

A peer-reviewed version of this preprint was published in PeerJ on 13 March 2014.

[View the peer-reviewed version](https://doi.org/10.7717/peerj.305) (peerj.com/articles/305), which is the preferred citable publication unless you specifically need to cite this preprint.

Joshi A, Esteve V, Buckroyd AN, Blatter M, Allain FH-, Curry S. 2014. Solution and crystal structures of a C-terminal fragment of the neuronal isoform of the polypyrimidine tract binding protein (nPTB) PeerJ 2:e305 <https://doi.org/10.7717/peerj.305>

Solution and crystal structures of a C-terminal fragment of the neuronal isoform of the polypyrimidine tract binding protein (nPTB)

Amar Joshi^{1*‡}, Vicent Esteve^{2#‡}, Adrian N. Buckroyd^{1‡}, Markus Blatter², Frédéric H.-T. Allain^{2†}, Stephen Curry^{1†}

¹Department of Life Sciences, Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom. ²Institute of Molecular Biology and Biophysics, ETH Zürich, Switzerland.

*Present address: Department of Biochemistry, Henry Wellcome Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, United Kingdom. #Present address: CIBER-BBN, Dep. Química Física, Universitat de Valencia, Dr. Moliner, 50 46100-Burjassot, Valencia, Spain. ‡Present address: Department of Biochemistry, University of Cambridge, Tennis Court Road, CB2 1QW, United Kingdom.

‡These authors contributed equally to this work.

†Corresponding authors:

Frédéric H.-T. Allain, Institute of Molecular Biology and Biophysics, ETH Zürich, Switzerland. Tel: +41 44 633 3940; email: allain@mol.biol.ethz.ch

Stephen Curry, Department of Life Sciences, Imperial College, Exhibition Road, London SW7 2AZ, United Kingdom. Tel: +44 20 7594 7632; email: s.curry@imperial.ac.uk

Abstract

The eukaryotic polypyrimidine tract binding protein (PTB) serves primarily as a regulator of alternative splicing of messenger RNA, but is also co-opted to other roles such as RNA localisation and translation initiation from internal ribosome entry sites. The neuronal paralogue of PTB (nPTB) protein is 75% identical in amino acid sequence with PTB. Although the two proteins have broadly similar RNA binding specificities and effects on RNA splicing, differential expression of PTB and nPTB can lead to the generation of alternatively spliced mRNAs. RNA binding by PTB and nPTB is mediated by four RNA recognition motifs (RRMs). We present here the crystal and solution structures of the C-terminal domain of nPTB (nPTB34) which contains RRM3 and 4. As expected the structures are similar to each other and to the solution structure of the equivalent fragment from PTB (PTB34). The result confirms that, as found for PTB, RRM3 and 4 of nPTB interact with one another to form a stable unit that presents the RNA-binding surfaces of the component RRM3s on opposing sides. The major differences between PTB34 and nPTB34 arise from amino acid side chain substitutions on the exposed β -sheet surfaces and adjoining loops of each RRM, which are likely to modulate interactions with RNA.

42 Introduction

43 Alternative splicing is the rule rather than the exception in the human genome. Over
44 95% of multi-exon genes produce alternatively spliced mRNAs (Nilsen & Graveley,
45 2010), yielding many more protein variants or isoforms than specified by the number
46 of genes. The process of alternative splicing is controlled by regulatory sequences
47 within pre-mRNA transcripts that recruit suites of proteins to determine whether
48 selected exons are included or excluded as mRNA is brought to maturity.

49 The polypyrimidine tract binding protein (PTB) is a well-characterized regulator of
50 alternative splicing (García-Blanco, Jamison & Sharp, 1989; Gil et al., 1991) although
51 it also has significant roles that affect the processing, localization and use of a variety
52 of mRNAs [reviewed by (Sawicka et al., 2008)]. PTB does not recognize a unique
53 RNA sequence; rather it is specific for short motifs (*e.g.* UCUCU, UCUU) within
54 pyrimidine rich regions of pre-mRNA (Singh, Valcárcel & Green, 1995; Pérez et al.,
55 1997). In alternative splicing, PTB binding is most commonly associated with
56 exclusion or skipping of the regulated exon, although more recent genome-wide
57 splicing studies have indicated that PTB binding can also promote exon *inclusion* in a
58 minority of cases (Xue et al., 2009; Llorian et al., 2010; Witten & Ule, 2011).

59 Three tissue-specific homologues of PTB have been identified — smPTB, ROD1 and
60 nPTB — all of which have high levels of amino acid sequence identity with the
61 prototypical protein (69-74%) and are therefore likely to have very similar structures.
62 smPTB (smooth muscle tissue) has so far only been found in rodents (Gooding, Kemp
63 & Smith, 2003), whereas the homologue ROD1 is expressed primarily in
64 hematopoietic tissues (Yamamoto et al., 1999). nPTB is found predominantly in brain
65 and testis but also, at low levels, in muscle cells (Markovtsov et al., 2000; Polydorides
66 et al., 2000).

67 PTB and nPTB bind to the same RNA sequences and have similar effects on
68 alternative splicing events for a number of transcripts (Markovtsov et al., 2000;
69 Pilipenko et al., 2001; Spellman, Llorian & Smith, 2007). However, for some
70 neuronal-specific alternative splicing events the expression of nPTB leads to distinct
71 outcomes. For example, the tyrosine kinase *c-src* has an *n-src* isoform found only in
72 brain tissue that requires nPTB for the selective inclusion of a neuronal-specific exon
73 (Markovtsov et al., 2000; Sharma, Falick & Black, 2005; Boutz et al., 2007).

PTB (and its homologues) contain 4 RNA recognition motif domains (RRMs) separated by long flexible linkers and arrayed, when not bound to RNA, in a relatively extended, linear conformation (Petoukhov et al., 2006). The solution structures of each RRM have been solved in the absence (Conte et al., 2000; Simpson et al., 2004; Vitali et al., 2006) and presence (Oberstrass et al., 2005) of RNA. While RRM1 and RRM4 adopt a canonical $\beta\alpha\beta\beta\alpha\beta$ topology with two helices packed against a four-strand β -sheet that forms the primary RNA binding surface, RRM2 and RRM3 have a modified architecture. In these domains a C-terminal extension beyond β 4 loops across the upper edge of the β -sheet and adds a fifth strand on the far side of the sheet; this expands the size of the RNA binding platform, while the β 4- β 5 loop provides additional points of interaction with RNA (Conte et al., 2000; Simpson et al., 2004; Oberstrass et al., 2005).

