Cell, Vol. 67, 1041–1046, December 20, 1991, Copyright © 1991 by Cell Press

RNA-Protein Interactions

Meeting Review

Alan D. Frankel,* Iain W. Mattaj,† and Donald C. Rio*‡
*Whitehead Institute for Biomedical Research
Cambridge, Massachusetts 02142
†European Molecular Biology Laboratory
6900 Heidelberg, Germany
‡Department of Biology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

The interaction of RNA and protein is critical to many biological processes. At a recent meeting on RNA–protein interactions (Workshop on RNA–Protein Interactions, Urbino, Italy, September 22–27, 1991; sponsored by the Human Frontier Science Program and organized by M. Wickens, O. Uhlenbeck, G. Tocchini-Valentini, and W. Keller), genetic, biochemical, and structural studies emphasized recent progress in understanding how proteins recognize specific RNA sequences. Several important themes emerged: RNA structure and dynamics play key roles in RNA–protein recognition, protein–protein and RNA–RNA interactions provide significant specificity in complex ribonucleoproteins, and specific RNA–protein interactions occur in a wide range of biologically important systems. A previous Workshop on RNA–Protein Interactions was reviewed by Wickens and Dahlberg (1987).

RNA-Protein Recognition: Structures of RNA-Binding Proteins

One of the most important recent advances in RNA–protein recognition has been the solution of several three-dimensional structures of RNA-binding proteins. The X-ray crystal structures of two tRNA synthetase–tRNA complexes and the crystal structure and NMR structure of a common RNA-binding domain have been solved. Detailed analyses of these structures are beginning to reveal which interactions are important for specific recognition of RNA.

In the case of E. coli glutaminyl-tRNA synthetase (T. Steitz, Yale University), an extensive set of interactions between protein and RNA allows discrimination between different tRNAs. The most important interactions appear to be base-specific hydrogen bonds in the RNA minor groove, fitting of single-stranded bases into hydrophobic binding pockets of the protein, and conformational changes in the tRNA acceptor stem and anticodon loop that allow specific interactions to form. Some interactions occur with non-Watson-Crick base pairs, which are only formed after the protein has bound. These non-Watson-Crick base pairs, as well as specific hydrogen bonds between bases and the protein, may be mediated through intervening water molecules. Contacts with the RNA backbone are extensive. Chemical footprinting experiments with aspartyl-tRNA synthetase (Giege, CNRS, Strasbourg; Eckstein, Max Planck Institute, Göttingen) also suggest that backbone contacts and conformational changes in the tRNA play important roles in discrimination. The location and number of identity elements varies between tRNAs. For example, experiments in which specificity determinants were transplanted between chargeable tRNA minihelices suggest that the anticodon bases in alanine tRNA contribute less to discrimination than does the acceptor stem (Schimmel, MIT); the opposite is true for valine tRNA (Giege). Discrimination may also result from exclusion of negative determinants, perhaps because bulky substitutions are not allowed at certain positions (Schimmel; Giege; T. Steitz).

The U1 snRNP A protein is a member of a large family of RNA-binding proteins that contain a conserved 80 amino acid RNA-binding domain variously known as the RNP-consensus (RNP-CS) RNA-binding domain, the RNA recognition motif (RRM), or the RNP motif. The crystal structure of the U1A domain shows a four-stranded β sheet and two α helices, with amino acids from the short conserved RNP-1 and RNP-2 motifs interacting across the β sheet (Nagai, MRC, Cambridge). The NMR structure of a related domain from the hnRNP C protein is similar, however, a loop connecting two of the β strands, which is implicated in specific RNA binding by U1A, is absent (Dreyfuss, University of Pennsylvania). Extensive mutagenesis studies of U1A suggest that basic amino acids clustered at one end of the domain, and amino acids within the RNP consensus regions, interact with the RNA (Nagai). Ethylation protection experiments using a short U1 RNA stem-loop show two regions of protection, consistent with a modeled structure of the RNA–protein complex (Nagai). In poly(A) binding protein, protein–protein interactions between adjacent RNP-CS domains may be required to achieve full RNA-binding specificity (Pieler, Max Planck Institute, Berlin).

