

# Molecular basis of double-stranded RNA–protein interactions: structure of a dsRNA-binding domain complexed with dsRNA

Jodi M.Ryter<sup>1</sup> and Steve C.Schultz<sup>2</sup>

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309-0215, USA

<sup>1</sup>Present address: Institute of Molecular Biology, 1229 University of Oregon, Eugene, OR 97403-1229, USA

<sup>2</sup>Corresponding author

**Protein interactions with double-stranded RNA (dsRNA) are critical for many cell processes; however, in contrast to protein–dsDNA interactions, surprisingly little is known about the molecular basis of protein–dsRNA interactions. A large and diverse class of proteins that bind dsRNA do so by utilizing an ~70 amino acid motif referred to as the dsRNA-binding domain (dsRBD). We have determined a 1.9 Å resolution crystal structure of the second dsRBD of *Xenopus laevis* RNA-binding protein A complexed with dsRNA. The structure shows that the protein spans 16 bp of dsRNA, interacting with two successive minor grooves and across the intervening major groove on one face of a primarily A-form RNA helix. The nature of these interactions explains dsRBD specificity for dsRNA (over ssRNA or dsDNA) and the apparent lack of sequence specificity. Interestingly, the dsRBD fold resembles a portion of the conserved core structure of a family of polynucleotidyl transferases that includes RuvC, MuA transposase, retroviral integrase and RNase H. Structural comparisons of the dsRBD–dsRNA complex and models proposed for polynucleotidyl transferase–nucleic acid complexes suggest that similarities in nucleic acid binding also exist between these families of proteins.**

**Keywords:** crystal structure/double-stranded RNA/double-stranded RNA-binding domain/protein–nucleic acid complex/protein–RNA interactions

## Introduction

The double-stranded RNA-binding domain (dsRBD) is a 65–70 amino acid sequence/structure motif that mediates dsRNA interactions in a large variety of proteins (St Johnston *et al.*, 1992) such as the dsRNA-dependent protein kinase PKR (Meurs *et al.*, 1990; Green and Mathews, 1992; Thomis *et al.*, 1992), *Drosophila* staufen protein (St Johnston *et al.*, 1991), *Escherichia coli* RNase III (March *et al.*, 1985), human Tar RNA-binding protein (Gatignol *et al.*, 1991), *Xenopus laevis* Xlrba (St Johnston *et al.*, 1992), monomeric RNA helicases (Gibson and Thompson, 1994), the N-terminal regions of eukaryotic RNases H1 (Cerritelli and Crouch, 1995; Cerritelli *et al.*, 1998) and dsRNA-dependent adenosine deaminases (Kim *et al.*, 1994; O'Connell *et al.*, 1995). Binding of the

dsRBD is highly specific for dsRNA, with little or no observable binding to ssRNA, dsDNA or ssDNA (St Johnston *et al.*, 1992; Bass *et al.*, 1994; Bevilacqua and Cech, 1996). Certain partially double-stranded RNA structures such as adenovirus VAI RNA (Galabru *et al.*, 1989; Mellitis *et al.*, 1990), Epstein–Barr virus (EBV) EBER-1 and EBER-2 RNAs (Clarke *et al.*, 1990; Sharp *et al.*, 1993) and RNase III processing signals (Krinke and Wulff, 1990; Chelladuri *et al.*, 1994) are also specifically recognized by certain dsRBD-containing proteins. Binding of the dsRBD appears to be independent of RNA sequence (Manche *et al.*, 1992; St Johnston *et al.*, 1992; Polson and Bass, 1994; Schweigert *et al.*, 1994; Krovat and Jantsch, 1996; Eckmann and Jantsch, 1997).

Protein structures of isolated dsRBDS from *E.coli* RNase III and *Drosophila* staufen protein as determined by NMR show an  $\alpha$ - $\beta$ - $\beta$ - $\alpha$  topology in which the N- and C-terminal  $\alpha$ -helices pack against one face of a three-stranded antiparallel  $\beta$ -sheet (Bycroft *et al.*, 1995; Kharrat *et al.*, 1995). The N-terminal domain of ribosomal protein S5 also contains this same basic fold (Ramakrishnan and White, 1992; Bycroft *et al.*, 1995). Residues implicated in RNA binding cluster on one face of the dsRBD near the beginning of the C-terminal  $\alpha$ -helix (Bycroft *et al.*, 1995; Kharrat *et al.*, 1995). We report here the crystal structure of the second dsRBD of Xlrba (Xlrba-2) complexed with dsRNA at 1.9 Å resolution. The structure of this complex provides a precise description of the dsRBD–dsRNA interaction.

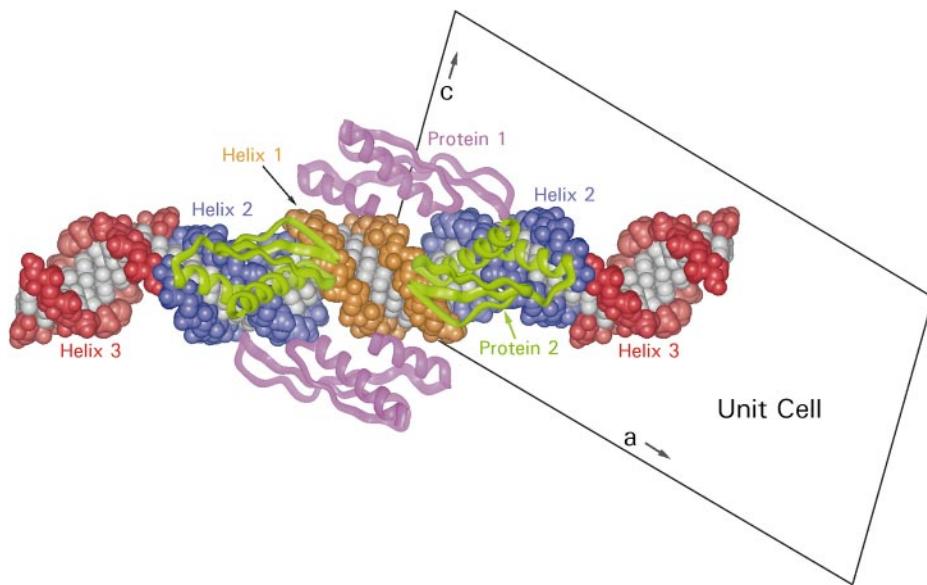
## Results and discussion

### Arrangement of molecules in the crystal

In crystallizing this presumably non-sequence-specific protein–dsRNA complex, RNAs of various lengths were explored with the hope that a crystal packing scheme would be obtained which would uniquely position the protein relative to the RNA. Crystals that diffract to 1.9 Å resolution were ultimately obtained with a self-complementary RNA 10mer (GGCGCGCGCC) which, indeed, provided a unique packing scheme.

Within the crystals, the 10 bp dsRNA helices stack end-to-end as a pseudo-continuous helix, with their long axes along the diagonal of the unit cell (Figure 1). The asymmetric unit contains one full RNA duplex and two ‘half-duplexes’, referred to as RNAs (or helices) 1, 2 and 3. Whereas all 10 bp of RNA 2 are contained within the asymmetric unit, RNAs 1 and 3 are bisected by crystallographic 2-fold rotation axes such that only half of each of these helices is contained in the asymmetric unit. RNAs 1 and 2 both interact with Xlrba-2 protein but, interestingly, RNA 3 does not.

