

# **RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence**

Grégoire Masliah<sup>1</sup>, Pierre Barraud<sup>1</sup> & Frédéric H.-T. Allain†

Institute of Molecular Biology and Biophysics, ETH Zurich, CH-8093 Zürich, Switzerland

<sup>1</sup> These authors contributed equally to this work

† Corresponding author: FH-T Allain, Institute of Molecular Biology and Biophysics, ETH Zurich, Schafmattstrasse 20, CH-8093 Zürich, Switzerland. Tel.: +41 44 633 3940, Fax: +41 44 633 1294. E-mail: [allain@mol.biol.ethz.ch](mailto:allain@mol.biol.ethz.ch)

## **ABSTRACT**

The double stranded RNA binding domain (dsRBD) is a small protein domain of 65-70 amino acids adopting an  $\alpha\beta\beta\alpha$  fold, whose central property is to bind to double stranded RNA (dsRNA). This domain is present in proteins implicated in many aspects of cellular life, including antiviral response, RNA editing, RNA processing, RNA transport and last but not least RNA silencing. Even though proteins containing dsRBDs can bind to very specific dsRNA targets *in vivo*, the binding of dsRBDs to dsRNA is commonly believed to be shape-dependent rather than sequence-specific. Interestingly, recent structural information on dsRNA recognition by dsRBDs opens the possibility that this domain performs a direct readout of RNA sequence in the minor groove, allowing a global reconsideration of the principles describing dsRNA recognition by dsRBDs. We review in this article the current structural and molecular knowledge on dsRBDs emphasizing the intricate relationship between the amino acid sequence, the structure of the domain and its RNA recognition capacity. We especially focus on the molecular determinants of dsRNA recognition and describe how sequence discrimination can be achieved by this type of domain.

## **Keywords**

Protein-RNA interactions; dsRBD; dsRBM; dsRNA; sequence specificity; protein-RNA recognition; nucleic acids recognition; structures; RNA minor groove; dsRBD extensions; RNA interference; RNA processing; RNA editing; nuclear retention ; nuclear localization signal.

## **Abbreviations**

dsRBD: double-stranded RNA binding domain

dsRNA: double stranded RNA

PKR: protein kinase RNA-activated

RHA: RNA helicase A

TRBP: HIV transactivation response RNA binding protein

PACT: PKR activator

ADAR: Adenosine deaminase acting on RNA

ILF3: Interleukin enhancer-binding factor 3

SPNR: Spindle perinuclear protein

DGCR8: DiGeorge syndrome critical region 8

HYL1: HYPONASTIC LEAVES1

## Introduction

Double-stranded RNA (dsRNA) is present in many biological processes. For example, many viruses carry their genetic information in the form of dsRNA, which may also be generated by the replication of single-stranded RNA viruses. Moreover, non-pathogenic cellular pathways constitute as well a massive source of dsRNA within the cell. Indeed, dsRNA elements can be formed by base pairing of complementary sequences within primary RNA transcripts. These dsRNA modules appear for example in untranslated regions of mRNAs [1,2], in mRNA precursors prior to intron removal where intronic sequences are found to base pair with complementary flanking exonic sequences [3], in stable RNAs and their precursors such as ribosomal RNAs and transfer RNAs [4], and last but not least in small RNA precursors such as short-interfering RNA (siRNA) and micro-RNA (miRNA) precursors [5-7]. Regarding the broad and different origins of dsRNA in the cell, dsRNA recognition plays critical roles in many cellular processes as diverse as response to viral infection, gene silencing through RNA interference pathways, regulation of translation, RNA processing, mRNA editing, export and localization (Table 1) [8-10].

In all these different cellular processes, dsRNA recognition is achieved by a superfamily of proteins having at least one double-stranded RNA binding domain (dsRBD; also referred to as dsRBM for double-stranded RNA binding motif). The dsRBD is one of the most abundant RNA binding domain after the RNA recognition motif (RRM) which is a well characterized single-stranded RNA binding domain [11] and zinc fingers which are best known for their ability to bind to DNA but which may also interact with dsRNA [12-14]. We also want here to mention that the dsRBD is not the only protein domain known to bind to dsRNA modules. Indeed, protein domains such as zinc fingers [12-14], SAM domains [15,16], Z-DNA/Z-RNA binding domains [17-19], or even RRM domains [20-23], are also known to recognize RNA modules containing dsRNA fragments, such as regular dsRNA fragments, left-handed Z-RNA helices or short RNA hairpins. This review concentrates on dsRBDs and readers interested in how other type of domains interact with particular dsRNA modules may refer to the existing literature on those domains (see references above).

Regarding the geometry of an A-form RNA helix and the nature of the chemical groups accessible to proteins in dsRNA [24-26], binding of dsRBDs to their dsR-

NA targets is commonly believed to be shape-dependent rather than sequence-specific [27-29]. However, *in vivo*, proteins containing dsRBDs bind to specific dsRNA targets [2,30] and despite a widespread significance in diverse cellular functions, our current understanding of how dsRBDs recognize these specific dsRNA targets remains imperfect [10,31]. Nevertheless, the recent increase of structural information on dsRNA recognition by dsRBDs allows a global reconsideration of the precepts describing dsRNA recognition by this class of domain.

We review here the current structural and molecular knowledge on the double-stranded RNA binding domain. As more comprehensive reviews on the function of dsRBD-containing proteins are available elsewhere [8-10,32], we focus in this article on the structural features of the dsRBD, the molecular determinants of dsRNA recognition and especially on how sequence discrimination can be achieved by this type of domain.

### **Structural characteristics of a dsRNA binding domain**

The dsRBD is a ~65-70 amino acids domain found in eukaryotic, prokaryotic and even viral proteins. This domain which specifically interacts with dsRNA (as compared with dsDNA or single-stranded nucleic acid molecules), was first recognized as a conserved functional domain in 1992 from sequence similarity between three different proteins: Staufen, a protein responsible for mRNA localization in *Drosophila*, human TAR-RNA binding protein (TRBP), a multi-functional protein first characterized as a an activator of HIV-1 gene expression, and *Xenopus leavis* RNA-binding protein A (Xlrpba), a homolog of human TRBP [33]. In the same study, database searches enabled the identification of several other proteins containing one copy of the domain. At the same time, different domains were characterized as dsRNA binding domains in proteins such as PKR, a dsRNA-dependent protein kinase [34-36], and human TRBP [1]. A sequence alignment of dsRBDs of diverse origins is presented on Figure 1. The sequence consensus is drawn below the alignment. Even if amino-acids are conserved along the entire domain, the last third at the C-terminal end is the most conserved part of the dsRBD. The first two third from the N-terminus are more divergent and some domains differing notably from the consensus in this

part have been referred to as type B dsRBDs and were shown to have rather low binding activity for dsRNA [33,37].

The first three-dimensional structures of dsRBDs, namely the *E. coli* RNase III dsRBD and the third dsRBD of *Drosophila* Staufen, were solved by solution NMR uncovering a mixed  $\alpha/\beta$  fold with a conserved  $\alpha\beta\beta\beta\alpha$  topology in which the two  $\alpha$ -helices are packed against a three-stranded anti-parallel  $\beta$ -sheet (Figure 2a-e) [38,39]. At the date of this review, the structure of about 30 dsRBDs have been determined from both NMR and X-ray crystallography (Table 2) [38-62]. These structures confirmed the conserved topology and the characteristic features of the dsRBD fold and uncovered the existence of some modest variations and some striking extensions to the canonical fold (see below).

The structure of the second dsRBD of Xlrpba constitutes the archetype of a canonical dsRBD domain [55]. This canonical dsRBD structure is represented on Figure 2a-b. The conserved residues matching the sequence consensus (> 40%) of the dsRBD sequence alignment of Figure 1 are labelled and shown as sticks on the structure. These residues appear to be conserved on the one hand for maintaining a stable hydrophobic core with the two  $\alpha$ -helices packed on the  $\beta$ -sheet surface (Figure 2c) and on the other hand for optimal dsRNA binding (Figure 2d):

(i) Conserved hydrophobic residues come from almost all secondary structured elements, namely  $\alpha 1$ ,  $\beta 1$ ,  $\beta 2$  and  $\alpha 2$  (Figure 1 and 2a-c). Aliphatic side-chains are mostly found in the two helices whereas aromatic rings are predominantly found in the  $\beta$ -strands. To be precise, aliphatic side chain include L6 and L9 in helix  $\alpha 1$  (residue numbering refers to the alignment of Figure 1), V39 and V41 in  $\beta 2$  and A58, A62, A63, A66, L67 and L70 in helix  $\alpha 2$ . Note that the small side chains of A58 and A62 in helix  $\alpha 2$  together with the conserved GxG motif at the end of the  $\beta 3$  strand allow a tight packing in this region with helix  $\alpha 2$  coming in remarkable close proximity with this part of the  $\beta$ -sheet, with for instance  $C\alpha$  carbons of G50 and A62 being less than 4 Å apart. In addition to the conserved aliphatic residues, two aromatic side chains are almost absolutely conserved in dsRBDs, namely Y21 in the  $\beta 1$  strand and F35 in the  $\beta 2$  strand (Figure 1 and 2a-c). These aromatic residues have been shown to be indirectly involved in RNA binding by maintaining key positively charged residues in

an optimal orientation for dsRNA binding [37,39,54,55] (see below). Indeed, mutations of these aromatic residues were reported to completely abolish dsRNA binding [37,39,54]. However, regarding the position of those aromatic rings, at the periphery of the dsRBD hydrophobic core, they most probably contribute in addition to the stability of the overall domain.

(ii) Systematic mutational analysis conducted in different dsRBDs uncovered three different regions important for dsRNA interaction [36,39,54,63-65]. These three regions are shown on Figure 1. Region 1 is located in helix  $\alpha_1$ , region 2 in the loop joining the  $\beta_1$  and  $\beta_2$  strands, and region 3 at the N-terminal tip of helix  $\alpha_2$ . In these three regions, conserved residues in the sequence consensus can therefore be explained by their involvement in dsRNA binding. To be precise, the side chain of E8 in helix  $\alpha_1$ , the GPxH motif in the  $\beta_1$ - $\beta_2$  loop and the positively charged residues in the KKxAK motif at the beginning of helix  $\alpha_2$  are highly conserved in dsRBDs as a consequence of their participation in dsRNA binding. Remarkably, the orientation of the side chains of K55 and K59 in the free protein is pre-organized for RNA binding [53,55]. The orientation of both side chains is stabilized by an extensive set of van der Waals interactions (Figure 2d). Namely, the first lysine interacts with three residues from the  $\beta$ -sheet whose side chains point down inside the hydrophobic core of the dsRBD: a residue in strand  $\beta_1$  (position 23), which is often a leucine or a valine, the conserved phenylalanine in strand  $\beta_2$  (position 35), and another less conserved residue in strand  $\beta_2$  (position 37). The third lysine interacts with a hydrophobic residue from helix  $\alpha_1$  (position 3), the aromatic cycle of the conserved tyrosine in strand  $\beta_1$  (position 21) and with the hydrophobic residue from strand  $\beta_2$  already mentioned (position 37). The third residue of the KKxAK motif is not conserved. It is exposed at the surface of the dsRBD and does not make any significant interaction with other part of the domain. The conserved alanine of the motif is packed against the  $\beta$ -sheet surface. A likely explanation for the conservation of this alanine is that a bulky side-chain at this position would cause steric clashes that would destabilize the  $\beta$ -sheet. In the third dsRBD of Staufen this alanine is replaced by a serine without causing any major rearrangements [41,54]. Note that molecular details of dsRNA recognition by these three conserved regions of the dsRBD will be precisely described in later paragraphs of this review.

### Some variations and extensions to the canonical dsRBD fold

Although dsRBD structures are in overall well conserved, some variations are found, especially in two regions of the domain. The first region of variability consists of the loop between helix  $\alpha 1$  and the  $\beta 1$  strand. This appears clearly in the alignment of Figure 1, where gaps are frequently needed to maintain a proper alignment in this region. In addition, the length of helix  $\alpha 1$  can sometimes be shorter like in *Yeast* RNase III, mammalian ADAR2 and *Drosophila* ADAR dsRBDs (Figure 1 and 3b-c). Yet no particular properties in terms of RNA binding or other function have been assigned to such dsRBDs containing a shorter helix  $\alpha 1$  [40,51]. The second region of variability consists of the loop connecting the  $\beta 1$  to the  $\beta 2$  strand (loop 2). This loop which is typically 6 amino-acids in length can accommodate some long insertion, like in the case of *S. pombe* Dcr1 and *A. thaliana* HEN1 (Figure 3e-f) [48,59]. In the case of *S. pombe* Dcr1, this longer loop 2 has been shown to be structurally heterogeneous in the free form of the protein and a deletion mutant with a shortened loop 2 failed to bind to dsRNA [48]. In the crystal structure of *A. thaliana* HEN1, conserved hydrophobic residues in this long loop 2 were found to interact with other conserved residues of the methylase catalytic domain [59]. This could suggest that long  $\beta 1$ - $\beta 2$  loops would have a dual role in both dsRNA binding and protein binding and could therefore serve to orient catalytic or auxiliary domains in respect to dsRNA substrate. This is an attractive possibility that would require further investigation.