Information about RNA–protein interactions is also being provided by systems in which structures are not yet available. In the HIV Tat–TAR interaction, a single arginine residue within a region of basic amino acids provides an essential sequence-specific RNA contact and may contact two phosphates at a bulge (Frankel). The free amino acid arginine binds specifically to the same site in TAR, emphasizing the importance of RNA structure in the interaction (Frankel). The key role of RNA structure in protein recognition is also implicated in the interaction of bacteriophage R17 coat protein with its RNA-binding site. Preliminary NMR studies indicate that bases required for binding are intimately tied to the overall shape of the RNA (Uhlenbeck, University of Colorado). Backbone-specific footprinting and interference experiments suggest a good correspondence between the RNA shape and protein contact points, perhaps reflecting specific interactions with the backbone. Protein–protein interactions can also contribute substantially to RNA binding; strong cooperativity is observed when two R17 coat protein dimers bind to adjacent RNA sites (Uhlenbeck).

Remarkable flexibility of protein sequence was reported for two RNA-binding proteins, supporting the view that RNA structure plays an important role in the interactions. In the Xenopus 5S rRNA–binding protein, TFIIIA, any of

the nine individual zinc fingers could be deleted without affecting RNA binding (Pieler). Even groups of three fingers could be deleted, suggesting that only a few fingers, regardless of their identity, may be sufficient for 5S RNA recognition. Ricin A chain, a toxin that specifically cleaves one N-glycosidic bond in a 17 nucleotide loop of eukaryotic 28S rRNA, also shows extraordinary sequence flexibility (Wool, University of Chicago). A large number of mutants, in which a combined total of 221 of 265 amino acids had been changed, were still capable of specific cleavage, suggesting that there is little essential amino acid sequence information required to specify a functional structure. Mutagenesis (Wool) and preliminary NMR data (Moore, Yale University) of a short RNA hairpin from 28S rRNA suggest that the 17 nucleotide loop is highly structured and that the region immediately surrounding the cleavage site may resemble a GAGA tetraloop. Tetraloops are among the most common loops found in RNA hairpins and have particularly stable structures (see below). Since the crystal structure of ricin is known, details of the RNA–protein interaction should be forthcoming.

RNA Structure and Conformational Change

The importance of RNA structure in RNA–protein interactions has been generally appreciated. However, because both the RNA and the protein must fold into specific structures, yet neither the RNA nor the protein folding problem has been solved, identifying the source of specificity has been difficult. To further complicate matters, several examples were presented at this meeting that suggest that the RNA, the protein, or both, can undergo conformational changes as the complex is formed.

Significant progress is being made in determining RNA structures, and unexpected interactions are being discovered. The NMR structures of two stable tetraloops have shown that non-Watson-Crick base pairs occur and that unusual base–backbone interactions stabilize the structure. Preliminary NMR data on a noncleavable analog of the hammerhead ribozyme (Uesugi, Osaka University) suggest that the ribozyme may also contain noncanonical base pairs and may be stabilized by a Mg^{2+} ion bound to a non-base-paired region (ribozymes have been reviewed by Cech, 1990). Another noncanonical base pair (G-G) was identified within a loop of the RRE, the HIV Rev-binding site, using an RNA selection method. The G-G base pair may distort the backbone, allowing recognition by Rev (Green, University of Massachusetts, Worcester).

Tertiary interactions can be crucial for stabilizing RNA structure. In the Tetrahymena ribozyme (Cech, 1990), a hydrogen bond was inferred between a specific 2′-hydroxyl on the RNA substrate and the N1 of an adenine distant in the secondary structure by systematically substituting 2′-hydroxyl groups and measuring cleavage kinetics of mutant RNAs (Cech, University of Colorado). This tertiary interaction contributes to holding the RNA substrate into the catalytic core of the ribozyme. Effects at other nucleotides near the critical adenine may indicate an induced fit as the substrate binds (Cech). Ribonuclease mapping of trypanosome spliced leader RNAs indicates that tertiary interactions are present, although the precise interactions have not yet been identified (Crothers, Yale University). Spliced leader RNAs are involved in trans-splicing, in which independent transcripts are joined to create mature mRNAs (see Agabian, 1990). Many tertiary interactions also exist within ribosomal RNAs, and one particular interaction in 26S rRNA was shown to be important for yeast L25 protein binding (Raué, Vrije University, Amsterdam).