Each asymmetric unit also contains two Xlrba-2 molecules, referred to as proteins 1 and 2 (Figure 1). The



**Fig. 1.** Packing of RNA and Xlrpbpa-2 in the unit cell. The view is along  $b$  with the  $a$  and  $c$  unit cell axes displayed. Individual RNA 10mer helices are shown as space fill representations in gold (helix 1), blue (helix 2) and red (helix 3).  $\alpha$ -Carbon traces of independent observations of the protein are shown in purple (protein 1) and green (protein 2). The asymmetric unit contains protein 1, protein 2, half of helix 1, all of helix 2 and half of helix 3. Protein 1 and protein 2 bridge the same RNA–RNA junction between helices 1 and 2 in precisely the same manner, but with opposite orientations. RNA 3 is not contacted by protein. The extended loop of protein 1 (which interacts with the minor groove of helix 2) is well ordered, whereas the analogous loop of protein 2 is disordered.

proteins are flipped head to tail and rotated 90° around the long axis of the RNA relative to each other. Both proteins span the junction between RNAs 1 and 2 in precisely the same manner.

#### Protein structure

Xlrpbpa-2 has the same  $\alpha$ - $\beta$ - $\beta$ - $\beta$ - $\alpha$  fold seen in the NMR structures of the dsRBD of RNase III and of the third dsRBD of *Drosophila* staufen protein (Bycroft *et al.*, 1995; Kharrat *et al.*, 1995). In each of these structures, the N- and C-terminal  $\alpha$ -helices pack against one face of a three-stranded antiparallel  $\beta$ -sheet (Figure 2A). The only significant difference between the NMR structures of the dsRBD and the crystal structure of Xlrpbpa-2 reported here is in the loop between  $\beta$ -strands 1 and 2, which is poorly structured/determined by NMR. In the crystals, this loop is disordered in protein 2, but is well ordered in protein 1 (Figure 1) and is observed to interact specifically with the dsRNA minor groove. The remaining portions of protein 1 and protein 2 are nearly identical, and superimpose with root mean square (r.m.s.) deviations of 0.41 Å for the peptide backbone atoms and 0.77 Å for all non-hydrogen atoms.

#### Protein–RNA interactions

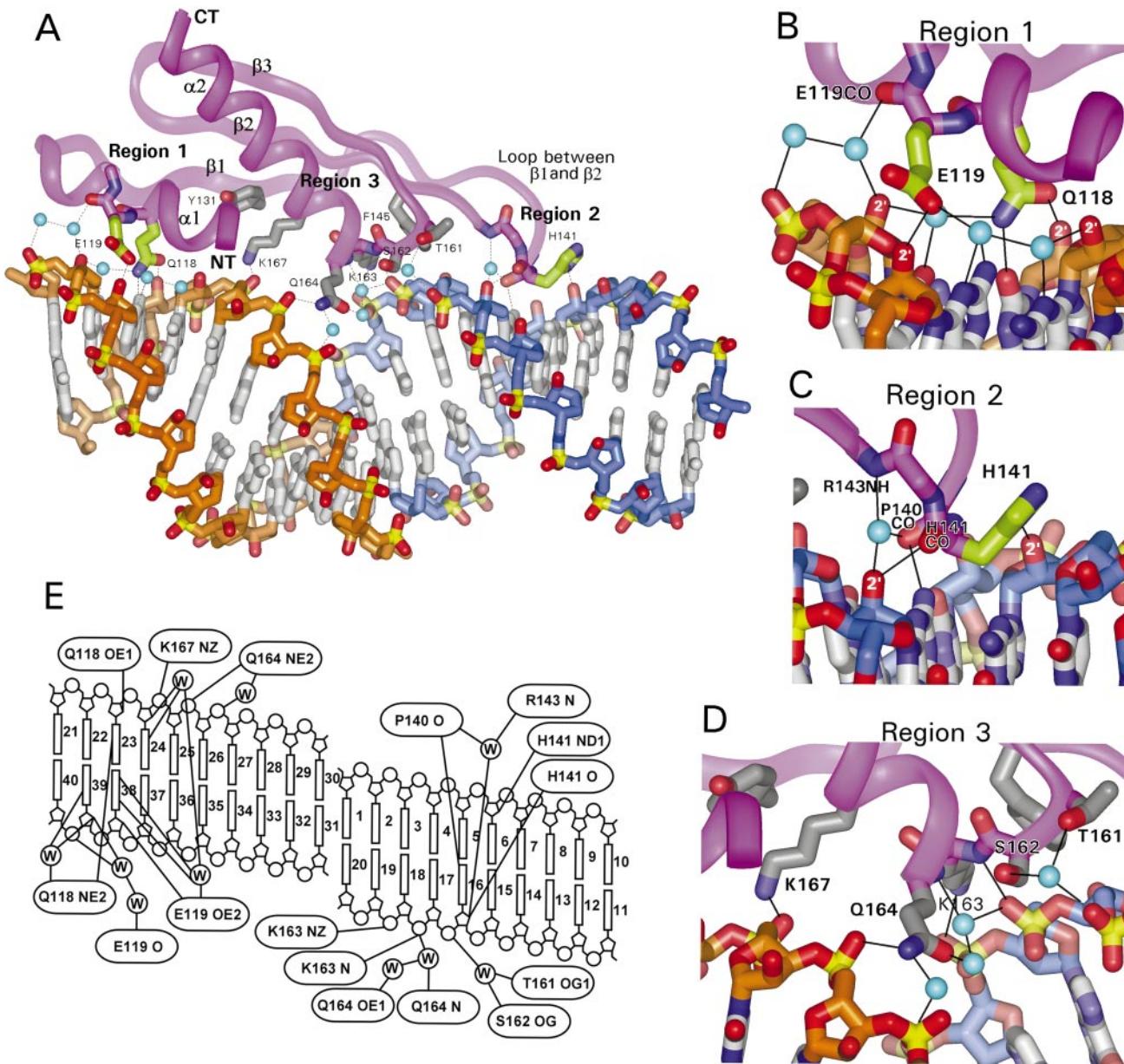
Xlrpbpa-2 interacts with two successive minor grooves and across the intervening major groove on one face of the dsRNA helix (Figures 1 and 2). These interactions span a total of 16 bp and collectively bury a surface area of 1680 Å<sup>2</sup>. The interactions can be divided into three regions (Figure 2): (region 1) interaction of the N-terminal  $\alpha$ -helix with the RNA minor groove; (region 2) interaction of the loop between  $\beta$ -strands 1 and 2 with the RNA minor groove; and (region 3) interaction of the C-terminal  $\alpha$ -helix across the RNA major groove.

In region 1, the N-terminal three-turn  $\alpha$ -helix interacts

with the minor groove of the RNA (Figure 2B and E). Two side chains (Q118 and E119) and one peptide backbone group (CO of E119) are involved in direct and water-mediated interactions with four 2'-OH groups and five base functional groups in the minor groove of the RNA (Figure 2B). These direct and water-mediated interactions satisfy all of the hydrogen-bonding potential in this region of the RNA minor groove. The N-terminal  $\alpha$ -helices of proteins 1 and 2, which were refined independently and are in different packing environments in the crystals, interact with RNAs 1 and 2, respectively, in precisely the same manner, including the water-mediated interactions.