In addition to these modest variations around the canonical dsRBD fold, some remarkable extensions have also been observed. So far, all of them are found as C-terminal extensions, thus occurring after helix  $\alpha 2$  (Figure 3c, d and e). The first C-terminal extension have been described in the *Yeast* RNase III (Rnt1p) and consists of a long  $\alpha$ -helix (helix  $\alpha 3$ ) that folds back after helix  $\alpha 2$  in the direction of helix  $\alpha 1$  and the loop joining helix  $\alpha 1$  with the  $\beta 1$  strand (Figure 3c) [40,41]. The additional helix  $\alpha 3$  has been shown to be important for the overall stability of the domain and to play a critical role in RNA binding. Even if helix  $\alpha 3$  is not in direct contact with RNA substrates, it has been proposed to contribute indirectly to RNA binding by helping positioning helix  $\alpha 1$ , which is the primary determinant of RNA recognition by Rnt1p [40,41]. In addition it has been noticed that helix  $\alpha 3$  would clash with the C-terminal end of helix  $\alpha 1$  if this latter one would not be shorter as compared to canonical dsR-



BDs (Figure 3a and c) [40]. Although shorter, a similar helix  $\alpha 3$  was found in the structure of *K. polysporus* Dcr1 (Figure 3d) [49]. Thus far, no particular role have been associated to helix  $\alpha 3$  in *K. polysporus* Dcr1. Indubitably, the most striking C-terminal extension found in dsRBDs resides in the C-terminal dsRBD of the fission yeast dicer (*Schizosaccharomyces pombe* Dcr1) [48]. Indeed, the C-terminal extension is composed of a short  $\alpha$ -helical turn (helix  $\alpha 3$ ) followed by a zinc-coordination site (Figure 3e). The short helix  $\alpha 3$  is clearly reminiscent of *S. cerevisiae* Rnt1p and *K. polysporus* Dcr1. But the CHCC zinc-binding motif has no equivalent in any dsRBD structure. What is specially remarkable is that not all four ligands composing this zinc-binding site (three cysteines and one histidine, CHCC) are part of the C-terminal extension. Only the last two ligands (chCC) are found within the extension, whereas the first two ligands (CHcc) are part of the dsRBD fold itself. These first two ligands are located in loop 1 between  $\alpha 1$  and  $\beta 1$  and in loop 3 between  $\beta 2$  and  $\beta 3$  respectively. This C-terminal extension has been shown to play a critical role in the sub-cellular localization of *S. pombe* Dcr1 [48,66]. Zinc coordination is indeed required for the proper folding of the domain and contributes to the formation of a protein-protein interaction surface that mediates nuclear localization of Dcr1. The C-terminal extension would allow the attachment of Dcr1 to a nuclear protein, resulting in nuclear accumulation of Dcr1 [48,67].

Attractively, other dsRBDs have been reported to mediate nucleocytoplasmic trafficking, promoting either nuclear import or nuclear export [68-72]. However, the exact mechanisms controlling the nucleocytoplasmic distribution of these dsRBD-containing proteins are poorly understood. Structural information would probably be needed to unravel the molecular basis of these mechanisms and particularly to determine whether extensions to the canonical dsRBD fold would also be present in these dsRBDs and would as well be involved in the regulation of the sub-cellular localization of these proteins.

Interestingly, several other dsRBDs have been reported to mediate interaction with protein domains. These interactions commonly involve other dsRBDs, but different types of domains have also been reported to mediate these dsRBD-protein interactions [59,73-81]. To date, structural information describing how dsRBD domains would interact with other protein partner is very limited. However, the crystal structure

of the two tandem dsRBDs of DGCR8 that extensively interact with each other might give an insight into dsRBD-dsRBD associations [60]. To date, the structure of DGCR8 dsRBDs is unique, as in the available solution structures of tandem dsRBDs, namely PKR and ADAR2, the linker between the two dsRBDs is flexible, leading to independent dsRBDs [51,52,61]. Additional structural studies of proteins containing multiple dsRBD domains would be critical to understand better how dsRBDs might associate with each other and how they collaborate for dsRNA binding, in terms of affinity, cooperativity and specificity.

### **dsRNA recognition by dsRNA binding domains**

#### *Structural characteristics of the A-form RNA helix.*

The most salient property of dsRBDs is their ability to discriminate among the structural and chemical variety of nucleic acid polymers to preferentially bind to dsRNA. The dsRNA helix structure has been extensively studied and the interested reader is referred to other reviews for further study [24,82]. Nevertheless we will point out a few aspects of dsRNA structure relevant to the specific recognition of dsRNA by dsRBDs. The conformation adopted in solution by dsRNA is the so-called A-form RNA double helix. The morphology of this helix is characterized by a wide and shallow minor groove where the edge of the bases is readily accessible and a deep and narrow major groove where access to the bases is hindered (Figure 4a). The other distinctive characteristic of dsRNA is the presence of 2'OH functional groups on the ribose sugars, lining up in the minor groove, which are absent in the DNA double helix. Discrimination between dsDNA and dsRNA is achieved by dsRBDs by probing chemical features (presence vs. absence of ribose 2'OH groups) and structural features (the width of the minor and major grooves). Due to the difference of grooves accessibility, contacts between RNA bases and dsRBDs occur in the minor groove. Figure 4c and d show the hydrogen bond donor/acceptor groups exposed in the minor grooves that can be probed by dsRBDs to distinguish an A-U base pair from a G-C base pair. As it can be seen, the only pattern difference that can be exploited to discriminate A-U from G-C base pairs, is the presence of one hydrogen bond donor in G-C base pairs (Figure 4c and d). In the next part of this review, we describe in detail the interactions observed in the high resolution structures of dsRBDs-RNA com-

plexes solved in the past 15 years (Table 3) that gave us insights into the molecular basis of dsRNA recognition. We will then discuss about the sequence specific contacts observed in three recent high resolution structures of dsRBD–RNA complexes, which gave new insights in the mechanisms of specific RNA target recognition by dsRBDs.

### *High resolution dsRBD-RNA structures*

At the date of writing, high resolution structures of nine different dsRBDs in complex with various types of RNA partners have been determined (Table 3) [41-46,52,54-56,59]. The proteins containing these dsRBDs are involved in a broad range of biological processes including RNA editing (ADAR2), RNA silencing (TRBP [81,83] and HYPONASTIC LEAVES1 [84]), ribosomal RNA processing (Rnt1p and *Aquifex aeolicus* RNase III), and mRNA localization (*Drosophila* Staufen). The RNA molecules used in these structures determination are of diverse types (RNA duplexes or RNA hairpins), origin (synthetic or natural target), size, and nucleotide sequence. For instance, Aa RNase III has been crystallized with five different types of RNA including coaxially stacked poly GC duplexes of synthetic origin, as well as RNA hairpins deriving from canonical substrates (See Table 3 and references for more details). This dataset of structures also provides interesting examples of specific RNA secondary structure recognition in four structures of dsRBDs in complex with RNA stem-loop (Staufen dsRBD3, Rnt1p dsRBD, ADAR2 dsRBD1). Interestingly, RNA sequence-specific contacts are observed in several structures. This raises the possibility that direct readout of the nucleotide sequence in the RNA minor groove modulates RNA recognition by dsRBDs. The majority of the structures of the dataset were determined by X-ray crystallography (thirteen structures) and five structures were solved by NMR spectroscopy (Table 3). In the next part of this paper, we present what we learnt from this set of structural data concerning two aspects of RNA recognition by dsRBDs: shape recognition (A-form RNA helix, hairpins apical loops) and direct readout of the nucleotide sequence.

## Description of the dsRBD-RNA binding interface

### *Global description of dsRBD-RNA interface.*

High resolution structures of dsRBDs in complex with dsRNA gave a precise description of the RNA binding surface and revealed how different parts of the domain combine together to form a surface ‘shaped’ to recognize specifically the A-form helix conformation adopted by dsRNA. DsRBDs achieve specific dsRNA recognition by making contacts to bases and ribose moieties located at two successive minor grooves and by contacting the phosphate backbone delimiting the intervening major groove (Figure 5a). As mentioned previously, three distinct regions of the protein participate in the recognition of dsRNA: (i) the N-terminal tip of helix  $\alpha_2$ , (ii) helix  $\alpha_1$ , located at the N-terminus of the domain, and (iii) the loop connecting the first and the second  $\beta$  strands (loop 2) (Figure 1 and 5a) [55]. The molecular basis of this recognition is described in detail below. All the positions mentioned in the text refer to the sequence alignment shown in Figure 1a.

### *Recognition of the dsRNA major groove by the N-terminal tip of helix $\alpha_2$ (region 3)*

The N-terminal tip of helix  $\alpha_2$  is part of the canonical dsRNA binding surface and was originally designated ‘region 3’ [55] (Figure 5a and b). The amino acids sequence corresponding to this region consists of the well conserved KKxAK motif, which is indeed part of the dsRBD consensus sequence originally identified [33] (Figure 1 and 5a). In the canonical RNA binding mode observed in the high resolution structures of ADAR2 dsRBD1 and 2, TRBP dsRBD2, Xlrbpa, and Aa RNase III, this surface contacts the phosphodiester backbone of both RNA strands across the major groove of the helix (Figure 5a and b). The side chains of the first and the third lysines of the motif point toward one of the RNA strands while the second lysine points toward the other strand, forming an arch spanning the width of the major groove (see K55, K56 and K59 in Figure 5a, and Figure 6c). At the atomic level, the amino group of each lysine contacts a non-bridging oxygen atom of the RNA backbone. In addition, the amide proton of the first lysine (K55) forms a direct hydrogen bond also with a non-bridging oxygen atom of the RNA backbone (Figures 5a, 6a and c). This remarkable set of interactions results from a particular orientation of the side chains of these three lysines, which is stabilized by: (i) the rigidity of the peptide backbone,

which is embedded into an  $\alpha$ -helical secondary structure (namely, N-terminal tip of helix  $\alpha$ 2) and (ii) an important set of van der Waals interactions between the side chains of the first and the third lysines (K55 and K59) and other side chains from the hydrophobic core of the domain (Figure 6c, see also the section concerning the structural characteristic of the dsRBD fold). It is noteworthy that this motif is not absolutely conserved, the most frequent variations being the substitution of one or more lysine by an arginine or a glutamine residue (e.g Xlrbpa dsRBD2, HsPKR dsRBD2) (Figure 1). These kind of variations are not expected to affect significantly the way region 3 interacts with dsRNA as it can be seen in the structure of Xlrbpa dsRBD2 [55]. In several cases, the third lysine is more drastically replaced by a negatively charged residue or a glycine (Figure 1). In all these cases, this substitution is compensated by the presence of a lysine in helix  $\alpha$ 1 (Figure 5a,b and 6a,c) as discussed in detail in the next paragraph.

*Recognition of the dsRNA minor groove by helix  $\alpha$ 1 (region 1).*

Helix  $\alpha$ 1 is part of the canonical dsRNA binding surface and was originally designated 'region 1' [55] (Figure 5a and b). It is an amphipathic  $\alpha$ -helix with a hydrophobic surface involved in the constitution of the hydrophobic core of the dsRBD, and a solvent-exposed hydrophilic surface making extensive contacts with bases and riboses in the minor groove of the dsRNA. Analysis of the structures of dsRBDs-dsRNA complexes (Table 3) shows that the interaction surface of helix  $\alpha$ 1 with RNA is constituted by four to five amino acids always at positions 3, 4, 7, 8 and 11 of the sequence alignment shown in Figure 1. The position of these residues at the surface of helix  $\alpha$ 1 is schematically shown in Figure 7. It is noteworthy that among these residues, only one is part of the dsRBD consensus sequence (position 8 in Figure 1). As it has already been noted, the relative lack of conservation of these exposed side chains could be one of the factors modulating the RNA binding properties of different dsRBDs [54]. The interactions involving these residues are described in the following paragraphs.

Residue found in position 3 (Figures 1 and 7) are almost invariably bulky and hydrophobic amino acids such as valine or isoleucine. As it can be seen in the structures of Xlrbpa dsRBD2, ADAR2 dsRBD1 and 2, this side chain has three

important involvements. First it makes van der Waals contacts to a ribose moiety of the RNA, contributing to helix  $\alpha 1$  positioning in the minor groove (see Val 3 in Figure 6c). Second, it interacts with the aliphatic portion of the third lysine of the KKxAK motif, forcing it to orient toward the RNA backbone (see Lys 59 in Figure 6c). Third, it interacts with the well conserved alanine located in helix  $\alpha 2$  (position 63 in Figure 1), participating to the formation of the hydrophobic core of the dsRBD fold. A striking exception is found in Aa RNase III dsRBD, where position 3 is occupied by a lysine. The high resolution structure of this dsRBD in complex with dsRNA reveals that the aliphatic moiety of the side chain of this lysine makes hydrophobic contacts with the core of the domain, whereas the amino group points toward the RNA backbone to contact the non-bridging oxygen atom that is normally interacting with the last lysine of the canonical KKxAK motif (see Lys 3 in Figure 5b and Figure 6a). Interestingly, RNase III dsRBD exhibits a non canonical KKxAE motif where the last lysine of the canonical KKxAK motif is replaced by a glutamate. In this case, the recognition of the RNA major groove is achieved by a bipartite motif where the first two lysines (Lys 55 and Lys 56) are coming from the N-terminal part of helix  $\alpha 2$  as it is usually the case, and the third lysine (Lys 3) is coming from helix  $\alpha 1$  (Figure 1, Figure 5 and Figure 6a). This situation occurs several times in the sequence alignment of Figure 1: the loss of the last lysine of the KKxAK motif (which is replaced by a glutamate or a glycine) is rescued by the presence of a lysine in the N-terminal part of helix  $\alpha 1$  (at position 3 in Figure 1). This functional complementation of the lysine at position 3 in helix  $\alpha 1$  and an incomplete KKxAx motif is supported by structural data from the different Aa RNase III-RNA complexes. Additional structural and biochemical data would be needed to know whether this observation can be generalized or not.

Amino acids found at position 4 in helix  $\alpha 1$  (Figures 1 and 7a) are Thr, Ser, Cys, Gly, Gln, Asn, Arg or Met. Except Gly or Met, all these side chains can potentially be involved in hydrogen bond interactions. The structures of Aa RNase III in complex with dsRNA show that the hydroxyl group of the threonine is involved in a hydrogen bonds network with the O4' atom and the 2'-OH group of two adjacent ribose sugars. In most structures, helix  $\alpha 1$  comes in close contact with the RNA ribose sugars, with a distance C $\alpha$  - 2'OH of about 3.5 Å. In both dsRBDs of ADAR2, this position is occupied by a methionine residue whose side chain points toward the

edge of the bases in the minor groove to make a sequence specific interaction [52] (Figure 6b and 7c). The impact of this interaction on RNA recognition will be discussed in detail later in the text.

Position 7 (Figures 1 and 7a) is occupied in most cases by either a glutamine or an asparagine residue (50% in the alignment of Figure 1). In several structures of Aa RNase III in complex with dsRNA, the side chain of the glutamine makes a bidentate hydrogen bond to the 2'-OH group of the ribose and to the base of the same nucleotide. The role of hydrogen bond acceptor can be fulfilled either by the O2 atom of a pyrimidine or the N3 atom of a purine (Figure 4c and d, and see Gln 7 in Figure 6a). Therefore, as noticed by Gan and coworkers, this interaction is not sequence specific since all four types of bases can support it [45]. The same interaction pattern is observed in the structure of TRBP dsRBD2 and Xlrbpa dsRBD2. However this interaction is not present in all known structures of dsRBD-RNA complexes: in both dsRBDs of ADAR2, the position is occupied by an asparagine which does not make such contacts, possibly due to its smaller length compared to the glutamine. In Staufen dsRBD3 and Rnt1p, position 7 is occupied by a histidine and a tyrosine respectively. In the NMR structure of Staufen dsRBD3, the histidine side chain is too far from the RNA to interact, due to the important curvature of the RNA helix [54]. In Rnt1p, the tyrosine residue is stacked on a ribose sugar at the 3' side of the apical loop of the RNA substrate [41,42].

