Biochemistry

 Volume 30, Number 15 April 16, 1991

Accelerated Publications

Structure of the 55-kDa Regulatory Subunit of Protein Phosphatase 2A: Evidence for a Neuronal-Specific Isoform[†,‡]

Regina E. Mayer,[§] Peter Hendrix,[∥] Peter Cron,[§] Renate Matthies,[§] Stuart R. Stone,[§] Jozef Goris,[∥] Wilfried Merlevede,[∥] Jan Hofsteenge,[§] and Brian A. Hemmings*[,§]

Friedrich Miescher-Institut, P.O. Box 2543, CH-4002 Basel, Switzerland, and Afdeling Biochemie, Faculteit Geneeskunde, Katholieke Universiteit te Leuven, Leuven, Belgium

Received December 26, 1990; Revised Manuscript Received February 1, 1991

ABSTRACT: The trimeric form of protein phosphatase 2A ($PP2A_1$ or polycation-stimulated protein phosphatase H1) was purified to homogeneity from rabbit skeletal muscle. Preparative SDS–polyacrylamide gel electrophoresis was used to purify the individual subunits with relative molecular masses of 36, 55, and 65 kDa. Sequence analysis of five peptides from the 65-kDa regulatory subunit (PR65) suggested that it was identical with the PR65 subunit derived from the dimeric protein phosphatase $2A_2$. Amino acid sequences derived from the 55-kDa regulatory subunit (PR55) were used to clone human and rabbit cDNAs encoding this protein. The PR55 subunit was found to be encoded by two genes, termed α and β. The open reading frames of the PR55 α and β cDNAs spanned 1341 and 1329 nucleotides, respectively, and predicted proteins with a molecular mass of about 52 kDa that are 86% identical. Comparison of the human PR55 amino acid sequences with the data obtained from the rabbit skeletal muscle protein and a partial rabbit PR55β cDNA clone indicated a high degree of conservation. Analysis of the mRNA expression in human cell lines revealed that the PR55α isoform was encoded by two transcripts of about 2.3 and 2.5 kb and a less abundant 4.4-kb mRNA. Whereas a PR55β transcript of about 2.3 kb was detected at high levels in the neuroblastoma derived cell line LA-N-1, the level of the mRNA was very low in the other human cell lines analyzed. Interestingly, the PR55 sequence showed limited homology to the catalytic domain (domains VI–IX) of the c-*abl* protein tyrosine kinase.

Recent evidence has implicated protein phosphatase 2A (PP2A,[1] also termed the polycation-stimulated protein phosphatase) in the regulation of a diverse number of systems [reviewed by Cohen et al. (1990), Prives (1990), and Lewin (1990)]. Analysis of protein factors required for in vitro SV40 DNA replication revealed that the catalytic subunit of PP2A stimulated the early stages of DNA replication (Virshup et al., 1989). A further link between viral function and PP2A was established by the identification of this phosphatase as one of the cellular proteins that forms stable complexes with the small t antigen of SV40 virus and the small t or middle T antigens of polyoma virus (Pallas et al., 1988, 1990; Walter et al., 1988, 1990). Analysis of the regulation of a number of protein kinases has revealed that PP2A can modulate the activity of phosphorylase *b* kinase (Ramachandran et al., 1987), casein kinase 2 (Agostinis et al., 1987), mitogen-stimulated S6 kinase (Ballou et al., 1988), and MAP-2 kinase (Sturgill et al., 1988; Anderson et al., 1990; Haccard et al., 1990) and thus also indirectly influence the phosphorylation state of their target proteins. On the basis of the effects of

[†] The work of P.H., J.G., and W.M. was partly funded by the Fonds voor Wetenschappelijk Geneeskundig Onderzoek and the G. O. A. from the Katholieke Universiteit Leuven. P.H. was supported for part of this study by a short-term EMBO fellowship.

* Author to whom correspondence should be addressed.

[‡] The nucleotide sequences reported in this paper have been submitted to GenBank under Accession Number J05328.

[§] Friedrich Miescher-Institut.

[∥] Katholieke Universiteit te Leuven.

