



Review

Reviewing PTBP1 Domain Modularity in the Pre-Genomic Era: A Foundation to Guide the Next Generation of Exploring PTBP1 Structure–Function Relationships

Christine Carico and William J. Placzek *

Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL 35294, USA

* Correspondence: placzek@uab.edu; Tel.: +1-205-975-2465

Abstract: Polypyrimidine tract binding protein 1 (PTBP1) is one of the most well-described RNA binding proteins, known initially for its role as a splicing repressor before later studies revealed its numerous roles in RNA maturation, stability, and translation. While PTBP1's various biological roles have been well-described, it remains unclear how its four RNA recognition motif (RRM) domains coordinate these functions. The early PTBP1 literature saw extensive effort placed in detailing structures of each of PTBP1's RRM, as well as their individual RNA sequence and structure preferences. However, limitations in high-throughput and high-resolution genomic approaches (i.e., next-generation sequencing had not yet been developed) precluded the functional translation of these findings into a mechanistic understanding of each RRM's contribution to overall PTBP1 function. With the emergence of new technologies, it is now feasible to begin elucidating the individual contributions of each RRM to PTBP1 biological functions. Here, we review all the known literature describing the apo and RNA bound structures of each of PTBP1's RRM, as well as the emerging literature describing the dependence of specific RNA processing events on individual RRM domains. Our goal is to provide a framework of the structure–function context upon which to facilitate the interpretation of future studies interrogating the dynamics of PTBP1 function.

Keywords: polypyrimidine tract binding protein 1 (PTBP1); RNA recognition motif (RRM); ribonucleoprotein 1 (RNP1); ribonucleoprotein 2 (RNP2); RNA; structure–function



Citation: Carico, C.; Placzek, W.J. Reviewing PTBP1 Domain Modularity in the Pre-Genomic Era: A Foundation to Guide the Next Generation of Exploring PTBP1 Structure–Function Relationships. *Int. J. Mol. Sci.* **2023**, *24*, 11218. <https://doi.org/10.3390/ijms241311218>

Academic Editor: Wolfgang Sippl

Received: 2 June 2023

Revised: 28 June 2023

Accepted: 6 July 2023

Published: 7 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

While it is now known as a ubiquitous regulator of RNA biogenesis, polypyrimidine tract binding protein 1 (PTBP1)—also known as hnRNP I (heterogeneous nuclear ribonucleoprotein 1) [1]—was initially identified as a sequence-specific splicing factor [1–5]. Early dogma in splicing biology suggested that the 5' splice site of RNA transcripts is recognized solely by sequence complementarity with U1 snRNA (small nuclear RNA) [6]; however, later studies suggested that other sequence-specific proteins were required for accurate and precise 5' splice site selection [7,8]. PTBP1 was one of these identified proteins. It was found to be a component of the spliceosome that binds pre-mRNAs specifically at the polypyrimidine tract, and its binding affinity directly correlated to splicing efficiency [1–5,9]. PTBP1's role in alternative splicing was well characterized for many gene targets [4,9–17] before its critical roles in other biological processes (e.g., polyadenylation [18], mRNA stability and transport [19–21], internal ribosome entry site (IRES)-mediated translation [22–24]) were later described, firmly establishing this protein as a central regulator of RNA biogenesis. One well-studied example is the PTBP1 splicing of its homolog, polypyrimidine tract binding protein 2 (PTBP2), that suppresses PTBP2 expression. In specific tissues, downregulation of PTBP1 occurs during differentiation, and causes the expression of the function PTBP2 mRNA and, ultimately, a PTBP2 protein that modulates a distinct profile of RNA regulation [11]. However, while these biological roles have been well described, the

mechanisms by which PTBP1 coordinates these diverse functions for a broad pool of RNA targets are still poorly understood.

Structurally, PTBP1 is comprised of an N-terminal bipartite nuclear localization and export signal (NLS/NES) [25–28] and four RNA binding domains (RBDs) of the RNA recognition motif (RRM) variety (designated RRM1–4) [29–34]. There are several types of RNA binding domains (e.g., K homology (KH) domain, zinc-finger domain), but the RRM is the most abundant and well-characterized, both structurally and biochemically [29,35–37]. Structural modularity is a common theme in RNA binding proteins (RBPs), as many of these proteins, like PTBP1, are comprised of multiple repeats of a few defined protein domains (e.g., KH domain, RRM, etc.) [37]. For many of these RBPs, studies have shown that these repeating domains can either have an independent function, or cooperate with one another to coordinate the ultimate biological function of the protein. For example, the identity and organization of domains in an RBD can increase RNA binding affinity through additive affinity [38], or create an extended RNA recognition interface via interaction of several domains [39–43]. Additionally, individual domains can act as tethers for adjacent domains, as short linker regions can increase the effective local concentration of a neighboring domain to augment functional affinity [44]. Finally, RNA binding domains (especially RRMs) can serve as protein–protein interaction interfaces that have important impacts on downstream protein function [45–47].

2. The RRM Domains of PTBP1 Are Structurally Heterogenous

As mentioned above, PTBP1 contains four RNA recognition motifs. Canonically, an RRM contains approximately 90 amino acids arranged in an $\alpha\beta$ -sandwich with $\beta 1\alpha 1\beta 2\beta 3\alpha 2\beta 4$ topology [29,35], in which the four anti-parallel β -strands form a β -sheet that is packed against the two α -helices [29,35,48,49]. RNA binding takes place on the β -sheet, and is coordinated by a series of intermolecular interactions (e.g., pi-stacking, hydrophobic, hydrogen-bonds, electrostatic, etc.) [29,35,37] between solvent exposed residues and target RNA. Within the β -sheet, there are two conserved RNA recognition sequences, termed RNP1 (ribonucleoprotein 1) on $\beta 3$ (canonical sequence: [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY]), and RNP2 (ribonucleoprotein 2) on $\beta 1$ (canonical sequence: [ILV]-[FY]-[ILV]-X-N-L), that are positioned to directly interact with target RNA (Table 1) [29,36,50,51]. In the protein–RNA interface, the RNA positions two nitrogenous bases to stack with conserved aromatic rings at position 5 of RNP1 and position 2 of RNP2 (Table 1). In addition, the RNA sugar rings contact a hydrophobic side chain at position 3 of RNP1 (also a conserved aromatic residue), and the phosphodiester group electrostatically interacts with a positively charged residue at position 1 of RNP1 (Table 1). Of note, although RRMs contain these conserved sequences, most do not make all four of these canonical contacts with target RNA dinucleotides, and typically only make between one and three of these interactions. RNA target specificity is further mediated by intermolecular interactions (e.g., hydrophobic interactions, hydrogen bonds, etc.) between target RNA and residues outside of these conserved sequences (especially on the $\beta 4$ and $\beta 2$ strands, as well as the N- and C-terminal regions), similar to an induced fit ligand–protein interaction.

PTBP1's four RRMs are arranged in a “beads on a string” orientation, where each RRM is joined to its neighbor by linker regions of variable length. The linker regions following RRM1 (42 residues) and RRM2 (58–84 residues) are long, and these domains have been shown to move independently of their neighbor in solution [1,5,30,33,52,53]. However, the linker region between RRM3 and RRM4 is short (24 residues [31,53]), and these domains interact with one another via their α -helical interface to form an intramolecular didomain complex with both β -sheets facing opposing directions [32,33]. This organization of tandem RRMs is unusual, as only a handful of RNA binding proteins contain RRMs that demonstrate stable intramolecular contacts [33,54–59], and even fewer make these stable contacts via their respective α -helices (hnRNP A1 [60,61], hnRNP L [59], PTBP1/PTBP2 [33]). Interestingly, the RRMs of PTBP1 also demonstrate significant divergence from the canonical RRM primary sequence and topologic organization [1,5,35,36]. Specifically, in each RRM of

PTBP1, key RNA-interacting aromatic residues in the consensus RNP1 and RNP2 sequences are often replaced by hydrophobic residues (discussed in detail below) [1,36,62].

Table 1. RNP1 and RNP2 sequences for all four RRMs of PTBP1. Conserved aromatic residues in the consensus and PTBP1 RRM sequences are shown in red. Substitutions of these residues in each PTBP1 RRM are in blue. Note that the RNP2 sequence of RRM2 (IIVENL) is shifted by two residues in RRM2's structure due to an insertion in the β 1 strand of RRM2.

Position	RNP2						RNP1									
	1	2	3	4	5	6	1	2	3	4	5	6	7	8		
Consensus sequence	L	F	V	G	N	L	K	G	F	G	F	V	X	F		
	I	Y	I	K	L	L	R	G	Y	A	F	V	X	Y		
RRM1 sequence	I	H	I	R	K	L	K	N	Q	A	F	I	E	M		
RRM2 sequence	L	R	I	I	V	E	N	L	Q	F	Q	A	L	L	Q	Y
RRM3 sequence	L	L	V	S	N	L	K	E	N	A	L	Q	V	M		
RRM4 sequence	L	H	L	S	N	L	R	K	M	A	L	I	Q	M		

PTBP1 was initially described to bind short, single-stranded pyrimidine sequences (such as UCUU, CUCUCU), including those found at splice junctions [10,11,63,64], and later structural interrogation using short oligos determined that each individual RRM demonstrates sequence context preferences using short oligos (in terms of sequence length, composition, and secondary structure). Here, we review the apo structure of each RRM of PTBP1, and the key residues involved in RRM: RNA interactions, and sequence and structure preferences of target RNA for each RRM. As RRM3 and RRM4 invariably exist as a didomain complex, they will be discussed together (and referred to as RRM3-4). We would like to note that there is extensive evidence that multiple RRMs (and even PTBP1 proteins) are involved in interactions with target RNA in the *in vivo* setting; however, to our knowledge, there are no detailed structural models of the full PTBP1 protein complexed with RNA. Therefore, the information summarized below reflects RRM: RNA interactions between each RRM (or RRM complex in the case of RRM3-4) and its minimal binding register.

