The EMBO Journal vol.14 no.14 pp.3563–3571, 1995

NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5

Mark Bycroft, Stefan Grünert[1], Alexey G.Murzin[2], Mark Proctor and Daniel St Johnston[1,3]

Cambridge Centre for Protein Engineering, Department of Chemistry, University of Cambridge, Lensfield Rd, Cambridge CB2 1EW, [1]Wellcome/CRC Institute and Department of Genetics, University of Cambridge, Tennis Court Rd, Cambridge CB2 1QR and [2]Cambridge Centre for Protein Engineering, Hills Rd, Cambridge CB2 2QH, UK

[3]Corresponding author

The double-stranded RNA binding domain (dsRBD) is an ~65 amino acid motif that is found in a variety of proteins that interact with double-stranded (ds) RNA, such as *Escherichia coli* RNase III and the dsRNA-dependent kinase, PKR. *Drosophila* staufen protein contains five copies of this motif, and the third of these binds dsRNA *in vitro*. Using multinuclear/multi-dimensional NMR methods, we have determined that staufen dsRBD3 forms a compact protein domain with an α-β-β-β-α structure in which the two α-helices lie on one face of a three-stranded anti-parallel β-sheet. This structure is very similar to that of the N-terminal domain of a prokaryotic ribosomal protein S5. Furthermore, the consensus derived from all known S5p family sequences shares several conserved residues with the dsRBD consensus sequence, indicating that the two domains share a common evolutionary origin. Using *in vitro* mutagenesis, we have identified several surface residues which are important for the RNA binding of the dsRBD, and these all lie on the same side of the domain. Two residues that are essential for RNA binding, F32 and K50, are also conserved in the S5 protein family, suggesting that the two domains interact with RNA in a similar way.

Key words: dsRNA binding/NMR/ribosomal protein/staufen/structure

Introduction

The discovery of conserved amino acid motifs in proteins with similar properties has led to the characterization of a number of functional protein domains. Several of such recurring motifs have been identified in proteins that interact with RNA (Mattaj, 1993; Burd and Dreyfuss, 1994). Three of these are >50 amino acids long and are therefore most likely to constitute independent structural motifs: the RNP domain, the hnRNP K homology domain (KH) and the double-stranded RNA binding domain (dsRBD). Different proteins contain various numbers and combinations of these motifs, stressing their modular character, and all three are found in a wide spectrum of organisms, suggesting that they have an ancient origin. The best characterized of these is the RNP motif, which has now been found in >150 polypeptides with a variety of different functions (Bandziulis *et al.*, 1989; Birney *et al.*, 1993). X-ray crystallography and NMR analysis have revealed that this motif encodes a protein domain with a β-α-β-β-α-β structure in which many of the conserved residues of the RNP consensus sequence contribute to the hydrophobic core (Nagai *et al.*, 1990; Hoffman *et al.*, 1991). The RNA binding site lies on one face of a four-stranded anti-parallel β-sheet, and several of the residues within this region of U1A protein make direct contacts to bases in its RNA target (Oubridge *et al.*, 1994). These direct contacts explain the sequence specificity of RNA binding that has been observed for many proteins in this family. NMR analysis indicates that the KH motif also corresponds to an independent protein domain (Gibson *et al.*, 1993; Siomi *et al.*, 1993; Castiglione Morelli *et al.*, 1995). Although the nature of its interaction with RNA remains to be elucidated, point mutations in the KH domains of hnRNP K and FMR-1, the product of the fragile-X mental retardation gene, impair the ability of these proteins to bind homopolymeric RNAs *in vitro*, suggesting that the domain binds RNA directly (Siomi *et al.*, 1994).

In contrast to the RNP and KH motifs, the ~65 amino acid long dsRBD motif has been found in a number of proteins that specifically recognize double-stranded (ds) RNAs (St Johnston *et al.*, 1992; Gibson and Thompson, 1994). These include *Escherichia coli* RNase III, a dsRNA-specific nuclease (Robertson *et al.*, 1968; March *et al.*, 1985); the dsRNA-dependent protein kinase, PKR (Meurs *et al.*, 1990; Green and Mathews, 1992); and the dsRNA-specific adenosine deaminase, DRADA (Bass and Weintraub, 1988; Kim *et al.*, 1994; O'Connell *et al.*, 1995). Furthermore, the motif seems to correspond to an independent structural domain that binds to dsRNA, since peptides containing just a single copy of the motif from either the *Xenopus* protein, Xlrbpa, or *Drosophila* staufen protein bind to dsRNA *in vitro*, but not to ssRNA (St Johnston *et al.*, 1992). The interaction of the dsRBD is unlikely to involve the recognition of specific sequences, as this *in vitro* binding is sequence independent, and three members of this protein family with well-defined substrates, RNase III, PKR and DRADA, all bind to any dsRNAs of sufficient length (Manche *et al.*, 1992; Polson and Bass, 1994; Schweisguth *et al.*, 1994). Nevertheless, multiple dsRBDs may be able to act in combination to recognize the structure of specific RNAs, as staufen protein, which contains five copies of the motif, associates specifically with *oskar* and *bicoid* mRNAs, and is required for their localization to opposite poles of the *Drosophila* egg (St Johnston *et al.*, 1989, 1991). Indeed, in the *Drosophila* embryo, full-length staufen protein associates specifically with a 400 nt region of the *bicoid* 3′ UTR

which forms a complex secondary structure, but does not bind to other dsRNAs (Ferrandon *et al.*, 1994).

The third dsRBD in staufen protein conforms well with the emerging consensus sequence for dsRBDs, and a peptide containing only this region of the protein binds dsRNA *in vitro*, suggesting that this copy of the motif folds into a typical dsRNA binding domain (St Johnston *et al.*, 1992). In order to gain an insight into how the dsRBD might interact with dsRNA, we have used multinuclear/multidimensional NMR methods to determine the three-dimensional structure of the third staufen domain. This structure shows a strong similarity to the N-terminal region of *Bacillus stearothermophilus* ribosomal protein S5. The two domains also share a number of conserved residues, indicating that they have a common evolutionary origin. In addition, we have used *in vitro* mutagenesis to examine which of the conserved residues on the surface of staufen domain 3 are required for dsRNA binding. These results lead us to propose a model for how the domain might contact dsRNA.

Results

Determination of the tertiary structure of staufen dsRBD3

We have previously assigned the ^{1}H,^{15}N and ^{13}C resonances of the staufen dsRBD3 and have determined that the domain has an α-β-β-β-α secondary structure (Bycroft *et al.*, 1995). To aid in the determination of the tertiary structure of the domain, stereospecific assignments for valine γ-protons and leucine δ-protons were obtained from the splitting patterns in a $^{1}H/^{13}C$ correlation spectrum of a 10% fractionally ^{13}C-labelled sample (Neri *et al.*, 1989). Nuclear Overhauser effects (NOEs) were identified from $^{1}H/^{1}H$ two-dimensional NOESY, $^{1}H/^{15}N$ three-dimensional HMQC NOESY and $^{1}H/^{13}C$ three-dimensional HMQC NOESY spectra (Figure 1) and were classified as strong, medium or weak, based on peak intensities. Long range NOEs were only observed for the region between residues P1 and L65, which correspond to P579 and L643 in the full-length protein. This agrees well with the limits of the domain based on sequence homology. A total of 621 NOE constraints were used in the calculations, including 153 sequential, 90 medium range [i to ≤(i+4)], and 178 long range NOEs [i to ≥(i+5)]. Backbone dihedral constraints for 28 residues were used based on $J_{\alpha N}$ coupling constants. Stereospecific assignments of β-protons and χ^1 torsion angle constraints for 21 residues were obtained from intra-residue NOEs and $J_{\alpha\beta}$ and $J_{N\beta}$ coupling constant information obtained from 2-D DQF COSY and 3-D HNHB spectra. Restraints corresponding to 30 hydrogen bonds identified from the presence of slowly exchanging amide protons and from characteristic secondary structure NOE patterns were included in the structure calculations.

