Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

Characterization of cDNAs encoding the polypyrimidine tract-binding protein

Anna Gil,[1] Phillip A. Sharp,[1,2] Sharon F. Jamison,[1,3] and Mariano A. Garcia-Blanco[1,3]

[1]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

[2]Corresponding author.
[3]Present address: Duke Comprehensive Cancer Center, Duke University Medical Center, Durham, North Carolina 27710 USA.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

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 splicing 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–Sephacel, 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 pre-mRNA (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-Phe-Gln-Ala-Leu-Leu-Gln-Tyr-Ala-Asp-Pro-Val-Ser-Ala-Gln-His-Ala-Lys-C. The sequence of a second oligopeptide revealed the 8 amino acids N-Ser-Gln-Pro-Ile-Tyr-

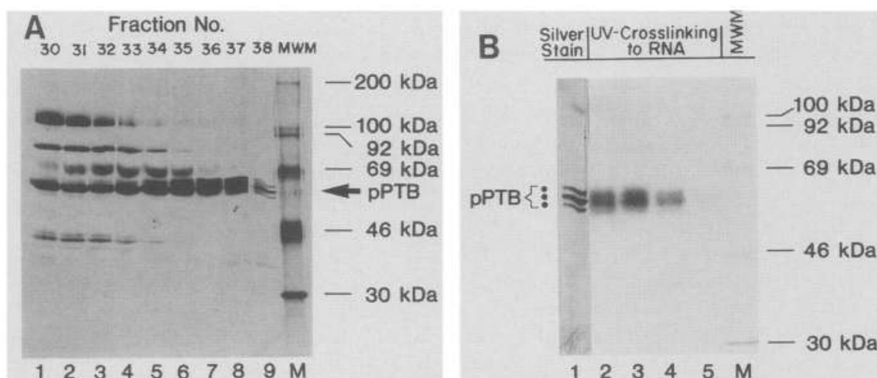
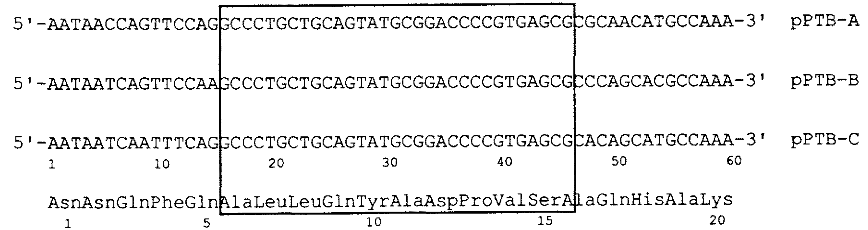


Figure 1. Purification of the 62-kD pPTB. (*A*) The pPTB was eluted from poly(U)–agarose with an increasing KCl gradient. The eluted protein in fractions 30–38 was resolved on an SDS–polyacrylamide gel and silver-stained (lanes *1-9*). pPTB resolved as a triplet of bands at 62 kD and is indicated. (*B*) 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 ^{14}C-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.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

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.



[Phe]-Gln-Phe-C. The phenylalanine in brackets was ambiguously assigned.

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 pBS$^-$. 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 digestions. Six clones, containing inserts ranging in size from 0.85 to 3.1 kb, were selected for further study.

Restriction endonuclease mapping of pC1, pC9, pC11, pC14, pC15, and pC19 revealed that the cDNA segments in all six clones were extensively related. Clone pC15 contained the longest cDNA insert (3.1 kb); thus, both strands of this segment were sequenced. The termini and the majority of the cDNA segments contained in the other five clones were also sequenced. The primary sequence analysis confirmed the overlapping nature of the six cDNA clones. In addition, the sequence of clone pC11 extended the 5′ end of the nucleotide sequence derived from clone pC15 by 23 nucleotides. As expected, all six cDNAs contained sequence encoding the 20 amino acids derived from purified pPTB.

