Characterization of cDNAs encoding the polypyrimidine tract-binding protein

Anna Gil, Phillip A. Sharp, Sharon F. Jamison, and Mariano A. Garcia-Blanco

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 USA

The polypyrimidine tract of mammalian introns is recognized by a 62-kD protein (pPTB). Mutations in the polypyrimidine tract that reduce the binding of pPTB also reduce the efficiency of formation of the pre-spliceosome complex containing U2 snRNP. The PTB protein was purified to homogeneity by affinity chromatography on a matrix containing poly(U), and peptide sequence was used to isolate several cDNAs. Because a variety of cell types express mRNA complementary to these cDNAs, PTB may be a ubiquitous splicing factor. Three classes of cDNAs were identified, on the basis of the presence of additional sequences at an internal position. This variation in sequence probably reflects alternative splicing of the PTB pre-mRNA and produces mRNAs encoding the prototype PTB protein, a form of PTB protein containing 19 additional residues, and a truncated form of PTB protein with a novel carboxyl terminus. A murine homolog of pPTB has been characterized previously as a DNA-binding protein. Sequence comparisons indicate that pPTB is distantly related to the hnRNP L protein and that these two proteins should be considered as members of a novel family of RNA-binding proteins.

[Key Words: Polypyrimidine tract; RNA binding; splicing; U2 snRNP; pre-spliceosome complex]

Received March 13, 1991; revised version accepted April 18, 1991.

The splicing of nuclear pre-mRNAs is a highly regulated process in which introns are recognized and removed to yield mature mRNAs [Green 1986; Padgett et al. 1986; Breitbart et al. 1987; Maniatis and Reed 1987; Sharp 1988]. Mammalian introns are characterized by three cis-acting elements: the 5'- and 3'-splice site consensus sequences, and the poorly conserved sequences at the branch site. A polypyrimidine tract typically precedes the AG dinucleotide at the 3'-splice site or immediately follows the branch site. Assembly of the spliceosome begins with the recognition of the 5'- and 3'-splice sites and the branchpoint of the pre-mRNA by U small nuclear ribonucleoprotein particles [snRNPs; Black et al. 1985; Brody and Abelson 1985; Frendeway and Keller 1985; Grabowski et al. 1985; Konarska and Sharp 1986]. Recognition of the branchpoint and 3'-splice-site region by U2 snRNP is enhanced by the presence of an adjacent polypyrimidine tract [Garcia-Blanco et al. 1989]. The 62-kD polypyrimidine tract-binding protein [pPTB] binds the pyrimidine tract with specificity and probably facilitates the binding of U2 snRNP.

pPTB was detected in HeLa cell nuclear extracts by UV cross-linking to pre-mRNAs [Garcia-Blanco et al. 1989]. pPTB specifically bound to the introns of pre-mRNAs that are efficiently spliced in vitro. The binding of this protein was mapped to the polypyrimidine tract of the intron by deletion analysis and by direct biochemical analysis of the protein–RNA adduct. Alterations in the polypyrimidine tract that reduce binding of pPTB result in a corresponding reduction in the formation of the pre-spliceosome complex containing U2 snRNP, as well as the spliceosome and spliced products [Garcia-Blanco et al. 1989]. The binding of pPTB to the polypyrimidine tract occurs rapidly and is independent of other components in nuclear extracts. Thus, pPTB exhibits characteristics strongly suggesting that it participates in the early stages of splice site recognition and pre-spliceosome assembly.

RNA-binding proteins other than pPTB have been implicated in pre-mRNA splicing. Either immunodepletion of heterogenous nuclear RNP [hnRNP] C or addition of a monoclonal antiserum specific for hnRNP C inhibited the splicing of pre-mRNA in vitro [Choi et al. 1986]. The mechanism of this inhibition is not known, but hnRNP C readily binds the polypyrimidine tract of the 3'-splice site region. Several other proteins also bind specifically to the polypyrimidine tract. The hnRNP proteins A1 and D recognize the conserved AG dinucleotide at the 3'-splice site [Swanson and Dreyfuss 1988], whereas the two related intron-binding proteins [IBPs] recognize both the 3'-splice site and polypyrimidine tract [Gerke and Steitz 1986; Tazi et al. 1986]. In addition, the 220-kD protein p220 binds pre-mRNAs under conditions for...
spliceosome formation [Garcia-Blanco et al. 1990]. The mammalian p220 is believed to be the human homolog of the yeast PRP8 splicing factor (Anderson et al. 1989; Pinto and Steitz 1989; Garcia-Blanco et al. 1990). Several reports suggest that mammalian protein factors recognize the 5'-splice site of introns [Mayeda et al. 1986; Zapp and Berget 1989; Siebel and Rio 1990]. These results are consistent with studies in yeast strongly suggesting that factors in addition to U1 and U2 snRNPs recognize the 5'-splice site and branchpoint (Seraphin et al. 1988; Ruby and Abelson 1989; Seraphin and Rosbash 1989, Siliciano and Guthrie 1989).

Biochemical complementation assays have been used to identify activities important for splicing or formation of specific complexes in the spliceosome pathway [Krainer and Maniatis 1985; Kramer and Keller 1985; Kramer 1988; Fu and Maniatis 1990; Utans and Kramer 1990]. The SF2 factor and alternative splicing factor (ASF) were originally identified by biochemical complementation and are probably related. SF2 is required for the formation or stabilization of a pre-spliceosome complex [Krainer et al. 1990a]. A purified SF2 preparation contains two related polypeptides of ~33 kD and binds RNA with no apparent sequence specificity [Krainer et al. 1990a]. Therefore, it is surprising that SF2 can influence the selection of a specific 5'-splice site in cases of pre-mRNAs containing multiple 5'-splice sites. High concentrations of SF2 favor splicing of the 5'-splice site most proximal to the 3'-splice site [Krainer et al. 1990b]. The 30- to 35-kD factor (ASF) has a very similar biochemical activity [Ge and Manley 1990]. The U2 snRNP auxiliary factor (U2AF) is a component required for the interaction of the U2 snRNP with the branch site and thus is required for pre-spliceosome formation [Ruskin et al. 1989; Zamore and Green 1989]. U2AF has been purified to near homogeneity and is probably a heterodimer with subunits of 68 and 35 kD [Zamore and Green 1989]. Several properties of U2AF suggest that it is distinct from pPTB [see Discussion].