"Bulged" nucleotides (unpaired bases within a duplex stem) are found in many biologically significant RNAs and are important for RNA–protein interactions. NMR studies of the bacteriophage R17 coat protein–binding site suggest that the single bulged adenosine required for protein recognition may actually be stacked into the RNA helix (Uhlenbeck). An experiment in which the R17 RNA-binding site was circularized and then cleaved to generate ends at all possible positions showed that the presence of a nick just 3′ to the "bulged" adenine did not reduce protein binding (Uhlenbeck). This suggests that the RNA structure in this region may be held rigidly, even in the absence of the phosphodiester bond.

Pseudoknots (structures in which nucleotides within a hairpin loop base pair to nucleotides outside the hairpin) continue to be identified frequently in regulatory systems. One interesting example is seen in autoregulation of E. coli threonyl-tRNA synthetase (Ehresmann, CNRS, Strasbourg). A pseudoknot resembling the structure of threonine tRNA is present in the synthetase mRNA; synthetase can bind to this site and repress its own translation. Pseudoknots have been identified as binding sites for the 30S E. coli ribosomal subunit and for HIV reverse transcriptase by RNA-binding site selection experiments (SELEX, see below) (Gold, University of Colorado). The ribosomal pseudoknots may be important in translational initiation; the pseudoknots that bind to reverse transcriptase function as enzyme inhibitors in vitro. The structure of the pseudoknot in turnip yellow mosaic virus RNA, originally proposed as a structural explanation for the tRNA-like properties of the viral RNA, has been largely confirmed by NMR experiments and by systematic mutagenesis using a valyl aminoacylation assay (Pleij, University of Leiden).

RNAs can often adopt more than one conformation, and transitions between conformations can play a role in RNA–protein interactions. For example, trypanosome spliced leader RNAs rapidly interconvert between two distinct conformations, as shown by temperature jump and gel electrophoresis experiments (Crothers). Point mutations that stabilize or destabilize particular base pairs can lock the RNA into one or the other conformer (Crothers). The S4 ribosomal protein binds to an RNA pseudoknot and induces a conformational change; this transition may convert between "closed" and "open" forms and allow enhanced ribosome binding (Draper, Johns Hopkins University). Changes in RNA structure are also seen when L11, S12, or S15 ribosomal proteins bind to their rRNA sites (Draper; Noller, University of California, Santa Cruz). In the ribosome, a loop in 16S RNA undergoes a conformational transition when a cognate tRNA binds to the anticodon (Noller). Conformational rearrangement of the loop is thought to be important for translational proofreading. Conformational changes that occur in tRNA synthetase–

tRNA interactions (T. Steitz; Giege) were discussed above. In the Tat peptide–TAR interaction, circular dichroism experiments suggest that both RNA and peptide may change conformation upon binding (Frankel). Even larger conformational transitions in RNA may be required during splicing or translation reactions, perhaps facilitated by ATP-dependent RNA helicases (Guthrie, University of California, San Francisco; Sonenberg, McGill University, Montreal).

Interaction between Complex RNPs: Pre-mRNA Splicing

The splicing of mRNA precursors requires the proper recognition, assembly, conformational changes, and dissociation of multiple ribonucleoprotein complexes (reviewed by Guthrie, 1991; Rosbash and Seraphin, 1991; Ruby and Abelson, 1991). The examples of RNA splicing presented at this meeting demonstrate several ways in which RNA–protein interactions can influence snRNP–pre-mRNA interactions, splice site selection, or spliceosome assembly.

The steps of the RNA-splicing reaction are generally conserved between yeast, Drosophila, and mammals, and many of the components are similar. For example, it has long been known that spliceosomal snRNAs are conserved between yeast and higher organisms. It appears that many protein components, either intrinsic snRNP proteins or extrinsic protein splicing factors, are also conserved. For example, a U5 snRNP protein from yeast, PRP8, is similar to that found in mammals, Drosophila, and plants (Beggs, Edinburgh University). PRP8 appears to be required for spliceosome assembly after formation of the early U1–U2 snRNP complex (Beggs) and, by genetic analysis, seems to interact with the *DED1* and *DBP1* genes. *DED1* and *DBP1* are members of the "DEAD" gene family, thought to encode ATP-dependent helicases (see Wassarman and Steitz, 1991). Biochemical experiments using the yeast PRP28 (DEAD motif) and PRP16 (DEAH motif) proteins demonstrated RNA-dependent ATPase activity in vitro (Guthrie). By analogy to translational proofreading, it was suggested that ATP hydrolysis may provide energy to ensure the fidelity of spliceosome assembly and splicing reactions (Guthrie). This could help explain how correct splice sites are chosen in complex pre-mRNAs apparently containing many potential splice sites.

