The EMBO Journal vol.14 no.14 pp.3572–3584, 1995

Structure of the dsRNA binding domain of *E.coli* RNase III

Abelhakim Kharrat, Maria J.Macias, Toby J.Gibson, Michael Nilges and Annalisa Pastore[1]

European Molecular Biology Laboratory, Postfach 102209, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

A.Kharrat and M.J.Macias contributed equally to this work

[1]Corresponding author

The double-stranded RNA binding domain (dsRBD) is a ~70 residue motif found in a variety of modular proteins exhibiting diverse functions, yet always in assocation with dsRNA. We report here the structure of the dsRBD from RNase III, an enzyme present in most, perhaps all, living cells. It is involved in processing transcripts, such as rRNA precursors, by cleavage at short hairpin sequences. The RNase III protein consists of two modules, a ~150 residue N-terminal catalytic domain and a ~70 residue C-terminal recognition module, homologous with other dsRBDs. The structure of the dsRBD expressed in *Escherichia coli* has been investigated by homonuclear NMR techniques and solved with the aid of a novel calculation strategy. It was found to have an α-β-β-β-α topology in which a three-stranded anti-parallel β-sheet packs on one side against the two helices. Examination of 44 aligned dsRBD sequences reveals several conserved, positively charged residues. These residues map to the N-terminus of the second helix and a nearby loop, leading to a model for the possible contacts between the domain and dsRNA.

Key words: NMR/protein–nucleic acid interactions/RNA/tertiary structure

Introduction

It is now widely understood that many eukaryotic proteins, especially those with regulatory or structural functions, are built from modular components, enabling specific tasks to be dispersed among different domains. Modular proteins in prokaryotes are rather less common. However, in the case of RNA metabolism, proteins in both prokaryotes and eukaryotes are frequently found to possess modular structures. Since aspects of RNA metabolism are of great antiquity, some of these modular proteins appear to be common to all living cells, while the known globular RNA binding modules are all found in both prokaryotes and eukaryotes. Currently, four domains are known whose primary functions are the recognition of RNA (for reviews, see Mattaj, 1993; Burd and Dreyfuss, 1994). The RNP (or RRM), KH and double-stranded RNA binding domain (dsRBD) modules are small globular domains, 50–100 residues in length. In contrast, the ~30–80 residue RGG/RGY motif is an extended module that can make extensive and repeated interactions with the RNA backbone (Ghisolfi *et al.*, 1992; Kiledjian and Dreyfuss, 1992). Other motifs, such as TFIIIA-type zinc fingers (Miller *et al.*, 1985) may recognize either DNA or RNA. Although there may be some overlap of the functional ranges of these modules, it is possible to discern clear differences in their usage. The best characterized and the most widespread is the RNP domain (reviewed by Birney *et al.*, 1993) which recognizes 5–10 nucleotides of single-stranded RNA, often in the context of a hairpin loop, and usually exhibits sequence specificity since it contacts several bases directly (Oubridge *et al.*, 1994). The KH domain is less well characterized but occurs in several proteins with non-specific single-stranded (ss) RNA binding activity, such as polynucleotide phosphorylase and nusA of *Escherichia coli*, and in numerous eukaryotic heterogeneous nuclear RNP proteins (Gibson *et al.*, 1993; Siomi *et al.*, 1993). The RGG motif is thought to bind ssRNA and may also act in destabilizing RNA secondary structure (Ghisolfi *et al.*, 1992).

In contrast to the ssRNA binding domains, the dsRBD module has been found exclusively in proteins which recognize dsRNA. It was first noted as a ~70 residue element repeated five times in the *Drosophila* developmental protein staufen and twice in the DAI protein kinase involved in the interferon response (Green and Matthews, 1992; St. Johnston *et al.*, 1992). Both these proteins recognize dsRNA, with staufen recognizing folded elements of *Drosophila* maternal RNAs and DAI binding to viral dsRNAs. Similar motifs were then detected in several other proteins, including RNase III, which cleaves at RNA hairpins, TRBP, a cellular protein which binds the HIV TAR hairpin and vaccinia E3L which can bind poly(rI)–poly(rC) (Green and Matthews, 1992; St. Johnston *et al.*, 1992). More recently, the domain has been reported in several more sequences (Bass *et al.*, 1994; Gibson and Thompson, 1994; Kao *et al.*, 1994), including human RNA helicase A (Lee and Hurwitz, 1993) and DRADA (dsRNA adenosine deaminase) which converts adenosine to inosine in dsRNA molecules, including as substrate the glutamate receptor mRNA which is edited in hairpin duplexes (Kim *et al.*, 1994; Polson and Bass, 1994; O'Connell *et al.*, 1995).

The *E.coli* enzyme RNase III (Robertson *et al.*, 1968) is the most extensively studied of the dsRBD-containing proteins. It acts directly on dsRNA to produce single-stranded nicks or staggered double-stranded breaks. Although it can degrade dsRNA, biological substrates for RNase III, termed RNase III processing signals, are usually hairpins with at least two turns of duplex (Schweisguth *et al.*, 1994). Cleavage at the processing signals is very precise, resulting in just one or a few nicks, necessitating exact recognition of substrate features: nevertheless, there

appears to be very limited sequence specificity (Chelladuri *et al.*, 1994). In the extensively studied phage T7 processing signal R1.1, cleavage occurs at one position in an internal loop in the middle of the hairpin (Chelladuri *et al.*, 1994). The cellular functions of RNase III include: processing of rRNA precursors; cleavage in the 5′ termini of many cellular mRNAs, affecting stability and translational efficiency; and cleavage within polycistronic transcripts produced by bacteriophages such as λ and T7, providing mature mRNAs (Belasco and Higgins, 1988; Court, 1993; Hajnsdorf *et al.*, 1994). Although the enzyme has only been studied in *E.coli*, it is now clear from DNA databases that homologues are widely distributed in both eubacteria and eukaryotes (Gibson and Thompson, 1994) and the enzyme seems likely to be present in all cells.

While the evidence is now overwhelming that the dsRBD functions in recognition of dsRNA, almost nothing is known about its detailed mode of binding. In order to learn more about dsRBDs, the *E.coli* RNase III system affords several advantages: it only has a single dsRBD; many substrate RNAs are known; and its dsRBD sequence has features that are helpful for structure determination by NMR spectroscopy. Here we report the solution structure of a bacterially expressed dsRDB peptide and consider the implications of the structure for the mode of binding to dsRNA substrates.

Results

Choice and expression of the RNase III dsRBD

The domain from *E.coli* RNase III was chosen for a number of reasons, not least the extensive experimental background. The protein is also quite small (226 amino acids) and the C-terminus of the dsRBD is the natural C-terminus of the protein: thus uncertainties about domain boundaries were minimized. A 74 residue peptide encompassing the dsRBD was chosen for expression in *E.coli*. This peptide was found to be well folded and soluble and was used for all the studies reported here.

NMR analysis

Optimal conditions for solubility of the dsRBD were found at pH 6 and low salt concentration. The spectra recorded under these conditions showed very good intrinsic resolution (Figure 1). The sample was found to be stable over the temperature range 20–40°C. Therefore, to reduce peak overlap further, spectra were collected at several different temperatures. The assignment was then obtained by combining the standard strategy based on the identification of spin systems with the nuclear Overhauser effect (NOE)-based sequential assignment procedure (Wüthrich, 1986).

Several independent entrance points helped in placing the connectivities along the sequence. The identification of the unique isoleucine (Ile38), five of the six alanines, most of the leucines, valines and of the AMX spin systems was achieved directly from two-dimensional TOCSY patterns. The rings of the three histidines (His16, 32 and 39) were identified by TOCSY peaks and then connected to their β-protons by NOEs. The identification of the unique Phe36 was complicated by degenerate ring *para*- and *meta*-protons and was achieved by sequential assignment. The remaining aromatic rings (Tyr11, 22) could then be identified and connected to their amide protons by NOESY connectivities. Identification of three proline spin systems for residues 18, 20 and 48 was straightforward, while Pro4 could be identified only in the final steps of the analysis. All were found to be in the *trans* configuration, from typical NOESY peaks between their δ-protons and the α-protons of Asp3, Leu17, Leu19 and Glu47 respectively. All residues have been assigned completely, with the exception of part of the Lys58 side chain. The full set of assignments is given in Table I.

Identification of the secondary structure

Several strong sequential peaks in the fingerprint region below 4.5 p.p.m., together with sequential H_N–H_N connectivities suggested immediately the presence of an αβ fold (Figure 1). A number of long-range H_α–H_α as well as H_N–H_N and H_N–H_β connectivities were all consistent with a three-stranded β-sheet, involving residues Thr21–Val27, Glu35–Val42 and Glu47–Ser54 (Figure 2A). The arrangement of the three strands is consecutive and antiparallel. In addition, two α-helical stretches are present (as shown by the presence of H_{Ni}–H_{Ni+1}, $H_{\alpha i}$–H_{Ni+3} and $H_{\alpha i}$–$H_{\beta i+3}$ connectivities) from Lys5 to Leu17 and from Ala59 to Glu74 (Figure 1). Both the secondary chemical shifts and amide proton exchange rates show very good correlation with the secondary structure elements. Amides in helix α1 exchange in general faster than in the other well defined secondary structure elements, which suggests a higher flexibility of the domain N-terminus (Figure 2B).

