clawson, G.A., eds. New York: Alan R. Liss, 1985.
47. Samarina, O.P., Molnar, J., Lukanidin, E.M., Bruskov, V.I., Krichevskaya, A.A., Georgiev, G.P. Reversible dissociation of nuclear ribonucleoprotein particles containing mRNA into RNA and protein. J. Mol. Biol. 27:189–191, 1967.
48. Samarina, O.P., Lukanidin, E.M., Molman, J., Georgiev, G.P. Structural organization of nuclear complexes containing DNA-like RNA. J. Mol. Biol. 33:251–263, 1968.
49. Martin, T.E., Billings, P., Levey A., Ozarsian, S., Quinlan, I., Swift, H., Urbas, L. Some properties of RNA: Protein complexes from the nucleus of eukaryotic cells. In: "Chromosome Structure and Function." Cold Spring Harbor, New York: The Cold Spring Harbor Laboratory. 1974:921–932.
50. Martin, T.E., Billings, P.B., Pullman, J.M., Stevens, B.J., Kinniburgh, A.J. Substructures of nuclear ribonucleoprotein complexes. In: "Chromatin." Cold Spring Harbor Symposia Vol. XLII. Cold Spring Harbor, New York: The Cold Spring Harbor Laboratory. 1978:899–909.
51. Beyer, A.L., Christensen, M.E., Walker, B.M., LeStourgeon, W.M. Identification and characterization of the packaging proteins of core 40S hnRNP particles. Cell 11:127–138, 1977.
52. LeStourgeon, W.M., Beyer, A.L., Christensen, M.E., Walker, B.W., Poupore, S.M., Daniels, L.P. The packaging proteins of core hnRNP particles and the maintenance of proliferative cell states. In: "Chromatin." Cold Spring Harbor Symposia Vol. XLII. Cold Spring Harbor, New York: The Cold Spring Harbor Laboratory. 1983:885–898.
53. Karn, J., Vidali, G., Boffa, L.C., Allfrey, V.G. Characterization of the non-histone nuclear proteins associated with rapidly labeled heterogenous nuclear RNA. J. Biol. Chem. 252:7307–7322, 1977.
54. Stevenin, J.H, Gallinaro-Matringe, R., Jacob, M. Complexity of the structure of particles containing heterogeneous nuclear RNA as demonstrated by ribonuclease treatment. Eur. J. Biochem. 74:589–602, 1977.
55. Gattoni, R., Stevenin, J., Devilliers, G., Jacob, M. Size heterogeneity of monoparticles from nuclear ribonucleoproteins containing premessenger RNA. FEBS Lett. 90:318–323, 1978.
56. Pederson, T. Proteins associated with heterogeneous nuclear RNA in eukaryotic cells. J. Mol. Biol. 83:163–183, 1974.
57. Mayrand, S., Setyono, B., Greenberg, J.R., Pederson, T. Structure of nuclear ribonucleoprotein: Identification of proteins in contact with Poly(A)$^+$ heterogeneous RNA in living HeLa cells. J. Cell Biol. 90:380–384, 1981.
58. Economides, I.V., Pederson, T. Structure of nuclear ribon-

ucleoprotein: Heterogeneous nuclear RNA is complexed with a major sextet of proteins in vivo. Proc. Natl. Acad. Sci. USA 80:1599–1602, 1983.
59. Choi, Y.D., Dreyfuss, G. Isolation of the heterogeneous nuclear RNA ribonucleoprotein complex: A unique supramolecular assembly. Proc. Natl. Acad. Sci. USA 81:7471–7475, 1984.
60. Lothstein, L., Arenstorf, H.P., Chung, S.Y., Walker, B.W., Wooley, J.C., LeStourgeon, W.M. General organization of protein in HeLa 40S nuclear ribonucleoprotein particles. Cell Biol. 100:1570–1581, 1985.
61. Wilk, H.E., Angeli, G., Schafer, K.P. In vitro reconstitution of 35S ribonucleoprotein complexes. Biochemistry 22:4592–4600, 1983.
62. Pullman, J.M., Martin, T.E. Reconstitution of nucleoprotein complexes with mammalian heterogenous nuclear nuclear ribonucleoprotein (hnRNP) core proteins. Cell Biol. 97:99–111, 1983.
63. Setyono, B., Greenberg, J.R. Proteins associated with Poly(A) and other regions of mRNA and hnRNA molecules as investigated by crosslinking. Cell 24:775–783, 1981.
