Defining the roles and interactions of PTB

Panagiota Kafasla§, Ian Mickleburgh, Miriam Llorian, Miguel Coelho, Clare Gooding, Dmitry Cherny¹, Amar Joshi*², Olga Kotik-Kogan², Stephen Curry², Ian Eperon¹, Richard Jackson & Christopher WJ Smith
Department of Biochemistry, University of Cambridge, Tennis Court Road,
Cambridge CB2 1QW
¹Department of Biochemistry, University of Leicester,
²Division of Cell and Molecular Biology, Imperial College, Exhibition Road,
London SW7 2AZ, UK

§ Current address: Institute of Immunology, BSRC “Al. Fleming”, Vari, Greece
* Current address: Henry Wellcome Laboratories of Structural Biology,
Department of Biochemistry, University of Leicester, Leicester LE1 9HN, UK

Tel: 01223-333655
Fax: 01223-766002
Email: cwjs1@cam.ac.uk

Keywords: alternative_splicing, Polypyrimidine_tract_binding_protein, hnRNP, RNA,
Abstract
Polypyrimidine tract binding (PTB) protein is an abundant and widely expressed RNA binding protein with four RNA Recognition Motif (RRM) domains. PTB is involved in numerous post-transcriptional steps in gene expression in both the nucleus and cytoplasm, but has been best-characterized as a regulatory repressor of some alternative splicing events (ASEs), and as an activator of translation driven by internal ribosome entry segments (IRESs). We have used a variety of approaches to characterize the activities of PTB and its molecular interactions with RNA substrates and protein partners. Using splice-sensitive microarrays we found that PTB acts not only as a splicing repressor but also as an activator, and that these two activities are determined by the location at which PTB binds relative to target exons. We have identified minimal splicing repressor and activator domains, and have determined high resolution structures of the second RRM domain of PTB binding to peptide motifs from the co-repressor protein Raver1. Using single-molecule techniques we have determined the stoichiometry of PTB binding to a regulated splicing substrate in whole nuclear extracts. Finally, we have used tethered hydroxyl radical probing to determine the locations on viral IRESs at which each of the four RRM domains bind. We are now combining tethered probing with single molecule analyses to gain a detailed understanding of how PTB interacts with pre-mRNA substrates to effect either repression or activation of splicing.
Polypyrimidine tract binding protein (PTB) is an abundant RNA binding protein of the hnRNP family, originally identified by its binding to the polypyrimidine tract at the 3’ splice site of mammalian introns [1, 2]. Structurally, PTB consists of four RNA Recognition Motif (RRM) domains [3], with three interdomain linkers and an N-terminal leader sequence containing nuclear localization and export signals (Fig. 1A). Early speculation that PTB was an essential pre-mRNA splicing factor was rapidly dispelled, and it was subsequently recognized as a repressive regulator of alternative splicing [4]. In vitro selection experiments showed that optimal binding substrates for PTB consisted of motifs such as UCUUC embedded within more extended pyrimidine-rich contexts [5, 6]. Such motifs were found in splicing silencer elements associated with many PTB repressed exons (Fig. 1B). PTB was also found to be involved in regulating numerous other post-transcriptional processes in both the nucleus and cytoplasm, including 3’-end processing, mRNA stability, internal ribosome entry segment (IRES) driven translation and mRNA localization [7]. Of these processes, most experimental attention has focused on the roles of PTB as a repressor of splicing and activator of IRES-mediated translation initiation. Here we discuss a range of approaches that we have deployed recently to analyze the roles of PTB and its interactions with RNA targets and protein partners.