In this study we describe the molecular cloning and structural characterization of cDNAs encoding the 62-kD pPTB.

Results

Purification and sequencing of pPTB

pPTB, from HeLa cell nuclear extracts, was purified to homogeneity by column chromatography over DEAE-Sephadex, heparin–agarose, and poly(U)–agarose matrices [Garcia-Blanco et al. 1989]. In the last step, a gradient of increasing KCl concentration was used to elute pPTB from the poly(U)–agarose matrix. The pPTB, assayed by UV cross-linking to the radiolabeled pre-mRNA splicing substrate Ad10, eluted from the affinity poly(U) matrix at 0.8 M KCl [Fig. 1A]. Following electrophoresis in a denaturing gel, purified pPTB consisted of a triplet of silver-stained bands that comigrated with a triplet of bands characterized by UV cross-linking to Ad10 premRNA [Fig. 1B]. As described previously, pPTB purified from different preparations of nuclear extract resolved either as a doublet or triplet of polypeptide bands [Garcia-Blanco et al. 1989].

To obtain peptide sequence, purified pPTB was resolved on a preparative denaturing gel, transferred to nitrocellulose membrane, and stained with Ponceau Red dye. pPTB from each band of the triplet was digested with trypsin [Aebersold et al. 1987]. The tryptic peptides were subsequently eluted from the nitrocellulose membrane and fractionated by reverse-phase high-performance liquid chromatography (HPLC). The peptide elution pattern for each of the three pPTB bands was nearly identical, suggesting that the proteins in the three bands shared identical sequences. The sequences of two peptides were determined. The amino-terminal sequence of one peptide revealed the 20 amino acids N-Asn-Asn-Gln-His-Ala-Lys-C. The sequence of a second oligopeptide revealed the 8 amino acids N-Ser-Gln-Pro-Ile-Tyr-Asp-Pro-Val-Ser-Ala.

Figure 1. Purification of the 62-kD pPTB. (A) The pPTB was eluted from poly(U)–agarose with an increasing KCl gradient. The eluted protein in fractions 30–38 was resolved on an SDS–polyacrylamide gel and silver-stained (lanes 1–9). pPTB resolved as a triplet of bands at 62 kD and is indicated. (B) Purified pPTB was silver-stained (lane 1), or UV cross-linked to the uniformly labeled RNAs: Ad10 [lane 2], PIP3 [lane 3], the BPY fragment that spans the Ad10 branchpoint and polypyrimidine tract [lane 4], and the RNase T1 fragment spanning the Ad10 polypyrimidine tract [lane 5]. In comparison to the other RNAs, the RNase T1 fragment had >10-fold lower specific activity. The sequences of the RNA substrates are described in Garcia-Blanco et al. [1989]. The 14C-labeled methylated protein molecular mass standards [myosin (200 kD), phosphorylase b (100 and 95 kD), bovine serum albumin (69 kD), ovalbumin (46 kD), and carbonic anhydrase (30 kD]; Amersham] are indicated.
Figure 2. The nucleic acid sequence of pPTB-A, pPTB-B, and pPTB-C cDNAs and encoded peptide sequence. The nucleic acid sequences are shown above the encoded peptide sequence. The name of each cDNA is listed to the right of the corresponding sequence. The nucleic and amino acid sequence in common between the three cDNAs is boxed, and the numerical position of the sequences is indicated.

Isolation and characterization of cDNAs encoding pPTB

A 60-bp cDNA segment encoding the 20-amino-acid tryptic peptide was isolated by amplification with the polymerase chain reaction (PCR) using degenerate oligonucleotides [see Materials and methods]. The 60-bp fragment, derived from PCR, was purified and subcloned into the plasmid vector pBS1. The sequences of three independent isolates, pPTB-A, pPTB-B, and pPTB-C, were determined [Fig. 2]. All three cDNA inserts shared identical sequence from nucleotide 16 to 46. As anticipated, this sequence encodes amino acids 6–15 of the 20-amino-acid tryptic peptide.

To isolate the full-length cDNA encoding pPTB, the unique sequence from amino acids 6 to 15 was used to synthesize a 32-nucleotide DNA probe [see Materials and methods]. The oligonucleotide probe was synthesized and used to screen a plasmid cDNA library prepared from human placenta [Simmons and Seed 1988]. Thirty-six positives were isolated from a primary screening of 500,000 bacterial colonies. Of the 36 primary isolates, 22 were colony-purified, and the DNAs from these isolates were compared by restriction endonuclease digestion. Six clones, containing inserts ranging in size from 1226 to 3090 bp in length (Fig. 3). The first initiation codon, which is in-frame with the sequences encoding pPTB, is at position 29 and represents a good consensus sequence for initiation [Kozak 1987]. Assuming that this specifies the amino terminus of the protein, the open reading frame (ORF) extends to position 1622 and encodes a protein of 572 kD. The 5'-untranslated region of the cDNA is only 28 nucleotides and does not contain an in-frame termination codon. Thus, a longer cDNA clone extending the 5' end of the cDNA might reveal the presence of another in-frame initiation codon. It is unlikely that the cDNA sequence is significantly longer than 3.1 kb, as this is the approximate length of the most abundant mRNA detected in a Northern blot analysis [see below]. The 3'-untranslated region is unusually long (1.47 kb) and contains the consensus AAUAAA polyadenylation signal [Proudfoot and Brownlee 1976], 14 nucleotides upstream of an abbreviated poly(A) tract.