64. Van Eekelen, C.A., Rieman, T., van Venrooij, W.J. Specificity in the interaction of hnRNA and mRNA with proteins as revealed by in vivo crosslinking. FEBS Lett. 130:223–226, 1981.
65. Dreyfuss, G.Y., Choi, Y.D., Adam, S.A. Characterization of heterogenous nuclear RNA-protein complexes *in vivo* with monoclonal antibodies. Mol. Cell. Biol. 4:1104–1114, 1984.
66. Isenberg, I., Histones. Ann. Rev. Biochem. 48:159–191, 1979.
67. Wilk, H.E., Werr, H., Friendrich, D., Kiltz, H.H., Schafer, K.P. The core proteins of 35S hnRNP complexes. Eur. J. Biochem. 146:71–81, 1985.
68. Holcomb, E.R., Friedman, D.L. Phosphorylation of the C proteins of HeLa cell hnRNP particles. J. Biol. Chem. 259:31–40, 1984.
69. Celis, J.E., Bravo, R., Arenstorf, H.P., LeStourgeon, W.M., Identification of proliferation-sensitive human proteins amongst components of the 40S hnRNP particles. FEBS Lett. 194:101–109, 1986.
70. Fuchs, J.P., Judes, C., Jacob, M. Characterization of glycine-rich proteins from the ribonucleoproteins containing heterogeneous nuclear ribonucleic acid. Biochem. 19:1087–1097, 1980.
71. Leser, G., Escara-Wilke, J., Martin, T. Monoclonal antibodies to hetergeneous nuclear RNA-protein complexes. J. Biol. Chem. 259:1827–1833, 1984.
72. Leser, G., Martin, T. The major protein components of hnRNP complexes. In: "DNA-Protein Interactions and Gene Regulation." Thompson, E.B., Papaconstantinous, J., eds. Univ. of Texas Press, in press.
73. Choi, Y.D., Dreyfuss, G. Monoclonal antibody characterization of C proteins of heterogenous nuclear ribonucleoprotein complexes in vertebrate cells. J. Cell Biol. 99:1997–2004, 1984.
74. Nakagawa, T.Y., Swanson, M.S., Wold, B.J., Dreyfuss, G. cDNA cloning of the nuclear RNP particle C proteins: A conserved gene family. Proc. Natl. Acad. Sci. USA 83:2007–2011, 1986.
75. Swanson, M.S., Nakagawa, T., Levan, K., Dreyfus, G. The complete primary structure of the human hnRNP C protein: C1 and C2 are produced from a single mRNA. Mol. Cell Biol. (in press).
76. Jones, R.E., Okamura, C.S., Martin, T.E. Immunofluorescent localization of the proteins of nuclear ribonucleoprotein complexes. J. Cell Biol. 86:235–243, 1980.
77. Cobianchi, F., Sen Gupta, D.N., Zmudzka, B.Z., Wilson, S.H. Structure of rodent helix-stabilizing protein revealed by cDNA cloning. J. Biol. Chem. 261:3536–3543, 1986.
78. Williams, K., Stone, K., LoPresti, M., Merrill, B., Planck, S. Amino acid sequence of the UP1 calf thymus helix-destabilizing protein and its homology to an analogous protein from mouse myeloma. Proc. Natl. Acad. Sci. USA 82:5666–5670, 1985.
79. Merrill, B.M., LoPresti, M.B., Stone, K.L., Williams, K.R. High pressure liquid chromatography purification of UP1 and UP2, two related single-stranded nucleic acid-binding proteins from calf-thymus. J. Biol. Chem. 261:878–883, 1986.
80. Riva, S., Morandi, C., Tsoulfus, P., Pandolfo, M., Biamonti, G., Merrill, B., Williams, K.R., Multhany, G., Bayreuther, K., Werr, H., Henrich, B., Schafer, K.P. Mammalian single-stranded DNA binding protein UP1 is derived from the hnRNP core protein A1. EMBO J. 5:2267–2274, 1986.
81. Kumar, A., Williams, K.R., Szer, W. Purification and domain structure of core hnRNP proteins A1 and A2, and their relation to ssDNA binding proteins. J. Biol. Chem. (in press).
82. Christensen, M.E., Beyer, A.L., Walker, B., LeStourgeon, W.M. Identification of N^G,N^G-Dimethylarginine in a nuclear protein from the lower eukaryote *physarum polycephalum* homologous to the major proteins of mammalian 40s ribonucleoprotein particles. Biochem. Biophys. Res. Commun. 74:621–629, 1977.
