ELSEVIER

FEBS Letters 340 (1994) 1–8

FEBS 13694

FEBS LETTERS

Minireview

New insights into the auxiliary domains of eukaryotic RNA binding proteins

Giuseppe Biamonti*, Silvano Riva

Istituto di Genetica Biochimica ed Evoluzionistica del CNR, Via Abbiategrasso 207, 27100 Pavia, Italy

Received 6 January 1994

Abstract

Eukaryotic RNA binding proteins (RBP) are key players in RNA processing and in post-transcriptional regulation of gene expression. By interacting with RNA and other factors and by modulating the RNA structure, they promote the assembly of a great variety of specific ribonucleoprotein complexes. Many RBPs are composed of highly structured and conserved RNA binding domains (RBD) linked to unstructured and divergent auxiliary domains; such modular structure can account for a multiplicity of interactions. In this context, the auxiliary domains emerge as essential partners of the RBDs in both RNA binding and functional specialisation. Moreover, the determinants of biologically important functions, such as strand annealing, protein–protein interactions, nuclear localization and activity in in vitro splicing, seem to reside in the auxiliary domains. The structural and functional properties of these domains suggest their possible derivation from ancestral non-specific RNA binding polypeptides.

Key words: RNA binding protein; Auxiliary domain; hnRNP protein; SR protein; RNA–protein complex

1. Introduction

A great deal of gene expression regulation in mammalian cells acts on the flow of post-transcriptional events that starts with the release of newly synthesised pre-mRNA from the active chromatin in the form of a ribonucleoprotein fibre, continues with processing in the nucleus, transport of mature mRNA in the cytoplasm and ends with translation into proteins. During each of these steps the RNA is bound to a plethora of proteins. Contrary to DNA, which is a relatively passive substrate of *trans*-acting factors, RNA participates as an active protagonist in its own fate. In fact each step of the pre-mRNA processing (capping, splicing, 3′ end formation etc.) entails the formation of specific RNA–protein assemblies involving both *trans*-acting interactions with proteins and ribonucleoprotein particles and *cis*-acting interactions within the RNA itself. The sequence diversity and structural versatility of RNA can account for an enormous number of specific interactions. Thus, the formation of specific RNA–protein complexes should be viewed as a dynamic process whereby RNA sequence and conformation direct the binding of *trans*-active factors which, upon binding, can in turn modulate the RNA structure towards an appropriate conformation. Along this line, it is tempting to consider certain ribonucleoprotein complexes as the modern version of the catalytic RNAs that might have populated a primordial 'RNA world' and some of the proteins that nowadays associate with the RNAs as 'enhancers' of the RNA potentialities [1]. This type of 'scenario' has in fact been invoked for the spliceosome [2,3] and for the ribosome [4].

In this perspective the identification and molecular characterisation of the RNA binding proteins (RBPs) is of outmost interest even if the number and diversity of such proteins constitutes a formidable experimental challenge. An initial rationalisation of the whole field was achieved through the identification of specific nuclear RNA-protein assemblies (such as the hnRNP particles, the snRNP and more recently the spliceosome complex) and the molecular characterisation of their protein constituents. The results of these studies revealed that many RBPs can be grouped into families and sub-families on the basis of common structural and functional domains [5–7]. Such domains, which are often conserved in evolution, have in turn been used as diagnostic motifs to identify other proteins and to expand the respective families [8]. One such motif was first identified through the comparison of the primary sequences of two nuclear proteins: the poly(A)-binding protein (PABP) and the hnRNP protein A1 [9] and was then found in other proteins involved in different steps of the RNA processing. This motif termed RBD from RNA binding domain (but

*Corresponding author. Fax (39) (382) 422 286. E-mail: BIAMONTI%39162@IMICLVX.BITNET

SSDI 0014-5793(94)00078-A

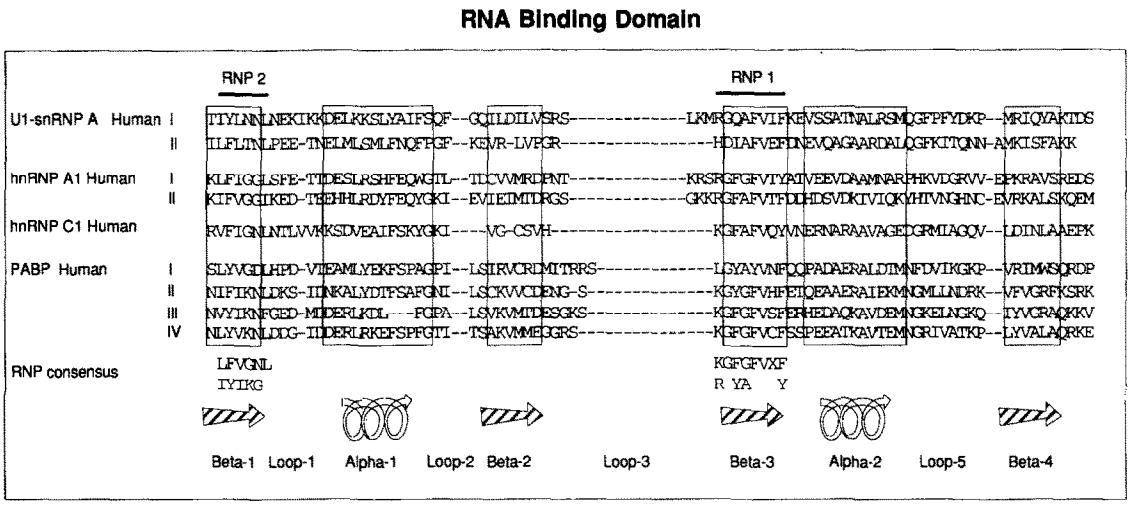

Fig. 1. Alignment of selected RBD sequences from four human RBPs. The common tertiary folding and the RNP 1 and RNP 2 consensus are outlined. For a more complete compilation and further details see [6].