Figure 2 shows 20 structures that were calculated with no NOE violations greater than 0.3 Å or angle violations greater than 5° (Table I). The r.m.s. deviation about the average structure for the domain is 1.02 Å for backbone atoms and 1.60 Å for all non-hydrogen atoms. No long range NOEs were observed from residues G25 to K30 and this region is disordered in the structures (Figure 3). When this region is excluded from the comparison, the r.m.s. deviation is 0.59 Å for backbone atoms and 1.29 Å for all non-hydrogen atoms.

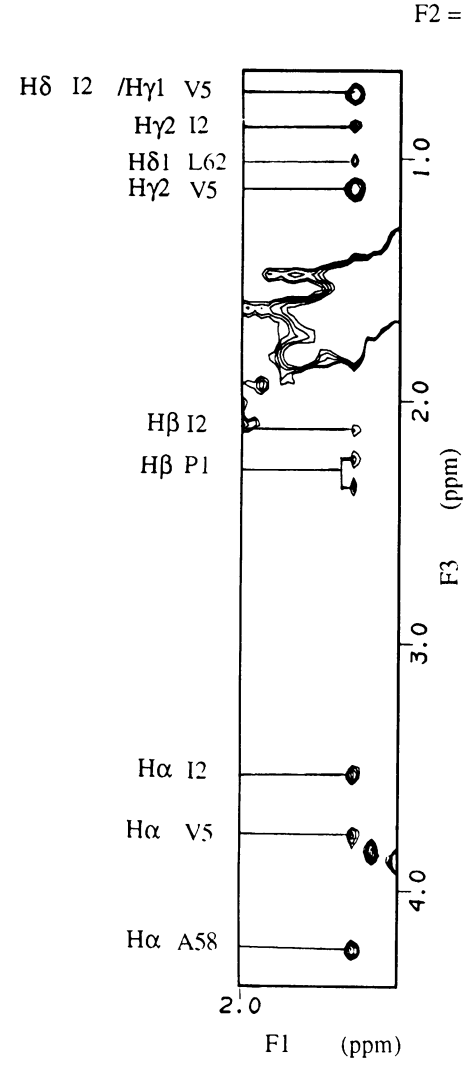

Fig. 1. F1/F3 plane from the 3-D HMQC NOESY spectrum of $^{15}N/^{13}C$-labelled staufen domain 3 showing NOEs from the β-protons of Ala58. The numbering in this figure refers to the positions of the residues within the domain, which extends from amino acids 579 (P1) to 643 (L65) in the full-length protein.

The structural analysis described above reveals that the α-β-β-β-α secondary structure of staufen domain 3 folds to form a tertiary structure in which the two α-helices lie on one side of a three-stranded anti-parallel β-sheet, formed by residues H17–E23, N31–V38 and I41–N48. The first β-strand contains a bulge. The first and second strands of the β-sheet are connected by a poorly defined loop, and the second and third strands are connected by a hairpin turn. The final strand of the sheet is followed by a turn into a C-terminal helix that runs from approximately residues K54 to K60. Residues K50–S53 have a high r.m.s. deviation, so there is some uncertainty as to the exact start of the helix. The C-terminal helix packs diagonally across the sheet, and the N-terminal helix packs against the edge of the sheet. Both of the helices in the dsRBD are amphipathic and residues from their hydrophobic faces pack onto hydrophobic residues on one side of the β-sheet to form a hydrophobic core.