2.1. RNA Recognition Motif 1 (RRM1)

RRM1 of PTBP1 demonstrates the canonical RRM topology of β 1 α 1 β 2 β 3 α 2 β 4 [30,65]; however, as introduced above, its RNP1 and RNP2 sequences reflect several notable deviations from the consensus sequence, namely the substitution of key aromatic residues [36,66] (Table 1, Figure 1a). Early work to characterize the structure of RRM1 bound to RNA revealed that RRM1, as was observed for the full length PTBP1 protein, binds short pyrimidine stretches [30,31,53,67]. Utilizing a short pyrimidine sequence (C1U2C3U4 [31,53]), specific RRM contacts with RNA were mapped, identifying the minimal binding register and mode by which it is coordinated on the RRM1 surface. RRM1 binds U2C3U4 and, consistent with canonical RRM:RNA interactions, RRM1 binds the C3 nucleotide by a pi-stacking interaction with H62 [68] (position 2 of RNP2). Notably, this H62 is a non-canonical substitution of the conserved aromatic residue of RNP2 at position 2 (Table 1); however, its planar aromatic ring can engage in stacking interactions with nitrogenous bases, thereby retaining the functional properties of this conserved position in RNP2. This interaction is further stabilized by a hydrogen bond with the main chain of N132 [68,69], and side chains of F130 and S131 [31,53,69]. These residues are located on β 4 outside of the consensus RNP1 and 2 sequences and, thus, are likely contributors to the induced fit specificity of this interaction. At the 3' end of this C3 nucleotide, the U2 nucleotide sits above β 4, is H-bonded to the side chain of Q129 (via its O2), and stacks with the side chain of R64 (position 4 of RNP2). At the 5' end of the C3 nucleotide, the U4 nucleotide sits above β 2 in a hydrophobic pocket formed by five protein side chains: L136 [68,69] (C ter-

minus), H133 [68] (β 4), F98 (β 3—RNP1 position 5), L91 [68] (β 2), and L89 [68] (β 2) [30,31] (Figure 1b). Importantly, subsequent studies have confirmed the involvement of many of these residues in binding RNA with different structural context (pyrimidine sequence in a loop region of IRES elements), and these are referenced above [68,69]. These studies identified additional RRM:RNA contacts, and it is thought that this is due to the presence of multiple distinct binding registers, as well as the structure of the RNA itself. Of note, of the two conserved aromatic residues in the consensus RNP1 sequence, the F98 residue at position 5 is the only retained aromatic, although rather than engaging in a stacking interaction as in most other RRMs, it contributes to hydrophobic pocket formation and, thus, the induced fit specificity of this interaction (Table 1).

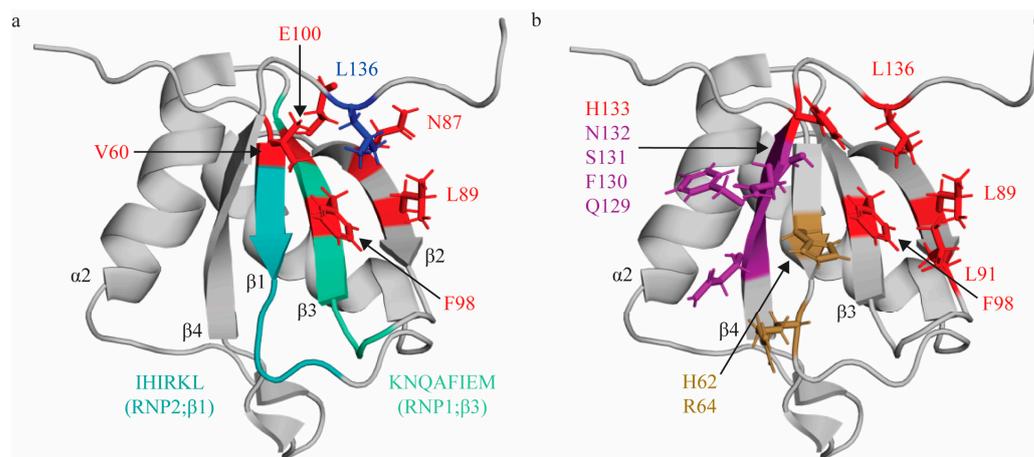


Figure 1. Topologic organization and key features of RRM1. (a) Apo structure of RRM1 (PDB: 1SJQ [30]) with residues comprising the RNP1 sequence on β 3 shaded green (K94, N95, Q96, A97, F98, I99, E100, M101) and residues of RNP2 on β 1 shaded in teal (I61, H62, I63, R64, K65, L66). C terminal residue (L136; blue) is stabilized by hydrophobic contacts with several residues across the β -sheet (V60, L89, F98, N87, E100; red). (b) Key residues that interact with the minimal RNA binding register YCN. Residues that make stacking interactions (H62, R64) are colored sand yellow. Residues that make hydrogen bonds (Q129, F130, S131, N132) are colored purple. Note that N132 also makes a stacking interaction with the C₃ nucleotide, but is colored based on hydrogen bond in this figure. Residues that engage in hydrophobic interactions (L89, L91, F98, H133, L136) are colored red.

Based on these data, the preferred binding register for RRM1 is YCU with only the C nucleotide engaging in a canonically described interaction with conserved features of the RNA binding interface. The first position of this binding register can be occupied by either pyrimidine (C or U), because Q129 on β 4 can act as either a hydrogen bond donor or acceptor, and can thus accommodate either pyrimidine [31]. Of note, although these structural studies mapped RRM:RNA contacts with a single-stranded tetranucleotide, several studies have demonstrated that RRM1 (and RRM2, as it was studied as a PTBP1 subdomain containing both RRM1 and RRM2) preferentially binds loop structures [67,70,71].

Intriguingly, several recent studies have demonstrated that the C-terminal region of the RRM1 domain is a critical allosteric regulator of RNA binding [68,72]. These studies identified a C-terminal α -helix (termed the α 3 helix) comprised of residues 144–154 that folds upon binding to the stem-loop region of the encephalomyocarditis virus (ECMV) IRES element, but does not directly interact with RNA. This α 3 helix is thought to serve as sensor of RNA secondary structure, and acts as an allosteric regulator of RNA binding—a phenomenon seen in other RRM-containing proteins [39–42]—suggesting that the C-terminal region of RRM1 is not simply displaced, but plays a significant role in allosterically regulating RNA binding in a structure-dependent context. This provides significant context for the studies that demonstrated that RRM1 preferentially binds loop structures [67,70,71].

2.2. RNA Recognition Motif 2 (RRM2)

Unlike RRM1, RRM2 demonstrates an extension of the canonical RRM topology with an additional fifth β -strand that sits adjacent and anti-parallel to β 2, stabilized by a stacking interaction between Y275 (β 5) and H201 (α 1 helix) [30,73]. β 4 and β 5 are connected by a long loop region (13 residues) that sits in a hydrophobic pocket created by several residues on the β -sheet interface (Figure 2a) [30]. Additionally, residues Y267–N269 are arranged in a “pseudo-sixth strand” configuration anti-parallel to β 5, with residues beyond N269 folding back to contact the β 4– β 5 loop [30] (Figure 2a). This β 1 α 1 β 2 β 3 α 2 β 4 β 5 topology with an additional pseudo-sixth β -strand extends the β -sheet RNA binding interface, as compared to canonical RRMs. In addition, the β 1, β 3 and β 4 strands are longer than in most RRM domains; taken together, these modifications significantly expand the available RNA binding interface [30].

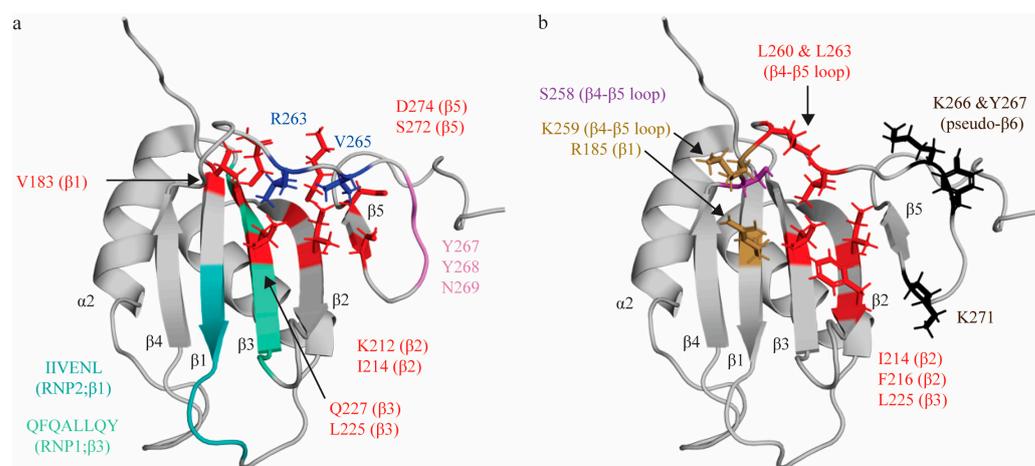


Figure 2. Topologic organization and key features of RRM2. (a) Apo structure of RRM2 (PDB: 1SJR [30]) with residues comprising the RNP1 sequence on β 3 shaded green (Q221, F222, Q223, A224, L225, L226, Q227, Y228) and residues of RNP2 on β 1 shaded teal (I186, I187, V188, E189, N190, L191). C-terminal residues (R263, V265; blue) are stabilized by hydrophobic contacts with several residues across the β -sheet (V183, I214, L225, K212, Q227, S272, D274; red). Residues Y267, Y268 and N269 form a pseudo- β 6 strand (pink). (b) Key residues that interact with the minimal RNA binding register CU(N)N. Residues that make stacking interactions (R185, K259) are colored in sand yellow. Residues that make hydrogen bonds (S258) are colored purple. Note that the main chain of K259 also forms an H-bond, but is colored based on stacking interaction in this figure. Residues that engage in hydrophobic interactions (I214, F216, L225, L260, L263) are colored red. Residues with undefined contacts with RNA (K66, Y267, K271) are colored black.