Interactions between snRNPs and pre-mRNAs were investigated by both genetic and biochemical experiments. Mutations in a conserved loop of U5 snRNA in yeast, previously isolated as suppressors of 5′ splice site mutations, were shown to activate aberrant 5′ cleavage sites (Newman, MRC, Cambridge). The mutations correlated with sequences upstream of the new 5′ cleavage sites, suggesting that the U5 loop might base pair with these regions of the pre-mRNA. Surprisingly, different mutations in the loop of U5 could suppress other pre-mRNA mutants blocked at the second step of splicing (cleavage at the 3′ splice site AG dinucleotide and exon ligation). This result led to the suggestion that U5 might also base pair with the first two nucleotides of the downstream exon (Newman). Experiments using mutant U1 snRNAs in Schizosaccharomyces pombe indicated possible base pairing between U1 snRNA and the AG dinucleotide at the 3′ splice site (Wise, University of Illinois). Although it is not entirely clear what roles these three proposed interactions might play in the splicing of nonmutant pre-mRNAs, the results strengthen the parallels between the splicing of nuclear pre-mRNAs and autocatalytic RNAs (Cech, 1990), where base pairing interactions at splice sites and adjacent exon sequences are important (Newman).

At the biochemical level, a sizable collection of mutants of both U4 and U6 snRNAs have been tested for their ability to complement splicing in Xenopus oocytes depleted of the corresponding endogenous snRNA (Mattaj). All of the noncomplementing mutants turned out to affect steps of either U4/U6 snRNP assembly or the normal assembly of U4, U6, or U5 RNA into splicing complexes on the pre-mRNA. In yeast, an ATP-independent complex of U1 snRNP with the 5′ splice site in pre-mRNA was identified (Rosbash, Brandeis University, Waltham, Massachusetts). Biochemical fractionation indicated that a non-snRNP-associated protein may mediate the interaction between U1 snRNP bound at the 5′ splice site and the intron branchpoint sequences (Rosbash). Biochemical complementation experiments identified a U5 snRNP protein that restores assembly of the U4–U6–U5 tri-snRNP complex and splicing activity to heat-inactivated splicing extracts (Krämer, University of Basel; Lührmann, Philipps University, Marburg, Germany). The mammalian splicing factor, U2AF, has been purified and exists as a heterodimer of 65 kd and 35 kd subunits (Green). U2AF binds to the intron polypyrimidine tract and is required early in spliceosome assembly for the ATP-dependent association of U2 snRNP with the intron branchpoint sequence. The isolated 65 kd subunit is sufficient for activity in vitro. The gene for the large subunit encodes a protein that contains three RNP-CS domains and an extensive arginine/serine-rich (R/S) domain that seems to be common to several splicing factors. In at least two cases, this region is essential for activity in vitro and may be involved in protein–protein or protein–RNA interactions (Green; Manley, Columbia University). All three U2AF RNA-binding domains are required for RNA binding (Green).

RNA–protein interactions figured prominently in examples of regulated pre-mRNA splicing (see Maniatis, 1991). In Drosophila P element transposition, splicing of the third intron of the transposase gene is specifically repressed in somatic cells. In vitro biochemical experiments suggest that a multiprotein complex binds to a site in the 5′ exon, thereby preventing U1 snRNP from binding to the nearby 5′ splice site (Rio). Somatic sex determination in Drosophila also involves alternative splicing events (see Baker, 1989; Maniatis, 1991). In vivo, sex-specific splicing of the Drosophila *transformer* gene by the RNA-binding protein, Sex-lethal, involves a conserved U_8C sequence in the non-sex-specific intron polypyrimidine tract as well as the intron branchpoint sequence (McKeown, Salk Institute, San Diego). In vitro, recombinant Sex-lethal protein blocks the interaction of the U2AF activator with the intron polypyrimidine tract, thereby causing a shift in 3′ splice site utilization and mimicking the regulation observed in vivo (Green). Cotransfection of the transformer and trans-