The prototype cDNA sequence is based on pC15 and is 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 57.2 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 PTB 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 AATAAA 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. (▼) 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 PTB3 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.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press


```
1                                                                                   CCGTCCCCCGCGGTCTGCTCTGTGTGCC   28

29     ATG GAC GGC ATT GTC CCA GAT ATA GCC GTT GGT ACA AAG CGG GGA TCT GAC GAG CTT TTC TCT ACT TGT GTC   100
       Met Asp Gly Ile Val Pro Asp Ile Ala Val Gly Thr Lys Arg Gly Ser Asp Glu Leu Phe Ser Thr Cys Val

101    ACT AAC GGA CCG TTT ATC ATG AGC AGC AAC TCG GCT TCT GCA GCA AAC GGA AAT GAC AGC AAG AAG TTC AAA   172
       Thr Asn Gly Pro Phe Ile Met Ser Ser Asn Ser Ala Ser Ala Ala Asn Gly Asn Asp Ser Lys Lys Phe Lys

173    GGT GAC AGC CGA AGT GCA GGC GTC CCC TCT AGA GTG ATC CAC ATC CGG AAG CTC CCC ATC GAC GTC ACG GAG   244
       Gly Asp Ser Arg Ser Ala Gly Val Pro Ser Arg Val Ile His Ile Arg Lys Leu Pro Ile Asp Val Thr Glu

245    GGG GAA GTC ATC TCC CTG GGG CTG CCC TTT GGG AAG GTC ACC AAC CTC CTG ATG CTG AAG GGG AAA AAC CAG   316
       Gly Glu Val Ile Ser Leu Gly Leu Pro Phe Gly Lys Val Thr Asn Leu Leu Met Leu Lys Gly Lys Asn Gln

317    GCC TTC ATC GAG ATG AAC ACG GAG GAG GCT GCC AAC ACC ATG GTG AAC TAC TAC ACC TCG GTG ACC CCT GTG   388
       Ala Phe Ile Glu Met Asn Thr Glu Glu Ala Ala Asn Thr Met Val Asn Tyr Tyr Thr Ser Val Thr Pro Val

389    CTG CGC GGC CAG CCC ATC TAC ATC CAG TTC TCT AAC CAC AAG GAG CTG AAG ACC GAC AGC TCT CCC AAC CAG   460
       Leu Arg Gly Gln Pro Ile Tyr Ile Gln Phe Ser Asn His Lys Glu Leu Lys Thr Asp Ser Ser Pro Asn Gln

461    GCG CGG GCC CAG GCG GCC CTG CAG GCG GTG AAC TCG GTC CAG TCG GGG AAC CTG GCC TTG GCT GCC TCG GCG   532
       Ala Arg Ala Gln Ala Ala Leu Gln Ala Val Asn Ser Val Gln Ser Gly Asn Leu Ala Leu Ala Ala Ser Ala

533    GCG GCC GTG GAC GCA GGG ATG GCG ATG GCC GGG CAG AGC CCT GTG CTC AGG ATC ATC GTG GAG AAC CTC TTC   604
       Ala Ala Val Asp Ala Gly Met Ala Met Ala Gly Gln Ser Pro Val Leu Arg Ile Ile Val Glu Asn Leu Phe

605    TAC CCT GTG ACC CTG GAT GTG CTG CAC CAG ATT TTC TCC AAG TTC GGC ACA GTG TTG AAG ATC ATC ACC TTC   676
       Tyr Pro Val Thr Leu Asp Val Leu His Gln Ile Phe Ser Lys Phe Gly Thr Val Leu Lys Ile Ile Thr Phe

677    ACC AAG AAC AAC CAG TTC CAG GCC CTG CTG CAG TAT GCG GAC CCC GTG AGC GCC CAG CAC GCC AAG CTG TCG   748
       Thr Lys Asn Asn Gln Phe Gln Ala Leu Leu Gln Tyr Ala Asp Pro Val Ser Ala Gln His Ala Lys Leu Ser

749    CTG GAC GGG CAG AAC ATC TAC AAC GCC TGC TGC ACG CTG CGC ATC GAC TTT TCC AAG CTC ACC AGC CTC AAC   820
       Leu Asp Gly Gln Asn Ile Tyr Asn Ala Cys Cys Thr Leu Arg Ile Asp Phe Ser Lys Leu Thr Ser Leu Asn

821    GTC AAG TAC AAC AAT GAC AAG AGC CGT GAC TAC ACA CGC CCA GAC CTG CCT TCC GGG GAC AGC CAG CCC TCG   892
       Val Lys Tyr Asn Asn Asp Lys Ser Arg Asp Tyr Thr Arg Pro Asp Leu Pro Ser Gly Asp Ser Gln Pro Ser

893    CTG GAC CAG ACC ATG GCC GCG GCC TTC GGC CTT TCC GTT CCG AAC GTC CAC GGC GCC CTG GCC CCC CTG GCC   964
       Leu Asp Gln Thr Met Ala Ala Ala Phe Gly Leu Ser Val Pro Asn Val His Gly Ala Leu Ala Pro Leu Ala

965    ATC CCC TCG GCG GCG GCG GCA GCT GCG GCG GCA GGT CGG ATC GCC ATC CCG GGC CTG GCG GGG GCA GGA AAT   1036
       Ile Pro Ser Ala Ala Ala Ala Ala Ala Ala Ala Gly Arg Ile Ala Ile Pro Gly Leu Ala Gly Ala Gly Asn

1037   TCT GTA TTG CTG GTC AGC AAC CTC AAC CCA GAG AGA GTC ACA CCC CAA AGC CTC TTT ATT CTT TTC GGC GTC   1108
       Ser Val Leu Leu Val Ser Asn Leu Asn Pro Glu Arg Val Thr Pro Gln Ser Leu Phe Ile Leu Phe Gly Val

1109   TAC GGT GAC GTG CAG CGC GTG AAG ATC CTG TTC AAT AAG AAG GAG AAC GCC CTA GTG CAG ATG GCG GAC GGC   1180
       Tyr Gly Asp Val Gln Arg Val Lys Ile Leu Phe Asn Lys Lys Glu Asn Ala Leu Val Gln Met Ala Asp Gly

1181   AAC CAG GCC CAG CTG GCC ATG AGC CAC CTG AAC GGG CAC AAG CTG CAC GGG AAG CCC ATC CGC ATC ACG CTC   1252
       Asn Gln Ala Gln Leu Ala Met Ser His Leu Asn Gly His Lys Leu His Gly Lys Pro Ile Arg Ile Thr Leu

1253   TCG AAG CAC CAG AAC GTG CAG CTG CCC CGC GAG GGC CAG GAG GAC CAG GGC CTG ACC AAG GAC TAC GGC AAC   1324
       Ser Lys His Gln Asn Val Gln Leu Pro Arg Glu Gly Gln Glu Asp Gln Gly Leu Thr Lys Asp Tyr Gly Asn

1325   TCA CCC CTG CAC CGC TTC AAG AAG CCG GGC TCC AAG AAC TTC CAG AAC ATA TTC CCG CCC TCG GCC ACT CTG   1396
       Ser Pro Leu His Arg Phe Lys Lys Pro Gly Ser Lys Asn Phe Gln Asn Ile Phe Pro Pro Ser Ala Thr Leu

1397   CAC CTC TCC AAC ATC CCG CCC TCA GTC TCC GAG GAG GAT CTC AAG GTC CTG TTT TCC AGC AAT GGG GGC GTC   1468
       His Leu Ser Asn Ile Pro Pro Ser Val Ser Glu Glu Asp Leu Lys Val Leu Phe Ser Ser Asn Gly Gly Val

1469   GTC AAA GGA TTC AAG TTC TTC CAG AAG GAC CGC AAG ATG GCA CTG ATC CAG ATG GGC TCC GTG GAG GAG GCG   1540
       Val Lys Gly Phe Lys Phe Phe Gln Lys Asp Arg Lys Met Ala Leu Ile Gln Met Gly Ser Val Glu Glu Ala

1541   GTC CAG GCC CTC ATT GAC CTG CAC AAC CAC GAC CTC GGG GAG AAC CAC CAC CTG CGG GTC TCC TTC TCC AAG   1612
       Val Gln Ala Leu Ile Asp Leu His Asn His Asp Leu Gly Glu Asn His His Leu Arg Val Ser Phe Ser Lys

1613   TCC ACC ATC TAG GGGCACAGGCCCACGGCCGGGCCCCCTGGCGACAACTTCCATCATTCCAGAGAAAAGCCACTTTAAAAACAGCTGAAGT   1703
       Ser Thr Ile End

1704   GACCTTAGCAGACCAGAGATTTTATTTTTTTAAAGAGAAATCAGTTTACCTGTTTTTAAAAAAATTAAATCTAGTTCACCTTGCTCACCCTGCGG   1798
1799   TGACAGGACAGCTCAGGCTCTTGGTGACTGTGGCAGCGGGAGTTCCCGGCCCTCCACACCCGGGGCCAGACCCTCGGGGCCATGCCTTGGTGGGG   1893
1894   CCTGTGTCGGGCGTGGGGCCTGCAGGTGGGCGCCCCGACCACGACTTGGCTTCCTTGTGCCTTAAAAAACCTGCCTTCCTGCAGCCACACACCCA   1988
1989   CCCGGGGTGTCCTGGGGACCCAAGGGGTGGGGGGGTCACACCAGAGAGAGGCAGGGGGCCTGGCCGGCTCCTGCAGGATCATGCAGCTGGGGCGC   2083
2084   GGCGGCCGCGCTGCGAGCACCCCAACCCCAGCCCTCTAATCAAGTCACGTGATTCTCCCTTCACCCCGCCCCCAGGGCCTTCCCTTCTGCCCCAG   2178
2179   CGGGCTCCCGCTGCTCAGCTGCGGAGCTGGTCGACATAATCTCTGTATTATATACTTTGCAGTTGCAGACGTCGTGTCCTAGCAATATTTCCAGT   2273
2274   TGACCAAATATTCTAATCTTTTTTCATTTATATGCAAAAGAAATAGTTTTAAGTAACTTTTTATAGCAAGATGATACAATGGTATGAGTGTAATC   2368
2369   TAAACTTCCTTGTGGTATTACCTTGTATGCTGTTACTTTTATTTTATTCCTTGTAATTAAGTCACAGGCAGGACCCAGTTTCCAGAGAGCAGCGG   2463
2464   GGCCGCCCAGTGGGTCAGGCACAGGGAGCCCCGGTCCTATCTTAGAGCCCCTGAGCTTCAGGGAAGGGCGGCGTGTCGCCGCTCTGCATCGCCTC   2558
2559   CGGTTGCCTTACACCACGCCTTCACCTGCAGTCGCCTAGAAAACTTGCTCTCAAACTTCAGGGTTTTTTCTTCCTTCAAATTTTGGACCAAAGTC   2653
2654   TCATTTCTGTGTTTTGCCTGCCTCTGATGCTGGGACCCGGAAGGCGGGCGCTCCTCCTGTCTTTGTGCTCTTTCTACGCCCGCGTCCTGTCCCGG   2748
2749   GGGCTCTCCTAGGATCCCCTTTCCGTAAAAGCGTGTAACAAGGGTGTAAATATTTATAATTTTTTATACCTGTTGTGAGACCCGAGGGGCGGCGG   2843
2844   CGCGGTTTTTTATGGTGACACAAATGTATATTTGCTAACAGCAATTCCAGGCTCAGTACCGCGACCGAGCCAGGGAACCCCACGCACATTCCGT   2938
2939   TGCCTTACCCGATGGCTTGTGACGCGGAGAGAACCGATTAAAACCGTTTGAGAAACTCCTCCCTTGTCTAGCCCTGTGTTCGCTGTGGACGCTGT   3033
3034   AGAGGCAGGTTGGCCAGTCTGTACCTGGACTTCGAATAAATCTTCTGTATCCTCAAA                                        3090
                                                  ******
```