The secondary structure topology of the dsRBD is therefore α-β-β-β-α.

Determination of the tertiary structure of the dsRBD

While assignment of sequence and secondary structure were straightforward, determination of the 3-D topology was complicated by the high degeneracy of the resonances between secondary structure elements (Figure 3A and Table I). The initial sets of both unambiguous and degenerate inter-residue contacts are summarized in Figure 3B. Only one tertiary contact could be identified unambiguously (between Tyr22 and Ala64) on the basis of chemical shift alone. We chose not to follow the usual strategy of attempting an initial placement of the helices onto the β-sheet that could provide entry points for further visual interpretation of the NOESY spectra. Instead, we adopted a fully automatic strategy using as input the initial set of unambigous and ambiguous NOEs. The procedure uses an iterative calculation/assignment strategy that relies heavily on distance restraints extracted from ambiguous cross-peaks (Nilges, 1993, 1995: see Materials and methods and Discussion). Most of the cross-peaks could be assigned in this way; a few cross-peaks remain ambiguous in the final set of structures (Figure 3C). Figure 4A shows a superposition of Cα coordinates from the 10 calculated best structures for the dsRBD domain. The quality of the geometry is good, as judged by a number of tests (see Table II for a list of the structure statistics). The structures are quite well ordered with the bundle superposition yielding a root-mean-square difference of 1.04 Å for the core regions 4–27, 35–74 .

The dsRBD structure has an approximately ellipsoidal shape (with the longest dimension ~3.7 nm and the

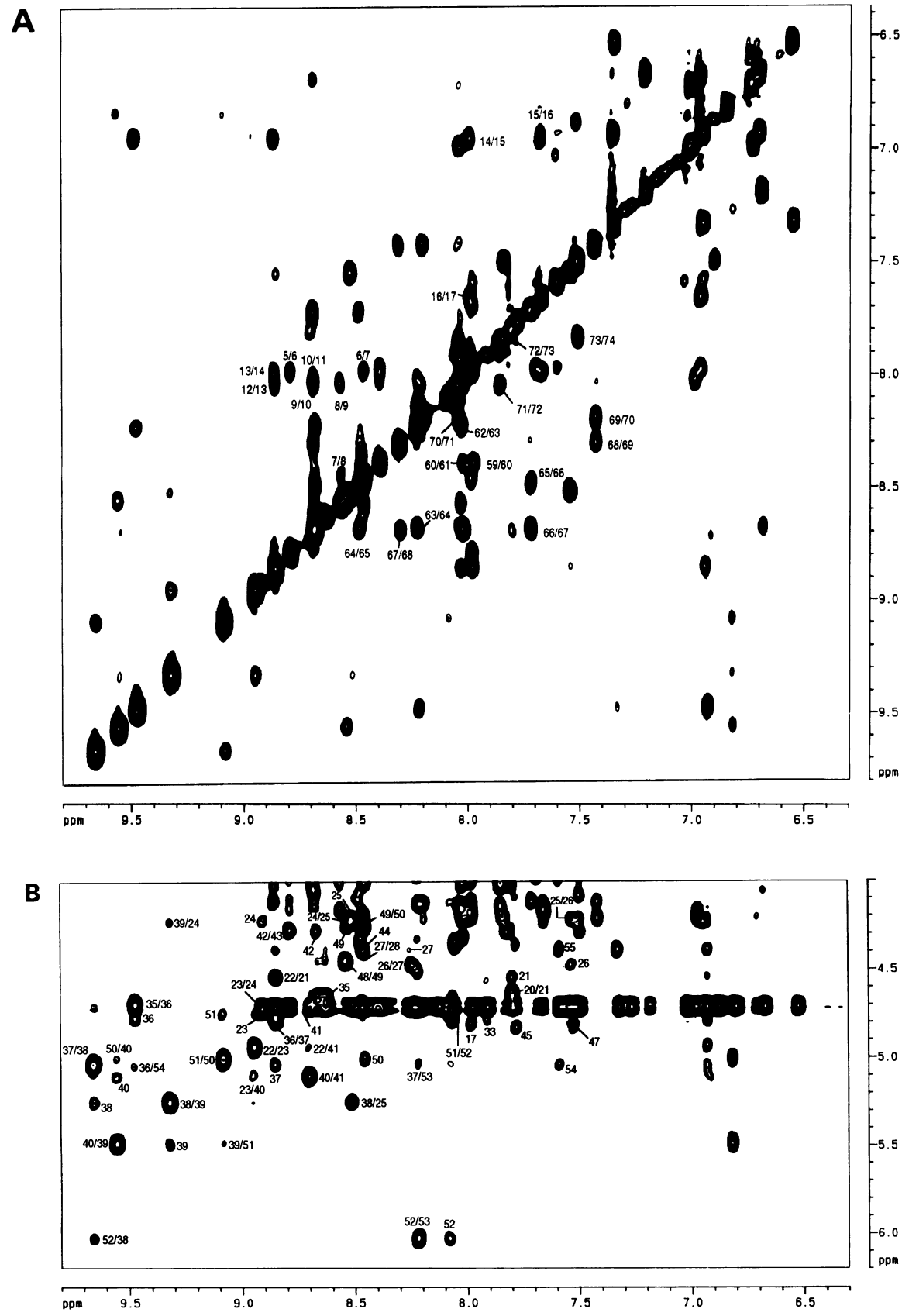

Fig. 1. NMR data and backbone assignments for the RNase III dsRBD. (**A**) NOESY spectrum at 300 K and 600 MHz of a 2 mM solution sample of the domain at pH 6.0. Numbers above the diagonal indicate the connectivities between contiguous amides in helix α1 (Lys5–Leu17). Numbers below the diagonal indicate the connectivities between amides in helix α2 (Ala59–Glu74). (**B**) Fingerprint region of the NOESY spectrum. Numbering indicates the sequential assignment of the three-stranded β-sheet.

Table I. Chemical shift in the assigned ^{1}H atoms of the dsRBD at 300 K, 600 MHz and pH 6.0

Amino acid	Chemical shifts			
	HN	HA	HB	Others
D 3	8.34	4.71	3.11–3.08	
P 4		4.15	2.43	HD 3.99 HD 4.05
K 5	8.96	3.88	2.1–1.75	HG 1.37 HD 1.60 HE 2.90
T 6	8.11	4.33	4.01	HG 1.33
R 7	8.59	4.15	2.04–1.85	HG1 1.94 HG2 1.62 HD 3.34
L 8	8.71	4.31	2.20–1.81	HG 1.54 HD1 0.98 HD2 1.08
Q 9	8.16	3.99	2.49	HG 1.92–1.85 ND 7.07–7.72
E 10	8.83	4.16	2.21–2.14	HG1 2.78 HG2 2.36
Y 11	8.12	4.36	3.31	HD 7.09 HE 6.82
L 12	8.15	3.84	1.89–1.64	HG 2.19 HD1 1.21 HD2 1.12
Q 13	8.99	4.26	2.35–2.15	HG1 2.70 HG2 2.54 ND 7.03–7.65
G 14	8.11	4.33–3.96		
R 15	7.09	4.38	1.54	HG 2.19 HD 2.94
H 16	7.78	4.31	3.52–3.40	
L 17	8.11	4.94	1.72	HG 1.50 HD1 1.08 HD2 0.97
P 18		4.47	2.14–1.83	HG 1.83–1.64 HD 3.92–3.60
L 19	8.21	4.49	2.04–1.79	HD 1.13
P 20		4.8	2.17	HG 1.77 HD1 3.64–3.74
T 21	7.91	4.69	4.17	HG 1.31
Y 22	9.01	5.06	3.01–2.79	HD 7.04 HE 6.79
L 23	9.05	4.88	1.76–1.71	HG 1.59 HD1 1.05 HD2 0.99
V 24	9.05	4.35	2.28	HG1 1.20 HG2 1.08
V 25	8.65	4.36	1.83	HG1 0.63 HG2 0.29
Q 26	7.65	4.61	2.36–2.01	HG1 2.26 HG2 2.13 ND 6.93–7.77
V 27	8.39	4.53	1.72	HG1 0.58 HG2 0.36
R 28	8.59	4.82	1.93–1.81	HG 1.66 HD 3.25
G 29	8.55	4.61–3.90		
E 30	8.27	4.51	2.40–2.24	HG1 1.88 HG2 1.95
A 31	8.35	4.02	1.55	
H 32	8.59	4.73	3.39–3.26	HD1 7.92 HD2 7.12
D 33	8.04	4.95	2.64–2.54	
Q 34	8.22	4.81	2.33–1.99	HG 1.80 ND 7.33–6.81
E 35	8.77	4.84	2.06–1.97	HG1 2.24 HG2 1.55
F 36	9.62	4.92	3.01–2.59	HE 7.45 HD 7.04
T 37	8.98	5.18	4.04	HG 1.25
I 38	9.79	5.39	2.01	HG1 1.55 HD2 1.09 HD1 1.14
H 39	9.46	5.62	3.11–2.87	HD1 7.40 HD2 6.94
C 40	9.71	5.27	2.69–2.47	
Q 41	8.81	4.81	2.33–2.15	HG 2.06–1.80 ND 7.48–6.61
V 42	8.83	4.41	2.01	HG1 1.11 HG2 1.01
S 43	8.94	4.23	3.95	
G 44	8.65	4.50–3.70		
L 45	7.91	4.98	2.15–1.69	HG1 1.59 HD 1.01
S 46	7.65	4.36	3.90–3.97	
E 47	7.66	4.96	2.16–1.96	HG 2.35
P 48		4.59	1.99	HG 1.84 HD 3.89 HD 3.74
V 49	8.71	4.38	2.11	HG1 1.11 HG2 1.06
V 50	8.61	5.14	2.1	HG1 1.12 HG2 0.97
G 51	9.2	4.88–3.99		
T 52	8.22	6.18	4.27	HG 1.38
G 53	8.34	4.65–4.09		
S 54	7.71	5.19	4.36–4.05	
S 55	7.71	4.52	3.99–3.89	
R 56	8.82	4	2.25–1.78	HG 1.88 HD 3.30
R 57	7.83	4.09	1.85	HG 1.63 HD 2.12
K 58	7.71			
A 59	8.09	4.09	1.53	
E 60	8.51	3.93	2.30–1.27	HG1 2.66 HG2 1.89
Q 61	8.14	4.3	2.23–2.23	HG 2.71
A 62	8.12	4.48	1.89	
A 63	8.34	4.09	1.62	
A 64	8.81	4.2	1.72	
E 65	8.6	4.07	2.41–2.35	HG 2.58
Q 66	7.84	4.27	2.40–2.27	HG 2.77 ND 7.16–7.74
A 67	8.82	4.04	1.62	
L 68	8.42	3.86	2.08–1.94	HG 1.44 HD1 1.09 HD2 0.81
K 69	7.53	4.36	2.09	HG 1.61 HD 1.78 HE 3.15
K 70	8.32	4.27	2.07	HG 1.82 HD 1.67 HE1 3.05 HE2 2.99