Many of the hydrophobic residues that are conserved

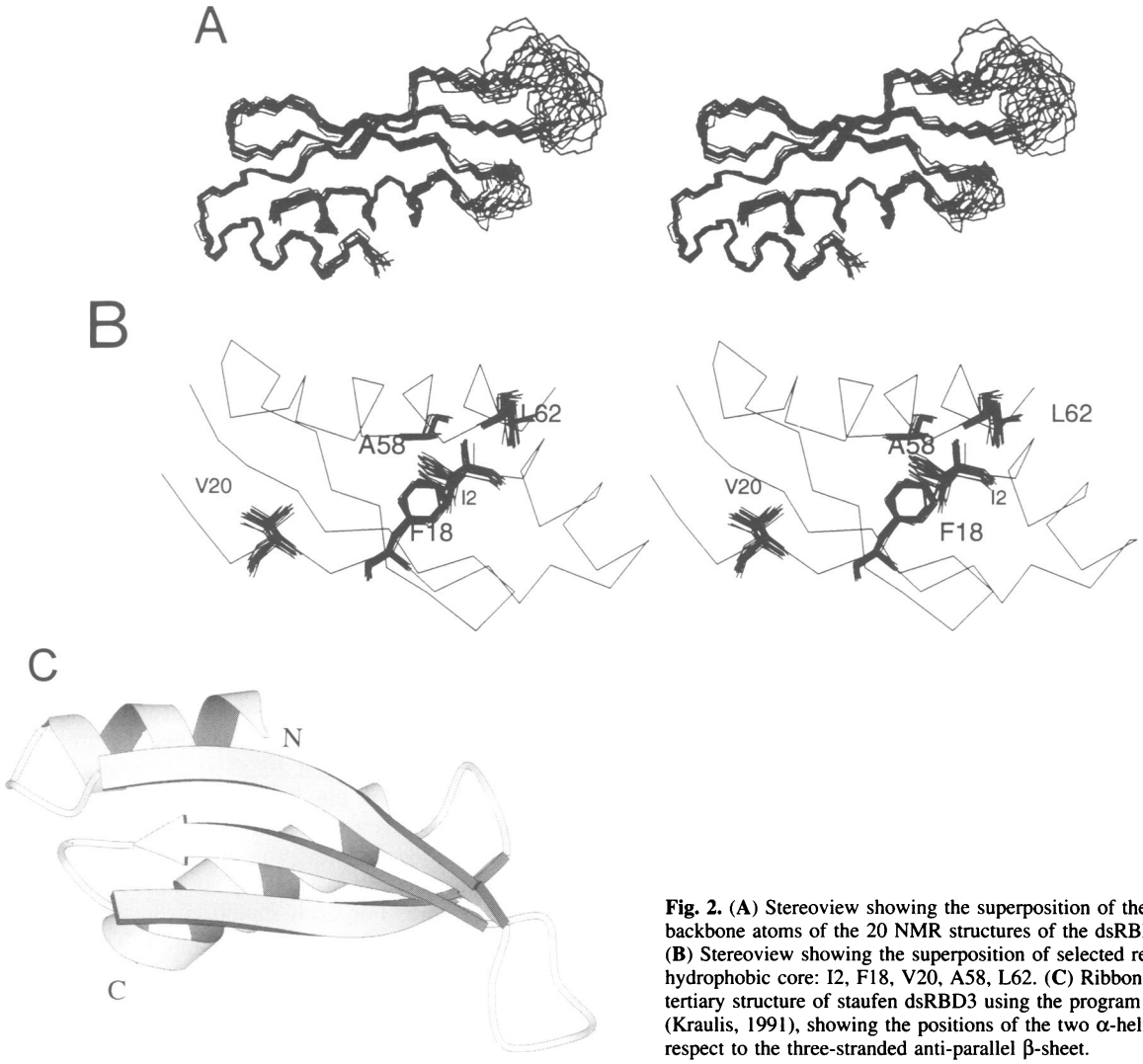

Fig. 2. (**A**) Stereoview showing the superposition of the Cα, C′ and N backbone atoms of the 20 NMR structures of the dsRBD. (**B**) Stereoview showing the superposition of selected residues in the hydrophobic core: I2, F18, V20, A58, L62. (**C**) Ribbon diagram of the tertiary structure of staufen dsRBD3 using the program MOLSCRIPT (Kraulis, 1991), showing the positions of the two α-helices with respect to the three-stranded anti-parallel β-sheet.

amongst dsRBDs correspond to amino acids in the hydrophobic core of staufen dsRBD3 (I2, V5, I8, F18, V20, L21, C36, V38, V42, M61, L62 and L65), indicating that their conservation reflects a role in the formation of the structure of this domain. In particular, F18, which is at the centre of the core, is a phenylalanine or tyrosine in all known domains, while bulky hydrophobic residues are usually found in the equivalent positions to V20 and L21, which form the bulge in the first strand of the β-sheet. The sequence G X G 5X A/S 3X AA, which spans the end of the third strand of the β-sheet and the start of the second helix, is also strongly conserved in other dsRBDs. From the structure of the domain in this region, it appears that these small residues are conserved to accommodate the close proximity of the C-terminal helix to the last strand of the β-sheet. The two glycine residues in the sheet pack onto the serine residue at the start of the helix, and the first of the two conserved alanine residues one helix turn away. The second conserved alanine packs onto the first helix. The second very highly conserved aromatic residue in the consensus sequence, F32, lies at the edge of the hydrophobic core, and its high conservation may reflect a role in the function of the domain rather than in its structure. Other conserved residues which lie on the surface of the domain are also likely to be important for function. The sequence KKxxK at the N-terminus of the second helix (50–54) is found in most members of this family, suggesting that the positively charged lysines may have a direct role in RNA binding. Another well conserved sequence is the GPxH (residues 25–28), which lies in the flexible loop connecting the first and second strands of the β-sheet.

Structural similarities to staufen dsRBD3

Comparison of the fold of the dsRBD with that of other known proteins revealed a strong resemblance to two other structures; the C-terminal domain of porphobilinogen deaminase (PDA) (Louie *et al.*, 1992), and the N-terminal domain of the ribosomal protein S5 (Ramakrishnan and White, 1992). In porphobilinogen deaminase, the C-terminal domain contains an essential cysteine residue covalently linked to a dipyrromethane cofactor; the rest of the structure shares a common fold with a periplasmic receptor family (Louie *et al.*, 1992). The PDA domain shares the same α-β-β-β-α topology with the staufen domain and can be superimposed onto it with an r.m.s. deviation of 2.24 Å for 43 Cα atoms in the common core. However, none of the five core residues that occupy identical positions in both domains is conserved between

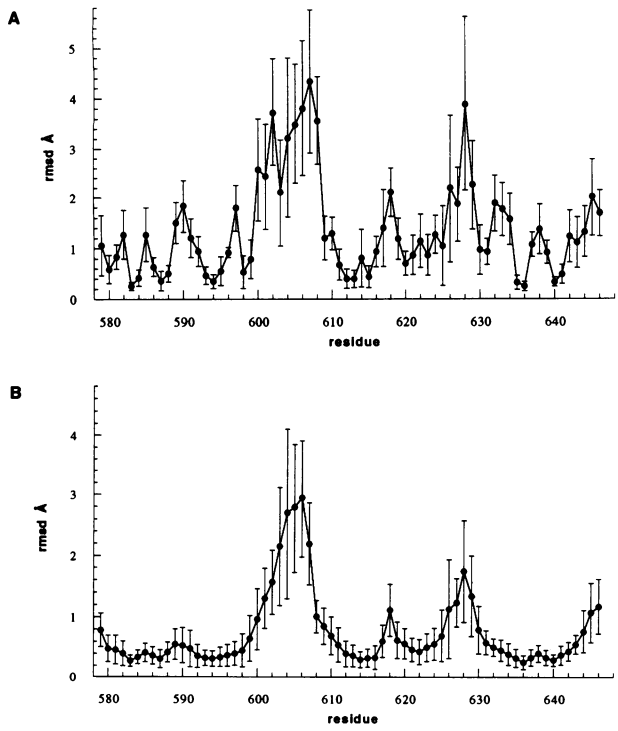

Fig. 3. Atomic r.m.s. distribution about the average structure as a function of residue number (within full-length staufen protein) for (**A**) the backbone atoms and (**B**) all non-hydrogen atoms.

the two protein families, and it seems most likely that this similarity is fortuitous.

In contrast to PDA, the structural similarity between the dsRBD and ribosomal protein S5 is likely to be significant. The crystal structure of *B.stearothermophilus* S5 has been solved, and consists of two distinct domains (Ramakrishnan and White, 1992), The C-terminal domain shares a common fold with elongation factor G and DNA gyrase B (Murzin, 1995). The N-terminal domain of S5 shares a common core of 46 residues with the staufen dsRBD, resulting in an r.m.s. deviation for the Cα atoms of 2.26 Å. The prokaryotic S5 proteins lack the N-terminal helix found in the dsRBD, but the eukaryotic and archaebacterial counterparts of S5 have an N-terminal extension that may correspond to this helix (All-Robyn *et al.*, 1990). The rest of the S5 domain has an identical topology to the dsRBD (Figure 4B and C). In both proteins, a C-terminal helix packs across a three-stranded anti-parallel β-sheet, in close proximity to the last strand. Furthermore, there is a β-bulge at the same position in the first strand of the β-sheet of each domain. The strands are also connected in a similar manner. The first two strands are linked by an extended loop, and the last two are connected by a hairpin turn.

When the consensus sequence of the dsRBD is aligned according to its structure with the consensus sequence of prokaryotic S5 proteins, and the consensus derived from archaebacterial and eukaryotic members of the S5 family, similar or identical conserved residues are found at 14 out of 50 positions in all three sequences (Figure 4A). These include the two residues that form the β-bulge, the glycines and alanines that allow the close approach of the C-terminal helix to the β-sheet, and several other residues in the hydrophobic core. In addition, the archaebacterial and eukaryotic ribosomal proteins have a proline in the equivalent position to the first proline of the dsRBD motif, and the turn that this introduces just before the first helix may allow the domain to fold without colliding with the N-terminal region of the protein. The extent of the sequence similarities between these two domains strongly suggests that the similarity between their structures reflects a common evolutionary origin.