83. Nowak, L., Marvil, D.K., Thomas, J.O., Boublik, M., Szer, W. A single-stranded nucleic acid-binding from artemia salina. J. Biol. Chem. 255:6473–6478, 1980.
84. Cruz-Alvarez, M., Szer, W., Pellicer, A. Cloning of cDNA sequences for an *artemia salina* hnRNP protein: Evidence for conservation through evolution. Nucleic Acids Res. 13:3917–3930.
85. Wilson, S.H., Cobianchi, Guy, H.R., cDNA cloning and structure-function relationships of a mammalian helix destabilizing protein: hnRNP particle core protein A1. In: "DNA: Protein Interaction and Gene Regulation." Thompson, E.B., Papaconstantinou, J., eds. Univ. Texas Press, 1986: in press.
86. Valenti, O., Biamonti, G., Pandolfo, M., Morandi, C., Riva, S. Mammalian single-stranded DNA binding proteins and heterogenous nuclear RNA proteins have common antigenic determinants. Nucleic Acids Res. 13:337–346, 1985.
87. Pandolfo, M., Valenti, O., Biamonti, G., Morandi, C., Riva, S. Single stranded DNA binding process derive from hnRNP proteins by proteolysis in mammalian cells. Nucleic Acids Res. 13:6577–6590.
88. Merrill, B., Stone, K.L., Kumar, A., Szer, W., Williams, K. The A1 hnRNP domain structure is conserved in many eukaryotic RNA binding proteins. In preparation.
89. Herrick, G., Alberts, B. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. J. Biol. Chem. 251:2124–2132.
90. Planck, S.R., Wilson, S.H. Studies on the structure of mouse helix-destabilizing protein. 1. DNA-binding and controlled proteolysis with trypsin. J. Biol. Chem. 255:11547–11546, 1980.
91. Chase, J.W., Williams, K.R. Single-stranded DNA binding proteins required for DNA replication. Ann Rev. Biochem. 55:103–136, 1986.
92. Lahiri, D.K., Thomas, J.O., A cDNA clone of the hnRNP C proteins and its homology with the single-stranded DNA binding protein UP2. Nucleic Acids Res. 14:4077–4094, 1986.
93. Sachs, A., Bond, M., Kornberg, R. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding protein: Domain structure and expression. Cell 45:827–835, 1986.
94. Adam, S., Nakagawa, T., Swanson, M., Woodruff, T., Dreyfuss, G. The mRNA polyadenylate-binding protein: Gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol. Cell Biol. 6:2932–2943, 1986.
95. Cobianchi, F., Notario, V., Williams, K.R., Wilson, S.H. Overexpression of mammalian hnRNP complex protein A*1* in *E. coli.* Manuscript in preparation.
96. McKeon, F.D., Kirschner, M.W., Caput, D. Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature 319:463–468, 1986.
97. Fisher, D.Z., Chaudhary, N., Blobel, G. The cDNA sequencing of nuclear lamins A and C reveal primary and secondary structural homology to intermediate filament proteins. Proc. Natl. Acad. Sci. USA 83:6450–6454, 1986.
98. Fuchs, E., Hanukoglu, J. Unraveling the structure of the intermediate filaments. Cell *34*:332–334, 1983.
99. Steinert, P.M., Rice, R.H., Roop, D.R., Tras, B.L., Steven, A.C. Complete amino acid sequence of a mouse keratin subunit and implications for the structure of intermediate filaments. Nature 302:794–800, 1984.
100. van Eekelen, C.A.G., van Venrooij, W.J. hnRNA and its attachment to a nuclear protein matrix. J. Cell. Biol. 88:554–563, 1981.
101. Fey, E.G., Krochmalnic, G., Penman, S. The nonchromatin substructures of the nucleus: The ribonucleoprotein (RNP)-

containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. J. Cell. Biol. 102:1654–1665, 1986.
102. Reeck, G.R., Teller, D.C. HMG proteins. In: "Progress in Nonhistone Proteins Research," Vol. 2. Bekhor, I, ed. Boca Raton, Florida: CRC Press, in press.
103. Raziuddin, Thomas, J.O., Szer, W. Nucleric acid binding properties of major proteins from the heterogeneous nuclear ribonucleoproteins of wheat. Nucleic Acids Res. 10:7777–7789, 1982.
104. Baer, B.W., Kornberg, R.D. Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. Proc. Natl. Acad. Sci. USA 77:1890–1892, 1980.
105. Baer, B.W., Kornberg, R.D. The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. J. Cell. Biol. 96:717–721, 1983.