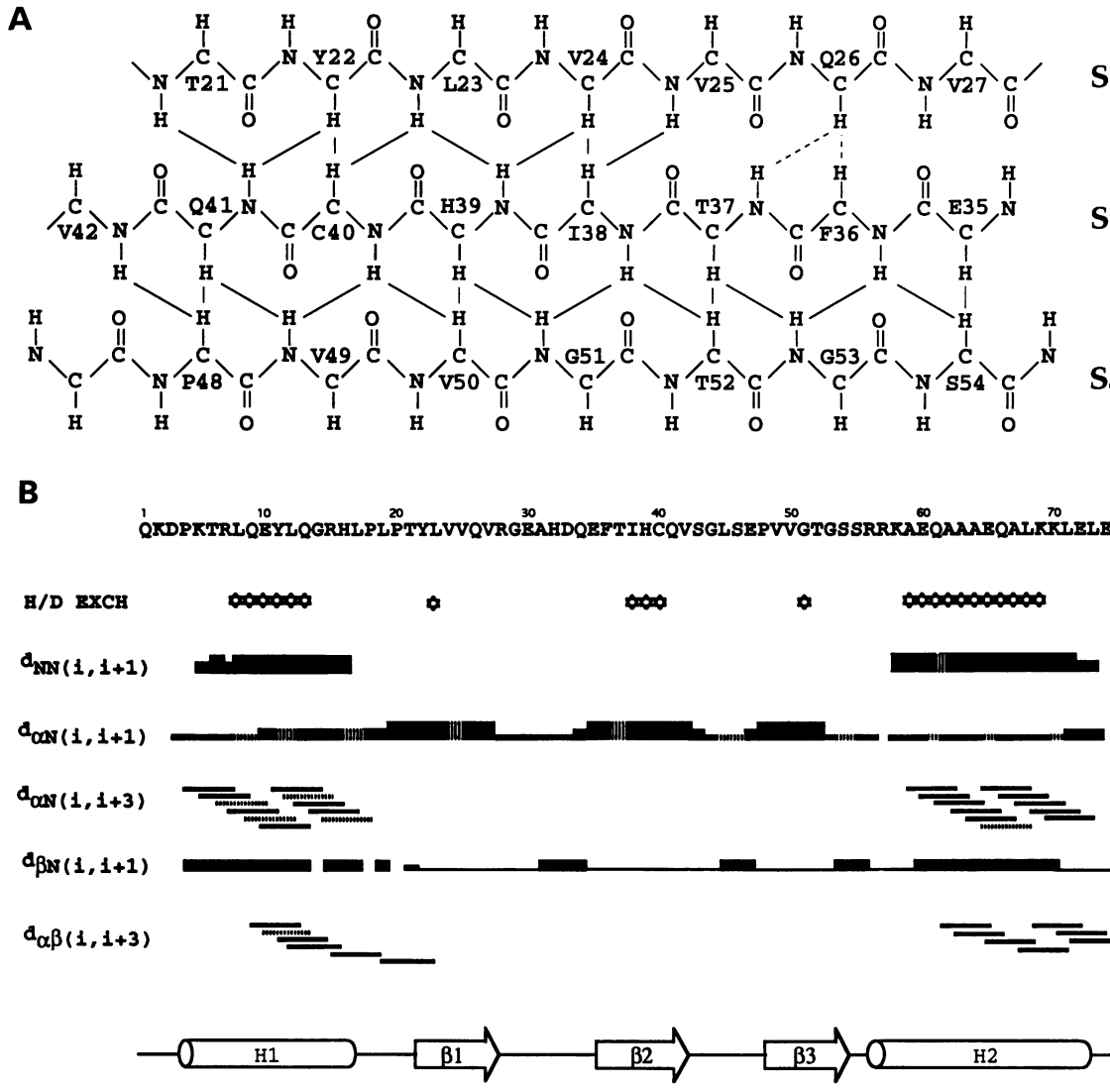

Fig. 2. Secondary structure of the RNase III dsRBD. (**A**) Scheme showing the three-stranded β-sheet. The strands are in consecutive order within the sheet. Sheet connectivities observed by NMR are indicated by interstrand lines. Dotted lines indicate distances longer than average for a sheet. (**B**) Survey of the observed short and medium range NOEs. For each residue, the amino acid is given together with the main short range contacts and slowly exchanging amide protons (marked by ✲ in H/D EXCH). Line thickness is proportional to peak size. Open boxes indicate superimposed peaks. Arrows and cylinders are used respectively to portray the α- and β-secondary structure elements.

other two dimensions ~2.0 nm in length). The secondary structure elements are all individually very well formed in the calculated solutions. There is some variability in the orientation of helix α1 with respect to the rest, which is not unexpected since it is oriented by only a few NOE contacts (Figure 3B and C) and has relatively rapid amide exchange rates. While the domain N-terminus is mainly disordered, the C-terminus shows only small divergence among the structures right to the end of the domain. Of the four loops connecting the β-sheet and the two helices, three are reasonably well ordered. In contrast, the longest loop, spanning residues 28–34, shows large conformational variability. The preceding strand, β1, ends with a β-bulge (Val24–Val27) that seems to define the initial direction of the loop. The three-stranded β-sheet possesses a typical left-handed twist which enables it to enwrap helix α2 partially. Both helices lie on the same side of the sheet. Helix α2 packs particularly closely against β3, with almost exclusively small residues in the interface (Gly51, Gly53 and Ser55 from β3, and Ala59 from α2). The connection between β3 and α1 is very tight, with Ser55 acting as helix cap by accepting H-bonds from the α2 N-terminus. Helix α1 packs loosely into a groove formed by α2 and the β-sheet, with only a few points of contact to the sheet (between Tyr22 and Leu8). A central hydrophobic core is formed by residues Leu8, Tyr22, Cys40 and Leu68 where the two helices and the sheet all pack together. Residues Tyr11, Leu17 and Leu45 form a second small patch of hydrophobic residues that closes the main cluster (at the lower part of the view in Figure 4B). The exposed side of the β-sheet consists of predominantly hydrophilic but uncharged residues. The opposite side (formed by the two helices) is however highly charged, especially near the N-terminus of α2.

