Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold

Maria R.Conte^{1,2}, Tim Grüne^{3,4},
Jamie Ghuman^{1,3}, Geoff Kelly¹,
Anastasia Ladas³, Stephen Matthews^{1,5} and
Stephen Curry^{3,5}

¹Department of Biochemistry, Imperial College of Science,
Technology and Medicine, Exhibition Road, London SW7 2AY and

³Biophysics Section, Blackett Laboratory, Imperial College of Science,
Technology and Medicine, Prince Consort Road, London SW7 2BW,
UK

²Present address: Biophysics Laboratories, School of Biological
Sciences, St Michael's Building, University of Portsmouth, White
Swan Road, Portsmouth PO1 2DT, UK

⁴Present address: EMBL Grenoble Outstation, 6 rue Jules Horowitz,
BP156, F-38042 Grenoble, Cedex 9, France

⁵Corresponding authors

e-mail: s.curry@ic.ac.uk or s.j.matthews@ic.ac.uk

Polypyrimidine tract binding protein (PTB), an RNA binding protein containing four RNA recognition motifs (RRMs), is involved in both pre-mRNA splicing and translation initiation directed by picornaviral internal ribosome entry sites. Sequence comparisons previously indicated that PTB is a non-canonical RRM protein. The solution structure of a PTB fragment containing RRM3 and 4 shows that the protein consists of two domains connected by a long, flexible linker. The two domains tumble independently in solution, having no fixed relative orientation. In addition to the $\beta\alpha\beta\alpha\beta$ topology, which is characteristic of RRM domains, the C-terminal extension of PTB RRM-3 incorporates an unanticipated fifth β -strand, which extends the RNA binding surface. The long, disordered polypeptide connecting β 4 and β 5 in RRM-3 is poised above the RNA binding surface and is likely to contribute to RNA recognition. Mutational analyses show that both RRM-3 and RRM-4 contribute to RNA binding specificity and that, despite its unusual sequence, PTB binds RNA in a manner akin to that of other RRM proteins.

Keywords: polypyrimidine tract binding protein/RNA binding/RNA recognition motif/splicing regulation/translation initiation

Introduction

The polypyrimidine tract binding protein (PTB), also known as hnRNP-I, was originally identified as a nuclear-localized splicing factor that binds specifically to the polypyrimidine tract within pre-mRNA introns (Gil *et al.*, 1991; Patton *et al.*, 1991; Ghetti *et al.*, 1992; for review, see Valcárcel and Gebauer, 1997). PTB is a homodimer (Pérez *et al.*, 1997a) and appears to function mainly as a negative regulator of pre-mRNA splicing by binding to

high-affinity sequence motifs in the introns upstream and downstream of regulated exons and repressing their incorporation into mature RNA (reviewed in Valcárcel and Gebauer, 1997). PTB has been shown to be involved in the regulation of splicing in a number of different genes, including α -tropomyosin (Pérez *et al.*, 1997a; Gooding *et al.*, 1998), α -actinin (Southby *et al.*, 1999), *c-src* (Chan and Black, 1997) and the γ 2 subunit of the GABA_A receptor (Ashiya and Grabowski, 1997). Recent studies have shown that PTB may have additional roles in the regulation of RNA polyadenylation (Moreira *et al.*, 1998; Lou *et al.*, 1999) and RNA localization (Cote *et al.*, 1999). The specific intron RNA sequence motifs that bind PTB are only partially characterized; in general, short RNA stretches such as UCUUC or UUCU/C within a longer pyrimidine segment represent the consensus binding site for the protein (Ashiya and Grabowski, 1997; Pérez *et al.*, 1997b).

Intriguingly, the protein also interacts with picornaviral internal ribosomal entry site (IRES) elements. These are large RNA segments (~450 bases) with extensive secondary structure found within the lengthy 5'-untranslated region of the picornavirus positive-sense single-stranded RNA genome (reviewed in Jackson, 1996). Picornaviral RNA lacks a 5'-cap structure, which is normally used by eukaryotic messages to recruit the initiation factors that enable the small ribosomal subunit to locate the initiation codon by scanning. Instead, the IRES structure permits the productive association of the viral message with the small ribosomal subunit by a cap-independent mechanism that largely avoids scanning (Jackson, 1996). IRES-dependent translation initiation requires a subset of canonical eukaryotic translation initiation factors (eIFs), but additional proteins are also needed (Belsham and Sonenberg, 1996). PTB was identified as one of these additional factors since it can stimulate the IRES-dependent translation initiation of most picornaviruses (Jang and Wimmer, 1990; Kaminski *et al.*, 1995; Kaminski and Jackson, 1998; Hunt and Jackson, 1999), and multiple PTB binding sites have now been mapped within the IRES elements of a number of different picornaviruses (Jang and Wimmer, 1990; Luz and Beck, 1991; Hellen *et al.*, 1994; Kolupaeva *et al.*, 1996).

The mechanisms of action for PTB in both splicing regulation and stimulation of translation initiation remain obscure, although both functions clearly require binding of PTB to RNA. PTB belongs to a family of RNA binding proteins characterized by possession of at least one ribonucleoprotein consensus sequence (RNP-CS) RNA binding domain. This RNA recognition motif (RRM) is ~90 amino acids long and contains conserved six- and eight-residue motifs called RNP-2 and RNP-1, respectively. The RRM domain folds into a $\beta\alpha\beta\alpha\beta$ structure that presents a four-stranded β -sheet for RNA binding

	RNP - 2	RNP - 1
PTB RRM1	IHIRKL	KNQAFIEM
PTB RRM2	IIVENL	QFQALLQY
PTB RRM3	LLVSNL	KENALVQM
PTB RRM4	LHLSNL	RKMALIQM
Family consensus	LFVGNL IYIKL	K ₂ FG ₂ FVX ₂ F ₂ R ₂ G ₂ YA ₂
	* * *	* * *

Fig. 1. Sequence comparison of RNP-2 and RNP-1 motifs for PTB with the RRM family consensus (Kenan *et al.*, 1991). The PTB motifs contain quite unusual amino acids relative to the consensus (indicated by shading). Positions of conserved hydrophobic positions are indicated by an asterisk.

(Nagai *et al.*, 1990). Each monomer of PTB is predicted to have four RRM domains but these are invariably non-canonical and contain unusual features in the sequences of the 'conserved' motifs (Kenan *et al.*, 1991). In particular, there is a general absence of aromatic residues at key positions in RNP-1 and RNP-2, which, in other RRM domains, have been shown to be responsible for non-specific contacts with the RNA (Oubridge *et al.*, 1994; Allain *et al.*, 1997; Price *et al.*, 1998; Deo *et al.*, 1999; Handa *et al.*, 1999) (Figure 1). Moreover, at position 2 in RNP-1 a conserved glycine residue is substituted by much larger side chains in all four RRM domains of PTB (Figure 1).