**PTB splicing maps**

PTB has been investigated as a splicing repressor in numerous model systems, allowing a number of general principles about its mode of action to be discerned. A number of models have been suggested to explain, in whole or part, PTB’s repressive activity including: direct binding competition with U2AF65 at the polypyrimidine tract; propagative binding between high affinity sites; looping of RNA between high affinity sites; inhibition of productive splicing complex assembly across introns or exons [8, 9]. In contrast, until recently only a couple of reports had suggested that PTB might be able to activate some splicing events. However, new insights have been facilitated by splice-sensitive microarrays, which allow quantitative profiling of large numbers of alternative splicing events (ASEs) in parallel, and by UV cross-linking immunoprecipitation (CLIP) followed by high-throughput sequencing, which allows a transcriptome-wide view of RNA-protein interactions [10]. Array analysis of cells in which PTB was knocked down identified large numbers of PTB-regulated ASEs [11, 12]. Although most of these ASEs were conventional targets of PTB-repression, there was a substantial minority of PTB-activated events. By combining functional PTB-target events from our array experiments with a transcriptome-wide PTB CLIP data-set [13] it was possible to decipher differences in the characteristics of PTB-repressed and PTB-activated cassette exons [12] (Fig. 2). In agreement with previous work, PTB-repressed exons were associated with PTB binding upstream of or within the regulated exon. However, in contrast to some well-studied PTB-repressed exons, including α-tropomyosin (Tpm1) exon 3 [14, 15] and SRC N1 exon [16], we found no general tendency for PTB binding downstream of repressed exons. In contrast, the newly characterized PTB-activated exons were clearly associated with PTB binding on the downstream side of the exon [12]. These exons had weaker 5’ splice sites than PTB-repressed or control alternative
exons, suggesting that PTB-binding in the downstream intron might be able to compensate for a weak 5′ splice site. However, by insertion of a PTB-binding site in the upstream intron, the exon became repressed by PTB, and the downstream site was converted from a PTB binding enhancer to a silencer [12]. PTB joins a growing list of splicing regulatory proteins showing a similar position-dependent activity [10]. This poses a series of questions about how the same protein can have opposite regulatory outcomes depending upon the location and number of binding sites, and also how several diverse proteins are able to activate splicing when bound downstream of a weak 5′ splice site [10]. The protein TIA1 can activate splicing by direct contact with U1C protein, thereby assisting U1 snRNP recruitment [17]. Perhaps other activators contact different components of U1 snRNP to assist its recruitment? Indeed PTB RRMs 1 and 2 have been shown to contact U1 snRNA, albeit in the context of repressing the SRC N1 exon [18].

**Functional domains of PTB**

Understanding the molecular function of biological macromolecules is invariably assisted by detailed structural information. Although full-length PTB has eluded high resolution structural determination, the structures of all four PTB RRMs have been determined by NMR, in both free form as well as bound to a hexameric CUCUCU RNA ligand [19-22]. This has revealed the basis of specificity of RNA recognition by each RRM (Fig. 1A), as well as interesting features of RRM organization. Notably, while the linkers between RRMs 1, 2 and 3 are flexible [23], RRMs 3 and 4 form a stable di-domain with back to back packing of the two RRMs involving the short linker [20, 22]. This di-domain structure necessitates a loop of at least 15 nt between the two pyrimidine tracts recognized by RRMs 3 and 4; targeted mutations to disrupt the didomain packing impair PTBs regulatory activity on SRC splicing [24]. Thus, it appears that the RRMs of PTB may serve not just to recognise specific sequence motifs, but also to bring about structural rearrangements of the RNA. Nevertheless, other lines of evidence suggest that PTB has additional effector functions that are distinct from RNA binding.