Figure 3. *(See facing page for legend.)*

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

nificant 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 *Eco*RI, *Dra*I, *Pst*I, *Hin*dIII, or *Bam*HI 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, pers. comm.). The 25-kD peptide probably represents a proteolytic fragment from the carboxyl terminus 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

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press


```
900             920                                                                              921                                     961
GACCATGGCCGCGGCCTTCG                                                                              GCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCT   PTB
 ThrMetAlaAlaAlaPhe                                                                              GlyLeuSerValProAsnValHisGlyAlaLeuAlaProLeu

GACCATGGCCGCGGCCTTCG                   CCTCTCCGTATGCAGGAGCTGGTTTCCCTCCCACCTTTGCCATTCCTCAAGCTGCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCT   PTB2
 ThrMetAlaAlaAlaPhe                   AlaSerProTyrAlaGlyAlaGlyPheProProThrPheAlaIleProGlnAlaAlaGlyLeuSerValProAsnValHisGlyAlaLeuAlaProLeu

GACCATGGCCGCGGCCTTCGGTGCACCTGGTATAACTCAGCCTCTCCGTATGCAGGAGCTGGTTTCCCTCCCACCTTTGCCATTCCTCAAGCTGCAGGCCTTTCCGTTCCGAACGTCCACGGCGCCCTGGCCCCCCT   PTB3
ThrMetAlaAlaAlaPheGlyAlaProGlyIleThrGlnProLeuArgMetGlnGluLeuValSerLeuProProLeuProPheLeuLysLeuGlnAlaPheProPheArgThrSerThrAlaProTrpProProTrp
```

