Cell, Vol. 51, 339–342, November 6, 1987, Copyright © 1987 by Cell Press

RNA–Protein Interactions

Meeting Report

Marvin P. Wickens* and James E. Dahlberg†
* Department of Biochemistry
† Department of Physiological Chemistry
University of Wisconsin
Madison, Wisconsin 53706

The importance of RNA–protein interactions has been appreciated for many years; yet until recently, these interactions have been difficult to analyze in detail. Progress has been limited by technical barriers, such as the inability to prepare large amounts of pure RNA and pure protein, or to determine their structures. Judging by reports at a recent workshop on RNA–Protein Interactions,* many of these barriers have now been overcome. Speakers examined not only the mechanism of the binding reactions, but also the functions performed by RNA–protein complexes.

How do proteins recognize and bind to specific RNA molecules? RNA molecules are idiosyncratic: each has its own distinctive three-dimensional shape and, as with proteins, that shape cannot yet be predicted from primary sequence. This complicates analysis of RNA–protein interactions. With DNA, the structure of at least one of the molecules is reasonably well known; generally, with RNA, neither structure is. In this respect, the interaction of a "globular" RNA with a protein is more analogous to protein–protein binding than to protein–DNA binding.

Simplified Systems: One RNA, One Protein

Some general features of RNA–protein interactions emerged from the analysis of simplified systems. Uhlenbeck discussed the highly specific binding of the coat protein of R17 coliphage to a 21-base sequence of the R17 RNA genome. The interaction requires a particular RNA conformation—a seven base-pair stem with a one base pucker, and a four nucleotide loop. The base-paired stem is essential though its sequence is irrelevant. In contrast, at several non-base-paired positions, the base *is* critical.

The same principles govern the interaction of aminoacyl-tRNA synthetases with tRNAs (Schimmel, Schulman, Uhlenbeck). Mutant RNAs that maintain the familiar L-shape of the tRNA still bind efficiently. In contrast, binding is prevented by disruption of tertiary interactions that help maintain the L-conformation. Single-stranded bases also contribute to specificity: changing the sequence of the anticodon in certain tRNAs results in recognition by the "wrong" synthetase and therefore in misacylation. The X-ray crystal structures of tRNAs alone or tRNAs cocrystallized with synthetases confirm direct contacts that had been predicted from genetic and nucleotide modification data (Giegé, Moras, T. Steitz). tRNA ligase and endonuclease, like the synthetases, recognize both pre-tRNA structure and a few specific, highly conserved nucleotides (Abelson, Tocchini-Valentini, Greer, Knapp).

Although only a small number of biologically significant RNA structures have been analyzed in detail, several important generalities are already emerging. For example, a so-called "bulged" adenosine is present in many of these RNAs. A bulged nucleotide is one that lies in the midst of a base-paired stem but lacks a partner on the opposite strand, and so could protrude from the helix. However, "bulge" may be a misnomer, since both U and A bulged bases in synthetic duplexes are actually stacked into the helix and cause it to bend by about 15° (Draper). In the R17 coat protein binding site, as in other translational operators, a bulged A interrupts an otherwise perfectly base-paired 7 nucleotide stem (Uhlenbeck). This A is essential for coat protein binding. Physical data indicate that it actually is intercalated, and may provide a specific contact with the protein. Similarly, in mRNA splicing, the adenosine which serves as the branch point in the intron is a bulge in a base-paired stem: this structure can be formed either intramolecularly, as in the self-splicing of group II introns, or by base-pairing between an mRNA precursor and U2 snRNA, as in the splicing of yeast nuclear mRNAs (Guthrie). The binding of ribosomal proteins S8 and S15 to 16S rRNA also requires stem-loops containing bulged adenosines (Draper).

Several cases in which RNA sequence is important in single-stranded regions were also discussed. The Sm antigen(s) of snRNPs recognizes and binds to a simple sequence present in snRNAs, $AU_{3\text{-}6}G$, even if that sequence is embedded in a foreign RNA molecule (Mattaj). Nonionic interactions between rho factor and non-base-paired cytidines contribute significantly to bacterial transcription termination (Richardson). Antitermination factors may affect this event by altering the duration of transcriptional pausing (Roberts).