Nucleotides 683–742 of the PTB cDNA encode the 20-amino-acid peptide sequence obtained from purified pPTB. Similarly, nucleotides 395–418 encode a nearly identical match to the 8-amino-acid peptide sequence. By peptide sequencing, the phenylalanine at position 6 was ambiguously assigned and is predicted to be an isoleucine. In addition, the amino-terminal residue was determined chemically to be serine but is predicted to be a glycine in the encoded protein sequence. This discrepancy between the cDNA and peptide sequence may reflect a modification of the amino-terminal glycine residue.

The predicted pPTB is 531 amino acids long and has an isoelectric point of 9.95. The encoded protein contains a large number of hydrophobic residues (44%) distributed over the entire length of the protein and contains a sig-

Figure 3. Nucleotide sequence of PTB cDNA and predicted amino acid sequence. The polyadenylation sequence ATTAAA at nucleotide 3068 is underscored with asterisks. The translation initiation codon AUG and the termination codon UAG are located at nucleotides 29 and 1625, respectively. The sequences encoding the 8 and 20 amino acids of pPTB, determined by microsequencing of tryptic peptides, are underlined. [V] The site of sequence variation between different cDNAs encoding pPTB begins at nucleotide 921 (see Fig. 4). The translation stop codon UAG of the deduced ORF encoding pPTB is located at nucleotide 1199 and is boxed. Except for the variation in sequence beginning at nucleotide 921, all six cDNA clones [pC1, pC9, pC11, pC14, pC15, and pC19] contained sequence identical to the prototype cDNA sequence. The 5' ends of the six cDNA clones were all located within the first 150 nucleotides of the PTB cDNA sequence and extended 0.9–3.1 kb toward the 3' end.
cDNAs encoding polypyrimidine tract-binding proteins

1

29 ATG GAC GCC ATT GTC CCA GAT ATA GCC GCT GTT GAC ACA CAG GGA TCT GAC GTT TTC TCT ACT TGT GTC

Met Asp Gly Ile Val Pro Asp Ala Val Ile Gly Thr Arg Ser Asp Leu Phe Ser Thr Cys Val

101 ACT AAC GGA CCG TTT ATC ATG AAC ACC AAG TCG TCT TCT GCA GCA AAG AAC AAG AAG AAA GAC TAC AAA

Thr Asn Gln Pro Phe Met Ser Ser Asn Ser Ser Ala Ser Ala Gly Asn Ser Asp Ser Lys Phe Lys

172

173GGT GAC GCC AGA CAC GTG GCC CCT TCT AGA GTG ATC ATG CAC ATC CGG AAG CTC TAC CCC ACC GAC TGC ACG GAG

Gly Asp Arg Ser Ala Gly Val Pro Ser Arg Ser Ser Ala Ser Ala Gly Asn Ser Lys Leu Arg Pro Ile Ser Asp Val Thr Glu

244

245GGG GAA GTC ATC TCC GTG GGG CCT TCT AGG AAG AAG GTC ACC ACC CTT CGT AAG ACG CAG TCG GGG AAC CTG GCC TTG GCT

Glu Gly Val Ile Ser Leu Gly Pro Leu Thr Leu Arg Ala Leu Met Thr Val Asn Tyr Thr Ser Val Thr Val Pro Val

316

317GCC TTC ATC GAG ATG AAC ACG GAG GAG GCT GCC AAC ACC ATG GTG AAC TAC TAC ACC TCG GTG ACC CCT GTG

Ala Phe Ile Glu Pro Ile Ser Ser Asp His Lys Leu Lys Thr Asp Ser Pro Asp Ala

388

461GCC GCG GACC CCC ACC TAC ATC GAG ATT GAC TGC TGC ACG TGG AAC TCG GTC CAG TCG GGG AAC CTG GCC TTG GCT

Leu Arg Gly Gln Pro Ile Val Ala Ser His Ser Ser Asp Asp Val Tyr Ser Lys Leu Ala Leu Ala Asp Ala

532

533GCC GCG CTG GAG GCC GGT GAG GCT GAC CCC GCG GCC CTG CAG CAC GCC AAG CTG TCG

Ala Arg Ala Glu Ala Leu Ala Leu Val Ser Gin Ser Gly Asn Leu Leu Ala Ala Ser Ala

604

676TAC CCT GTG ACC CAT GTG CAG CAC ATT TTC GAG ATG AAC ACG GAG GAG GCT GCC AAC ACC ATG GTG AAC TAC TAC ACC TCG GTG ACC CCT GTG

Thr Tyr Val Thr Leu Asp Leu Val Ala Gly Ser His Ser Asp His Lys Leu Lys Ser Tyr Ser

748

749CTG GAC CAG ACC ATC TAC ACC CCC TCT GTG CAC ATG GCC CTC ATT GAC CTG CAC AAC CAC GAG GTC TCC GGT GTC

Leu Asp Gln Thr Met Ala Ala Ala Phe Gly Leu Ser Val Pro Asn Val His Gly Asn Leu Pro Leu Ala

820

892GGCGGTTTTTTATGGTGAcACAAATGTATATTTTGCTAACAGcAATTCCAGGCTCAGTACCGCGACCGAGCCAGGGAACCCCACGCACATTCCGT

Figure 3. (See facing page for legend.)

966
significant number of charged residues in the carboxy-terminal region.