The strongest determinants for RNA binding have been identified by deletion mutagenesis to reside in the two C-terminal RRM domains of PTB (Kaminski *et al.*, 1995; Pérez *et al.*, 1997a; Oh *et al.*, 1998). To gain a clearer understanding of the mode of action of the protein in splicing and translation initiation and of the specificity determinants of its unusual RRM domains, we have undertaken NMR-based structural studies of a C-terminal fragment that contains these two domains, designated PTB-34. The three-dimensional solution structure of PTB-34 reveals an unanticipated C-terminal extension of the β -sheet RNA binding surface of RRM-3 by an additional strand and shows that the two linked RRM-3 and RRM-4 domains tumble independently in solution, having no fixed relative orientation. In addition, site-directed mutagenesis and RNA binding experiments have been used to map the interaction of PTB with IRES RNA.

Results and discussion

Structure determination

The expression and sequence-specific ^1H N, ^{15}N , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and $^{13}\text{C}'$ assignments of PTB-34 (residues 335–531) were performed as described previously (Conte *et al.*, 1999). The side chain resonance assignments were mainly achieved using HCCH-TOCSY experiments (Bax *et al.*, 1990). The distance restraints used in the structure calculation were obtained from 3D $^1\text{H}/^{15}\text{N}$ and $^1\text{H}/^{13}\text{C}$ -edited NOESY-HSQC experiments (Fesik and Zuiderweg, 1988; Norwood *et al.*, 1990), as well as from $^{15}\text{N}/^{13}\text{C}$ and $^{13}\text{C}/^{13}\text{C}$ -edited HMQC-NOESY-HSQC experiments (Vuister *et al.*, 1993). The dihedral angle restraints were determined using coupling constants and the program TALOS (Cornilescu *et al.*, 1999). The structures were

Fig. 2. Superposition of 20 refined structures for PTB RRM-3 and PTB RRM-4 domains. The backbone traces for residues 334–433 in RRM-3 and 449–531 in RRM-4 are shown. The N- and C-termini are indicated for each domain and β -strands are numbered.

calculated with the program X-PLOR using a dynamic simulated annealing protocol (Nilges *et al.*, 1988; Brünger, 1993) on the basis of 702 total nuclear Overhauser effect (NOE) conformational sensitive distance restraints, 120 H-bond distance restraints and 223 dihedral angle restraints. The relatively low number of long-range NOEs is caused by severe spectral overlap from the unstructured and extensive linker region. Despite this, the backbone atoms of the secondary structure elements are well defined.

Figure 2 shows a final family of 20 structures for RRM-3 and RRM-4. The two domains were superimposed separately, since in solution they tumble independently with respect to each other (see below). The overall root mean square deviation (r.m.s.d.) between the family and

Table I. Summary of structural statistics for PTB-34

NMR restraints	RRM-3 ^a	RRM-4 ^a
Total distance restraints	349	343
Intraresidue	5	1
Interresidue	344	342
short-medium range (residue <i>i</i> to <i>i</i> + <i>j</i> , <i>j</i> = 2–4)	221	237
long range (residue <i>i</i> to <i>i</i> + <i>j</i> , <i>j</i> > 4)	123	115
Hydrogen bond	62	58
Total dihedral angle restraints	115	108
ϕ	58	54
ψ	57	54
Restraints violations		
Distance restraint violation > 0.2 Å	none	none
Dihedral restraint violation > 4.0°	none	none
Average r.m.s.d. (Å) among the 20 refined structures		
Backbone of secondary structure elements ^b	0.62	0.60
All heavy atoms of secondary structure elements	1.30	1.30
Ramachandran statistics of 20 structures ^c		
Percentage residues in		
most favoured regions	72.8%	77.5%
additional allowed regions	16.6%	16.8%
generously allowed regions	8.1%	4.8%
disallowed regions	2.4%	0.9%

^aPTB RRM-3, residues 338–430; PTB RRM-4, residues 449–531.

^bSecondary structure elements: RRM-3 (338–343, 352–360, 364–369, 375–380, 386–395, 404–408, 426–430); RRM-4 (454–459, 467–476, 483–487, 494–498, 504–513, 523–528).

^cAnalysis of the Ramachandran plot was performed by PROCHECK-NMR (Laskowski *et al.*, 1996).

the mean co-ordinate position is 0.62 and 0.60 Å for the backbone atoms in regions of secondary structure for RRM-3 and RRM-4, respectively. None of the 20 structures has any distance violation >0.2 Å or dihedral angle violation >4.0°. The structural calculation statistics are shown in Table I.

Tertiary structure of PTB-34

PTB-34 has a tandem domain structure in which the RRM domains are tethered by a linker peptide of 25 residues in length. While the conformations of the secondary structural elements within RRM domains 3 and 4 are well defined, the inter-domain linker region of PTB-34 was mostly unassigned (Conte *et al.*, 1999). Attenuation of resonances for most of the 25 residues forming the linker region could be caused by internal mobility on a slow or intermediate timescale. In addition, we found no evidence of contact between the two RRM domains, which appear to tumble independently in solution.

Although RRM-4 contains the $\beta\alpha\beta\alpha\beta$ topology expected for an RRM domain, the C-terminal extension of RRM-3 incorporates an additional fifth β -strand (Figures 2–4), a feature that is unprecedented for this family of proteins. The inter-residue NOE network reveals that this strand—designated β_5 (residues 426–430)—lies in an antiparallel fashion to β_2 on one side of the β -sheet and is connected to β_4 on the opposite side of the domain by an extended loop (Figure 3). This loop (residues 409–425) appears to be very mobile and is unstructured apart from a single helical turn close to the start of β_5 . In our construct the linker makes no observable contact with the rest of the protein; however, it is positioned above the RNA binding surface and may play a role in RNA recognition (see below).

Helices α_1 and α_2 of RRM-3 are very well defined and pack against the β -sheet on the opposite side to the RNA

binding surface, creating an extensive hydrophobic core (Figures 2 and 3). Several loop residues participate in this hydrophobic interaction, in particular loops α_1 – β_2 and β_3 – α_2 , which appear to be very well defined. To characterize the dynamic aspects of PTB-34, the $\{^1\text{H}\}$ – ^{15}N NOE values were determined (data not shown). The observation of significantly smaller $\{^1\text{H}\}$ – ^{15}N NOE values (0.1–0.6) in the extended β_4 – β_5 linker indicates that this region undergoes fast internal motion in the pico- to nanosecond timescale, so the poor convergence of the β_4 – β_5 loop and the lack of observable NOEs result from intrinsic flexibility. Smaller than average NOE values are also observed for some residues in loops β_1 – α_1 and α_2 – β_4 , which are less well defined than the core of RRM-3. The β_2 – β_3 loop also appears to be disordered in solution. Several of the resonances in this loop were very weak (L370, N372, K373) or unassigned (F371), preventing measurement of their $\{^1\text{H}\}$ – ^{15}N NOE values; however, attenuation of these signals could be caused by internal mobility of these regions on a slow or intermediate timescale (μs – ms).

The structure of RRM-4 is very similar to that of canonical RRM motifs (Figures 2–4) and contains just the standard four-strand β -sheet within a $\beta\alpha\beta\alpha\beta$ topology—there is no additional fifth β -strand. The inner surface of the four-stranded β -sheet forms an extensive hydrophobic core with the inner surface of the two α -helices. As observed for RRM-3 the secondary structure elements are very well defined, as is loop β_3 – α_2 . In contrast, loops β_1 – α_1 , α_1 – β_2 and α_2 – β_4 of RRM-4 have rather small $\{^1\text{H}\}$ – ^{15}N NOE values and appear to be more disordered in solution than the corresponding regions in RRM-3. As was found for RRM-3, signals in the β_2 – β_3 loop of RRM-4 were weak (K489, D490) or missing (Q488), suggesting an intrinsic flexibility for this loop in the micro- to millisecond timescale.