The amino acid at position 8 (Figures 1 and 7) is a remarkably well conserved glutamate which is indeed part of the consensus sequence identified for dsRBDs [33]. It is therefore expected to be important for RNA binding, and actually mutation of this residue has been shown to abolish RNA binding [54]. In the set of dsRBDs having the shorter version of helix  $\alpha 1$  (e.g. ADAR2 dsRBDs and Rnt1p) this is the last residue of the helix which interacts with dsRNA. In ADAR2 dsRBD2, Xlrbpa dsRBD2, TRBP dsRBD2 and RNase III dsRBD, the carboxylic group of this glutamate makes a hydrogen bond with the 2'-OH of a ribose sugar of the RNA. In ADAR2 dsRBD1 and in Staufen dsRBD3, this glutamate interacts with a nucleotide located in the apical loop of the RNA stem-loop binding partner. Rnt1p is one of the few examples where this glutamate is not conserved. It is replaced by a serine which also interacts with a 2'-OH group in the apical loop.

Position 11 is present only in dsRBDs that have the longer version of helix  $\alpha 1$  (Figure 1 and compare Figure 7c and 7d). It is in an *ad hoc* position to contact the RNA minor groove. Hydrophobic residues (Val or Ile) or glutamine residues are frequently found at this position. In the second dsRBDs of TRBP and Xlrpba a valine is present while an isoleucine is found in the third dsRBD of Staufen. In these three structures the hydrophobic side chain makes a van der Waals interaction with an RNA ribose. Aa RNase III provides an example where this position on helix  $\alpha 1$  is occupied by a glutamine residue. Remarkably, the side chain of the glutamine points toward the minor groove and interacts with the edge of a nucleotide base. This interaction has been reported to be sequence specific and will be further discussed in detail later in the review.

#### *Interactions from region 2 (loop 2) to the dsRNA minor groove.*

The last region of the dsRBDs composing the canonical RNA binding surface is the loop connecting the two strands  $\beta 1$  and  $\beta 2$  (also designated 'region 2' [55]). In the canonical mode of binding, this loop inserts into the RNA minor groove one helix turn away from the minor groove interacting with the helix  $\alpha 1$  (Figure 5a and b, Figure 6c and d). This loop has a well defined length (6 residues, positions 28-33 in Figure 1) and amino acid composition matching the pattern GPxHxx (Figure 1). A notable exception is found in the second dsRBD of *A. thaliana* HEN1 (Figure 3f) where the loop differs both in size and amino acid composition. Interestingly, the crystal structure of HEN1 in complex with RNA [58] shows that the loop makes extensive contacts with other parts of the protein and interacts only marginally with the RNA. The RNA binding mode of loop 2, originally described in the structure of Xlrpba dsRBD2 [55] relies on a small set of interactions which are conserved in the other high resolution structures (TRBP dsRBD2, ADAR2 dsRBDs, and Aa RNase III). Namely, the carbonyl group of the peptide backbone of the conserved His 31 (Figure 1) makes a direct hydrogen bond to a 2'OH group on one strand whereas the imidazol cycle stacks on one ribose and makes a hydrogen bond to the 2'OH of the previous ribose on the other strand (Figure 6d). In many dsRBD-RNA structures the peptide backbone carbonyl group of residue 30 makes a hydrogen bond with the amino group of a guanine. This contact is sequence specific and will be discussed in a later paragraph.



The side chain of residue 30 contacts the edge of a base in the minor groove and might also play a role in RNA recognition, although experimental evidence is still lacking here. For instance in HYL1 dsRBD1, this position is occupied by a serine which makes a hydrogen bond with the N3 atom of a guanine while in ADAR2 dsRBD1 this position is occupied by a valine, whose bulky side chain faces a guanine in the *syn* conformation (Figure 6c). The conservation of loop 2 binding mode results from the particular conformation adopted by the peptide backbone. Although there is a certain degree of variability in loop 2 sequences (Figure 1) a set of interactions organized around the highly conserved His 31 is almost systematically observed in the different high resolution structures. The side chain of His 31 makes van der Waals contacts with the side chains of the conserved proline (position 29) and of the next residue (position 32), which often has a significant aliphatic moiety (Figure 1). The side chain of residue 33, which is often a lysine or an arginine folds back on the side of the loop facing the RNA minor groove. Altogether this set of interactions forms a small hydrophobic cluster which must have a stabilizing effect on loop 2 conformation. The central role of the conserved His 31 in RNA binding has also been demonstrated by mutation studies [37].

## **RNA shape recognition by dsRBDs**

### *A-form helix recognition*

Several studies have shown that dsRBDs specifically bind dsRNA as opposed to other types of nucleic acids (namely ssRNA, RNA:DNA hybrid, dsDNA). Dissociation constants on the micromolar range have been reported for dsRNA binding whereas dissociation constants higher by an order of magnitude or more have been reported for other types of nucleic acid [1,48,57,83,85]. The interactions described previously confer dsRBDs the singular property to bind specifically to the A-form RNA helix. Two important characteristics of the dsRNA helix are recognized by this set of interactions: the width of the major groove and the presence of hydroxyl functional groups on the ribose sugars. The major groove of the A-form dsRNA helix has a width of about 10 Å (defined as the interstrand phosphate-phosphate distance between base pair  $i$  and  $i+6$ ) and a minor groove width of about 15 Å. For comparison, the major and minor grooves in a B-form DNA helix have a width of about 17 Å and

11 Å, respectively [24]. The width of the RNA major groove is probed by dsRBDs with the N-terminal tip of helix  $\alpha 2$  (region 3; KKxAK motif) which interacts with the non bridging oxygen atoms of the phosphodiester backbone which point outward the dsRNA axis. The residues of the protein involved in these interactions are located in a well structured region of the protein and hence are not prone to undergo important structural changes. The particular spatial arrangement of the amino and amide functional groups of the lysines present in dsRNA binding region 3 results from the tight packing of the lysine side chains with elements of the hydrophobic core of the domain coming from helix  $\alpha 1$  and from the  $\beta$ -sheet, emphasizing the intricate dependence between the fold of the domain and the formation of an operational binding surface. This rigidity results in a strict recognition of the major groove width leading to dsRNA specific recognition. The second distinctive chemical characteristic of RNA which is recognized by dsRBDs is the presence of hydroxyl functional groups at the 2' position of the ribose sugars. In the dsRNA these functional groups are located in the minor groove of the helix (Figure 4). They are recognized by the helix  $\alpha 1$  and loop 2 regions of the dsRBDs through direct and water mediated hydrogen bonds as described previously.

#### *Apical-loop recognition*

RNA hairpins are structural motifs frequently found in cellular RNA including mRNA, tRNA and rRNA [86]. Internal loops are formed in non complementary regions of dsRNA whereas apical loops are the RNA structures capping RNA hairpins [87]. Biological relevance of RNA apical loop recognition by a dsRBD has been demonstrated for Rnt1p. It was shown that AGNN tetraloop recognition by Rnt1p dsRBD was essential for proper RNA substrate recognition [88-90]. To date, it is not clearly understood whether apical loop specific recognition is a property conserved among all dsRBDs or whether it only concerns a subset of the dsRBD family. In the set of structures shown in Table 3, contacts between dsRBDs and RNA apical loops are observed for Staufen dsRBD3, Rnt1p dsRBD and ADAR2 dsRBD1 [41,42,52,54] (Figure 8a) whereas no contacts are observed in the structures of RNase III dsRBD and ADAR2 dsRBD2, although both dsRBDs are bound to RNA hairpins (Figure 8b).

The structures of three different dsRBDs in complex with RNA hairpins (Staufen, Rnt1p, and ADAR2 dsRBD1), revealed a conserved mode of binding in which helix  $\alpha$ 1 contacts the minor groove of the RNA apical loop, with the N and C-termini of the helix oriented toward the 5' side and the 3' side of the RNA loop respectively (Figure 8a). In all these structures, the helix contacts the sugars and bases of the loop and of the base pairs located immediately below it. Most of the side chains of helix  $\alpha$ 1 contacting the RNA apical loop occupy the same positions than those involved in double strand minor groove recognition (positions 3, 4, 7, 8 in Figure 7). For instance Asn 87 and Glu 88 (positions 7 and 8) in rat ADAR2 dsRBD1 and Glu 585 (position 8) in Staufen dsRBD3 are interacting with the RNA apical loop. In Staufen dsRBD3, Gln 582 and Lys 589 (positions 5 and 12) are making contacts to the RNA loop although they do not occupy one of the standard positions shown in Figure 7 and mutation of Gln 582 (position 5) abolishes RNA binding [54]. Rnt1p dsRBD uses side chains Lys 371, Arg 372, Tyr 375, Ser 376 (corresponding to the standard positions 3, 4, 7, 8) to contact the RNA apical loop and the closing base pairs. One of the most exciting features found in Rnt1p is the presence of an additional  $\alpha$  helix at the C terminus of the dsRBD (Figure 3c) which was shown to be essential for apical loop recognition [40,41]. However, the two solution structures of Rnt1p dsRBD in complex with RNA stem loop show that this terminal helix does not interact directly with the RNA [41,42]. Its influence on RNA binding and RNA loop recognition was proposed to result from an indirect effect resulting from the particular orientation imposed to helix  $\alpha$ 1 and loop 1 [40,42]. Surprisingly, the two conserved A and G bases of the AGNN-type RNA apical loop are pointing in the major groove, and thus are not contacted by the dsRBD. To account for that, it was proposed that Rnt1p dsRBD recognizes the particular fold of AGNN tetraloop [91,92] rather than the nucleotide sequence [41].

The impact of internal loops on dsRNA recognition by dsRBDs has been less studied than apical loop recognition. Nevertheless it is expected that disturbance of the regular A-form RNA double helix will impact dsRBD binding. It was reported accordingly that the presence of internal loops has a negative impact on RNA binding affinity [85].

## The sequence-specificity paradox

The structures of dsRBDs in complex with RNA (Table 3) revealed the canonical mode of dsRNA recognition. For a long time, this canonical mode of interaction was described with essentially an absence of any sequence specificity, even if few sequence specific contacts have been noticed in some structures [43,55]. In fact, as the majority of dsRBD-RNA interactions involve contacts with non-bridging oxygen of the phosphodiester backbone or with 2'-hydroxyl groups of the ribose sugar rings, the dsRBD has been described and often restricted to a non sequence-specific dsRNA binding domain [27-29]. In addition, very few sequence-specific contacts have been observed in the first structures of dsRBDs in complex with dsRNA, which was actually not completely unexpected regarding the fact that these structures have been solved with non-natural targets [54,55].

On the other hand, multiple examples exist of dsRBD-containing proteins showing a high degree of specificity in their interaction with RNAs. For example, the *Drosophila* Stauf protein, involved in mRNA transport, contains five dsRBDs and binds specifically to the 3'-UTR of certain mRNAs like *bicoid*, *oskar* and *prospero* mRNAs [2,74,93-97]. Furthermore, pre-mRNA editing enzymes from the ADAR family, that contain up to three dsRBDs essential for RNA substrate recognition [17,30], can be highly specific and RNA editing catalyzed by this family of enzymes is critical for normal life in mice and *Drosophila* [98-102]. In addition, bacterial and yeast RNase III contain a single dsRBD domain and cleave their substrates in a highly site-specific manner, which is required for optimal RNA function [103,104].

The necessity to reconcile these two points led to the conclusion that dsRBDs must recognize their specific substrates through their shape, meaning that a specific recognition would occur at particular distortions points of the A-form RNA helix introduced by specific RNA elements such as apical or internal loops, base-pair mismatches or bulges [27-29]. This notion was also promoted by the many structural examples in which these type of structural imperfections are largely exploited by proteins or peptides to achieve a sequence-specific recognition of the RNA [105,106]. Indeed, such imperfections in the dsRNA helix widen the major groove therefore opening an access to the higher sequence information content of the major groove, as compared to the minor groove [25,26]. Interestingly, in the case of dsRBDs, all the

structures solved in complex with RNA have clearly established that contacts with RNA bases can only occur in the minor groove. There is to date no structure of dsRBDs bound to an irregular RNA helix containing for instance bulges or internal loops that could reveal how these elements would indeed target certain dsRBDs to specific sites. The only structural elements that are clearly involved in targeting dsRBDs to specific RNAs are apical loop structures found as essential capping elements for certain dsRNA substrates. As already mentioned, the shape of the A/uGNN or the AAGU tetraloops is indeed a primary element for the specific binding and cleavage of RNA substrates by the yeast RNase III [41,42,89,90,107,108].

However, some key examples resist to this simple description of dsRBDs as shape-dependent dsRNA recognition domains. For example, the molecular features controlling the site-specific cleavage of bacterial RNase III consist of a subtle combination of determinants and antideterminants favouring or impairing the dsRBD binding to its RNA substrates [109-112]. Importantly, even if the minor groove has smaller information content than the major groove, a sequence-specific recognition of dsRNA in the minor groove is still possible [25]. Indeed, recent structures of ADAR2 dsRBDs bound to a natural and specific RNA substrate have revealed that the binding is achieved via a direct readout of the RNA sequence in the minor groove [52]. The recent increase of structural information on dsRNA recognition by dsRBDs allows an extension of the currently admitted RNA shape recognition model, to include the effect of RNA sequence on binding affinity.

### **Molecular basis for RNA sequence recognition**

Recent high resolution structures of ADAR2 dsRBDs and Aa RNase III in complex with dsRNA have shown the presence of a few RNA sequence specific contacts between the peptide backbone functional groups or side chains, and the edge of RNA bases in the minor groove. These contacts involve two regions of the dsRBDs: helix  $\alpha 1$  and loop 2 which contact the RNA at two successive minor grooves (Figure 5a and b). We discuss in the remaining parts the recent structural and biochemical finding supporting a direct readout of the RNA minor groove by certain dsRBDs.