former-2 proteins appears to activate a female-specific 3′ splice site in the Drosophila *double-sex* gene (Shimura, Kyoto University). Transformer-2 is known to bind to repeated sequence elements in the female-specific exon in vitro via an RNP-CS domain. In Xenopus, splicing of two of the nine introns of the ribosomal L1 protein pre-mRNA is inhibited when L1 protein is present in excess (Bozzoni, University of La Sapienza). One of the two introns in the L1 pre-mRNA is also subject to specific cleavage, presumably further down-regulating expression (Bozzoni).

RNA structure can influence splicing. In dicotyledenous plants, AU-rich sequences in introns are required for their removal while in monocots, AU-rich sequences enhance intron removal (Filipowicz, Friedrich Miescher Institute, Basel). It was proposed that AU-richness might help minimize formation of RNA secondary structure. An AU-containing hairpin inserted into a dicot intron inhibited splicing, supporting this proposal (Filipowicz). Mutational analysis of the chicken β-tropomyosin pre-mRNA indicates that specific base-paired structures can be important in alternative splicing (Brody, CNRS, Gif sur Yvette). In this system, splice site selection is tissue specific, suggesting that transacting factors may be present in different cell types that influence the stability of the RNA structure. Previous work on tRNA splicing showed that a splicing endonuclease recognized a structure common to both the precursor and mature tRNAs and then selected cleavage sites based on their distance from that structure. New data suggest that this may not be the whole story (Tocchini-Valentini, CNR, Rome). The identity of two nucleotides, one close to the 5′ end and one adjacent to the 3′ end of the intron, was shown to be important for cleavage of the phenylalanyl tRNA precursor in Xenopus. Although the intron sequence had not been thought to influence cleavage, mutational analysis supported a role for base pairing between an intron base and the first (5′) nucleotide.

snRNPs and hnRNPs: Transport and Other Functions

The assembly, processing, and transport of snRNPs and hnRNPs was discussed extensively (reviewed by Lamond, 1991). U6 snRNA was found to be modified at the 3′ end in most species; a 2′-3′ cyclic phosphate exists at the 3′ end of U6 in vertebrates and higher plants and either a noncyclic phosphate or another, as yet unidentified, blocking group was found in lower eukaryotes (Dahlberg, University of Wisconsin, Madison). Of the species examined, only trypanosome U6 had an unmodified 3′-hydroxyl terminus. Although the significance of the modification is unknown, the observation that the ratio of modified to unmodified U6 varies in a cell-type- and stage-specific way suggests that modification may be regulated (Dahlberg).

The final stage of U1 snRNA 3′ end formation involves removal, in the nucleus, of one or two extra nucleotides left on the 3′ tail after cytoplasmic processing of the precursor RNA (Lund, University of Wisconsin, Madison). Inhibitors have been used to analyze the transport of U snRNA precursors out of the nucleus and their return as partially mature U snRNPs. Wheat germ agglutinin inhibits U snRNA export from the nucleus (Lund) but, while it blocks protein and U6 snRNA import into the nucleus essentially quantitatively, its effect on the reimport of U1-U5 snRNPs is much less pronounced (Lund; Lührmann). Other competitors also have differential effects on the transport of U1-U5 snRNPs and nucleophilic proteins (Lührmann), suggesting that there are differences in the cellular components mediating the two processes. Part of the localization signal required for nuclear accumulation of some U snRNPs in oocytes is the 5′-trimethylguanosine (TMG) cap structure. An activity capable of specific hypermethylation of the U snRNA TMG cap structure was identified in HeLa cell cytoplasmic extracts and does not appear to be an integral part of the mature snRNP particles (Lührmann).