In region 2, the loop between  $\beta$ -strands 1 and 2 interacts with the adjacent minor groove of the RNA (Figure 2C and E). One amino acid side chain (H141) and three peptide backbone groups (CO of P140, CO of H141 and NH of R143) make direct and water-mediated interactions with two 2'-OH groups and one base functional group in the RNA minor groove (Figure 2C). At the very tip of the peptide loop, the backbone CO of His141 interacts with a 2'-OH group on one side of the minor groove, and the side chain interacts with a 2'-OH on the other side such that this one residue bridges the minor groove to interact with 2'-OH groups in both strands of the dsRNA. Mutational analyses have demonstrated that His141 is important for the dsRNA–Xlrpbpa interaction (Krovat and Jantsch, 1996). Interestingly, these His141 interactions require an RNA phosphodiester backbone conformation different from that of ideal A-form RNA, as will be described later.

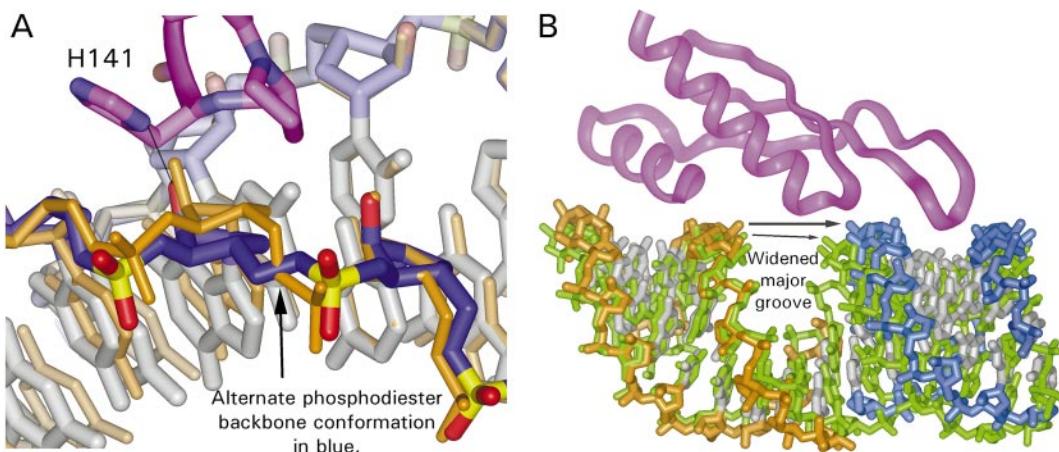
In region 3, the protein interacts across the major groove of the RNA. A total of six non-bridging oxygens of the phosphodiester backbone are contacted either directly or via water molecules (Figure 2D and E). A surprisingly short peptide segment just preceding and including the N-terminal portion of the C-terminal  $\alpha$ -helix (residues 161–



**Fig. 2.** Interactions between the Xlrbp-2 dsRBD and dsRNA. (A) Overview of the dsRNA-Xlrbp-2 interactions. The protein interacts with two successive minor grooves and across the intervening major groove on one face of the dsRNA helix. Protein 1 is shown as an  $\alpha$ -carbon trace in purple. Side chain residues interacting with the RNA minor groove in regions 1 and 2 are colored green, and side chains interacting with the RNA major groove in region 3 are colored gray. Also displayed in gray are the side chains of Y131 and F145, which serve to position K167 and K163, respectively. Oxygen atoms are shown in red, nitrogen atoms in blue and phosphorus in yellow. Hydrogen bonds are indicated as black lines. Residue numbers are those of the full-length Xlrbp-2 protein. (B) Expanded view of region 1 in which the N-terminal  $\alpha$ -helix interacts with the RNA minor groove. (C) Expanded view of region 2 in which the loop between  $\beta$ -strands 1 and 2 interacts with the minor groove. (D) Expanded view of region 3 in which the residues within and just preceding the N-terminal end of the C-terminal helix interact across the major groove. (E) Schematic detailing the interactions between Xlrbp-2 and dsRNA.

167) provides all of the major groove interactions. On the side of the major groove proximal to region 2, the backbone amide NH groups of residues 163 and 164, which are at the N-terminal end of the five-turn  $\alpha$ -helix (the positive end of the helical dipole), interact with a non-bridging oxygen of the phosphodiester backbone. The side chain of K163 hydrogen-bonds directly to an adjacent phosphodiester group on this same side of the major groove, and amino acid side chains further along in the  $\alpha$ -helix (Q164 and K167) interact on the other side of the major groove (proximal to region 1) such that the  $\alpha$ -helix bridges the major groove. Q164 is generally Lys in other

dsRBPs, which would probably interact either with the same phosphate as Q164 or with the adjacent phosphate group. The side chains of K163 and K167 pack against the aromatic side chains of F145 and Y131, respectively, in a manner that appears to position stably these long, flexible lysine side chains. Mutational analyses of Xlrbp-2 demonstrate that F145 and the C-terminal  $\alpha$ -helix are indeed critical for RNA binding (Krovat and Jantsch, 1996), and mutations in the third dsRBD of *Drosophila* staufen protein at positions analogous to F145, K163, Q164 and K167 of Xlrbp-2 disrupt RNA binding (Bycroft *et al.*, 1995).



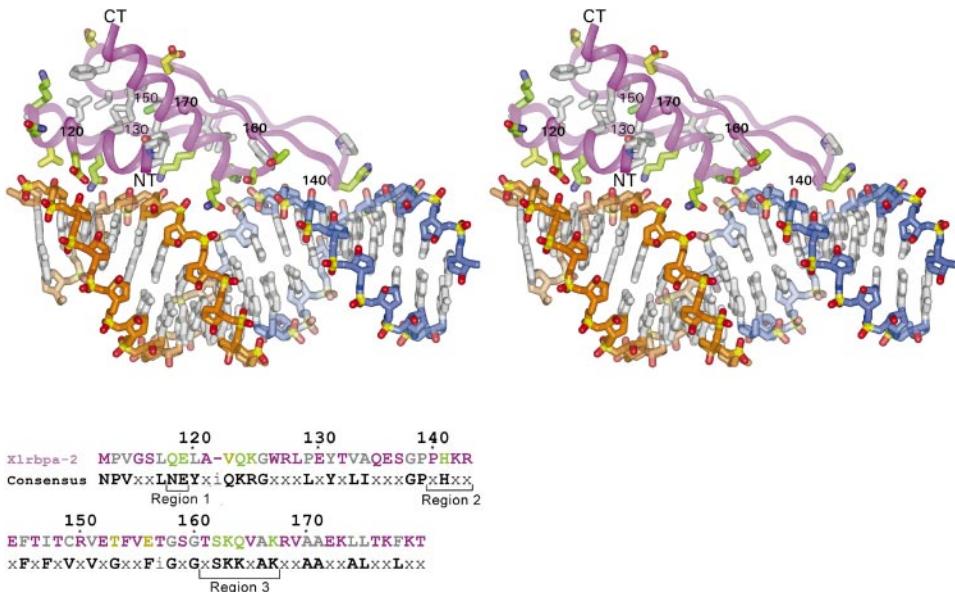
**Fig. 3.** Deviations from normal A-form RNA as observed in the structure of the dsRNA–Xlrpba-2 complex. (A) Alternative phosphodiester backbone conformation at nucleotide 6. The backbone of RNA helix 2 is shown in blue and the bases are shown in white. Helix 1, as superimposed on the first 5 bp of RNA 2, is shown in gold. Rotations around  $\alpha$  and  $\gamma$  of nucleotide 6 of helix 2 alter the position of the 2'-OH of this nucleotide so that it can interact with H141. The 2'-OH of the analogous nucleotide in helix 1 (in gold), which is in an A-form conformation, would collide with the H141 side chain. (B) Protein 1 is shown in purple, RNA helix 1 in gold, RNA helix 2 in blue and idealized A-form RNA (superimposed on helix 1) in green. RNA helices 1 and 2 are stacked in a way that expands the major groove 2–3 Å relative to an A-form RNA helix, suggesting that a widened major groove is important for interaction of Xlrpba-2 with dsRNA.

