Commentary

# Delivering messages from the 3′ end

**Gabriele Varani***

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom

RNA-binding proteins perform many essential functions during basal gene expression and its regulation through their RNA-binding activity mediated by conserved domains (1, 2). The majority of these domains belong to a dozen or so RNA-binding motives that are now well characterized structurally and biochemically (3, 4). Their function is to target the protein to specific cellular RNAs: prem-RNAs, mature mRNAs, or components of ribonucleoproteins such as the ribosome and the spliceosome. Many RNA-binding proteins also contain additional modules, so-called "auxiliary" domains (5) that mediate other aspects of their function (Fig. 1). These misnamed "auxiliary" domains often represent the business end of the protein, mediating the enzymatic activity of enzymes that operate on RNA (6, 7) or protein–protein interactions (8–10). The last few years have seen remarkable progress in the elucidation of the structure and RNA recognition principles of RNA-binding domains. In contrast, "auxiliary" domains have so far largely escaped the attention of structural biologists, but not of molecular biologists who have been dissecting their function. Until now, questions such as the structure of auxiliary domains or how they mediate protein–protein interactions have largely remained unanswered. In the present issue of PNAS, these questions are addressed in two manuscripts reporting the NMR (11) and x-ray crystallographic (12) structures of the C-terminal domain of human poly(A) binding protein and of a highly homologous domain derived from the human hyperplastic discs (HYD) protein.

The cytoplasmic poly(A) binding protein (PABP) is the product of a highly conserved, essential gene. PABP becomes associated with mature mRNAs after export to the cytoplasm and forms a stable ribonucleoprotein complex on the 3′ end of the mature mRNAs. A function of this complex is to protect mRNAs from degradation, which requires shortening of the poly(A) tail. This was surprising, because degradation proceeds in the 5′→3′ direction and requires the activity of an enzyme that removes the methylated guanosine analogue that marks the 5′ end of mature mRNAs (13–15). It has also been known for about 10 years that the poly(A) tail synergistically promotes initiation of protein synthesis with the cap (16). Messenger RNAs that are both capped and polyadenylated are translated much more efficiently *in vitro* than mRNAs that are only capped or polyadenylated (17, 18). This synergy is likely to be even more important *in vivo*, because translational initiation factors are limiting. For example, the expression of maternal mRNAs that are under translational control in early development is activated by developmental signals through the stimulation of the activity of a cytoplasmic enzyme that elongates the poly(A) tail (19). How can translational initiation, which requires recognition of the cap at the 5′ end of the mRNA and 5′→3′ scanning to find the initiation codon, be affected by the length of the poly(A) tail and by the poly(A)–PABP complex? How does the PABP–poly(A) complex assembled on the 3′ end of mRNAs affect events that take place at the 5′ end?

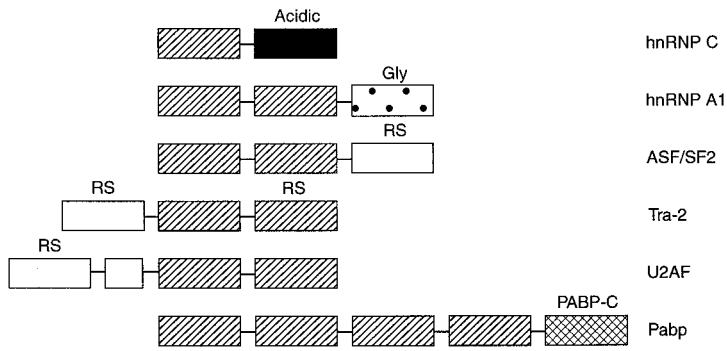


**Fig. 1.** RNA-binding proteins have modular structure and contain auxiliary domains that mediate protein–protein interactions, as well as highly conserved RNA-interaction domains such as the RRM.

These questions began to be addressed by the discovery that yeast PABP interacts with a key component of the translational initiation apparatus (20), the adaptor protein eIF-4G (16). Later studies identified additional interacting partners for human PABP among other proteins involved in translational control and mRNA stability, such as Paip1 (PABP-interacting protein 1) and Paip2 (21). The long-observed circularization of mRNAs was shown to be mediated by the eIF-4G-PABP interaction (22), providing a physical mechanism for the 3′–5′ communication and a framework to understand the functional interaction between opposite ends of mRNAs. The stability, localization, and efficiency of protein synthesis for individual mRNAs all depend on signals located within the 3′-untranslated region (3′-UTR) of mRNAs, just upstream of the Poly(A) tail (23). The observation of a physical interaction between the PABP–poly(A) complex and the protein complex assembled at the 5′ end of the mRNA also provided a framework to understand how signals within the 3′-UTR affect gene expression in many diverse functional contexts. The two structures presented here (11, 12) and the seminal structure of the PABP–poly(A) complex (24), provide some of the structural insight needed to understand how this communication occurs.