also RNA recognition motif (or RRM [10]) or RNP 80 [11] motif by some authors) consists of 80–90 conserved amino acids containing two stretches of 8 and 6 highly conserved residues called RNP 1 and RNP 2 respectively [12] (see Fig. 1). Several lines of evidence point to the RBD domain as being one of the diagnostic determinants of RNA binding [8]. The finding that specific residues in the RNP sequences make direct contacts with nucleic acid [13] brings further support to this contention, even if the involvement of flanking sequences seems likely, at least in some cases [6,11]. The tertiary structure of two RBDs, one from protein U1A of U1 snRNP and one from the hnRNP protein C1, was determined by X-ray crystallography and NMR, respectively [14,15]. In both cases the structure consists of 4 anti-parallel β-strands forming a β-sheet connected by 2 α-helices on one side. On the basis of these and other data a model was devised by which the β-sheet constitutes a binding surface where the bound RNA is exposed to the solvent in a configuration available for other interactions. It is now generally accepted that the RBD is an important determinant of RNA binding present in a great number of proteins and its importance is strengthened by the observation that, at least in some cases, it contains also the determinant for sequence/structure specificity. Thus the RBD could be envisaged as the 'RNA world' counterpart of DNA binding motifs such as, for example, the helix-loop-helix or Zn-finger structures of transcription factors. It is interesting to note that the analogy between some DNA and RNA binding proteins extends further to overall structural organisation which in both cases is a modular assembly of different domains [16,17]. Moreover, much in the same way as transcription factors contain different activating domains linked to similar DNA binding motifs, the RBPs of the RBD family are also characterised by a variety of auxiliary domains. Although the function of these domains is still largely unknown, increasing evidence points to them as important determinants in the formation of specific supramolecular complexes.

2. Structure and specificity of the RBD proteins

As mentioned above a distinctive feature of RBD proteins is a modular structure in which one or more RBD domains are associated with one or more auxiliary domains. Some proteins possess only one RBD and one auxiliary domain while others have multiple RBDs and auxiliary domains assembled in different ways [16]. While RBD domains are usually rather well conserved in sequence, the auxiliary domains are widely divergent. However, some proteins appear to have similar auxiliary domains and, on this basis, a classification can be proposed. A schematic representation of the structure of the most representative RBD polypeptides is shown in Fig. 2.

What is the role of the two types of domains? The molecular dissection of a few well-characterised RBPs and a number of in vitro studies performed with their recombinant counterparts have provided some answers but at the same time have raised new questions. In particular, the schematic view that considers the RBD domain as the main RNA binding determinant, relegating the auxiliary domains to undefined interactions, should be reconsidered (see below) and, moreover, a differentiation of functions between apparently similar RBDs has emerged. In fact, in the case of the yeast poly(A)-binding protein (PABP), containing 4 RBDs, initial in vivo studies indicated that only one of the 4 RBDs is essential for viability and at the same time sufficient for stable RNA binding in vitro [18]. However, the single RBDs are conserved in evolution [12] suggesting a specific role for each of them.

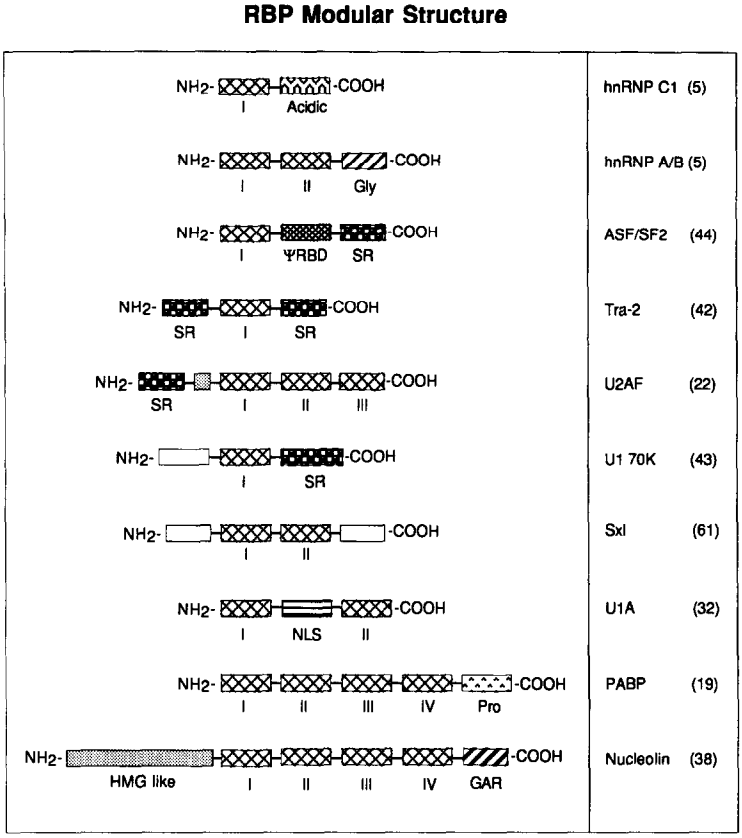

Fig. 2. Schematic representation of the modular structure of some representative RBD-type RNA binding proteins. Boxes represent structural and/or functional proteins domains. Conserved RBD domains are indicated by roman numbering (I–IV). Auxiliary domains are designated on the basis of their distinctive features (see text). White boxes represent uncharacterized regions. For further details see references given in parentheses in the figure.