Fig. 3. Ribbon diagrams for PTB-34 and Sex-lethal. (A) Comparison of tandem domain structures of PTB and Sex-lethal. The relative orientation of the two domains shown for PTB is arbitrary, as is the structure of the inter-domain linker. The structure of Sex-lethal was solved crystallographically in the presence of bound RNA (Handa *et al.*, 1999), which has been omitted from the figure. (B) Comparison of RRM-3 and RRM-4 domains from PTB with RRM-1 of Sex-lethal. PTB RRM-3 contains an additional strand ($\beta 5$) on one side of the RNA binding surface. Note that the conformation shown for the $\beta 4$ – $\beta 5$ loop is only one of many conformations that are consistent with the data

Fig. 4. Structure-based sequence alignment of RRM-3 and RRM-4 from PTB. Secondary structure elements are depicted above the sequences. Identical residues are shaded and the RNP-1 and RNP-2 motifs are boxed.

Analysis of RNA binding

The structure of PTB-34 reveals that the RNA binding domains, despite their unusual RNP-1 and RNP-2 motifs, have much in common with other RRM proteins. To examine the role of the unusual motif sequences in RNA binding and to probe for additional determinants of binding specificity, mutagenesis of RRM-3 and RRM-4 was performed. Filter binding assays were used to measure the affinity of the various PTB fragments and mutants (Figure 5A) for radiolabelled RNA transcripts containing either the full-length FMDV or EMCV IRES, or domain 1 from the EMCV IRES (Figure 5B). Domain 1 contains nucleotides 260–444 of the EMCV IRES and incorporates stem-loops D–H (Kaminski *et al.*, 1995); stem-loop H is considered to have the highest affinity PTB binding site (Jang and Wimmer, 1990; Witherall *et al.*, 1993).

The binding affinities of hisPTB for domain 1 and the full-length EMCV IRES are comparable (Table II), confirming that domain 1 contains the most important PTB binding site. The full-length hisPTB protein binds IRES RNA with dissociation constants in the range 0.03–0.06 μ M, which is consistent with previous data (Witherall *et al.*, 1993). Truncation of hisPTB to hisPTB-34a, which includes just the third and fourth RRM domains, reduced binding to EMCV IRES targets

Fig. 5. Overview of protein and RNA constructs used in this study. (A) Schematic depiction of PTB constructs; RRM domains are indicated by shading. (B) Schematic diagram of the EMCV IRES; domain 1 is boxed.

~7-fold (Figure 6), although the loss of affinity was somewhat less for FMDV IRES (Table II). This may be due in part to loss of the dimerization function of the N-terminal domain of PTB (Pérez *et al.*, 1997a), but it may also suggest that the N-terminal RRMs contribute more to RNA binding than was previously thought. The single-domain constructs hisPTB-3a, hisPTB-3b and hisPTB-4a had even lower affinities for EMCV IRES domain 1, although the third RRM binds significantly more tightly

Table II. Dissociation constants (μM) for binding of deletion mutants of PTB to IRES RNA

Protein	Residue range	K_D (μM)		
		EMCV domain I	EMCV	FMDV
hisPTB	1–531	0.026	0.038	0.061
hisPTB-34	324–531	0.18	0.30	0.15
hisPTB-34a	335–531	0.23	0.21	0.15
hisPTB-3a	324–422	0.82	–	–
hisPTB-3b	324–433	0.62	–	–
HisPTB-4a	442–531	3.20	–	–

Fig. 6. Binding curves for the interaction of full-length and truncated PTB constructs with EMCV IRES domain 1 RNA. Experimental details are given in Materials and methods. The data for each protein have been normalized. Analysis of such curves was used to determine dissociation constants for all the PTB constructs and mutants used in this study (Tables II and III).

than the fourth. This is somewhat at odds with UV cross-linking experiments, which found no RNA binding activity associated with single-domain constructs of PTB (Kaminski *et al.*, 1995; Pérez *et al.*, 1997a), although the discrepancy is probably attributable to the fact that constructs used in UV experiments contained neither the fourth nor the fifth β -strands of the domain.

Constructs hisPTB-3a and hisPTB-3b (Figure 5A) differ in that only the latter fragment contains the $\beta 5$ strand but the proteins have the same affinity for the EMCV IRES, suggesting that $\beta 5$ and the $\beta 4$ – $\beta 5$ linker do not contribute to RNA binding in the context of these single-domain fragments. Nonetheless, these features may well play a role in RNA recognition in the presence of the fourth domain or in the context of the intact protein.

To probe the contribution to RNA recognition of individual amino acids on the binding surface, single and double amino acid substitutions were introduced into hisPTB-34a. Binding experiments with these mutants gave a number of surprising results, summarized in Table III. Several of the mutations within the core RNP-1 and RNP-2 motifs, including N343A, Q380A, N460A and Q497A, have almost no effect on binding. These residues lie towards the top and bottom edges of the RNA binding surface (Figure 7) and do not always contact the nucleic acid ligand (Ding *et al.*, 1999; Handa *et al.*, 1999). The mutation R491S at the start of the RNP-1 motif in RRM-4 has little, if any, effect on RNA binding, while the more drastic double mutation K373E/K374E at the equivalent position in RRM-3 did reduce the RNA affinity 4-fold. These findings stand in marked contrast to results for other RRM proteins such as U1A or Sex-lethal, which lose all RNA binding affinity if the basic residue conserved at the start of RNP-1 is mutated (Nagai *et al.*, 1990; Lee *et al.*, 1997).

The unusual RNP-2 and RNP-1 sequences in PTB mean that, unlike most other RRM proteins, the RNA binding surface is almost completely devoid of aromatic residues that might stack with the RNA bases. Nevertheless, the unusual residues that occur within the RNP-1 and RNP-2 motifs of PTB (Figure 1) do contribute to RNA binding. For example, mutations L378S and H457A, which are in positions normally occupied by conserved aromatic amino acids in the vast majority of RRM proteins, both significantly reduce IRES binding. These data confirm that, despite possessing unusual RNP-1 and RNP-2 motifs, the

Table III. Dissociation constants (μM) for binding of mutants of hisPTB-34a to EMCV IRES RNA

RRM-3 mutation	K_D (μM)		RRM-4 mutation	K_D (μM)		Structural location
	EMCV domain I	EMCV		EMCV domain I	EMCV	
Wild type	0.21	0.23				
L340S	0.28	0.39	H457A	0.52	0.46	$\beta 1$ (RNP-2)
N343A	0.24	0.19	N460A	0.18	0.16	$\beta 1$ (RNP-2)
K368E	0.56	1.01	K485E ^a	0.37	0.67	$\beta 2$
L378S	0.46	0.55	–	–	–	$\beta 3$ (RNP-1)
K373E/K374E	0.70	0.78	R491S	0.35	0.19	$\beta 3$ (RNP-1)
Q380A	0.17	–	Q497A	0.16	–	$\beta 3$ (RNP-1)
O380A/Q497A	–	0.19				$\beta 3$ (RNP-1)
K439E/K440E	0.55	0.73				linker
K439S/K440S	0.74	0.59				linker

^aAlso contains the substitution Q365R as a result of a PCR error. However, the solvent-exposed side chain of Q365, which lies at the beginning of $\beta 2$, points away from the RNA binding surface and is unlikely to affect RNA binding.