All PABP proteins from yeast to human share the same domain structure, with four RNA-binding domains within the N-terminal half of the protein and a conserved C-terminal domain that is the subject of the two articles presented in this issue of PNAS (Fig. 1). The RNA-binding domains belong to the RRM superfamily, by far the largest RNA-binding protein family and also the best understood (3). Each domain is individually conserved: yeast RRM1, for example, is more similar to human RRM1 than to yeast RRM2, and so on. Systematic biochemical and genetic studies identified RRM1 and RRM2 as

See companion articles on pages 4409 and 4414.

*E-mail: gv1@mrc-lmb.cam.ac.uk.



www.pnas.org/cgi/doi/10.1073/pnas.091108098

responsible for the RNA-binding activity of PABP (25, 26) and for mediating its effect on translational initiation (27). The structure of the complex of RRM1–RRM2 bound to Poly(A) revealed the molecular basis for the specificity of the PABP–Poly(A) interaction (24) and how the opposite face of RRM1–RRM2, when bound to Poly(A), forms a phylogenetically conserved surface that defines the eIF-4G-binding surface of PABP (24, 28). The two structures presented here provide additional insight into PABP structure/function by revealing the structure of its C-terminal domain, by defining interaction surfaces with protein components of the translational apparatus, and by discovering some unexpected homologies.

The carboxy-terminus domain of PABP contains a conserved, proteolitically stable domain (PABP-C) of ≈60–70 aa (Fig. 1) that is essential for yeast normal growth (27). PABP-C is a protein–protein interaction domain: it interacts with Paip1, Paip2, and eukaryotic release factor 3 (eRF3), another component of the translational apparatus; with hnRNPE, a factor involved in the control of mRNA stability, and with a viral RNA polymerase. In higher eukaryotes, PABP and Paip1 are part of a multiprotein complex that stabilizes the c-fos mRNA against degradation. Kozlov and colleagues report in this issue of PNAS (11) the structure of PABP-C determined in solution by using NMR. Because this same domain failed to yield suitably diffracting crystals, Deo and coworkers (12) scanned the sequence database for homologues, identifying a closely related sequence within the so-called human hyperplastic disk protein (HYD), a ubiquitin ligase, and determined its structure. PABP-C represents a four-helical right handed supercoil with no homologues in the protein data bank. The NMR investigation of the complex of PABP-C with a peptide derived from Paip2, one of several proteins that interact with this domain, identified a hydrophobic interacting patch (11). This putative protein–protein interaction surface is highly conserved (12).

The similarities with the component of the translational initiation apparatus responsible for delivering the 5′ end of mRNAs to the ribosome, the cap-binding protein eIF-4E, are striking. eIF-4E interacts with the adaptor protein eIF-4G and a set of proteins, 4E-BPs, that inhibit the 4E–4G interaction. The structure of the complex between eIF-4E and peptides derived from the 4E-BPs revealed that this protein–protein interaction occurs on the helical surface of eIF-4E. The 4E-binding peptides are unfolded (29), but they form a short $\alpha$-helix when bound to eIF-4E (30). The NMR study presented here shows that Paip2 is also unfolded in the absence of PABP-C, but forms a helical structure on binding to PABP-C. Each of these proteins provide examples of the growing list of proteins that function in the genuinely unfolded state when isolated, but fold up on binding their intracellular target. Whether this stimulating observation is also functionally significant remains unclear.

The observation that PABP-C is an independently folded protein interaction domain connected through a proteolytically sensitive linker to the four RRMs at the N-terminal half of PABP suggests a model for the function of the PABP–Poly(A) complex (11). The two N-terminal RRMs of PABP anchor the protein to the 3′ end of the mRNA through a direct interaction with the Poly(A) tail (24). The $\alpha$-helical surface opposite to RNA binding remains available to interact with eIF-4G. Because the 4G-interacting surface is optimally formed when PABP is bound to Poly(A), binding to Poly(A) favors the interaction with eIF-4G, thereby enhancing the affinity of polyadenylated mRNAs for the apparatus responsible for initiation of protein synthesis. The functional roles of RRM3 and RRM4, which are similar to RRM1 and RRM2, respectively, remain unclear. They may play a role in increasing the affinity of the PABP–Poly(A) interaction and in forming a complex on the poly(A) tail containing multiple copies of PABP. In this model, PABP-C is available to form protein–protein interactions with other components of the translation or mRNA stability apparatus. Interactions with eRF3, hnRNP-E, Paip1, and Paip2 have already been established, but other PABP-C-interacting factors may yet be discovered. Thus, the PABP–Poly(A) complex forms a scaffold on which to assemble, presumably in a regulated fashion, functional protein complexes on the 3′ end of mature mRNAs. Perhaps a primary role for this complex is to mediate the communication between opposite ends of a mRNA, thereby allowing the information encoded with the 3′-UTR of mRNAs to be delivered where it can be used.