[1] Abbreviations: PP2A, protein phosphatase type 2A; PR55, 55-kDa regulatory subunit; PR65, 65-kDa regulatory subunit; PVDF, poly(vinylidene difluoride).

0006-2960/91/0430-3589$02.50/0

okadaic acid (a tumor promoter and relatively specific inhibitor of PP2A), PP2A has also been identified as a negative regulator of the dephosphorylation and activation of the $p34^{cdc2}$ protein kinase and therefore as a suppressor of the G2/M transition in the cell cycle (Goris et al., 1989; Felix et al., 1990).

Several holoenzyme forms of PP2A have been purified [reviewed by Ballou and Fischer (1986) and Cohen (1989)] that have a common core structure consisting of the 36-kDa catalytic subunit and a 65-kDa (PR65) regulatory subunit ($PP2A_2$ or PCS_L). In the other forms this core dimer is associated with a 55-kDa (PR55) subunit ($PP2A_1$ or PCS_{H1} and $PP2A_0$) or with a 72-kDa subunit (PCS_M) (Crouch & Safer, 1980; Tamura & Tsuiki, 1980; Tamura et al., 1980; Pato & Adelstein, 1980, 1983; Li, 1981; Paris et al., 1984; Tung et al., 1985; Mumby et al., 1987; Waelkens et al., 1987; Usui et al., 1988). Analysis of the trimeric holoenzymes, $PP2A_0$ and $PP2A_1$, revealed two types of PR55 subunits that apparently lead to different holoenzyme characteristics (Tung et al., 1985). Some evidence has been presented that the PR65 subunit can modulate the activity of the catalytic subunit in vitro depending on the substrate used (Imaoka et al., 1983; Takeda et al., 1985; Chen et al., 1989). The PR55 subunit either does not affect or reduces the phosphatase activity of the dimeric complex of the catalytic and PR65 subunits (Imaoka et al., 1983). The exact role of the different regulatory subunits in vivo has not yet been established; possible roles include suppression of catalytic activity, modulation of substrate specificity, or subcellular localization.

Molecular cloning of the cDNAs encoding the catalytic and PR65 subunits has revealed a complex picture regarding the structure of PP2A holoenzymes (Stone et al., 1987a; Hemmings et al., 1990). There are at least two different genes encoding both subunits that are apparently expressed in a tissue-specific manner (Khew-Goodall & Hemmings, 1988; Mayer et al., manuscript in preparation). In order to learn more about the role of the regulatory subunits in the control of PP2A, we have determined the structure of the PR55 subunit. In this paper we present the complete cDNA sequence of two isoforms encoding the human PR55 subunit. Northern analysis showed that the β isoform was highly expressed in a cell line of neuronal origin whereas the expression of the α isoform had the characteristics of a housekeeping gene. Database searches indicated that the α and β isoforms showed a limited homology to the c-*abl* tyrosine protein kinase.

Experimental Procedures

Protein Purification. The trimeric protein phosphatase $2A_1$ (PCS_{H1}) was purified from rabbit skeletal muscle on the basis of a method originally described for purification of $PP2A_2$ (PCS_L) from *Xenopus* oocytes (Hermann et al., 1988) with the following modifications. Rabbit skeletal muscle (300 g) was homogenized in 1 L of 50 mM Tris-HCl, pH 8.0, 250 mM sucrose, 1 mM EDTA, 1 mM EGTA, 0.5 mM benzamidine, 0.1 mM TLCK (*N*-α-tosyl-L-lysine chloromethyl ketone), 0.1 mM phenylmethanesulfonyl fluoride, and 0.5 mM DTT buffer and centrifuged at 5000*g* for 1 h. The supernatant was absorbed batchwise onto 200 mL of tyrosine–agarose (Sigma) in the presence of 20% (w/v) $(NH_4)_2SO_4$, and the column was washed extensively with buffer A (20 mM Tris-HCl, pH 7.4, 0.5 mM DTT, 1 mM EDTA, 1 mM EGTA, and 0.5 mM benzamidine) containing 20% $(NH_4)_2SO_4$, and eluted with a decreasing salt gradient (500 mL) of 20%–0% $(NH_4)_2SO_4$ in buffer A. The fractions containing PP2A activity were pooled and diluted with H_2O until a conductivity of 10 mS/cm was reached and further purified by chromatography on a DEAE-Sepharose CL-6B column (2 × 15 cm) equilibrated with buffer A, with a 1-L gradient of 0–0.5 M NaCl in buffer A used for elution. The active fractions were pooled, applied to a polylysine–Sepharose-4B (Yang et al., 1980) column (0.6 × 10 cm) and eluted with a 200-mL linear gradient of 0.2–0.7 M NaCl in buffer A. After concentration, the sample was applied to a Mono Q FPLC column (Pharmacia) and eluted with a 0.2–0.5 M NaCl gradient (30 mL) in buffer A. Two peaks of PP2A activity were observed, the first corresponding to $PP2A_1$ and the second to dimeric forms of PP2A. The $PP2A_1$ peak, eluting at 290 mM NaCl, was concentrated and further purified by gel filtration on a Superose 12 column (Pharmacia) (1 × 30 cm; flow rate 0.3 mL/min) equilibrated with 80 mM NH_4HCO_3, pH 7.8.