Studies characterizing RRM2's interactions with various short pyrimidine sequences (C1U2C3U4, C1U2C3U4C5, C1U2C3U4C5U6) identified that RRM2 binds the C3U4 doublet, as well as U6 [31,53]. As with RRM1, the C3 nucleotide is sandwiched between two sidechains: R185 on β 3, and K259 on β 4- β 5 loop. Intriguingly, while R185 is not part of RNP2, sequence alignment reveals that the R185 residue sits in the location that is canonically occupied by position 2 of RNP2, as the conserved RNP2 sequence of RRM2 is shifted by two positions second to an insertion in the β 1 strand. Ultimately, although the functional properties of the residue and mechanism of interacting with target RNA is not retained, the position of the interacting residue is conserved (Figure 2a). As seen in RRM1, the U4 nucleotide is bound in a hydrophobic pocket (created by side chains I214 (β 2), F216 (β 2), L225 (β 3—RNP1 position 5), L260 and L263 (both in the β 4- β 5 loop), with both bases being sequence-specifically recognized by both the main chain of the protein and by S258 (Figure 2b). A subsequent study demonstrated that mutation of the I214 and F216 (as well as K218) residues robustly ablated binding to target sequence within the ECMV and poliovirus-1 (PV-1) IRESs, confirming the role of these residues in RNA binding,

even in different structural contexts [69]. L225 is at position 5 of RNP1, which canonically contains an aromatic residue that stacks with RNA bases. However, as seen in RRM1, the residue at this position contributes to hydrophobic pocket formation rather than aromatic stacking. RRM2 does not contact the immediately adjacent nucleotide (C5) and, instead, the U6 nucleotide is in contact with K266, Y267 (located immediately adjacent to and in the pseudo-sixth strand, respectively), and K271 (which is beyond the pseudo-sixth strand proximal to β 5) (Figure 2b) [31,53].

Based on these data, the minimal binding register of RRM2 is CU(N)N, which is extended by an additional nucleotide as compared to the minimal binding register for RRM1 due to the U6 interaction with the C-terminal extension of this domain (the β 5 and pseudo-sixth strand) not seen in RRM1. An additional study using longer sequences than the minimal binding register has also confirmed that there are significant chemical shift perturbations upon RNA binding in the β 4- β 5 loop, the β 5 strand, and the pseudo-sixth strand, as reported above [30]. Although these structural studies demonstrate that RRM2 canonically binds single-stranded RNA, as with RRM1, subsequent studies have suggested RRM2 preferentially binds pyrimidine-rich sequences in loops and bulges [67,70,71,74].

2.3. RRM2-RRM3 Linker

The linker region C-terminal to RRM2 (between RRM2 and RRM3) has important structural considerations. Notably, this linker region appears to be arranged in a relatively compact globular conformation, as compared to the RRM1-RRM2 linker, such that RRM2 and RRM3 are in closer proximity than RRM2 and its N-terminally adjacent domain RRM1 [52]. Alternative 3' splice site selection at exon 9 of this linker region produces three isoforms of PTBP1 (listed from shortest to longest): PTB1, PTB2, and PTB4 [1,5,75]. PTB1 is the shortest isoform, with exon 9 completely excluded from the final transcript, whereas PTB2 contains a truncated exon 9 (and thus an additional 19 residues in this linker region), and PTB4 contains the full exon 9 (an additional 26 residues in this linker region) [76]. This difference in RRM2-RRM3 linker length appears to have important biological implications, as these isoforms have cell-type specific expression [77], and have been shown to have differential splicing repression activity on α -tropomyosin exon 3 (but not α -actinin, another known PTBP1 target) and IRES-driven translation of human rhinovirus-2 [78–80]. The differing biological enrichments and activities of these PTBP1 isoforms that differ only at this linker region indicates that its length and conformation have important functional consequences.

2.4. RNA Recognition Motifs 3 and 4 (RRM3-4)

Note that all residue numbers correspond to the longest isoform (PTB-4) that contains the entirety of exon 9 (and, thus, the full linker region between RRM2 and RRM3). RRM3 and RRM4 of PTBP1 (RRM3-4), like their N-terminal neighbors, demonstrate several non-canonical modifications. RRM3, like RRM2, contains a β 5 strand that sits antiparallel to β 2, and is connected to β 4 by a long loop that is positioned above the RNA binding interface. RRM4, like RRM1, contains the canonical RRM topology, with four anti-parallel β -strands comprising the β -sheet [31–33]. However, unlike RRM1 and RRM2, which are followed at their C-termini by long linker regions that allow their independent movement in solution, RRM3-4 domains are separated by a short linker region of only 25 residues. There are extensive interdomain contacts between the α -helical interfaces of each domain that positions them with a fixed orientation relative to one another, in which their β -sheets point in opposing directions (Figure 3). This large interdomain interface is largely hydrophobic, and involves both RRM-RRM and RRM-linker contacts. The α 2 helix of RRM4 is positioned perpendicularly to the α 1 helix of RRM3, and interacts with the α 2- β 4 loop of RRM3 (see Figure 3 for specific residue interactions). A large portion of the interdomain linker contributes to interdomain interactions, and a majority of these are with α 1 and α 2 of RRM3, with two additional contacts with α 2 and β 4 of RRM4 (see Figure 3 for specific residues). While these interdomain contacts (both between RRMs and between

the linker region and either RRM) are largely hydrophobic, the interaction is also stabilized by an ion pair (K424 on RRM3 and E528 on RRM4) [33,66] (Figure 3). Notably, mutation of three side chain residues on the $\alpha 2$ helix of RRM4 (E528, V531 and I535) was sufficient to ablate this interdomain interaction, indicating that this region of RRM4 is particularly critical in forming the interdomain interface [31].