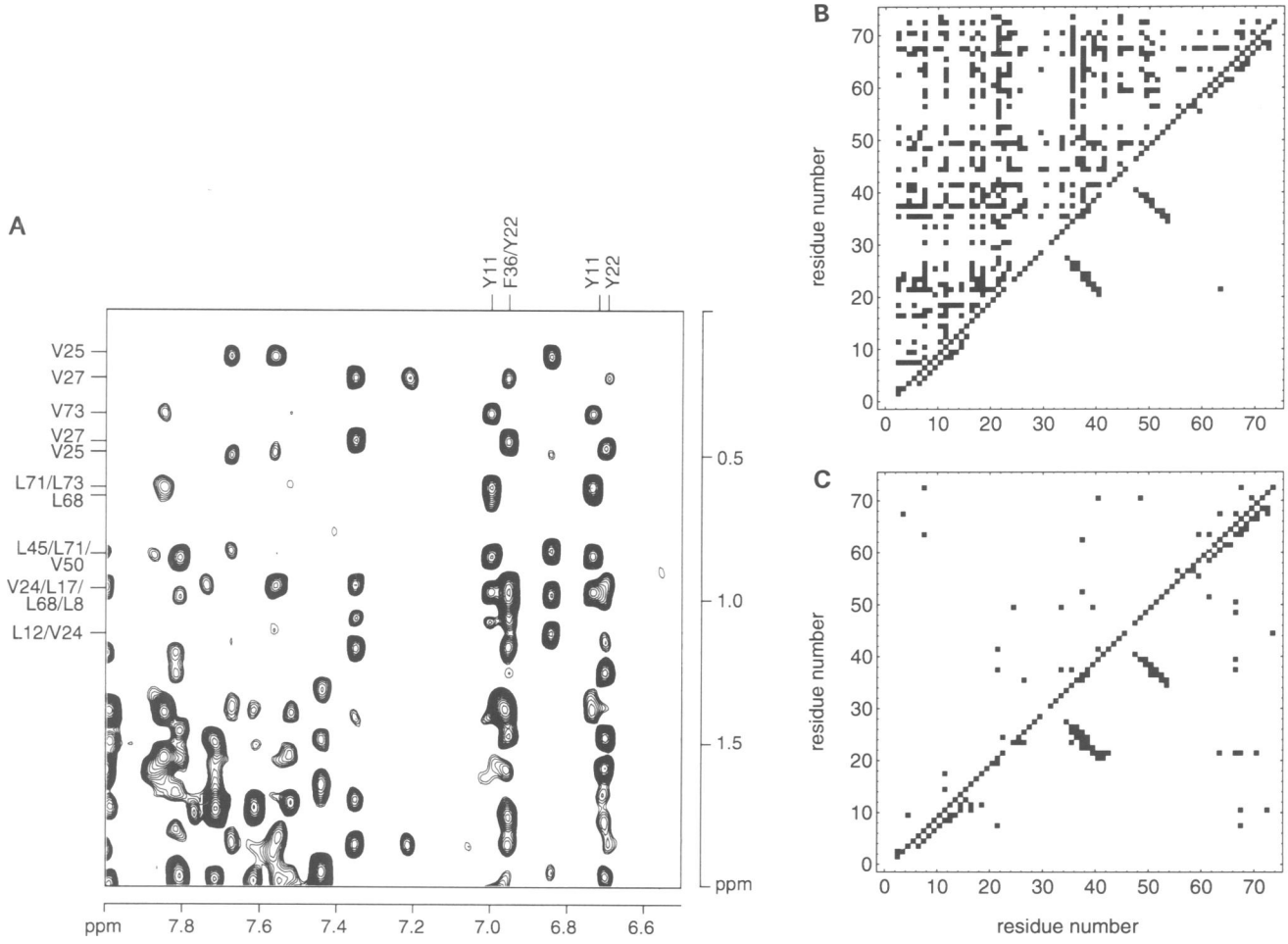

Fig. 3. (**A**) Region of the same NOESY spectrum as Figure 1 which contains contacts between aromatic and methyl protons. Ambiguous assignment of some of the resonances is indicated. (**B**) 2-D plot of the sequence against itself, summarizing the initial set of inter-residue restraints. Above and left of the diagonal are plotted the ambiguous restraints from degenerate NOEs. Below and right of the diagonal are plotted the unambiguous restraints derived from non-degenerate NOEs. (**C**) 2-D plot summarizing the final set of inter-residue restraints. Above and left of the diagonal are plotted the ambiguous restraints still remaining after the calculation. Most of these are truly ambiguous NOEs, having several contributions, while a few correspond to alternate assignments in the calculated structures. Below and right of the diagonal are plotted the unambiguous restraints consisting of the calculated non-degenerate restraints added to the initial list from the non-degenerate NOEs.

DsRBD-containing proteins

Figure 5 shows an alignment of 44 dsRBDs, including seven newly identified sequences. The solved structure has been used to ensure that the core-forming positions have appropriate residues. In particular, the alignment of strand β3 has been improved since the last comprehensive alignment (Gibson and Thompson, 1994). Some of the newly detected sequences are in cosmids or ESTs (expressed sequence tags) generated by large sequencing projects and are of unknown function. One EST, CEK019CYR, is homologous to human Son-A. The dsRBD in the cosmid T20H4 (Wilson *et al.*, 1994) is adjacent to an identified similarity to human DRADA and is likely to be an additional exon. The *Caenorhabditis elegans* sequence only has a single dsRBD, unlike DRADA: the similarity of the three DRADA dsRBDs clearly indicates that they arose by domain duplication events from a single ancestral domain.

Two of the new sequences are RNase III homologues. One is an open reading frame in MO15224, a mycoplasma from potato. The other is composed of three overlapping ESTs from a human library. These ESTs are known to be contaminated by bacterial and yeast sequences. The reconstructed RNase III is clearly bacterial (data not shown) and is not the human homologue. Since there are now seven RNase III sequences known (three eubacterial, two mycoplasma and two eukaryotic), the enzyme appears to be ubiquitous. It should be emphasized that this is at odds with the fact that RNase III has not been thought relevant to RNA processing in eukaryotes, with the single exception of the enzyme encoded by the essential gene *pac1* in *Schizosaccharomyces pombe* (Iino *et al.*, 1991). In particular, RNase III has not been considered to have a role in eukaryotic rRNA processing, despite its importance in *E.coli*.

Inspection of the aligned sequence set reveals that all the conserved hydrophobic residues are involved in the structural core. There are also seven positions which show considerable positive charge conservation that map to the surface of the structure (Figure 5). Therefore, these residues are likely to be functional.

Table II. Structural statistics for the 10 best calculated structures

R.m.s. difference from ideality	
Bonds (Å)	0.0024 ±0.0001
Angles (deg)	0.57 ± 0.012
Impropers (deg)	0.43 ± 0.014
Quartic vdW term (kcal/mol)	16.2 ± 3.1
vdW energy (kcal/mol)[a]	−133.5 ± 13.1
Fit to experimental data	
NOE distances (Å)	0.039 ± 0.0032
Violations >0.4 Å	0 ± 0
Violations >0.3 Å	0.9 ± 0.74
Quality control	
Prosa average energy[b]	−1.26 ± 0.16
Whatif average energy[c]	−2.17 ± 0.14
R.m.s. difference from average structure[d]	
C, Cα, N for range 4–27, 35–74 (Å)	1.04 ± 0.19
C, Cα, N of all residues (Å)	2.30 ± 0.44
All non-hydrogen atoms (Å)	2.90 ± 0.41

[a]van der Waal's energy calculated with the standard Lennart–Jones parameters of the CHARM parmallh6 force field (Brooks *et al.*, 1983). This energy term is used for analysis only and is not part of the target function.
[b]Prosa evaluation is according to Sippl (1993). The dominating positive terms are from solvent-accessible, unconserved residues Leu19 and Leu45.
[c]Whatif evaluation is according to Vriend and Sander (1993).
[d]The average structure was calculated by a best fit of residues 4–27 and 35–74. The 10 best structures were superimposed.

Discussion

Direct tertiary structure calculation using an ambiguous NMR dataset

The high solubility and small size of the RNase III dsRBD resulted in spectra with generally good dispersion, which enabled a rapid and straightforward assignment of the sequence and the secondary structure elements. At this stage it became apparent that NOEs defining tertiary folds were less well resolved, requiring either a difficult and laborious analysis using 2-D NMR data or else costly isotopic labelling and 3-D NMR techniques. Instead, the degeneracies in the NOEs determining the packing were dealt with efficiently by an iterative calculation strategy utilizing ambiguous restraints.

Several years ago, it became clear that a calculation strategy using ambiguous restraints was needed to overcome the difficulty of solving symmetrical dimers with NMR data. In such a case, it is impossible to make an *a priori* distinction between inter- and intra-molecular NOEs (Saudek *et al.*, 1991). One such calculation strategy, in which all possibilities for the ambiguous NOEs were kept open, was developed using a MetJ repressor homodimer model system (Nilges, 1993). More recently, this strategy has been extended to the general case for NOE ambiguities using pancreatic trypsin inhibitor as a monomeric model system (Nilges, 1995). The approach has proven useful in refinement of the M13 phage gene 5 homodimer (Folkers *et al.*, 1994) and the pleckstrin homology domain monomer (Macias *et al.*, 1994) and as an aid to a difficult assignment problem with the p53 tetramerization domain (Lee *et al.*, 1994).

In the work reported here, a modified strategy has been used whereby an ensemble of structures is iteratively refined and, after each iteration, the list of possibilities for each ambiguous NOE may be reduced by discarding the most unlikely alternatives. To our knowledge, this is the first time such a strategy has been applied directly to the automatic determination of a structure from the initial NOE dataset.

As well as a substantial increase in computational efficiency, no manual intervention is required in the method. Furthermore, subjective decisions made during analysis of the NMR spectra can be wholely avoided, minimizing the likelihood of human error degrading the final structure. The method allows an interpretation of the complete NOE dataset, whereas a typical NMR solution will have data selectively omitted by human decision. The new strategy is also likely to be more powerful than automated assignment strategies which reject the ambigous NOEs before the initial model construction (Guenthert *et al.*, 1993; Meadows *et al.*, 1994). It is unlikely that such strategies could have worked with the present dataset.