Fig. 7. Mapping of sites of mutation onto the structure of PTB-34. The RNP-1 and RNP-2 motifs are indicated on the protein structure by dark shading. Large spheres (with underlined labels), positions of mutations that caused >1.5-fold decrease in binding; small spheres, positions of mutations that caused <1.5-fold decrease in binding (see Table II).

β -sheet surface on RRM-3 and RRM-4 forms an RNA binding site. Mutations K368E and K485E, which lie in equivalent positions in $\beta 2$ of RRM-3 and RRM-4, respectively, both substantially decrease RNA binding. This may be due to the loss of direct contacts with the RNA ligand. However, the structure shows that the K368 side chain can interact with E424 within the helical turn at the C-terminal end of the $\beta 4$ – $\beta 5$ loop, so the impact of the substitution of this residue on RNA binding may be due to induced conformational changes in the protein. On balance, substitutions of amino acids in RRM-4 are as deleterious as substitutions in RRM-3, confirming that RRM-4 is important for RNA binding (Table III; Figure 7). The linker connecting the two RRM domains also contributes to RNA binding, since the double mutation K439E/K440E significantly reduces the affinity. This is consistent with crystallographic structural analyses of Sex-lethal and poly(A)-binding protein (PABP) complexed with their respective RNA targets (Deo *et al.*, 1999; Handa *et al.*, 1999), which revealed specific contacts between the inter-domain linker and the bound RNA, although the topology of the linker in PTB is likely to be very different because of the presence of the $\beta 5$ strand, which does not occur in Sex-lethal or PABP.

The observation that the effects of amino acid substitutions on the binding of hisPTB-34a to IRES RNA are

usually rather modest contrasts sharply with studies on other RRM proteins, such as U1A and Sex-lethal (Nagai *et al.*, 1990; Lee *et al.*, 1997), which found more severe reductions in binding associated with single amino acid substitutions. It may be that the PTB–IRES interaction is less specific than those of other RRM proteins or that PTB relies on dimeric, multi-domain contacts for specific recognition.

We also assessed the affinity of PTB for an intron target, an RNA oligomer containing 15 bases from the sequence of the polypyrimidine tract of the downstream regulatory element of exon 3 from α -tropomyosin (Pérez *et al.*, 1997b). This oligomer (Dpy) has the sequence UUUCUCCUCUUCUUU and is bound by hisPTB and hisPTB-34a with dissociation constants of 0.17 and 1.3 μ M, respectively. Using a different assay, the affinity of PTB for a slightly longer DPy oligomer was previously reported to be 0.003 μ M (Pérez *et al.*, 1997b) but at present we have no satisfactory explanation for the discrepancy. Our finding that PTB binds much more tightly to IRES targets may be due to the presence of more than one binding site on the significantly larger viral RNA.

Comparison with other modular RRM proteins

Structural similarities of PTB-34 to other RRM proteins were assessed using the program DALI (Holm and Sander, 1993). Apart from the novel fifth strand revealed in PTB RRM-3 and despite the presence of unusual residues in the RNP-1 and RNP-2 motifs, the remainder of domain 3 and the whole of RRM-4 are very similar to the classical RRM fold (Figure 3). The closest structural homologues of PTB RRM-3 were found to be Sex-lethal RRM-1 (Handa *et al.*, 1999; r.m.s.d. of 3.7 Å over 82 equivalent C_{α} atoms, with a Z score of 7.0) and U2AF⁶⁵ RRM-2 (Ito *et al.*, 1999; 2.6 Å over 75 sites, Z = 6.8). For PTB RRM-4 the closest homologues were again Sex-lethal RRM-1 (2.8 Å over 67 sites, Z = 6.1) and U2AF⁶⁵ RRM-2 (3.3 Å over 72 sites, Z = 6.0). Curiously, both Sex-lethal and U2AF⁶⁵, like PTB, recognize polypyrimidine sequences.

PTB RRM-3 differs from other members of the RRM family in that it possesses an additional, fifth β -strand and a long flexible loop between $\beta 4$ and $\beta 5$. A survey of PTB homologues in the sequence database indicates that the $\beta 5$ strand in RRM-3 is conserved in mammals, *Caenorhabditis elegans* and plants (Figure 8). Homologous proteins include NPTB (a recently identified neurone-specific variant of PTB; 74% identity; Doug Black, personal communication), ROD1 (a mammalian protein implicated in differentiation control; 74% identity; Yamamoto *et al.*, 1999) and hnRNP-L (23% identity), which has already been noted as similar to PTB (Ghetti *et al.*, 1992). In all cases the homologues have four RRM domains and, as in PTB, the predicted fifth strand occurs only in the third domain. The presence of a fifth β -strand in more distantly related RRM proteins has not yet been detected; several are known that have longer inter-domain linkers than PTB-34 (Shamoo *et al.*, 1995) but the sequence determinants for this feature are not yet sufficiently well defined to permit an effective database search.

N- and C-terminal extensions to the core $\beta\alpha\beta\beta\alpha\beta$ structure of canonical RRM domains are important for RNA recognition (Query *et al.*, 1989; Scherly *et al.*, 1989;

		$\beta 4$	$\beta 5$	
PTB	Human	KPIRITLSKHQNVQ-LPREGQEDQG-LTKDYGNSPLHRFKKP	441	
PTB	Pig	KPVRIITLSKHQNVQ-LPREGQEDQG-LTKDYGNSPLHRFKKP	467	
PTB	Rat	KSVRIITLSKHQSVQ-LPREGQEDQG-LTKDYGSSPLHRFKKP	465	
NPTB	Human	KIIRVTLSKHQTVQ-LPREGLDDQG-LTKDFGNSPLHRFKKP	442	
PTB	Nematode	RLIRVAPSKHTNVQ-MPKEGQPDAG-LTRDYAHSTLHRFKKP	494	
PTB	Arabidopsis	CKRLLSYSRHTDLN--VKAFSD---KSRDYTLPLDLSLLVAQ	348	
ROD1	Human	KVLRATLSKHQAVQ-LPREGQEDQG-LTKDFGNSPLHRFKKP	431	
hnRNP-L	Human	QKLNVCVSKQPAIMPQGSYGLEDGSCSYKDFSESRNNRSTP	457	

Fig. 8. Alignment of PTB and homologous sequences in the region of $\beta 4$ and $\beta 5$ of RRM-3, indicating the conservation of $\beta 5$ in eukaryotes. DDBJ/EMBL/GenBank accession Nos for PTB and homologous sequences are: human (X62006), pig (X93009), rat (Q00438), *C.elegans* (Z36948), *Arabidopsis thaliana* (AF076924), NPTB (AF176085), ROD1 (AB023967) and hnRNP-L (X16135).