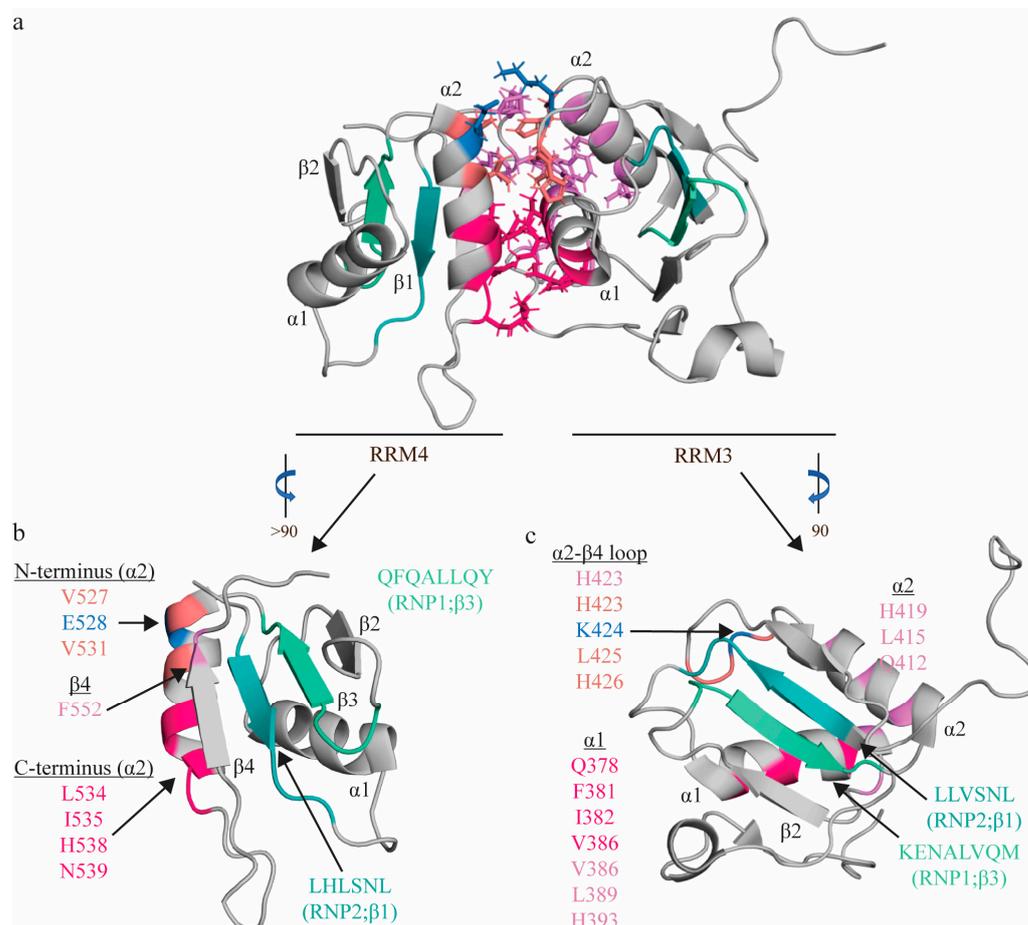


Figure 3. Topologic organization and key features of apo RRM3-4. Apo structure of the RRM3-4 didomain (PDB: 2EVZ [33]) (a), with individual views of RRM4 (b) and RRM3 (c). Residues comprising the RNP1 sequences on $\beta 3$ shaded green (RRM3: K400, E401, N402, A403, L404, V405, Q406, M407; RRM4: R517, K518, M19, A520, L521, I522, Q523, M524) and residues of RNP2 on $\beta 1$ shaded teal (RRM3: L365, L366, V367, S368, N369, L370; RRM4: L482, H483, L484, S485, N486, L487). The C-terminal region of RRM4 $\alpha 2$ (L534, I535, H538, N539; hot pink) interacts with L461 (hot pink) of the interdomain linker, as well as with the N-terminal region of RRM3 $\alpha 1$ (Q378, F381, I382, V386; hot pink), positioning these helices perpendicularly to one another. The N-terminal region of RRM4 $\alpha 2$ (V527 (salmon), E528 (blue), V531 (salmon)) interacts with the $\alpha 2$ - $\beta 4$ loop of RRM3 (H423 (salmon), K424 (blue), L425 (salmon), H426 (salmon)). Most contacts between the interdomain linker (F464, F472, I475, P477, and P478; not individually identified in this figure) are between $\alpha 1$ (V386, Y387, D389; pink) and $\alpha 2$ of RRM3 (Q412, L415, H419; purple), along with H423 (salmon) of the following loop, with an additional contact with RRM4 (F552 of $\beta 4$ of RRM4; purple). Finally, this interaction is stabilized by an ion pair (K424 on RRM3 and E528 on RRM4; blue). Note that colors correspond to residues that interact with one another, and residues with multiple contacts (V386, H423) are reported twice in the appropriate colors.

Due to the heterodimeric form adopted by RRM3-4, RRM: RNA interactions for the individual RRM3 and RRM4 domains were determined using the RRM3-4 didomain, in order to provide a more accurate structural interrogation of RNA interactions. As with RRM1 and RRM2, short pyrimidine sequences (C1U2C3U4C5U6) were utilized to identify

the minimal binding register of each tandem domain and map specific interactions [31,53]. RRM3 interacts with the U2C3U4C5U6 quintet (with its minimal binding register defined as YCUNN), and this longer binding register, as compared to either RRM1 or RRM4 (discussed below), is facilitated by the extended β -sheet. While there is no aromatic residue at RNP2 position 2, as with all other RRM of PTBP1, the C3 nucleotide still sits above this canonical position on the β -sheet of RRM3 (occupied in RRM3 by L366 [81]), and is coordinated by two hydrogen bonds from residues on the β 4 strand (S435 [82] and K436) (Figure 4). The two uracil nucleotides flanking this C3 nucleotide (U2 and U4) are coordinated by a series of hydrogen bonds and hydrophobic interactions. U2 is H-bonded to T433, and sits above R431 on β 4. U4 hydrogen bonds with β 4- β 5 loop residues H437 [82] and V440, and forms hydrophobic interactions with β 3 residues K394, L396 [81] and L404 [81] (position 5 of RNP1) (Figure 4). Notably, these three nucleotides, which also make up the minimal binding register of RRM1, are coordinated on RRM3 in a similar manner as RRM1. The C-terminal C5 nucleotide is extensively coordinated, and its base interacts with L396 and stacks on F397, while its sugar interacts with L452, and its phosphate oxygen contacts K394 (Figure 4). A subsequent study demonstrated that mutation of K394, L396, and F397 significantly ablated RRM3 binding to ECMV and PV-1 IRESs, confirming the importance of these residues in driving target RNA binding [69]. U6 is less specifically coordinated [82], and is contacted by P443 and R444, while its phosphate oxygen also contacts K394 (Figure 4). Note that all residues are derived from Oberstrass et al. (2005) [31] and Auweter et al. (2007) [53], while additional references reflect subsequent studies confirming these interactions.

RRM4, like RRM1, binds the U2C3U4 triplet in a similar fashion but with several key differences (hence its minimal binding register—YCN—is slightly different from RRM1). Like RRM1, C3 stacks on H483 [82] (RNP2 position 2), and is H-bonded to two residues on the β 4 strand (S553 [82] and K554 [82]) (Figure 4). A later study confirmed that mutation of S553 and K554 significantly reduced binding to ECMV and PV-1 IRESs, further confirming the importance of these residues in RNA binding [69]. However, in contrast with RRM1, U2 is not bound as tightly, and stacks and forms a hydrogen bond with the N474 side chain (interdomain linker) (Figure 4). Finally, the U4 nucleotide is coordinated in a hydrophobic pocket formed by L521 (RNP1 position 5), F513 and the aliphatic portions of K511 [81] and K515 [81] (Figure 4). Again, in contrast with RRM1, this U4 nucleotide is flipped 180 degrees, as the shorter C-terminal extension only allows one hydrogen bond with this nucleotide (I557 [81]), rather than the two seen in RRM1 (Figure 4). Intriguingly, a later study modeling changes in domain backbone dynamics upon formation of the RRM3-4 complex have indicated that there are changes in backbone dynamics upon association of RRM3 and RRM4 to form RRM3-4 (as compared to the individual domains) that could be important factors in ligand (e.g., RNA) selection [66]. Specifically, RRM3 appears to lose conformational flexibility upon association with RRM4.

The organization of these RRM (juxtaposed so that their β -sheets face in opposing directions) is unique among RNA binding proteins with tandem RRMs, and has important functional implications. All other RNA binding proteins with tandem domains organize these domains such that they bind immediately adjacent stretches of nucleotides on target RNA [39–41]. The organization of RRM3-4 requires separation of the binding register for each domain by at least 15 nucleotides [31], thus RRM3-4 preferentially binds longer single-stranded RNA [67,70]. Importantly, this complex has been shown to facilitate induction of RNA looping [31,83], which is particularly relevant when considering PTBP1's roles in alternative splicing and IRES-mediated translation, as both processes require RNA looping (either for exon exclusion or ribosomal binding, respectively).

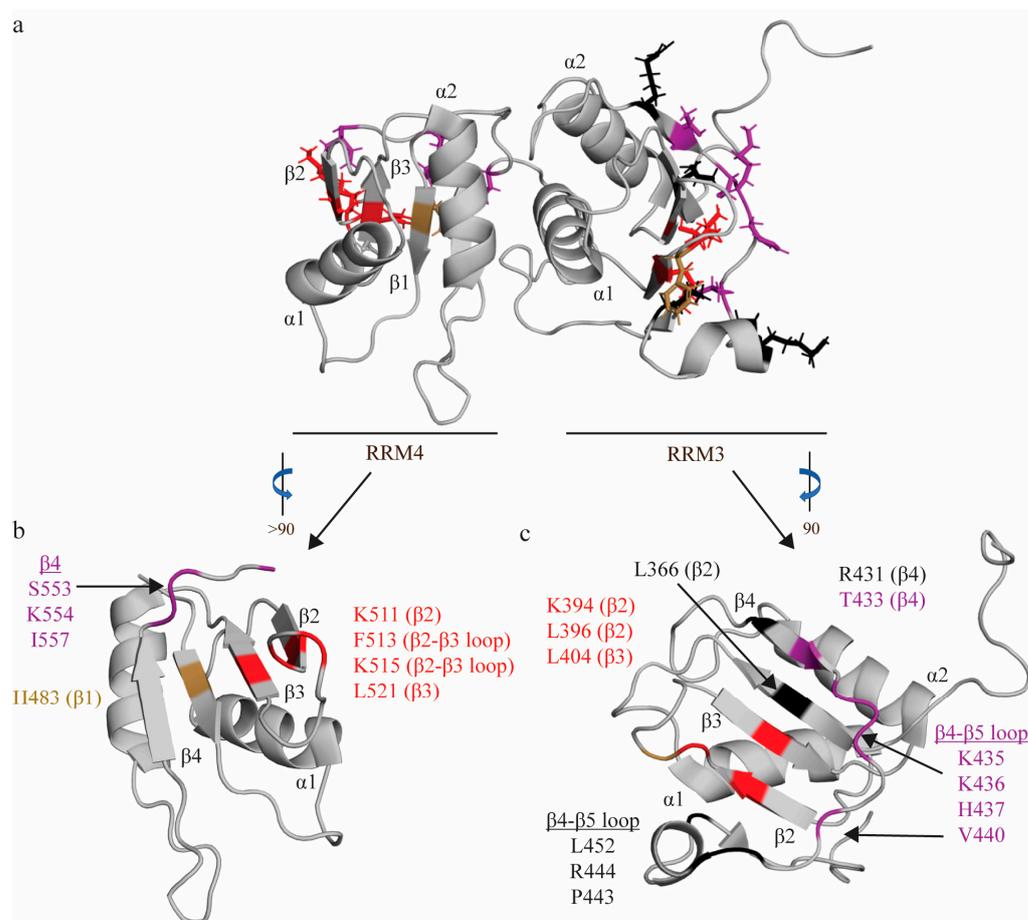


Figure 4. RNA interactions of RRM3-4. (a) Structure of RRM3-4 didomain (PDB: 2EVZ [33]) with key residues that interact with the minimal binding register of RRM4 (YCN) and RRM3 (YCUUN), highlighted in colors that reflect the nature of their chemical interaction with RNA nucleotides. Residues that make stacking interactions are colored sand yellow, residues that engage in hydrophobic interactions are colored red, and residues that make hydrogen bonds are colored purple (see descriptions below). (b) Key residues of RRM4 that interact with the minimal RNA binding register (YCN). Residue H483 makes a stacking interaction. Residues N448, S553, K554, I557 make hydrogen bonds. Residues K511, F513, K515, L521 engage in hydrophobic interactions. (c) Key residues of RRM3 that interact with the minimal RNA binding register (YCUNN). Residue F397 makes a stacking interaction. Residues T433, S435, K436, H437, V440 make hydrogen bonds. Residues K394, L396, L404 engage in hydrophobic interactions. Residues L366—RNP2 position 2, L396, R431, L452, P443, and R444 have undefined contacts with RNA and are colored black.