Three classes of cDNAs encoding different forms of pPTB

The six cDNA clones were divided into three classes on the basis of a variation in sequence beginning at nucleotide 921. The three classes are represented by the prototype cDNA sequences contained in clones pC15, pC19, and pC9. Hereafter, the corresponding cDNA structure will be referred to as PTB, PTB2, and PTB3, respectively (Fig. 4). The variation in sequences most likely reflects alternative splicing of PTB pre-mRNA and results in the extension or truncation of the ORF contained in the prototype sequence. The cDNA sequence of PTB2 contains an additional 57 nucleotides that are not present in the prototype cDNA sequence. The additional 57 nucleotides extends the ORF of the prototype cDNA by 19 amino acids. The cDNA sequence of clone PTB3 contains an additional 77 nucleotides of sequence not present in the prototype PTB clone. This sequence extends the 57 nucleotides of additional sequence present in PTB2 by 20 nucleotides. The additional nucleotides in PTB3 shifts translation, at position 921, from the long ORF encoded by the prototype cDNA to a second ORF specifying a unique carboxy-terminal sequence of 95 amino acids. The predicted full-length ORF in PTB3 terminates at nucleotide 1199 and encodes a protein of 42.8 kD.

The nucleotide sequences at the boundaries where PTB2 and PTB3 diverge from the prototype cDNA suggest that the three types of cDNAs are generated by alternative splicing. The trinucleotide CAG is found at the 3′ boundary of the additional sequences in PTB2 and is part of the consensus 3′-splice-site sequence. If this CAG sequence represents a 3′-splice site, an intron would have been excised from PTB pre-mRNA in forming the bond between nucleotides 920 and 921 of the prototype PTB sequence. The same trinucleotide CAG is also present immediately upstream of the 5′ terminus of the additional sequences in PTB2. Similarly, if this CAG sequence represents a 3′-splice site, an intron would have been excised from the PTB pre-mRNA in forming the bond at nucleotide 920 in PTB2. The dinucleotide GT is present at the 5′ terminus of the additional sequences in PTB3 and could represent part of the highly conserved sequences found in 5′-splice sites. However, the hypothetical intron that would be excised to generate PTB2 would have to be only 20 nucleotides in length. Because the shortest intron that can be spliced efficiently is ~65 nucleotides in length (Wieringa et al. 1984), it is unlikely that the GT dinucleotide is recognized as part of a 5′-splice site. It is more likely that additional intron sequences separate the sequences at position 921 in PTB3 and that the presence of the GT dinucleotide immediately downstream of position 920 is fortuitous. Therefore, the synthesis of the PTB, PTB2, and PTB3 mRNAs could represent the alternative utilization of a staggered set of 3′-splice sites.

Analysis of PTB mRNA

The most abundant mRNA detected by hybridization to a fragment of the PTB cDNA insert is ~3.5 kb in length and is present in a variety of human cell lines [Fig. 5]. The 1.4-kb cDNA insert of clone pC14 was used to probe Northern blots containing poly[A]+ RNA isolated from human B (BJAB), T (Jurkat), and epithelial (HeLa) cell lines. The blot was reprobed with a fragment of the human β-actin gene (Gunning et al. 1983) to control for the level of PTB mRNA. The RNA blot analysis indicated that PTB mRNA is abundantly expressed in all three cell types. The three classes of PTB cDNA may be present in the population of 3.5-kb transcripts detected in the Northern analysis. The lengths of the different PTB mRNAs differ by <100 nucleotides, and under the experimental conditions used, individual species would not have been resolved. The poly[A] tract in mammalian mRNA is typically 200–300 nucleotides in length. Thus, comparison of the length of the 3.1-kb prototype cDNA to the 3.5-kb transcript detected by Northern analysis suggests that the cDNA is probably full length.

The PTB mRNA is transcribed from a single- or low-copy gene. The 1.4-kb cDNA contained in clone pC14 was used to probe a Southern blot containing HeLa genomic DNA digested with either EcoRI, Dral, PstI, HindIII, or BamHI restriction endonuclease. The limited number and simple pattern of bands resulting from hybridization of the probe under stringent conditions suggests that PTB mRNA is encoded by a single- or low-copy gene [data not shown].

Murine homolog of pPTB

The deduced pPTB sequence was compared to sequences in the PIR, SWISSPROT, and GENPEPT data bases available through the National Center for Biotechnology Information, using the BLASTP 1.1.114MP algorithm (Altschul et al. 1990). The search revealed that the sequence of a 25-kD mouse nuclear protein is 98% identical to the deduced pPTB sequence in contiguous regions >100 amino acids in length (Bothwell et al. 1990). The mmp25K protein was isolated from mouse plasmacytoma cells by affinity chromatography and characterized as a single-stranded DNA-binding protein (A.L.M. Bothwell et al., pers. comm.). The 25-kD peptide probably represents a proteolytic fragment from the carboxyl terminal of a larger protein. The murine cDNA encoding mmp25K is 3.08 kb in length, which closely approximates the size of the full-length mammalian PTB cDNA. Although there is some homology within the sequence preceding the translation start codon, the region of strikingly high homology is coincident with the deduced ORF of PTB. The sequences are widely divergent in the 3′-untranslated regions.

PTB homology to other RNA-binding proteins

Data base sequence comparisons also revealed that the deduced pPTB sequence contained significant regions of
cDNAs encoding polyadenylation tract-binding proteins

Figure 4. Nucleic acid sequences of PTB, PTB2, and PTB3 cDNAs and predicted amino acid sequences. The variation in sequence represented in the three classes of cDNAs begins at nucleotide 921 and is shown. The sequences of these cDNAs are otherwise identical to the prototype sequence (see Materials and methods). The GT and AG consensus splice site dinucleotides are underlined, and the predicted amino acid sequence encoded by each cDNA is shown below the nucleotide sequence.