### *Sequence specific interactions from helix $\alpha 1$*

These sequence specific interactions involve the side chains of residues located at position 4 and 11 in the sequence alignment shown in Figure 1 (see also Figure 7). In both dsRBDs of ADAR2 position 4 (Figure 1 and 7c) is occupied by a methionine residue. The high resolution solution structures of both dsRBDs in complex with an RNA stem loop deriving from a natural substrate of ADAR2, show that the side chain of the methionine extends into the RNA minor groove where the methyl group makes a hydrophobic contact with the edge of an adenine (Met 4 in Figure 6b). If this adenine were to be replaced by a guanine, a steric clash is predicted to occur between the amino group of the guanine and the methyl group of the methionine (compare Figure 4c and d) leading to a decrease in the affinity. To test this hypothesis, Stefl and coworkers studied the impact of the presence of the exocyclic amino group in the minor groove on binding affinity by replacing the adenine contacted by the methionine by a guanine. Indeed, these results show an affinity decrease of 4 to 5 fold when the adenine is substituted by a guanine [52]. No steric clash is predicted to occur if the adenine is replaced by a pyrimidine base, however no structural or biochemical data is available to see the impact of such a mutation. These structural and biochemical data suggest that the guanine acts as a binding antideterminant, which restricts the number of binding sites accessible to ADAR2. In dsRBD1 of *Drosophila* ADAR, the equivalent methionine is replaced by an alanine, the side chain of which is too short to influence any kind of binding register [53]. This could explain why *Drosophila* ADAR is less selective than mammalian ADARs and leads to extensive site-specific editing events [113,114]. The observation of such contacts for the two dsRBDs of ADAR2 is remarkable because it provides a basis to explain the mechanism of substrate selectivity by ADAR2.

Compared to a G-C base pair, it has already been recognized that an A-U base pair lacks hydrogen bonds donor/acceptor groups in the minor groove to be unambiguously recognized by proteins (Figure 4b). Thus the use of an aliphatic side chain could be a way for a protein to scan the RNA minor groove for the presence of the small hydrophobic surface from the adenine H2 proton. Interestingly it can be seen in several entries of the sequence alignment (Figure 1) that a glutamine or an asparagine is found at the same position (e.g. PACT dsRBD2). One would expect

that the side chain of these residues would point toward the minor groove and could contact the edge of an RNA base. As we shall see for the dsRBD of Aa RNase III, this type of configuration would favour the presence of a guanine. However some structural and biochemical data would be needed here to assess that possibility.

The second sequence specific contact involving helix  $\alpha$ 1 has been observed in several crystal structures of Aa RNase III in complex with RNA [43,45,46]. It involves the amino acid at position 11 in the sequences alignment shown in Figure 1 (see also Figure 7b and d). This position corresponds to the C-terminal extremity of helix  $\alpha$ 1 and it is not present in dsRBDs having a shorter version of helix  $\alpha$ 1 (e.g. ADAR2 dsRBDs, Rnt1p, compare also Figure 7c and d). The most prevalent amino acid found at this position is glutamine (42% in the set of sequences of Figure 1), followed by isoleucine or valine (Figure 1 and 7b). Quite interestingly, in several structures of Aa RNase III in complex with RNA, the side chain of the glutamine points directly into the minor groove of the RNA to contact the edge of a guanine. The amide side chain of this residue makes a pair of hydrogen bonds to the N3 atom (hydrogen bond acceptor) and to the exocyclic amino group (hydrogen bond donor) of the guanine edge exposed in the minor groove (see Gln 11 in Figure 6a and 7d). As noted previously, the amino group of guanine bases is the only hydrogen bond donor present in the RNA minor groove (compare Figure 4c and d), therefore this contact is sequence specific and likely has an effect on binding site selection by dsRBDs. Several reports pointed out that bacterial RNase III preferentially cleaves dsRNA at specific sites and that the sequence of the RNA region in contact with helix  $\alpha$ 1 was a major determinant [109,110,112]. In this context, this interaction between the glutamine and the guanine may be important to understand how RNA sequence affects target sites selection.

Overall, two kinds of sequence specific contacts involving amino acid side chains located in helix  $\alpha$ 1 have been observed in several high resolution structures of dsRBDs in complex with RNA: namely a methionine at position 4 contacting an adenine (ADAR2 dsRBD1, dsRBD2 [52]) (Figures 6b and 7c), and a glutamine at position 11 contacting a guanine (Aa RNaseIII [45,46]) (Figures 6a and 7d). In PACT dsRBD1 these positions in helix  $\alpha$ 1 are occupied by a glutamine and a methionine respectively (Figure 7a and b). It would be interesting to know whether these two side

chains could also make sequence specific contacts to the RNA, thereby increasing nucleotide sequence specificity for binding.

### *Sequence specific interactions in loop 2*

In twelve X-ray structures out of thirteen and in two NMR structures out of five (Table 3), an RNA sequence specific interaction involving region 2 of the dsRBDs canonical binding surface has been observed. This contact involves the carbonyl group of the peptide backbone of the third residue in loop 2 (position 30 in Figure 1), which makes a hydrogen bond with the exocyclic amino group of a guanine base in the RNA minor groove (Figure 6c and d) [43-46,52,55,56]. Since guanine is the only base exhibiting an amino group in the minor groove, the formation of this hydrogen bond is sequence specific. Both G-C and C-G base pairs can fulfill this requirement for a guanine because the exocyclic amino group occupies virtually the same position in the minor groove in both cases [25]. Actually this interaction has been observed in several different base pairing contexts, including Watson-Crick G-C (Xlrbpa, ADAR2 dsRBD2) and C-G base pairs (HYL1 dsRBD1, TRBP dsRBD2, Aa RNase III), G-U wobble pair (Aa RNase III), and G-G mismatch (ADAR2 dsRBD1). It is noteworthy that the structure of HYL1 dsRBD1 shows an additional hydrogen bond between the hydroxyl group of the serine (third residue of loop 2, position 30 in Figure 1) and the N3 atom of a guanine. This bidentate hydrogen bond involving the carbonyl group and the side chain of Ser 30 HYL1 dsRBD1 (Figure 1) could in principle discriminate a G-C from a C-G base pair but this remains to be tested.

To assess the importance of this interaction for ADAR2 dsRBDs binding affinity, Stefl and co-workers performed a series of binding assays with several mutants of the *GluR2* R/G RNA substrate of ADAR2, in which specific GC base pairs were substituted by AU. These results show that the loss of this interaction results in a significant decrease of binding affinity (1.5 to 6.6 fold for dsRBD1 and dsRBD2 respectively) [52]. The biological relevance of this interaction has also been tested *in vivo*, showing that mutations affecting loop 2 of either dsRBDs lead to a decrease of ADAR2 editing activity by 80-90% [52].

The formation of this CO-H<sub>2</sub>N hydrogen-bond highly depends on the conformation of loop 2 in order to have the peptide carbonyl group pointing in the proper



direction inside the minor groove. Indeed the conformation of loop 2 is quite well conserved in all the structures where this interaction is observed. The presence of the consensus Gly 28 and His 31 residues seems to be absolutely required whereas the consensus Pro 29 (Figure 1) which is not present in ADAR2 dsRBD2 seems not to be essential.

### **Sequence preference: register of binding**

Two RNA sequence specific contacts involving residues of helix  $\alpha 1$  and loop 2 have been observed in high resolution structures of three different dsRBDs-RNA complexes [43,52]. Both dsRBDs of human ADAR2 and the dsRBD of Aa RNase III use a carbonyl group from the peptide backbone in loop 2 to make a hydrogen bond to the amino group of a guanine in the minor groove (Figure 6c and d). Each of these three dsRBDs also use a side chain located in helix  $\alpha 1$  to make a sequence specific contact with a base in the RNA minor groove: a methionine located in the N terminal half of helix  $\alpha 1$  interacts with an adenine in ADAR2 dsRBD1 and dsRBD2, whereas a glutamine located two helical turn after, in the C terminal half of helix  $\alpha 1$ , interacts with a guanine in the dsRBD of RNase III (Figure 6a and b, Figure 7c and d). As a consequence of differences in the position and the orientation of these side chains in helix  $\alpha 1$ , the number of RNA base pairs between the two sequence specific contacts, also referred to as the 'register length', is different for these three dsRBDs-RNA complexes (Figure 9). In the structure of Aa RNase III in complex with RNA, the register length between the guanine contacted by loop 2 and the guanine contacted by helix  $\alpha 1$  is ten base pairs. In the structures of ADAR2 dsRBD1 the register between the guanine contacted by loop 2 and the adenine contacted by helix  $\alpha 1$  is nine base pairs whereas it is eight base pairs for dsRBD2 (Figure 9). The register difference observed for ADAR2 dsRBD1 and dsRBD2 results from a slightly different orientation of helix  $\alpha 1$  with respect to helix  $\alpha 2$ , emphasizing the importance of the residues of helix  $\alpha 1$  involved in the hydrophobic core of the protein [51,52]. How do these sequence specific contacts observed in high resolution structures actually affect dsRNA binding by dsRBDs? Stefl and co-workers reported that mutation of any of the RNA bases contacted specifically by the dsRBDs of ADAR2 lead to an affinity decrease from 2 to 5 fold [52] for the individual dsRBDs. These results indicate that the individual dsR-

BDs of ADAR2 are actually able to discriminate different RNA stem-loops based on their nucleotide sequence. When the two dsRBDs are associated in tandem as it is the case in the full length ADAR2, one can speculate that even more stringent RNA sequence recognition could arise, since four sequence specific interactions would then be present. Deletion of ADAR2 dsRBDs or mutation of the residues involved in sequence specific contacts with RNA leads to a marked decrease in site specific RNA editing *in vivo* [52]. It was also reported that ADAR1 and ADAR2 edit RNA with a different specificity [115,116]. Interestingly,  $\alpha$ 1 helices of the dsRBDs present in these proteins show differences predicted to affect RNA binding specificity. Namely, the methionine at position 4 in ADAR2 dsRBDs is replaced by a serine or a threonine in ADAR1, whereas ADAR1 dsRBD1 has a longer helix with a glutamine at position 11 (Figure 7a-d).

## **Concluding remarks**

Over the past fifteen years, since the determination of the first structure of a dsRBD by NMR, a wealth of structural data has accumulated, providing useful insights on the molecular basis of dsRNA specific recognition. Originally thought to recognize exclusively the A-form nature of the RNA double helix, dsRBDs have been later found to be sensitive to additional RNA determinants. We reviewed in this paper the recent structural data showing how dsRBDs can recognize, beyond the A-form RNA helix, other RNA features, such as apical loops and nucleotide sequences. This subtle modulation of dsRBD binding is likely to play a crucial role for targeting dsRBD containing proteins to their specific RNA substrates *in vivo*. A better understanding of these mechanisms should provide a basis to help predicting the optimal RNA binding sites for a given dsRBD. Although our understanding of RNA recognition by dsRBDs has made a great progress in the past few years, some particular aspects like the influence of RNA mismatches and internal loops on dsRBD binding would need further studies to be better understood. A particularly interesting point which has not been addressed so far is the possible interplay between dsRBDs in proteins harboring multiple copies of this domain. This could result in an increase of binding specificity or it might enable dsRBDs to recognize complex RNA tertiary structures. Another fascinating aspect of dsRBDs that has recently emerged is the role played by some of them in the sub-cellular localization of protein. This control occurs through interactions with other protein partners and sometimes with dsRNA. In conclusion, we believe dsRBDs are fascinating protein domains that concentrate a lot of possibilities in a very compact fold leaving still many aspects of their functions to be discovered.

## **Acknowledgment**

We sincerely apologize to the colleagues whose important work is not cited because of space limitation, or unfortunately because of our negligence. This work was supported by the Swiss National Science Foundation Nr. 31003AB-133134 and 310030E-131031, the SNF-NCCR structural biology and a KTI grant 11329.1 PFLS-LS. GM was supported by grant from the "Fondation pour la Recherche Médicale". PB was supported by the Postdoctoral ETH Fellowship Program.

## References

1. Gagnon A, Buckler-White A, Berkhout B, Jeang KT (1991) Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* 251 (5001):1597-1600.
2. Ferrandon D, Elphick L, Nüsslein-Volhard C, St Johnston D (1994) Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner. *Cell* 79 (7):1221-1232.
3. Higuchi M, Single FN, Köhler M, Sommer B, Sprengel R, Seeburg PH (1993) RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* 75 (7):1361-1370.
4. Nicholson AW (1999) Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol Rev* 23 (3):371-390.
5. Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409 (6818):363-366.
6. Hutvagner G, McLachlan J, Pasquinelli AE, Bálint E, Tuschl T, Zamore PD (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293 (5531):834-838.
7. Carthew RW, Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136 (4):642-655.
8. Fierro-Monti I, Mathews MB (2000) Proteins binding to duplexed RNA: one motif, multiple functions. *Trends Biochem Sci* 25 (5):241-246.
9. Saunders LR, Barber GN (2003) The dsRNA binding protein family: critical roles, diverse cellular functions. *FASEB J* 17 (9):961-983.
10. Tian B, Bevilacqua PC, Diegelman-Parente A, Mathews MB (2004) The double-stranded-RNA-binding motif: interference and much more. *Nat Rev Mol Cell Biol* 5 (12):1013-1023.
11. Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272 (9):2118-2131.
12. Brown RS (2005) Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* 15 (1):94-98.
13. Hall TM (2005) Multiple modes of RNA recognition by zinc finger proteins. *Curr Opin Struct Biol* 15 (3):367-373.
14. Lu D, Searles MA, Klug A (2003) Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* 426 (6962):96-100.
15. Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH (2006) Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. *Nat Struct Mol Biol* 13 (2):160-167.
16. Aviv T, Lin Z, Ben-Ari G, Smibert CA, Sicheri F (2006) Sequence-specific recognition of RNA hairpins by the SAM domain of Vts1p. *Nat Struct Mol Biol* 13 (2):168-176.
17. Barraud P, Allain FH-T (2012) ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr Top Microbiol Immunol* 353:35-60.
18. Schwartz T, Rould MA, Lowenhaupt K, Herbert A, Rich A (1999) Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science* 284 (5421):1841-1845.