2.5. Summary

In all RRMs of PTBP1, although there are substitutions of aromatic amino acids at key RNA interacting positions within RNP1 and RNP2, the residues at these conserved positions still contact target RNA as reflected in seminal studies mapping RNA: RRM interactions [31,53]. RNP2 position 2 and RNP1 position 5 of all four RRMs contact target RNA, as demonstrated in the studies referenced above [31,33,53,68,81,82]. Specificity of RNA: RRM interactions arise from the H-bond and hydrophobic pocket networks that coordinate each respective binding register, as well as the unique topologies between RRMs (extra $\beta 5$ on RRM2 and RRM3) [29,35]. Each RRM has a different minimal binding register, different secondary structure preference and, in the case of RRM3-4, unique ability to manipulate the structure of RNA [31].

3. The RRM Domains of PTBP1 Have a Unique Contribution to PTBP1 Function

The idea of separation of function among the RRM domains of PTBP1 was proposed in the early PTBP1 literature, and has been refined by numerous subsequent studies over the roughly four decades since PTBP1's initial discovery. These early PTBP1 studies had suggested that RRM3 and RRM4 (RRM3-4) were the major drivers of RNA binding (particularly RRM3 [28,32]), while RRM1 and RRM2 served as protein interacting domains and had little role in RNA binding [23,28,84]. Indeed, Oh et al. had designated RRM1 as a HeLa factor responding domain, for its apparent ability to enhance RNA binding through interaction with HeLa cell cytoplasmic factors [84]. This study and another [28] also suggested that RRM2 drove PTBP1 dimerization rather than RNA binding—a phenomenon that was later found to be an artifact of the size exclusion chromatography methodology [30,52,85,86].

Subsequent biochemical studies revealed that all four RRMs of PTBP1 bind RNA, demonstrating that the separation of functions within this protein is not as clearly delineated as “protein interacting” domains and “RNA binding” domains. Instead, these studies suggest that separation of functions within the PTBP1 protein is a result of unique sequence and structure preferences of each RRM for target RNA. While all four RRMs of PTBP1 utilize the residues at RNP2 position 2 (or in the case of RRM2, the canonical position) and RNP1 position 5 to coordinate RNA, the network of hydrogen bonds and hydrophobic interactions that drive the specificity of RRM: RNA interactions are unique between the RRMs. Therefore, as described in the previous sections of this review, each RRM has a unique binding register, secondary structure preference and, especially in the case of RRM3-4, the ability to manipulate the structure of target RNA. This, combined with the solution behavior of the full PTBP1 protein (RRM1 and RRM2 move independently in solution due to long flexible C-terminal linker regions, whereas RRM3-4 is in a constitutive didomain conformation), supports the idea that each RRM may provide unique contributions to independently modulate RNA specificity and, ultimately, PTBP1 function. In this section, we summarize the current evidence for domain-specific functions of each of PTBP1's RRM domains (or didomain in the case of RRM3-4).

3.1. RNA Recognition Motif 1 (RRM1)

Of the four RRMs of PTBP1, RRM1 has the most numerous examples in the literature of discrete domain-specific functions. As described above, a C-terminal $\alpha 3$ helix folds upon binding to an RNA hairpin in the ECMV IRES, and this helix was critical for PTBP1-mediated enhancement of IRES activity in vitro [68]. While this has not yet been explored in vivo, this study and another [72] have revealed a potential mechanism of domain-specific RRM1 function in PTBP1-dependent IRES initiated translation, the functional implications of which are important to elucidate in an in vivo setting.

It is known that PTBP1 regulates multiple aspects of the biogenesis of several apoptotic genes [87–90]. RRM1, specifically, has been shown to be important for maintaining the expression of Caspase-9, BAX, and BID in differentiating cardiomyocytes, as deletion of this domain resulted in a decreased expression of these transcripts. Zhang and colleagues further demonstrated that RRM1 plays a role in regulating exon 10 inclusion in its homolog PTBP2, as truncation of the protein to exclude RRM1 and the N-terminal NLS resulted in increased exon 10 skipping in PTBP2 [91]. This is particularly biologically relevant given that this is the mechanism by which PTBP1 represses PTBP2 signaling: exon 10 skipping leading to nonsense mediated decay of the PTBP2 transcript [92].

Finally, RRM1 has also been shown to have a critical function in regulating genes that control invasive potential in cancer. A study by Wang et al. demonstrated that PTBP1 binds the 5' UTR (untranslated region) of the hypoxia inducible factor 1 α (HIF-1 α) transcript via its RRM1 and RRM3 domains, destabilizing HIF-1 α and promoting an invasive phenotype in a non-small cell lung cancer (NSCLC) model [93]. More recently, this group also demonstrated that PTBP1 negatively regulates the AXL tyrosine kinase transcript by binding its 5' UTR, resulting in reduced invasive potential, and that RRM1 is critical for this binding interaction [94]. Although RRM1 has conflicting effects on invasive potential

in these two model systems, these studies demonstrate that RRM1 can impact cellular phenotype through regulation of distinct RNA targets. Importantly, supplementing this growing literature, we have recently shown that RRM1 contains a reverse Bcl-2 homology domain 3 (rBH3) motif regulatory sequence. This sequence allows the anti-apoptotic protein, MCL1, to displace target RNA from RRM1, establishing a mechanism by which these emerging functions of RRM1 can be independently regulated [95].

3.2. RNA Recognition Motif 2 (RRM2)

It has been well described that RRM2s can serve as protein–protein interaction interfaces in addition to their canonical RNA binding function [35,62,96]. In the early PTBP1 literature, RRM2 was initially thought to be a protein interacting domain of PTBP1, with a specific role in mediating oligomerization of PTBP1 [28,84]. Although later studies revealed that PTBP1 was a monomer in solution and did not dimerize via its RRM2 domain, the concept that RRM2 can serve as a protein–protein interaction interface in addition to its RNA binding properties was demonstrated in subsequent studies. Raver1 was initially identified as a PTBP1 binding partner via yeast two-hybrid screen [97], and it was later shown that this interaction with PTBP1 was necessary for repression of exon 3 in α -tropomyosin [98]. Subsequent studies by the same group identified that RRM2 and the adjacent linker region were the minimal repressor domain for α -tropomyosin exon 3 repression [99], and that Raver1 interacts with a hydrophobic groove on the α -helical surface of RRM2 via a defined peptide motif ([S/G][I/L]LGxPP) [100]. Of note, this motif is also present on Raver2 and Matrin3, suggesting that these proteins could also play a role in regulation of other RRM2-dependent functions [100,101]. Additionally, Kafasla et al. demonstrated that deletion of RRM2 ablated ECMV IRES activity (but not poliovirus IRES activity), indicating that interrogation into impact of RRM2 protein–protein interactions on IRES function is necessary [70].

3.3. RNA Recognition Motifs 3 and 4 (RRM3-4)

While the three isoforms of PTBP1 formed by alternative splicing at exon 9 are well described, there is also a fourth PTBP1 isoform—formed by the removal of exons 2–10—that contains only RRM3 and RRM4 [1,3,5,75,102]. The presence of an endogenous independent RRM3-4 didomain suggests these two RRM2s have biologically relevant independent functions. Both the hepatitis A virus [103] and poliovirus 3C [104] proteinases cleave PTBP1 at the linker region between RRM2 and RRM3 to release the RRM3-4 didomain, further underscoring the independent biological function of this didomain.

As described above, the geometric orientation of these domains (in which their RNA binding interfaces face in opposite directions) is well poised to induce RNA looping and, indeed, prior studies have confirmed this ability of the RRM3-4 didomain [31,83]. It has been suggested that this RNA looping function is important in processes such as alternative splicing (looping out of a repressed exon) and IRES-mediated translation (looping to create a ribosomal binding site). Accordingly, several studies have demonstrated a role of this didomain in regulation of alternative splicing. One study revealed that RRM4 was required for switching from a neural to non-neural splicing pattern in a C28 splicing assay, as mutation of several key RNA-interacting residues of RRM4 uncoupled RNA binding from splicing activity [105]. In another study, RRM3 and RRM4 were found to be required to mediate exon 11 repression of PTBP1 itself in a negative feedback loop, as a truncated protein expressing just the RRM1 and RRM2 domains failed to repress exon 11 [106]. A similarly truncated PTBP1 protein (containing only RRM3 and 4) was also found to have reduced splicing capacity of a cardiac troponin exon 5 minigene reporter in vivo [107]. Finally, RRM3-4 function was found to be essential for splicing repression in a splicing reporter assay (using the DS9-175 minigene) in vivo, and this was likely due to its ability to induce RNA looping [83].

4. Perspectives and Conclusions

Since its initial identification in 1988, PTBP1 has taken a center stage as a model protein for studying RBP biology and function. Its roles in virtually all stages of RNA maturation and utilization (e.g., 3' and 5' end processing, splicing, IRES translation) have provided a rich substrate for study, as by these ubiquitous functions PTBP1 impacts a wider range of biological processes than any other RBP. It has been described as a critical regulator of gene expression homeostasis in various tissues, has been implicated in regulating development in multiple tissue types, and its dysregulation has been shown to drive phenotypes in multiple disease states including, but not limited to, various cancers.