Figure 5. RNA analysis. (A) Northern blot containing 5 µg of poly(A)+ RNA isolated from human cell lines. [Lane 1] Epithelial (HeLa); [lane 2] T (Jurkat); [lane 3] B (BJAB). The blot was probed with the 1.4-kb cDNA insert of clone pC14. The size of the PTB mRNA detected in each lane is ~3.5 kb. The positions of the 28S and 18S rRNA bands are indicated. (B) The Northern blot in A was reprobed with a 0.7-kb fragment of the human β-actin cDNA to determine the relative level of PTB mRNA expressed in each cell line. Lanes 1, 2, and 3 correspond to the same lanes denoted in A.

Discussion

The polyadenylation tract is essential for recognition of the 3'-splice-site region of pre-mRNAs that are spliced efficiently.
Figure 6. Alignment of domains in RNA-binding proteins potentially homologous to the predicted pPTB sequence. Residues potentially homologous to pPTB are numbered. Alignment of protein domains with the BLASTP 1.1.114MP algorithm derived from the data base search of the right shown below the pPTB sequence and are representing potential homology from one {see Materials and methods). Sequences PIR, SWISSPROT, and GENPEPT libraries matches are denoted by the single-letter contiguous region are bracketed. Exact amino acid code; similar matches are denoted by a plus sign (+). The identities of previously defined RNP-CS present in the sogn are listed on the same line and to the proteins containing sequence homology of the RNA-binding proteins is 387, 407-422, 463-483, 337-359, and 372-419 of Nucleolin; 127-157, 219-246, and 211-238 of Sex-lethal; and 123-138 of U1 70K. These homologies are presented in the order of their alignment from amino to carboxyl terminus of the pPTB sequence.

Purified pPTBs resolved into a triplet of bands with an average molecular mass of 62 kD. This variation in the electrophoretic mobility could be caused by chemical modifications of the proteins or could represent different forms generated by alternative splicing of the PTB pre-mRNA. Three classes of cDNA sequences have been characterized and most likely represent alternatively spliced forms of the PTB pre-mRNA. The prototype sequence is based on the longest cDNA segment characterized and contains an ORF encoding a 57.2-kD protein. The sequence present in the PTB2 cDNA contains an additional 57 nucleotides beginning at position 921 and,
thus, extends the prototype ORF by 19 amino acids. These additional 19 residues are primarily hydrophobic and contain 4 proline residues that could disrupt many secondary structures. Such alterations in pPTB sequence could modify the RNA-binding properties of the protein or modulate potential protein–protein interactions. Small differences between two otherwise identical protein sequences are not unusual features of hnRNP proteins (Burd et al. 1989). The sequences of the hnRNP C1 and C2 proteins differ by an additional 13 amino acids in the mid-region of the C2 sequence. Similarly, the sequences of hnRNP A2 and B1 differ by an additional 12 amino acids in B1. In the latter case, the difference is generated by the alternative splicing of mini-exons.

In contrast, the PTB3 cDNA contains 77 additional nucleotides at position 921, which shifts the prototype ORF to a second reading frame. The resulting ORF encodes a truncated protein with a unique 95-amino-acid carboxyl terminus. The carboxy-terminal sequence is punctuated with a large number of proline and basic residues. Interestingly, the sequence includes an 11-amino-acid peptide, PRRRQQLRRQV, which is similar to the arginine-rich consensus sequence found in bacteriophage antiterminators and conserved in several other RNA-binding proteins that recognize specific RNA stem–loop structures (Lazinski et al. 1989). Thus, the predicted PTB3 protein may have functional properties that are different from the prototype pPTB.

Surprisingly, a cDNA encoding a protein highly homologous to the pPTB had been isolated previously from mouse plasmacytoma cells (Boothwell et al. 1990). This murine cDNA was isolated with peptide sequence derived from a 25-kD protein purified by DNA affinity chromatography. This purification is probably due to the binding of the protein to single-stranded DNA, which is a property shared by many hnRNP proteins. The murine cDNA is highly homologous throughout the PTB-coding sequence but dramatically diverges within the 3′-untranslated sequence. Similar to the human PTB cDNA, the murine cDNA contains a very short 5′ (34 nucleotide)-untranslated sequence and an unusually long 3′-untranslated region (1.5 kb). Thus, the murine cDNA is undoubtedly the homolog of the human PTB.

The predicted pPTB sequence also contained significant regions of homology to the human hnRNP L (Pirollo-Roma et al. 1989) and limited regions of homology to the Drosophila elav proteins (Robinow et al. 1988), yeast PABP (Sachs et al. 1986; Adam et al. 1986), chicken Nucleolin (Maridor et al. 1990), Drosophila Sex-lethal (Bell et al. 1988), and Xenopus U1 70K (Etzerodt et al. 1988) proteins. The elav, PABP, Nucleolin, Sex-lethal, and U1 70K proteins contain at least one example of a 90-amino-acid domain conserved in many RNA-binding proteins (Dreyfuss et al. 1988; Bandziulis et al. 1989). Both binding and UV cross-linking studies show that peptide fragments consisting largely of this domain can bind RNA (Herrick and Alberts 1976; Bugler et al. 1987; Query et al. 1989). The most conserved feature of this RNA-binding domain is the 8-amino-acid consensus sequence (R/K)GF(G/A)FVX(F/Y) (Adam et al. 1986; Swanson et al. 1987). Recent structural studies show that the entire domain consists of two α-helices and four β-sheets (Nagai et al. 1990). The RNP–CS is contained in one of the β-sheets with the basic terminal residue in a loop connecting a β-sheet and an α-helix. This conserved domain binds a specific RNA stem–loop structure and recognizes specific nucleotides within the loop. Interestingly, the PTB sequence does not contain a discernible match to the RNP–CS (see Materials and methods). Instead, the limited regions of homology between the encoded pPTB and the elav, PABP, Sex-lethal, Nucleolin, and U1 70K proteins are restricted to a pattern of hydrophobic residues that are similar in character to those loosely conserved residues within the 90-amino-acid RNA-binding domain. This may reflect a common structural organi-
zation between the carboxyl region of pPTB and the conserved RNA-binding domain. It is likely that the carboxy-terminal half of the encoded pPTB contains one or more RNA-binding domains, because the corresponding domain of the murine protein (25-kD peptide) binds nucleic acid (A.L.M. Bothwell, pers. comm.). Thus, pPTB is probably a member of a new family of RNA-binding proteins that have a distant relationship to RNA-binding proteins containing the RNP–CS.