The transport of mRNA from the nucleus to the cytoplasm is less well understood, although there are indications that particular classes of hnRNP proteins might be involved (Dreyfuss). Previous work revealed that following mitosis, certain hnRNP proteins, for example, the C class, return directly to the nucleus, whereas others, for example, hnRNP A1, require RNA polymerase II transcription before they can return. It was shown that hnRNP A1 shuttles between the nucleus and cytoplasm and that transcription is also required for nuclear return during interphase (Dreyfuss). The A1 protein, when in the cytoplasm, is attached to poly(A)-containing RNA and may leave the nucleus complexed with mRNA (Dreyfuss). The unique C-terminal domain of hnRNP A1 was shown to be an in vitro substrate for protein kinase A and casein kinase II, but not protein kinase C (Riva, CNR, Pavia). The isolated C-terminal domain could interact with DNA oligonucleotides only in the unphosphorylated form. Phosphorylation may also affect RNA binding, since different RNAs were found to associate with the A1 protein in vitro in the presence or absence of the C-terminal domain (Riva).

The HSURs (Herpes virus saimiri U RNAs) are the first virally encoded members of the U snRNA class. In common with many of the lower abundance cellular U snRNAs, their function has not yet been determined. Three of the five HSURs thus far examined contain one or more copies of the sequence AUUUA (J. Steitz, Yale Medical School). These sequences are also found in the 3′ UTRs of short-lived cellular mRNAs, including several growth factors and proto-oncogenes, and it is known that the AUUUA sequence can confer instability on heterologous stable mRNAs and allow binding of a 32 kd protein (for a review on mRNA stability, see Peltz et al., 1991). HSURs bind the same 32 kd protein, suggesting that they might compete for binding of the protein and thereby affect degradation of important short-lived cellular mRNAs (J. Steitz).

The enzyme telomerase is a nuclear RNP required to maintain the ends of chromosomes (see Blackburn, 1991). Telomerase was shown to be processive in vitro though it appears to be distributive in vivo (Greider, Cold Spring Harbor). A DNA primer anneals to the internal telomerase RNA template to initiate elongation, but it seems that the 3′ end of the primer does not have to base pair with the RNA, resulting in the ability to add telomeres to a variety of DNA fragments with unrelated 3′ terminal sequences. This may be important if telomerase has to recognize and add telomeres to broken chromosome ends. Interestingly,

telomeres tend to shorten as a cell ages and a corresponding decrease in telomerase activity is also observed. However, telomeres do not shorten in transformed cells (Greider).

In addition to the small nuclear RNPs, vertebrates contain small cytoplasmic RNPs (Ro RNPs) whose function is unknown. The four human Ro RNAs—hY1, 3, 4, and 5—have been found to associate with two Ro antigens, Ro60 and Ro52, and with the La protein, which also transiently associates with other RNA polymerase III transcripts. Biochemical experiments showed that both La and Ro60 bind directly to hY RNAs, but that Ro52 could bind RNA only in the presence of Ro60 (van Venrooij, University of Nijmegen, The Netherlands). The two Ro proteins form a complex in the absence of RNA, suggesting that Ro52 may interact with the RNP primarily through protein–protein interaction with Ro60 (van Venrooij).

Another cytoplasmic ribonucleoprotein complex, the signal recognition particle (SRP), contains six proteins bound to a large structured RNA. In S. pombe, substituting one stable tetraloop in the RNA with a totally different tetraloop did not affect function, suggesting that the two tetraloops (AAAG and UUCG) can adopt similar structures in vivo (Wise).

RNA in Early Development

In many organisms, the earliest stages of development proceed without transcription. After fertilization, but prior to the onset of zygotic transcription, embryos rely exclusively on the regulation of pre-existing maternally synthesized mRNAs to modulate protein synthesis. Proper control and localization of maternal mRNAs is critical for the execution of important developmental decisions, including the specification of body axes and control of the cell cycle. Thus, RNA–protein interactions must be of fundamental importance in early development.

Many maternal mRNAs are translationally repressed ("masked") until a specific stage in development. In the clam, specific sequences in the 3′ untranslated region (3′ UTR) of ribonucleotide reductase and cyclin A mRNAs mediate translational inhibition both in vivo and in vitro (Standart, Cambridge University). Masking can be reversed by deleting the 3′ UTR sequences or with antisense RNA. Masking of mouse protamine mRNAs during spermatogenesis also requires sequences in the 3′ UTR (Hecht, Tufts University, Medford, Massachusetts). In both cases, specific proteins can be crosslinked to the 3′ UTR RNAs (Standart; Hecht). Although the involvement of these proteins in translational control has not yet been demonstrated, a regulatory role is suggested by the observations that both the clam and mouse proteins are subject to phosphorylation (Standart; Hecht) and that crosslinking to the mouse protein disappears during spermatogenesis (Hecht).