The α -helical regions of RRM3 and RRM4 pack together to form a stable unit in which the RNA-binding β -sheets are exposed on opposite sides; it has been proposed that this configuration enforces looping of RNA that plays an important role in defining exon structures to be excised from pre-mRNA (Oberstrass et al., 2005; Vitali et al., 2006; Lamichhane et al., 2010). Although early studies had suggested that RRM3 and RRM4 might not interact at physiological salt concentrations (Conte et al., 2000; Clerte & Hall, 2006), it is now generally accepted that this pair of RRMs have fixed relative orientations (Lamichhane et al., 2010; Maynard & Hall, 2010). Evidence suggests that the RRM1 and RRM2 bind RNA as individual domains although RRM2 is also involved in protein-protein interactions with the splicing co-regulator Raver1 (Rideau et al., 2006). Crystallographic analysis has revealed how the helical face of the PTB RRM2 provides a binding surface for specific recognition of specific peptide motifs in Raver1 (Joshi et al., 2011).

Although PTB and nPTB are likely to have similar structures because of their high sequence identity (75%) (Fig. 1a), to begin to tease out the structural basis for observed differences in splicing regulation we have determined the structure of a C-terminal fragment of nPTB that contains RRM3 and RRM4 (nPTB34). We have solved the structures both crystallographically and by NMR. Here we compare those structures with each other and with the solution structure of the equivalent fragment from PTB isoform1 (called here PTB34).

Results and Discussion

Crystal Structure of nPTB34

nPTB34, a C-terminal nPTB fragment that contains RRM3 and 4 (residues 336-531) was over-expressed in *E. coli*, purified and crystallized by sitting drop vapor diffusion in 20% PEG 6000, 0.1 Tris pH 8.0 with 2 mM ZnCl₂ as an additive (see Materials and Methods). The crystals diffracted X-rays to 1.7 Å and were found to belong to space-group P1. Attempts to phase the diffraction data by molecular replacement using the solution structure of PTB34 (Vitali et al., 2006) were unsuccessful. Instead, phases were obtained by multi-wavelength anomalous dispersion (MAD) using data collected from a crystal of Se-Met-labeled nPTB34. This produced a high-quality electron density map that revealed a total of eight nPTB34 molecules in the asymmetric unit (Fig. 1b). The eight polypeptide chains were built almost in their entirety; due to poor electron density residue 336 at the start of nPTB34 was omitted, as were residues 420-423 (and one or two flanking residues in some chains) at the C-terminal end of the β4-β5 loop of RRM3. The final model, which incorporates 1099 water molecules, was refined to an R_{free} of 27.3% with good stereochemistry (Fig. 1c). Full data collection and refinement statistics are given in Table 1.

The structures of the eight copies of nPTB34 within the asymmetric unit of our crystals are very similar to one another (Fig. 1d). Pairwise superpositions of the eight chains give an average root-mean-square deviation (RMSD) for C_α atoms of only 0.17 Å. This is not surprising since the unit cell parameters deviate only marginally from a P2₁2₁2 space group (see Materials and Methods); what this means is that there are only two positions in the asymmetric unit, exemplified by chains A, B, E and F on the one hand and C, D, G and H on the other, that have significantly different packing environments.

Each molecule in the asymmetric unit is associated with a Zn²⁺ ion that mediates crystal-packing contact between His 412 from one chain and His 491 of a neighboring molecule (*e.g.* between chains A and C within the asymmetric unit). The tetrahedral coordination required by zinc appears to be completed by a Cl⁻ ion and a water molecule. Four additional crystal contacts are bridged by Zn²⁺ ions linking His 520 of one molecule (chains A, B, E and F) and the same residue in a second molecule of

PTB34 either inside or outside the asymmetric unit. In this case, the two additional ligands for Zn^{2+} are presumed to be water molecules though the electron density was too weak to incorporate them in the refined model. These zinc-mediated interactions explain why the cation is required to obtain crystals (see Materials and Methods) but, since only two of the four ligands are provided by amino acids from nPTB34, they are not considered to be physiologically significant.

NMR structure of nPTB34

The solution NMR structure of nPTB34 was obtained from a construct containing residues 325-531 and is very similar to the one used for crystallography (see Materials and Methods). Using a total of 3622 nOe derived distance restraints and 55 hydrogen bond restraints identified from slow exchanging amides we derived a fairly good ensemble of nPTB34 conformations in solution with a heavy atom RMSD of 0.71 Å of the structured region (Fig. 2a). None of the restraints were violated by more than 0.3 Å and 67% of backbone torsions of the structured residues lay in the most favored region defined by the Ramachandran plot. Due to signal overlaps and broad linewidths the loops (β 2- β 3-loops, β 4- β 5 loop RRM3) and part of the second beta strand in RRM3 could not be completely assigned. Missing assignments in β 2 of RRM3 appear to be the reason for the slightly different packing of the first alpha helix (α 1) in RRM3 against the β -sheet compared to the crystal structure. Since α 1 is part of the domain interface this different packing results also in a slight difference in the domain interface compared to the nPTB34 crystal structure and RNA bound and free PTB34 solution structures.

Comparison of Crystal and NMR structures of nPTB34

Overall there is very good correspondence between the structures of nPTB34 determined crystallographically and by NMR. Superposition of the two models using just C_α atoms gives an RMSD of 3.0 Å, and RMSDs of 1.8 and 2.6 Å for RRMs 3 and 4 respectively (Fig. 2b). These values are slightly higher than for the superposition of the crystal structure of nPTB34 onto the solution structures of PTB34 in the absence or presence of RNA, which both give overall RMSDs of 1.3 Å (Fig. 2c). This largely

reflects the fact that the solution structure of nPTB34 (unlike PTB34 (Vitali et al., 2006)) was determined without the aid of segmental labeling and thus chemical shift completeness for protons of only 77% could be achieved. The relative orientation of RRM3 and 4 differs slightly between the crystal and solution structures of nPTB34 but this is unlikely to be significant since there is no significant difference in the domain orientations between the crystal structure of nPTB34 and the solution structure of PTB34 (see below); moreover superposition of the crystal structures of the PTB34 fragments from nPTB and hnRNP-L (Zhang et al., 2013) (RMSD = 1.3 Å) shows that the relative disposition of RRM3 and RRM4 is highly conserved between these two proteins (Fig. 2d).

Within each RRM of nPTB34 there is good correspondence of secondary structure features between the crystal and solution structures. Those loops in the crystal structure with the highest temperature factors (RRM3 - β 2- β 3 and the C-terminal end of β 4- β 5; RRM4 - β 1- α 2, β 2- β 3, α 2- β 4; and the inter-domain linker) are those exhibiting greatest mobility in the solution structure (Fig. 2a, b). Notably, the β 4- β 5 loop is not disordered in the crystal structure of RRM2 but this may simply be because it is involved in more crystal contacts (Joshi et al., 2011).