Mutations affecting the RNA binding of the dsRBD

To investigate which amino acids in the dsRBD participate in RNA binding, we have used a PCR-based *in vitro* mutagenesis procedure to generate a number of point mutations in staufen domain 3 (underlined in Figure 4A). Since the major groove of dsRNA is too narrow to allow direct contacts between amino acid side chains and the bases (Alden and Kim, 1979), the best candidates for residues involved in RNA binding are positively charged amino acids that can interact with the phosphate backbone. We have therefore generated mutants in five such residues that lie on the surface of the domain: H28 and K30, which lie in the flexible loop between the first and second strands of the β-sheet; K50 and K51, which fall in the loop between the β-sheet and the C-terminal helix; and K54, which forms part of this C-terminal helix (Figure 5C). In addition, we have mutated F32, which protrudes from the second strand of the β-sheet, since the very high conservation of this residue in both dsRBD-containing proteins and ribosomal S5 proteins cannot be explained easily by a structural role, as it is not buried in the hydrophobic core. For each mutation, the single amino acid residue was substituted by alanine. The resulting modified domains were then expressed as glutathione-S-transferase (GST) fusion proteins, and tested for their ability to bind dsRNA when renatured on North-Western blots (Figure 5A and B). The results of these experiments show that the mutations at either F32 or K50 completely abolish RNA binding, the K51 and K54 mutations cause a partial loss of activity, while the H28 and K30 substitutions bind as well as the wild-type domain in this assay.

Although it is possible that the mutants could disrupt RNA binding indirectly, by perturbing the structure of the domain, this seems unlikely for two reasons. First, all of these mutations affect residues that lie on the surface of the domain or, in the case of F32, at least partially on the surface and, since alanine is a fairly neutral substitution, it should not interfere with the folding of the domain. Secondly, domains with mutations in residues within the hydrophobic core are insoluble when expressed in *E.coli*, presumably because they are misfolded, whereas these mutant peptides and the wild-type domain are all completely soluble, suggesting that they have folded properly (S.G. and D.St J., unpublished results). Thus, this mutational analysis strongly suggests that dsRNA binds in the cleft between the N-terminus of the second helix and one face of the β-sheet, with F32 and K50 making essential contacts with the RNA (Figure 5C).

Discussion

The three-dimensional structure of the single dsRBD in *E.coli* RNase III has been solved by Kharrat *et al.* (1995)

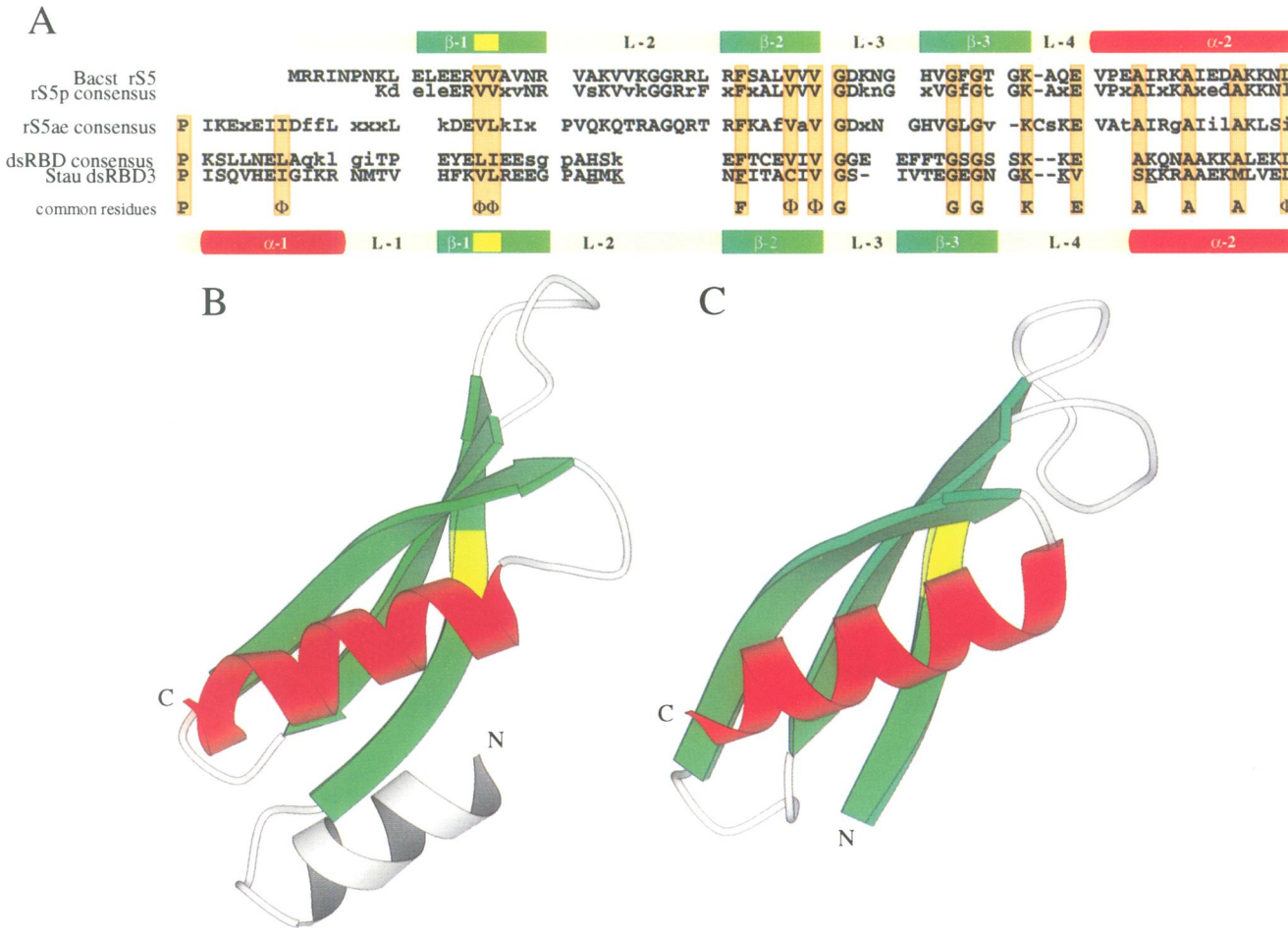

Fig. 4. Comparison between the dsRBD and the N-terminal domain of ribosomal protein S5. (**A**) The primary and secondary structures of the N-terminal domain of the *Bacillus stearothermophilus* rS5 (Bacst rS5) and of the dsRBD3 of staufen (Stau dsRBD3). Above and below the amino acid sequences, we indicate the secondary structure of the rS5 and Stau dsRBD3 domain, respectively. Also shown are the consensus sequences for prokaryotic ribosomal S5 proteins (rS5p consensus), archaebacterial and eukaryotic members of the S5 family (rS5ae consensus) and the dsRBD consensus (after Gibson and Thompson, 1994). A capital letter represents an amino acid that occurs at the same position in >70% of the sequences, and a lower case letter an amino acid present in the same position in >50% of the sequences: an x is used for all other positions. Residues shared by all three consensus sequences are highlighted with coloured blocks, and shown in the lowest line. ϕ indicates large hydrophobic residues. Underlined amino acids in the Stau dsRBD3 sequence correspond to the residues changed to alanines in the mutation analysis. (**B**) Ribbon diagram of the tertiary structure of staufen dsRBD3 using the program MOLSCRIPT (Kraulis, 1991). (**C**) The tertiary structure of rS5, as determined by Ramakrishnan and White (1992). β-Sheets are indicated by green arrows, and the position of the β-bulge in the N-terminal β-sheet is shown in yellow. α-Helices are indicated by red spirals (the N-terminal helix of dsRBD3 is shown in grey, as it has no counterpart in the bacterial protein).