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.

homology to the human hnRNP L (Piñol-Roma et al. 1989; Fig. 6). In addition, the pPTB sequence contained limited homology to a number of other proteins containing RNA-binding domains [yeast polyadenylate-binding protein (PABP; Adam et al. 1986; Sachs et al. 1986), chicken Nucleolin (Maridor et al. 1990), *Drosophila* Sex-lethal (Bell et al. 1988), and *Xenopus* U1 70K (Etzerodt et al. 1988)]. pPTB contained limited but broadly distributed regions of homology to the *Drosophila* neuronal protein elav (Robinow et al. 1988). The sequences of both elav and the group of proteins with limited homology to the encoded pPTB contain at least one copy of the 8-amino-acid ribonucleoprotein consensus sequence (R/K)GF(G/A)FVX(F/Y), (RNP–CS; Adam et al. 1986; Swanson et al. 1987). This highly conserved sequence is part of a 90-amino-acid domain, which is loosely conserved in many RNA-binding proteins (Dreyfuss et al. 1988; Bandziulis et al. 1989). A second consensus sequence, LFYGNL (RNP2; Dreyfuss et al. 1988), is part of this domain and is located ~37 residues amino-terminal to the RNP–CS. The predominant regions of sequence homology shared between conceptual pPTB and the PABP, Nucleolin, Sex-lethal, and U1 70K proteins are more related to the RNP2 and the domain between RNP2 and the RNP–CS than to the RNP–CS (see Discussion).

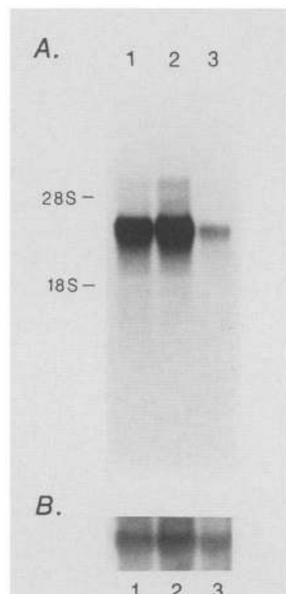


Figure 5. RNA analysis. (*A*) Northern blot containing 5 μg of poly(A)$^+$ RNA isolated from human cell lines. (Lane *1*) Epithelial (HeLa); (lane *2*) T (Jurkat); (lane *3*) B (BJAB). The blot was probed with the 1.4-kb cDNA insert of clone pC14. The size of the PTB mRNA detected in each lane is ~3.5 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*.

Immunoprecipitation of pPTB cross-linked to the polypyrimidine tract of a pre-mRNA splicing substrate

The PTB cDNA encodes a peptide sequence derived from a protein purified on the basis of its ability to bind RNA. Therefore, the method of protein purification and the structural features of the PTB cDNA strongly indicate that it is an RNA-binding protein. Direct evidence was obtained by the use of specific antiserum that the isolated cDNA encodes the pPTB. The specific antiserum was prepared by immunization of rabbits with a synthetic peptide derived from a 20-amino-acid region of the pPTB sequence (see Materials and methods). PTB polypeptide was originally characterized by its specific UV cross-linking to the polypyrimidine tract of the 3′-splice-site region. Radiolabeled pre-mRNA was transcribed from the pPIP7.A plasmid (see Materials and methods) and was incubated with nuclear extract from HeLa cells under splicing conditions (Konarska et al. 1984). pPTB bound to the polypyrimidine tract of this splicing substrate was then cross-linked by UV irradiation. The reaction was digested with RNase A and incubated with either control preimmune or immune serum. As anticipated, a prominent 62-kD protein was specifically labeled in the cross-linking reaction (Fig. 7, lane 1). This labeled polypeptide was immunoprecipitated efficiently by the immune serum (lane 3) but not by the control preimmune serum (lane 2). This strongly suggests that the 62-kD pPTB, characterized previously by its specific binding to the polypyrimidine tract of efficiently spliced pre-mRNAs, corresponds to the polypeptides encoded by the PTB cDNAs.

Discussion

The polypyrimidine tract is essential for recognition of the 3′-splice site of pre-mRNAs that are spliced effi-