Within the inter-domain linker (residues 433-453), the central portion (residues 439-448) exhibits considerable conformational variation in the solution structure of nPTB34 (Fig. 2a). A similar but slightly smaller region (residues 443-449), which protrudes away from the body of the domain, is variable between the eight chains present in the asymmetric unit of the nPTB34 crystals (Fig. 1d). Again, this likely reflects packing variations since there are effectively only two distinct conformers evident for residues 443-449, which partition into the two distinct packing environments of the pseudo $P2_12_12_1$ space group (see above). The N- and C-terminal portions of the linker are involved in a number of specific polar and nonpolar interactions that help to stabilize the packing of RRM3 and RRM4 and are described in more detail below.

Comparison of nPTB34 with PTB34

PTB and nPTB have 75% sequence identity overall but this is focused in the RRMs where the sequence identity between equivalent domains ranges from 80% (RRMs 3

and 4) to 88% (RRM2) (Fig. 1a). As expected, the three-dimensional structures of nPTB34 and PTB34 are very similar; differences are largely confined to the local effects of amino acid substitutions.

The crystal structure of nPTB34 reveals that the interaction between RRM3 and RRM4 is stabilized by the exclusion of water molecules, which allows direct contact between a conserved set of hydrophobic residues — Ile 505, Leu 508, Ile 509 and Phe 526 from RRM4 and the apolar portions of Ser 354, Thr 357, Leu 358, Val 361 and Met 400 from RRM3 (Fig. 3a). Ringing this hydrophobic core are direct and water-mediated hydrogen bonds between the RRMs and from the RRMs to the linker polypeptide. Within the RRM3-RRM4 interface there are only five amino acid changes between PTB34 and nPTB34 and these are generally of a conservative character; the apolar residues Ile 356, Met 400 and Val 505 in PTB34 are replaced by Thr 357, Leu 399 and Ile 505 respectively in nPTB34, while His 397 and His 400 in PTB34 are altered to Gln 398 and Tyr 401 (Table 3). The water-mediated hydrogen bond interaction from Tyr 401 of RRM3 to the side chain of Asp 501 in RRM4 of nPTB34 is therefore lost in PTB34 but neither this nor the other amino acid changes appear sufficient to significantly alter the interaction between the two domains. A similar degree of conservation of the apolar residues at the hydrophobic interface between RRM3 and RRM4 is also seen in hnRNP-L (Zhang et al., 2013) (Table 3). The relative orientation of these domains is therefore a conserved feature of PTB, nPTB and hnRNP-L. The incorporation of a pair of His residues within the inter-domain interface of PTB34 (H397, H400) may affect the stability of the module at low pH (Nordlund et al., 2003), but we are not aware of any circumstances where pH regulation of the conformation would be applied.

Although the interaction of RRM4 with the helical face of RRM3 within PTB34 and nPTB34 is much more extensive than the interaction of PTB-binding peptides in Raver1 with the equivalent face of RRM2 (Joshi et al., 2011), there is one interesting parallel. Superposition of the RRM2-raver1 co-crystal structure onto RRM3 of nPTB34 reveals that the Raver1 peptide occupies essentially the same position as helix $\alpha 2$ of RRM4 (Fig. 3b,c). Although the Raver1 peptide is non-helical — in fact it has a pinched but largely extended conformation — it inserts a pair of Leu residues (Leu 500 and Leu 501 in PDB entry 3zzy) into a hydrophobic pocket formed between

the helices $\alpha 1$ and $\alpha 2$ of RRM2 that effectively mimics the positioning of Ile 505 and Ile 509 from RRM4 into the equivalent pocket on RRM3 within nPTB34.

Although the RRMs are the most highly conserved features between PTB and nPTB, there are some notable differences on their RNA binding surfaces. These differences appear to be more concentrated on RRM3. While there are a total of five amino acid substitutions in nPTB RRM3 of residues identified to contact RNA in the solution structure of the complex of PTB34 with a CUCUCU oligonucleotide (Oberstrass et al., 2005), there are only two such amino acid changes in RRM4 (Fig. 4).

The replacement of Phe 371 in the $\beta 2$ - $\beta 3$ loop of RRM3 in PTB34 by Tyr 372 in nPTB34 is conservative and likely to preserve the stacking interaction with the cytosine base at position 5 in CUCUCU RNA oligomer. However, the substitution of Asn 376 (to Ser 377) at the start of the $\beta 3$ strand and a triplet of replacements in the $\beta 4$ - $\beta 5$ loop (Asn 413, Gln 421 and Glu 422 in PTB34 to Thr 414, Leu 422 and Asp 423 respectively in nPTB34) may modulate RNA binding (Fig. 4a). Likewise the $\beta 2$ - $\beta 3$ loop of RRM4 in PTB34 is altered in nPTB34 by deletion of Lys 489 and replacement of Arg 491 by His, a pair of changes that may also alter RNA interactions (Fig. 4b). The idea that these structural changes affect RNA affinity is supported by mutagenesis experiments which showed that residues in the $\beta 2$ - $\beta 3$ loops of RRMs 3 and 4 and in the $\beta 4$ - $\beta 5$ loop of RRMs are important for RNA binding (Conte et al., 2000; Yuan et al., 2002). The observed amino acid changes are likely to contribute to the differences observed in the interactions of PTB and nPTB with larger RNA targets (Pilipenko et al., 2001). However, it should not be forgotten that sequence differences between PTB and nPTB are focused in regions *outside* the RRM domains and these may well also be important for differences in splicing activity (Markovtsov et al., 2000; Robinson & Smith, 2005), perhaps by recruiting different protein cofactors.

Materials and Methods

Plasmid construction and protein expression and purification of nPTB34 for crystal structure determination

The cDNA coding for residues 336-531 of nPTB were amplified by PCR and ligated into the *E. coli* expression vector pETM-11 (Zou et al., 2003), which adds an N-

terminal hexahistidine tag that could be removed using the Tobacco Etch Virus N1a protease. The fragment was cloned to position the first residue (G336) as the Gly in the TEV protease recognition site (ENLYFQ/G); protease processing therefore removed all vector-derived amino acids.

Expression of nPTB34 was induced in *E. coli* BL21 (DE3) cells at 37°C for four hours by addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG). Cells were re-suspended in 0.1 M NaCl, 25 mM Tris HCl pH7.5, 0.1% Triton X-100, 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mg/mL lysozyme and lysed by sonication. Cell debris and nucleic acid were removed by centrifugation at 25,000g for 40 minutes at 4°C in the presence of 1mg/mL protamine sulfate. nPTB34 was purified from the supernatant by gravity flow through TALON Metal Affinity Resin (Clontech) with stepped elution. The hexahistidine tag was cleaved by addition of 1mg Tobacco Etch Virus N1a protease for 20 mg of nPTB34 and dialysed overnight into 50 mM Tris pH7.8, containing 0.1 M NaCl, 10% glycerol, 3mM β -mercaptoethanol and 0.5 μ M Ethylenediamine tetraacetic acid (EDTA). Cleaved nPTB34 was purified away from the tag, uncleaved nPTB34 and N1a protease by a second application to TALON resin. Proteins were further purified by size exclusion chromatography on a Superdex 75 (GE Healthcare) using an ÄKTA FPLC system in 25 mM Tris pH 7.8, 0.1 M NaCl and 0.5 mM DTT. Peak fractions were concentrated and stored at -80°C. Expression of Se-Methionine (Se-Met) labeled nPTB34 was performed as above with the following differences: proteins were expressed in *E. coli* B834 (DE3) cells in minimal media supplemented with Se-Met (Molecular Dimensions). Purification proceeded as for unlabeled protein with the addition of 5 mM β -mercaptoethanol to all buffers. Incorporation of Se-Met was determined to be 100% by MALDI-TOF mass spectrometry (data not shown).