This general scheme for interaction with the dsRNA major groove, in which backbone NH groups at the N-terminal end of an  $\alpha$ -helix hydrogen-bond to non-bridging oxygens of the phosphodiester backbone, and side chains emanating from the first 1–2 turns of this same helix interact with phosphodiester groups on the other side of the major groove, appears to represent a feature that can discriminate between different major groove conformations in double-helical nucleic acids. Although this type of major groove interaction has not yet been observed in structures of other protein–RNA complexes, it does exist in the structure of the histone core particle (Luger *et al.*, 1997). In the histone core particle, backbone NH groups at the N-termini of seven different  $\alpha$ -helices hydrogen-bond to DNA phosphodiester groups, but only for  $\alpha$ 1 of H3 and  $\alpha$ 1 of H2B do side chain residues emanating from the same helix interact across the DNA major groove. These interactions occur near regions of greatest DNA bending and appear to require the narrowed major groove created by the bend. The shape of the major groove in straight, canonical B-form DNA would not permit such interactions. Conformation-specific recognition of the narrow major groove of an A-form RNA double helix (versus the wider, shallower major groove of straight, canonical B-DNA) via this  $\alpha$ -helix interaction probably accounts, in part, for the binding preference of the dsRBD for dsRNA. Type B dsRBDs, which contain amino acid sequence homologies only in the C-terminal one-third of the dsRBD rather than throughout the ~70 amino acid motif (St Johnston *et al.*, 1992; Krovat and Jantsch, 1996), contain all of the residues involved in these major groove interactions (region 3), but are not conserved in the regions involved in minor groove interactions.

Discrimination between dsRNA and dsDNA via the minor groove (Bevilacqua and Cech, 1996) would appear to involve direct and water-mediated interactions with 2'-OH groups. The 2'-OH groups of RNA line the entire minor groove and present a very different chemical character from that which exists in the minor groove of 2'-deoxyribonucleic acids. The many direct interactions

between Xlrpba-2 and 2'-OH groups in the minor groove would be expected to strongly favor binding to dsRNA over dsDNA. Other structures of proteins bound to double-helical RNAs, including those of tRNA synthetase–tRNA complexes (Rould *et al.*, 1989; Ruff *et al.*, 1991), flock house virus (Fisher and Johnson, 1993) and satellite tobacco mosaic virus (Larson *et al.*, 1993, 1998) show binding primarily to the minor grooves of double-helical RNA segments. In each of these complexes, the proteins interact with 2'-OHs, phosphodiester backbone groups and base functional groups within and along the dsRNA minor groove.

dsRBD–dsRNA interactions are presumably sequence independent (Manche *et al.*, 1992; St Johnston *et al.*, 1992; Polson and Bass, 1994; Schweiguth *et al.*, 1994; Krovat and Jantsch, 1996; Eckmann and Jantsch, 1997), although as yet unidentified specificities might still exist. A majority of the interactions in regions 1, 2 and 3 involve the phosphodiester backbone and 2'-OH groups, which would be expected to be sequence independent. The minor groove–base interactions might yield sequence information, but most of those observed in the crystals are via water molecules which could probably adjust their minor groove hydrogen-bonding patterns to accommodate various sequences. Of the two direct interactions with base groups in the minor groove, Q118 interacts with a hydrogen bond acceptor group that would be presented similarly by any of the four nucleotides at this position. However, the other direct base interaction, in which a backbone carbonyl of Pro140 interacts with the exocyclic amino group of a G in region 2, would appear to be highly specific for a GC base pair at this position and warrants further investigation. The structures of intact RNA viruses have provided additional examples of presumably non-sequence-specific protein–ssRNA complexes (Chen *et al.*, 1989) and protein–dsRNA complexes (Fisher and Johnson, 1993; Larson *et al.*, 1993). As in the dsRNA–Xlrpba-2 complex, these proteins primarily utilize interactions with 2'-OH and phosphodiester backbone groups, but also



**Fig. 4.** dsRBD consensus sequence as it relates to the Xlrbp-a-2 structure and RNA interactions. Residues of Xlrbp-a-2 that are conserved as hydrophobic/non-hydrogen-bonding residues are shown in white, and residues that are conserved as hydrophilic/hydrogen-bonding residues are shown in green. Residues in yellow are those of Xlrbp-a-2 that do not match the consensus sequence. Side chain oxygen atoms are shown in red and nitrogens in blue. The CO<sub>2</sub>s of residues 120, 130, 140, 150, 160 and 170 are labeled and shown as purple spheres within the  $\alpha$ -carbon backbone trace. Color coding of the Xlrbp-a-2 amino acid sequence aligned with the consensus sequence is the same as in the structure, with purple representing the non-conserved residues. In the consensus sequence, x indicates non-conserved residues, i indicates positions in which additional residues are present in certain dsRBDs, and (–) indicates positions commonly deleted in dsRBDs. Residues involved in interactions in regions 1, 2 and 3 (Figure 2) are indicated.

include limited base interactions along the RNA minor groove.

Sequence-dependent structural features of the RNA are also possible sources of specificity. Interestingly, we observe two deviations from idealized A-form RNA in the structure of the dsRNA-Xlrbp-a-2 complex, which suggest that novel sources of specificity might indeed exist.

## *RNA structure*

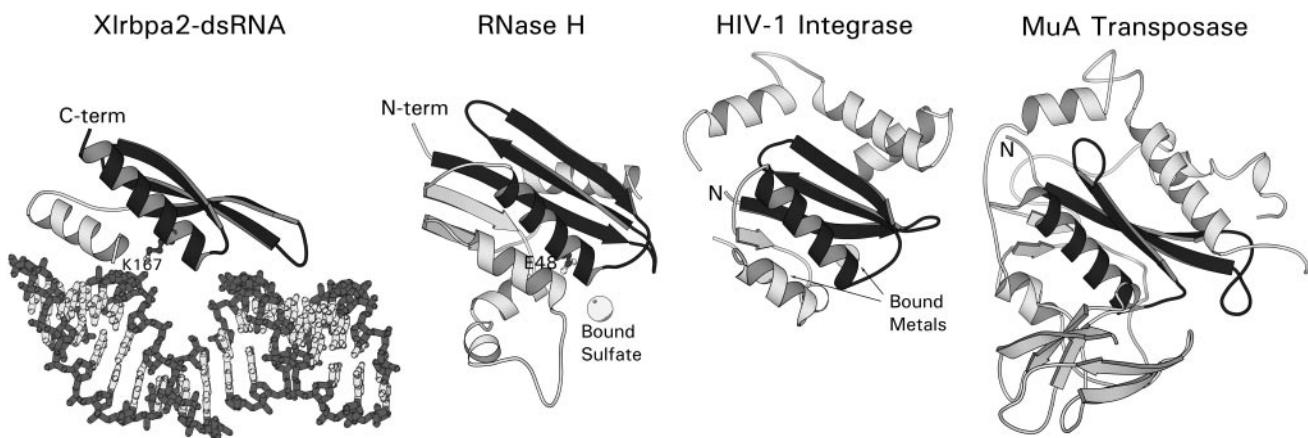
The three dsRNAs in the asymmetric unit individually form nearly ideal A-form helices, although they are underwound slightly (11.8, 12.5 and 11.4 bp per turn for helices 1, 2 and 3, respectively, compared with 11.0 for A-form RNA). Idealized A-form RNA helices superimpose on helices 1, 2 and 3 with r.m.s. deviations of 1.4, 2.3 and 2.0 Å, respectively (for the phosphodiester backbone groups).

