Journal of the Korean Magnetic Resonance Society **1998**, *2*, 152-157

RNA-Protein Interactions and Protein-Protein Interactions during Regulation of Eukaryotic Gene Expression

Luca Varani, Andres Ramos, Paul T. Cole, David Neuhaus and Gabriele Varani*

MRC Laboratory of Molecular Biology
Hills Road - Cambridge CB2 2QH - UK
Received September 16, 1998

Abstract : The diversity of RNA functions ranges from storage and propagation of genetic information to enzymatic activity during RNA processing and protein synthesis. This diversity of functions requires an equally diverse arrays of structures, and, very often, the formation of functional RNA-protein complexes. Recognition of specific RNA signals by RNA-binding proteins is central to all aspects of post-transcriptional regulation of gene expression. We will describe how NMR is being used to understand at the atomic level how these important biological processes occur.

RNA-binding proteins have a modular structure comprising domains that achieve RNA recognition, and other domains that perform additional functions;[1,2] these additional functions often involve binding other proteins. The RNA-binding regions are composed of single or multiple copies of common RNA recognition units that sometimes constitute independent structural domains. The three most common RNA-binding domains are the RBD (or RRM) motif, the KH motif and the double-stranded RNA-binding domain. [1,2] Components of the RNA processing machinery are targeted to specific mRNAs through the recognition of specific RNA sequences and structures by RNA-binding regions or domains. Complexes functional in RNA processing or translational initiation are then formed by networks of RNA-dependent multiprotein complexes, targeted to mRNAs by specific interactions involving one or more of the components

of the assembly. Thus, biological specificity (identification of a specific intracellular RNA target) and physical chemical specificity (RNA-protein molecular recognition) are intimately related.

The three major eukaryotic RNA-binding protein motifs, RBD, KH-domain and dsRBD, all have a compact stable fold. All three classes of proteins display a common $\alpha\beta$ structure shared by many other prokaryotic and eukaryotic proteins involved in RNA recognition.[1,2] However, the protein fold is different for each of these proteins, suggesting convergent evolution to a common structure. This convergence could imply that the exposed β-sheet surface common to all such motives represent a convenient structural platform for RNA recognition. Interactions between residues from the β-sheet surface and RNA bases are important determinants of specificity in RNP proteins.[3] However, dsRBD proteins bind RNA through the face opposite the β-strand. Thus, it appears that the common $\alpha\beta$ structure may just reflect the high stability of this small protein folds and the ancient evolutionary origin of RNA-binding proteins. Recognition of single-stranded nucleotides is not the only mechanism by which specificity is achieved. RNA-binding proteins utilize the structural diversity of RNA to identify specific RNA structures to achieve effective discrimination. The variety of RNA structure defines many diverse, unique shapes for molecular recognition by proteins and other molecules that bind RNA. Formation of hydrogen bonds and van der Waals contacts with base functionality exposed on single stranded nucleotides allow additional fine tuning in the identification of the RNA target.

Despite intensive efforts, relatively few structures of RNA-protein complexes have been determined either by NMR or X-ray crystallography. In particular, NMR studies of protein-RNA complexes have been severely limited by the relatively large molecular weight of such complexes, despite some notable success.[4,5] The traditional method to determine structure of intermolecular complexes by NMR has been based on recording 1/2X filtered NOESY experiments on complexes containing isotopically labeled protein and unlabelled DNA or RNA.[6,7] The availability of isotopically labeled RNA provides the opportunity to do the reverse, i.e. mix labeled RNA with unlabelled protein, thereby facilitating extraction of distance constraints and the assignments of intermolecular NOE interactions to specific nucleic acid resonances.[8] NOE-based methods to determine structure of protein-nucleic acids complexes by NMR work well for systems characterized by tight binding (K_d 10^{-9} M), highly specific recognition and molecular weights of 40 kDa or less. However, these methods become increasingly less effective as the molecular weight increases and for systems (of great biochemical interest) characterized by weaker binding and poorer specificity. We have used several novel approaches to extend the range of RNA-protein structures that can be determined by NMR.