Crystal structure determination of nPTB34

Purified recombinant nPTB34 was concentrated to $\approx$ 5 mg/mL in 100 mM NaCl, 20 mM Tris (pH 7.8) and 0.5 mM DTT. nPTB34 was crystallised by sitting drop vapor diffusion with a reservoir solution containing 0.1 M Tris (pH 8.0), 2 mM ZnCl_2 and 20% polyethylene glycol (PEG) 6000. Crystals appeared at first as stacked plates, but single crystals were obtained after micro-seeding into equilibrated drops of

nPTB34 in 0.1 M Tris (pH 8.0), 0.2 mM ZnCl₂ and 12% PEG 6000. No crystals were obtained in the absence of ZnCl₂ — the structure revealed that this is because two crystal–packing interactions are mediated by zinc ions (see results). Crystals were harvested and briefly incubated in mother liquor supplemented with 20% glycerol before flash cooling in liquid N₂.

X-ray diffraction data were collected from a single crystal (60 μm × 40 μm × 20 μm) at beamline I02 at the Diamond Light Source (Didcot, UK). Diffraction data were initially indexed in iMosflm and scaled in space group P2₁2₁2 using SCALA (Collaborative Computer Project No. 4, 1994). Closer inspection of the data in POINTLESS suggested the space group was more likely to be P2₁. At this stage molecular replacement was attempted in PHASER (McCoy et al., 2007) using the solution structure of nPTB34 as a search model to locate the four molecules in the asymmetric unit expected from the Matthews co-efficient of 2.08 Da Å⁻³ but failed to find any acceptable solutions.

We therefore switched to phasing by multi-wavelength anomalous dispersion (MAD) using isomorphous crystals grown from Se-Met labeled nPTB34 (see above). Diffraction data were collected at three wavelengths at beamline I03 at the Diamond Light Source and processed in space-group P2₁. Heavy atom sites were identified using SHELXD within autoSHARP (Bricogne et al., 2003; Vonrhein et al., 2007; Sheldrick, 2008) and following density modification the figure of merit was 0.802. Initial model building was performed automatically in BUCCANEER (Cowtan, 2006) where two nPTB34 molecules were built in the asymmetric unit. We used this initial model in Phaser 2.1 to find the remaining pair of molecules in the asymmetric unit. However, maps calculated using the ‘completed’ model remained noisy in the region corresponding to this second pair of polypeptide chains and refinement stalled at an R_{free} value of over 37%. At this stage the data were re-indexed in P1 and a molecular replacement search performed using the tetrameric arrangement of crystallographic models generated as described above was used to locate all eight copies of nPTB34 in the larger asymmetric unit. This model refined without any further difficulty. Initially, strict NCS restraints were employed during refinement in CNS v1.2 (Brunger et al., 1998). TLS refinement and latter cycles of refinement were performed in Phenix without NCS restraints (Adams et al., 2010). Model building was performed

in COOT (Emsley & Cowtan, 2004). Stereochemistry and clashes were assessed using MolProbity (Adams et al., 2010).

Plasmid construction, expression and purification of nPTB34 for solution structure.

A 207-residue C-terminal fragment (residues 325-531) of human nPTB (Swissprot Q9UKA9), which contain the third and fourth RRM, was sub-cloned into a pET-28a(+) vector, using the restriction sites Nde I and Not I. The sequence contained an N-terminal 6xHis-tag. The proteins were over-expressed in BL21(+) E.coli cultures grown in M9 minimal medium containing 50 µg/mL Kanamycin. The expression was induced by IPTG at an OD of 0.4. For ¹⁵N-labelling ¹⁵N-NH₄Cl was added as the only nitrogen source and for ¹³C-labelling normal glucose was replaced by ¹³C labeled glucose. The cell pellets were harvested by centrifugation and resuspended in lysis buffer containing 300 mM NaCl, 50 mM NaH₂PO₄ and 1 mM imidazole at pH8 and lysed by 3 passages through a cell-cracker (EmulsiFlex-C5 High Pressure Homogenizer from Avestin, Canada). The protein was purified by two successive Ni-NTA columns (1ml Ni-NTA resin per liter of culture). Elution occurred between 40 mM and 200 mM imidazole. After the second column the protein was dialyzed against the NMR buffer containing 20mM NaCl and 10 mM NaH₂PO₄ at pH 5.8 and concentrated by centricon ultra centrifugation to 1 to 2 mM. The purity of the protein sample was tested by SDS gels. The final yield per liter of culture was estimated to be around 20 mg protein. Mass spectroscopy was used to characterize the proteins.

NMR spectroscopy

Shigemi NMR tubes with 250 µL sample solution (10% or 100% D₂O) were used. NMR spectra were acquired at 303 K on a Bruker DRX-600 and a Avance-900 spectrometer. NMR data were processed using Topspin 3.0 (Bruker) and analyzed using Sparky (Goddard & Kneller, 2007).

For the sequence specific backbone assignment HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, ¹⁵N HSQC, ¹³C HSQC as well as ¹⁵N TROSY spectra were recorded. The aliphatic side chains were assigned based on a ¹⁵N NOESY and a ¹³C NOESY with the help of a ¹⁵N TOCSY and a HCCH-TOCSY. The assignment of side chain amides

was achieved by analyzing the ^{15}N NOESY and the ^{15}N HSQC, while for the assignment of the aromatic side chains a 2D TOCSY and a 2D NOESY in D_2O were recorded. Distance restraints used in structure calculations were extracted from ^{15}N -NOESY-HSQC and ^{13}C -NOESY-HSQC in H_2O and 2D NOESY in D_2O . Hydrogen-bonded NH groups were identified by the presence of amide resonances in ^{15}N HSQC spectra, which were recorded immediately after lyophilizing and dissolving the sample in D_2O .