On the other hand more recent in vitro studies with constructs consisting of different parts of the yeast PABP and with different pair-wise combinations of the 4 RBDs showed that none of the individual RBDs binds poly(A) specifically or efficiently while contiguous two-domains are required for efficient binding. Moreover, each pair-wise combination has distinct RNA binding activity and, while the two amino-terminal RBDs are dispensable for viability of yeast, the two carboxy-terminal ones are required [19]. It appears, therefore, that the 4 RBDs of PABP constitute a multifunctional polypeptide with at least two separable activities: the two amino-terminal RBDs, most likely involved in specific binding to the poly(A) tail, and the other two RBDs, which are engaged in interactions either with a different part of the same mRNAs or with other factors. Most likely the main role of the RBDs domains is to confer some kind of binding specificity, as also indicated by the fact that most RBD containing proteins (e.g. PABP, hnRNP A1, C1 and snRNA binding proteins, etc.) show a clear preference for certain sequences [6,11,20,21]. Moreover, in the case of proteins containing multiple RBDs (PABP, U2AF^{65}), each repetition confers different sequence specificity [19,22]. A rather convincing demonstration of sequence specificity was provided in the case of proteins U1A and U2B″ of snRNP U1 and U2, respectively. It was shown, in fact, that the replacement of 8 amino acids in the RBD of U1A with the corresponding ones of the homologous protein U2B″ confers to the former the sequence specificity of the latter [23]. On the other hand, as will be shown below, the RBD does not account for the whole binding capacity of the protein since some auxiliary domains contribute significantly to the overall RNA binding free energy, thus putting the overall functional organisation of RNA binding proteins into a new perspective.

3. Types of auxiliary domains

It is difficult to classify the auxiliary domains of the RBP proteins on the basis of their amino acid sequence and, moreover, the almost complete lack of functional data hampers the identification of distinctive structural elements. Interestingly, however, the analysis of certain RBP auxiliary domains reveals the presence of features commonly found in the activating domains of transcription factors. For example, hnRNP C has a very acidic auxiliary domain [24] while those of other proteins are

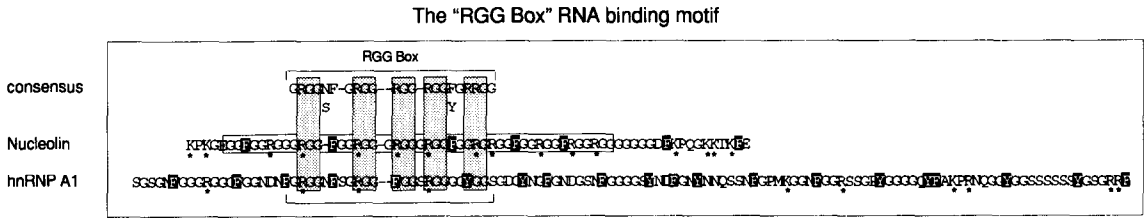

Fig. 3. Visualisation of the 'RGG box' consensus [52] in the auxiliary domains of hnRNP protein A1 and nucleolin. The GAR domain of nucleolin is boxed. Recurrence of aromatic residues (F,Y) is indicated. Asterisks mark the position of basic residues (K,R).

very rich in Q, P, A or M residues [9,18,25–30]. By analogy with transcription factors, it was suggested that these types of RBP domains could also be involved in protein–protein interactions. In this regard the auxiliary domain of the 64 kDa subunit of the human polyadenylation factor seems to fit this prediction since it contains a twelve-fold repetition of an MEARA/G consensus forming an extended α-helix which is likely to interact electrostatically with other proteins [31]. Altogether, our understanding of structure–function relationships in these domains is still scanty and complicated by the possibility that they might be involved in unpredictable functions. The case of protein U1A of snRNP U1 is emblematic of this situation. Its 100 amino acid long auxiliary domain was recently dissected by deletion mutagenesis and it was found that the entire region is required for the nuclear localisation of the protein [32]. In effect, this domain constitutes a completely new type of nuclear localisation signal. Moreover, the considerable size (110 amino acids) makes it unlikely that nuclear localisation is its only function.

It is not easy, therefore, to rationalise these results. However, in spite of these difficulties, two broad groups of RBP auxiliary domains can be envisaged, characterised by significant structural and functional similarities. The first group includes the so-called Gly-rich domains found in most basic hnRNP proteins, in nucleolin, in fibrillarin and in several yeast RBPs [5,33–36]. These domains have an abnormally high C content and are also characterised by the presence of more-or-less regularly interspersed basic and aromatic residues [33,37].

Many of these proteins have been studied in detail and, along with remarkable similarities, significant differences have emerged. For example, in human nucleolin, in fibrillarin and in yeast NSR1 protein, all involved in the processing of pre-rRNA, the Gly-rich region is organised into the so-called GAR domain [33] characterised by a repetition of the RGGXGGR sequence (where X is generally F or less frequently S, Y, A) that can form a series of β-turns producing a β-spiral that is likely to interact with RNA [38,39]. In contrast, no such defined structure can be seen in the Gly-rich domains of the basic hnRNP proteins [33] which, however, as will be shown below, have significant RNA binding activity further evidencing the difficulty of relating function to structure.