19. Placido D, Brown BA, 2nd, Lowenhaupt K, Rich A, Athanasiadis A (2007) A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1. *Structure* 15 (4):395-404.
20. Skrisovska L, Bourgeois CF, Stefl R, Grellscheid SN, Kister L, Wenter P, Elliott DJ, Stevenin J, Allain FH (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep* 8 (4):372-379.
21. Oubridge C, Ito N, Evans PR, Teo CH, Nagai K (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372 (6505):432-438.
22. Allain FH, Gubser CC, Howe PW, Nagai K, Neuhaus D, Varani G (1996) Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* 380 (6575):646-650.
23. Allain FH, Bouvet P, Dieckmann T, Feigon J (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J* 19 (24):6870-6881.
24. Saenger W (1984) *Principles of nucleic acid structure*. Springer-Verlag New York,
25. Seeman NC, Rosenberg JM, Rich A (1976) Sequence-specific recognition of double helical nucleic acids by proteins. *Proc Natl Acad Sci U S A* 73 (3):804-808.
26. Steitz TA (1990) Structural studies of protein-nucleic acid interaction: the sources of sequence-specific binding. *Q Rev Biophys* 23 (3):205-280.
27. Chang K-Y, Ramos A (2005) The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J* 272 (9):2109-2117.
28. Stefl R, Skrisovska L, Allain FH-T (2005) RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle. *EMBO Rep* 6 (1):33-38.
29. Chen Y, Varani G (2005) Protein families and RNA recognition. *FEBS J* 272 (9):2088-2097.
30. Nishikura K (2010) Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 79:321-349.
31. Carlson CB, Stephens OM, Beal PA (2003) Recognition of double-stranded RNA by proteins and small molecules. *Biopolymers* 70 (1):86-102.
32. Hall KB (2002) RNA-protein interactions. *Curr Opin Struct Biol* 12 (3):283-288.
33. St Johnston D, Brown NH, Gall JG, Jantsch M (1992) A conserved double-stranded RNA-binding domain. *Proc Natl Acad Sci U S A* 89 (22):10979-10983.
34. McCormack SJ, Thomis DC, Samuel CE (1992) Mechanism of interferon action: identification of a RNA binding domain within the N-terminal region of the human RNA-dependent P1/eIF-2 alpha protein kinase. *Virology* 188 (1):47-56.
35. Feng GS, Chong K, Kumar A, Williams BR (1992) Identification of double-stranded RNA-binding domains in the interferon-induced double-stranded RNA-activated p68 kinase. *Proc Natl Acad Sci U S A* 89 (12):5447-5451.
36. Green SR, Mathews MB (1992) Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Genes Dev* 6 (12B):2478-2490.
37. Krovat BC, Jantsch MF (1996) Comparative mutational analysis of the double-stranded RNA binding domains of *Xenopus laevis* RNA-binding protein A. *J Biol Chem* 271 (45):28112-28119.

38. Kharrat A, Macias MJ, Gibson TJ, Nilges M, Pastore A (1995) Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J* 14 (14):3572-3584.
39. Bycroft M, Grünert S, Murzin AG, Proctor M, St Johnston D (1995) NMR solution structure of a dsRNA binding domain from *Drosophila* staufer protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J* 14 (14):3563-3571.
40. Leulliot N, Quevillon-Cheruel S, Graille M, van Tilbeurgh H, Leeper TC, Godin KS, Edwards TE, Sigurdsson STL, Rozenkrants N, Nagel RJ et al. (2004) A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *EMBO J* 23 (13):2468-2477.
41. Wu H, Henras A, Chanfreau G, Feigon J (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci U S A* 101 (22):8307-8312.
42. Wang Z, Hartman E, Roy K, Chanfreau G, Feigon J (2011) Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure* 19 (7):999-1010.
43. Blaszczyk J, Gan J, Tropea JE, Court DL, Waugh DS, Ji X (2004) Noncatalytic assembly of ribonuclease III with double-stranded RNA. *Structure* 12 (3):457-466.
44. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X (2005) Intermediate states of ribonuclease III in complex with double-stranded RNA. *Structure* 13 (10):1435-1442.
45. Gan J, Tropea JE, Austin BP, Court DL, Waugh DS, Ji X (2006) Structural insight into the mechanism of double-stranded RNA processing by ribonuclease III. *Cell* 124 (2):355-366.
46. Gan J, Shaw G, Tropea JE, Waugh DS, Court DL, Ji X (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol Microbiol* 67 (1):143-154.
47. Du Z, Lee JK, Tjhen R, Stroud RM, James TL (2008) Structural and biochemical insights into the dicing mechanism of mouse Dicer: a conserved lysine is critical for dsRNA cleavage. *Proc Natl Acad Sci U S A* 105 (7):2391-2396.
48. Barraud P, Emmerth S, Shimada Y, Hotz H-R, Allain FH-T, Bühler M (2011) An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *EMBO J* 30 (20):4223-4235.
49. Weinberg DE, Nakanishi K, Patel DJ, Bartel DP (2011) The inside-out mechanism of Dicers from budding yeasts. *Cell* 146 (2):262-276.
50. Mueller GA, Miller MT, Derose EF, Ghosh M, London RE, Hall TMT (2010) Solution structure of the Drosha double-stranded RNA-binding domain. *Silence* 1 (1):2.
51. Stefl R, Xu M, Skrisovska L, Emeson RB, Allain FH-T (2006) Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 14 (2):345-355.
52. Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L, Maris C, Peng L, Hofr C, Emeson RB et al. (2010) The solution structure of the ADAR2 dsRBD-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* 143 (2):225-237.
53. Barraud P, Heale BS, O'Connell MA, Allain FH (2012) Solution structure of the N-terminal dsRBD of *Drosophila* ADAR and interaction studies with RNA. *Biochimie* 94 (7):1499-1509.
54. Ramos A, Grünert S, Adams J, Micklem DR, Proctor MR, Freund S, Bycroft M, St Johnston D, Varani G (2000) RNA recognition by a Staufer double-stranded RNA-binding domain. *EMBO J* 19 (5):997-1009.

55. Ryter JM, Schultz SC (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* 17 (24):7505-7513.
56. Yang SW, Chen H-Y, Yang J, Machida S, Chua N-H, Yuan YA (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure* 18 (5):594-605.
57. Yamashita S, Nagata T, Kawazoe M, Takemoto C, Kigawa T, Guntert P, Kobayashi N, Terada T, Shirouzu M, Wakiyama M et al. (2011) Structures of the first and second double-stranded RNA-binding domains of human TAR RNA-binding protein. *Protein Sci* 20 (1):118-130.
58. Rasia RM, Mateos J, Bologna NG, Burdisso P, Imbert L, Palatnik JF, Boisbouvier J (2010) Structure and RNA interactions of the plant MicroRNA processing-associated protein HYL1. *Biochemistry* 49 (38):8237-8239.
59. Huang Y, Ji L, Huang Q, Vassilyev DG, Chen X, Ma J-B (2009) Structural insights into mechanisms of the small RNA methyltransferase HEN1. *Nature* 461 (7265):823-827.
60. Sohn SY, Bae WJ, Kim JJ, Yeom K-H, Kim VN, Cho Y (2007) Crystal structure of human DGCR8 core. *Nat Struct Mol Biol* 14 (9):847-853.
61. Nanduri S, Carpick BW, Yang Y, Williams BR, Qin J (1998) Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J* 17 (18):5458-5465.
62. Nagata T, Tsuda K, Kobayashi N, Shirouzu M, Kigawa T, Guntert P, Yokoyama S, Muto Y (2012) Solution structures of the double-stranded RNA-binding domains from RNA helicase A. *Proteins* 80 (6):1699-1706.
63. Green SR, Manche L, Mathews MB (1995) Two functionally distinct RNA-binding motifs in the regulatory domain of the protein kinase DAI. *Mol Cell Biol* 15 (1):358-364.
64. McMillan NA, Carpick BW, Hollis B, Toone WM, Zamanian-Daryoush M, Williams BR (1995) Mutational analysis of the double-stranded RNA (dsRNA) binding domain of the dsRNA-activated protein kinase, PKR. *J Biol Chem* 270 (6):2601-2606.
65. Patel RC, Stanton P, Sen GC (1996) Specific mutations near the amino terminus of double-stranded RNA-dependent protein kinase (PKR) differentially affect its double-stranded RNA binding and dimerization properties. *J Biol Chem* 271 (41):25657-25663.
66. Emmerth S, Schober H, Gaidatzis D, Roloff T, Jacobeit K, Bühler M (2010) Nuclear retention of fission yeast dicer is a prerequisite for RNAi-mediated heterochromatin assembly. *Dev Cell* 18 (1):102-113.
67. Woolcock KJ, Stunnenberg R, Gaidatzis D, Hotz HR, Emmerth S, Barraud P, Buhler M (2012) RNAi keeps Atf1-bound stress response genes in check at nuclear pores. *Genes Dev* 26 (7):683-692.
68. Brownawell AM, Macara IG (2002) Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J Cell Biol* 156 (1):53-64.
69. Strehblow A, Hallegger M, Jantsch MF (2002) Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol Biol Cell* 13 (11):3822-3835.

70. Gwizdek C, Ossareh-Nazari B, Brownawell AM, Evers S, Macara IG, Dargemont C (2004) Minihelix-containing RNAs mediate exportin-5-dependent nuclear export of the double-stranded RNA-binding protein ILF3. *J Biol Chem* 279 (2):884-891.
71. Macchi P, Brownawell AM, Grunewald B, DesGroseillers L, Macara IG, Kiebler MA (2004) The brain-specific double-stranded RNA-binding protein Staufen2: nucleolar accumulation and isoform-specific exportin-5-dependent export. *J Biol Chem* 279 (30):31440-31444.
72. Fritz J, Strehblow A, Taschner A, Schopoff S, Pasierbek P, Jantsch MF (2009) RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Mol Cell Biol* 29 (6):1487-1497.
73. Patel RC, Sen GC (1998) PACT, a protein activator of the interferon-induced protein kinase, PKR. *EMBO J* 17 (15):4379-4390.
74. Schuldt AJ, Adams JH, Davidson CM, Micklem DR, Haseloff J, St Johnston D, Brand AH (1998) Miranda mediates asymmetric protein and RNA localization in the developing nervous system. *Genes Dev* 12 (12):1847-1857.
75. Micklem DR, Adams J, Grunert S, St Johnston D (2000) Distinct roles of two conserved Staufen domains in oskar mRNA localization and translation. *EMBO J* 19 (6):1366-1377.
76. Zhang F, Romano PR, Nagamura-Inoue T, Tian B, Dever TE, Mathews MB, Ozato K, Hinnebusch AG (2001) Binding of double-stranded RNA to protein kinase PKR is required for dimerization and promotes critical autophosphorylation events in the activation loop. *J Biol Chem* 276 (27):24946-24958.
77. Tremblay A, Lamontagne B, Catala M, Yam Y, Larose S, Good L, Elela SA (2002) A physical interaction between Gar1p and Rnt1pi is required for the nuclear import of H/ACA small nucleolar RNA-associated proteins. *Mol Cell Biol* 22 (13):4792-4802.
78. Gupta V, Huang X, Patel RC (2003) The carboxy-terminal, M3 motifs of PACT and TRBP have opposite effects on PKR activity. *Virology* 315 (2):283-291.
79. Hitti EG, Sallacz NB, Schoft VK, Jantsch MF (2004) Oligomerization activity of a double-stranded RNA-binding domain. *FEBS Lett* 574 (1-3):25-30.
80. Haase AD, Jaskiewicz L, Zhang H, Laine S, Sack R, Gatignol A, Filipowicz W (2005) TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep* 6 (10):961-967.
81. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, Shiekhattar R (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature* 436 (7051):740-744.
82. Dickerson RE, Drew HR, Conner BN, Wing RM, Fratini AV, Kopka ML (1982) The anatomy of A-, B-, and Z-DNA. *Science* 216 (4545):475--485.
83. Gredell JA, Dittmer MJ, Wu M, Chan C, Walton SP (2010) Recognition of siRNA asymmetry by TAR RNA binding protein. *Biochemistry* 49 (14):3148--3155.
84. Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* 12 (2):206-212.
85. Bevilacqua PC, Cech TR (1996) Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry* 35 (31):9983-9994.



86. Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc Natl Acad Sci U S A* 87 (21):8467-8471.
87. Varani G (1995) Exceptionally stable nucleic acid hairpins. *Annu Rev Biophys Biomol Struct* 24:379-404.
88. Chanfreau G (2003) Conservation of RNase III processing pathways and specificity in hemiascomycetes. *Eukaryot Cell* 2 (5):901--909.
89. Chanfreau G, Buckle M, Jacquier A (2000) Recognition of a conserved class of RNA tetraloops by *Saccharomyces cerevisiae* RNase III. *Proc Natl Acad Sci U S A* 97 (7):3142--3147.
90. Nagel R, Ares J, M (2000) Substrate recognition by a eukaryotic RNase III: the double-stranded RNA-binding domain of Rnt1p selectively binds RNA containing a 5'-AGNN-3' tetraloop. *RNA* 6 (8):1142-1156.
91. Wu H, Yang PK, Butcher SE, Kang S, Chanfreau G, Feigon J (2001) A novel family of RNA tetraloop structure forms the recognition site for *Saccharomyces cerevisiae* RNase III. *EMBO J* 20 (24):7240--7249.
92. Lebars I, Lamontagne B, Yoshizawa S, Aboul-Elela S, Fourmy D (2001) Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. *EMBO J* 20 (24):7250--7258.
93. Macdonald PM, Struhl G (1988) cis-acting sequences responsible for anterior localization of bicoid mRNA in *Drosophila* embryos. *Nature* 336 (6199):595-598.
94. Kim-Ha J, Webster PJ, Smith JL, Macdonald PM (1993) Multiple RNA regulatory elements mediate distinct steps in localization of oskar mRNA. *Development* 119 (1):169-178.
95. Wagner C, Palacios I, Jaeger L, St Johnston D, Ehresmann B, Ehresmann C, Brunel C (2001) Dimerization of the 3'UTR of bicoid mRNA involves a two-step mechanism. *J Mol Biol* 313 (3):511-524.
96. Broadus J, Fuerstenberg S, Doe CQ (1998) Staufer-dependent localization of prospero mRNA contributes to neuroblast daughter-cell fate. *Nature* 391 (6669):792-795.
97. Doyle M, Kiebler MA (2011) Mechanisms of dendritic mRNA transport and its role in synaptic tagging. *EMBO J* 30 (17):3540-3552.
98. Higuchi M, Maas S, Single FN, Hartner J, Rozov A, Burnashev N, Feldmeyer D, Sprengel R, Seeburg PH (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 406 (6791):78-81.
99. Walkley CR, Liddicoat B, Hartner JC (2012) Role of ADARs in mouse development. *Curr Top Microbiol Immunol* 353:197-220.
100. Palladino MJ, Keegan LP, O'Connell MA, Reenan RA (2000) A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. *Cell* 102 (4):437-449.
101. Keegan LP, Brindle J, Gallo A, Leroy A, Reenan RA, O'Connell MA (2005) Tuning of RNA editing by ADAR is required in *Drosophila*. *EMBO J* 24 (12):2183-2193.
102. Paro S, Li X, O'Connell MA, Keegan LP (2012) Regulation and functions of ADAR in *drosophila*. *Curr Top Microbiol Immunol* 353:221-236.
103. Nicholson AW (2011) Ribonuclease III and the Role of Double-Stranded RNA Processing in Bacterial Systems. *Ribonucleases*:269--297.