The hnRNP L, a 65-kD protein, shares significant regions of homology over the entire length of the pPTB sequence and also does not contain a strong match to the RNP–CS (Piotol-Roma et al. 1989). PTB and hnRNP L are most likely members of the same family of RNA-binding proteins. We propose that the hnRNP L protein will bind single-stranded RNA in a sequence-specific manner similar to the pPTB. This possibility may explain the observation that antisera directed against the hnRNP L protein binds to the giant loops of amphibian lampbrush chromosomes (Piotol-Roma et al. 1989).

Mutational analysis suggests that the pPTB is important for formation of the pre-spliceosome complex containing U2 snRNP (Garcia-Blanco et al. 1989). Biochemical fractionation studies suggest that U2AF is also essential for the binding of U2 snRNPs to pre-mRNA (Zamore and Green 1989). Fractionation of cellular proteins on a poly(U) affinity column in the presence of 1 M NaCl separates the U2AF component from U2 snRNPs and other splicing factors. The most prominent proteins in the U2AF fraction have molecular masses of 65 and 35 kD. However, only the 65-kD protein is critical for the reconstitution of splicing activity (Zamore and Green 1991). Although important for a sequence-specific activity, the p65 protein of U2AF has not been shown to recognize RNA in a sequence-specific manner. The 62-kD pPTB was also purified by poly(U) affinity chromatography and is most likely not related to the 65-kD U2AF. For example, U2AF elutes from the poly(U) matrix only in the presence of 2 M urea, whereas the pPTB elutes at 0.8 M monovalent salt. Both pPTB and U2AF are probably RNA-binding components essential for the recognition of some pre-mRNAs by U2 snRNP and therefore could simultaneously bind pre-mRNA in the 3'-splice-site region.

The presence of a polypyrimidine tract is essential for lariat formation in the first step of splicing. The polypyrimidine tract may also be critical for the second step in splicing [Reed 1989]. When the branch site is not adjacent to the AG dinucleotide, exon ligation occurs more efficiently if a polypyrimidine tract is located immediately upstream of the 3'-splice-site junction. Similarly, the presence of a polypyrimidine tract preceding the AG dinucleotide greatly enhanced the utilization of an alternative 3'-splice site in yeast (Patterson and Guthrie 1991). Some alternatively spliced pre-mRNAs contain branch sites at distances >100 nucleotides upstream of the 3'-splice-site junction (Helfman and Ricci 1989; Smith and Nadal-Ginard 1989). Efficient utilization of these branch sites requires the presence of a flanking polypyrimidine tract. Therefore, in addition to being essential for 3'-splice-site recognition, pPTB may also modulate the utilization of these sites in alternatively spliced pre-mRNAs.

Sexual differentiation in Drosophila melanogaster is determined by a hierarchy of genes that are regulated at the level of alternative splicing of pre-mRNA. These genes include Sex-lethal (Sxl; Bell et al. 1988) and transformer (tra; Boggs et al. 1987). The alternative splicing of Sxl and tra pre-mRNAs is thought to be regulated by the binding of Sxl protein to RNA sequences within the 3'-splice site, which consequently blocks utilization of this site (Bell et al. 1988; Sosnowski et al. 1989; Inoue et al. 1990). Interestingly, the alternative splicing of the Sxl and tra pre-mRNA controls the synthesis of mRNA encoding either a full-length functional RNA-binding protein or a truncated nonfunctional form. Alternative splicing of Sxl pre-mRNA is probably autoregulated by the binding of the Sxl protein. A similar type of autoregulation may control the processing of PTB pre-mRNA, because it may be alternatively spliced to generate mRNAs encoding either the full-length 62-kD protein or a shorter truncated form. Characterization of the genomic PTB sequences will be necessary to determine the regulatory processes responsible for specifying this pattern of splicing.

Materials and methods

Purification of pPTB

pPTB was purified, on a preparative scale, from 35 ml of HeLa cell nuclear extract (~600 mg of protein) as described in Garcia-Blanco et al. (1989). pPTB was resolved through DEAE–Sephacel, heparin–agarose, and poly(U)–agarose columns using a fourfold larger volume of matrix than that described previously. pPTB eluted from the poly(U)–agarose matrix at 0.8 M KCl, and the peak fractions containing primarily homogenous pPTB were collected.