and has a very similar topology to that of the third staufen domain described here. Since many of the conserved residues that define the dsRBD motif correspond to residues in the hydrophobic cores of these two structures, it seems probable that the other members of this family will also fold in the same way, and that the motif represents a well-defined protein domain. Although the individual dsRBDs share many conserved residues, the overall length of the domain is somewhat variable. For example, the second staufen domain contains a 91 amino acid insertion between the residues that correspond to positions 25 and 26 in staufen domain 3 (Gibson and Thompson, 1994). Based on the structure of domain 3, this insertion occurs in the extended loop between the first two strands of the β-sheet. Similarly, the second domains of RNA helicase A and *Drosophila* male-less protein contain five extra amino acids between positions 40 and 41, which are predicted to fall in the loop that connects the second and third strands of the sheet. Thus, the length variations between family members fall in the less ordered regions of the structure and are unlikely to affect the overall folding of the domain.

The structure of the dsRBD does not resemble that of any of the three other RNA binding domains for which structural information is available. The RNP domain has an β-α-β-β-α-β arrangement of secondary structure, and the overall topology is quite different from the dsRBD (Nagai *et al.*, 1990; Hoffman *et al.*, 1991; Oubridge *et al.*, 1994). The KH domain also has an αβ fold, but with a β-α-α-β-β arrangement (Castiglione Morelli *et al.*, 1995). The third RNA binding domain is the cold shock domain found in the *Xenopus* Y box proteins, which bind both RNA and DNA (Ladomery and Sommerville, 1994; Murray, 1994; Wolffe, 1994). The three-dimensional structures of two bacterial cold shock proteins have been solved recently, and these domains have a β-barrel structure which is again very different to the fold of the dsRBD (Schindelin *et al.*, 1993, 1994; Schnuchel *et al.*, 1993;

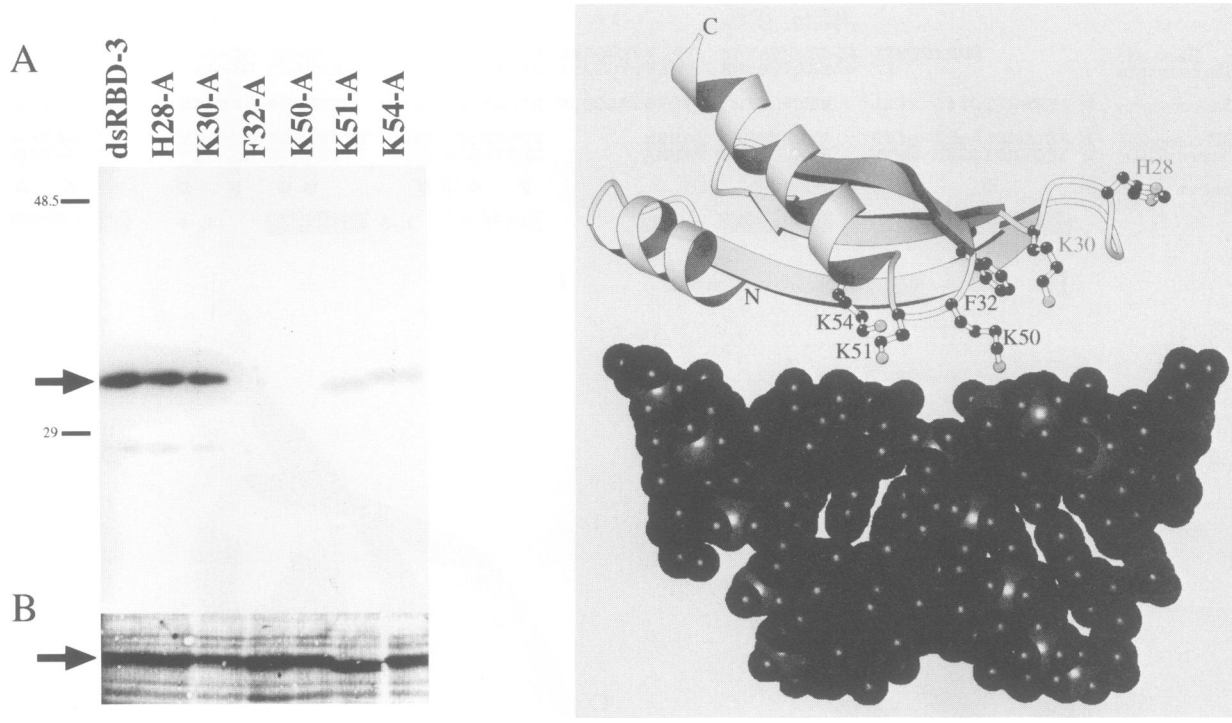

Fig. 5. The effect of mutations in the dsRBD3 of staufen on RNA binding. Fusion proteins containing mutations in the dsRBD3 were overexpressed in *E.coli* and assayed on North-Western blots. (**A**) An autoradiogram of a typical North-Western experiment probed with ^{32}P-labelled dsRNA (VAI RNA) The positions of two marker proteins (29 and 48.5 kDa) are shown on the left, and the overexpressed fusion peptide is indicated by an arrowhead. (**B**) The corresponding CPTS-stained blot shows an approximately even loading of fusion peptides. The K51-A peptide shows a slightly increased mobility in SDS–PAGE compared with the other constructs, even though sequence analysis shows the construct does not carry any other mutations. (**C**) The side chains of the six mutated amino acids were modelled onto the three-dimensional representation of the dsRBD3. Light grey indicates amino acids in which mutations have a weak or no effect on RNA binding, and black indicates those that have a strong effect. The positions of the known residues affecting RNA binding suggest that RNA interacts with one face of the domain. The dsRNA shown in the lower half of the figure is a space filling representation of a 14 bp A-type helix (Dock-Bregeon *et al.*, 1989).

Newkirk *et al.*, 1994). The dsRBD does, however, have a very similar topology to the N-terminal domain of ribosomal protein S5 (Ramakrishnan and White, 1992). Although none of the S5 family of ribosomal proteins were detected in the database searches with the dsRBD motif, the two domains also share a number of conserved residues, indicating the importance of structure determination in revealing relationships that are not readily apparent at the primary sequence level. The association of S5 with the 30S ribosomal subunit depends on other ribosomal proteins, but it seems likely that the protein binds to RNA directly, since it protects a number of bases in 16S rRNA from chemical modification, and can be UV-cross-linked to several regions of 16S rRNA, and to mRNA in the elongating ribosome (Stern *et al.*, 1988, 1989; Brimacombe, 1991; Rinke-Appel *et al.*, 1991). Thus, the structural and sequence similarities between these two domains are likely to be reflected in similarities in their functions.