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press


```
1     MDGIVPDIAVGTKRGSDELFSTCVTNGPFIMSSNSASAANGNDSKKFKGDSRSAGVPSRVI   61
                          [+ CV++ P IM++ S +  +G+]           [+S V+  hnRNP L
                                  [FIM++ +A+++   ++     + + +A+V++]  ELAV
                                                                     PABP
                                                                     SEX-LETHAL
                                                              [+R +  NUCLEOLIN
                                                                     U1 70k

62    HIRKLPIDVTEGEVISLGLPFGKVTNLLMLKGKNQAFIEMNTEEAANTMVNYYTSVTPVLR  122
      HIR L   V E++++    FG ++ ++++  K QA +E +   +A   VNY +]   [+   hnRNP L
         [LP  +TE E+ SL  + G + ++ +++ K+Q +I+]                       ELAV
                                      [K  +FI  + E A    +    + TP      "
                                                                [++  PABP
                                                                     SEX-LETHAL
       +++LP  VTE E+]                [+K  A+IE +TE  A+]              NUCLEOLIN
                                                                     U1 70K

123   GQPIYIQFSNHKELKTDSSPNQARAQAALQAVNSVQSGNLALAASAAAVDAGMAMAGQSPV  183
      G+P ++ +S     +    + ++++R+]                            [++V   hnRNP L
                                                                     ELAV
                                                                       "
       +PI + FSN]                                                    PABP
      G P+Y+  + +K+++    + Q +A+  ++]           [A+AAA  A+ +M+GQ]     SEX-LETHAL
                                                                     NUCLEOLIN
                                                                     U1 70K

184   LRIIVENLFYPVTLDVLHQIFSKFGTVLKIITFTKNNQFQALLQYADPVSAQHAKLSLDGQ  244
      L    + N  Y++T DVL  I +   G V +I+ F KN]                        hnRNP L
                                  [I    ++N  QA++++    SAQ+AK SL+G    "
                                                                     ELAV
       [I + NL    +     L + FS FG +L]                                PABP
                                                                     SEX-LETHAL
       [+IV+NL Y+ + + L+++F K]                                       NUCLEOLIN
                                                                     U1 70K

245   NIYNACCTLRIDFSKLTSLNVKYNNDKSRDYTRPDLPSGDSQPSLDQTMAAAFGLSVPNVH  305
                                                                     hnRNP L
      +IY++CCTL+I+++K T LNV  N++ ++DYT P+L++]                          "
                                          [D+Q+ L Q+ AAA +++  N      ELAV
                                  [Y  P L+  + ++ L++T A    L V N++   PABP
                                                                     SEX-LETHAL
           [+L+++  KL +L++K ++ KS++]                                 NUCLEOLIN
                                                                     U1 70K

306   GALAPLAIPSAAAAAAAAGRIAIPGLAGAGNSVLLVSNLNPERVTPQSLFILFGVYGDVQR  366
                                  [A ++VL+V  L+  ++   + +F +F +YG+V++  hnRNP L
      +A     A + AAAA]                              [+L  LFG +G VQ   ELAV
      +]                                            [+L+  F+V+GD+    PABP
                                           [L+ N  P+ +T + L+ LF   G +  SEX-LETHAL
                                                [P  +T + L  +FG YG + +   "
                                                                     NUCLEOLIN
                                                         [F VYG + R  U1 70K

367   VKILFNKKENALVQMADGNQAQLAMSHLNGHKLHGKPIRITLSKHQNVQLPREGQEDQGLT  427
      VK + +K   A+V+MADG   + A++HLN + + G  + + +SK+  +]              hnRNP L
      VKI+] [K  + V M++ ++A +A+  LNG  +   + ++++]                    ELAV
                           [+A+ A++ LNG +L+ K I+++ ++]                 "
       KI]                                                           PABP
       +I+]                                                          SEX-LETHAL
       +IL]                                                            "
                                                                     NUCLEOLIN
         I++NK]                                                      U1 70K

428   KDYGNSPLHRFKKPGSKNFQNIFPPSATLHLSNIPPSVSEEDLKVLFSSNGGVVKGFKFFQ  488
     [KD+++S  +RF  P         I  PS  LH  N P  V+EE+]                  hnRNP L
                                     [+A L +S +P ++++++L  +F++ G+++]  ELAV
                        [Q++  +SA+L ++++ PSVSE  L   +FS+ G+V]         PABP
                                     [+ L + NI +  ++E+   LF+  G +V]     "
                                     [+ L ++N+P ++++++L  +F+  G++V]   SEX-LETHAL
                                      [TL + N+P  V+E+++K +F]          NUCLEOLIN
                                                                     U1 70K

489   KDRKMALIQMGSVEEAVQALIDLHNHDLGENHHLRVSFSKSTI                    531
```

Figure 6. Alignment of domains in RNA-binding proteins potentially homologous to the predicted pPTB sequence. Residues 1–531 of the predicted pPTB sequence are numbered. Alignment of protein domains potentially homologous to pPTB are shown below the pPTB sequence and are derived from the data base search of the PIR, SWISSPROT, and GENPEPT libraries with the BLASTP 1.1.114MP algorithm (see Materials and methods). Sequences representing potential homology from one contiguous region are bracketed. Exact matches are denoted by the single-letter amino acid code; similar matches are denoted by a plus sign (+). The identities of the proteins containing sequence homology are listed on the same line and to the *right* of the sequence alignments. The previously defined RNP–CS present in the sequences of the RNA-binding proteins is boxed. The domains containing potential homology to pPTB are derived from amino acid positions 419–440, 69–128, 129–159, 161–199, 193–261, 348–428, and 444–485 of hnRNP L; 3–32, 156–190, 290–328, 15–50, 443–476, 215–239, and 247–275 of elav; 386–418, 128–153, 197–228, 141–157, 31–64, and 219–246 of PABP; 370–387, 407–422, 463–483, 337–359, and 372–391 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.

ciently in vitro. Complementary DNAs encoding a protein (pPTB) that specifically binds to the polypyrimidine tract have been isolated. The mRNAs corresponding to these cDNAs are abundantly expressed in a variety of cell types, suggesting that pPTB is a ubiquitous splicing factor. This is consistent with the hypothesis, derived from mutational analysis, that pPTB promotes the binding of U2 snRNP to pre-mRNA (Garcia-Blanco et al. 1989). The pPTB remains bound to pre-mRNA during formation of the spliceosome.

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,

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

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, PRRRRQLRRQV, 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 (Bothwell 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 (Piñol-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-