One of the two deviations from idealized A-form RNA observed in the complex involves the phosphodiester backbone conformation at nucleotide 6 of helix 2, where the  $\alpha$  and  $\gamma$  torsion angles are  $159^\circ$  and  $176^\circ$ , respectively, rather than  $\sim 300^\circ$  and  $\sim 50^\circ$  for an A-form conformation (Figure 3A). This alternative conformation appears to be induced by interaction of the 2'-OH of this nucleotide with the side chain of H141. In an idealized A-form conformation, the 2'-OH would be  $\sim 0.5$  Å closer to the H141 side chain, causing a steric clash (Figure 3A). None of the other 39 nucleotides in the asymmetric unit exhibits a backbone conformation that differs significantly from that of A-form, including the position in helix 1 that corresponds to nucleotide 6 of helix 2. Accordingly, an interaction is not observed between the His141 side chain of protein 2 and RNA 1. Rather, the loop containing

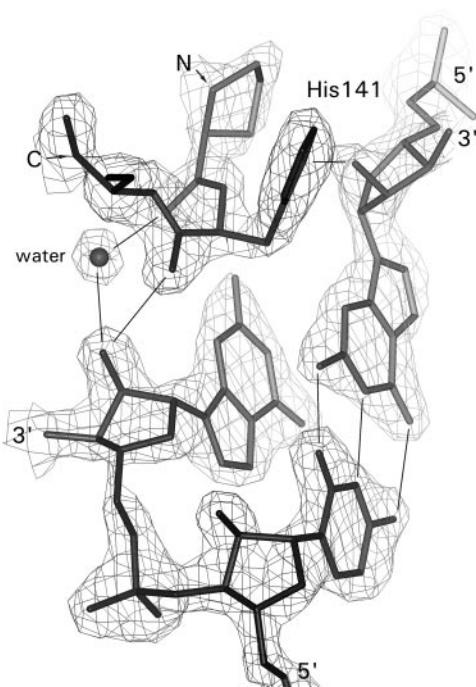
His141 in protein 2 is disordered in the crystals. The H141 side chain interaction apparently induces/requires this local deviation from an A-form conformation.

A second feature of the dsRNA-Xlrbp-2 interaction that involves a departure from A-form RNA occurs at the stacking junction between adjacent 10 bp dsRNA helices where the protein spans the RNA major groove. This junction between RNA duplexes involves a cross-strand purine-purine stack in which the 5'-guanine bases of adjacent duplexes stack on each other and the bridging oxygens of the ribose rings stack on the 3'-cytosine bases of the adjacent duplex. The two independent RNA-RNA junctions in the crystals (helix 1-helix 2 and helix 2-helix 3) exhibit identical stacking arrangements. Since helix 3 does not interact with protein, the structure of these junctions is apparently determined by the forces involved in the stacking interactions and not by interactions with the protein.

The cross-strand purine-purine stack of these RNA-RNA junctions results in a major groove that is 2-3 Å wider than that of an idealized A-form dsRNA helix (Figure 3B). This widened major groove creates the binding site for Xlrbp-2. Whether interactions with a continuous dsRNA helix would involve altering the structures of the protein, the RNA or both relative to their conformations in the crystal is an important question. We believe that it is likely that a continuous segment of dsRNA would need to be distorted from an ideal A-form helix into a structure that contains a widened major groove. The protein side chain and peptide backbone groups implicated by sequence comparisons, mutational studies and structural studies as important for dsRNA interactions are rather rigidly positioned in the structure, so it seems



**Fig. 5.** Comparison of the dsRBD and polynucleotidyl transferases. The  $\beta$ - $\beta$ - $\beta$ - $\alpha$  folds of Xlrpbpa-2, RNase H, HIV-1 integrase and MuA transposase are shown in black. Notice that the same portion of this fold (the loop between  $\beta$ -strands 1 and 2 and the N-terminal portion of the C-terminal  $\alpha$ -helix) interacts or is proposed to interact with nucleic acid substrate. The bound sulfate in RNase H is thought to represent a phosphate-binding site, and the bound metals of HIV-1 integrase are thought to participate in catalysis. The side chains of the conserved Lys (K167) of the dsRBD (which is Glu in RNaseIII) and the conserved Glu (E48) of RNase H are displayed and labeled.



**Fig. 6.** Simulated-annealed omit electron density map calculated at 1.9 Å resolution. Contoured at  $2\sigma$ .

unlikely that they could be repositioned readily to accommodate a narrower major groove.

The potential importance of a widened major groove in the dsRNA-Xlrpbpa-2 interaction gives rise to the interesting possibility that this is an RNA structure-specific feature of the dsRBD-dsRNA interactions. Since the central nucleotides of the 16 bp dsRNA-binding site are not contacted by the protein (Figure 2E), a variety of local structural variations, such as slight local underwinding, could give rise to a widened major groove. Regions of non-Watson-Crick base pairing might result in conformations that define dsRBD-binding sites. Natural substrates for specific cleavage by RNase III contain duplex regions of ~10 bp separated by internal loops without Watson-Crick base pairing (Dunn and Studier, 1983; March *et al.*,

1985; Krinke and Wulff, 1990; Chelladuri *et al.*, 1994). PKR specifically interacts with structured RNAs that contain only short segments of uninterrupted Watson-Crick base pairing, such as adenovirus VAI RNA (Galabru *et al.*, 1989; Mellitis *et al.*, 1990) and the EBV RNAs EBER-1 and EBER-2 (Clarke *et al.*, 1990; Sharp *et al.*, 1993). RNAs selected for binding to the dsRBD-binding domain of PKR sometimes contain GA:AG base pairs (Bevilacqua *et al.*, 1998) which can widen the RNA major groove (Wu *et al.*, 1997). Perhaps non-Watson-Crick base-paired regions between helical segments allow formation of structures important for specific dsRBD-dsRNA interactions.