A first method uses paramagnetic spin labels to extract long-range intermolecular distance information in protein-RNA complexes (Ramos and Varani, submitted). This approach is based on electron-proton dipolar relaxation and is applicable also to DNA-protein complexes. We have inserted proxyl paramagnetic spin labels at specific sites on an RNA substrate. The presence of the unpaired electron on the nitroxide spin label increases the relaxation of NMR resonances in their vicinity. This effect can be quantitated by recording heteronuclear correlation spectra of a sample containing spin-labeled RNA and isotopically labeled protein, and is proportional to the inverse sixth power of the distance between the label and the reporter nucleus. This approach was applied to the complex between Drosophila Staufen protein and double stranded RNA. Staufen is an intensively studied protein containing multiple dsRBD domains, and its interaction with the bicoid mRNA 3'-untranslated region represents a paradigm in developmental biology. Selected resonances in the protein are broadened by the spin label and map to specific regions of the protein. We have been able to collect 12 long range distance constraints (up to approximately 15 Å) between the RNA and the protein from two test samples, using ^{15}N-labeled protein and sample concentrations of only 0.05 mM. The sensitivity of the method is very high and this approach should remain effective at very high molecular weight, particularly using random fractional deuterated samples and/or line-narrowing techniques.[9] This technique can provide additional, longer-range information for systems that can be studied using NOE-based methods, but can also be applied to complex assemblies with molecular weight well in excess of the current limits for NMR structure determination.

A second approach we have employed is based on recording residual dipolar interactions. We have followed the approach introduced by Bax and coworkers to induce partial orientation in biological samples by immersing the sample in a liquid crystalline phase.[10,11] We have recorded coupled heteronuclear spectra of samples containing either labeled protein or labeled RNA, and measured residual dipolar splittings in dilute liquid crystalline solutions. In order to test that the method could work with RNA-protein complexes as well, we have measured residual dipolar splitting for the 22 kDa complex of U1A protein, the structure of which has been determined by us.[4,5,8] The introduction of these additional constraints did not improve the structure of the RNA-bound protein in any significant way, but the structure was fully consistent with all recorded dipolar splittings.

The greatest limitation of NMR-derived structures of RNA and RNA-protein complexes is the long-range order of the RNA in the structure. Locally the RNA conformation is nearly as well defined as a protein structure, but the paucity of long-range NOE distance constraints in RNA (as

compared to proteins) makes the overall structure much less precisely determined.[8,12] In order to overcome this limitation, we have measured residual dipolar splittings for the RNA component of the U1A complex. Since there are relatively few NH resonances in an RNA molecule, we measured both NH and CH residual dipolar couplings. We were only able to record about 40 residual dipolar splittings due to the severe spectral overlap in the sugar region and to line broadening due to increased aggregation of the complex in the liquid crystalline solution. We are currently refining the structure using these additional constraints to verify whether their introduction leads to a significant improvement in the long-range definition of the RNA structure.

We have also used residual dipolar couplings to determine the relative orientation of protein and RNA components in the dsRBD-RNA complex. Residual dipolar couplings are positive for NH residues located on β-sheet residues, and negative for residues from α-helical regions. NH and CH groups on the RNA bases are approximately perpendicular to the long-axis of the RNA double helix and he residual dipolar coupling constants are also positive. Therefore, the two α helices that determine the long axis of this elongated protein, must be roughly parallel to the long axis of the RNA double helix. In this way, we have been able to determine the relative orientation of protein and RNA in the structure of this complex.

The human U1A protein represents the best understood example of RNA recognition. U1A is a component of the RNA splicing machinery and the prototypical member of the RNP protein superfamily.[13] U1A recognizes a stem-loop within U1 snRNA during splicing and an internal loop element during regulation of polyadenylation of the mRNA coding for U1A itself. Crystallographic [3] and NMR [4] structures of U1A alone and in complex with both target sites have provided a very close insight into how specificity is achieved. U1A recognizes the identity of exposed single stranded nucleotides through hydrogen bonding and van der Waals interactions, but selectivity against non-cognate RNAs is provided primarily by fit between the protein structure and the unique shape and charge distribution of the RNA.