The second group comprises the so-called SR domains that are found in splicing factors of the SR family (SF2, SC35, etc.) [40], in U2AF [22], in splicing regulators of *Drosophila* (*tra* and *tra-2*) [41,42] and also in the 70 kDa protein of snRNP U1 [43]. It is now believed that the presence of an SR motif is diagnostic of activity in splicing [44]. In the SR protein family these domains constitute SR dipeptide runs the length of which is a distinctive feature of each member. On the other hand, in *Drosophila* factors *tra* and *tra-2*, the SR dipeptides are embedded in a basic milieu (K, R) [41,42], while in U1 70K protein, SR runs are mixed with RE and RD dipeptides and G hinges [45]. The SR region of the splicing factor U2AF is even more complex than the previous ones [22]. Thus, while the occurrence of SR domains in different proteins suggests a common biochemical function [44], their different sizes and spatial relationship with respect to neighbouring motifs, could determine the different functional roles. Such a possibility is further supported by the observation that the SR domains (as well as some Gly-rich domains) are subject to regulated alternative splicing and to post-translational modifications ([46–48], see below).

4. RNA binding properties of the RBPs auxiliary domains

The role of the RBD domains in RNA binding is confirmed by numerous experimental data [10,11,49,50]. Moreover, as previously mentioned, the tertiary structure of two such domains provides a plausible model for such interaction [14]. Surprisingly, however, the RBD domain is not the only portion of the protein that makes contacts with the RNA since, in many cases, the Gly-rich and the SR auxiliary domains were found to contribute significantly to the binding free energy of the entire protein. It was shown, for example, that the Gly-rich domain of hnRNP protein A1 contributes about half of the binding free energy of the entire protein [51]. Also the GAR domains of nucleolin and fibrillarin were extensively characterised; as already mentioned they can adopt a β-spiral structure able to bind and unfold the RNA [38]. Accordingly, a synthetic polypeptide consisting of a 7-fold repetition of the GNFGGGRGGNYGGSRG consensus, matching both A1 and nucleolin, binds RNA

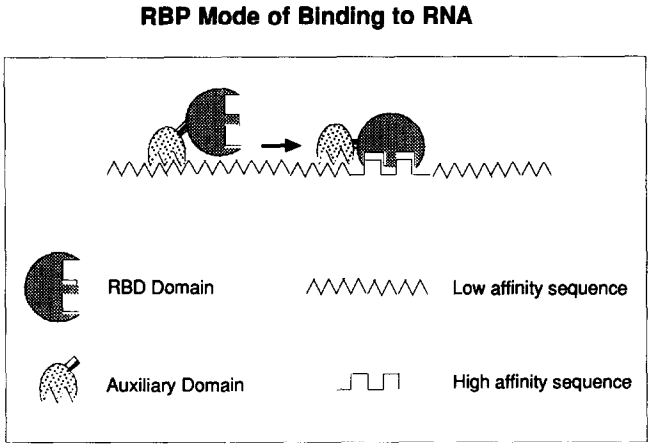

Fig. 4. A tentative model of RBD domain and auxiliary domain contribution to RNA binding and sequence specificity. The protein binds aspecifically to RNA with the auxiliary domain, then migrates (arrow) to an high affinity site where a conformational change could take place (for further details see text).

with high affinity and causes unfolding of bases [37]. Thus, in spite of substantial sequence divergence, the auxiliary domains of A1 and nucleolin must actually encode similar binding determinants. Such determinants have not been fully elucidated; interestingly however, both proteins show similarities to the so-called 'RGG box', a motif demonstrated to be the key RNA binding element in the hnRNP protein U [52] (see Fig. 3).

Concerning the SR domains, a direct demonstration of RNA binding was initially attained in the splicing factor, U2AF, where it was actually shown to contribute most of the RNA binding activity [53]. More recently, also the SR domain of SF2/ASF splicing factor was found to exhibit efficient RNA binding [44,45] while, surprisingly, the RBD domain of this protein binds RNA only inefficiently [54]. On the basis of several preliminary indications it seems likely that other SR domains will turn out to bind RNA. It should be noticed that both the Gly-rich and the SR auxiliary domains seem to bind RNA aspecifically. Thus, a general mechanism could be proposed by which the auxiliary domains play the important role of aspecifically 'anchoring' the protein to the RNA, allowing the cognate RBD domain to make more specific contacts (see Fig. 4).

5. Involvement of auxiliary domains in protein–protein contacts

As already mentioned, the auxiliary domains of RBPs have often been proposed to mediate protein–protein interactions, by analogy with the activating domains of transcription factors. In many cases such interactions are deduced simply on the basis of co-immunoprecipitation or co-localization experiments [55,56] while in others the evidence is more direct. For example, several results suggest that some SR-type splicing factors specifically interact with the splicing regulators *tra* and *tra-2* of *Drosophila* to form a splicing enhancer complex that controls alternative splicing of *doublesex* pre-mRNA [57]. Another likely candidate is protein U1A, a constituent of U1 snRNP that interacts with a specific segment of the snRNA. Interestingly, when such a segment is deleted, U1A still binds to the particles, presumably via contacts with the protein moiety [11]. A further example is the interaction between the two subunits (35 and 65 kDa) of the splicing factor, U2AF, where a detailed deletion analysis identified two protein domains with opposite charges, presumably mediating electrostatic interactions [58].