Despite PTBP1's position as a main character in the field of RBP biology, there has been little description of the dynamic anatomy of this protein and how its various components (here, individual RRM) coordinate its diverse functions. In the early PTBP1 literature, the available genomic techniques did not allow for high-resolution and throughput analyses in cellular-based assays to elucidate PTBP1's intra-protein functional dynamics in an endogenous environment. However, recent advances in genomic technology now provide a toolbox to answer these previously unanswered questions. The advent of CRISPR-based technology now allows the capability of deleting portions of or entire protein domains to interrogate their biological function endogenously. More efficient and cost-efficient immunoprecipitation and sequencing techniques allow higher resolution mapping of binding sites and even interrogation of RNA structure, when it is bound to proteins of interest. These approaches allow us the enhanced molecular resolution necessary to begin mapping modular roles of individual domains of RBPs on a target-specific basis.

The current lack of knowledge of the dynamics of RBP—and specifically PTBP1—coordination of functions has rendered these proteins essentially “undruggable” and, therefore, most effort in therapeutic development has focused on interventions downstream of RBPs—e.g., targeting the dysregulated RNA or protein product. Deconstructing the modularity of RBPs will allow more targeted upstream interventions of dysregulated RBP function. For example, if a particular domain of an RBP drives a pathologic splicing event—e.g., RRM2 and α -tropomyosin—a small molecule inhibitor can be designed for the specific ligand (RNA): RRM that does not impact the function of the RBP at other targets. Our goal is that the information contained in this review will frame the structure–function context of each RRM of PTBP1 and facilitate the next generation of PTBP1 interrogation—moving from studying the macroscopic cellular function of the entire protein, to a higher-resolution view of how the protein coordinates these observed functions. Ultimately, we hope that this will provide a foundation for understanding how individual RBD domains (and their specific RNA interactions) can be targeted to treat developmental defects and disease.

Author Contributions: Conceptualization, C.C. and W.J.P.; formal analysis, C.C.; funding acquisition, W.J.P.; writing—original draft preparation, C.C. and W.J.P.; writing—review and editing, C.C. and W.J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported, in part, by funding from the National Institutes of Health Grants R01GM117391 (to W.J.P.) and National Institutes of Health Grants T32GM008361 (to C.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ghetti, A.; Pinol-Roma, S.; Michael, W.M.; Morandi, C.; Dreyfuss, G. hnRNP I, the polypyrimidine tract-binding protein: Distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.* **1992**, *20*, 3671–3678. [[CrossRef](#)] [[PubMed](#)]
2. Garcia-Blanco, M.A.; Jamison, S.F.; Sharp, P.A. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes. Dev.* **1989**, *3*, 1874–1886. [[CrossRef](#)]
3. Patton, J.G.; Mayer, S.A.; Tempst, P.; Nadal-Ginard, B. Characterization and molecular cloning of polypyrimidine tract-binding protein: A component of a complex necessary for pre-mRNA splicing. *Genes. Dev.* **1991**, *5*, 1237–1251. [[CrossRef](#)]

4. Mullen, M.P.; Smith, C.W.; Patton, J.G.; Nadal-Ginard, B. Alpha-tropomyosin mutually exclusive exon selection: Competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes. Dev.* **1991**, *5*, 642–655. [[CrossRef](#)] [[PubMed](#)]
5. Gil, A.; Sharp, P.A.; Jamison, S.F.; Garcia-Blanco, M.A. Characterization of cDNAs encoding the polypyrimidine tract-binding protein. *Genes. Dev.* **1991**, *5*, 1224–1236. [[CrossRef](#)]
6. Padgett, R.A.; Grabowski, P.J.; Konarska, M.M.; Seiler, S.; Sharp, P.A. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **1986**, *55*, 1119–1150. [[CrossRef](#)]
7. Seraphin, B.; Kretzner, L.; Rosbash, M. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *EMBO J.* **1988**, *7*, 2533–2538. [[CrossRef](#)]
8. Mayeda, A.; Tatei, K.; Kitayama, H.; Takemura, K.; Ohshima, Y. Three distinct activities possibly involved in mRNA splicing are found in a nuclear fraction lacking U1 and U2 RNA. *Nucleic Acids Res.* **1986**, *14*, 3045–3057. [[CrossRef](#)] [[PubMed](#)]
9. Mulligan, G.J.; Guo, W.; Wormsley, S.; Helfman, D.M. Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. *J. Biol. Chem.* **1992**, *267*, 25480–25487. [[CrossRef](#)]
10. Perez, I.; Lin, C.H.; McAfee, J.G.; Patton, J.G. Mutation of PTB binding sites causes misregulation of alternative 3' splice site selection in vivo. *RNA* **1997**, *3*, 764–778.
11. Singh, R.; Valcarcel, J.; Green, M.R. Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **1995**, *268*, 1173–1176. [[CrossRef](#)] [[PubMed](#)]
12. Lin, C.H.; Patton, J.G. Regulation of alternative 3' splice site selection by constitutive splicing factors. *RNA* **1995**, *1*, 234–245. [[PubMed](#)]
13. Chan, R.C.; Black, D.L. 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.* **1997**, *17*, 4667–4676. [[CrossRef](#)]
14. Ashiya, M.; Grabowski, P.J. 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* **1997**, *3*, 996–1015. [[PubMed](#)]
15. Lou, H.; Gagel, R.F.; Berget, S.M. An intron enhancer recognized by splicing factors activates polyadenylation. *Genes. Dev.* **1996**, *10*, 208–219. [[CrossRef](#)]
16. Southby, J.; Gooding, C.; Smith, C.W. Polypyrimidine tract binding protein functions as a repressor to regulate alternative splicing of alpha-actinin mutually exclusive exons. *Mol. Cell. Biol.* **1999**, *19*, 2699–2711. [[CrossRef](#)]
17. Zhang, L.; Liu, W.; Grabowski, P.J. Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA* **1999**, *5*, 117–130. [[CrossRef](#)]
18. Castelo-Branco, P.; Furger, A.; Wollerton, M.; Smith, C.; Moreira, A.; Proudfoot, N. Polypyrimidine tract binding protein modulates efficiency of polyadenylation. *Mol. Cell. Biol.* **2004**, *24*, 4174–4183. [[CrossRef](#)]
19. Millevoi, S.; Decorsiere, A.; Loulergue, C.; Iacovoni, J.; Bernat, S.; Antoniou, M.; Vagner, S. A physical and functional link between splicing factors promotes pre-mRNA 3' end processing. *Nucleic Acids Res.* **2009**, *37*, 4672–4683. [[CrossRef](#)]
20. Gama-Carvalho, M.; Barbosa-Morais, N.L.; Brodsky, A.S.; Silver, P.A.; Carmo-Fonseca, M. Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling mammalian splicing factors. *Genome Biol.* **2006**, *7*, R113. [[CrossRef](#)]
21. Knoch, K.P.; Bergert, H.; Borgonovo, B.; Saeger, H.D.; Altkruger, A.; Verkade, P.; Solimena, M. Polypyrimidine tract-binding protein promotes insulin secretory granule biogenesis. *Nat. Cell Biol.* **2004**, *6*, 207–214. [[CrossRef](#)] [[PubMed](#)]
22. Jackson, R.J.; Kaminski, A. Internal initiation of translation in eukaryotes: The picornavirus paradigm and beyond. *RNA* **1995**, *1*, 985–1000. [[PubMed](#)]
23. Kaminski, A.; Hunt, S.L.; Patton, J.G.; Jackson, R.J. Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA. *RNA* **1995**, *1*, 924–938.
24. Niepmann, M. Porcine polypyrimidine tract-binding protein stimulates translation initiation at the internal ribosome entry site of foot-and-mouth-disease virus. *FEBS Lett.* **1996**, *388*, 39–42. [[CrossRef](#)] [[PubMed](#)]
25. Matera, A.G.; Frey, M.R.; Margelot, K.; Wolin, S.L. A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J. Cell Biol.* **1995**, *129*, 1181–1193. [[CrossRef](#)]
26. Huang, S.; Deerinck, T.J.; Ellisman, M.H.; Spector, D.L. The dynamic organization of the perinucleolar compartment in the cell nucleus. *J. Cell Biol.* **1997**, *137*, 965–974. [[CrossRef](#)]
27. Li, B.; Yen, T.S. Characterization of the nuclear export signal of polypyrimidine tract-binding protein. *J. Biol. Chem.* **2002**, *277*, 10306–10314. [[CrossRef](#)]
28. Perez, I.; McAfee, J.G.; Patton, J.G. Multiple RRM domains contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry* **1997**, *36*, 11881–11890. [[CrossRef](#)]
29. Maris, C.; Dominguez, C.; Allain, F.H. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **2005**, *272*, 2118–2131. [[CrossRef](#)]
30. Simpson, P.J.; Monie, T.P.; Szendroi, A.; Davydova, N.; Tyzack, J.K.; Conte, M.R.; Read, C.M.; Cary, P.D.; Svergun, D.I.; Konarev, P.V.; et al. Structure and RNA interactions of the N-terminal RRM domains of PTB. *Structure* **2004**, *12*, 1631–1643. [[CrossRef](#)]
31. Oberstrass, F.C.; Auweter, S.D.; Erat, M.; Hargous, Y.; Henning, A.; Wenter, P.; Reymond, L.; Amir-Ahmady, B.; Pitsch, S.; Black, D.L.; et al. Structure of PTB bound to RNA: Specific binding and implications for splicing regulation. *Science* **2005**, *309*, 2054–2057. [[CrossRef](#)] [[PubMed](#)]