Regulation in Simple Systems

Regulation can be achieved by variations on the single protein-single RNA paradigm. E. coli threonyl synthetase represses translation of its own mRNA, as do several ribosomal proteins (Draper, B. Ehresmann, C. Ehresmann, Leffers, Springer, Zimmermann). In each case, the structure recognized in the mRNA is similar to that in the tRNA or rRNA to which the protein also binds. For example, threonine synthetase binds to a region in its own mRNA that is similar in both sequence and secondary structure to the anticodon stem and loop of several threonyl tRNAs. Moreover, synthetase mutants that overrepress translation bind tRNA more tightly than normal, while mutants that derepress translation bind tRNA less tightly (C. Ehresmann, Springer). In other mRNAs, translational repression (Brody) or activation (Oppenheim) can be achieved by interactions with specific ribonucleases at or near ribosome binding sites.

Not only can the same protein recognize the same structure in two different RNAs, but different proteins with

* Workshop on RNA–Protein Interactions, held June 29–July 1, 1987 at the Sogesta Conference Center near Urbino, Italy. The workshop was cosponsored by the European Molecular Biology Organization, the National Science Foundation, the Italian chemical company EniChem, and the Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie of CNR.

very different functions apparently can recognize the same RNA structure. The interaction of ribosomal protein L5 and transcription factor IIIA (TFIIIA) with 5S RNA reveals what may be a versatile regulatory network that coordinates levels of L5 and 5S RNA (Romaniuk, J. Steitz, Wool). Both proteins bind to the same region of 5S RNA, leading to the speculation that L5 could displace bound TFIIIA. In the cell, excess 5S RNA is found in a "storage particle" with TFIIIA. Free L5 might displace the TFIIIA from this particle, and thereby direct the 5S RNA to the nucleolus for ribosome assembly and at the same time liberate the transcription factor to stimulate 5S RNA transcription. J. Steitz speculated that viroids, small infectious RNAs which destroy host plant cells, may interfere with assembly of host ribosomes by sequestering L5. Viroid RNAs are nucleolar, are associated with a protein of the same apparent molecular weight as L5 (Klaff), and possess a structure very similar to that found in the 5S RNA binding site (J. Steitz).

More Complex Systems: Multiple Proteins Bound to One RNA

Our understanding of more complex systems is well advanced in the signal recognition particle (SRP), a complex of six proteins and a 300 nucleotide RNA (Walter). This particle targets polyribosomes synthesizing secretory proteins to the endoplasmic reticulum and, in addition, arrests translational elongation until the ribosomes associate with a receptor there. By reconstituting SRPs in the absence of one protein, mutant SRPs have been formed which can perform one or the other activity, both, or neither. Thus each protein may have a unique role in the function of this particle, rather than the particle functioning only when intact.

snRNPs, which participate in mRNA splicing, are analogous to the SRP, in that each snRNP contains a single RNA and a small number of proteins. snRNPs can now be reconstituted in vitro using crude extracts of either HeLa cells (Green) or frog oocytes (Mattaj) and synthetic snRNAs made in vitro by phage polymerases. The reconstituted snRNPs possess the correct physical properties, contain expected proteins, and, at least for human U1 (Green), bind to the correct sequence in the mRNA precursor. It is not yet known whether reconstituted snRNPs can support splicing.

Mattaj analyzed the assembly of U2 snRNPs using RNAs from which various portions had been deleted. The binding of each protein is relatively independent of the binding of the others, suggesting that, unlike ribosome assembly, there may be no obligatory ordered pathway for the assembly of a snRNP. Yet, based on the properties of double deletion mutants, the proteins appear to interact once in an assembled particle. Green demonstrated that a truncated U1 snRNA containing only the first 40 nucleotides of the 164-nucleotide RNA can still assemble into a particle that binds efficiently to the 5′ splice site. This domain of the RNA contains the binding site for the 70 kd U1-specific protein.

The assembly of snRNPs containing mutant proteins will be facilitated by the cloning of genes encoding these polypeptides—work that is underway in several laboratories. The genes for the U1-specific 70 kd, A, and C proteins, for the U2-specific A′ and B″ proteins, and for protein D (present in all snRNPs), have each been cloned and at least partially sequenced (Spritz, van Venrooij, Lührmann, Hoch). One provocative point has emerged from a comparison of their inferred amino acid sequences. The 70 kd, U1-A, and U2-B″ proteins all contain a consensus sequence—the so-called "RNP consensus" of Arg/Lys–Gly–Phe/Tyr–Gly/Ala–Phe/Tyr–Val–X–Phe/Tyr—which is also found in the poly(A) binding protein, hnRNP proteins A1, A2, and C, and the pre-rRNA binding protein C23 (Dreyfuss). This octapeptide may be diagnostic of proteins that bind RNA, but its function, if any, is unknown.