Recent experiments strongly suggest a direct interaction between protein PTB (pyrimidine tract binding; 55 kDa, identical to hnRNP I) and the splicing factor, PSF (which binds to the 3′ splice site and is required for the first step of spliceosome assembly [28]. In this case, the P/Q-rich domains of the two proteins are probably involved, as in the case of certain trandscription factors [59,60].

A role of the SR domain in protein–protein interactions is also deduced from experiments performed on the *Drosophila* splicing regulator, *sxl* (that lacks an SR domain), fused to a heterologous SR domain. Unlike the wild-type protein, the hybrid protein can substitute the splicing factor, U2AF, in bringing the U2 snRNP to the branch site on the *tra* gene pre-mRNA [61]. This result is interpreted by assuming that the SR domain confers the novel capacity to bind a snRNP particle.

The Gly-rich auxiliary domain of hnRNP protein A1 provides the most convincing evidence of protein binding capacity. Such a property was initially suggested by the observation that its presence greatly enhances the cooperativity of binding to the RNA compared to that of the protein lacking such a domain. (UP1) [62]. Moreover, a synthetic oligopeptide of 42 amino acids covering part of the Gly-rich domain was found to bind RNA cooperatively [51]. We have recently investigated the interactions mediated by the auxiliary domain of A1 with a more direct assay. For this, a hybrid protein consisting of the Gly-rich domain fused to GST and bound to glutathione–agarose beads was used to assay the retention of other proteins in affinity chromatography. We found that the entire A1 protein is specifically retained even at high salt concentration (0.4 M NaCl) while the protein lacking the Gly-rich domain (UP1) is not. Thus, this domain is absolutely required for protein–protein contacts (in preparation). Initial deletion analysis indicated that protein binding determinants are distributed over the entire length of the domain and are probably identifiable in the periodic recurrence of aromatic residues, in accord with the hydrophobic nature of interaction. In the same type of experiments we observed that proteins such as the hnRNP basic proteins A2, B1, B2 and hnRNP C,

are efficiently retained by the auxiliary domain of A1 while other hnRNP proteins (e.g. hnRNP I) are not bound (in preparation). Obviously the notion that an auxiliary domain can establish specific contacts with itself and with other RBPs in the absence of RNA has important implications in explaining the assembly of specific multi-protein complexes on the pre-mRNA. Moreover, because of their plasticity and multiplicity, protein–protein interactions could also play an unexpected role in other processes, as discussed in the next sections.

6. Annealing activity

In the last few years it has become increasingly evident that many RBPs can efficiently promote intra-strand RNA annealing in vitro. Interestingly, such activity seems to involve primarily the auxiliary domains. In nucleolin, for example, re-annealing depends on the presence of the GAR domain which, incidentally, is also responsible for the helix-destabilizing activity of this protein [38,39,63]. Similarly, in hnRNP protein A1, the Gly-rich domain is entirely responsible for reannealing. Actually the isolated Gly-rich domain of A1 (without any RBD) is able to efficiently promote reannealing [64–66]. More recently other hnRNP proteins [67] and several SR proteins [53,68] have been shown to promote annealing. Protein A1, which is the best studied case, increases up to 500-fold the rate of renaturation [69]. Protein A1 acts both by lowering the annealing temperature and by shifting the DS/SS equilibrium toward the DS form. In fact, in the presence of A1 there is no complete duplex denaturation even at very high temperatures [69]. All these effects could be accounted for by the ability of the Gly-rich domain to establish protein–protein and protein–RNA interactions through basic and aromatic residues regularly interspersed in a flexible structure. In fact, the cooperative binding of many such flexible structures to the RNA can both cause the shielding of phosphates and the stabilisation of relatively open RNA structures in which nucleotides are exposed and available for interaction. In support of this mechanism is the observation that certain long-chain cationic detergents can mimic the reannealing activity of the Gly-rich domain [70]. In the case of A1, protein–protein interactions mediated by the Gly-rich domain could further accelerate renaturation by bringing the covered complementary strands in close proximity. Such interaction, however, although probably relevant, seems not strictly required. In fact, the isolated RBD domain of hnRNP protein C1, presumably not involved in protein–protein interactions, has limited but significant reannealing activity, indicating that RNA binding proteins could reanneal RNA strands simply by acting as 'chaperons' of RNA conformation [67].

As already mentioned, several splicing factors containing SR motifs exhibit strand-annealing activity [53,68] but only for the U2AF factor was such activity unequivocally attributed to the SR region [53].

What is the biological significance of the reannealing capacity of many RBPs? Although there is no evidence that such a reaction is actually relevant in vivo it is tempting to speculate that the hnRNPs and other RBPs, in addition to recruiting factors through protein–protein interactions, favour the correct folding of the RNA during processing by acting as chaperons and match makers. Indeed it is difficult to envisage the assembly of the spliceosome and the splicing of introns without invoking reactions of the kind described above. One significant example in this regard could be the ennealing of U4 and U6 snRNPs promoted by the yeast protein PRP24 [71].

Further support for the biological role of annealing activity derives from the observation that some RBPs (including hnRNP protein A1) can enhance hammerhead ribozyme catalysis in vitro [1]. Ribozymes act by recognising and annealing to an RNA target sequence the length of which usually determines a strong binding and a slow dissociation rate which limits the specificity for the substrate. Thus, it is conceivable that proteins able to accelerate RNA association and/or dissociation can enhance ribozyme activity and specificity. This further example of protein–RNA synergism brings new support to the idea that non-specific RNA binding peptides were introduced into a 'RNA world' early in evolution, probably to enhance the RNA potentialities. From this point of view, the auxiliary domains of RBP could be the molecular remnants of such aspecific RNA binding peptides to which more organised domains (RBD) were subsequently joined.