32. Conte, M.R.; Grune, T.; Ghuman, J.; Kelly, G.; Ladas, A.; Matthews, S.; Curry, S. Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold. *EMBO J.* **2000**, *19*, 3132–3141. [[CrossRef](#)]
33. Vitali, F.; Henning, A.; Oberstrass, F.C.; Hargous, Y.; Auweter, S.D.; Erat, M.; Allain, F.H. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **2006**, *25*, 150–162. [[CrossRef](#)]
34. Dreyfuss, G.; Swanson, M.S.; Pinol-Roma, S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* **1988**, *13*, 86–91. [[CrossRef](#)] [[PubMed](#)]
35. Clery, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290–298. [[CrossRef](#)] [[PubMed](#)]
36. Birney, E.; Kumar, S.; Krainer, A.R. Analysis of the RNA-recognition motif and RS and RGG domains: Conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **1993**, *21*, 5803–5816. [[CrossRef](#)]
37. Lunde, B.M.; Moore, C.; Varani, G. RNA-binding proteins: Modular design for efficient function. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 479–490. [[CrossRef](#)]
38. Wang, X.; McLachlan, J.; Zamore, P.D.; Hall, T.M. Modular recognition of RNA by a human pumilio-homology domain. *Cell* **2002**, *110*, 501–512. [[CrossRef](#)]
39. Allain, F.H.; Bouvet, P.; Dieckmann, T.; Feigon, J. Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J.* **2000**, *19*, 6870–6881. [[CrossRef](#)]
40. Deo, R.C.; Bonanno, J.B.; Sonenberg, N.; Burley, S.K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **1999**, *98*, 835–845. [[CrossRef](#)]
41. Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **1999**, *398*, 579–585. [[CrossRef](#)] [[PubMed](#)]
42. Perez-Canadillas, J.M. Grabbing the message: Structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J.* **2006**, *25*, 3167–3178. [[CrossRef](#)]
43. Wang, X.; Tanaka Hall, T.M. Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat. Struct. Biol.* **2001**, *8*, 141–145. [[CrossRef](#)]
44. Shamoo, Y.; Abdul-Manan, N.; Patten, A.M.; Crawford, J.K.; Pellegrini, M.C.; Williams, K.R. Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry* **1994**, *33*, 8272–8281. [[CrossRef](#)]
45. Kielkopf, C.L.; Rodionova, N.A.; Green, M.R.; Burley, S.K. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer. *Cell* **2001**, *106*, 595–605. [[CrossRef](#)]
46. Kielkopf, C.L.; Lucke, S.; Green, M.R. U2AF homology motifs: Protein recognition in the RRM world. *Genes. Dev.* **2004**, *18*, 1513–1526. [[CrossRef](#)]
47. Selenko, P.; Gregorovic, G.; Sprangers, R.; Stier, G.; Rhani, Z.; Kramer, A.; Sattler, M. Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* **2003**, *11*, 965–976. [[CrossRef](#)]
48. Nagai, K.; Oubridge, C.; Jessen, T.H.; Li, J.; Evans, P.R. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **1990**, *348*, 515–520. [[CrossRef](#)]
49. Query, C.C.; Bentley, R.C.; Keene, J.D. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **1989**, *57*, 89–101. [[CrossRef](#)] [[PubMed](#)]
50. Adam, S.A.; Nakagawa, T.; Swanson, M.S.; Woodruff, T.K.; Dreyfuss, G. mRNA polyadenylate-binding protein: Gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* **1986**, *6*, 2932–2943. [[CrossRef](#)] [[PubMed](#)]
51. Swanson, M.S.; Nakagawa, T.Y.; LeVan, K.; Dreyfuss, G. Primary structure of human nuclear ribonucleoprotein particle C proteins: Conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol. Cell. Biol.* **1987**, *7*, 1731–1739. [[CrossRef](#)]
52. Petoukhov, M.V.; Monie, T.P.; Allain, F.H.; Matthews, S.; Curry, S.; Svergun, D.I. Conformation of polypyrimidine tract binding protein in solution. *Structure* **2006**, *14*, 1021–1027. [[CrossRef](#)]
53. Auweter, S.D.; Oberstrass, F.C.; Allain, F.H. Solving the structure of PTB in complex with pyrimidine tracts: An NMR study of protein-RNA complexes of weak affinities. *J. Mol. Biol.* **2007**, *367*, 174–186. [[CrossRef](#)] [[PubMed](#)]
54. Afroz, T.; Skrisovska, L.; Belloc, E.; Guillen-Boixet, J.; Mendez, R.; Allain, F.H. A fly trap mechanism provides sequence-specific RNA recognition by CPEB proteins. *Genes. Dev.* **2014**, *28*, 1498–1514. [[CrossRef](#)] [[PubMed](#)]
55. Bae, E.; Reiter, N.J.; Bingman, C.A.; Kwan, S.S.; Lee, D.; Phillips, G.N., Jr.; Butcher, S.E.; Brow, D.A. Structure and interactions of the first three RNA recognition motifs of splicing factor prp24. *J. Mol. Biol.* **2007**, *367*, 1447–1458. [[CrossRef](#)] [[PubMed](#)]
56. Barraud, P.; Allain, F.H. Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: How the relative orientation between RRMs influences the nucleic acid binding topology. *J. Biomol. NMR* **2013**, *55*, 119–138. [[CrossRef](#)]
57. Cukier, C.D.; Hollingworth, D.; Martin, S.R.; Kelly, G.; Diaz-Moreno, I.; Ramos, A. Molecular basis of FIR-mediated c-myc transcriptional control. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1058–1064. [[CrossRef](#)] [[PubMed](#)]
58. Mackereth, C.D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcarcel, J.; Sattler, M. Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **2011**, *475*, 408–411. [[CrossRef](#)]

59. Zhang, W.; Zeng, F.; Liu, Y.; Zhao, Y.; Lv, H.; Niu, L.; Teng, M.; Li, X. Crystal structures and RNA-binding properties of the RNA recognition motifs of heterogeneous nuclear ribonucleoprotein L: Insights into its roles in alternative splicing regulation. *J. Biol. Chem.* **2013**, *288*, 22636–22649. [[CrossRef](#)]
60. Xu, R.M.; Jokhan, L.; Cheng, X.; Mayeda, A.; Krainer, A.R. Crystal structure of human UPI, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* **1997**, *5*, 559–570. [[CrossRef](#)]
61. Shamoo, Y.; Krueger, U.; Rice, L.M.; Williams, K.R.; Steitz, T.A. Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat. Struct. Biol.* **1997**, *4*, 215–222. [[CrossRef](#)] [[PubMed](#)]
62. Kenan, D.J.; Query, C.C.; Keene, J.D. RNA recognition: Towards identifying determinants of specificity. *Trends Biochem. Sci.* **1991**, *16*, 214–220. [[CrossRef](#)] [[PubMed](#)]
63. Kolupaeva, V.G.; Hellen, C.U.; Shatsky, I.N. 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* **1996**, *2*, 1199–1212.
64. Pilipenko, E.V.; Viktorova, E.G.; Guest, S.T.; Agol, V.I.; Roos, R.P. Cell-specific proteins regulate viral RNA translation and virus-induced disease. *EMBO J.* **2001**, *20*, 6899–6908. [[CrossRef](#)] [[PubMed](#)]
65. Simpson, P.J.; Davydova, N.; Curry, S.; Matthews, S. Resonance assignment and topology of the 2H, 13C, 15N labelled 29 kDa N-terminal fragment of the polypyrimidine tract binding protein (PTB). *J. Biomol. NMR* **2002**, *24*, 79–80. [[CrossRef](#)]
66. Maynard, C.M.; Hall, K.B. Interactions between PTB RRM1 induce slow motions and increase RNA binding affinity. *J. Mol. Biol.* **2010**, *397*, 260–277. [[CrossRef](#)]
67. Clerte, C.; Hall, K.B. The domains of polypyrimidine tract binding protein have distinct RNA structural preferences. *Biochemistry* **2009**, *48*, 2063–2074. [[CrossRef](#)]
68. Maris, C.; Jayne, S.; Damberger, F.F.; Beusch, I.; Dorn, G.; Ravindranathan, S.; Allain, F.H. A transient alpha-helix in the N-terminal RNA recognition motif of polypyrimidine tract binding protein senses RNA secondary structure. *Nucleic Acids Res.* **2020**, *48*, 4521–4537. [[CrossRef](#)]
69. Kafasla, P.; Lin, H.; Curry, S.; Jackson, R.J. Activation of picornaviral IRESs by PTB shows differential dependence on each PTB RNA-binding domain. *RNA* **2011**, *17*, 1120–1131. [[CrossRef](#)]
70. Kafasla, P.; Morgner, N.; Robinson, C.V.; Jackson, R.J. Polypyrimidine tract-binding protein stimulates the poliovirus IRES by modulating eIF4G binding. *EMBO J.* **2010**, *29*, 3710–3722. [[CrossRef](#)]
71. Sharma, S.; Maris, C.; Allain, F.H.; Black, D.L. U1 snRNA directly interacts with polypyrimidine tract-binding protein during splicing repression. *Mol. Cell* **2011**, *41*, 579–588. [[CrossRef](#)]
72. Han, Z.; Wu, Z.; Gong, W.; Zhou, W.; Chen, L.; Li, C. Allosteric mechanism for SL RNA recognition by polypyrimidine tract binding protein RRM1: An atomistic MD simulation and network-based study. *Int. J. Biol. Macromol.* **2022**, *221*, 763–772. [[CrossRef](#)] [[PubMed](#)]
73. Auweter, S.D.; Allain, F.H. Structure-function relationships of the polypyrimidine tract binding protein. *Cell. Mol. Life Sci.* **2008**, *65*, 516–527. [[CrossRef](#)]
74. Dorn, G.; Leitner, A.; Boudet, J.; Campagne, S.; von Schroetter, C.; Moursy, A.; Aebersold, R.; Allain, F.H. Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat. Methods* **2017**, *14*, 487–490. [[CrossRef](#)]
75. Romanelli, M.G.; Lorenzi, P.; Morandi, C. Organization of the human gene encoding heterogeneous nuclear ribonucleoprotein type I (hnRNP I) and characterization of hnRNP I related pseudogene. *Gene* **2000**, *255*, 267–272. [[CrossRef](#)]
76. Sawicka, K.; Bushell, M.; Spriggs, K.A.; Willis, A.E. Polypyrimidine-tract-binding protein: A multifunctional RNA-binding protein. *Biochem. Soc. Trans.* **2008**, *36*, 641–647. [[CrossRef](#)]
77. Wagner, E.J.; Carstens, R.P.; Garcia-Blanco, M.A. A novel isoform ratio switch of the polypyrimidine tract binding protein. *Electrophoresis* **1999**, *20*, 1082–1086. [[CrossRef](#)]
78. Wollerton, M.C.; Gooding, C.; Robinson, F.; Brown, E.C.; Jackson, R.J.; Smith, C.W. Differential alternative splicing activity of isoforms of polypyrimidine tract binding protein (PTB). *RNA* **2001**, *7*, 819–832. [[CrossRef](#)] [[PubMed](#)]
79. Gueroussov, S.; Gonatopoulos-Pournatzis, T.; Irimia, M.; Raj, B.; Lin, Z.Y.; Gingras, A.C.; Blencowe, B.J. An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **2015**, *349*, 868–873. [[CrossRef](#)] [[PubMed](#)]
80. Keppetipola, N.M.; Yeom, K.H.; Hernandez, A.L.; Bui, T.; Sharma, S.; Black, D.L. Multiple determinants of splicing repression activity in the polypyrimidine tract binding proteins, PTBP1 and PTBP2. *RNA* **2016**, *22*, 1172–1180. [[CrossRef](#)]
81. Yuan, X.; Davydova, N.; Conte, M.R.; Curry, S.; Matthews, S. Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein. *Nucleic Acids Res.* **2002**, *30*, 456–462. [[CrossRef](#)]
82. Schmid, N.; Zagrovic, B.; van Gunsteren, W.F. Mechanism and thermodynamics of binding of the polypyrimidine tract binding protein to RNA. *Biochemistry* **2007**, *46*, 6500–6512. [[CrossRef](#)] [[PubMed](#)]
83. Lamichhane, R.; Daubner, G.M.; Thomas-Crusells, J.; Auweter, S.D.; Manatschal, C.; Austin, K.S.; Valniuk, O.; Allain, F.H.; Rueda, D. RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4105–4110. [[CrossRef](#)] [[PubMed](#)]
84. Oh, Y.L.; Hahm, B.; Kim, Y.K.; Lee, H.K.; Lee, J.W.; Song, O.; Tsukiyama-Kohara, K.; Kohara, M.; Nomoto, A.; Jang, S.K. Determination of functional domains in polypyrimidine-tract-binding protein. *Biochem. J.* **1998**, *331 Pt 1*, 169–175. [[CrossRef](#)]
85. Monie, T.P.; Hernandez, H.; Robinson, C.V.; Simpson, P.; Matthews, S.; Curry, S. The polypyrimidine tract binding protein is a monomer. *RNA* **2005**, *11*, 1803–1808. [[CrossRef](#)]