#### Consensus sequence

The amino acid residues conserved among various dsRBDS correlate extremely well with important features of the dsRNA-Xlrpbpa-2 complex as observed in our structure. The amino acid sequence of Xlrpbpa-2 is either identical or similar to the dsRBD consensus sequence in 32 of 37 residues (Figure 4) (Krovat and Jantsch, 1996). The hydrophobic/non-hydrogen-bonding residues of the consensus sequence comprise the core of Xlrpbpa-2 and are likely to be important for folding and stability. Strikingly, nearly all of the conserved hydrophilic/hydrogen-bonding residues are positioned along the surface of Xlrpbpa-2 that interacts with RNA and include all of the residues that directly contact the RNA (Figure 4). Of the five residues of Xlrpbpa-2 that do not match the consensus sequence, one is the N-terminal Met, which is an Asn in the consensus sequence, and another is a Pro, which in several dsRBDS is inserted between residues analogous to E156 and T157. The remaining three residues of Xlrpbpa-2 that do not correspond to the consensus sequence are shown in yellow in Figure 4. Val122 of Xlrpbpa-2, which would correspond to a Gln in the consensus sequence, packs against a sugar ring in the minor groove of the RNA. One could easily imagine that a Gln located at this position could hydrogen-bond to the 2'-OH of this same ribose ring. The other two residues of Xlrpbpa that do not match the consensus sequence are on the surface of the protein opposite the RNA-binding surface and are present near

**Table I.** Crytallographic data

Native data	
Completeness (99–1.89 Å)	90.8%
Completeness (99–2.04 Å)	97.2%
Completeness (1.96–1.89 Å)	50.0%
(Ave $I$ )/(Ave error $I$ )	
(99–1.89 Å)	2543.8/176.6 = 14.4
(1.96–1.89 Å)	272.0/110.1 = 2.5
$R_{\text{sym}}$ (99–1.89 Å)	6.3%
$R_{\text{sym}}$ (1.96–1.89 Å)	28.7%
Derivative data: r(GGCGCGd <sup>t</sup> CGCC)	
Resolution	15.0–2.6 Å
$R_{\text{sym}}$	8.0%
$R_{\text{iso}}$	26.5%
No. of sites	4
$R_{\text{Cullis}}$	0.75
Phasing power	0.82
Refinement statistics ( $F \geq 0$ )	
Resolution	15.0–1.9 Å
Unique reflections	24 336
No. of atoms	1884 + 373 oxygens from H <sub>2</sub> O
$R$ -value/ $R_{\text{free}}$ (10% of data)	22.7%/25.8%
R.m.s.d. from ideal:	
Bond lengths	0.009 Å
Bond angles	1.304°
Ramachandran analysis (PROCHECK)	
	Protein 1      Protein 2
Most favored regions	94.8%      94.0%
Allowed regions	3.4%      4.0%
Generously allowed	0.0%      2.0% (E152)
Disallowed	1.7% (E152)      0.0%

$$R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$$

$$R_{\text{iso}} = \sum |F_{\text{PH}} - F_{\text{P}}| / \sum F_{\text{PH}}$$

$$R_{\text{Cullis}} (\text{centric}) = \sum |F_{\text{PH}} - |F_{\text{P}} + F_{\text{Hcalc}}|| / \sum |F_{\text{PH}} - F_{\text{P}}|$$

$$\text{Phasing power} = |F_{\text{Hcalc}}| / |F_{\text{PH}} - |F_{\text{P}} + F_{\text{Hcalc}}||$$

$$R\text{-value} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$$

regions that contain insertions or deletions in other dsRBDS.

There is a very satisfying correspondence between residues found to be conserved in dsRBDS, mutational analysis of the dsRBD and the structure of the complex. The dsRNA–Xlrbp-2 structure presented here illustrates how the residues conserved in dsRBDS provide a stable protein structure that specifically binds dsRNA in a sequence-independent manner, as is required for the many diverse biological functions performed by dsRNA-binding proteins.

The structure of the dsRNA–Xlrbp-2 complex also presents a possible scheme by which multiple dsRBDS could cooperate to bind dsRNA. Since the dsRBD–dsRNA interaction involves only one face of the dsRNA helix, additional interactions could occur around the remaining surfaces of a dsRNA helix. In the crystals, protein 1 and protein 2 interact with different sides of the same dsRNA helices (Figure 1). The N-terminus of protein 1 is <25 Å from the C-terminus of protein 2, such that a relatively short polypeptide segment would be required to link these separate dsRBDS. Proteins that contain multiple dsRBDS might use a similar scheme in which the dsRBDS would ‘wrap’ around the dsRNA helix to enhance and extend the interactions with dsRNA.

### Similarities between the dsRBD and polynucleotidyl transferases

The β-β-β-α portion of the dsRBD resembles the N-terminal portion of the recently described conserved core

structure of polynucleotidyl transferases such as MuA transposase, retroviral integrases, RuvC and RNase H (Figure 5) (Ariyoshi *et al.*, 1994; Rice and Mizuuchi, 1995; Yang and Steitz, 1995). The structure of the dsRBD–dsRNA complex suggests that this folding motif interacts similarly with double-helical nucleic acids in these different proteins. For example, in the β-β-β-α fold of the RNase H domain of HIV-1 reverse transcriptase, the N-terminal end of the α-helix is positioned near the phosphate backbone, and the loop between β-strands 1 and 2 appears to interact with the minor groove of the dsDNA (Jocobo-Molina *et al.*, 1993) (although actual substrates for RNase H are RNA–DNA hybrids and not dsDNA). This same general arrangement is observed in the dsRBD–dsRNA complex. Proposed models for the nucleic acid complexes of *E.coli* RNase H (Yang *et al.*, 1990) and RuvC (Ariyoshi *et al.*, 1994) also place the loop between β-strands 1 and 2 near the minor groove of a nucleic acid helix and, for RNase H, the N-terminal end of the α-helix is proposed to interact with the phosphate backbone. Perhaps the dsRBD/polynucleotidyl transferase β-β-β-α fold provides an especially effective means of positioning the N-terminal end of an α-helix immediately adjacent to a β-loop in a manner that can be adapted readily for interactions with double-helical nucleic acids.

In addition to these general similarities, a short segment of amino acids in *E.coli* RNase H (**AAIVALEAL**) precisely matches the C-terminal portion of the dsRBD consensus sequence (**AAXXXALXXL**), and these residues align surprisingly well in the structures of Xlrbp-2 and RNase H. Intriguingly, three residues prior to this sequence, RNase III contains a Glu rather than the dsRBD consensus Lys. This Glu corresponds to a conserved Glu of RNase H (Yang *et al.*, 1990) which binds a metal that is involved in RNA cleavage (Davies *et al.*, 1991). The close resemblance of Xlrbp-2 and the dsRBD of RNase III enables one to reliably position this Glu in place of K167 immediately adjacent to a dsRNA phosphodiester backbone (Figure 2). This would be an excellent location for Glu (together with residues from the N-terminal domain) to participate in metal binding and the RNase III function of cleaving the phosphodiester backbone of certain dsRNAs. These potential similarities in RNA binding and cleavage by RNase III and RNase H relate the diverse families of polynucleotidyl transferases and dsRNA-interacting proteins on a more fundamental level than previously anticipated.

## Materials and methods

### Protein purification

Xlrbp-2 was expressed as inclusion bodies in *E.coli* Bl21(DE3) pLysS using an inducible pET3 T7-based expression system [a generous gift from Michael Jantsch, University of Vienna (Krovat and Jantsch, 1996)]. Following cell lysis (by sonication), the inclusion bodies were isolated by centrifugation and extracted with lysozyme and deoxycholate as described by Langley *et al.* (1987). The protein was solubilized in 7 M guanidine hydrochloride and refolded by dialysis into 20 mM HEPES (pH 7.5), 50 mM NaCl, 0.25 mM EDTA, 0.02% NaN<sub>3</sub> and 2.0 mM dithiothreitol (DTT). Refolded protein was purified using an S-Sepharose (Pharmacia) FPLC column and a Superdex FPLC gel filtration column (Pharmacia). Final yields were typically 25 mg/l of bacterial cell culture.