The addition or removal of the poly(A) tail at the 3′ end of mRNAs is a widespread means of regulating translation of maternal messages. Here again, critical signals for this type of control lie in the 3′ UTR (Wickens, University of Wisconsin, Madison). mRNAs containing appropriate sequences (e.g., UUUUUUAU) receive poly(A) and are activated, while mRNAs that lack these signals lose their tails and are translationally shut off. These reactions are regulated, both with respect to when they occur during development and to the number of A residues gained or lost. Polyadenylation of maternal mRNAs in the cytoplasm may be similar to nuclear polyadenylation, involving separable polymerase and specificity components and similar sequence signals, and may be required to regulate the meiotic cell cycle (Wickens).

Many of the RNA processing/modification reactions that occur in early embryos are found throughout development. For example, removal of poly(A) occurs in all eukaryotic cells, including yeast (see below). Similarly, an RNA modification/unwinding activity first identified in frog eggs has been detected in a variety of somatic cells. This enzyme binds specifically to double-stranded RNA and converts adenosine (A) residues to inosine (I). Mechanistic studies indicate that A to I modification occurs by direct deamination and that there is a substrate preference for partially unwound duplex RNAs (Bass, University of Utah).

It is clear that localization of specific maternal mRNAs to particular poles of Drosophila or Xenopus eggs plays a key role in the specification of axial polarity. Localization of maternal mRNAs to the anterior or posterior pole in Drosophila embryos involves the 3′ UTR of target mRNAs. Genetic and molecular analyses have begun to identify transacting factors in Drosophila that affect RNA localization, some of which presumably bind directly to target mRNAs (MacDonald, Stanford University).

Polyadenylation and Deadenylation: Specificity through Protein-Protein Interactions

Several talks described the formation and processing of mRNA 3′ ends (reviewed by Wickens, 1990; Sachs, 1991). As with pre-mRNA splicing and some of the other systems, the general results seem to indicate that multiple proteins cooperate, either to provide RNA-binding specificity or to create appropriate substrates for the RNA processing reactions.

The RNA polymerase that adds poly(A) tails onto mRNAs in the nucleus has been cloned from mammals and yeast (Manley; Wahle, University of Basel; Keller, University of Basel). The mammalian enzyme contains, at its N-terminus, RNP-CS domains and a polymerase module found in RNA-dependent polymerases, and multiple potential phosphorylation sites at its C-terminus (Manley). Poly(A) polymerase (PAP) will add a stretch of adenines to the 3′ end of any RNA; however, specific addition to RNAs containing the polyadenylation signal, AAUAAA, requires another protein composed of 3–4 polypeptides. This multisubunit cleavage–polyadenylation specificity factor (CPSF) binds directly to the AAUAAA site and to PAP (Keller; Manley). CPSF is also necessary for the 3′ cleavage reaction, but here other proteins appear to form a bridge between it and the actual endonuclease (Manley). The sequences of the bridging proteins are intriguing and may point to specific regions involved in protein–protein interaction; one protein contains 12 repeats of the sequence MEARA, which may form a long α helix, and another contains transducin repeats found in the β subunit of G-proteins and in

the yeast pre-mRNA splicing protein PRP4 (Manley). After a short poly(A) tail is added, an additional protein may bind to poly(A) and interact with PAP (Wahle).

Maturation of mRNAs in the cytoplasm requires degradation of the poly(A) tail to a specific length. Interestingly, the signal for specific deadenylation is not merely RNA itself, but rather an RNA–protein complex containing the poly(A)-binding protein (PAB) and poly(A) tail (Sachs, Whitehead Institute, Cambridge, Massachusetts). The poly(A) nuclease (PAN) from yeast has been purified and requires that PAB be bound to poly(A) for RNA degradation in vitro (Sachs). In addition to their roles in general poly(A) shortening, PAB and PAN may also play a role in rapid degradation of specific mRNAs. Yeast mating-type a factor mRNA is a short-lived mRNA in vivo. A 12 nucleotide sequence in the 3′ UTR seems to target the mRNA for rapid degradation in vitro, in a PAB- and PAN-dependent reaction (Sachs). In contrast to polyadenylated mRNAs, histone mRNAs contain no poly(A) tail and are degraded in a cell cycle–specific manner (Marzluff, University of North Carolina, Chapel Hill). A stem-loop near the 3′ end of histone mRNAs has been identified as the crucial element and a specific RNA-binding protein has been UV cross-linked to this element (Marzluff). For a recent review of mRNA turnover and its regulation, see Peltz et al. (1991).