Examination of the protein and RNA composition of snRNPs raises questions concerning both the heterogeneity within each class of snRNP and the similarities between classes (Jeanteur, Lührmann, Spritz, Van Venrooij). Lührmann suggested that the 70 kd protein present in U1 snRNPs is multiply phosphorylated to generate six isoforms distinguishable by their pI. U1 snRNPs are also heterogeneous in that U1 snRNA is actually a family of closely related but different RNAs (J. Dahlberg). Are these heterogeneities in protein and RNA functionally significant? All U-snRNPs, regardless of the RNA they contain, share at least seven proteins. What are the functions of these common proteins?

Interaction between Complex RNPs: mRNA Processing

Several speakers moved to the next higher level of complexity—the formation and function of multi-RNP assemblies. In mRNA splicing, for example, no less than five different snRNPs participate in the assembly of a particle capable of splicing mRNA precursors.

It is now clear that the differences noted between yeast and mammalian snRNPs are only superficial. Each of the major mammalian snRNAs has a counterpart in yeast, and although the yeast U2 homologue is much longer than mammalian U2 RNA, it fulfills the same function and is not a poly-snRNA, as had once been speculated. Moreover, the pathways used to form a pre-splicing complex, or spliceosome, appear to be similar, with the sequential binding of U1, U2, and then a U4/U5/U6 multi-snRNP particle. Since the pathways are similar, yeast genetics should provide a powerful means of elucidating the mechanism of eukaryotic mRNA splicing in general. For example, Beggs described experiments indicating that the *RNA8* gene product is probably in the yeast U5 snRNP, and Abelson indicated that the *RNA11* product is present in the mature splicing complex. The genetic approach may also bootstrap its way into new problems which are not yet easily assayed in vitro—RNA transport, for example.

The term spliceosome connotes homogeneity—that the complex which catalyzes splicing is a single, static entity. However, multiple complexes exist, since both extrinsic factors and snRNPs interact transiently with the assembling, fully competent complex. For example, although U4 and U6 may be extensively base-paired before associa-

tion with the mRNA precursor (Guthrie), U4 but not U6 snRNP is released prior to cleavage at the 5′ splice site and lariat formation. Since not all components present in the complex that catalyzes splicing have been identified, the term spliceosome, though useful, is ambiguous.

Recognition of the 3′ splice site and adjacent polypyrimidine tract may involve both U2 snRNP and other factors. It was suggested that two apparently different entities, IBP (Jeanteur, J. Steitz) and hnRNP-C (Dreyfuss), bind to this region, while another two, U2BF (Green) and U1 snRNP (Berget), facilitate binding of U2 snRNP to the prospective branch point. Keller described formidable efforts to purify all the components necessary for splicing from a crude mammalian nuclear extract. As yet, four separable fractions are required. A different approach, discussed by Reed, is to isolate intact, complete splicing complexes for analysis in the electron microscope.

The formation of mRNA 3′ termini appears to be a simpler process than splicing, but the two bear certain similarities. mRNA precursors are cut to generate a new 3′ end, to which poly(A) is added. Both cleavage and polyadenylation depend on the same conserved sequence, AAUAAA. Like splicing, 3′ end formation is preceded by the formation of a large complex (Keller, Wickens). An snRNP may participate in this reaction, but it has not yet been characterized (Keller). After a certain length of poly(A) has been added, the catalytic components are released and recycle (Wickens).

Regulation of mRNA processing was discussed by several speakers. In immunoglobulin pre-mRNAs containing more than one potential polyadenylation site, the relative affinity of each site for cleavage factors may determine which site is used (Galli). Regulation of splice site choice in mammals (Hartmuth, Brody, Spritz, Keller), frogs (Bozzoni), and yeast (Domdey) was also described. As yet, the regulation observed in vivo has not been reconstituted in cell-free systems.