The homology with HYD protein, a ubiquitin ligase that also interacts with Paip1 (12), suggests that PABP-C may also target PABP for degradation by targeted protein destruction. If this hypothesis is proven, then it would provide the first example of regulation of protein synthesis by this mechanism (12). Kozlov and colleagues (11) make yet another surprising connection by identifying a conserved sequence within proteins that interact with PABP-C. The PABP-C-interacting motif was identified in ataxin-2 (and other related proteins), a protein associated in type-2 spinocerebellar ataxia, a familial neurodegenerative disease of unknown molecular mechanism. The interaction between ataxin-2 and PABP-C suggests a connection between ataxia and the control of mRNA stability or translatability, and provides a new avenue to dissect the molecular mechanisms leading to neurodegeneration.

COMMENTARY


1. Burd, C. G. & Dreyfuss, G. (1994) *Science* **265,** 615–621.
2. Mattaj, I. W. (1993) *Cell* **73,** 837–840.
3. Varani, G. & Nagai, K. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27,** 407–445.
4. Draper, D. E. (1999) *J. Mol. Biol.* **293,** 255–270.
5. Biamonti, G. & Riva, S. (1994) *FEBS Lett.* **340,** 1–8.
6. Liu, Y., Lei, M. & Samuel, C. E. (2000) *Proc. Natl. Acad. Sci.* **97,** 12541–12546.
7. Elela, S. A., Igel, H. & Ares, M. J. (1996) *Cell* **85,** 115–124.
8. Wu, J. Y. & Maniatis, T. (1993) *Cell* **75,** 1061–1070.
9. Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Lührmann, R., Garcia-Blanco, M. A. & Manley, J. L. (1994) *Nature (London)* **368,** 119–124.
10. Manley, J. L. & Tacke, R. (1996) *Genes Dev.* **10,** 1569–1579.
11. Kozlov, G., Trempe, J.-F., Khaleghpour, K., Kahvejian, A., Ekiel, I. & Gehring, K. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4409–4413. (First Published April 13, 2001; 10.1073/pnas.071024998)
12. Deo, R. C., Sonenberg, N. & Burley, S. K. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 4414–4419. (First Published April 3, 2001; 10.1073/pnas.071552198)
13. Sachs, A. B. (1993) *Cell* **74,** 413–421.
14. Beelman, C. A. & Parker, R. (1995) *Cell* **81,** 179–183.
15. Caponigro, G. & Parker, R. (1995) *Genes Dev.* **9,** 2421–2432.
16. Sachs, A. B., Sarnow, P. & Hentze, M. W. (1997) *Cell* **89,** 831–838.
17. Preiss, T. & Hentze, M. W. (1998) *Nature (London)* **392,** 516–520.
18. Otero, L. J., Ashe, M. P. & Sachs, A. B. (1999) *EMBO J.* **18,** 3153–3163.
19. Sheets, M. D., Wu, M. & Wickens, M. (1995) *Nature (London)* **374,** 511–516.
20. Tarun, S. Z. J. & Sachs, A. B. (1996) *EMBO J.* **15,** 7168–7177.
21. Craig, A. W. B., Haghigat, A., Yu, A. T. K. & Sonenberg, N. (1998) *Nature (London)* **392,** 520–523.
22. Wells, S. E., Hillner, P. E., Vale, R. D. & Sachs, A. B. (1998) *Mol. Cell* **2,** 135–140.
23. Wickens, M., Anderson, P. & Jackson, R. J. (1997) *Curr. Opin. Genet. Dev.* **7,** 220–232.
24. Deo, R. C., Bonanno, J. B., Sonenberg, N. & Burley, S. K. (1999) *Cell* **98,** 835–845.
25. Burd, C. G., Matunis, E. L. & Dreyfuss, G. (1991) *Mol. Cell. Biol.* **11,** 3419–3424.
26. Kühn, U. & Pieler, T. (1996) *J. Mol. Biol.* **256,** 20–30.
27. Kessler, S. H. & Sachs, A. B. (1998) *Mol. Cell. Biol.* **18,** 51–57.
28. Sachs, A. & Varani, G. (2000) *Nat. Struct. Biol.* **7,** 356–360.
29. Fletcher, C. M., McGuire, A. M., Gingras, A.-C., Li, H., Matsuo, H., Sonenberg, N. & Wagner, G. (1998) *Biochemistry* **37,** 9–15.
30. Marcotrigiano, J., Gingras, A.-C., Sonenberg, N. & Burley, S. K. (1999) *Mol. Cell.* **3,** 707–716.