Lutz-Freyermuth *et al.*, 1990; Görlach *et al.*, 1994; Deo *et al.*, 1999; Handa *et al.*, 1999). The $\beta 4$ – $\beta 5$ loop of PTB RRM-3 occupies a similar position to the C-terminal helix of U1A protein, which is implicated in RNA binding (Scherly *et al.*, 1989; Lutz-Freyermuth *et al.*, 1990), although there are differences between the two structures. In the absence of bound RNA, the C-terminal helix in U1A packs in a well defined manner across the RNA binding β -sheet (Avis *et al.*, 1996) and prevents solvent exposure of hydrophobic residues. However, binding of RNA displaces the helix to allow the ligand to form a close network of interactions involving residues of the β -sheet and the C-terminal helix (Oubridge *et al.*, 1994; Allain *et al.*, 1997). In contrast, the loop $\beta 4$ – $\beta 5$ of PTB RRM-3 is highly disordered in solution, and no interaction was found with the rest of the molecule. Nevertheless, its position with respect to the central β -sheet suggests that it may function as determinant for RNA specificity, perhaps making specific hydrogen bonds with RNA bases that help to clamp the bound RNA in position. Consistent with this notion, the $\beta 4$ – $\beta 5$ region is rich in glutamine residues, which, along with asparagines, can make specific hydrogen bonds with RNA bases (Handa *et al.*, 1999). It is interesting to note that the C-terminus of PTB RRM-4 extends for five amino acids beyond the end of $\beta 4$ and lies close to the RNA binding surface, and thus may also contribute to RNA recognition by that domain. The loops joining the conserved secondary structure elements of RRM domains are the most variable stretches between different RRM proteins, differing in length, amino acid sequence, conformation and flexibility (Kenan *et al.*, 1991). Both crystallographic and NMR data indicate that the $\beta 2$ – $\beta 3$ loop, which can play a central role in RNA recognition, is rather flexible (Nagai *et al.*, 1990; Allain *et al.*, 1997; Shamoo *et al.*, 1997; Handa *et al.*, 1999; Nagata *et al.*, 1999); this flexibility appears to allow induced fit of the RNA ligand (Mittermaier *et al.*, 1999). The $\beta 2$ – $\beta 3$ loops in both RRM-3 and RRM-4 of PTB also appear flexible, and may well be important for RNA binding, although our mutagenesis results suggest that they make less of a contribution to RNA binding than has been observed in other cases.

PTB is predicted to be a modular RRM protein containing four independently folded domains (Ghetti *et al.*, 1992), two of which occur in PTB-34. To our knowledge, this is the first report of the solution structure

Fig. 9. The role of sequences flanking the core RRM domain in RNA recognition. (A) The structure of U2B'' complexed with its cognate RNA (Price *et al.*, 1998). For ease of comparison with the other structures only a portion of the stem-loop RNA (nucleotides 8–14) is shown and U2A' (which does not interact with the portion of RNA included in the figure) has been omitted. The C-terminal helix is crucial for high-affinity RNA binding. (B) Structure of RRM-2 of PABP bound to poly(A) (nucleotides 1–7; Deo *et al.*, 1999). The figure shows one RRM from the two-domain fragment that was present in the crystal. The N-terminal helical turn preceding $\beta 1$ interacts specifically with the bound RNA. (C) Structure of RRM-2 of Sex-lethal bound to its cognate RNA (nucleotides 1–11; Handa *et al.*, 1999). RRM-1, which lies on the right and also interacts with the RNA, has been omitted from the figure. (D) Superposition of the RNAs from U2B'', PABP and sex-lethal onto PTB RRM-3. The matrix required to superpose each protein domain on PTB RRM-3 was determined and applied to the RNA. In each case the conformation of the bound RNA clashes with the C-terminal end of the $\beta 4$ – $\beta 5$ loop in PTB RRM-3 (indicated by arrows).

of an RRM protein containing more than one domain, although the structures of double-domain fragments from three modular RRM proteins have been determined by X-ray crystallography. Crystallographic analyses of bound and free forms of such fragments have already suggested that in the absence of bound nucleic acid, the pair of RRM domains has some relative freedom of movement (Shamoo *et al.*, 1997; Xu *et al.*, 1997; Crowder *et al.*, 1999; Ding *et al.*, 1999; Handa *et al.*, 1999). RNA binding is coupled with stabilization of RRM inter-domain contacts and rigidification of the inter-domain linkers. In our present analysis of the NMR structure of PTB-34, the extent of flexibility of the inter-domain linker is particularly striking (Figure 3): there are no inter-domain contacts whatsoever and the two RRM domains appear to tumble independently, the long linker region making no fixed contacts with the rest of the protein, presumably due to disorder. As with other modular RRM proteins, a fixed conformation of PTB-34 may only occur upon binding, although it is worth noting that PTB has a 25-residue linker, which is significantly longer than those found in Sex-lethal (10), PABP (12) or hnRNP-A1 (16) (Shamoo *et al.*, 1995).

Superposition of the RRM–RNA or RRM–DNA crystal structures that have been published to date (U1A, U2Bⁿ, Sex-lethal, PABP, hnRNP-A1) shows that, despite the diversity of nucleic acid ligands, there are common features to the mode of binding. The centre of the RRM binding surface generally interacts with single-stranded RNA or DNA lying in a 5′ to 3′ direction across β 1– β 3– β 2 at an angle of $\sim 45^\circ$ to the direction of the β -strands. Moreover, the central pair of bound nucleotides, which most directly overlie the RNP-2 and RNP-1 motifs (on β 1 and β 3, respectively), occur in more or less the same position in all the structures, the planes of the bases being approximately parallel to the surface of the β -sheet. In marked contrast, beyond this central region there are variations in the conformations of the 3′ ends of the bound nucleic acid and these can be attributed primarily to the disposition of the N- or C-terminal extensions to the RRM domain (Figure 9). When these structures are superposed on PTB RRM-3 it is clear that the PTB domain may well bind the central pair of nucleotides of its cognate RNA in a manner very similar to that found for other RRM proteins, a proposal supported by our mutagenesis data (Figure 7). However, the superposition also shows that the 3′ end of the ligands observed for the other RRM proteins would clash with the C-terminal end of the β 4– β 5 loop (Figure 9). Thus, in PTB RRM-3 either the bound RNA adopts a different path to that observed previously or binding of RNA would affect the conformation adopted by the β 4– β 5 loop and possibly displace β 5 itself.

A more detailed understanding of PTB–RNA interactions will require structural analysis of protein–RNA complexes. Work is in progress to optimize RNA targets, which will permit a high resolution structural analysis of PTB-34 complexed with RNA.

Materials and methods

Plasmid construction and protein expression

A 198-residue C-terminal fragment of human PTB-1 (Gil *et al.*, 1991), which contains the third and fourth RRM domains, was subcloned by PCR and ligated into pET-15b (Novagen) using engineered *Nco*I and *Nde*I restriction sites (Conte *et al.*, 1999). The resulting fragment (residues 335–531; designated PTB-34) has the authentic C-terminus and an additional methionine at the N-terminus. The protein was over-expressed in BL21 (DE3) *Escherichia coli*. Cell pellets were lysed in 25 mM HEPES pH 7.25, 1.5 mM MgCl₂, 0.2 mM EDTA (buffer A) containing 0.2 M KCl. The protein was applied to a 20 ml cation exchanger column (POROS HS50; Perkin-Elmer Biosystems) and eluted with a linear 0.2–1.0 M KCl gradient in buffer A; PTB-34 eluted at 0.45 M KCl. The peak fractions were pooled, diluted to 0.2 M KCl with buffer A and loaded onto a 5 ml HiTrap Blue column (Cibacron Blue 3G; Amersham-Pharmacia Biotech). PTB-34 eluted as an almost homogenous protein at 1.3 M KCl. Peak fractions were pooled, dialysed into 20 mM sodium acetate pH 5.2 and concentrated to 1 mM.