Structure of the RNase III dsRBD

The domain adopts a tertiary fold consisting of five secondary structure elements in the order α-β-β-β-α. These coalesce into an anti-parallel three-stranded β-sheet which is solvent exposed on one face but packed against the two helices on the other (Figure 4). Helix α2 packs against all three strands, while α1 only has contacts to β1 and α2. Thus, for most of its length, α2 is the more deeply buried of the two helices. This is reflected in greater hydrophobic residue conservation in α2 than in α1 in all the dsRBDs (Figure 5). The helical axes diverge at an angle of ~40° from parallel, so that their N-termini are found close by on the same end of the structure. This arrangement of the helices, together with the presence of three positively charged residues on the α2 N-terminus, brings a large array of partial and full positive charges into proximity on one face of the structure. This striking arrangement is difficult to reconcile with proposals that α-helical dipoles are important for structure stability (Hol *et al.*, 1981). On the other hand, it is likely to be very important for function. The topology of the helices also has the consequence that the N- and C-termini of the dsRBD are at opposite ends of the structure: therefore, in proteins with dsRBD repeats, the domains can be arranged consecutively in space.

The topology of the RNase III domain is very similar overall to that of the homologous domain, from the *Drosophila* staufen protein, solved by Bycroft *et al.* (1995) in an accompanying paper. These authors also report that the N-terminal domain of ribosomal protein S5 (Ramakrishnan and White, 1992) shares the same fold (except that it lacks the N-terminal helix) and is therefore likely to be a distant homologue of the dsRBD. Our dsRBD sequence profile searches were unable to pick up any S5 entries, which may imply a limit to the sensitivity of the current method.

Implications for double-stranded RNA recognition

The dsRBD structure reveals an array of positive charges and hydrogen bond donors, situated on one surface of the domain, that seem well suited for contacting backbone

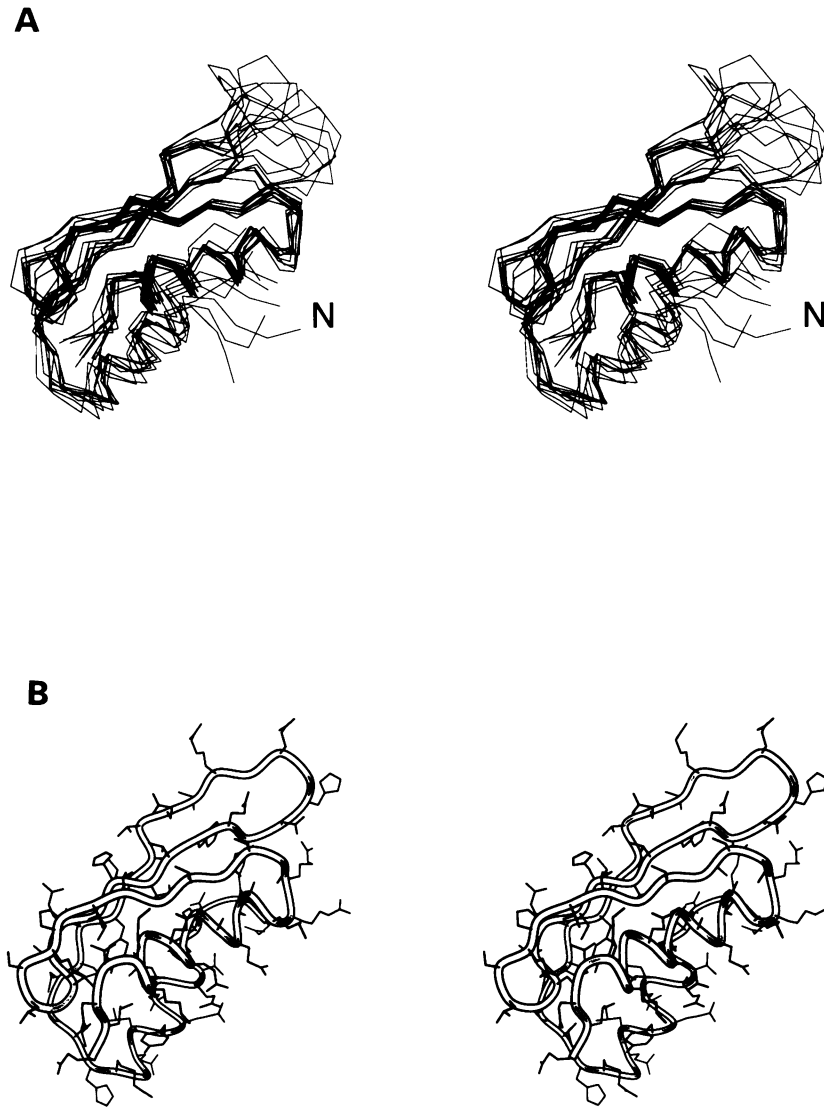

Fig. 4. Tertiary structure of the RNase III dsRBD. The orientation is the same in all views. (**A**) Stereo pair showing the Cα traces in the superimposed bundle of the 10 best calculated structures. (**B**) Stereo pair of the top structure, including side chains. The backbone is represented as a tube.

phosphates in RNA. Three semi-conserved positive charges (residues 32, 34 and 35) are found on loop 2 between β1 and β2, while three strongly conserved charges (residues 56, 57 and 60) plus three conserved H bond donors (54, 55 and 61) are found at the α2 N-terminus (see Figure 5). Since the α2 N-terminus itself is in the centre of this charge array, it is in an excellent position to donate peptide backbone H-bonds to phosphates—which would impart greater rigidity to the complex than the interactions with flexible side chains. It is also possible that backbone amides from loop 2 could donate H-bonds to RNA phosphates. Loop 2 is extruded and quite flexible in solution but would then be stabilized by an induced fit to the RNA upon binding.

Figure 6 shows, for comparison, the proposed binding surface of the dsRBD structure oriented towards an A-form RNA duplex. With the domain positioned so that the conserved surface residues around the α2 N-terminus are oriented toward the RNA, it can be seen that the domain provides sufficient ligands to make contacts to both strands, by crossing either the narrow major groove (as shown) or else the wider minor groove. Although this might imply that sequence-specific base contacts in one of the grooves could also be made, there is no evidence to suggest that this happens. In the case of DRADA, the shorter the dsRNA substrate, the greater the specificity of adenosine modification (Polson and Bass, 1994). It has been found that the three DRADA dsRBD domains can each bind independently to dsRNA (Kim *et al.*, 1994), although with lower affinity than all three domains together. This suggests that, *in vivo*, there will be a minimum length of duplex for DRADA binding, imposed by the triple repeat of dsRBD domains.

Recently, the effects of alanine-scanning mutagenesis for several C-terminal residues in dsRBD/1 of the human DAI kinase (also called PKR) have been presented (McMillan *et al.*, 1995). Using the numbering in the Figure 5 alignment, it was found that Gly53 and Lys56 substitutions lead to complete loss of dsRNA binding, while substitutions at Lys60 and Leu71 lead to severe