106. Sachs, A.B., Kornberg, R.D. Nuclear polyadenylate-binding protein. Mol. Cell. Biol. 5:1993–1996, 1985.
107. Thomas, J.O., Raziuddin, Sobata, A., Boublik, M., Szer, W. A RNA helix-destabilizing protein is a major component of artemia salina nuclear ribonucleoproteins. Proc. Natl. Acad. Sci. USA 78:2888–2892, 1981.
108. Thomas, J.O., Glowack, S.K., Szer, W. Structure of complexes between a major protein of heterogeneous nuclear ribonucleoprotein particles and polyribonulceotides. J. Mol. Biol. 171:439–455, 1983.
109. Wooley, J., Chung, S.Y., Wall, J., LeStourgeon, W. Architecture of pre-messenger, nuclear ribonucleoprotein monoparticles. Biophys. J. 49:17–19, 1986.
110. Lukanidin, E.M., Zalmanzon, E.S., Komaromi, L., Samarina, O.P., Georgiev, G.P. Structure and function of informofers. Nature 238:193–197, 1972.
111. Prosvirnin, V., Ruzidic, S., Samarina, O. Cross-linked informofers. Nucleic Acids Res. 7:1649–1661, 1979.
112. Kulguskin, V.V., Krichevskaya, A.A., Lukandin, E.M., Georgiev, G.P. Studies on dissociation and reconstitution of nuclear 30s ribonucleoprotein particles containing pre-mRNA. Biochim. Biophys. Acta 609:410–424, 1980.
113. Calvet, J.P., Pedersen, J. Nucleoprotein organization of inverted repeat DNA transcripts in heterogeneous nuclear RNA ribonucleoprotein particles in HeLa cells. J. Mol. Biol. 122:361–378, 1978.
114. Robertson, H.D., Dickson, E., Dunn, J.J. A nucleotide sequence from a ribonuclease III processing site in bacteriophage T7 RNA. Proc. Natl. Acad. Sci. USA 74:822–826, 1977.
115. Paoletti, J. Rech, J., Brunel, C., Jenateur, P. Configuration of double-stranded ribonucleic acid in hela hnRNP and its relaxation by ribonuclease D. Biochemistry 19:5223–5228, 1980.
116. Steitz, J.A., Kamen, R. Arrangement of 30s heterogeneous nuclear ribonucleoprotein on polyoma virus late nuclear transcripts. Mol. Cell. Biol. 1:21–34, 1981.
117. Ohlsson, R.I., van Eekelen, C., Philipson, L. Non-random localization of ribonucleoprotein (RNP) structures within an adenovirus mRNA precursor. Nucleic Acids Res. 10:3053–3068, 1982.
118. Patton, J.A., Chae, C.B. A method for mapping RNA initiation, termination, splice, and protein binding sites: Ribosome binding sites on B-globin messenger RNA. J. Biol. Chem. 258:3991–3995, 1983.
119. Sievakowska, H., Szer, W., Furdon, P.J., Kole, R. Antibodies to hnRNP core proteins inhibit *in vitro* splicing of human B-globin pre-mRNA. Nucleic Acids Res. 14:5241–5254, 1986.
120. Sperling, R., Sperling, J., Levine, A., Spann, P., Stark, G., Kornberg, D. Abundant nuclear ribonucleoprotein form of CAD RNA. Mol. Cell. Biol. 5:569–575, 1985.
121. Skoglund, U., Andersson, K., Strandberg, B., Daneholt, B. Three-dimensional structure of a specific pre-messenger RNP particle established by electron microscope tomography. Nature 319:560–564, 1986.
122. Lamb, M.M., Daneholt, B. Characterization of active transcription units in balbani rings of chironomus tentans. Cell 17:835–848, 1979.
123. Skoglund, U., Andersson, K., Bjorkroth, B., Lamb, M., Daneholt, B. Visualization of the formation and transport of a specific hnRNP particle. Cell 34:847–855, 1983.
124. Olins, D.E., Olins, A.L., Levy, H.A., Durfee, R.C., Margle, S.M., Tinnel, E.P., Dover, S.D. Electron microscope tomography: transcription in three dimensions. Science 220:498–500, 1983.
125. Stevens, B.J., Swift, H. RNA transport from nucleus to cytoplasm in chironomus salivary glands. J. Cell Biol. 31:55–78, 1966.
126. Harrison, S.C. Multiple modes of subunit association in the structure of simple spherical virsuses. Trends Biochem. Sci. 9:345–351, 1984.