104. Lamontagne B, Larose S, Boulanger J, Elela SA (2001) The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. *Curr Issues Mol Biol* 3 (4):71-78.
105. Draper DE (1995) Protein-RNA recognition. *Annu Rev Biochem* 64:593-620.
106. Draper DE (1999) Themes in RNA-protein recognition. *J Mol Biol* 293 (2):255-270.
107. Gaudin C, Ghazal G, Yoshizawa S, Elela SA, Fourmy D (2006) Structure of an AAGU tetraloop and its contribution to substrate selection by yeast RNase III. *J Mol Biol* 363 (2):322-331.
108. Ghazal G, Elela SA (2006) Characterization of the reactivity determinants of a novel hairpin substrate of yeast RNase III. *J Mol Biol* 363 (2):332-344.
109. Zhang K, Nicholson AW (1997) Regulation of ribonuclease III processing by double-helical sequence antideterminants. *Proc Natl Acad Sci U S A* 94 (25):13437-13441.
110. Pertz AV, Nicholson AW (2006) Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res* 34 (13):3708-3721.
111. Nathania L, Nicholson AW (2010) *Thermotoga maritima* ribonuclease III. Characterization of thermostable biochemical behavior and analysis of conserved base pairs that function as reactivity epitopes for the *Thermotoga* 23S rRNA precursor. *Biochemistry* 49 (33):7164-7178.
112. Shi Z, Nicholson RH, Jaggi R, Nicholson AW (2011) Characterization of *Aquifex aeolicus* ribonuclease III and the reactivity epitopes of its pre-ribosomal RNA substrates. *Nucleic Acids Res* 39 (7):2756-2768.
113. Keegan LP, McGurk L, Palavicini JP, Brindle J, Paro S, Li X, Rosenthal JJ, O'Connell MA (2011) Functional conservation in human and *Drosophila* of Metazoan ADAR2 involved in RNA editing: loss of ADAR1 in insects. *Nucleic Acids Res* 39 (16):7249-7262.
114. Graveley BR, Brooks AN, Carlson J, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. (2011) The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471 (7339):473-479.
115. Bass BL (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 71:817-846.
116. Lehmann KA, Bass BL (2000) Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 39 (42):12875-12884.
117. Corpet F (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 16 (22):10881-10890.
118. Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: network protein sequence analysis. *Trends Biochem Sci* 25 (3):147-150.

## FIGURES LEGENDS

### Figure 1

Sequence alignment of various double-stranded RNA binding domains.

Multiple sequence alignment of various dsRBD from human (*Homo sapiens*, Hs), fruitfly (*Drosophila melanogaster*, Dm), baker's yeast (*Saccharomyces cerevisiae*, Sc), frogs (*Xenopus leavis*, Xl), plants (*Arabidopsis thaliana*, At) and bacteria (*Escherichia coli*, Ec and *Aquifex aeolicus*, Aa). Alignment was done with Multalin [117,118] and manually optimized using 3D structural information. For each sequence, the name of the protein and the dsRBD number are given in the first item. The second item corresponds to the accession code in the UniProt database (<http://www.uniprot.org>). The third item corresponds to the range of amino-acid composing the dsRBD in the numbering of the full-length protein. The alignment is coloured by amino-acid conservation (> 40%) and properties. The sequence consensus (> 40%), the residues conserved for the fold and/or dsRNA binding and the canonical secondary structured elements are shown below the alignment. The three regions of interaction with dsRNA are also indicated.

### Figure 2

Structure of the double-stranded RNA binding domain.

(a-b) Three-dimensional structure of a dsRBD shown in two different orientations. The structure of XIRBPA-ds2 (PDB code 1DI2) is shown as a cartoon representation with  $\alpha$ -helices *in blue* and  $\beta$ -strands *in yellow*. The secondary structure elements composing the  $\alpha\beta\beta\beta\alpha$  fold are labelled on the structure. Conserved residues matching the sequence consensus of Figure 1 are shown as sticks and labelled with the residue number corresponding to the sequence alignment of Figure 1. (c) Conserved residues important for the fold of the domain. (d) Conserved residues important for dsRNA binding. Note that Y21 and F35 are important for both the fold and dsRNA binding by orienting the long lysine side-chains of K59 and K55, respectively. (e) Topology of a dsRBD. Residue numbers corresponding to the sequence alignment of Figure 1 are indicated.

### Figure 3

Variations and extensions to the canonical dsRBD fold.

Various dsRBDs structures are shown as cartoons with  $\alpha$ -helices *in blue* and  $\beta$ -strands *in yellow*. Variations and extensions to the fold are shown *in red* or *in green*. (a) *Xenopus leavis* RNA binding protein A (Xlrpba) dsRBD2 constitute the archetype of a canonical dsRBD (PDB code 1DI2). (b) *Mammalian* ADAR2 dsRBD2 with a shorter helix  $\alpha_1$  (PDB code 2L2K). (c) *Budding yeast* RNase III (*S. cerevisiae* Rnt1p) with a shorter helix  $\alpha_1$  and an  $\alpha$ -helical C-terminal extension (helix  $\alpha_3$ ) (PDB code 2LBS). (d) *Budding yeast* Dicer (*K. polysporus* Dcr1) with a short  $\alpha$ -helical C-terminal extension (helix  $\alpha_3$ ) (PDB code 3RV0). (e) *Fission yeast* Dicer (*S. pombe* Dcr1) with

a C-terminal extension composed of a short helix  $\alpha 3$  and a CHCC zinc-binding motif (*in green*). It also bears a long insertion in loop 2 (PDB code 2L6M). (f) *Plant* small RNA methyltransferase (*A. thaliana* HEN1) with a long insertion in loop 2 (PDB code 3HTX).

#### Figure 4

Sequence discrimination in the minor groove.

(a-b) Differences in the topology of the A-form dsRNA helix and the B-form dsDNA helix. Note especially the differences in the dimension and accessibility of the minor and major grooves. (c-d) Chemical groups of an A-U pair (c) and a G-C pair (d) lying in the minor and major grooves. Discrimination in the minor groove relies on the appreciation of the chemical group in position 2 of purine rings.

#### Figure 5

Global view of the canonical RNA binding surface of dsRBDs. Region 1 (helix  $\alpha 1$ ) and region 2 (loop 2) insert in two successive RNA minor grooves, region 3 (loop 4) contacts the phosphodiester backbone of the intervening RNA major groove. The residues shown in red sticks correspond to (a) the canonical KKxAK motif of the ADAR2 dsRBD1-RNA stem-loop complex (PDB code 2L3C) and (b) to the bipartite motif found in the RNase III-RNA complex (PDB code 2NUE) where one of the lysine is coming from helix  $\alpha 1$ . The numbering of the lysines corresponds to the sequences alignment of Figure 1. See text for details.

#### Figure 6

Sequence specific interactions in the dsRNA minor grooves

Amino acids side chains important for RNA binding are represented as sticks. (a) Recognition of the minor groove by helix  $\alpha 1$  (region 1), showing the sequence specific hydrogen bond between the Gln 11 side chain (Gln 161 in the wild type sequence) and the atom N3 and the exocyclic amino group of a guanine in the minor groove. (*A. aeolicus* RNase III dsRBD, PDB code 2EZ6). (b) Recognition of the dsRNA minor groove by helix  $\alpha 1$  (region 1), showing the sequence specific van der Waals contact between the methyl group of Met 4 (Met 238 in the wild type sequence) and the H2 atom of an adenine in the minor groove. (ADAR2 dsRBD2, PDB code 2L2K). (c) Recognition of the major and minor grooves by region 3 and 2. Lys 55, Lys 56 and Lys 59 (Lys 127, Lys 128 and Lys 131 in the wild type sequence) interact with the phosphodiester backbone on opposite strands, across the major groove. The conformation of Lys 55 and Lys 59 side chains is stabilized by hydrophobic contacts with Val 3, Tyr 21, Leu 23, Phe 35 and Met 37. The main chain carbonyl group of Val 30 makes a sequence specific hydrogen bond to the amino group of a guanine in the minor groove (ADAR2 dsRBD2, PDB code 2L2K). (d) Recognition of the minor groove by loop 2 (Xlrbpa dsRBD2, PDB code 1DI2). The two RNA strands are

'bridged' by hydrogen bonds involving the peptide backbone carbonyl and the N $\delta$  atom of the imidazol ring of His 31 (His 141 in the wild type sequence) and one hydroxyl group on each strand. The peptide backbone carbonyl group of P30 makes a sequence specific hydrogen bond with the amino group of a guanine.

### Figure 7

(a) Schematic representation of helix  $\alpha$ 1 showing the position of the residues involved in dsRNA binding. The numbers inside the circles correspond to the positions in the sequence alignment of Figure 1 and 7b. (b) Alignment of helix  $\alpha$ 1 sequences emphasizing the variability of the residues at positions 3, 4, 7 and 11, involved in RNA binding. (c) Schematic representation of helix  $\alpha$ 1 of ADAR2 dsRBD2 showing the location of the methionine making a sequence specific contact to the H2 proton of an adenine. (d) Schematic representation of helix  $\alpha$ 1 of the dsRBD of Aa RNase III showing the location of the glutamine making a pair of sequence specific contact to the N3 and to the exocyclic amino group of a guanine.

### Figure 8

Interaction of dsRBDs with RNA apical loops.

(a) Structure of Rnt1p dsRBD with helix  $\alpha$ 1 interacting with the RNA AGAA apical loop (PDB code 1T4L). The surface of the RNA hairpin is shown and colored in white for the RNA stem and in red for the RNA loop. The protein is shown as a cartoon, with  $\alpha$  helices represented as blue cylinders. (b) Structure of ADAR2 dsRBD2 showing no interaction between the protein and the RNA apical loop (PDB code 2L2K).

### Figure 9

Register variability between sequence specific contacts from helix  $\alpha$ 1 and loop 2. Structure superposition of ADAR2 dsRBD1 (blue) and dsRBD2 (red) and of the dsRBD of Aa RNase III (grey) (PDB codes 2L3C, 2L2K and 2EZ6). The glutamine side chain (Aa RNase III dsRBD), the methionine side chain (ADAR2 dsRBD1 and 2) from helix  $\alpha$ 1 and the peptide backbone carbonyl of loop 2, which make sequence specific contacts with the RNA bases in the minor groove are represented as sticks. The variety of registers observed in these three structures mainly results from the different position of the glutamine (position 11) and the two methionines (position 4) on helix  $\alpha$ 1 (Figure 7c and d).

## TABLES

**Table1**

Functions of double-stranded RNA binding proteins described in this review.

Protein name	Principal Functions
RNase III family	dsRNA specific endoribonucleases
Bacterial RNase III	Processing of rRNA and tRNA
Yeast Rnt1p	Processing of rRNA
Dicer	Processing of siRNA and miRNA
Drosha	Processing of rRNA and miRNA
ADAR family	A-to-I editing of viral and cellular RNAs
Staufen	Localization of mRNAs
PKR	Serine/Threonine kinase
	Anti-viral defense
	Cellular stress response
	Control of protein synthesis
TRBP/PACT	Modulator of PKR/RNA interference
TRBP	Inhibition of PKR
	Translation activation
	Component in RNAi pathway
PACT	Activation of PKR
HYL1/HEN1	Maturation of miRNA in plants

**Table 2**

dsRBD structures present in the protein data bank.

Protein name	PDB accession code <sup>a</sup>	References <sup>b</sup>
<b>RNase III family</b>		
<i>E. coli</i> RNase III	n.d.	[38]
<i>S. cerevisiae</i> Rnt1p	1T4O, 1T4N, 1T4L, 2LBS	[40-42]
<i>A. aeolicus</i> RNase III	1RC7, 1YZ9, 1YYK, 1YYO, 1YYW, 2EZ6, 2NUG, 2NUF, 2NUE	[43-46]
<i>T. maritima</i> RNase III	1O0W	JCSG
<i>M. musculus</i> Dicer	3C4B, 3C4T	[47]
<i>S. pombe</i> Dcr1	2L6M	[48]
<i>K. polysporus</i> Dcr1	3RV0	[49]
<i>H. sapiens</i> Drosha	2KHX	[50]
<b>ADAR family</b>		
<i>R. norvegicus</i> ADAR2	2B7T, 2B7V, 2L3C, 2L2K, 2L3J	[51,52]
<i>D. melanogaster</i> ADAR	2LJH	[53]
<b>Staufen</b>		
<i>D. melanogaster</i> Staufen	1STU, 1EKZ	[39,54]
<i>M. musculus</i> Stau2	1UHZ	RIKEN
<b>TRBP/PACT</b>		
<i>X. leavis</i> RBPA	1DI2	[55]
<i>H. sapiens</i> TRBP	3ADL, 3LLH, 2CPN	[56,57]
<i>H. sapiens</i> PACT	2DIX	RIKEN
<b>HYL1/HEN1</b>		
<i>A. thaliana</i> HYL1	3ADI, 3ADJ, 3ADG, 2L2N, 2L2M	[56,58]
<i>A. thaliana</i> HEN1	3HTX	[59]
<b>DGCR8</b>		
<i>H. sapiens</i> DGCR8	2YT4, 1X47	[60] and RIKEN
<b>PKR</b>		
<i>H. sapiens</i> PKR	1QU6	[61]
<i>M. musculus</i> PKR	1X48, 1X49	RIKEN
<b>ILF3/SPNR</b>		
<i>H. sapiens</i> ILF3	3P1X, 2L33	NESGC
<i>H. sapiens</i> SPNR	2DMY	RIKEN
<b>RHA</b>		
<i>M. musculus</i> RHA	2RS6, 2RS7, 1UIL, 1WHQ	[62] and RIKEN

<sup>a</sup> Apart for *E. coli* RNase III structure for which no coordinates have been deposited (n.d.: not deposited), PDB accession codes are given.

<sup>b</sup> Primary references related to each structure are given. In the case of structures solved by structural genomics centres and not associated with a publication, the name of the structural genomic centre is given.