7. Nuclear localisation

Many RBD-type RBPs contain canonical nuclear localisation signals identified by homology with the SV40 large T-antigen motif. In some RBPs, however, no such signals are recognisable while, interestingly, atypical localisation determinants seem to be located in auxiliary domains. Several examples in this regard have already been reported. As previously mentioned, the entire auxiliary domain of protein U1A is implicated in the nuclear import of this protein [32]. The SR domains of *Drosophila* splicing regulators *tra* and *su(w^a)* are responsible not only for nuclear localisation but also for the typical 'speckled' intranuclear pattern, probably due to stable association with insoluble nuclear structures [72]. We have recently demonstrated by means of transfection experiments with appropriate constructs, that the Gly-rich domain of hnRNP protein A1 contains all the determinants for the nuclear import of this protein. Interestingly, such determinants overlap those of RNA binding and protein–protein interactions (paper in preparation). Although no precise definition of the minimal sequences

required for nuclear localisation was attained in any of these proteins, the above reported findings further expand the spectrum of activities of the auxiliary domains of RNA binding proteins and raise the possibility that a single RBP might be active in different steps and at different times during RNA processing.

8. Auxiliary domain modifications

The auxiliary domains of many RBPs undergo a number of post-transcriptional and post-translational modifications that modulate their structural and functional properties. For example, the SR dipeptides in the SR auxiliary domains are often phosphorylated, as demonstrated for several splicing factors [73] and for the 70 kDa protein of the U1 snRNP [74]. The effect of such modification is not yet fully understood. However, it was shown that U1 snRNP containing a phosphorylated 70 kDa protein is inactive in in vitro splicing reactions [74]. A protein kinase associated to the U1 snRNP and responsible for 70 kDa protein phosphorylation was recently described [75]. It was proposed that the same kinase could phosphorylate other spliceosome proteins such as SR proteins and splicing regulators. Many of these factors, in fact, are phosphorylated in vivo, as indicated by their aberrant electrophoretic mobility. In the case of the splicing factor, SF2/ASF, the most phosphorylated region (residues 200–224) is the same as that which is necessary and sufficient to provide the SR function in constitutive splicing assay [44,75]. Also, many hnRNP proteins (e.g. A/B, C, U) are phosphorylated in vivo and in some instances (e.g. hnRNP A2 and C) in a cell cycle-dependent fashion [76,77]. A specific kinase that phosphorylates hnRNP protein C was described [77] and it was proposed that specific phosphorylation/dephosphorylation steps could control the binding and the release of hnRNP C from complexes assembled on the pre-mRNA [78]. Interestingly, several experiments indicate that phosphorylation/dephosphorylation cycles could direct the sequential binding of splicing factors and snRNPs to the pre-mRNA and influence specific catalytic steps of the splicing reaction [79].

Concerning hnRNP A1, we showed that the in vitro phosphorylation of Ser^{199} in the Gly-rich domain by protein kinase A (PKA) reduces the affinity of the protein for nucleic acid and abrogates strand annealing activity [80]. It should be pointed out, however, that the site of in vivo phosphorylation of this protein has not yet been determined.

In addition to phosphorylation, another in vivo modification, i.e. the dimethylation of one or more R residues, is often observed in the Gly-rich domains of many basic hnRNP proteins and of nucleolin [34,81]. The effect of this modification is not known; it is possible, however, that, in analogy to what is observed with the TAT protein of HIV, it might influence RNA binding [82].

9. Conclusions

Most eukaryotic RNA binding proteins appear to be an assembly of different structural modules. One such module, the RBD domain, is a conserved highly structured RNA binding motif, widely shared among proteins involved in different steps of RNA processing. The RBD is responsible, at least in part, for the sequence–structure specificity of binding. Linked to the RBDs are highly variable modules, the so called auxiliary domains, that identify different groups of proteins. Contrary to the RBD, the auxiliary domains are largely unstructured and characterised by monotonous repetitions of distinctive amino acids. Quite surprisingly, however, the auxiliary domains appear to be key functional constituents of RNA binding proteins since they: (i) contribute to non-specific RNA binding and modulate RNA conformation; (ii) mediate the interactions with other proteins; (iii) determine the intracellular localisation of the proteins.

The binding specificity of RBDs and the plasticity of auxiliary domain interactions seem suitable to account for the assembly of specific ribonuclear complexes along the RNA. In the hypothesis of a transition from a primordial 'RNA world' to the present situation it is tempting to view the auxiliary domains as the descendants of the ancestral simple polypeptides that must have non-specifically bound the RNA to enhance its potentiality; the present form of RBP could therefore be the result of the addition later in evolution of more specific binding determinants such as the RBDs.

Acknowledgements: Work in our laboratory was supported by the Progetto Finalizzato 'Ingegneria Genetica' CNR (Rome).