Purified pPTB was concentrated, resolved on an SDS–polyacrylamide gel, and transferred to nitrocellulose. The immobilized pPTB was trypsinized in situ, and the tryptic peptides were eluted onto a narrow-bore reverse-phase HPLC and fractionated. The amino-terminal sequence for two tryptic peptides was determined by William S. Lane of the Harvard Microchemistry Laboratory.

cDNA isolation and DNA analysis

Peptide sequence derived from purified pPTB was used to synthesize three degenerate oligonucleotides, A [5'-GGGGAATTCAATT/GCAAT/CICAI/GATTC/GTCA/GAGC/G-3'], B [5'-GGGAAACCTTGAGA/GITC/G/C/GTCG/A/GGT/GC-3'], and C [5'-GGGAAACCTTGC/A/G/C/G/TCA/G/TGGC/AT/GC-3']. Complementary DNA was synthesized from HeLa poly(A) RNA using random hexanucleotides (Pharmacia LKB Biotechnology, Inc.) to prime the reverse transcription reaction. The single-stranded DNA preparation was then added to a PCR containing 100 ng/ml of oligonucleotides A and C, 200 μM of each deoxynucleotide triphosphate, and 1 μl of Taq polymerase (Perkin Elmer Cetus). The mixture was overlaid with Gil et al.
Restriction endonuclease cleavage of the PCR products generated a 60-bp DNA fragment. The 60-bp fragment was analyzed by Southern blot analysis (Maniatis et al. 1982) using the radiolabeled oligonucleotide B to probe for specific internal sequences. The 60-bp fragment that hybridized specifically to oligonucleotide B was purified and subcloned into the plasmid vector pBS- (Stratagene). The sequences of three independent isolates, pPTB-A, pPTB-B, and pPTB-C, were determined by the dioxygenucleotide chain-termination reaction (Sanger et al. 1977; U.S. Biochemical) primed with oligonucleotides complementary to the pBS- vector sequences flanking the cDNA insert.

The cDNA clones were isolated from a human placenta cDNA library provided by Brian Seed (Simmons and Seed 1988). Bacterial colonies (50,000) from the plasmid library were screened at a density of 20,000/150-mm plate. The plates were overlayed with nitrocellulose filters, and the primary colonies were grown directly on the filters. The primary filters were transferred in duplicate to nitrocellulose filters. To prepare the filters for hybridization they were soaked for 5 min in 0.5 M NaOH, neutralized for 5 min in 1 M Tris-Cl (pH 7.5), washed 5 min in 0.5 M Tris-Cl (pH 7.5), 1.25 M NaCl, and baked at 80°C for 90 min. The filters were prewashed in 3x SSC [0.45 M NaCl, 0.045 M sodium citrate], 0.1% SDS, prehybridized for 16 hr at 37°C in 6x SSC, 5x Denhardt's solution [0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin], 0.05% Na2HPO4 (pH 6.8) 100 μg/ml of herring sperm DNA, 0.5% SDS, and hybridized with a radiolabeled probe [see below] at 42°C for 16 hr in 1x Denhardt's solution, 0.02 M Na2HPO4 (pH 6.8) 100 μg/ml of boiled herring sperm DNA, 10% dextran sulfate, and 50% formamide. Following hybridization, the membrane was rinsed at room temperature in 2x SSC, 0.1% SDS, and washed twice at 42°C for 30 min in 0.1x SSC, 0.1% SDS. These washes were followed by two washes at 60°C for 30 min in 0.1x SSC, 0.1% SDS.

The DNA of clone pC14 was digested with the restriction endonuclease XhoI, and the 1.4-kb cDNA insert was isolated after resolution on a 1% agarose gel. The DNA fragment was uniformly labeled with [α-32P]dCTP by DNA polymerase I (Klenow fragment of Escherichia coli) primed with random hexanucleotides. The reagents were obtained from Pharmacia LKB Biotechnology, Inc., and used as recommended. The labeled DNA probe was added to the hybridization solution at a final concentration of 10^6 cpm/ml.

The Northern blot probe with the 1.4-kb fragment of pC14 was reprobed with a fragment of the human β-actin cDNA (Gunning et al. 1983). The plasmid pHFA-1, containing the human β-actin cDNA insert, was hybridized with the restriction endonucleases EcoRI and BamHI. The 0.7-kb DNA fragment was isolated, labeled, and hybridized to the Northern blot as described above except that the membrane was prewashed at 65°C for 30 min in 1x SSC, 0.5% SDS, prior to hybridization.

### Antisera production and immunoprecipitation assays

Antisera directed against the 20-amino-acid peptide encoded in the sequence of the PTB cDNA was raised in three New Zealand white rabbits. A 21-amino-acid peptide was synthesized and used as the immunogen. The peptide consisted of the 20-amino-acid sequence derived from purified pPTB and a carboxy-terminal cysteine residue. The cysteine residue was added to the carboxyl terminus of the peptide so that it could be specifically coupled to a preactivated carrier protein cBSA (Pierce). The conjugated peptide was then mixed with aluminum hydroxide adjuvant under conditions recommended by the manufacturer (Pierce). The rabbits were initially injected with 400 μg of crude peptide and subsequently boosted with 200 μg of HPLC-purified peptide at 5- to 6-week intervals. The rabbits were bled 10 days after the initial immunization and after each boost. A 10-ml sample of preimmune sera was extracted from the three rabbits immediately prior to the initial injection.

Sera from the second bleed was used in the immunoprecipitation assays. The p62 was UV cross-linked to the radiolabeled pre-mRNA substrate pPIP7.A, in nuclear extracts from HeLa cells, as described by Garcia-Blanco et al. (1989). The pPIP7.A clone (M.J. Moore and P.A. Sharp, in prep.) is a derivative of the pIP3 clone used to characterize p62 (Garcia-Blanco et al. 1989) and contains an extended polypyrimidine tract (21 nucleotides in length). Pre-mRNA transcribed from the pPIP7.A plasmid is efficiently spliced and specifically binds the 62-kD pPTB (M.J. Moore and P.A. Sharp, in prep.) under conditions established previously (Garcia-Blanco 1989). The pPIP7.A pre-mRNA was transcribed and radiolabeled to high specific activity under standard conditions (Konarska et al. 1984).