**Table 3**

## Structures of dsRBDs in complex with RNA

Protein name	PDB code	Method	RNA substrates	References
<b>RNase III family</b>				
<i>S. cerevisiae</i> Rnt1p	1T4L, 2LBS	NMR	Single hairpin <sup>a</sup>	[41,42]
<i>A. aeolicus</i> RNase III	1RC7	X-ray	Coaxially stacked duplexes <sup>b</sup>	[43]
	1YZ9,		Coaxially stacked duplexes <sup>c</sup>	[44]
	1YYW,		Coaxially stacked duplexes <sup>d</sup>	[44]
	1YYO, 1YYK,		Coaxially stacked duplexes <sup>e</sup>	[44]
	2EZ6		Coaxially stacked hairpins <sup>f</sup>	[45]
	2NUF		Coaxially stacked hairpins <sup>f</sup>	[46]
	2NUE		Single hairpin <sup>f</sup>	[46]
2NUG,	Coaxially stacked duplexes <sup>f</sup>	[46]		
<b>ADAR family</b>				
<i>R. norvegicus</i> ADAR2	2L3C,	NMR	Single hairpin <sup>g</sup>	[52]
	2L2K			[52]
	2L3J			[52]
<b>Staufen</b>				
<i>D. melanogaster</i> Staufen	1EKZ	NMR	Single hairpin <sup>h</sup>	[54]
<b>TRBP/PACT</b>				
<i>X. leavis</i> RBPA	1DI2	X-ray	Coaxially stacked duplexes <sup>b</sup>	[55]
<i>H. sapiens</i> TRBP	3ADL	X-ray	Coaxially stacked duplexes <sup>i</sup>	[56]
<b>HYL1/HEN1</b>				
<i>A. thaliana</i> HYL1	3ADI	X-ray	Coaxially stacked duplexes <sup>j</sup>	[56]
<i>A. thaliana</i> HEN1	3HTX	X-ray	Coaxially stacked duplexes <sup>k</sup>	[59]

<sup>a</sup> small nucleolar RNA snR47 capped by an AGAA (1T4L) or AAGU (2LBS) tetraloop.

<sup>b</sup> Non natural substrate [GGCGCGCGCC]<sub>2</sub>

<sup>c</sup> Non natural substrate [CGAACUUCGCG]<sub>2</sub>

<sup>d</sup> Non natural substrate [AAAUAUAUAUUU]<sub>2</sub>

<sup>e</sup> Non natural substrate [CGCGAAUUCGCG]<sub>2</sub>

<sup>f</sup> Derived from R1.1, a canonical substrate of *E. coli* RNase III [109]

<sup>g</sup> RNA hairpins derived from the R/G editing site of mammalian *GluR2*

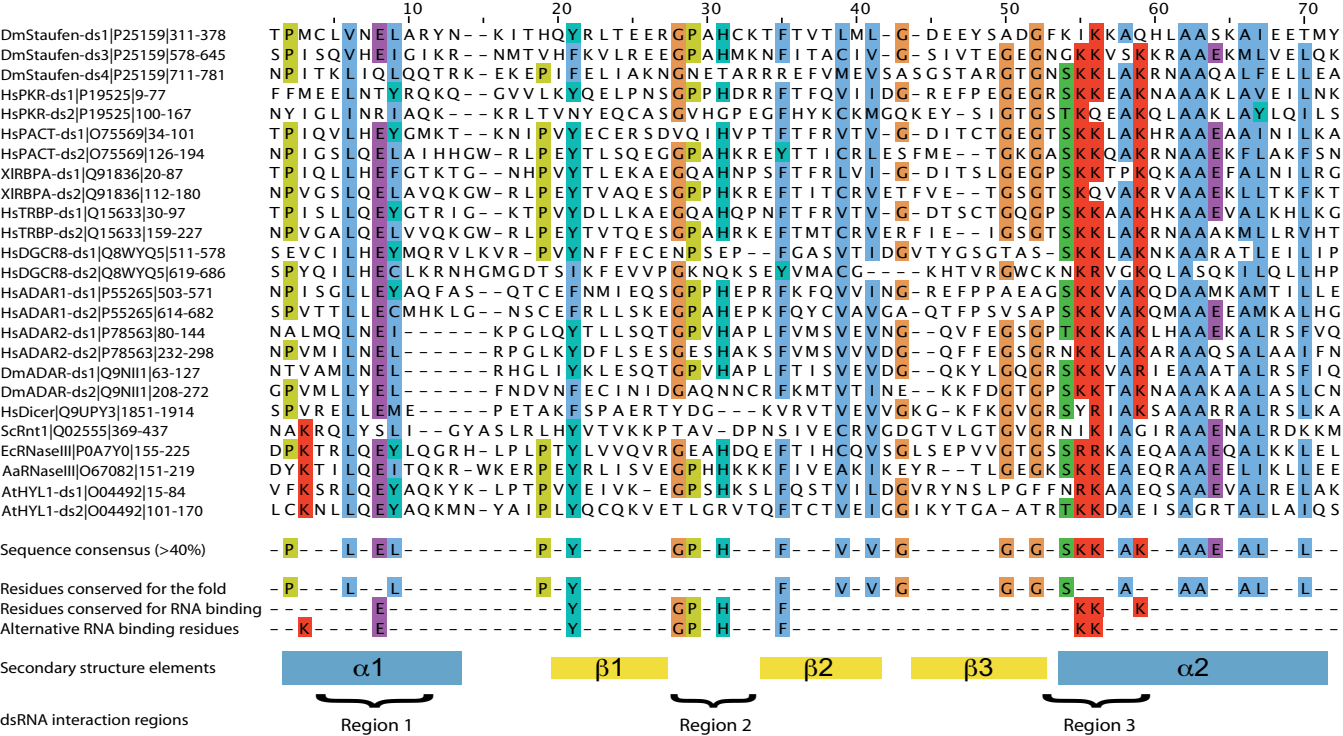
<sup>h</sup> Non natural substrate (the tetraloop is underlined) G G A C A G C U G U C C C U U C G G G G A C  
A G C U G U C C

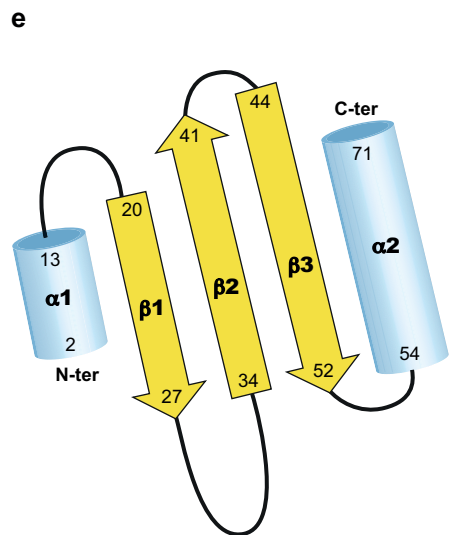
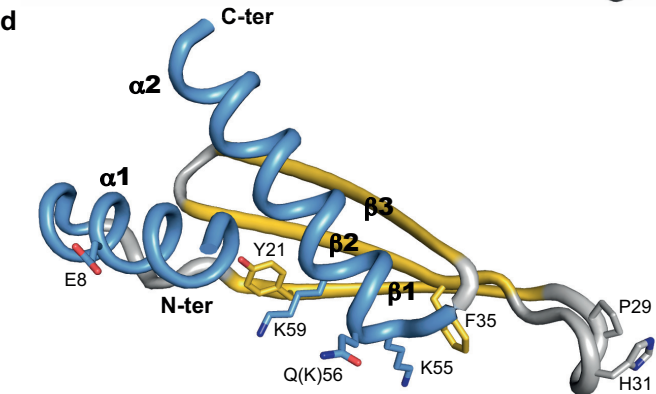
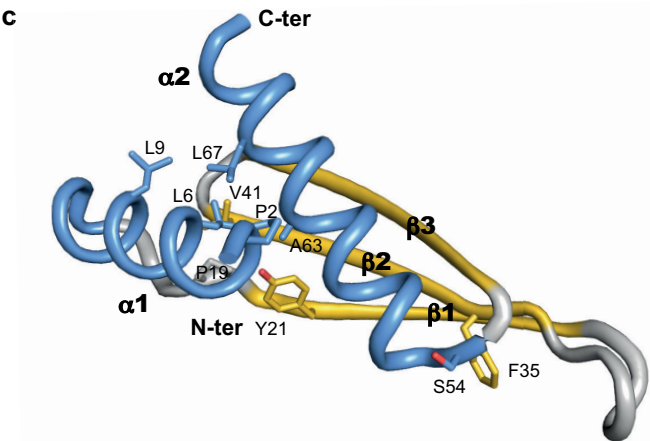
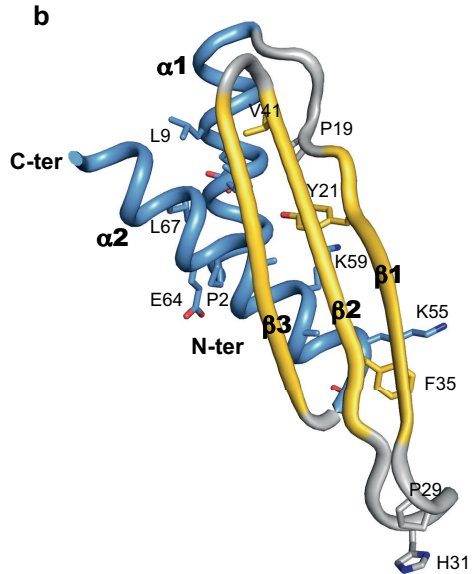
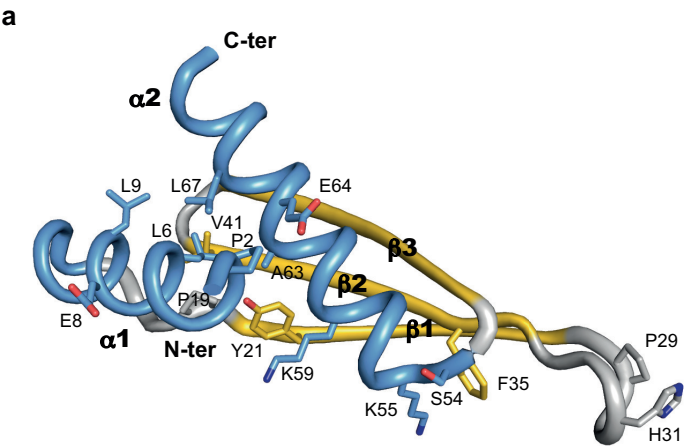
<sup>i</sup> Non natural substrate [CGCGCGCGCG]<sub>2</sub>

<sup>j</sup> Non natural substrate [CUCGAUAACC]/[GGUUAUCGAG]

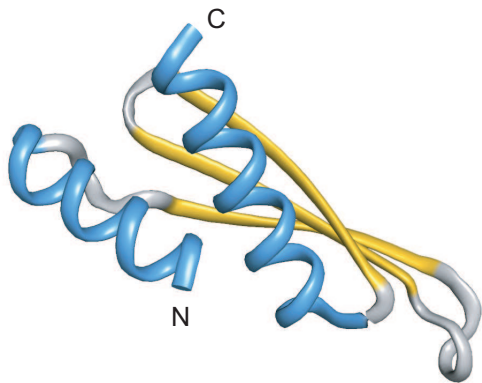
<sup>k</sup> Derived from miR173/miR173\*, a natural substrate of HEN1



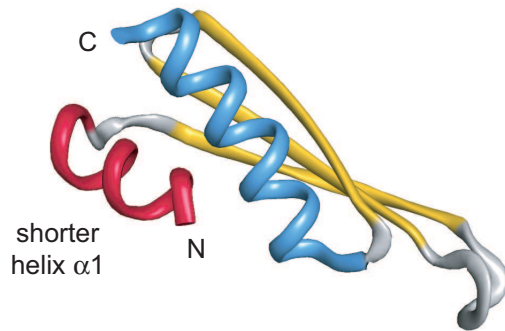




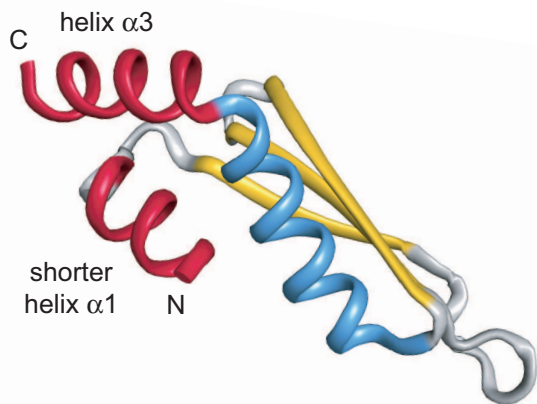
**a** *X. leavis* rbpA  
(canonical dsRBD)