The complexity of RNA–protein interactions was further illustrated by study of termination of vaccinia virus transcription, in which the same two proteins that add the 5′ cap to new initiated RNA also facilitate termination by the elongating RNA polymerase (Shuman, Sloan-Kettering Institute, New York). Termination requires that the capping enzymes recognize the sequence UUUUUNU as the RNA emerges from the transcription complex and that they contact the polymerase. Thus, the same two proteins bind to the polymerase, the cap, and the nascent RNA and catalyze the three sequential enzymatic reactions of capping (Shuman).

New Methods for Studying RNA Structure, Function, and Protein Interactions

As many of the structural talks indicated, probes of the RNA backbone will be very useful in studies of RNA–protein interactions. In vitro transcription can be used to replace phosphorus with sulfur, creating phosphorothioate-substituted RNAs, and the modified positions can be efficiently cleaved by iodine (Eckstein). This method has been used to footprint aminoacyl-tRNA synthetase–tRNA complexes (Eckstein; Giege). RNA 2′-hydroxyl groups can be substituted with 2′-fluoro or 2′-deoxy groups by chemical synthesis. These substitutions have been used to study the kinetics, mechanism, and structure of the Tetrahymena and hammerhead ribozymes (Cech; Eckstein). For studies of RNA folding and RNA tertiary structure, Fe(II)-EDTA has been a useful cleavage reagent to probe accessibility and has been used in studies of the Tetrahymena ribozyme (Cech).

Methods were described for selecting RNA molecules from large random pools in which repeated cycles of purification and amplification are used to isolate protein-binding sites (SELEX; Gold). Examples were presented using R17 coat protein (Uhlenbeck; Gold), HIV Rev (Gold; Green), and the rho bacterial transcription termination factor (Platt, University of Rochester, New York). Selection of the rho binding site yielded an unexpected result. From past work, rho appeared to bind to unstructured C-rich RNAs. However, the SELEX experiment identified a stem-loop structure next to a C-rich region that binds to rho with high affinity and may have a counterpart in natural RNAs (Platt). Thus, it seems likely that by carefully designing selection experiments, it will be possible to identify important determinants of RNA–protein recognition and RNA structure.

These new methods will undoubtedly be applied to many of the biologically interesting systems presented at this meeting. Problems as diverse as mRNA transport, regulation of pattern formation, and tRNA identity find common ground in the specific interactions of RNAs and proteins. We are optimistic that the next few years will bring rapid progress in the broad and emerging field of RNA–protein interactions.

References

Agabian, N. (1990). Cell *61*, 1157–1160.

Baker, B. S. (1989). Nature *340*, 521–524.

Blackburn, E. H. (1991). Nature *350*, 569–573.

Cech, T. R. (1990). Annu. Rev. Biochem. *59*, 543–568.

Guthrie, C. (1991). Science *253*, 157–163.

Lamond, A. I. (1991). Curr. Opin. Cell Biol. *3*, 493–501.

Maniatis, T. (1991). Science *251*, 33–34.

Peltz, S. W., Brewer, G., Bernstein, P., Hart, P. A., and Ross, J. (1991). Crit. Rev. Euk. Gene Exp. *1*, 99–126.

Rosbash, M., and Seraphin, B. (1991). Trends Biochem. Sci. *16*, 187–190.

Ruby, S. W., and Abelson, J. (1991). Trends Genet. *7*, 79–85.

Sachs, A. (1990). Curr. Opin. Cell Biol. *2*, 1092–1098.

Wassarman, D. A., and Steitz, J. A. (1991). Nature *349*, 463–464.

Wickens, M. (1990). Trends Biochem. Sci. *15*, 277–281.

Wickens, M. P., and Dahlberg, J. E. (1987). Cell *51*, 339–342.