NMR structure calculation

Initial peak picking and nOe assignments were performed using the ATNOSCANDID package (Herrmann, Guntert & Wuthrich, 2002b; Herrmann, Guntert & Wuthrich, 2002a). Peak lists of the final seventh cycle were used as an input for the program CYANA 3.0 (Guntert, 2004). The “noeassign” protocol of CYANA was used to reassign and calibrate the nOe signals of the given peak lists. These lists were manually reviewed and the distance restraint list were further used by CYANA to calculate 250 structures by a simulated annealing protocol (20,000 MD steps). Based on the target function the 50 best structures were selected for refinement using a simulated annealing protocol with the AMBER 9 suite (Pearlman et al., 1995) and against the ff99 force field (Lindorff-Larsen et al., 2010) using implicit water (Bashford & Case, 2000). The 20 final conformers were selected using a combined AMBER energy and violation energy as selection criteria.

Acknowledgements

We thank staff at the Diamond Light Source (Didcot, UK) for assistance with X-ray data collection. This work was supported in part by grant support from the Biotechnological and Biological Sciences Research Council (BBSRC) to SC and from SNF-NCCR Structural Biology to FHTA. AJ is grateful for the award of a BBSRC PhD studentship.

Figures

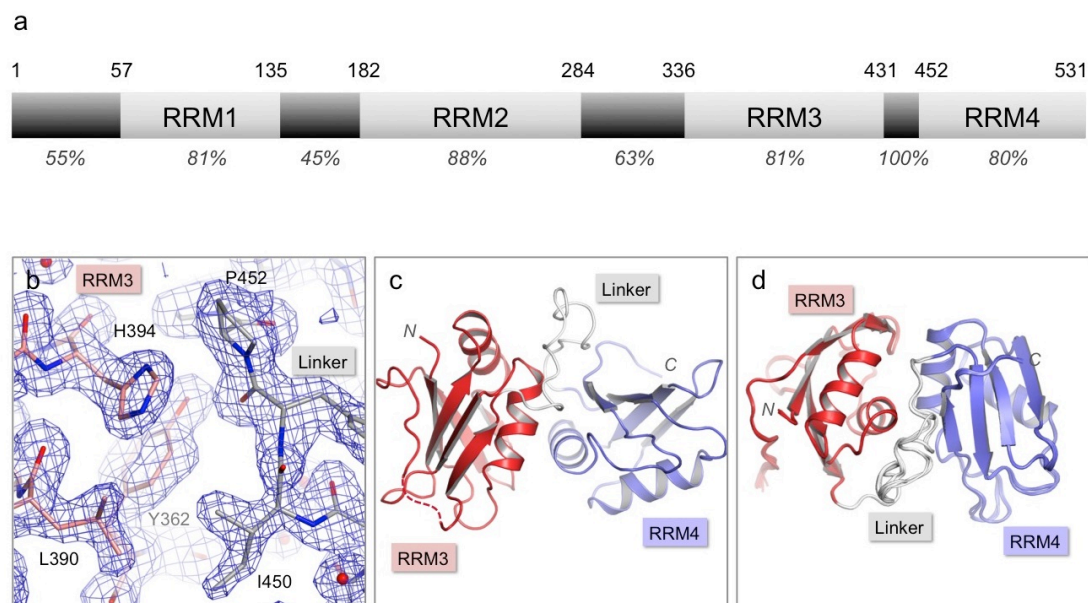


Figure 1: Crystal structure of nPTB34. (a) Domain structure of nPTB. Residue numbers for domain boundaries are given along the top; percentages below indicate the sequence identity with PTB within each defined domain or linker region. (b) $3F_o - 2F_c$ electron density map contoured at 1σ for the refined crystal structure of nPTB34 (shown in stick representation). Carbon atoms in RRM3, RRM4 and the inter-domain linker are coloured pink, light-blue and grey respectively, a colour-scheme that is maintained throughout for nPTB34; nitrogen and oxygen atoms are coloured blue and red respectively. (c) The overall fold of the crystal structure of nPTB34 showing secondary structure features. (d) Alignment of all eight chains within the crystal asymmetric unit. View rotated with respect to panel c showing good overall alignment of secondary structure elements but differences in the linker.

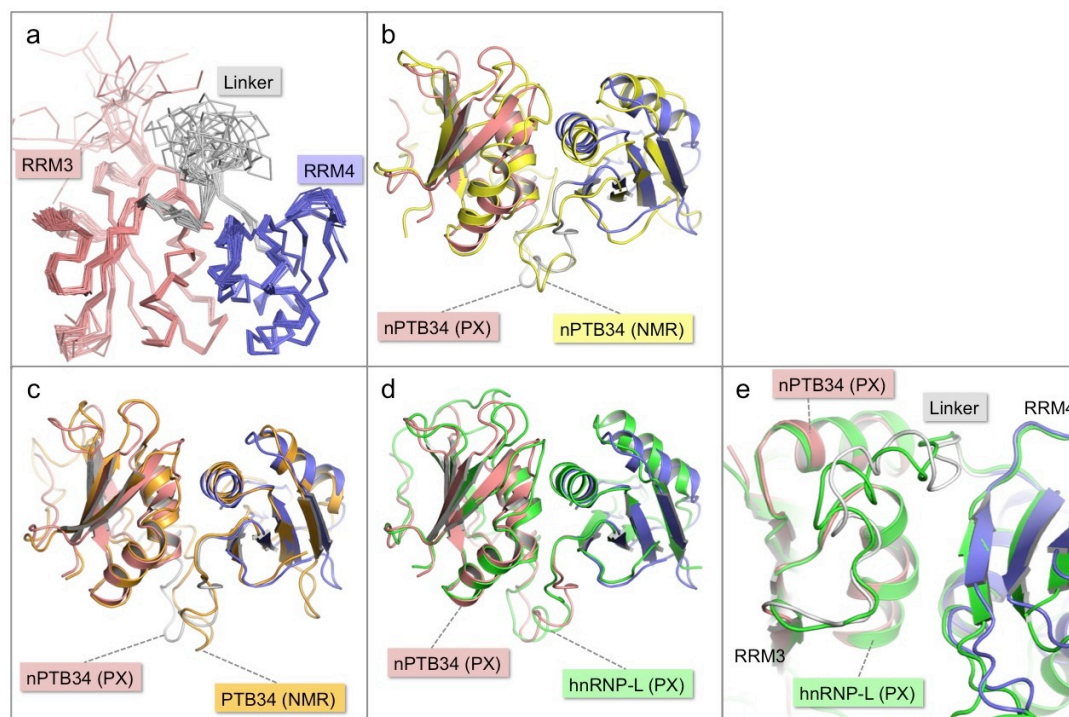


Figure 2: Solution structure of nPTB34 and comparison with related structures.

(a) Ensemble of the 20 lowest energy solution structures of nPTB34 shown as C_{α} traces coloured by domain. (b) Ribbon representation of the superposition of the crystal structure of nPTB34 (coloured by domain) on the nPTB34 solution structure (yellow). (c) Superposition of the crystal structure of nPTB34 (coloured by domain) on the solution structure of PTB34 (PDB ID 2evz; orange) (Vitali et al., 2006). (d) Superposition of the crystal structure of nPTB34 (coloured by domain) on the crystal structure of the equivalent domain from hnRNP-L (PDB ID 3to8; green) (Zhang et al., 2013). (e) Close up of the superposition shown in panel d to illustrate the similarities and differences in the linker regions between RRM3 and RRM4 in nPTB34 and hnRNP-L.