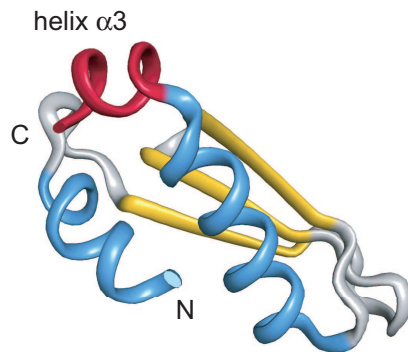
**b** *Mammalian* ADAR2



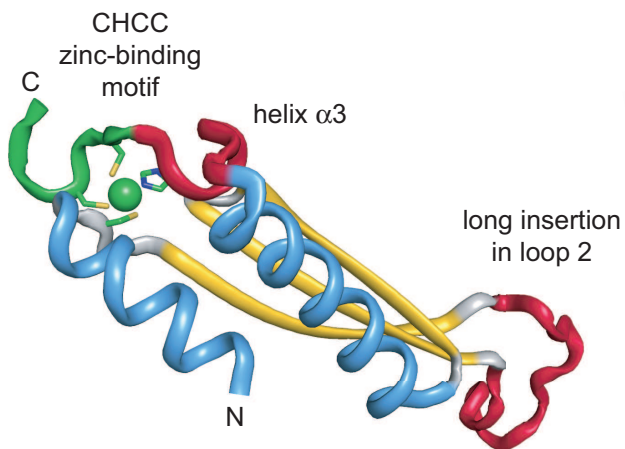
**c** *S. cerevisiae* Rnt1p



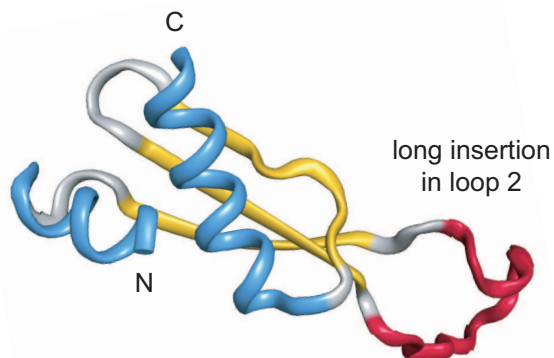
**d** *K. polysporus* Dcr1



**e** *S. pombe* Dcr1

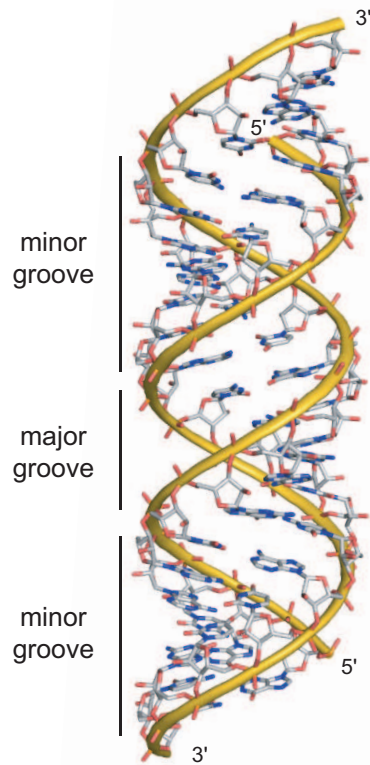


**f** *A. thaliana* HEN1

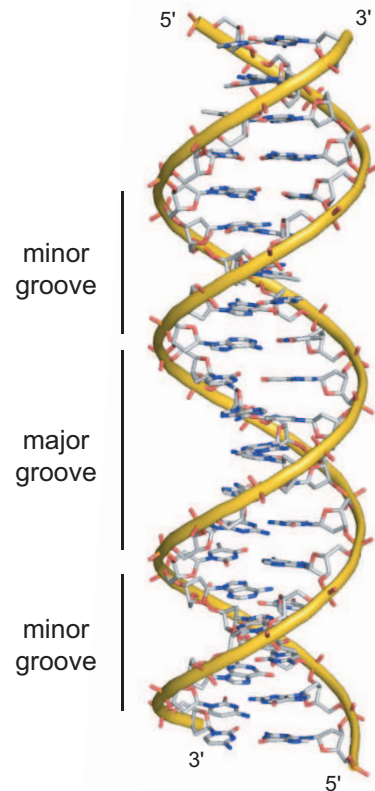


**a**

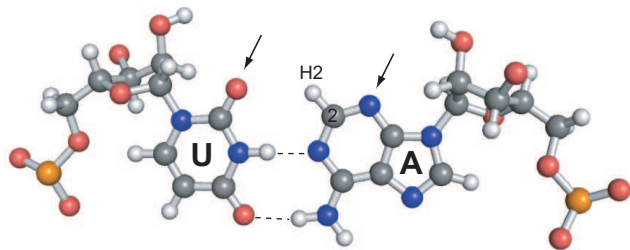
## A-form RNA helix

**b**

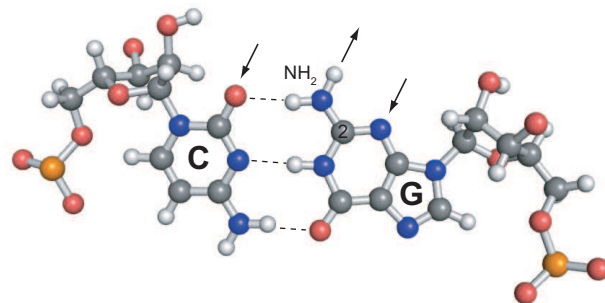
## B-form DNA helix

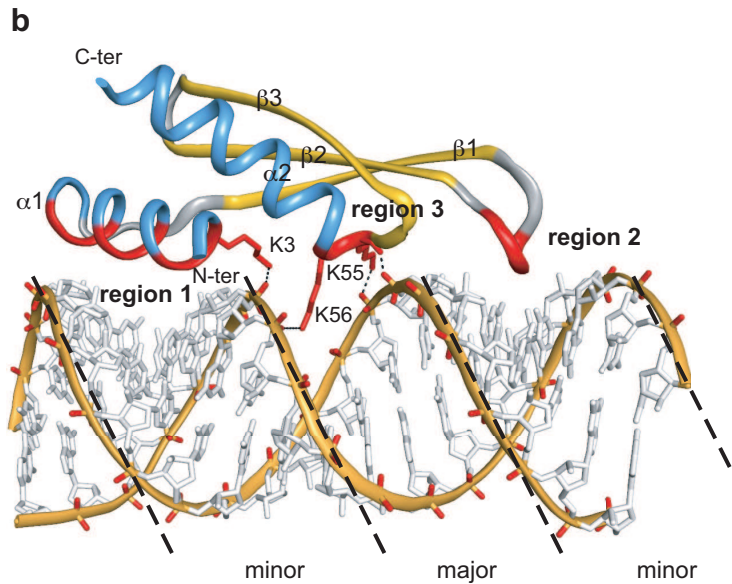
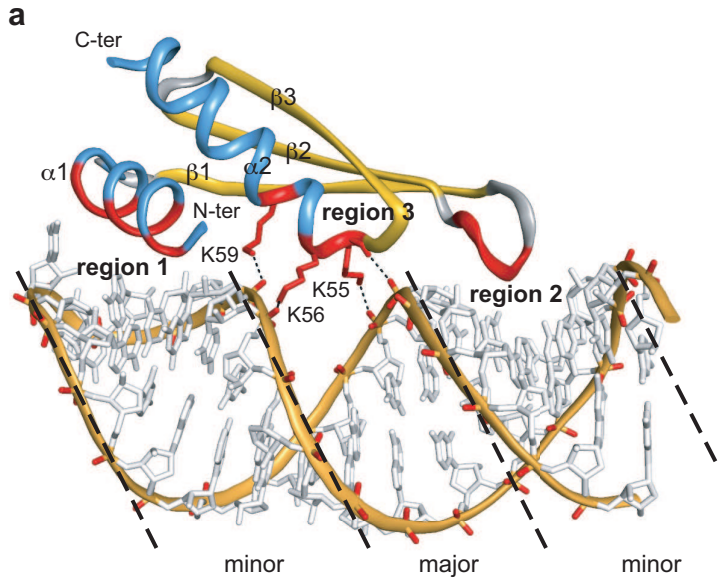
**c**

## minor groove

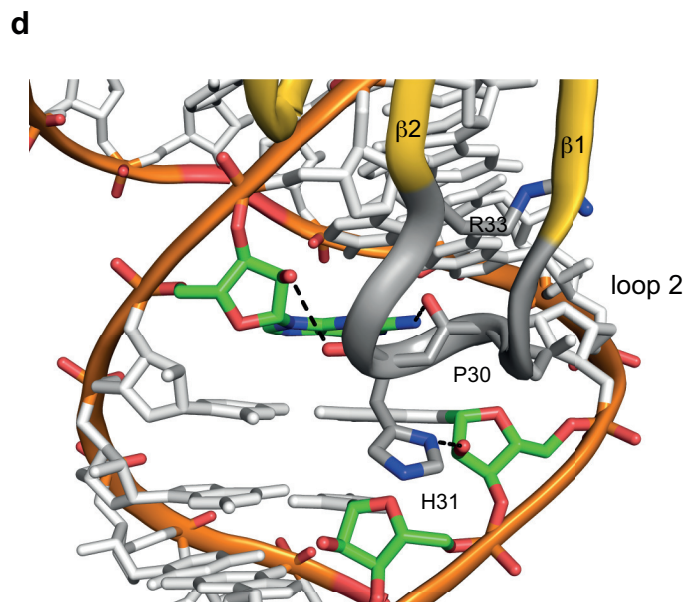
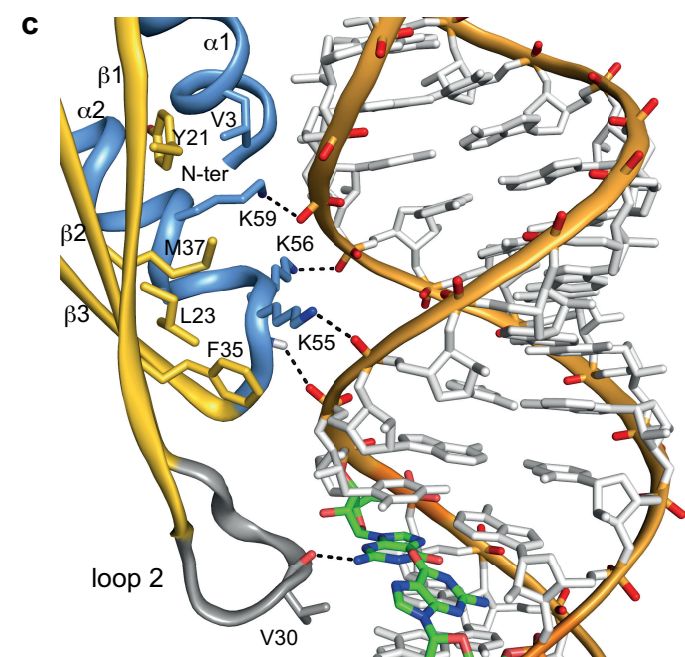
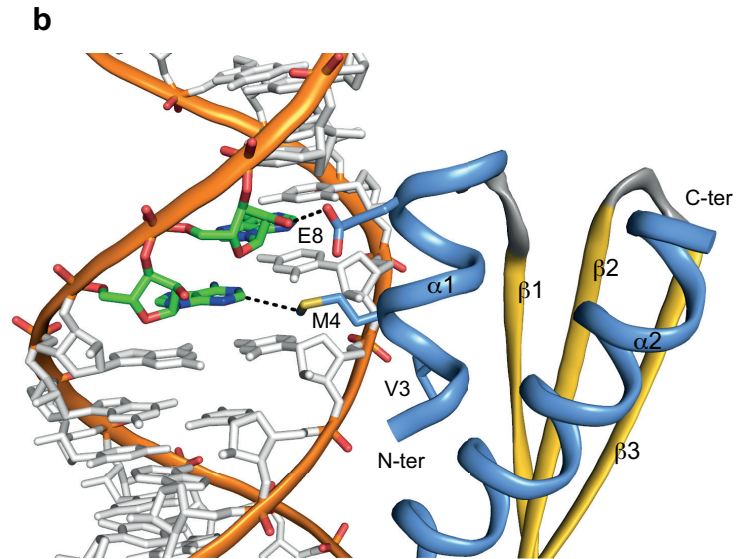
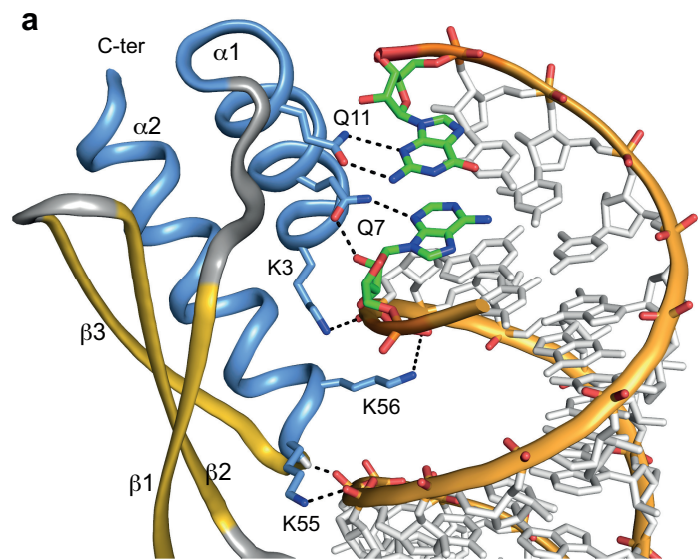
**d**

## minor groove



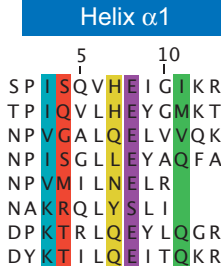
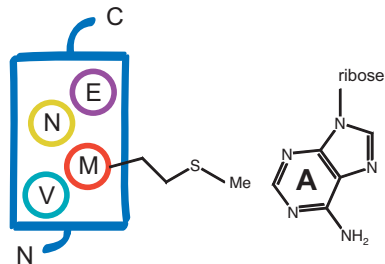




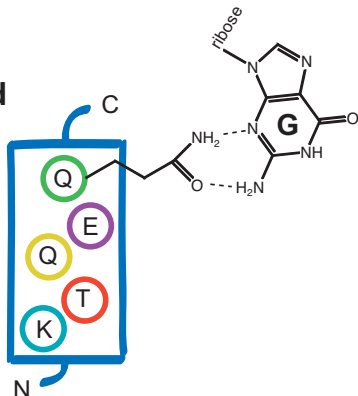


**a**Helix  $\alpha 1$ **b**

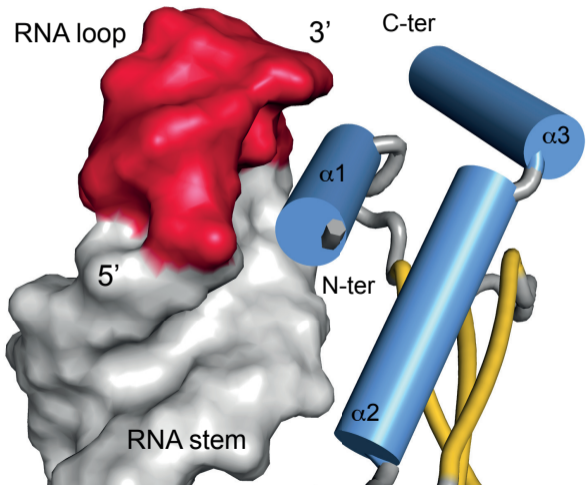
DmStaufen-ds3|P25159|578-590  
 HsPACT-ds1|O75569|34-46  
 HsTRBP-ds2|Q15633|159-171  
 HsADAR1-ds1|P55265|503-515  
 HsADAR2-ds2|P78563|232-241  
 ScRnt1|Q02555|369-378  
 EcRNaseIII|P0A7Y0|155-167  
 AaRNaseIII|O67082|151-163

**c**

HsADAR2-ds2

**d**

Aa RNase III

**a****b**