### RNA purification

RNAs were generated by transcription of synthetic DNA oligonucleotides with T7 RNA polymerase (Milligan *et al.* 1987). Following transcription,

NTPs and salts were removed by gel filtration chromatography, and the 5'-triphosphate groups were removed by treatment with calf intestinal alkaline phosphatase. The RNAs were purified using a Nucleopac PA-100 anion exchange HPLC column (Dionex) heated to 85°C. The RNAs were then concentrated and exchanged into 1 mM Tris-HCl (pH 7.5), 0.1 mM EDTA in a Centricon-3 spin concentrator, and further desaltsed using an HPLC SEC-250 gel filtration column (Bio-Rad). Once crystals were obtained, RNAs used in subsequent crystallizations were synthesized using an Applied Biosystems DNA/RNA synthesizer and purified by HPLC as described above.

### Crystallization

Crystals were grown by hanging drop vapor diffusion using 0.7 mM Xlrba-2, 0.7 mM r(GGCGCGGCC), 225 mM KCl, 50 mM NaCl, 1% polyethylene glycol (PEG) 4000, 12% ethylene glycol, 100 mM MES (pH 5.4), 10 mM DTT and 10 mM  $\beta$ -mercaptoethanol (final conditions). Crystals were harvested into 12% PEG 4000, 12% ethylene glycol, 100 mM MES (pH 6.5), 100 mM KCl, 10 mM MgCl<sub>2</sub> and 11 mM DTT prior to flash freezing in liquid propane. The crystals diffracted to 1.9 Å resolution and were space group C2 with  $a = 110.4$ ,  $b = 58.5$ ,  $c = 58.9$  Å and  $\beta = 105.2^\circ$ .

### Data collection

Native and derivative data were collected from crystals frozen at 113 K using an R-AXIS IIC mounted on a Rigaku RU200 rotating anode generator equipped with the MSC/Yale design focusing mirrors. Data were reduced to reflection intensities using DENZO and SCALEPACK (Otwinowski and Minor, 1996). Data were scaled using SCALEIT in the CCP4 suite of crystallographic programs (CCP4, 1994). Statistics for native and derivative data are listed in Table I.

### Structure determination and refinement

The structure was solved using a single isomorphous heavy atom derivative which was obtained by crystallizing Xlrba-2 with an RNA containing a 5-iododeoxycytidine substitution at position 7 of the self-complementary RNA 10mer [r(GGCGCGd<sup>I</sup>C GCC)]. The heavy atoms were located using difference Patterson and difference Fourier methods, and the positions and occupancies were refined using MLPHARE (Otwinowski, 1991) (Table I). The initial non-solvent flattened single isomorphous replacement (SIR) map calculated at 2.6 Å resolution was of sufficient quality to fit all 20 RNA base pairs into the asymmetric unit [using O (Jones *et al.*, 1991)]. Individual nucleotide positions were refined as single rigid bodies in idealized A-form conformations using XPLOR (Brünger, 1992). Phases calculated from this rigid body refined model (using XPLOR) were combined with the SIR phases using SigmaA (Read, 1986) with damp = 0.1, and these combined phases were modified by solvent flattening using DM (Cowtan, 1994). Maps calculated using the combined/solvent-flattened phases were of sufficient quality to fit portions of the protein molecules using the NMR structure of Bycroft *et al.* (1995) as a guide. The remaining portions of the protein were fit during subsequent iterations of phase combination and solvent flattening. The model was refined at 1.9 Å resolution using XPLOR (Brünger, 1992) and CNS (Brünger *et al.*, 1998). The data were scaled anisotropically in CNS for the final refinement. A total of 373 well-defined water molecules that were within hydrogen-bonding distance of appropriate hydrogen-bonding groups (protein, RNA and/or other water molecules) were added to the structure during refinement. Because of the unusually large number of water molecules, their positions were verified and adjusted using simulated-annealed omit electron density maps calculated using CNS (Brünger *et al.*, 1998). The structure currently is refined to an  $R$ -value of 22.7% and an  $R_{\text{free}}$  of 25.8%. Refinement statistics are listed in Table I. A portion of a simulated-annealed omit electron density map is shown in Figure 6. The coordinates have been deposited in the Brookhaven Protein Data Bank.

All protein residues fall in most favored or allowed regions of Ramachandran plots except for E152. E152 refines to  $\phi =$  approximately  $-120^\circ$  and  $\psi = \sim 60^\circ$  in both protein 1 and protein 2, a generously allowed/disallowed region of Ramachandran plots. This configuration fits the electron density very well. The reason for this unusual configuration is unclear. Perhaps the positioning of this residue between two  $\beta$ -branched residues at the tip of a loop results in unusual configurational constraints.

### Acknowledgements

We thank Michael Jantsch for generously providing Xlrba-2 expression plasmids and for helpful discussions during the course of this work;

Vasili Carperos, Jeff Hansen, Martin Horvath, Olve Peersen and Andrew Berglund for their assistance with the structural work and in preparing this manuscript; and Tom Cech, Olke Uhlenbeck and Deborah Wuttke for critical reading of the manuscript. This work was funded by the Colorado Advanced Technology Institute through a grant received from the Colorado RNA Center.

### References

- Ariyoshi,M., Vassylyev,D.G., Iwasaki,H., Nakamura,H., Shinagawa,H. and Morikawa,K. (1994) Atomic structure of the RuvC resolvase: a Holliday junction-specific endonuclease from *E.coli*. *Cell*, **78**, 1063–1072.
- Bass,B.L., Hurst,S.R. and Singer,J.D. (1994) Binding properties of newly identified *Xenopus* proteins containing dsRNA-binding motifs. *Curr Biol*, **4**, 301–314.
- Bevilacqua,P.C. and Cech,T.R. (1996) Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR. *Biochemistry*, **35**, 9983–9994.
- Bevilacqua,P.C., George,C.X., Samuel,C.E. and Cech,T.R. (1998) Binding of the protein kinase PKR to RNAs with secondary structure defects: role of tandem A–G mismatch and noncontiguous helices. *Biochemistry*, **37**, 6303–6316.
- Brünger,A.T. (1992) *X-PLOR (Version 3.1) Manual*. Yale University Press, New Haven, CT.
- Brünger,A.T. *et al.* (1998) Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr*, **D54**, 905–921.
- Bycroft,M., Grunert,S., Murzin,A.G., Proctor,M. and St Johnston,D. (1995) NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.*, **14**, 3563–3571.
- CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr*, **D50**, 760–763.
- Cerritelli,S.M. and Crouch,R.J. (1995) The non-RNase H domain of *Saccharomyces cerevisiae* RNase H1 binds double-stranded RNA: magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA*, **1**, 246–259.
- Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J. (1998) A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucleic Acids Res.*, **26**, 1834–1840.
- Chelladuri,B.S., Li,H. and Nicholson,A.W. (1994) A conserved sequence element in ribonuclease III processing signals is not required for accurate *in vitro* cleavage. *Nucleic Acids Res.*, **19**, 1759–1766.
- Chen,Z., Stauffacher,C., Li,Y., Schmidt,T., Bomu,W., Kamer,G., Shanks,M., Lomonosoff,G. and Johnson,J.E. (1989) Protein–RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science*, **245**, 154–159.
- Clarke,P.A., Sharp,N.A. and Clemens,M.J. (1990) Translational control by the Epstein–Barr virus small RNA EBER-1. *Eur. J. Biochem.*, **193**, 635–641.
- Cowtan,K.D. (1994) DM: an automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newslett. Protein Crystallogr.*, **31**, 34–38.
- Davies,J.F., Hostomska,Z., Hostomsky,Z., Jordon,S.R. and Matthews,D.A. (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science*, **252**, 88–95.
- Dunn,J.J. and Studier,F.W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.*, **166**, 477–535.
- Eckmann,C.R. and Jantsch,M. (1997) Xlrba, a double-stranded RNA-binding protein associated with ribosomes and heterogeneous nuclear RNAs. *J. Cell Biol.*, **138**, 239–253.
- Fisher,A.J. and Johnson,J.E. (1993) Ordered duplex RNA controls capsid architecture in an icosahedral aminal virus. *Nature*, **361**, 176–179.
- Galabru,J., Katze,M.G., Robert,N. and Hovanessian,A.G. (1989) The binding of double-stranded RNA and adenovirus VAI RNA to the interferon-induced protein kinase. *Eur. J. Biochem.*, **178**, 581–589.
- Gatignol,A., Buckler-White,A., Berkhout,B. and Jeang,K.-T. (1991) Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science*, **251**, 1597–1600.
- Gibson,T.J. and Thompson,J.D. (1994) Detection of dsRNA-binding domains in RNA helicase A and *Drosophila* maleless: implications for monomeric RNA helicases. *Nucleic Acids Res.*, **22**, 2552–2556.