Structure and sequence similarities between other RNA binding domains and ribosomal proteins have been noted. The RNP domain has some topological similarity to a number of ribosomal proteins, in particular to the S6 protein, and the cold shock domain has a similar fold to that of S17 (Hoffman *et al.*, 1991; Golden *et al.*, 1993; Lindahl *et al.*, 1994). In both these cases, however, the structural resemblance is not reflected in an obvious homology at the sequence level. The similarity between the dsRBD and the N-terminal domain of the S5 family of ribosomal proteins is more extensive, since the two domains not only share the same topology, but also a number of conserved residues in their hydrophobic cores. This indicates that the dsRBD has an ancient origin, and supports the idea that it may have evolved from a primitive ribosomal protein. This may also turn out to be the case

Table I. Structural statistics for the 20 dsRBD structures

Structural statistics	
R.m.s. deviations from experimental restraints	
Distance (Å)	0.0382 ± 0.0012
Dihedral angle (deg)	0.67 ± 0.19
XPLOR energies (kcal/mol)	
F_{vdW}[a]	45.46 ± 6.61
F_{L-J}[b]	−130.5 ± 31.5
Deviations from ideal covalent geometry	
Bonds (Å)	0.0034 ± 0.00016
Angles (deg)	0.38 ± 0.04
Impropers (deg)	0.35 ± 0.01

All variances are quoted ± one standard deviation.
[a]The quadratic van der Waals' term was calculated with a force constant of 4 kcal/mol[4] with the van der Waals' radii set to 0.8 times the standard value used in the CHARM.
[b]F_{L-J} was calculated using the CHARM empirical energy function. Note that this Lennard-Jones term is not used as part of the target function in any part of the structure calculations.

for the KH domain, which has sequence homology with the ribosomal protein S3 (Gibson *et al.*, 1993; Siomi *et al.*, 1993).

The *in vitro* mutagenesis of staufen dsRBD3 has led to the identification of four amino acids that are important for the RNA binding of the domain, F30, K50, K51 and K54. Similar mutagenesis studies have been carried out on the first dsRBD of PKR and, although the interpretation of these experiments is complicated by the fact that the mutants were introduced into peptides that contain two dsRBDs, the results are in very good agreement with those presented here (Green and Mathews, 1992; Green *et al.*, 1995; McMillan *et al.*, 1995). The two mutations that cause the strongest reduction in the RNA binding of PKR without altering structurally important residues correspond to F32 and K50, while the equivalent mutations to K51 and K54 also reduce binding, but less dramatically. The mutations which abolish the RNA binding of staufen domain 3 fall in two of the most highly conserved positions within the dsRBD consensus sequence, F32 and K50, and both of these amino acids are also conserved in the S5 proteins. K51 and K54, which play a less important role in RNA binding, are not as well conserved in the dsRBD family and, while K51 corresponds to a conserved residue in the archaebacterial and eukaryotic S5 consensus sequence but not the prokaryotic consensus, K54 is not present in either. Finally, the two mutations that have no noticeable effect on RNA binding are only weakly conserved amongst dsRBDs and have no obvious counterparts in the ribosomal proteins. Thus, there is a strong correlation between conservation of an amino acid and the effect of mutation of this residue on RNA binding. Futhermore, the conservation of the two amino acids that are essential for RNA binding between the dsRBD and the S5 proteins strongly suggests that the ribosomal proteins bind to RNA in the same way as the dsRBD.

In a number of other RNA binding proteins, conserved basic residues bind to RNA through electrostatic interactions with the phosphate backbone, and this is also likely to be the case for the conserved lysines in staufen dsRBD3 (Biou *et al.*, 1994; Oubridge *et al.*, 1994; Valegård *et al.*, 1994). Stacking interactions between aromatic residues and the bases of the RNA have also been shown to be important for the recognition of specific RNA sequences by ssRNA binding proteins (reviewed in Moras and Poterszman, 1995). This is unlikely to be the case for F32 of the dsRBD, since the major groove of the A form helix of dsRNA is too narrow to permit direct contacts between the amino acid side chains and the bases. However, the interaction of TATA binding protein (TBP) with its dsDNA target, which has an A-like structure, may provide a hint as to how F32 contacts dsRNA (Kim and Burley, 1994). Two conserved phenylalanines in TBP are inserted between the bases in the minor groove of the DNA, and form Van der Waal's and π–π interactions with the bases which disturb the base stacking, causing a kink in the DNA helix.

The positions of the four residues that are important for RNA binding are indicated in a three-dimensional representation of the domain in Figure 5C. The minimum segment of dsRNA that can bind to staufen domain 3 is 11 bp (S.Grünert and D.St Johnston, unpublished observations), and we have used a slightly longer (14 bp) dsRNA of known structure (Dock-Bregeon *et al.*, 1989) to model how the domain might interact with RNA. Of the three theoretical possibilities for positioning this length of RNA next to the domain, only the arrangement shown in Figure 5C brings all four residues important for RNA binding activity into close proximity to the RNA. Although the H28 and K30 mutations in the flexible second loop do not affect RNA binding, they lie on the same side of the domain as the RNA in this model, albeit further away, and it is possible that they also contact the RNA without contributing significantly to the binding affinity. Alternatively, the conservation of these residues may reflect a different role in the function of the domain. The equivalent loop within ribosomal protein S5 also seems to have a functional role, since a number of spectrinomycin resistance mutations map to this region of the protein (Wittman-Liebold and Greuer, 1978; Ramakrishnan and White, 1992). The fact that these mutations are not lethal, however, suggests that they are unlikely to disrupt the interaction of S5 with 16S rRNA. While this model is consistent with our data so far, it assumes that neither the RNA nor the protein change conformation on binding. A more complete understanding of the molecular basis for this interaction must therefore await the solution of the structure of the dsRBD–dsRNA complex.

Materials and methods

NMR sample preparation and NMR spectroscopy.

The purification of uniformly ^{15}N- and $^{15}N/^{13}C$-labelled dsRBD3 of staufen protein has been described previously (Bycroft *et al.*, 1995). NMR samples were 2 mM protein in 50 mM potassium phosphate pH 6.8, 3 mM deuterated dithiothreitol (DTT) (Isotec). Spectra were recorded at 298 K on a Bruker AMX 500 spectrometer. Data were processed using the Felix 2.10 software package. The ^1H carrier was placed on the water resonance which was suppressed by pre-saturation. 3-D HNHB and $^1H/^{15}N$ 3-D HMQC NOESY spectra were recorded on a ^{15}N-enriched sample with 64 real points in the nitrogen dimension, 128 or 256 real points in the indirect proton dimension and 512 complex points in the acquired dimension. The ^{15}N carrier was placed at 118 p.p.m. and a spectral width of 40 p.p.m. was used in the ^{15}N dimension. $^1H/^1H$ 2-D NOESY and $^1H/^1H$ DQF COSY spectra were recorded with 512 real points in f1, and 2000 complex points in f2 with a 12 p.p.m. spectral width in both dimensions. $^1H/^{13}C$ 3-D HMQC NOESY spectra were acquired with 256 real points in f1, 32 complex points in f2 and 512 complex points in f3. The carbon carrier frequency was placed at 39 p.p.m. A 20 p.p.m. spectral width was used in the carbon dimension resulting in folding of the spectrum. Quadrature detection in the indirect dimensions was achieved using the TPPI method, except for the t_2 dimension of the $^1H/^{13}C$ 3-D HMQC NOESY which was acquired using the TPPI/STATES method. $^1H/^{13}C$ correlation specra were recorded on a 10% fractionally ^{13}C-labelled sample as described in Neri *et al.* (1989)

Structures were calculated using the DG/SA protocol of the program XPLOR 3.1 (Brünger, 1988). The coordinates of the structure have been deposited in the Brookhaven database under 1stu.