86. Amir-Ahmady, B.; Boutz, P.L.; Markovtsov, V.; Phillips, M.L.; Black, D.L. Exon repression by polypyrimidine tract binding protein. *RNA* **2005**, *11*, 699–716. [[CrossRef](#)]
87. Bielli, P.; Bordi, M.; Di Biasio, V.; Sette, C. Regulation of BCL-X splicing reveals a role for the polypyrimidine tract binding protein (PTBP1/hnRNP I) in alternative 5' splice site selection. *Nucleic Acids Res.* **2014**, *42*, 12070–12081. [[CrossRef](#)] [[PubMed](#)]
88. Cui, J.; Placzek, W.J. PTBP1 enhances miR-101-guided AGO2 targeting to MCL1 and promotes miR-101-induced apoptosis. *Cell Death Dis.* **2018**, *9*, 552. [[CrossRef](#)]
89. Li, X.; Han, F.; Liu, W.; Shi, X. PTBP1 promotes tumorigenesis by regulating apoptosis and cell cycle in colon cancer. *Bull. Cancer* **2018**, *105*, 1193–1201. [[CrossRef](#)]
90. Juan, W.C.; Roca, X.; Ong, S.T. Identification of cis-acting elements and splicing factors involved in the regulation of BIM Pre-mRNA splicing. *PLoS ONE* **2014**, *9*, e95210. [[CrossRef](#)]
91. Zhang, J.; Bahi, N.; Llovera, M.; Comella, J.X.; Sanchis, D. Polypyrimidine tract binding proteins (PTB) regulate the expression of apoptotic genes and susceptibility to caspase-dependent apoptosis in differentiating cardiomyocytes. *Cell Death Differ.* **2009**, *16*, 1460–1468. [[CrossRef](#)] [[PubMed](#)]
92. Spellman, R.; Llorian, M.; Smith, C.W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* **2007**, *27*, 420–434. [[CrossRef](#)] [[PubMed](#)]
93. Wang, M.J.; Lin, S. A region within the 5'-untranslated region of hypoxia-inducible factor-1alpha mRNA mediates its turnover in lung adenocarcinoma cells. *J. Biol. Chem.* **2009**, *284*, 36500–36510. [[CrossRef](#)] [[PubMed](#)]
94. Cho, C.Y.; Chung, S.Y.; Lin, S.; Huang, J.S.; Chen, Y.L.; Jiang, S.S.; Cheng, L.C.; Kuo, T.H.; Lay, J.D.; Yang, Y.Y.; et al. PTBP1-mediated regulation of AXL mRNA stability plays a role in lung tumorigenesis. *Sci. Rep.* **2019**, *9*, 16922. [[CrossRef](#)]
95. Carico, C.C.J.; Acton, A.; Placzek, W.J. Polypyrimidine tract binding protein 1 (PTBP1) contains a novel regulatory sequence, the rBH3, that binds the pro-survival protein MCL1. *J. Biol. Chem.* **2023**, in press.
96. Amrein, H.; Hedley, M.L.; Maniatis, T. The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell* **1994**, *76*, 735–746. [[CrossRef](#)]
97. Huttelmaier, S.; Illenberger, S.; Grosheva, I.; Rudiger, M.; Singer, R.H.; Jockusch, B.M. Raver1, a dual compartment protein, is a ligand for PTB/hnRNPI and microfilament attachment proteins. *J. Cell Biol.* **2001**, *155*, 775–786. [[CrossRef](#)]
98. Gromak, N.; Rideau, A.; Southby, J.; Scadden, A.D.; Gooding, C.; Huttelmaier, S.; Singer, R.H.; Smith, C.W. The PTB interacting protein raver1 regulates alpha-tropomyosin alternative splicing. *EMBO J.* **2003**, *22*, 6356–6364. [[CrossRef](#)]
99. Robinson, F.; Smith, C.W. A splicing repressor domain in polypyrimidine tract-binding protein. *J. Biol. Chem.* **2006**, *281*, 800–806. [[CrossRef](#)]
100. Rideau, A.P.; Gooding, C.; Simpson, P.J.; Monie, T.P.; Lorenz, M.; Huttelmaier, S.; Singer, R.H.; Matthews, S.; Curry, S.; Smith, C.W. A peptide motif in Raver1 mediates splicing repression by interaction with the PTB RRM2 domain. *Nat. Struct. Mol. Biol.* **2006**, *13*, 839–848. [[CrossRef](#)]
101. Henneberg, B.; Swiniarski, S.; Sabine, B.; Illenberger, S. A conserved peptide motif in Raver2 mediates its interaction with the polypyrimidine tract-binding protein. *Exp. Cell Res.* **2010**, *316*, 966–979. [[CrossRef](#)] [[PubMed](#)]
102. Hamilton, B.J.; Genin, A.; Cron, R.Q.; Rigby, W.F. Delineation of a novel pathway that regulates CD154 (CD40 ligand) expression. *Mol. Cell. Biol.* **2003**, *23*, 510–525. [[CrossRef](#)] [[PubMed](#)]
103. Kanda, T.; Gauss-Muller, V.; Cordes, S.; Tamura, R.; Okitsu, K.; Shuang, W.; Nakamoto, S.; Fujiwara, K.; Imazeki, F.; Yokosuka, O. Hepatitis A virus (HAV) proteinase 3C inhibits HAV IRES-dependent translation and cleaves the polypyrimidine tract-binding protein. *J. Viral Hepat.* **2010**, *17*, 618–623. [[CrossRef](#)] [[PubMed](#)]
104. Back, S.H.; Kim, Y.K.; Kim, W.J.; Cho, S.; Oh, H.R.; Kim, J.E.; Jang, S.K. Translation of polioviral mRNA is inhibited by cleavage of polypyrimidine tract-binding proteins executed by polioviral 3C(pro). *J. Virol.* **2002**, *76*, 2529–2542. [[CrossRef](#)]
105. Liu, H.; Zhang, W.; Reed, R.B.; Liu, W.; Grabowski, P.J. Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein. *RNA* **2002**, *8*, 137–149. [[CrossRef](#)]
106. Wollerton, M.C.; Gooding, C.; Wagner, E.J.; Garcia-Blanco, M.A.; Smith, C.W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **2004**, *13*, 91–100. [[CrossRef](#)]
107. Charlet, B.N.; Logan, P.; Singh, G.; Cooper, T.A. Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol. Cell* **2002**, *9*, 649–658. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.