Assay Methods. Protein phosphatase activity was determined by using phosphorylase *a* as substrate in the presence or absence of 33 μg/mL protamine and 16 mM $(NH_4)_2SO_4$ according to standard procedures (Jessus et al., 1989). Protein concentration was determined by the method of Bradford (1976) with bovine serum albumin as a standard.

Subunit Separation and Protein Sequence Analysis. Purified $PP2A_1$ (80 μg) was separated into its constituent subunits by preparative 10% SDS–polyacrylamide gel electrophoresis followed by electroblotting onto PVDF membranes (Immobilon, Millipore) for 70 min at 70 V in a buffer containing 25 mM Tris, 192 mM glycine, pH 8.3, and 20% (v/v) methanol. The PVDF membranes were stained with 0.1% (w/v) Amido Black in 45% (v/v) methanol and 7% (v/v) acetic acid for 30 s and destained in H_2O. The areas of membrane corresponding to the individual subunits were excised. Proteolytic digestion of membrane-bound subunits was carried out according to the procedure described by Bauw et al. (1989). The PVDF membranes were treated with 500 μL of 0.2% (w/v) poly(vinylpyrrolidone) in methanol for 15 min, followed by addition of 500 μL of H_2O and further incubation for 10 min. After the membranes were washed with digestion buffer, proteolytic cleavage was performed overnight at 37 °C with 2 μg of TLCK-treated trypsin (Cooper Biomedical) in a total volume of 100 μL of 100 mM Tris-HCl, pH 8.0, and 2 mM $CaCl_2$ buffer. The supernatant was removed and the PVDF membrane pieces were washed with 100 μL of 80% formic acid, followed by four washes with 100 μL of H_2O, and all fractions were pooled. Tryptic peptides were separated on a Vydac C_{18} column equilibrated in 0.1% (v/v) trifluoroacetic acid by using a 3-h linear gradient from 0 to 42% (v/v) acetonitrile in the same solvent. Amino acid sequence analysis of the purified peptides was performed on a Applied Biosystems Model 470 or 473 gas-phase sequencer with on-line PTH analysis.

Molecular Cloning and DNA Sequence Analysis. The methods used for the molecular cloning of the PR55 subunit cDNAs have been previously described (Stone et al., 1987a; Hemmings et al., 1990). Two overlapping predicted 27-mer oligonucleotide probes, based on codon usage (Lathe, 1985), corresponding to a 14-residue peptide (STFQSHEPEFDYLK) were annealed and filled in by using [α-^{32}P]dATP with the Klenow fragment of DNA polymerase. The probe was used to screen human and rabbit fetal brain cDNA libraries (Clontech, Palo Alto, CA) as described previously. Positive clones were plaque purified and sequenced. Additional clones were isolated by rescreening these libraries, as well as a human WI38 lung fibroblast cDNA library (Stratagene, La Jolla, CA), using sequenced cDNAs.