The reaction containing UV cross-linked p62 was treated...
with RNase A and precleared by spinning for 5 min at 12,000g before incubation with immune or preimmune antiserum. The precleared reaction (25 μl) was mixed with a solution containing 25 μl of immune preimmune antiserum and 450 μl of PBSA [phosphate-buffered saline, 0.01% CaCl₂, 0.01% MgCl₂, 1% NP-40 (Sigma Chemical Co.). This mixture was incubated for 30 min on ice and subsequently transferred to a tube containing 50 μl [packed volume] of protein A–agarose beads [protein A–Sepharose CL-4B, Pharmacia LKB Biotechnology, Inc.]. This tube was incubated at 4°C for 30 min on a rotating rack. The p62-antibody complex bound to protein A beads was pelleted by centrifugation for 5 sec on an Eppendorf microcentrifuge at 4°C, and the supernatant was discarded. Unbound proteins were removed from the pellet with six successive washes at 4°C in 1 ml of chilled PBSA, 1% NP-40. The beads were resuspended in the wash buffer, sedimented for 5 sec on the Eppendorf microcentrifuge, and the supernatant from each wash was discarded.

The beads were resuspended in 50 μl of SDS-PAGE loading buffer [0.05 M Tris-Cl (pH 6.8), 0.1 MDTT, 2% SDS, 0.1% bromophenol blue, 10% glycerol], boiled for 10 min, and centrifuged at 12,000g for 1 min, and the supernatant was loaded on a 10% SDS–polyacrylamide gel. The p62 protein was resolved using conditions described by Garcia-Blanco et al. (1989). The 14C-labeled methylated proteins [myosin (200 kD), phosphorylase b (100 and 95 kD), bovine serum albumin (69 kD), ovalbumin (46 kD), and carbonic anhydrase (30 kD)] were obtained from Amersham and were used as molecular mass standards on the SDS–polyacrylamide gels.

Computer analysis of nucleic and amino acid sequences

The predicted amino acid sequence of PTB was compared to the PIR, SWISSPROT, and GENPEPT data bases of the National Center for Biotechnology Information. The BLASTP 1.114MP algorithm, updated January 28, 1991, was used to determine the sequence alignments (Altschul et al. 1990). The computer program PEPTIDESTRUCTURE (Genetics Computer Group, Inc., Madison, WI) was used to determine the predicted secondary structure of pPTB.

To determine whether pPTB contained a significant match to the RNP–CS, the BESTFIT program (Genetics Computer Group, Inc.) was used to find the best alignments to each of the eight possible permutations of the consensus motif [[R/K][GF][G/A][F/V/X][F/Y]]. The quality of the matches was given a numerical value based on exact identities and similarities to the aligned sequences. The significance of these matches was compared to matches found in randomized sequence containing the same length and base composition of the pPTB sequence. Each permutation of the RNP–CS was aligned against 100 different randomized sequences. The defined matches of the RNP–CS found in hnRNP L, elav, PABP, Sex-lethal, Nucleolin, and U1 70K were also analyzed in the same manner. The scores for the matches against the pPTB sequence were compared to the average scores for matches found against 100 trials of randomized sequence and the scores for the defined matches found in hnRNP L, elav, PABP, Sex-lethal, Nucleolin, and U1 70K.

The quality of matches against the RNP–CS found in the specific pPTB sequence ranged between 5.4 and 6.2 and was no better than the quality of matches obtained from nonspecific randomized sequence. The average quality of these matches ranged between 5.8 and 6.3 with a s.d. of 0.5–0.7. Similar to the alignments found in PTB, the quality of matches found against the two RNP–CS present in the hnRNP L protein ranged between 5.0 and 6.1 and was no better than the quality of matches found in randomized sequence. In contrast, the quality of matches to the defined RNP–CS present in elav, PABP, Sex-lethal, Nucleolin, and U1 70K were several s.d. higher and ranged between 8.5 and 10.5.

Acknowledgments

We thank Brian Seed for kindly providing the human placenta cDNA library and Alfred Bothwell for generously sharing unpublished results. We thank Tom Kristie, David Fisher, Jorgen Kiem, Jack Keene, and Melissa Moore for helpful discussion and critical reading of the manuscript; Margarita Siafaca for her generous support and assistance in the preparation of this manuscript; Richard Cook, Michael Kelley, Heather LeBlanc, and Sandy Schultz for their invaluable and expert assistance in synthesis of oligonucleotides and peptides, Will Gilbert and Brian Seed for assistance in the data base computer searches; Melissa Moore for providing the pPIP7.A plasmid, and Lana Parent and Courtney Wood for excellent technical assistance. A.G. and M.G.-B. acknowledge support from National Institutes of Health [NIH] postdoctoral fellowships 1F32GM12426 and 1F32CA08342, respectively. This work was supported by U.S. Public Health Service grants RO1-GM32467 and RO1-GM34277 and partially by Cancer Center Support (core) grant P30-CA14051 from the NIH to P.A.S.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 USC section 1734 solely to indicate this fact.

Note added in proof

The PBT nucleotide and amino acid sequence data described in this paper have been submitted to the EMBL, GenBank, and DDBJ nucleotide sequence data bases.

References


——. 1990b. The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell* 62: 35–42.


Characterization of cDNAs encoding the polypyrimidine tract-binding protein.

A Gil, P A Sharp, S F Jamison, et al.

*Genes Dev.* 1991 5: 1224-1236
Access the most recent version at doi:10.1101/gad.5.7.1224

**References**
This article cites 68 articles, 19 of which can be accessed free at:
http://genesdev.cshlp.org/content/5/7/1224.full.html#ref-list-1

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](http://genesdev.cshlp.org/subscriptions).