A variety of histidine-tagged PTB constructs were prepared for use in binding assays. These constructs were generated by PCR to introduce 5′ *Bam*HI and 3′ *Hind*III restriction sites, which were used to incorporate the insert into expression vector pQE-9 (Qiagen), which adds an N-terminal histidine tag to the expressed protein. The details of the various constructs are summarized in Figure 5A. HisPTB-34 is simply a tagged version of the protein PTB-34 that was used for all the NMR studies. HisPTB-34a has a slightly longer N-terminus than hisPTB-34; this construct starts at residue 324 and incorporates an Arg to Ser mutation at residue 325 due to the introduction of a *Bam*HI site. The single-domain constructs for RRM-3 (hisPTB-3a and hisPTB-3b) have the same N-terminus as hisPTB-34a, since attempts to generate these constructs with the same N-terminus as hisPTB-34 resulted in very poor expression. HisPTB-4a was engineered with a Phe to Tyr substitution at residue 446 (just

upstream of the RRM domain) to give the protein a non-zero absorbance at 280 nm and thereby facilitate accurate quantitation of the concentration of the purified protein. Site-directed mutagenesis was performed in the context of the hisPTB-34a construct using overlap PCR. All mutations were confirmed by cDNA sequencing.

His-tagged PTB fragments and mutants were expressed in SG13009 *E. coli* (Qiagen). Cell pellets were lysed into 20 mM Tris pH 7.7, 250 mM NaCl (buffer B) and purified on TALON® metal affinity resin (Clontech) using the manufacturer's protocol. The purified protein was eluted with buffer B containing 100 mM imidazole. In all cases except one the protein was at least 95% pure as judged by Coomassie staining of SDS polyacrylamide gels. The mutant hisPTB-34a(L378S) gave a relatively low yield of soluble protein that was $\sim 90\%$ pure.

NMR spectroscopy

NMR sample preparation and sequence-specific ¹HN, ¹⁵N, ¹³C $_{\alpha}$, ¹³C $_{\beta}$ and ¹³C′ assignments have been reported previously (Conte *et al.*, 1999). Side chain resonance assignments have been achieved using HCCH-TOCSY experiments (Bax *et al.*, 1990). NMR spectra were acquired at 302 K on a four channel Bruker DRX500 equipped with a z -shielded gradient and triple resonance probe. Distance restraints used in structure calculations were obtained from ¹H/¹⁵N and ¹H/¹³C-edited NOESY-HSQC experiments (Fesik and Zuiderweg, 1988; Norwood *et al.*, 1990), as well as from ¹⁵N/¹³C and ¹³C/¹³C-edited HMQC-NOESY-HSQC experiments (Vuister *et al.*, 1993). Dihedral ϕ angles were obtained estimating ³J_{HN α} coupling constants from HNHA experiments (Kuboniwa *et al.*, 1994). Hydrogen-bonded NH groups were identified by the presence of amide proton resonances in HSQC spectra recorded 12 h after dissolving in D₂O. [¹H]-¹⁵N NOE experiments with minimal water saturation were acquired using the pulse-sequence of Farrow *et al.* (1994). All experiments incorporated gradient sensitivity enhancement (Kay *et al.*, 1992). NMR data were processed using XWINNMR and analysed using AURELIA (Neidig *et al.*, 1995).

Structure calculation

NOE cross peak intensities were measured at 100 and 120 ms mixing times. Fifty structures were calculated from random starting co-ordinates on the basis of 702 NOE distance restraints, composed of six intraresidue, 464 short range (residue *i* to residue *i* + *j*, where $1 < j \leq 4$) and 238 long range (residue *i* to residue *i* + *j*, where $j > 4$) connectivities, 223 dihedral angle restraints, composed of 112 ϕ and 111 ψ angles, and 120 H-bond distance restraints. We used only conformational-sensitive NOEs, i.e. intraresidue conformational insensitive distance restraints were not employed. A dynamics simulated annealing protocol executed within the program X-PLOR (Nilges *et al.*, 1988; Brünger, 1993) was employed. The restraints were distributed as follows: five intraresidue, 221 short range and 123 long range distance restraints, 115 dihedral angle and 62 H-bond restraints in RRM-3, and one intraresidue, 237 short range and 115 long range distance restraints, 108 dihedral angle and 58 H-bond restraints in RRM-4. The distance restraints were calibrated internally using known sequential distances. NOEs observed at 120 ms mixing time were placed in three categories on the basis of estimated cross peak intensities: strong (< 2.8 Å), medium (< 3.5 Å) and weak (< 5.0 Å). The dihedral angle restraints were calculated from ³J_{HN α} coupling constants and by using the backbone torsion angle prediction package TALOS (Cornilescu *et al.*, 1999). Excellent agreement was found for ϕ backbone angle values from ³J_{HN α} measurements and from TALOS. Of the 97 good ϕ/ψ predictions only one ϕ angle was found in disagreement with the ³J_{HN α} coupling constant. Interactive analysis of the TALOS predictions, together with ³J_{HN α} coupling constant data, allowed additional dihedral restraints (14 ϕ/ψ predictions and one ϕ angle) to be included in the structure calculation. The final family comprised the 20 lowest total energy structures, which contain no distance violation > 0.2 Å and no dihedral violation $> 4.0^\circ$.

Preparation of RNA

EMCV IRES plasmids were the kind gift of Richard Jackson and Ann Kaminski. The construction of plasmid pSG6A, which contains nucleotides 259–848 of the EMCV IRES fused to the open reading frame of poliovirus 2A protein, has been described elsewhere (Kaminski *et al.*, 1994; Kaminski and Jackson, 1998). pSG6A is derived from plasmid pSG1 and contains six adenines in the ‘A bulge’ within the JK stem-loop (Kaminski and Jackson, 1998). Plasmid pGEM1-Dom1 contains nucleotides 260–444 corresponding to domain 1 of the EMCV IRES (Kaminski *et al.*, 1995). A plasmid containing the full-length IRES of FMDV strain O₁K was kindly provided by Graham Belsham [pSP64-FMD(polyA)]. Transcription reactions were performed using methods

similar to those described previously (Kaminski *et al.*, 1990). Typically, 20 µl reactions, programmed with 1 µg of linearized DNA, were performed in 40 mM Tris pH 8, 15 mM MgCl₂, 5 mM dithiothreitol (DTT), 6 U of RNAGuard (Amersham-Pharmacia Biotech), 1 mM each of ATP, CTP and GTP, 0.5 mM UTP, 50 µCi of [α -³²P]UTP (at >3000 Ci/mmol) and 1 µl of purified T7 RNA polymerase (provided by A.Kaminski). The reactions were incubated for 30 min at 37°C; unincorporated nucleotides were removed using NucTrap gel filtration columns (Stratagene).