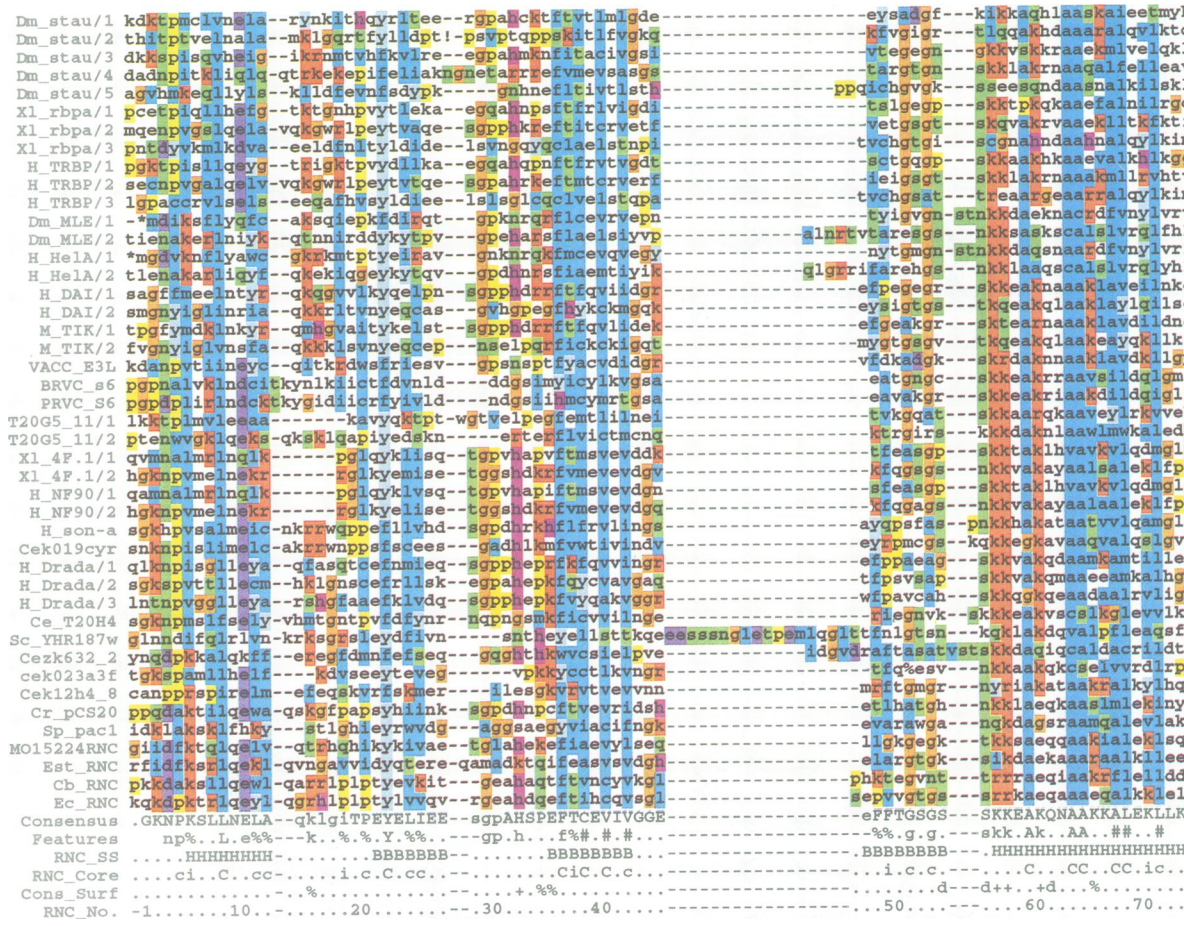

Fig. 5. Alignment of 44 dsRBD sequences, colour-coded to highlight similarities, together with a summary of structural details. An ! marks a 92 residue insertion in staufen repeat 2. Summary lines are: consensus representing the most frequent residue at each position; features, the strongly conserved features (symbols: # strongly conserved hydrophobicity; % semi-conserved hydrophobicity); RNC_SS, the observed secondary structure of RNase III; RNC_Core, categories of packing residue (i, interface; c, core and C, deep core); Cons_Surf, conserved surface positive charges (+, or % if weak), and H-bond donors (d). All G (orange) and P (yellow) residues are coloured. Other colouring is by conserved property in >40% of a column: uncoloured residues lack a sufficiently conserved property. Blue, hydrophobic; light blue, partially hydrophobic; red and pink, positive; purple, negative; green, hydrophilic. Database entries and accession numbers are listed in Materials and methods or in Gibson and Thompson (1994). The figure was prepared with the GDE alignment editor (S.Smith, Harvard University) and COLORMASK (J.Thompson, EMBL).

reduction in dsRNA binding. In the absence of either a solved structure or the application of methods such as circular dichroism, which can monitor structural stability, it is not possible to distinguish whether loss of function mutations arise from a structural pertubation or removal of a critical ligand from the interaction surface. With the aid of the dsRBD structure, it is clear that the substitutions at positions 53 and 71 affect important core packing residues, while the mutations at positions 56 and 60 affect exposed, conserved positive charges. Thus the latter results emphasize the importance of the helix α2 N-terminal region in dsRNA recognition.

If the dsRBD spans one RNA groove and contacts each backbone once only, a single dsRBD would require a minimum substrate duplex of not more that one turn, i.e. ≤12 bp. This contrasts with the observed size of RNase III substrates which are normally two turns of duplex. However, in the case of the R1.1 processing signal, the cleavage site is in a loop with about one helical turn of duplex to either side. It is possible that the dsRBD domain only binds to one side of the cleavage site. Determination of the minimal RNA duplex that will efficiently bind the dsRBD will be an important first step in establishing specific dsRBD/dsRNA complexes for structural analysis.

Comparison to other classes of RNA binding domain

The three RNA binding domains, RNP, dsRBD and KH all have αβ folds with a β-sheet packed against a pair of α helices (Oubridge *et al.*, 1994; Castiglione Morelli *et al.*, 1995). The RNP domain is currently the best studied small RNA binding module. It has the secondary structure pattern β-α-β-β-α-β, which it shares with the small enzyme acylphosphatase (Pastore *et al.*, 1992). The structure is topologically non-equivalent to the α-β-β-β-α structure of the dsRBD, despite the common theme of a β-sheet packed against a pair of helices. The high resolution crystal structure of an RNP domain complexed with an RNA hairpin reveals that the RNA binding surface is formed by the surface of the β-sheet (Oubridge *et al.*, 1994). Individual bases of ssRNA lie against the β-sheet in a groove bordered by loop regions and the domain C-terminus. This arrangement allows sequence-specific recognition of ssRNA.

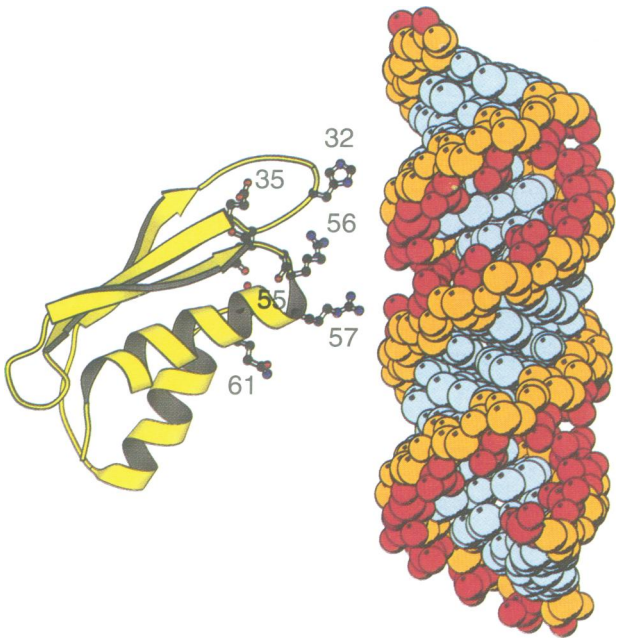

Fig. 6. Model of the interaction between the dsRBD and 20 bp of A-form RNA. Only exposed conserved (predominantly positively charged) side chains are shown. These are likely to be part of the RNA recognition surface. The domain is placed so that it could access both backbone strands, across the narrow major groove. RNA colour scheme: red, phosphates; orange, sugars; light blue, bases. The figure was produced with MOLSCRIPT (Kraulis, 1991).

The KH domain has the arrangement β-α-α-β-β(-α) which is topologically non-equivalent to the other two domains (Castiglione Morelli *et al.*, 1995). While little is known about the RNA interaction, the available evidence suggests that the KH is likely to interact with ssRNA via the helical side (Gibson *et al.*, 1993).

Thus, despite a superficial structural resemblance, these small domains appear to represent three quite different evolutionary solutions to the problem of RNA recognition.

Materials and methods

Sample preparation

The DNA sequence coding for residues 153–226 of RNase III, spanning the dsRBD, was amplified by PCR using *E.coli* DNA as a template. The PCR primers were designed to fuse fMet and Gly residues to the N-terminus of the construct. The PCR primers were 5′-TAT ATT CAT GAG TCA TCA CCA TCA CCA TCA GTC CAT GGG TCA AAA AGA TCC GAA AAC GCG CTT GCA AGA A 3′ and 5′ ATA TAA GCT TAT TCC AGC TCC AGT TTT TTC AAC GCC TGT TCG-3′

The amplified fragment was cloned into the modified expression vector pET3d (van der Oost *et al.*, 1992) and expressed in *E.coli* strain BL21.

The pellet from a 4 l culture was suspended with 100 ml of bis-Tris propane buffer 5 mM (pH 6) containing the proteolytic inhibitor phenylmethyl sulfonylfluoride (0.11 mM). Cells were lysed by French press at 4°C. The viscosity of the suspension was reduced by adding DNase (2 mg) and $MgCl_2$ (3 mM). The suspension was centrifuged for 1 h at 40 000 r.p.m. in a Sorvall H6000A rotor, and the clear supernatant was recovered and filtered. The filtrate was applied to a Q-Sepharose FastFlow anion-exchange column equilibrated with bis-Tris propane 5 mM (pH 6.0). The column was developed with a linear NaCl gradient from 0 to 0.5 M NaCl. Peak fractions containing the domain (which elutes at 0.26 M NaCl) were pooled, desalted and diluted with bis-Tris propane 5 mM (pH 9.0). For the last step of purification, a monoQ column equilibrated with the same buffer was used. The protein was then dialysed against 100 mM NaCl solution (pH 6.0) and brought to a final concentration of 2 mM. The mass of the fragment was checked by electrospray mass spectroscopy carried out with a Sciex API III instrument. The found mass is 8350.2, which corresponds to the construct where the initial methionine is cleaved away. This was also confirmed by the absence of the ε-proton in the 1-D NMR spectrum. The Gly could be identified in the 2-D spectra but, since it is not part of the domain sequence, is not considered further in the manuscript.