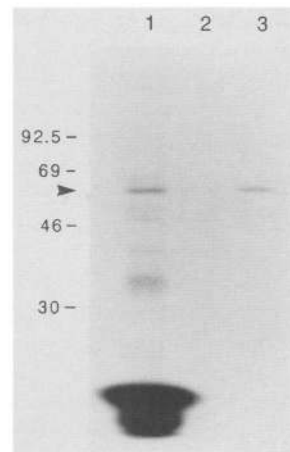


Figure 7. Immunoprecipitation of p62 labeled by UV cross-linking to pre-mRNA. The 62-kD pPTB was cross-linked by UV irradiation to the pPIP7.A RNA splicing substrate in nuclear extract from HeLa cells (see Materials and methods). An aliquot of this reaction (lane *1*) was mixed with either preimmune or immune serum directed against a 20-amino-acid peptide encoded in the PTB cDNA. Protein bound to the antibody was pelleted with protein A–Sepharose (protein A–Sepharose CL-4B, Pharmacia LKB Biotechnology, Inc.). The reactions containing preimmune and immune serum were resolved in lane *2* and lane *3*, respectively. Measurements determined by laser densitometry (LKB Ultroscan XL, Pharmacia LKB Biotechnology, Inc.) indicated that the signal from the immunoprecipitation of p62 protein by immune serum was fourfold higher than the background signal detected in the control reaction containing preimmune sera (cf. lanes *2* and *3*). The immune serum almost quantitatively immunoprecipitated the labeled p62 polypeptide present in the reaction. The level of specific immunoprecipitation reflects the high background of p62 precipitated by the preimmune serum. The same high level of background was observed with many batches of rabbit preimmune sera as well as mouse preimmune sera. The specific immunoprecipitation of the p62 polypeptide was observed repeatedly; it was the only polypeptide specifically enriched in immunoprecipitations by the immune sera. The position of the ^{14}C-labeled methylated protein 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, and p62 is denoted by an arrowhead.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

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, PTB 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 (Piñol-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 (Pinol-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 snRNP 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 snRNP 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'-GGGGAATTCAA(T/C)AA(T/C)CA(G/A)TT(C/T)CA(G/A)GC-3'], B [5'-GGGAAGCTTGG(A/G)TC(A/G/C/T)GC(A/G)TA(C/T)TG-3'], and C [5'-GGGAAGCTTTT(A/G/C/T)GC(A/G)TG(C/T)TG(A/G/C/T)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 overlayed with

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

mineral oil, and the DNA in the mixture denatured at 94°C, reannealed at 35°C, and extended at 72°C for a total of 45 cycles. 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 dideoxynucleotide 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 (500,000) 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 colonies 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 prerinsed in 3× SSC (0.45 M NaCl, 0.045 M sodium citrate), 0.1% SDS; prehybridized for 16 hr at 37°C in 6× SSC, 5× Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 0.05% Na_2HPO_4, 100 μg/ml of herring sperm DNA, 0.5% SDS; and hybridized with a radiolabeled probe (see below) for 16 hr at 50°C in 6× SSC, 1× Denhardt's solution, and 100 μg/ml yeast tRNA. Following hybridization, the filters were washed twice at 50°C for 30 min and twice at 60°C for 30 min in 6× SSC, 0.05% Na_2HPO_4 (pH 6.8).

The 32-nucleotide oligo 5'-GCCCTGCTGCAGTATGCGGACCCCGTGAGCGC-3' spans the common sequence shared between clones pPTB-A, pPTB-B, and pPTB-C and encodes residues 6–15 of the 20-amino-acid tryptic peptide. The oligonucleotide (150 ng) was end-labeled with [γ-^{32}P]ATP by T4 polynucleotide kinase (Boehringer Mannheim) and added to the hybridization solution at a final concentration of 10^6 cpm/ml.

Primary positives (36) were rescreened in the same manner described above. Of these, 22 were colony-purified in a tertiary screen. The DNAs from the 22 positive clones were digested with a limited number of restriction endonucleases to determine the length and orientation of the cDNA insert. The cDNA inserts from six clones (pC1, pC9, pC11, pC14, pC15, and pC19), ranging in size from 0.9 to 3.1 kb, were further characterized by mapping of restriction endonuclease sites.

Both strands of the 3.1-kb cDNA insert of clone pC15 were sequenced. In addition, the termini and >80% of both strands of the five shorter cDNA inserts were sequenced. The dideoxy chain-termination reaction was used to sequence the cDNA inserts (Sanger et al. 1977). Reagents for the reactions were obtained from U.S. Biochemical and used as recommended by the manufacturer. Oligonucleotides complementary to the strands of the cDNA insert were used to prime the sequencing reaction. Oligonucleotides complementary to the vector sequence flanking the cDNA insert were also used to prime the reactions at the termini of the cDNA segment.

Southern blot analysis was performed under the same conditions described for the RNA blot analysis (see below). The genomic DNA was isolated from human epithelial (HeLa) and B (BJAB)-cell lines, and was digested with *Eco*RI, *Dra*I, *Pst*I, *Hin*dIII, and *Bam*HI restriction endonuclease by standard procedures (Maniatis et al. 1982).

RNA analysis

Poly(A)$^+$ RNA was isolated from human epithelial (HeLa), T (Jurkat)-, and B (BJAB)-cell lines with reagents obtained from In Vitrogen; the reagents were used as directed by the manufacturer. Poly(A)$^+$ RNA (5 μg) was loaded per slot and resolved on a 0.8% agarose gel. The RNA was transferred to nylon membrane by capillary blotting for 16 hr in 20× SSC. The membrane was then prehybridized and hybridized to the radiolabeled probe (see below) at 42°C for 16 hr in 1× Denhardt's solution, 0.02 M Na_2HPO_4 (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 2× SSC, 0.1% SDS, and washed twice at 42°C for 30 min in 0.1× SSC, 0.1% SDS. These washes were followed by two washes at 60°C for 30 min in 0.1× SSC, 0.1% SDS.

The DNA of clone pC14 was digested with the restriction endonuclease *Xho*I, and the 1.4-kb cDNA insert was isolated after resolution on a 1% agarose gel. The DNA fragment was uniformly labeled with [α-^{32}P]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 probed with the 1.4-kb fragment of pC14 was reprobed with a fragment from the human β-actin cDNA (Gunning et al. 1983). The plasmid pHFβA-1, containing the human β-actin cDNA insert, was digested with the restriction endonucleases *Eco*RI and *Bam*HI. The 0.7-kb cDNA 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 1× 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 pPIP3 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

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

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% $CaCl_2$, 0.01% $MgCl_2$), 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 M DTT, 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 ^{14}C-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 1.1.114MP 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)FVX(F/Y)]. The quality of the matches was given a numerical value based on exact identities and similarties 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 Kjems, 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 1F32CAO8342, 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 PTB nucleotide and amino acid sequence data described in this paper have been submitted to the EMBL, GenBank, and DDBJ nucleotide sequence data bases.