Synthetic RNA oligonucleotides were prepared and purified by HPLC by Oswell Research Products Ltd. For use in binding assays the RNA oligonucleotides were 5'-end-labelled with [γ -³²P]ATP in reactions catalysed by T4 polynucleotide kinase (USB). Unincorporated nucleotides were removed by the use of Microspin G-25 spin columns (Amersham-Pharmacia Biotech).

Binding assays

The affinity of the various PTB fragments and mutants for radiolabelled RNA targets was assessed using filter binding assays. PTB-RNA binding reactions (75 µl) were prepared and incubated for at least 15 min at room temperature (24°C) in 10 mM HEPES pH 7.25, 100 mM KCl, 3 mM MgCl₂, 5% glycerol, 1 mM DTT, 50 µg/ml yeast tRNA (Boehringer Mannheim), 50 µg/ml human serum albumin (Delta Biotechnology). The RNA concentration was typically fixed at 4 nM. Assays were performed using the protein-binding Protran BA-85 nitrocellulose membrane (Schleicher and Schuell). The membrane was washed extensively in 10 mM HEPES pH 7.25, 3 mM MgCl₂, 5% glycerol, 1 mM DTT and mounted on a 96-well dot-blotter (Bio-Rad). Before and after application of 65 µl of the binding reaction, the membrane was washed with 180 µl of wash buffer. Following the experiment, the membrane was dried and the quantity of bound PTB-RNA complex determined by scintillation counting of Cerenkov radiation.

The results of binding assays were analysed using simple equilibrium binding models for the protein-RNA interaction. For the truncated proteins, a single binding site was assumed. HisPTB, which is a dimer in solution (Pérez *et al.*, 1997a), was assumed to bind in either of two orientations to a single site on the RNA target. Such a scheme may be an over-simplification of the true interaction but provides a reasonable fit to the data. The values reported for dissociation constants in Tables II and III are the average of at least two independent experiments; variation between individual experiments was typically <20%.

Co-ordinates

Structure co-ordinates have been submitted to the Protein Data Bank, accession code 1qm9.

Acknowledgements

We are grateful to Professor P.A.Sharp (MIT, USA) for the gift of human PTB cDNA, to Richard Jackson and Ann Kaminski (Cambridge, UK) for the gift of EMCV IRES plasmids, to Graham Belsham (Pirbright, UK) for the gift of a plasmid containing the full-length IRES of FMDV strain O₁K, to Doug Black (UCLA, USA) for providing us with information prior to publication, to Chris Smith (Cambridge, UK) for valuable discussions and to Peter Brick for critical reading of the manuscript. The authors are indebted to The Wellcome Trust and the BBSRC for financial support. S.M. and S.C. are members of the Imperial College Centre for Structural Biology. J.G. acknowledges the award of an MRC studentship. A list of NMR assignments and restraints is available from M.R.C. (sasi.conte@port.ac.uk).

References

Allain,F.H.-T., Howe,P.W.A., Neuhaus,D. and Varani,G. (1997) Structural basis of the RNA-binding specificity of human U1A protein. *EMBO J.*, **16**, 5764–5774.
 Ashiya,M. and Grabowski,P.J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA*, **3**, 996–1015.
 Avis,J.M., Allain,F.H.-T., Howe,P.W.A., Varani,G., Nagai,K. and Neuhaus,D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.
 Bax,A., Clore,G.M. and Gronenborn,A.M. (1990) ¹H-¹H correlation via

isotropic mixing of ¹³C magnetization, a new 3-dimensional approach for assigning ¹H and ¹³C spectra of ¹³C enriched proteins. *J. Magn. Reson.*, **88**, 425–431.
 Belsham,G.J. and Sonenberg,N. (1996) RNA-protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.*, **60**, 499–511.
 Brünger,A.T. (1993) *X-Plor Manual version 3.1*. Yale University, New Haven, CT.
 Chan,R. and Black,D. (1997) The polypyrimidine tract binding protein binds upstream of neural cell-specific *c-src* exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.*, **17**, 4667–4676.
 Conte,M.R., Grüne,T., Curry,S. and Matthews,S. (1999) Resonance assignment and topology of a 22 kD C-terminal fragment of the polypyrimidine tract binding protein (PTB) containing two RNA binding domains. *J. Biomol. NMR*, **14**, 383–384.
 Cornilescu,G., Delaglio,F. and Bax A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
 Cote,C.A., Gautreau,D., Denegre,J.M., Kress,T.L., Terry,N.A. and Mowry,K.L. (1999) A *Xenopus* protein related to hnRNP I has a role in cytoplasmic RNA localization. *Mol. Cell*, **4**, 431–437.
 Crowder,S.M., Kanaar,R., Rio,D.C. and Alber,T. (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. *Proc. Natl Acad. Sci. USA*, **96**, 4892–4897.
 Deo,R.C., Bonanno,J.B., Sonenberg,N. and Burley,S.K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell*, **98**, 835–845.
 Ding,J., Hayashi,M.K., Zhang,Y., Manche,L., Krainer,A.R. and Xu,R.M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.*, **13**, 1102–1115.
 Farrow,N.A. *et al.* (1994) Backbone dynamics of a free and phosphopeptide complexed src homology-2 domain studied by ¹⁵N NMR relaxation. *Biochemistry*, **33**, 5984–6003.
 Fesik,S.W. and Zuiderweg,E.R.P. (1988) Heteronuclear 3-dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two dimensional NMR spectra. *J. Magn. Reson.*, **78**, 588–593.
 Ghatti,A., Piñol-Roma,S., Michael,W.M., Morandi,C. and Dreyfuss,G. (1992) hnRNP-I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.*, **20**, 3671–3678.
 Gil,A., Sharp,P.A., Jamison,S.F. and Garcia Blanco,M.A. (1991) Characterization of cDNAs encoding the polypyrimidine tract binding protein. *Genes Dev.*, **5**, 1224–1236.
 Gooding,C., Roberts,G.C. and Smith,C.W. (1998) Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated α -tropomyosin exon. *RNA*, **4**, 85–100.
 Görlach,M., Burd,C.G. and Dreyfuss,G. (1994) The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein-C proteins. *J. Biol. Chem.*, **269**, 23074–23078.
 Handa,N., Nureki,O., Kurimoto,K., Kim,I., Sakamoto,H., Shimura,Y., Muto,Y. and Yokoyama,Y. (1999) Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature*, **398**, 579–585.
 Hellen,C.U.T., Pestova,T.V., Litterst,M. and Wimmer,E. (1994) The cellular polypeptide p57 (pyrimidine tract binding protein) binds to multiple sites in the poliovirus 5' untranslated region. *J. Virol.*, **68**, 941–950.
 Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
 Hunt,S.L. and Jackson,R.J. (1999) Polypyrimidine-tract binding protein (PTB) is necessary, but not sufficient, for efficient internal initiation of translation of human rhinovirus-2 RNA. *RNA*, **5**, 344–359.
 Ito,T., Muto,Y., Green,M.R. and Yokoyama,S. (1999) Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF⁶⁵). *EMBO J.*, **18**, 4523–4534.
 Jackson,R.J. (1996) A comparative view of initiation site selection mechanisms. In Hershey,J.W.B., Mathews,M.B. and Sonenberg,N. (eds), *Translational Control*. Cold Spring Harbor Laboratory Press, New York, NY, pp. 71–112.
 Jang,S.J. and Wimmer,E. (1990) Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA binding protein. *Genes Dev.*, **4**, 1560–1572.
 Kaminski,A. and Jackson,R.J. (1998) The polypyrimidine tract binding protein (PTB) requirement for internal initiation of translation of