Table III. Categories of restraints used in the calculations

Initial data list	No. of restraints
Ambiguous	106
Intra-residue	0
Sequential	211
Medium range	47
Long range	36
Final data list	**No. of restraints**
Ambiguous	21
Intra-residue	1
Sequential	223
Medium range	65
Long range	90

Nuclear magnetic resonance

The NMR measurements were carried out using 2 mM samples in 90% H_2O/10% D_2O on either a BRUKER AMX-500 or an AMX-600 spectrometer. All 2-D NMR spectra were acquired in phase-sensitive mode (TPPI) (Marion and Wüthrich, 1983) either with pre-irradiation of the water resonance or with selective excitation (WATERGATE pulse sequence, Piotto *et al.*, 1992). Clean TOCSY spectra (Griesinger *et al.*, 1988) were measured using the TOWNY composite pulse cycle (Kadkhodaei *et al.*, 1993). Mixing times used were in the 30–60 ms range for the TOWNY and 60–180 ms for the NOESY experiments. 2-D spectra were recorded at 290, 303 and 310 K, with 2048 data points in the acquisition domain and 700–1024 data points in t_1. Prior to Fourier transformation, the data were zero-filled to 2048 points in the t_1 dimension and weighted with a Gaussian window in t_2 and a sine window in t_1. A baseline correction was performed in both dimensions using a polynomial. Data were processed on a Bruker X-32 station using the UXNMR program. The AURELIA program (Peter Neidig, Bruker, Karlsruhe; a gift of Bruker) was used for measuring peak volumes.

Structure calculations

NOE volumes were integrated with AURELIA. Only inter-residue NOEs were applied in the calculations. In total, 400 NOE-derived restraints were used (see Table III for the list of restraint categories). The calculation used upper bounds only, which were set to 20% higher than the quantitated distance and rounded to the nearest 0.1 Å. Many of the NOEs determining the packing of the two helices and the β-sheet could not be assigned unambiguously (Figure 3B). These NOEs were quantitated as for the unambiguously assigned ones and used in the calculations in the form of ambiguous distance restraints (Nilges, 1993, 1995). In the automated procedure used in this paper (see below), the NOEs were classified as ambiguous solely on the basis of chemical shifts, apart from intra-residue and secondary structure NOEs. Floating assignment (Weber *et al.*, 1988; Holak *et al.*, 1989) was used for the methylene and propyl groups. Methyl groups and equivalent aromatic protons were treated like ambiguous NOEs, i.e. in the form suggested by Levy *et al.* (1989). The NOE-derived distance restraints were complemented by distance restraints for hydrogen bonds for the slowly exchanging amide protons. The donor–acceptor distance was restrained between 2.7 Å and 3.2 Å, and the hydrogen–acceptor distance between 1.7 Å and 2.2 Å. The hydrogen bonds at the N-terminal end of helix α2 were assigned ambiguous distance restraints with several possibilities for the hydrogen bond acceptor, in order to allow for possible irregularities due to end effects. The carbonyl and side chain oxygens of residues 54–56 and residues 55–57 were possibly acceptable for hydrogen bonds from the amides of residues 59 and 60 respectively.

Structures were generated using simulated annealing protocols similar to those published (Nilges *et al.*, 1988, 1991) with an extended version of X-PLOR version 3.1 (Brünger, 1989). Modifications were introduced

Table IV. Parameters used in the refinement protocol

Stage	Search[a]	Cool1	Cool2/mini
Temperature (K)[b]	2000.0	2000.0→1000.0	1000.0→100.0
Masses (a.m.u.)	100.0	100.0	100.0
Energy constants			
K_{bonds} [(kcal/(mol/Å^2)]	1000.0	1000.0	1000.0
K_{angles} [(kcal/(mol rad^2)]	500.0	500.0	500.0
K_{planar} [kcal/(mol rad^2)]	500.0	500.0	500.0
K_{chiral} [kcal/(mol rad^2)]	500.0	500.0	500.0
K_{repel} [kcal/(mol Å^4)]	0.02→0.1[c]	0.003→4.0	4.0
K_{float}[d] [kcal/(mol rad^2)]	5.0	5.0→500.0	500.0
K_{NOE} [kcal/(mol rad^2)]	1.0	50.0	50.0
K_{ambig}[e] [kcal/(mol Å^2)]	10.0→50.0	1.0→50.0	50.0
Asymptote[f]	2.0	2.0	2.0
S (Å)[f]	1.0	1.0	1.0
Steps	6500	5000	2000/250

[a]The search phase was omitted in the refinement cycles.
[b]The temperature is maintained by coupling to a heat bath (Berendsen *et al.*, 1984) with a coupling constant of 10 ps^{-1}.
[c]The non-bonded interactions are computed only between Cα atoms and one carbon atom for each side chain, with van der Waals radii of 2.25 Å.
[d]Angle energy constant for diastereospecifically unassigned methylene and propyl groups, which were treated with a floating chirality assignment approach (Holak *et al.*, 1989).
[e]Energy constant for ambiguous distance restraints and distance restraints derived from hydrogen bonds. An ambiguous distance restraint derived from an ambiguous NOE is applied by restraining the appropriate sum over all possible distances D_i between protons with appropriate chemical shifts to upper and lower bounds U and L derived from the size of the NOE cross-peak: $L<(\Sigma_i D_i^{-6})^{-1/6}<U$.
[f]Asymptotic slope and inflexion point of the flexible distance restraining potential (Nilges *et al.*, 1988).

in the protocols to include the ambiguous distance restraints, floating assignment for prochiral groups and a reduced representation for non-bonded interactions for part of the calculation to increase efficiency. The extensions of X-PLOR were necessary for the floating assignment approach used, the analysis and partial automatic assignment of the ambiguous NOEs and the best-fitting procedure (see below).

Fifty structures were calculated starting from conformations with random backbone torsion angles. The data sets and the structures were then refined in an iterative procedure. In each iteration, the ambiguous NOEs were partially assigned by keeping only those possibilities that were predominant in the 10 structures with lowest energy for the next iteration of refinement. This was done in the following way. For each possible assignment, the minimum distance, D_{min}, in the ensemble of 10 structures was determined. These distances were then sorted according to size. The Np possibilities corresponding to the shortest distances were selected such that,

$$\sum_{k}^{Np} D_{min\,k}^{-6} > p \sum_{i}^{N} D_{min\,i}^{-6}$$

where p was reduced from 0.99 for the first iterations to 0.80 for the last and N is the total number of possible assignments. A total of seven iterations of refinement/assignment were performed. Each refinement consisted of the cool1/cool2/mini phases in Table IV, and took ~15 min on an R4400 processor per structure.

A detailed description of the procedure and a comparison to other strategies (Günthert *et al.*, 1993; Meadows *et al.*,, 1994) will be published elsewhere.

The structures were superimposed using an iterative multiple-structure best-fitting procedure (which has been interfaced with X-PLOR) that automatically searches for a well defined region in an ensemble of structures (Nilges *et al.*, 1988). The final set of 10 best structures superimposes with an RMSD of 1.04 Å for the backbone atoms of residues 4–27, 35–74. The structures have good stereochemistry: e.g. average bond angle deviations of 0.57° (see Table II for the full structure statistics).

DsRBD profiles and database searches

Thirty-three previously aligned dsRBD sequences (Gibson and Thompson, 1994) were used to prepare a profile (Gribskov *et al.*, 1987) using the program PROFILEWEIGHT (Thompson *et al.*, 1994). The profile was calculated with the BLOSUM45 residue substitution matrix (Hennikoff and Hennikoff, 1992), weighting for sequence divergence, and excision of INDEL columns if gaps were present in >50% of the sequences. At INDEL sites, gap opening and extension penalties were set to 1/10th the penalty for other columns, except at the restricted 4th INDEL (Figure 5), where the gap extension penalty was not reduced.

Profile searches of protein databases were conducted with SearchWise, and output alignments were generated with PairWise (Ewan Birney, Oxford, unpublished; as outlined in Gibson *et al.*, 1994). These programs compare a protein sequence against all three reading frames of a DNA sequence in a single pass, allowing for frameshifts. The algorithm is an extension of the Smith–Waterman (1981) local similarity comparison, with three additional penalties: (i) for initiating a frame-jump; (ii) for extending the frameshift and (iii) for aligning onto a stop codon. The algorithm efficiently detects frameshifts due to sequencing errors and can traverse introns. The penalty values were usually set to: gap opening, 900; gap extension, 100; frame opening, 1000; frame extension, 2; stop codon, 500.

Database searches with single sequences were conducted against SWISS-PROT v. 30 (Bairoch and Boeckmann, 1993) using the EMBL BLITZ network service (Rice *et al.*, 1993) or against DNA databases using GCG TFASTA (Devereux *et al.*, 1984; Pearson and Lipman, 1988). The BLOSUM 62 matrix with gap penalty set to 8 was used in all BLITZ searches.

Previously known dsRBD-containing sequences are recorded in Gibson and Thompson (1994), Kao *et al.* (1994) and Polson and Bass (1994). Six newly identified dsRBDs are: bases 3475–4266 within ORF2 of CEZK632 (Ac# Z22181); complemented 2940–3162 adjacent to ORF4 in CET20H4 (Ac# U00037); 8402–8669, part of ORF YHR187W in SCH9998 (Ac# U00030); complemented 126–348 in CEK019CYR (Ac# D32592); 45–243 in CEK023A3F (Ac# D35594); 600–819 (part of an unidentified reading frame encoding an RNase III homologue) in MO15224 (Ac# U15224). Three overlapping ESTs together encoding an RNase III homologue of bacterial origin are HSA64D051 (Ac# 216130), HSA52B011 (Ac# 215957) and HSAADZFX (Ac# Z21338). The new sequences were aligned by profile into the existing alignment. These were then inspected by eye for optimal alignment of the structurally important core-forming positions and revised if necessary.