We previously exploited “artificial tethering” to map a minimal repressor domain of PTB [25]. The PTB binding silencer downstream of *Tpm1* exon 3 was replaced by a binding site for bacteriophage MS2 coat protein. Repression of the exon by PTB was thereby abrogated, but could be restored by expressing a fusion of PTB-MS2 coat protein. Consistent with the MS2 protein now providing RNA-binding function, deletion analysis showed that three of PTB’s four RRMs were dispensable. A minimal repressor domain consisted of just RRM2 and the following inter-RRM linker [25]. We had also found that the protein Raver1, initially identified by yeast 2-hybrid assays as a PTB interactor [26], was able to strongly promote skipping of *Tpm1* exon 3 [27]. We identified four short peptide motifs in Raver1, of the form [S/G][I/L]LGxxP, which could interact with PTB. The motifs were subsequently referred to as PTB RRM interacting (PRI) motifs [28]. NMR experiments showed that the high affinity PRI3 motif interacted with PTB RRM2 on its dorsal surface, and that RRM2 was able to form a ternary complex with RNA bound to the β-sheet surface and the PRI3 peptide bound in a shallow hydrophobic groove on the dorsal surface [29]. Much more detailed
information about the Raver1-PTB interaction has now been obtained by X-ray crystallography of PTB RRM2 linked via a flexible linker to Raver1 PRI motifs [30]. Fortuitously, the two fusion proteins that crystallized (diffracting to <1.6 Å) contained PRI3 and 4, which are the highest and lowest affinity PRI motifs from Raver1. The structural data therefore provided a wealth of information about the recognition of PRIs by PTB RRM2. In particular, the side-chain of PTB Tyr 247 is a focal point for apolar contacts from a pair of hydrophobic side-chains in the PRI motifs (L3 and P6 of PRI3, L3 and L5 of PRI4). The importance of Tyr 247 was confirmed by a Y247Q mutation, which caused a loss of Raver1 binding in vitro and a 30% reduction in the activity of the MS2 minimal PTB repressor domain. Perhaps more surprising was that mutation of the RNA binding surface of RRM2 had an even larger effect in the MS2 assay (60-90% loss of activity), even though the MS2 protein brings the fusion protein to the location of the original PTB binding site [30]. This raises the question of the RNA site that must be bound by PTB RRM2. One possibility, suggested by in vitro experiments in HeLa extracts, is that the tethered PTB interacts with the pyrimidine tract upstream of exon 3, thereby forming a protein-bridged RNA loop (M Coelho & CWJS, unpublished observations). Alternatively, it is possible that the tethered RRM2 interacts with U1 snRNA, as demonstrated for the SRC N1 exon [18].

We have also used the MS2 approach with PTB activated exons, initially focusing on an exon in the KTN1 gene [12]. Both full-length PTB as well as RRM2 and the following linker were able to activate inclusion of the KTN1 exon. In contrast, RRMs 3 and 4 fused to MS2 had the opposite effect, slightly increasing exon skipping. Thus, at least for the Tpm1 and KTN1 exons, the minimal repressor and activator domains comprise RRM2 and the following linker. However, the KTN1 exon is not an ideal experimental model since the amplitude of response to PTB knockdown or MS2-PTB recruitment is quite modest. We are therefore repeating these analyses with other exons that are more responsive to PTB, with a view to carrying out a more detailed comparison of minimal repressor and activator domains. We will combine this approach with analyses of protein (and RNA) binding targets of the minimal effectors, as described recently for Rbfox proteins [31]. Using in vitro pull-down assays with recombinant RRM2 wild type and Y247Q mutant proteins we have already found that a number of other nuclear proteins interact with the surface of PTB RRM2 recognised by PRI motifs (M. Coelho & CWJS, unpublished observations). These proteins are candidate splicing co-regulators and/or targets of PTB’s repressor domain.

**Counting proteins in complexes**

PTB typically interacts with more than one pyrimidine tract in the region of exons that it represses. To fully understand how PTB functions, it is important to know how many PTB molecules bind to target RNAs. Conventional biochemical assays, such as electrophoretic mobility shift or filter-binding, have shown that splicing substrates typically bind more than one PTB molecule [14-16, 32]. However, these analyses have used pure recombinant PTB, which could easily lead to overestimates of the true number of molecules bound in vivo, where non-specific binding will be suppressed by the numerous other proteins that coat the RNA. To circumvent this problem,
Cherny et al [14] devised a single molecule approach, using total internal reflection fluorescence (TIRF) microscopy to image the binding of GFP-tagged PTB to substrate RNAs immobilized on a slide using Cy-5 labelled complementary oligonucleotides. The PTB-GFP was present in nuclear extracts from transiently transfected HeLa cells, and PTB-RNA interactions were observed by colocalization of Cy5 and GFP fluorescence. The number of bound PTBs was determined either by analysis of successive photobleaching steps or by modelling cumulative distributions of total photon counts. This approach was used to analyse binding of PTB to substrates containing Tpm1 exon 3, which is flanked by two long pyrimidine tracts that act as PTB-binding silencer elements (Fig. B, [5, 15]). The data indicated that 5-6 PTB molecules bound around Tpm1 exon 3, thereby ruling out models for PTB repression involving propagative binding of PTB between the distantly separated high affinity binding tracts [8, 9]. Using the specificities of individual RRM binding to RNA (Fig. 1A, [20]), and assuming a short but flexible separation of sites bound by RRMs 1, 2 and 3, we were able to model how 3 PTBs would bind to the upstream and 2 to the downstream pyrimidine tract. An important insight was that not only could both of the high affinity tracts accommodate 2 or 3 PTB molecules, in agreement with experimental observations, but that there were numerous ways in which this could be achieved, with multiple overlapping arrangements for the specific RRM-RNA contacts. We suggest that the potential for multiple binding configurations may be a hallmark of high affinity binding sites. Indeed, many motifs enriched in the vicinity of PTB regulated exons (e.g. UCUCU)[12] have embedded overlapping motifs for individual PTB RRMs. Our model required that the RRM4 domains would contact the RNA outside of the P3 and DY PTB-binding pyrimidine tracts. An additional C-rich tract upstream of exon 3 is a candidate site (Fig. 1B), which is consistent with the observed specificity of RRM4 (Fig 1A). In order to test this model, we need an experimental approach that can facilitate systematic determination of the sites contacted by individual RRMs. Such an approach has been developed to interrogate the interaction of PTB with viral IRESs.