A Paradigm for Complex Particles: The Ribosome

Perhaps the best understood particle containing multiple RNAs and proteins is the E. coli ribosome. A variety of methods—footprinting, chemical cross-linking, and chemical protection—have been used to position ribosomal proteins with respect to one another and to rRNA. These very different methods converge on a single ribosomal architecture (Brimacombe, Noller), a satisfying result that tends to validate the methods used. The use of simplified systems provides a detailed picture of how individual proteins interact with specific sites within the RNA molecule (Draper, B. Ehresmann, Leffers, Raué, Wool, Zimmermann). With this information, and with cloned ribosomal protein genes, it may be possible to approach long-standing questions in translation, such as the mechanisms of peptidyl transfer and chain elongation. Protein–protein interactions, which are not analyzed in the simplified systems, will undoubtedly be essential for a full understanding of these processes.

As several speakers pointed out, the analogy between the activities of the ribosome and the spliceosome may be instructive. Do both particles generate active sites through the interaction of multiple proteins and RNAs? What roles do extrinsic factors and ATP or GTP play in the formation and disassembly of these particles? To what extent do the RNAs play active, ribozymal roles, as opposed to their providing a scaffold to position proteins appropriately for catalysis?

Probing RNA Structure and Interactions

Perhaps the most obvious and pervasive question at the meeting was: how can we predict the structure of an RNA from its sequence? An answer will require a considerably expanded library of known RNA structures. A major limitation in applying crystallography or other physical methods (e.g., NMR) to RNA has been the difficulty of obtaining sufficient amounts of pure RNAs. For at least short RNAs, this problem may now have been circumvented. DNA oligonucleotides have been prepared which, when mixed with nucleoside triphosphates and large amounts of T7 RNA polymerase, direct the synthesis of up to 24 mg of a 21-nucleotide RNA (Uhlenbeck).

The identification of molecules that interact with an RNA sequence has become a prime objective in several areas represented at the meeting. Many of the methods used to study DNA–protein interactions have been applied successfully to RNA–protein interactions. These include enzymatic and chemical footprinting, and gel retardation assays. Chemical modification of RNA can be used to determine quickly which bases are essential for any reaction of interest, and so accelerate mutagenesis studies

The synthesis and use of photoactivatable nucleotides for specifically cross-linking RNAs to nearby proteins was discussed by Expert-Bezançon and Hanna. These analogues are reactive, contain relatively small chemical substitutions, and can be incorporated into RNA by phage RNA polymerases. Since a tRNA precursor containing 5-bromouridine can be specifically cross-linked to the yeast tRNA splicing ligase, the new reagents appear very promising (Abelson, Hanna). The ribosome, for which a rather detailed architecture has now been mapped (Noller, Brimacombe), may provide an excellent model system with which to evaluate further the utility of these reagents.

Horizons

As the techniques for analyzing RNA–protein interactions become more sophisticated and widespread, the investigation of new biological problems will become experimentally tractable. Intracellular transport of RNA is one such problem. The analysis of interactions between tRNA and pre-tRNA with the tRNA endonuclease and ligase (Abelson, Greer, Knapp, Tocchini-Valentini) has led to the speculation that these enzymes may function in the export of mature tRNAs from the nucleus. In contrast, snRNAs are exported from the nucleus as unprocessed transcripts (J. Dahlberg) that bind to Sm antigens in the cytoplasm and are processed in that cell compartment (J. Dahlberg, Mattaj). Subsequent to the binding of Sm antigen to specific sites in the RNA, the resulting RNP is transported back to the nucleus. Perhaps the most surprising case of intracellular RNA shuttling is that of the nuclear-encoded RNA which leaves the nucleus, enters the mitochondrion, and there enters an smRNP (small mitochondrial ribonucleo-

protein) that cleaves RNA primers synthesized during mitochondrial DNA replication (Clayton). What role do specific RNA–protein complexes play in each of these intracellular movements?

In addition to these new problems, each subject explicitly discussed at the meeting is likely to progress rapidly. A major focus in simplified systems will be structural analysis of RNA, using short RNA substrates and direct physical measurements of RNA conformation. Perhaps by elucidating and manipulating the structure of the proteins as well, it will become possible to develop an understanding of the biochemistry underlying the binding reactions. In more complex systems, the finding that RNPs can be made to form in vitro using synthetic RNAs is likely to accelerate work both on their assembly and on their function. Reports at this meeting make us optimistic that the assembly and function of even the most complicated, multicomponent RNA–protein particles—like ribosomes and splicing complexes—will be understood sooner and in greater detail than was expected just a few years ago.