- Green,S.R. and Mathews,M. (1992) Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Genes Dev.*, **6**, 2478–2490.
- Jocobo-Molina,A. et al. (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc. Natl Acad. Sci. USA*, **90**, 6320–6324.
- Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kharrat,A., Macias,M.J., Gibson,T.J., Nilges,M. and Pastore,A. (1995) Structure of the dsRNA binding domain of *E.coli* RNase III. *EMBO J.*, **14**, 3572–3584.
- Kim,U., Wang,Y., Sanford,T., Zeng,Y. and Nishikura,K. (1994) Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc. Natl Acad. Sci. USA*, **91**, 11457–11461.
- Krinke,L. and Wulff,D.L. (1990) The cleavage specificity of RNase III. *Nucleic Acids Res.*, **18**, 4809–4815.
- Krovat,B.C. and Jantsch,M.F. (1996) Comparative mutational analysis of the double-stranded RNA binding domains of *Xenopus laevis* RNA-binding protein A. *J. Biol. Chem.*, **271**, 28112–28119.
- Langley,K.E., Berg,T.F., Strickland,T.W., Fenton,D.M., Boone,T.C. and Wypych,J. (1987) Recombinant-DNA-derived bovine growth hormone from *Escherichia coli*. 1. Demonstration that the hormone is expressed in reduced form and isolation of the hormone in oxidized, native form. *Eur. J. Biochem.*, **163**, 313–321.
- Larson,S.B., Koszelak,S., Day,J., Greenwood,A., Dodds,J.A. and McPherson,A. (1993) Double-helical RNA in satellite tobacco mosaic virus. *Nature*, **361**, 179–182.
- Larson,S.B., Day,J., Greenwood,A. and McPherson,A. (1998) Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J. Mol. Biol.*, **277**, 37–50.
- Luger,K., Mader,A.W., Richmond,R.K., Sargent,D.F. and Richmond,T.J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
- Manche,L., Green,S.R., Schmedt,C. and Mathews,M.B. (1992) Interactions between double-stranded RNA regulators and the protein kinase DAI. *Mol. Cell. Biol.*, **12**, 5238–5248.
- March,P.E., Ahn,J. and Inouye,M. (1985) The DNA sequence of the gene (*rnc*) encoding ribonuclease III of *Escherichia coli*. *Nucleic Acids Res.*, **13**, 4677–4685.
- Mellitis,K.H., Kostura,M. and Mathews,M.B. (1990) Interaction of adenovirus BA RNA I with the protein kinase DAI: nonequivalence of binding and function. *Cell*, **61**, 843–852.
- Meurs,E., Chong,K., Galabru,J., Thomas,N.S.B., Kerr,I.M., Williams,B.R.G. and Hovanessian,A.G. (1990) Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell*, **62**, 379–390.
- Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.*, **15**, 8783–8798.
- O'Connell,M.A., Krause,S., Higuchi,M., Hsuan,J.J., Toty,N.F., Jenny,A. and Keller,W. (1995) Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol. Cell. Biol.*, **15**, 1389–1397.
- Otwinowski,Z. (1991) Maximum likelihood refinement of heavy atom parameters. In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Daresbury Laboratory, Daresbury, UK, pp. 80–86.
- Otwinowski,Z. and Minor,W. (1996) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
- Polson,A.G. and Bass,B.L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.*, **13**, 5701–5711.
- Ramakrishnan,V. and White,S.W. (1992) The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature*, **358**, 768–771.
- Read,R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr.*, **A42**, 140–149.
- Rice,P. and Mizuuchi,K. (1995) Structure of the bacteriophage Mu transposase core: a common structural motif for DNA transposition and retroviral integration. *Cell*, **82**, 209–220.
- Rould,M.A., Perona,J.I., Söll,D. and Steitz,T.A. (1989) Structure of *E.coli* glutamyl-tRNA synthetase complexed with tRNA-Gln and ATP at 2.8 Å resolution. *Science*, **246**, 1135–1142.
- Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitshler,A.,
- Podjarny,A., Rees,B., Thierry,J.C. and Moras,D. (1991) Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA-ASP. *Science*, **252**, 1682–1689.
- Schweiguth,D.C., Chelladurai,B.S., Nicholson,A.W. and Moore,P.B. (1994) Structural characterization of a ribonuclease III processing signal. *Nucleic Acids Res.*, **22**, 604–612.
- Sharp,T.V., Schwemmle,M., Jeffrey,I., Liang,K., Mellor,H., Proud,C.G., Hilse,K. and Clemens,M.J. (1993) Comparative analysis of the regulation of the interferon-inducible protein kinase PKR by Epstein-Barr virus RNAs EBER-1 and EBER-2 and adenovirus VAI RNA. *Nucleic Acids Res.*, **21**, 4483–4490.
- St Johnston,D., Beuchle,D. and Nusslein-Volhard,C. (1991) *Staufen*, a gene required to localize maternal RNAs in the *Drosophila* egg. *Cell*, **66**, 51–63.
- St Johnston,D., Brown,N.H., Gall,J.G. and Jantsch,M. (1992) A conserved double stranded RNA-binding domain. *Proc. Natl Acad. Sci. USA*, **89**, 10979–10983.
- Thomis,D.C., Doohan,J.P. and Samuel,C.E. (1992) Mechanism of interferon action: cDNA structure, expression and regulation of the interferon-induced, RNA-dependent P1/eIF-2α protein kinase from human cells. *Virology*, **188**, 33–46.
- Wu,M., SantaLucia,J., Jr and Turner,D.H. (1997) Solution structure of (rGGCAGGCC)<sub>2</sub> by two dimensional NMR and iterative relaxation matrix approach. *Biochemistry*, **36**, 4449–4460.
- Yang,W. and Steitz,T.A. (1995) Recombining the structures of HIV integrase, RuvC and RNase H. *Structure*, **3**, 131–134.
- Yang,W., Hendrickson,W.A., Crouch,R.J. and Satow,Y. (1990) Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science*, **249**, 1398–1405.

Received September 11, 1998; revised and accepted October 16, 1998