Acknowledgements

We thank Mark Bycroft, Daniel St. Johnston and their colleagues for communicating their results in advance. We also thank Pekka Lappalainen, José Castresana and Gilles Travé for useful advice and the Matthias Mann group for mass spectroscopy.

References

Bairoch,A. and Boeckmann,B. (1993) The SWISS-PROT protein sequence databank. *Nucleic Acids Res.*, **21**, 3093–3096.
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.
Belasco,J.G. and Higgins,C.F. (1988) Mechanism of mRNA decay in bacteria: a perspective. *Gene*, **72**, 15–23.
Berendsen,H.J.C., Postma,J.P.M., van Gunsteren,W.F., DiNola,A. and Haak,J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
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.
Brooks,B.R., Bruccoleri,R.E., Olafson,B.D., States,D.J., Swaminathan,S. and Karplus,M. (1983) CHARMM. A program for macromolecular energy, minimisation and dynamics calculations. *J. Comp. Chem.*, **4**, 187–217.
Brünger,A.T. (1989) X-PLOR. *A System for X-ray Crystallography*. 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., Grünert,S., Murzin,A.G., Proctor,M. and St. Johnston,D. (1995) NMR solution structure of a double-stranded RNA-binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.*, **14**, 3563–3571.
Castiglione Morelli,M.A., Stier,G., Gibson,T., Joseph,C., Musco,G., Pastore,A. and Travé,G. (1995) The KH module has an αβ fold. *FEBS Lett.*, **358**, 193–198.
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.
Court,D (1993) RNA processing and degradation by RNase III. In Belasco,J.G. and Brawerman,G. (eds), *Control of mRNA Stability*, Academic Press Inc., NY, pp. 71–116.
Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the Vax. *Nucleic Acids Res.*, **12**, 387–395.
Folkers,P.J.M., Nilges,M., Folmer,R.H.A., Konings,R.N.H. and Hilbers, C.W. (1994) The solution structure of the Tyr41-His mutant of the single-stranded DNA binding protein encoded by Gene V of the filamentous bacteriophage M13. *J. Mol. Biol.*, **236**, 229–246.
Ghisolfi,L., Kharrat,A., Joseph,G., Amalric,F. and Erard,M. (1992) Concerted activities of the RNA-recognition and the glycine-rich C-terminal domains of nucleolin are required for efficient complex formation with pre-ribosomal RNA. *Eur. J. Biochem.*, **209**, 541–548.
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.*, **3**, 361–366.
Gibson,T.J., Hyvönen,M., Birney,E., Musacchio,A. and Saraste,M. (1994) PH domain: the first anniversary. *Trends Biochem. Sci.*, **19**, 349–353.
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.
Gribskov,M., McLachlan,A.D. and Eisenberg,D. (1987) Profile analysis: detection of distantly related proteins. *Proc. Natl Acad. Sci. USA*, **84**, 4355–4358.
Griesinger,C., Otting,G., Wuethrich,K. and Ernst,R.R. (1988) Clean TOCSY for 1H spin system identification in macromolecules. *J. Am. Chem. Soc.*, **110**, 7870–7872.
Guenthert,P., Berndt,K.D. and Wuethrich,K. (1993) The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. *J. Biomol. NMR*, **3**, 601–606.
Hajnsdorf,E., Carpousis,A.J. and Régnier,P. (1994) Nucleolytic inactivation and degradation of the RNase III processed *pnp* message encoding poylynucleotide phosphorylase of *Escherichia coli*. *J. Mol. Biol.*, **239**, 439–454.
Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
Hol,W.G.J., Halie,L.M. and Sander,C. (1981) Dipoles of the α-helix and β-sheet: their role in protein folding. *Nature*, **294**, 532–536.
Holak,T.A., Nilges,M. and Oschkinat,H. (1989) Improved strategies for the determination of protein structures from NMR data. The solution structure of acyl carrier protein. *FEBS Lett.*, **242**, 649–654.
Iino,Y., Sugimoto,A. and Yamamoto,M. (1991) *S. pombe pac1+*, whose over-expression inhibits sexual development, encodes a ribonuclease III-like RNase. *EMBO J.*, **10**, 221–226.
Kadkhodaei,M., Hwang,T.-L., Tang,J. and Shaka,A.J. (1993) A simple windowless mixing sequence to suppress cross-relaxation in TOCSY experiments. *J. Magn. Res.*, **105**, 104–107.
Kao,P.N., Chen,L., Brock,G., Ng,J., Kenny,J., Smith,A.J. and Corthesy,B. (1994) Cloning and expression of cyclosporin A and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.*, **269**, 20691–20699.
Kiledjian,M. and Dreyfuss,G. (1992) Primary structure and binding sequence of the hnRNP U protein: binding RNA through RGG box. *EMBO J.*, **11**, 2655–2644.
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. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
Lee,C.-G. and Hurwitz,J. (1993) Human RNA helicase A is homologous to the maleless protein of *Drosophila*. *J. Biol. Chem.*, **268**, 16822–16830.
Lee,W., Harvey,T.S., Yin,Y., Yau,P., Litchfield,D. and Arrowsmith,C.H. (1994) Solution structure of the tetrameric minimum transforming domain of P53. *Struct. Biol.*, **1**, 877–890.
Levy,R.M., Bassolino,D.A., Kitchen,D.B. and Pardi,A. (1989) Solution structures of proteins from NMR data and modeling: alternative folds for neutrophil peptide 5. *Biochemistry*, **28**, 9361–9372.
Macias,M.J., Musacchio,A., Ponstingl,H., Nilges,M., Saraste,M. and Oschkinat,H. (1994) Structure of the pleckstrin homology domain from β-spectrin. *Nature*, **369**, 6785–677.
Marion,D. and Wüthrich,K. (1983) Application of phase-sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^{1}H-^{1}H spin-spin coupling constants in proteins. *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
Mattaj,I.W. (1993) RNA recognition: a family matter? *Cell*, **73**, 837–840.
McMillan,A.J.N., Carpick,B.W., Hollis,B., Toone,W.M., Zamanian-Daryoush,M. and Williams,B.R.G. (1995) Mutational analysis of the double-stranded (dsRNA) binding domain of the dsRNA-activated protein kinase PKR. *J. Biol. Chem.*, **270**, 2601–2606.
Meadows,R.P., Olejniczak,E.T. and Fesik,S.W. (1994) A computer-based protocol for semi-automated assignments and 3D structure determination of proteins. *J. Biomol. NMR*, **4**, 79–96.
Miller,J., McLachlan,A.D. and Klug,A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.*, **4**, 1609–1614.
Nilges,M. (1993) A calculation strategy for the structure determination of symmetric dimers by 1H NMR. *Proteins*, **17**, 297–309.
Nilges,M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.*, **245**, 645–660.
Nilges,M., Gronenborn,A.M., Brünger,A.T. and Clore,G.M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints: application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.*, **2**, 27–38.
Nilges,M., Kuszewski,J. and Brünger,A.T. (1991) Sampling properties of simulated annealing and distance geometry. In Hoch,J.C., Poulsen,F.M. and Redfield,C. (eds), *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy*. Plenum Press, New York, pp. 451–455.
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.
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 spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.
Pastore,A., Saudek,V., Ramponi,G. and Williams,R.J.P. (1992) Three-dimensional structure of acylphosphatase: refinement and structure analysis. *J. Mol. Biol.*, **224**, 427–440.
Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR spectra of aqueous solutions. *J. Biomol. NMR*, **2**, 661–664.

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.

Rice,C.M., Fuchs,R., Higgins,D.G., Stoehr,P.J. and Cameron,G.N. (1993) The EMBL data library. *Nucleic Acids Res.*, **21**, 2967–2971.

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.

Saudek,V., Pastore,A., Castiglione Morelli,M.A., Frank,R., Gausepohl,H. and Gibson,T. (1991) The solution structure of a leucine-zipper motif peptide. *Protein Eng.*, **4**, 519–529.

Schweisguth,D.C., Chelladuri,B.S., Nicholson,A.W. and Moore,P.B. (1994) Structural characterisation 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**, 1163–1198.

Sippl,M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins*, **17**, 355–362.

Smith,T.F. and Waterman,M.S. (1981) Comparison of bisosequences. *Adv. Appl. Math.*, **2**, 482–489.

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.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) Improved sensitivity of profile searches through the use of sequence weights and gap excision. *Comput. Appl. Biosci.*, **10**, 19–29.

van der Oost,J., Lappalainen,P., Mussacchio,A., Warne,A., Lemieux,L., Rumbley,J., Gennis,R.B., Aasa,R., Pascher,T., Malmstrom,B.G. and Saraste,M. (1992) Restoration of a lost metal-binding site. Construction of two different copper sites into a sub-unit of the *E.coli* cytochrome O quinol oxidase complex. *EMBO J.*, **11**, 3209–3217.

Vriend,G. and Sander,C. (1993) Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.*, **26**, 47–60.

Weber,P.L., Morrison,R. and Hare,D. (1988) Determining stereo-specific 1H NMR assignments from distance geometry calculations. *J. Mol. Biol.*, **204**, 483–487.

Wilson,R. *et al.* (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.elegans*. *Nature*, **368**, 32–38.

Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. Wiley Press.

Received on March 7, 1995; revised on April 26, 1995