All cDNAs were sequenced on both strands by using the dideoxy method (Sanger et al., 1977) and Sequenase (U.S. Biochemical, Cleveland, OH) according to the manufacturer's

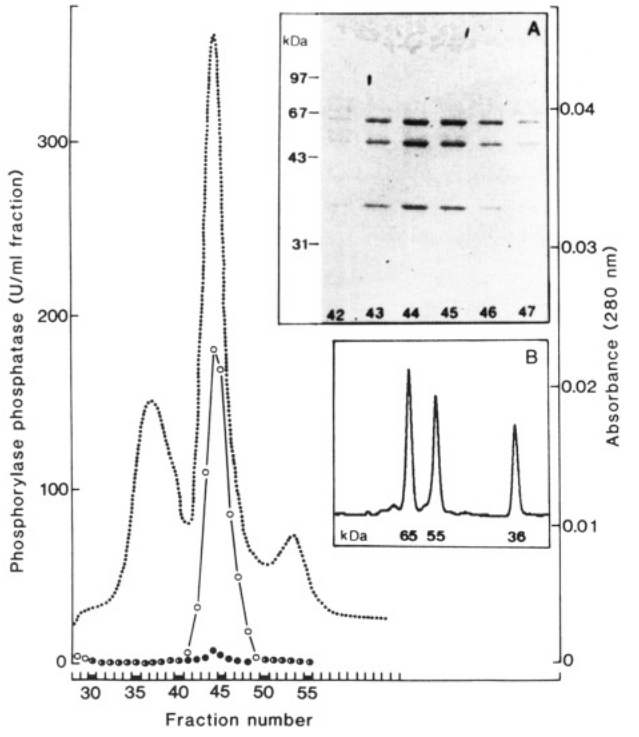


FIGURE 1: Gel-filtration chromatography of protein phosphatase $2A_1$ on Superose 12. Fractions were assayed for phosphorylase phosphatase activity in the absence (closed circles) or presence (open circles) of protamine; the dotted line indicates absorbance at 280 nm. Inset A shows the analysis of individual fractions by SDS–polyacrylamide gel electrophoresis, visualized by Coomassie blue staining. Lane 44 was quantitated by densitometric scanning (inset B).

instructions. Sequence analysis was carried out by using the GCG software package (Devereux et al., 1984) and the FASTA and TFASTA programs (Pearson & Lipman, 1988).

RNA Analysis. RNA isolation from human cell lines (A1146, A-431, Bowes, HeLa, LA-N-1, IMR32, and MCF7) was performed as described previously (Khew-Goodall & Hemmings, 1988). For size estimation, RNA markers of 0.24–9.5 kb (BRL) were used. Total RNA (20 μg) was fractionated on 1% formaldehyde–agarose gels and transferred onto Zeta-Probe nylon membrane (Bio-Rad) by using 20× SSC (1× SSC = 150 mM NaCl and 15 mM sodium citrate, pH 7.0). Gene-specific probes are described in the figure legends. Hybridization conditions were essentially as described previously (Khew-Goodall & Hemmings, 1988). The loading and transfer efficiency was monitored by staining ribosomal RNA on the nylon membrane with 0.05% methylene blue in 0.5 M sodium salicylate, pH 5.2.

RESULTS

Purification of the Trimeric Form of Protein Phosphatase 2A. A modified procedure was developed to isolate the trimeric form of PP2A. The purification involved the use of tyrosine–agarose gel chromatography, followed by conventional steps previously used for the isolation of the dimeric $PP2A_2$ (Hermann et al., 1988; Hemmings et al., 1990). Approximately 60–70 μg of protein was obtained from ~300 g of rabbit skeletal muscle. The final purification step involved size fractionation on a Superose 12 column (Figure 1). $PP2A_1$ eluted as a symmetrical peak of protein and protamine-stimulated phosphorylase phosphatase activity. Analysis of the peak fractions by SDS–polyacrylamide gel electrophoresis (Figure 1, inset A) showed three protein bands of 65, 55, and 36 kDa with an equimolar ratio. This preparation of $PP2A_1$ was subsequently used to generate partial amino acid sequence data for the two regulatory subunits, PR55 and PR65.