References

Adam, S.A., T.Y. Nakagawa, M.S. Swanson, T. Woodruff, and G. Dreyfuss. 1986. mRNA polyadenylate-binding protein: Gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* **6:** 2932–2943.

Aebersold, R.H., J. Leavitt, R.A. Saavedra, L.E. Hood, and S.B. Kent. 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci.* **84:** 6970–6974.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Anderson, G., M. Bach, R. Luhrmann, and J.B. Beggs. 1989. Conservation of a U5 snRNP protein from yeast to man. *Nature* **342:** 819–821.

Bandziulis, R.J., M.S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes & Dev.* **3:** 431–437.

Bell, L.R., E.M. Maine, P. Schedl, and T.W. Cline. 1988. Sex-lethal, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55:** 1037–1046.

Black, D.L., B. Chabot, and J.A. Steitz. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. *Cell* **42:** 737–750.

Boggs, R.T., P. Gregor, S. Idriss, J.M. Belote, and M. McKeown. 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the transformer gene. *Cell* **50:** 739–747.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

Bothwell, A.L.M., D.W. Ballard, and W.M. Philbrick. 1990. 25 kd Nuclear protein found in plasmacytoma. EMBL/Genbank Databases EMBL; X52101; MMP25K.

Breitbart, R.E., A. Andreadis, and B. Nadal-Ginard. 1987. Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* **56:** 467–495.

Brody, J. and J. Abelson. 1985. The "spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228:** 1344–1349.

Bugler, B., H. Bourbon, B. Lapeyre, M.O. Wallace, J.H. Chang, F. Amalric, and M.O.J. Olson. 1987. RNA binding fragments from nucleolin contain the ribonucleoprotein consensus sequence. *J. Biol. Chem.* **262:** 10922–10925.

Burd, C.G., M.S. Swanson, M. Gorlach, and G. Dreyfuss. 1989. Primary structures of the hnRNP A2, B1, and C2 proteins: A diversity of RNA binding is generated by small peptide inserts. *Proc. Natl. Acad. Sci.* **86:** 9788–9792.

Choi, Y.D., P.J. Grabowski, P.A. Sharp, and G. Dreyfuss. 1986. Heterogeneous nuclear ribonucleoproteins: Role in RNA splicing. *Science* **231:** 1534–1539.

Dreyfuss, G., M.S. Swanson, and S. Piñol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem. Sci.* **13:** 86–91.

Etzerodt, M., R. Vignali, G. Ciliberto, D. Scherly, I.W. Mattaj, and L. Philipson. 1988. Structure and expression of a Xenopus gene encoding an snRNP protein (U1 70K). *EMBO J.* **7:** 4311–4321.

Frendeway, D. and W. Keller. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42:** 355–367.

Fu, X.-D. and T. Maniatis. 1990. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343:** 437–441.

Garcia-Blanco, M.A., S. Jamison, and P.A. Sharp. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes & Dev.* **3:** 1874–1886.

Garcia-Blanco, M.A., G.J. Anderson, J. Beggs, and P.A. Sharp. 1990. A mammalian protein of 220 kDa binds pre-mRNAs in the spliceosome: A potential homologue of the yeast PRP8 protein. *Proc. Natl. Acad. Sci.* **87:** 3082–3086.

Ge, H. and J.L. Manley. 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. *Cell* **62:** 25–34.

Gerke, V. and J.A. Steitz. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3′ splice site of pre-messenger RNA. *Cell* **47:** 973–984.

Grabowski, P.J., S.R. Seiler, and P.A. Sharp. 1985. A multicomponent complex is involved in the splicing of precursor to mRNAs. *Cell* **42:** 345–353.

Green, M.R. 1986. Pre-mRNA splicing. *Annu. Rev. Genet.* **20:** 671–708.

Gunning, P., P. Ponte, H. Okayama, J. Engel, H. Blau, and L. Kedes. 1983. Isolation and characterization of full-length cDNA clones for human α-, β-, and γ-actin mRNAs: Skeletal but not cytoplasmic actins have an amino-terminal cysteine that is subsequently removed. *Mol. Cell. Biol.* **3:** 787–795.

Helfman, D.M. and W.M. Ricci. 1989. Branch point selection in alternative splicing of tropomyosin pre-mRNAs. *Nucleic Acids Res.* **17:** 5633–5650.

Herrick, G. and B. Alberts. 1976. Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J. Biol. Chem.* **251:** 2124–2132.

Inoue, K., K. Hoshijima, H. Sakamoto, and Y. Shimura. 1990. Binding of the *Drosophila* Sex-lethal gene product to the alternative splice site of transformer primary transcript. *Nature* **344:** 461–463.

Konarska, M.M. and P.A. Sharp. 1986. Electrophoretic separation of complexes involved in the splicing of precursors to mRNA. *Cell* **46:** 845–855.

Konarska, M.M., R.A. Padgett, and P.A. Sharp. 1984. Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell* **38:** 731–736.

Kozak, M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8133.

Krainer, A. and T. Maniatis. 1985. Multiple factors including the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* **42:** 725–736.

Krainer, A.R., G.C. Conway, and D. Kozak. 1990a. Purification and charcterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes & Dev.* **4:** 1158–1171.

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

Kramer, A. 1988. Pre-splicing complex formation requires two proteins and U2 snRNP. *Genes & Dev.* **2:** 1155–1167.

Kramer, A. and W. Keller. 1985. Purification of a protein required for the splicing of premRNA and its separation from the lariat debranching enzyme. *EMBO J.* **4:** 3571–3581.