Mutagenesis

Mutagenesis was performed using a modified two-step PCR mutagenesis protocol (Good and Nazar, 1994). Briefly, we used two primers flanking the dsRBD3 with *Bam*HI and *Eco*RI restriction enzyme sites, which enabled us to clone the PCR fragment into the polylinker of pGEX-2T (Promega), fusing the dsRBD3 in-frame with GST. For each mutation, we synthesized a third primer with the desired point mutation, which also carried silent mutations, to create a novel restriction enzyme site in the mutant PCR product. PCR was performed with *Pfu* (Stratagene) to reduce the risk of undesired random mutations. The primary PCR was performed in a 20 µl volume for 20 cycles with 100 nM of the mutant primer and 200 nM of the matching flanking primer under conditions as described in Picard *et al.* (1994). To this PCR, we added 5 µl containing 0.5 µl reaction buffer, 500 µM dNTPs and 200 nM final

concentration of the third primer. This sample was subjected to 15 more cycles, and the DNA was phenol/chloroform extracted, ethanol precipitated and digested to subclone the resulting fragment into appropriately digested pGEX-2T.

The following primers were used for the PCR: 5′ flanking, 5′-CGAGGGATCCATGGATGAGGGTGACAAGA; 3′ flanking, 5′-ATGGGAATTCCTTGGTGGGCGTAAGGGG; H28-A antisense, 5′-GTTCTTCATGGCCGCCGGGCCCTCCTCGC; K30-A antisense, 5′-TATAAAGTTGGCCATGTGCGCCGGGCCCTCCTCGC; F32-A antisense, 5′-CCACAATGCATGCTGTTATAGCGTTCTTCATG; K50-A sense, 5′-GGAAATGGCGCCAAAGTGTCC; K51-A antisense, 5′-ACGCTTCTTCGAAACTGCTTTGCCATTTCC; K54-A antisense, 5′-TTCCGCGGCCCGCTTGGCGGACACTTTTTT.

Mini-DNA preparations of single colonies were screened for the presence of the newly introduced restriction sites, and the presence of the desired mutation was verified by dideoxy sequencing.

RNA binding assay

Plasmid DNA of the appropriate constructs were transformed into *E.coli* DH5α. An overnight culture from a single colony was diluted 1:10 (v/v) and grown for a further 2 h at 37°C. Fusion protein production in 12.5 ml culture was induced by the addition of 0.2 mM IPTG and, after a further 2 h, the bacteria were harvested by centrifugation. The pellets were taken up in 500 μl SDS sample buffer and boiled for 1 min followed by sonication. The North-Western assay was performed as described (St Johnston *et al.*, 1992). Briefly, 1.5 μl of cell extract was separated by SDS–PAGE and blotted onto Immobilon P membranes. Samples were loaded in duplicate on the gels and one half of the membrane was stained for proteins in 0.05% copper phthalocyanine 3,4′,4″,4‴-tetrasulfonic acid tetrasodium salt (CPTS) in 12 mM HCl to check transfer and loading (Bickar and Reid, 1992). Proteins on the other half of the membrane were used for the RNA binding assay and renatured from 8 M urea in Tris-buffered saline (TBS) by 10 stepwise 2:3 v/v dilutions into TBS. After rinsing the blot in TBS, the membrane was blocked for 1 h in N/W buffer containing 5% dried milk. The block was diluted 1:1 in N/W buffer containing ^{32}P-labelled VAI *in vitro* transcript, an RNA that contains a number of double-stranded regions (Akusjärvi *et al.*, 1980). After 30 min, the membrane was washed 3×5 min in N/W buffer, dried and exposed to X-ray film.

Sequence alignments

The prokaryotic ribosomal S5 protein consensus sequence was derived from the S5 sequences from *E.coli*, *B.subtilis*, *B.stearothermophilus*, *Micrococcus mycoplasma* and an algal plastid sequence. For the eukaryotic and archaebacterial members of the S5 family, the consensus is derived from the ribosomal protein S4 sequences from mouse, human, *Dictyostelium* and *Saccharomyces cerevisiae*, the S2 sequences from rat and *Drosophila*, and the S5 sequences from two archaebacteria, *Methanococcus* and *Halobacterium*. In Figure 4A, the structural similarities between S5 and the dsRBD were used to align these consensus sequences with the dsRBD consensus. This initial alignment was then optimized locally. In the region of β-strands 2 and 3, the optimization resulted in a two amino acid shift of the dsRBD consensus with respect to the S5 sequences. Since this is a region of β-sheet, the shift does not change the orientation of the amino acid side chains.

Acknowledgements

M.B. and M.P. were supported by a ZENECA/MRC/DTI LINK program. D.St J. is the recipient of a Wellcome Trust Senior Fellowship.

References

Akusjärvi,G., Mathews,M.B., Andersson,P., Vennström,B. and Pettersson,U. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 2424–2428.

Alden,C.J. and Kim,S.-H. (1979) Solvent-accessible surfaces of nucleic acids. *J. Mol. Biol.*, **132**, 411–434.

All-Robyn,J.A., Brown,N., Otaka,E. and Liebman,S.W. (1990) Sequence and functional similarity between a yeast ribosomal protein and the *Escherichia coli* S5 *ram* protein. *Mol. Cell. Biol.*, **10**, 6544–6553.

Bandziulis,R.J., Swanson,M.S. and Dreyfuss,G. (1989) RNA-binding proteins as developmental regulators. *Genes Dev.*, **3**, 431–437.

Bass,B.L. and Weintraub,H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, **55**, 1089–1098.

Bickar,D. and Reid,P.D. (1992) A high-affinity protein stain for Western blots, tissue prints and electrophoretic gels. *Anal. Biochem.*, **203**, 109–115.

Biou,V., Yaremchuck,A., Tukalo,M. and Cusack,S. (1994) The 2.9 Å crystal structure of *T.thermophilus* seryl-tRNA synthetase complexed with tRNA^{Ser}. *Science*, **263**, 1404–1410.

Birney,E., Kumar,S. and Krainer,A.R. (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.*, **21**, 5803–5816.

Brimacombe,R. (1991) RNA–protein interactions in the *Escherichia coli* ribosome. *Biochimie*, **73**, 927–936.

Brünger,A.T. (1988) *Xplor Manual*. Yale University Press, New Haven, CT.

Burd,C.G. and Dreyfuss,G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Bycroft,M., Proctor,M., Freund,S. and St Johnston,D. (1995) Assignment of the backbone ^{1}H, ^{15}N, ^{13}C NMR resonances and secondary structure of a double-stranded RNA-binding domain from the *Drosophila* protein staufen. *FEBS Lett.*, in press.

Castiglione Morelli,M.A., Stier,G., Gibson,T., Joseph,C., Musco,G., Pastore,A. and Travè,G. (1995) The KH module has an αβ fold. *FEBS Lett.*, **358**, 193–198.

Dock-Bregeon,A.C., Chevrier,B., Podjarny,A., Johnson,J., de Bear,J.S., Gough,G.R., Gilham,P.T. and Moras,D. (1989) Crystallographic structure of an RNA helix: $[U(UA)_6 A]_2$. *J. Mol. Biol.*, **209**, 459–474.