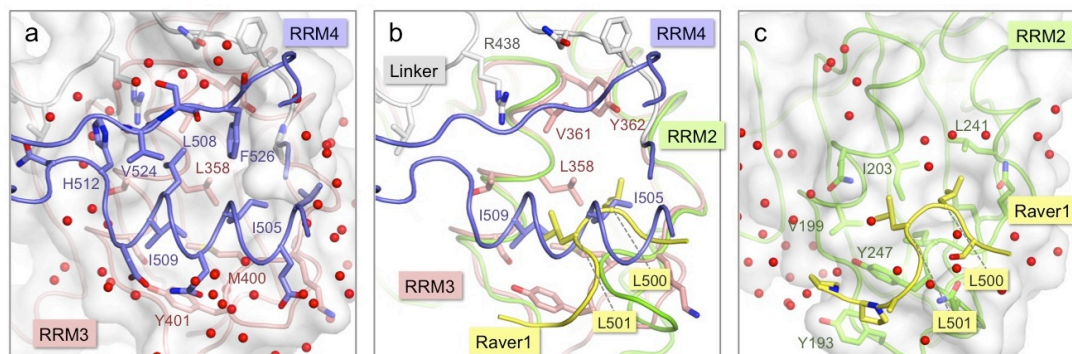


Figure 3: Comparison of the inter-domain interface within nPTB34 with the PTB RRM2-Raver1 interaction. (a) The inter-domain interface in the crystal structure of nPTB34. RRM3 is shown with pink backbone and side-chain carbon atoms along with a semi-transparent rendering of its molecular surface (light-grey). The portion of RRM4 that contacts RRM3 within nPTB34 is shown without its molecular surface. The oxygen atoms of bound water molecules located in the crystal structure are indicated by red spheres. (b) Superposition of the nPTB34 crystal structure (coloured as in panel a) with the structure of the PTB RRM2-Raver1 complex (PDB ID 3zzy; RRM2 – green; Raver1 – yellow) (Joshi et al., 2011). Selected side chains are indicated. (c) The interaction between PTB RRM2 (green) and Raver1 (yellow). As for RRM3 in panel a, the molecular surface of RRM2 is indicated by a semi-transparent rendering.

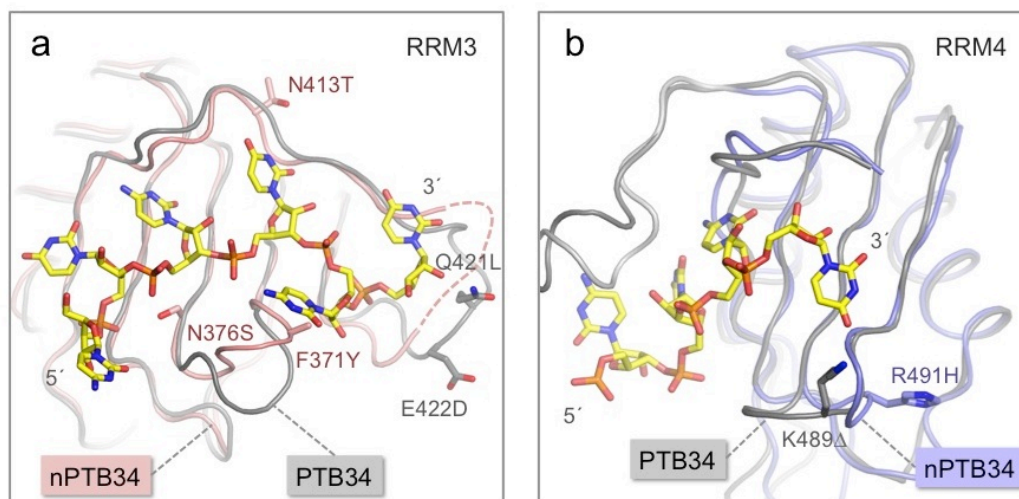


Figure 4: Location of amino acid differences on the RNA binding surfaces of PTB34 and nPTB34. (a) Superposition of the solution structure of RRM3 from PTB34 in complex with a hexameric CUCUCU RNA oligomer (PDB ID 2adc; grey) (Oberstrass et al., 2005) onto RRM3 from the crystal structure of nPTB34 (this work; pink). The disordered portion of the β 4- β 5 loop in RRM3 of nPTB34 is indicated by a dashed line. The side chains of amino acids that differ between the two structures (discussed in the text) are labeled to indicate the change from PTB34 to nPTB34 and shown, where possible, for nPTB34 (note that the Tyr side chain of the F317Y substitution is disordered in the crystal structure); the two amino acid differences that occur in a disordered loop of nPTB34 RRM3 are indicated by showing their positions and conformations of their side chains in PTB34 RRM3 (Q421, E422). (b) Superposition of the solution structure of RRM4 with the hexameric CUCUCU RNA oligomer from (PDB ID 2adc; grey) with RRM4 from the nPTB34 crystal structure (this work; light-blue). To avoid cluttering the figure only the four RNA nucleotides that make contact with RRM4 are shown. The side chains of amino acids that differ between the two structures are shown for the substitution R491H and the deletion of Lys 489 (PTB34) in nPTB34, which is labeled K491 Δ .

References

- Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW et al. . 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst D* 66:213-221.
- Bashford D, and Case DA. 2000. Generalized born models of macromolecular solvation effects. *Annual review of physical chemistry* 51:129-152.
- Boutz PL, Stoilov P, Li Q, Lin C-H, Chawla G, Ostrow K, Shiue L, Ares M, and Black DL. 2007. A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* 21:1636-1652.
- Bricogne G, Vonrhein C, Flensburg C, Schiltz M, and Paciorek W. 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Cryst D* 59:2023-2030.
- Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS et al. . 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Cryst D* 54:905-921.
- Clerte C, and Hall KB. 2006. Characterization of multimeric complexes formed by the human PTB1 protein on RNA. *RNA (New York, NY)* 12:457-475.
- Collaborative Computer Project No. 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Cryst D* 50:760-763.
- Conte MR, Grüne T, Ghuman J, Kelly G, Ladas A, Matthews S, and Curry S. 2000. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J* 19:3132-3141.
- Cowtan K. 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Cryst D* 62:1002-1011.
- Emsley P, and Cowtan K. 2004. Coot: model-building tools for molecular graphics. *Acta Cryst D* 60:2126-2132.
- García-Blanco MA, Jamison SF, and Sharp PA. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev* 3:1874-1886.