References

[1] Tsuchihashi, Z., Khosla, M. and Herschlag, D. (1993) Science 262, 99–101.
[2] Ruby, S.W. and Abelson, J. (1991) Trends Gene 7, 79–85.
[3] Wassarman, D.A. and Steitz, J.A. (1992) Science 257, 1918–1925.
[4] Noller, H.F. (1991) Annu. Rev. Biochem. 60, 191–227.
[5] Dreyfuss, G., Matunis, M.J., Piñol-Roma, S. and Burd, C.G. (1993) Annu. Rev. Biochem 62, 289–321.
[6] Kenan, J.D., Query, C.C. and Keene, J.D. (1991) Trends Biochem. Sci. 16, 214–220.
[7] Mattaj, I.W. (1993) Cell 73, 837–840.
[8] Kim, Y.J. and Baker, B.S. (1993) Mol. Cell. Biol. 13, 174–183.
[9] Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. and Dreyfuss, G. (1986) Mol. Cell. Biol. 6, 2932–2943.
[10] Query, C.C., Bentley, R.C. and Keene, J.D. (1989) Cell 57, 89–101.
[11] Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. and Mattaj, I.W. (1989) EMBO J. 8, 4163–4170.
[12] Dreyfuss, G., Swanson, M.S. and Piñol-Roma, S. (1988) Trends Biochem. Sci. 13, 86–91.