A further level of biological selectivity, beyond RNA-binding, is provided by RNA-dependent protein-protein interactions. U1A autoregulates its own production by forming an RNA-dependent interaction with Poly(A) polymerase (PAP), the enzyme responsible for formation of the mature 3'-end of almost all eukaryotic mRNAs.[14] PAP is an essential enzyme and its activity must not be indiscriminately affected by U1A. In the 40 kDa ternary complex between two U1A proteins and the U1A polyadenylation regulatory element RNA, interaction with RNA induces a conformational rearrangement in U1A. The conformational shift allows protein-protein interactions to occur between U1A monomers to allow cooperative binding, and a productive interaction with the Poly(A)-polymerase enzyme. Protein and RNA conformational

rearrangements are very common during molecular recognition and central to the formation of specific intermolecular complexes. The conformational changes occurring upon binding can activate RNA-binding proteins and allow the formation of productive interactions with other proteins or RNAs.

The methods described in the previous pages have also been applied to study this complex multimolecular assembly. We have used site-directed mutagenesis to introduce Cysteine residues at specific sites within U1A; the sites were chosen to correspond with flexible loops of the protein and to be close enough to the intermolecular protein-protein interface, but not at the interface itself. Proxyl spin labels were attached to the protein and heteronuclear correlation spectra were recorded on a complex containing 50% isotopically labeled U1A, 50% spin-labeled U1A and its RNA substrate. Statistically, there are three species in the NMR tube; two correspond to spin-labeled U1A with ^{15}N-labeled U1A and another contains just ^{15}N-labeled U1A. Residues of the ^{15}N-labeled protein in proximity (<25 Å) of the spin label experience a decrease in T_2, resulting in broadening of the signal and loss of intensity in heteronuclear correlated spectra. The signal is recovered upon reduction of the unpaired electron. Thus, comparison of spectra of spin-labeled and reduced U1A complex, recorded on the same sample before and after reduction, immediately identifies residues that are in close proximity to the spin label. In this way, we have been able to obtain approximately 15 protein-protein intermolecular long-range distance constraints, that are currently been used to determine the structure of this complex of nearly 40 Kda.

RNA-protein recognition events regulate gene expression and are essential for the replication of many human and animal pathogens. Regulatory RNA elements are remarkably conserved in structure and sequence, highlighting their essential role in the viral life cycle. The progress in understanding the structural basis for RNA recognition suggest that it will be possible to disrupt these essential interactions by targeting the RNA regulatory elements. Combinatorial chemistry approaches have yielded compounds that disrupt trans-activation in HIV-1 both *in vitro* and *in vivo*, and that shut off viral replication in human lymphocytes.[15] The challenge ahead is to use our growing understanding of the structural and thermodynamic basis of RNA recognition to improve the selectivity and affinity of these compounds.

REFERENCES

1. G. Varani, *Acc. Chem. Res.* **30**, 189 (1997).
2. K. Nagai, *Curr. Op. Struct. Biol.* **6**, 53 (1996).
3. C. Oubridge, N. Ito, P.R. Evans, C. H. Teo, K. Nagai, *Nature* **372**, 432 (1994).
4. F. H. T. Allain, C. C. Gubser, P. W. A. Howe, K. Nagai, D. Neuhaus, G. Varani, *Nature* **380**, 646 (1996).
5. F. H. T. Allain, P. W. A. Howe, D. Neuhaus, G. Varani, *EMBO J.* **16**, 5764 (1997).
6. Y. Q. Qian, G. Otting, M. Billeter, M. Müller, W. Gehring, K. Wüthrich, *J. Mol. Biol.* **234**, 1070 (1993).
7. G. Otting, K. Q. Wüthrich, *Rev. Biophys.* **23**, 39 (1990).
8. P.W. A. Howe, F. H. T. Allain, G. Varani, D. Neuhaus, *J. Biomol. NMR* **11**, 59 (1998).
9. K. Pervushin, R. Riek, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **94**, 12366 (1997).
10. N. Tjandra, A. Bax, *Science* **278**, 1111 (1997).
11 N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, A.Bax, *Nature Struc. Biol.* **4**, 732 (1997).
12 F. H. T. Allain, G. Varani, *J. Mol. Biol.* **267**, 338 (1997).
13 G.Varani, K. Nagai, *Ann. Rev. Biophys. Biomol. Struct.* **27**, 407 (1998).
14 S. I. Gunderson, K. Beyer, G. Martin, W. Keller, W. C. Boelens, I. W. Mattaj, *Cell* **76**, 531 (1994).
15 F. Hamy, E. R. Felder, G. Heizmann, J. Lazdins, F. Aboul-ela, G. Varani, J. Karn, T. Klimkait, *Proc. Natl. Acad. Sci. USA* **94**, 3548 (1997).