**Mapping the sites of PTB RRM contacts with IRESs**

In contrast to the unpaired pyrimidine tracts that typically mediate splicing repression by PTB, IRESs are highly structured. PTB contacts short pyrimidine rich loops as well as sequences that are predicted to be at least partially base-paired [7]. Despite the higher degree of structural order in IRES-PTB complexes, high resolution structures of full length PTB bound to IRES substrates have not been obtained. Kafasla and Jackson therefore adapted the tethered hydroxyl radical probing approach, originally used to map binding sites of ribosomal proteins on rRNAs [33]. A battery of PTB mutants were generated, each with a single cysteine residue inserted close to the RNA binding surface of one of the RRM domains [34]. Modification of cysteine with Fe(II)BABE allows the localized generation of short-lived reactive hydroxyl radicals, which can cleave the RNA backbone within a limited distance (~10-35 Å). Mapping the cleavage sites by reverse transcription primer extension reveals the region of the RNA contacted by individual RRMs. This approach was applied to the highly structured encephalomyocarditis virus (EMCV) and poliovirus (PV) IRESs [34, 35].
Combined with non-denaturing mass-spectrometry, which revealed that two PTBs bind to the EMCV IRES, but only one to the poliovirus IRES, this allowed for the first time the positions of individual PTB RRsMs to be mapped onto the distinct RNA structural domains of the two IRESs. The importance of individual RRM-RNA contacts was next tested by combining mutations designed to disrupt RNA binding by individual RRMs, with the single cysteine mutants to report on the RNA contacts by the mutated RRM and the remaining functional RRMs [36]. Inactivation of any individual RRM domain reduced affinity modestly but did not impair or reposition RNA contacts by the remaining RRMs. However, the effects upon IRES activity were variable. For EMCV, only mutation of RRM2 led to full loss of activity; mutation of RRM1, 3 or 4 had moderate effects, but combined mutation of RRMs 3 and 4 eliminated activity. In contrast, the PV IRES was inactivated by mutations of RRMs 1, 2 or 4, but tolerated mutation of RRM3 surprisingly well. RRMs 1 and 2 mapped to the structured region of domain V of the PV IRES, adjacent to the site of eIF4G binding; in contrast, RRMs 3 and 4 contacted the single stranded regions flanking the base of domain V. This agrees with independent reports that RRMs 1 and 2 favour structured RNAs while RRMs 3 and 4 bind better to single stranded pyrimidine tracts [37]. Moreover, PTB and eIF4G were seen to reciprocally modulate each others' position of binding, suggesting that PTB activation of the PV IRES involves optimizing the position of eIF4G for translation initiation just downstream of domain V [35].