- 488 Gil A, Sharp PA, Jamison SF, and Garcia-Blanco MA. 1991. Characterization of
489 cDNAs encoding the polypyrimidine tract-binding protein. *Genes Dev* 5:1224-
490 1236.
- 491 Goddard TD, and Kneller DG. 2007. SPARKY 3.114. San Francisco: University of
492 California.
- 493 Gooding C, Kemp P, and Smith CWJ. 2003. A novel polypyrimidine tract-binding
494 protein paralog expressed in smooth muscle cells. *J Biol Chem* 278:15201-15207.
- 495 Guntert P. 2004. Automated NMR structure calculation with CYANA. *Methods in*
496 *molecular biology* 278:353-378.
- 497 Herrmann T, Guntert P, and Wuthrich K. 2002a. Protein NMR structure
498 determination with automated NOE assignment using the new software CANDID
499 and the torsion angle dynamics algorithm DYANA. *Journal of molecular biology*
500 319:209-227.
- 501 Herrmann T, Guntert P, and Wuthrich K. 2002b. Protein NMR structure
502 determination with automated NOE-identification in the NOESY spectra using the
503 new software ATNOS. *Journal of biomolecular NMR* 24:171-189.
- 504 Joshi A, Coelho MB, Kotik-Kogan O, Simpson PJ, Matthews SJ, Smith CW, and
505 Curry S. 2011. Crystallographic analysis of polypyrimidine tract-binding protein-
506 raver1 interactions involved in regulation of alternative splicing. *Structure*
507 19:1816-1825.
- 508 Lamichhane R, Daubner GM, Thomas-Crusells J, Auweter SD, Manatschal C, Austin
509 KS, Valniuk O, Allain FH-T, and Rueda D. 2010. RNA looping by PTB:
510 Evidence using FRET and NMR spectroscopy for a role in splicing repression.
511 *Proceedings of the National Academy of Sciences of the United States of America*
512 107:4105-4110.
- 513 Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, and Thornton JM. 1996.
514 AQUA and PROCHECK-NMR: programs for checking the quality of protein
515 structures solved by NMR. *Journal of biomolecular NMR* 8:477-486.
- 516 Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, and Shaw
517 DE. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein
518 force field. *Proteins* 78:1950-1958.
- 519 Llorian M, Schwartz S, Clark TA, Hollander D, Tan L-Y, Spellman R, Gordon A,
520 Schweitzer AC, De La Grange P, Ast G et al. . 2010. Position-dependent

- 521 alternative splicing activity revealed by global profiling of alternative splicing
522 events regulated by PTB. *Nat Struct Mol Biol*.
- 523 Markovtsov V, Nikolic JM, Goldman JA, Turck CW, Chou MY, and Black DL. 2000.
524 Cooperative assembly of an hnRNP complex induced by a tissue-specific
525 homolog of polypyrimidine tract binding protein. *Mol Cell Biol* 20:7463-7479.
- 526 Maynard CM, and Hall KB. 2010. Interactions between PTB RRM1s induce slow
527 motions and increase RNA binding affinity. *Journal of molecular biology*
528 397:260-277.
- 529 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ.
530 2007. Phaser crystallographic software. *J App Crystallogr* 40:658-674.
- 531 Nilsen TW, and Graveley BR. 2010. Expansion of the eukaryotic proteome by
532 alternative splicing. *Nature* 463:457-463.
- 533 Nordlund HR, Hytönen VP, Laitinen OH, Uotila STH, Niskanen EA, Savolainen J,
534 Porkka E, and Kulomaa MS. 2003. Introduction of histidine residues into avidin
535 subunit interfaces allows pH-dependent regulation of quaternary structure and
536 biotin binding. *FEBS letters* 555:449-454.
- 537 Oberstrass FC, Auweter SD, Erat M, Hargous Y, Henning A, Wenter P, Reymond L,
538 Amir-Ahmady B, Pitsch S, Black DL et al. . 2005. Structure of PTB bound to
539 RNA: specific binding and implications for splicing regulation. *Science* 309:2054-
540 2057.
- 541 Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham Iii TE, DeBolt S,
542 Ferguson D, Seibel G, and Kollman P. 1995. AMBER, a package of computer
543 programs for applying molecular mechanics, normal mode analysis, molecular
544 dynamics and free energy calculations to simulate the structural and energetic
545 properties of molecules. *Computer Physics Communications* 91:1-41.
- 546 Pérez I, Lin CH, McAfee JG, and Patton JG. 1997. Mutation of PTB binding sites
547 causes misregulation of alternative 3' splice site selection in vivo. *RNA* 3:764-778.
- 548 Petoukhov MV, Monie TP, Allain FH, Matthews S, Curry S, and Svergun DI. 2006.
549 Conformation of polypyrimidine tract binding protein in solution. *Structure*
550 14:1021-1027.
- 551 Pilipenko EV, Viktorova EG, Guest ST, Agol VI, and Roos RP. 2001. Cell-specific
552 proteins regulate viral RNA translation and virus-induced disease. *EMBO J*
553 20:6899-6908.

- 554 Polydorides AD, Okano HJ, Yang YY, Stefani G, and Darnell RB. 2000. A brain-
555 enriched polypyrimidine tract-binding protein antagonizes the ability of Nova to
556 regulate neuron-specific alternative splicing. *Proc Natl Acad Sci USA* 97:6350-
557 6355.
- 558 Rideau AP, Gooding C, Simpson PJ, Monie TP, Lorenz M, Hüttelmaier S, Singer RH,
559 Matthews S, Curry S, and Smith CWJ. 2006. A peptide motif in Raver1 mediates
560 splicing repression by interaction with the PTB RRM2 domain. *Nature structural*
561 *& molecular biology* 13:839-848.
- 562 Robinson F, and Smith CW. 2005. A splicing repressor domain in polypyrimidine
563 tract binding protein. *J Biol Chem*.
- 564 Sawicka K, Bushell M, Spriggs KA, and Willis AE. 2008. Polypyrimidine-tract-
565 binding protein: a multifunctional RNA-binding protein. *Biochem Soc Trans*
566 36:641-647.
- 567 Sharma S, Falick AM, and Black DL. 2005. Polypyrimidine tract binding protein
568 blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E
569 complex. *Mol Cell* 19:485-496.
- 570 Sheldrick GM. 2008. A short history of SHELX. *Acta Cryst A* 64:112-122.
- 571 Simpson PJ, Monie TP, Szendrői A, Davydova N, Tyzack JK, Conte MR, Read CM,
572 Cary PD, Svergun DI, Konarev PV et al. . 2004. Structure and RNA interactions
573 of the N-terminal RRM domains of PTB. *Structure* 12:1631-1643.
- 574 Singh R, Valcárcel J, and Green MR. 1995. Distinct binding specificities and
575 functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*
576 268:1173-1176.
- 577 Spellman R, Llorian M, and Smith CWJ. 2007. Crossregulation and functional
578 redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1.
579 *Mol Cell* 27:420-434.
- 580 Vitali F, Henning A, Oberstrass FC, Hargous Y, Auweter SD, Erat M, and Allain FH-
581 T. 2006. Structure of the two most C-terminal RNA recognition motifs of PTB
582 using segmental isotope labeling. *EMBO J* 25:150-162.
- 583 Vonrhein C, Blanc E, Roversi P, and Bricogne G. 2007. Automated structure solution
584 with autoSHARP. *Methods Mol Biol* 364:215-230.
- 585 Witten JT, and Ule J. 2011. Understanding splicing regulation through RNA splicing
586 maps. *Trends Genet* 27:89-97.