- cardiovirus RNAs is conditional rather than absolute. *RNA*, **4**, 626–638.
- Kaminski, A., Howell, M.T. and Jackson, R.J. (1990) Initiation of encephalomyocarditis virus RNA translation: the authentic initiation site is not selected by a scanning mechanism. *EMBO J.*, **9**, 3753–3759.
- Kaminski, A., Belsham, G.J. and Jackson, R.J. (1994) Translation of encephalomyocarditis virus RNA: parameters influencing the selection of the internal initiation site. *EMBO J.*, **13**, 1673–1681.
- Kaminski, A., Hunt, S.L., Patton, J.G. and Jackson, R.J. (1995) Direct evidence that the polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *RNA*, **1**, 924–938.
- Kay, L.E., Keifer, P. and Saarinen, T.J. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.*, **114**, 10663–10665.
- Kenan, D.J., Query, C.C. and Keene, J.D. (1991) RNA recognition: towards identifying determinants for specificity. *Trends Biochem. Sci.*, **16**, 214–220.
- Kolupaeva, V.G., Hellen, C.U.T. and Shatsky, I.N. (1996) Structural analysis of the interaction of the pyrimidine tract-binding protein with the internal ribosomal entry site of encephalomyocarditis virus and foot-and-mouth disease virus RNAs. *RNA*, **2**, 1199–1212.
- Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) Measurement of H-N-H- α J-couplings in calcium-free calmodulin using new 2D and 3D water-flip-back methods. *J. Biomol. NMR*, **4**, 871–878.
- Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*, **8**, 477–486.
- Lee, A.L., Volkman, B.F., Robertson, S.A., Rudner, D.Z., Barbash, D.A., Cline, T.W., Kanaar, R., Rio, D.C. and Wemmer, D.E. (1997) Chemical shift mapping of the RNA binding interface of the multiple-RBD protein Sex-lethal. *Biochemistry*, **36**, 14306–14317.
- Lou, H., Helfman, D.M., Gagel, R.F. and Berget, S.M. (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol. Cell. Biol.*, **19**, 78–85.
- Lutz-Freyermuth, C., Query, C.C. and Keene, J.D. (1990) Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl Acad. Sci. USA*, **87**, 6393–6397.
- Luz, N. and Beck, E. (1991) Interaction of a cellular 57kDa protein with the internal translation initiation site of foot-and-mouth disease virus. *J. Virol.*, **65**, 6486–6494.
- Mittermaier, A., Varani, L., Muhandiram, D.R., Kay, L.E. and Varani, G. (1999) Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. Mol. Biol.*, **294**, 967–979.
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J.L. and Proudfoot, N.J. (1998) The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev.*, **12**, 2522–2534.
- Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Crystal structure of the RNA-binding domains of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Nagata, T., Kanno, R., Kurihara, Y., Uesugi, S., Imai, T., Sakakibara, S., Okano, H. and Katahira, M. (1999) Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1. *J. Mol. Biol.*, **287**, 315–330.
- Neidig, K.-P., Geyer, M., Gorler, A., Antz, C., Saffrich, R., Beneicke, W. and Kalbitzer, H.R. (1995) Aurelia, a program for computer aided analysis of multidimensional NMR spectra. *J. Biomol. NMR*, **6**, 255–270.
- Nilges, M., Gronenborn, A. and Clore, G.M. (1988) Determination of 3-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.
- Norwood, T.J., Boyd, J., Heritage, J.E., Soffe, N. and Campbell, I. (1990) Comparison of techniques for ^1H detected heteronuclear ^1H and ^{15}N spectroscopy. *J. Magn. Reson.*, **87**, 488–501.
- Oh, Y.L. et al. (1998) Determination of functional domains in polypyrimidine-tract binding protein. *Biochem. J.*, **331**, 169–175.
- 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.
- Patton, J.G., Mayer, S.A., Tempst, P. and Nadalginard, B. (1991) Characterisation and molecular cloning of polypyrimidine tract binding protein. A component of a complex necessary for pre-messenger-RNA splicing. *Genes Dev.*, **5**, 1237–1251.
- Pérez, I., McAfee, J.G. and Patton, J.G. (1997a) Multiple RRM contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry*, **36**, 11881–11890.
- Pérez, I., Lin, C.-H., McAfee, J.G. and Patton, J.G. (1997b) Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection *in vivo*. *RNA*, **3**, 764–778.
- Price, S.R., Evans, P.R. and Nagai, K. (1998) Crystal structure of the spliceosomal U2B''–U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature*, **394**, 645–650.
- Query, C.C., Bentley, R.C. and Keene, J.D. (1989) A common RNA recognition motif identified within a defined U1 RNA-binding domain of the 70k U1 snRNP protein. *Cell*, **57**, 89–101.
- Scherly, D., Boelens, W., van Venrooij, W.J., Dathan, N.A., Hamm, J. and Mattaj, J.W. (1989) Identification of the RNA-binding segment of human U1 A-protein and definition of its binding-site on U1 snRNA. *EMBO J.*, **8**, 4163–4170.
- Shamoo, Y., Abdul-Manan, N. and Williams, K.R. (1995) Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res.*, **23**, 725–728.
- Shamoo, Y., Krueger, U., Rice, L.M., Williams, K.R. and Steitz, T.A. (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nature Struct. Biol.*, **4**, 215–222.
- Southby, J., Gooding, C. and Smith, C.W.J. (1999) Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of α -actinin mutually exclusive exons. *Mol. Cell. Biol.*, **19**, 2699–2711.
- Valcárcel, J. and Gebauer, F. (1997) Post-transcriptional regulation: the dawn of PTB. *Curr. Biol.*, **7**, R705–R708.
- Vuister, G.W., Clore, G.M., Gronenborn, A., Powers, R., Garrett, D.S., Tschudin, R. and Bax, A. (1993) Increased resolution and improved spectral quality in 4-dimensional $^{13}\text{C}/^{13}\text{C}$ separated HMQC-NOESY-HMQC spectra using pulsed-field gradients. *J. Magn. Reson. B*, **101**, 210–213.
- Witherall, G.W., Gil, A. and Wimmer, E. (1993) Interaction of polypyrimidine tract binding protein with the encephalomyocarditis virus mRNA internal ribosomal entry site. *Biochemistry*, **32**, 8268–8275.
- Xu, R.-M., Jokhan, L., Cheng, X., Mayeda, A. and Krainer, A.R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure*, **5**, 559–570.
- Yamamoto, H., Tsukahara, K., Kanaoka, Y., Jinno, S. and Okayama, H. (1999) Isolation of a mammalian homologue of a fission yeast differentiation regulator. *Mol. Cell. Biol.*, **19**, 3829–3841.

Received March 17, 2000; revised and accepted April 20, 2000