[13] Merrill, B.M., Stone, K.L., Cobianchi, F., Wilson, S.H. and Williams, K.R. (1988) J. Biol. Chem. 263, 3307–3313.
[14] Nagai, K., Outbridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Nature 346, 515–520.
[15] Wittekind, M., Görlach, M., Friedrichs, M., Dreyfuss, G. and Mueller, L. (1992) Biochemistry 31, 6254–6265.
[16] Bandziulis, R.J., Swanson, M.S. and Dreyfuss, G. (1989) Genes Dev. 3, 431–437.
[17] Frankel, A.D. and Kim, P.S. (1991) Cell 65, 717–719.
[18] Sachs, A.B., Davis, R.W. and Kornberg, R.D. (1987) Mol. Cell. Biol. 7, 3268–3276.
[19] Burd, C.G., Matunis, E.L. and Dreyfuss, G. (1991) Mol. Cell. Biol. 11, 3419–3424.
[20] Buvoli, M., Cobianchi, F., Biamonti, G. and Riva, S. (1990) Nucleic Acids Res. 18, 6595–6600.
[21] Swanson, M.S. and Dreyfuss, G. (1988) Mol. Cell. Biol. 8, 2237–2241.
[22] Zamore, P.D., Patton, J.G. and Green, M.R. (1992) Nature 355, 609–614.
[23] Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J. and Mattay, I.W. (1990) Nature 345, 502–506.
[24] Swanson, M.S., Nakagawa, T.Y., LeVan, K. and Dreyfuss, G. (1987) Mol. Cell. Biol. 7, 1731–1739.
[25] Datar, K.V., Dreyfuss, G. and Swanson, M.S. (1993) Nucleic Acids Res. 21, 439–446.
[26] Lee, F.-J. and Moss, J. (1993) Biol. Chem. 268, 15080–15087.
[27] Matunis, M.J., Matunis, E.L. and Dreyfuss, G. (1993) Mol. Cell. Biol. 13, 6114–6123.
[28] Patton, J.G., Porro, E.B., Galceran, J., Tempst, P. and Nadal-Ginard, B. (1993) Genes Dev. 7, 393–406.
[29] Robinow, S., Campos, A.R., Yao, K.-M. and White, K. (1988) Science 242, 1570–1572.
[30] Sillekens, P.T.G., Habets, W.J., Beijer, R.P. and van Venrooij, W.J. (1987) EMBO J. 6, 3841–3848.
[31] Takagaki, Y., MacDonald, C.C., Shenk, T. and Manley, J.L. (1992) Proc. Natl. Acad. Sci. USA 89, 1403–1407.
[32] Kambach, C. and Mattaj, I.W. (1992) J. Cell. Biol. 118, 11–21.
[33] Girard, J.-P., Lehtonen, H., Caizergues-Ferrer, M., Amalric, F., Tollervey, D. and Lapeyre, B. (1992) EMBO J. 11, 673–682.
[34] Lapeyre, B., Amalric, F., Ghaffari, S.H., Rao, S.V.V., Dumbar, T.S. and Olson, M.O.J. (1986) J. Biol. Chem 261, 9167–9173.
[35] Lapeyre, B., Mariottini, P., Mathieu, C., Ferrer, P., Amaldi, F., Amalric, F. and Caizergues-Ferrer, M. (1990) Mol. Cell. Biol. 10, 430–434.
[36] Riva, S., Morandi, C., Tsoulfas, P., Pandolfo, M., Biamonti, G., Merril, B., Williams, K.R., Multhaup, G., Beyreuther, K., Werr, H., Henrich, B. and Schafer, K.P. (1986) EMBO J. 5, 2267–2273.
[37] Casas-Finet, J.R., Smith, J.D., Kumar, A., Kim, J.G., Wilson, S.H. and Karpel, R.L. (1993) J. Mol. Biol. 229, 873–889.
[38] Ghisolfi, L., Joseph, G., Amalric, F. and Erard, M. (1992) J. Biol. Chem. 267, 2955–2959.
[39] Ghisolfi, L., Kharrat, A., Joseph, G., Amalric, F. and Erard, M. (1992) Eur. J. Biochem. 209, 541–548.
[40] Zahler, A.M., Lane, W.S., Stolk, J.A. and Roth, M.B. (1992) Genes Dev. 6, 837–847.
[41] Amrein, H., Gorman, M. and Nöthiger, R. (1988) Cell 55, 1025–1035.
[42] Goralski, T.J., Edström, J.E. and Baker, B.S. (1989) Cell 56, 1011–1018.
[43] Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Lührmann, R. and Philipson, L. (1986) EMBO J. 5, 3209–3217.
[44] Zuo, P. and Manley, J.L. (1993) EMBO J. 12, 4727–4737.
[45] Krainer, A.R., Mayeda, A., Kozak, D. and Binns, G. (1991) 66, 383–394.
[46] Biamonti, G., Buvoli, M., Bassi, M.T., Morandi, C., Cobianchi, F. and Riva S. (1989) J. Mol. Biol. 207, 491–503.
[47] Buvoli, M., Cobianchi, F., Bestagno, M.G., Mangiarotti, A., Bassi, M.T., Biamonti, G. and Riva, S. (1990) EMBO J. 9, 1229–1235.
[48] Ge, H., Zuo, P. and Manley, J.L. (1991) Cell 66, 373–382.
[49] Görlach, M., Wittekind, M., Beckman, R.A., Mueller, L. and Dreyfuss, G. (1992) EMBO J. 11, 3289–3295.
[50] Lutz-Freyermuth, C., Query, C.C. and Keene, J.D. (1990) Proc. Natl. Acad. Sci. USA 87, 6393–6397.
[51] Nadler, S.G., Merrill, B.M., Roberts, W.J., Keating, K.M., Lisbin, M.J., Barnett, S.F., Wilson, S.H. and Williams, K.R. (1991) Biochemistry 30, 2968–2976.
[52] Kiledjian, M. and Dreyfuss, G. (1992) EMBO J. 11, 2655–2664.
[53] Lee, C.-G., Zamore, P.D., Green, M.R. and Hurwitz, J. (1993) J. Biol. Chem. 268, 13472–13478.
[54] Cáceres, J.F. and Krainer, A.R. (1993) EMBO J. 12, 4715–4726.
[55] Fu, X.-D. and Maniatis, T. (1992) Proc. natl. Acad. Sci. USA 89, 1725–1729.
[56] Fu, X.-D., Mayeda, A., Maniatis, T. and Krainer, A.R., (1992) Proc. Natl. Acad. Sci. USA 89, 11224–11228.
[57] Tian, M. and Maniatis, T. (1993) Cell 74, 105–114.
[58] Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I. and Green, M.R. (1992) Proc. Natl. Acad. Sci. USA 89, 8769–8773.
[59] Courey, A.J. and Tjian, R. (1988) Cell 55, 887–898.
[60] Tanese, N., Pugh, F. and Tjian, R. (1991) Genes Dev. 5, 2212–2224.
[61] Valcárcel, J., Singh, R., Zamore, P.D. and Green, M.R. (1993) Nature 362, 171–175.
[62] Cobianchi, F., Karpel, R.L., Williams, K.R., Notario, V. and Wilson, S.H. (1988) J. Biol. Chem. 263, 1063–1071.
[63] Sapp, M., Richter, A., Weisshart, K., Cazergues-Ferrer, M., Amalric, F., Wallace, M., Kirstein, M. and Olson, M. (1989) Eur. J. Biochem. 179, 541–548.
[64] Kumar, A. and Wilson, S.H. (1990) Biochemistry 29, 10717–10722.
[65] Munroe, S.H. and Dong, X.F. (1992) Proc. Natl. Acad. Sci. USA 89, 895–899.
[66] Pontius, B.W. and Berg, P. (1990) Proc. Natl. Acad. Sci. USA 87, 8403–8407.
[67] Portman, D.S. and Dreyfuss, G. (1994) EMBO J. (in press).
[68] Krainer, A.R., Conway, G.C. and Kozak, D. (1990) Genes Dev. 4, 1158–1171.
[69] Pontius, B.W. and Berg, P. (1992) J. Biol. Chem. 267, 13815–13818.
[70] Pontius, B.W. and Berg, P. (1991) Proc. Natl. Acad. Sci. USA 88, 8237–8241.
[71] Shannon, K.W. and Guthrie, C. (1991) Genes Dev. 5, 773–785.
[72] Li, H. and Bingham, P.M. (1991) Cell 67, 335–342.
[73] Zahler, A.M. Neugebauer, K.M., Stolk, J.A. and Roth, M.B. (1993) Mol. Cell. Biol. 13, 4023–4028.
[74] Tazi, J., Kornstädt, U., Rossi, F., Jeanteur, P., Cathala, G., Brunel, C. and Lührmann, R. (1993) Nature 363, 283–286.
[75] Woppmann, A., Will, C.L., Kornstädt, U., Zuo, U., Manley, J.L. and Lührmann, R. (1993) Nucleic Acids Res. 21, 2815–2822.
[76] Leser, G.P. and Martin, T.E. (1987) J. Cell Biol. 105, 2083–2094.
[77] Piñol-Roma, S. and Dreyfuss, G. (1993) Mol. Cell. Biol. 13, 5762–5770.
[78] Mayrand, S.H., Dwen, P. and Pederson, T. (1993) Proc. Natl. Acad. Sci. USA 90, 7764–7768.
[79] Mermoud, J.E., Cohen, P. and Lamond, A.I. (1992) Nucleic Acids Res. 20, 5263–5269.
[80] Cobianchi, F., Calvio, C., Stoppini, M., Buvoli, M. and Riva, S. (1993) Nucleic Acids Res. 21, 949–955.
[81] Najbauer, J., Johson, B.A., Young, A.L. and Aswad, D.W. (1993) J. Biol. Chem. 268, 10501–10509.
[82] Tao, J. and Frankel, A.D. (1992) Proc. Natl. Acad. Sci. USA 89, 2723–2726.