Ferrandon,D., Elphick,L., Nüsslein-Volhard,C. and St Johnston,D. (1994) Staufen protein associates with the 3′ UTR of *bicoid* mRNA to form particles which move in a microtubule-dependent manner. *Cell*, **79**, 1221–1232.

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.

Gibson,T.J., Thompson,J.D. and Heringa,J. (1993) The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid. *FEBS Lett.*, **324**, 361–366.

Golden,B.L., Hoffmann,D.W., Ramakrishnan,V. and White,S.W. (1993) Ribosomal protein S17: characterization of the three-dimensional structure by 1H and 15N NMR. *Biochemistry*, **32**, 12812–12820.

Good,L. and Nazar,R.N. (1994) An improved thermal cycle for two-step PCR-based targeted mutagenesis. *Nucleic Acids Res.*, **20**, 4934.

Green,S.R. and Mathews,M.B. (1992) Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Genes Dev.*, **6**, 2478–2490.

Green,S.R., Manche,L. and Mathews,M.B. (1995) Two functionally distinct RNA-binding motifs in the regulatory domain of the protein kinase DAI. *Mol. Cell Biol.*, **15**, 358–364.

Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) RNA-binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.

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,J.L. and Burley,S.K. (1994) 1.9 Å resolution structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.*, **1**, 638–653.

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.

Kraulis,P.J. (1991) MOLSCRIPT—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Ladomery,M. and Sommerville,J. (1994) Binding of Y-box proteins to RNA—involvement of different protein domains. *Nucleic Acids Res.*, **22**, 5582–5589.

Lindahl,M. *et al.* (1994) Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.*, **13**, 1249–1254.

Louie,G.V., Brownlie,P.D., Lambert,R., Cooper,J.B., Blundell,T.L., Wood,S.P., Warren,M.J., Woodcock,S.C. and Jordan,P.M. (1992) Structure of porphobilinogen deaminase reveals a flexible multidomain polymerase with a single catalytic site. *Nature*, **359**, 33–39.

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., Ahnn,J. and Inouye,M. (1985) The DNA sequence of the

gene (*rnc*) encoding ribonuclease III of *Escherichia coli*. *Nucleic Acids Res.*, **13**, 4677–4685.
Mattaj,I.W. (1993) RNA recognition: a family matter. *Cell*, **73**, 837–840.
McMillan,N.A.J., Carpick,B.W., Hollis,B., Toone,W.M., Zamanian-Daryoush,M. and Williams,B.R.G. (1995) Mutational analysis of the double-stranded RNA (dsRNA) binding domain of the dsRNA-activated protein kinase, PKR. *J. Biol. Chem.*, **270**, 2601–2606.
Meurs,E., Chong,K., Galabru,J., Thomas,N.S., Kerr,I.M., Williams,B.R. and Hovanessian,A.G. (1990) Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon. *Cell*, **62**, 379–90.
Moras,D. and Poterszman,A. (1995) RNA–protein interactions: diverse modes of recognition. *Curr. Biol.*, **5**, 249–251.
Murray,M.T. (1994) Nucleic acid-binding properties of the *Xenopus* oocyte Y-box protein Mrnp(3+4). *Biochemistry*, **33**, 13910–13917.
Murzin,A.G. (1995) A ribosomal protein module in EF-G and DNA gyrase. *Nature Struct. Biol.*, **2**, 25–26.
Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
Neri,D., Szyperski,T., Otting,G., Senn,H. and Wüthrich,K. (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry*, **28**, 7510–7516.
Newkirk,K., Feng,W., Jiang,W., Tejero,R., Emerson,S.D., Inouye,M. and Montelione, G.T. (1994) Solution NMR structure of the major cold shock protein (CspA) from *Escherichia coli*: identification of a binding epitope for DNA. *Proc. Natl Acad. Sci. USA*, **91**, 5114–5118.
O'Connell,M.A., Krause,S., Higuchi,M., Hsuan,J.J., Totty,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.
Oubridge,C., Ito,N., Evans,P.R., Teo,C.-H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A splicesomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
Picard,V., Erdsal-Badju,E., Lu,A. and Clark Bock,S. (1994) A rapid and efficient one-tube PCR-based mutagenesis technique using *Pfu* DNA polymerase. *Nucleic Acids Res.*, **22**, 2587–2591.
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.
Rinke-Appel,J., Jünke,N., Stade,K. and Brimacombe,R. (1991) The path of mRNA through the *Escherichia coli* ribosome; site-directed cross-linking of mRNA analogues carrying a photo-reactive label at various points 3′ to the decoding site. *EMBO J.*, **10**, 2195–2202.
Robertson,H.D., Webster,R.E. and Zinder,N.D. (1968) Purification and properties of ribonuclease III from *Escherichia coli*. *J. Biol. Chem.*, **243**, 82–91.
Schindelin,H., Marahiel,M.A. and Heinemann,U. (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B.subtilis* major cold-shock protein. *Nature*, **364**, 164–168.
Schindelin,H., Jiang,W., Inouye,M. and Heinemann,U. (1994) Crystal structure of CspA, the major cold shock protein of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **91**, 5119–5123.
Schnuchel,A., Wiltscheck,R., Czisch,M., Herrier,M., Willimsky,G., Graumann.,P., Marahiel,M.A. and Holak,T.A. (1993) Structure in solution of the major cold-shock protein from *Bacillus subtilis*. *Nature*, **364**, 169–171.
Schweisguth,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.
Siomi,H., Matunis,M.J., Michael,W.M. and Dreyfuss,G. (1993) The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.*, **21**, 1193–1198.
Siomi,H., Choi,M., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1994) Essential role for KH domains in RNA binding: impaired RNA binding by a mutation in the KH domain of FMR1 that causes fragile X syndrome. *Cell*, **77**, 33–39.
St Johnston,D., Driever,W., Berleth,T., Richstein,S. and Nüsslein-Volhard,C. (1989) Multiple steps in the localization of *bicoid* RNA to the anterior pole of the *Drosophila* oocyte. *Development Suppl.*, **107**, 13–19.
St Johnston,D., Beuchle,D. and Nüsslein-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.
Stern,S., Powers,T., Changchien,L.-M. and Noller,H.F. (1988) Interaction of ribosomal proteins S5, S6, S11, S12, S18 and S21 with 16 S rRNA. *J. Mol. Biol.*, **201**, 683–695.
Stern,S., Powers,T., Changchien,L.-M. and Noller,H.F. (1989) RNA–protein interactions in 30S ribosomal subunits: folding and function of 16S rRNA. *Science*, **244**, 783–790.
Valegård,K., Murray,J.B., Stockley,P.G., Stonehouse,N.J. and Liljas,L. (1994) Crystal structure of an RNA bacteriophage coat protein–operator complex. *Nature*, **371**, 623–626.
Wittman-Liebold,B. and Greuer,B. (1978) The primary structure of the protein S5 from the small subunit of the *Escherichia coli* ribosome. *FEBS Lett.*, **95**, 91–98.
Wolffe,A.P. (1994) Structural and functional properties of the evolutionarily ancient Y-box family of nucleic-acid binding-proteins. *Bioessays*, **16**, 245–251.

Received on March 23, 1995; revised on April 19, 1995