- 587 Xue Y, Zhou Y, Wu T, Zhu T, Ji X, Kwon Y-S, Zhang C, Yeo G, Black DL, Sun H et
588 al. . 2009. Genome-wide analysis of PTB-RNA interactions reveals a strategy
589 used by the general splicing repressor to modulate exon inclusion or skipping. *Mol*
590 *Cell* 36:996-1006.
- 591 Yamamoto H, Tsukahara K, Kanaoka Y, Jinno S, and Okayama H. 1999. Isolation of
592 a mammalian homologue of a fission yeast differentiation regulator. *Mol Cell Biol*
593 19:3829-3841.
- 594 Yuan X, Davydova N, Conte MR, Curry S, and Matthews S. 2002. Chemical shift
595 mapping of RNA interactions with the polypyrimidine tract binding protein.
596 *Nucleic Acids Res* 30:456-462.
- 597 Zhang W, Zeng F, Liu Y, Zhao Y, Lv H, Niu L, Teng M, and Li X. 2013. Crystal
598 structures and RNA-binding properties of the RNA recognition motifs of
599 heterogeneous nuclear ribonucleoprotein L: insights into its roles in alternative
600 splicing regulation. *Journal of Biological Chemistry* 288:22636-22649.
- 601 Zou P, Gautel M, Geerlof A, Wilmanns M, Koch MHJ, and Svergun DI. 2003.
602 Solution scattering suggests cross-linking function of telethonin in the complex
603 with titin. *J Biol Chem* 278:2636-2644.
604

Table 1: Crystal structure — data collection and refinement statistics

DATA COLLECTION	Native nPTB34	Se-Met labeled nPTB34		
Space-group	P1	P2 ₁		
a, b, c (Å)	60.98, 65.81, 99.55	65.81, 60.95, 99.55		
α, β, γ (°)	89.99, 90.00, 90.00	90.00, 90.01, 90.00		
		Peak	Inflection	Remote
Wavelength (Å)	0.9795	0.9808	0.9813	0.9790
Resolution range (Å)	44.83-1.69 (1.78-1.69)	55.10-2.30 (2.42-2.30)	55.14-2.30 (2.42-2.30)	55.27-2.50 (2.64-2.50)
No. of independent reflections	150219 (20577)	-	-	-
Multiplicity ¹	1.9 (1.9)	1.7 (1.6)	1.7 (1.7)	1.7 (1.7)
Completeness (%)	86.4 (81.0)	89.9 (84.8)	95.9 (96.1)	95.7 (93.5)
I/σ _I	4.2 (1.5)	10.0 (4.3)	8.7 (3.0)	8.1 (2.4)
R _{merge} (%) ²	8.8 (54.4)	6.7 (21.2)	8.6 (37.9)	9.4 (47.8)
MODEL REFINEMENT				
No. of Non-hydrogen atoms/waters	12571			
R _{model} (%) ³	22.8			
R _{free} (%) ⁴	27.4			
RMSD bonds (Å) ⁵	0.009			
RMSD bond angles (°)	1.296			
Ramachandran plot (% favoured/allowed)	97.4/1.9			
PDB Identifier	4cko			

¹Values for highest resolution shell given in parentheses

²R_{merge} = 100 × $\sum_{hkl} |I_j(hkl) - \langle I_j(hkl) \rangle| / \sum_{hkl} \sum_j I_j(hkl)$, where $I_j(hkl)$ and $\langle I_j(hkl) \rangle$ are the intensity of measurement j and the mean intensity for the reflection with indices hkl, respectively.

³R_{work} = 100 × $\sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$.

⁴R_{free} is the R_{model} calculated using a randomly selected 5% sample of reflection data that were omitted from the refinement.

⁵RMSD, root-mean-squared deviations.

616 **Table 2:** Statistics for the solution structure of nPTB34

Number of Distance Restraints	3677
intraresidual	657
sequential (i-j =1)	941
medium range (1< i-j <5)	684
long range (i-j >=5)	134
hydrogen bonds ¹	55
Energy Statistics ²	
Average distance constraint violations (Å)	
0.1-0.2 Å	52.4 +/- 4.2
0.2-0.3 Å	0.4 +/- 0.6
>0.3 Å	0.0 +/- 0.0
Maximal (Å)	0.20 +/- 0.03
Mean AMBER distance violation Energy (kcal mol ⁻¹)	100.9 +/- 2.5
Mean AMBER Energy (kcal mol ⁻¹)	-6170.0 +/- 15.5
Mean Deviation from ideal covalent geometry	
Bond Length (Å)	0.0042 +/- 0.0000
Bond Angle (degrees)	1.796 +/- 0.007
Ramachandran plot Statistics ^{2,3,4}	
Most favoured regions (%)	67.1 +/- 2.0
Additionally allowed regions (%)	29.2 +/- 1.9
Generously allowed regions (%)	3.5 +/- 0.8
Disallowed regions (%)	0.1 +/- 0.2
RMS Deviations from mean structure Statistics ^{2,3}	
Backbone atoms (Å)	0.37 +/- 0.05
Heavy atoms (Å)	0.71 +/- 0.05
PDB identifier	To be deposited

617
618 ¹Hydrogen bond constraints were identified from slow exchanging amide
619 protons in D₂O.

620 ²Statistics computed for the deposited bundle of 20 violation energy best
621 structures selected out of 30 energy best structures refined in Amber
622 (Pearlman et al., 1995).

623 ³Based on structured residue range as defined by cyana command
624 overlay: 9-47 , 52-86 , 91-113 , 126-163 , 168-207.

625 ⁴Ramachandran plot, as defined by the Procheck (Laskowski et al.,
626 1996).

Table 3: Conservation of amino acids within the interface between RRM3-RRM4 in PTB, nPTB and hnRNP-L

	PTB	nPTB	hnRNP-L
RRM3			
	S353	S354	R398
	I356	T357	N401
	L357	L358	V402
	V360	V361	L405
	Y361	Y362	Y406
	H397	Q398	N442
	L399	M400	M444
	H400	Y401	F445
RRM4			
	V501	V501	K552
	V505	I505	L556
	L508	L508	L559
	I509	I509	G560
	F526	F526	F582