Lazinski, D., E. Grzadzielska, and D. Asis. 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. *Cell* **59:** 207–218.

Maniatis, T. and R. Reed. 1987. The role of small nuclear ribonucleoprotein particles in pre-RNA splicing. *Nature* **325:** 673–678.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Maridor, G., W. Krek, and E.A. Nigg. 1990. Structure and developmental expression of chicken nucleolin and NO38: Coordinate expression of two abundant non-ribosomal nucleolar proteins. *Biochim. Biophys Acta* **1049:** 126–133.

Mayeda, A., K. Tatei, K. Kitayawa, K. Takeuma, and Y. Ohshima. 1986. Three distinct activities possibly involved in mRNA splicing are found in a nuclear fraction lacking U1 and U2 RNA. *Nucleic Acids Res.* **14:** 3045–3057.

Nagai, K., C. Oubridge, T.H. Jessen, J. Li, and E. Philip. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348:** 515–520.

Padgett, R.A., P.J. Grabowski, M.M. Konarska, S.R. Seiler, and P.A. Sharp. 1986. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55:** 1119–1150.

Patterson, B. and C. Guthrie. 1991. A U-rich tract enhances usage of an alternative 3′ splice site in yeast. *Cell* **64:** 181–187.

Piñol-Roma, S., M.S. Swanson, J.G. Gall, and G. Dreyfuss. 1989. A novel hnRNP protein with a unique distribution on nascent transcripts. *J. Cell. Biol.* **109:** 2575–2587.

Pinto, A. and J. Steitz. 1989. The mammalian analogue of the yeast prp8 splicing protein is present in the U4/5/6 small nuclear ribonucleoproteins. *Proc. Natl. Acad. Sci.* **86:** 8742–8746.

Proudfoot, N.J. and Brownlee G.G. 1976. 3′ Non-coding region sequences of eukaryotic messenger RNA. *Nature* **263:** 211–214.

Query, C.C., R.C. Bentley, and J.D. Keene. 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. *Cell* **57:** 89–101.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

Reed, R. 1989. The organization of 3′ splice-site sequences in mammalian introns. *Genes & Dev.* **3:** 2113–2123.

Robinow, S., A.R. Campos, K.-M. Yao, and K. White. 1988. The elav gene product of *Drosophila*, required in neurons, has three RNP consensus motifs. *Science* **242:** 1570–1572.

Ruby, S. and Abelson, J. 1989. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosomal assembly. *Science* **242:** 1028–1035.

Ruskin, B., P.D. Zamore, and M.R. Green. 1989. A factor U2AF is required for the U2 snRNP binding and splicing complex assembly. *Cell* **52:** 207–219.

Sachs, A.B., M.W. Bond, and R.D. Kornberg. 1986. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: Domain structure and expression. *Cell* **45:** 827–835.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:** 5463–5467.

Seraphin, B. and M. Rosbash. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell* **59:** 349–358.

Seraphin, B., L. Kretzner, and M. Rosbash. 1988. 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.* **7:** 2533–2538.

Siebel, C.W. and D.C. Rio. 1990. Regulated splicing of the Drosophila P transposable element third intron in vitro: Somatic repression. *Science* **248:** 1200–1208.

Sharp, P.A. 1988. RNA splicing and genes. *J. Am. Med. Assoc.* **260:** 3035–3041.

Siliciano, P. and C. Guthrie. 1989. 5′-Splice site selection in yeast: Genetic alterations in base-pairing with U1 reveal additional requirements. *Genes & Dev.* **2:** 1258–1267.

Simmons, D. and B. Seed. 1988. The Fcγ-receptor of natural killer cells is a phospholipid-linked membrane protein. *Nature* **333:** 568–570.

Smith, C.W.J. and B. Nadal-Ginard. 1989. Mutually exclusive splicing of α-tropomyosin exons enforced by an unusual lariat branchpoint location: Implications for constitutive splicing. *Cell* **56:** 749–758.

Sosnowski, B.A., J.M. Belote, and M. McKeown. 1989. Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell* **58:** 449–459.

Swanson, M.S. and G. Dreyfuss. 1988. RNA binding specificity of hnRNP proteins: A subset binds to the 3′ end of introns. *EMBO J.* **7:** 3519–3529.

Swanson, M.S., T.Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. 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.* **7:** 1731–1739.

Tazi, J., C. Alibert, J. Temesamani, I. Reveilland, G. Cathala, C. Brunel, and P. Jeanteur. 1986. A protein that specifically recognizes the 3′-splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. *Cell* **47:** 755–766.

Utans, U. and A. Kramer. 1990. Splicing factor SF4 is dispensable for the assembly of a functional splicing complex and participates in the subsequent steps of the splicing reaction. *EMBO J.* **9:** 4119–4126.

Wieringa, B., E. Hofer, and C. Weissman. 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. *Cell* **37:** 915–925.

Zamore, P.D. and M.R. Green. 1989. Identification, purification, and biochemical characterization of U2 small ribonucleoprotein. *Proc. Natl. Acad. Sci.* **86:** 9243–9247.

Zamore, P.D. and M.R. Green. 1991. Biochemical characterization of U2 snRNP auxiliary factor: An essential pre-mRNA splicing factor with a novel intranuclear distribution. *EMBO J.* **10:** 207–214.

Zapp, M. and S. Berget. 1989. Evidence for nuclear factors involved in recognition of 5′ splice sites. *Nucleic Acids Res.* **17:** 2655–2674.

Downloaded from genesdev.cshlp.org on August 25, 2022 - Published by Cold Spring Harbor Laboratory Press

Characterization of cDNAs encoding the polypyrimidine tract-binding protein.

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

Genes Dev. 1991, **5:**
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

License

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.**

horizon
a PerkinElmer company
Streamline your research with
Horizon Discovery's ASO tool

Copyright © Cold Spring Harbor Laboratory Press