**Going forward – tethered probing of splicing substrates**

The set of PTB single cysteine mutants for tethered hydroxyl radical probing, combined with the RRM mutations to selectively impair RNA binding by individual RRMs, can potentially be applied to various PTB substrate RNAs involved in other steps of gene expression, including PTB regulated ASEs. We initially attempted to use tethered hydroxyl radical probing to investigate the sites at which individual PTB RRMs contact the Tpm1 RNA. However, under a wide range of conditions it was difficult to detect specific signals (PK unpublished observations), despite the fact that we know that 5-6 PTB molecules bind to this substrate [14]. However, Tpm1 might be a particularly challenging substrate for this kind of analysis due to the large number of bound PTBs. It is likely that there are numerous different binding arrangements for the multiple RRMs interacting with the extended pyrimidine tracts, which would produce poor signal to noise ratio in the tethered probing assay. For this reason, we are applying tethered probing to FAS exon 6 [38], in which a single 16 nt exonic sequence (referred to as URE6) is the only known element required for PTB-induced exon skipping (Fig. 1B). Skipping of FAS exon 6 in HeLa nuclear extracts is promoted by the addition of recombinant PTB, thereby allowing monitoring of functional activity, as well as RRM-RNA contacts by the modified single-cysteine PTBs. Moreover, the single-molecule TIRF approach can be used to determine the stoichiometry of PTB-RNA interactions [14], which is necessary for interpretation of the probing data. As discussed at the RNA UK 2012 meeting, these combined approaches are now allowing us to build a picture of how the RRMs of PTB interact with the FAS RNA, both within URE6 and at previously unidentified locations, to promote exon skipping.
Acknowledgements
Work in the authors’ laboratories was funded by grant support from the BBSRC to SC (Ref: BB/E02209X/1) and CWJS/RJJ (BB/H004203/1), and from the Wellcome Trust to CWJS (092900). AJ is grateful for the award of a BBSRC PhD studentship.

References
analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. Mol Cell. 36, 996-1006


17 Forch, P., Puig, O., Martinez, C., Seraphin, B. and Valcarcel, J. (2002) The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J. 21, 6882-6892


protein raver1 regulates alpha-tropomyosin alternative splicing. EMBO J. 22, 6356-6364
36 Kafasla, P., Lin, H., Curry, S. and Jackson, R. J. (2011) Activation of picornaviral IRESs by PTB shows differential dependence on each PTB RNA-binding domain. RNA. 17, 1120-1131
Figure legends

Figure 1. Schematic representation of PTB and PTB-repressed exons. A) Domain organization of PTB. The four RNA Recognition Motif (RRM) domains are shown, with the specific RNA sequence motifs recognized by each RRM shown below [20], and the PRI peptide motif recognized by the dorsal surface of RRM2 shown above [28-30]. B) Depiction of rat Tpm1 exon 3 and human FAS exon 6. These are well characterized PTB-repressed exons, as discussed in the text. Known PTB binding motifs (P3, DY, URE6) are indicated by black boxes, with the sequences shown above. Sequence motifs resembling the optimal selected PTB-binding motifs are shown bold and underlined. In addition, the sequence of an additional C-rich tract upstream of Tpm1 exon 3 is shown. The grey circles represent additional elements needed for repression of the exon [39], and which bind Mbnl proteins (C.G. & CWJS, unpublished observations). The black circle upstream of P3 is the branch point for exon 3.

Figure 2. Schematic PTB splicing map as revealed by transcriptome wide analysis of PTB-regulated splicing events and binding targets [12]. Top: exons that are repressed by PTB are associated with PTB-binding on the upstream intronic flank and/or within the exon. Bottom: exons that are activated by PTB are associated with PTB-binding in the immediate downstream intronic flank.

Figure 3. Schematic representation of the interaction of PTB with picornaviral IRESs as mapped by tethered hydroxyl radical probing combined with mass spectrometric analysis [34, 35]. The four RRM of PTB are depicted by oval lines of different shading as shown. Top: Two PTB molecules interact with the EMCV IRES. Bottom: One PTB molecule binds on the PV IRES. The dashed oval line shows the proposed position of the central domain of eIF4G [40].
A

[B/G][I/L]LGxΦP

RRM1  RRM2  RRM3  RRM4

YCU  CUNN  YCUNN  YCN

B

P3

Tpm1 exon 3

CCCCACCCCCUCCCCCU
UCCUCCCCCACCACCC

URE6

FAS exon 6

[Kafasla et al Fig1]
Kafasla et al Fig 3