Previously we demonstrated that the N-termini of both the catalytic and PR65 subunits were blocked (Stone et al., 1987a; Hemmings et al., 1990). Therefore, the trimeric holoenzyme was used for direct protein sequence analysis. No significant sequence was obtained by using about 50 pmol of $PP2A_1$ indicating that all three subunits were N-terminally blocked. Thus the $PP2A_1$ (~80 μg) was fractionated into its constituent subunits by preparative SDS–polyacrylamide gel electrophoresis. Following transfer to PVDF membranes, both the PR55 and PR65 subunits were digested with trypsin. The resulting tryptic peptides from each subunit were resolved by reverse-phase HPLC. The overall yield of this procedure was about 5–10%.

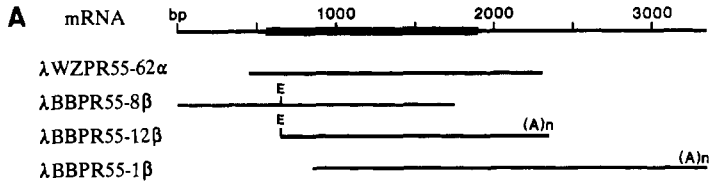
Five peptides isolated from the PR65 subunit were analyzed and yielded the following sequence data: PILDNSTLQSEXVPILE, LAGGDWFTSR, EWAHATIIP, ISHEHSPSDLEAH, and YFAQE. All of the 53 residues determined matched the sequence previously reported for the human PR65 α isoform (Hemmings et al., 1990). The α and β isoforms of PR65 differ at 15 positions in these peptides. Thus the PR65 from both $PP2A_2$ and $PP2A_1$ appears to be the same protein.

Eighteen of the peptides obtained from the PR55 subunit were sequenced with high confidence to yield 140 amino acid residues (PR55 protein sequence data are presented in Figure 4A). Comparison of these sequence data with the Swissprot data base revealed no homology with any known protein.

Molecular Cloning of PR55 Isoforms from Human and Rabbit cDNA Libraries. Two overlapping oligonucleotides, based on codon preference, corresponding to the tryptic peptide STFQSHEPEFDYLK (residues 75–88, Figure 2B) were synthesized and used to screen human and rabbit fetal brain cDNA libraries (see Experimental Procedures for details). These libraries were chosen since previous results showed high levels of the transcripts for the PP2A catalytic and PR65 subunits in brain tissue (Khew-Goodall & Hemmings, 1988; Mayer et al., manuscript in preparation). Several positive clones were identified, plaque-purified, and sequenced.

The deduced amino acid sequences of the human and the partial rabbit cDNA clones were identical (see Figure 4A). However, they showed only 86% identity to the peptide sequences obtained from the rabbit skeletal muscle protein, indicating that a related form of the rabbit muscle PR55 subunit had been cloned. Subsequently, a related cDNA (λWZPR55-62) was isolated from a library derived from WI38 human lung fibroblasts by screening with cDNA probes. The deduced amino acid sequence was identical with the peptide sequences obtained from the rabbit skeletal muscle protein in all but one residue (Figure 4A), suggesting that this cDNA encoded the human homologue of the isolated rabbit protein. The isoform corresponding to the rabbit protein sequence was termed PR55α, and the isoform isolated from the human and rabbit fetal brain cDNA libraries was consequently termed PR55β. Further rescreening of the human fetal brain cDNA library resulted in the isolation of several α isoform cDNAs but they were all found to be derived from partially processed mRNAs.

DNA Sequence Analysis of the Human PR55 Isoforms. The PR55α clone λWZPR55-62 had a noncoding region of 105 bp followed by an open reading frame of 1341 bp and a 3′-nontranslated region of 685 bp (Figure 2B). The open reading frame predicted a protein of 51 745 Da, which was slightly smaller then the mass determined for the isolated rabbit protein. However, SDS–polyacrylamide gel electro-



B

```
-105 CCGCCGCCATCCGCCCTCTCTACCCCCCCATCCCCAGGTGAGGGGGGTGAGTTCAGGAAGCGGAGACCCCGAGGAACCCAGCAGGGTCAC  - 16

 - 15 CATTTGCAGCGCAACATGGCAGGAGCTGGAGGAGGGAATGATATTCAGTGGTGTTTTTCTCAGGTGAAAGGAGCAGTAGATGATGATGTA   75
                      M  A  G  A  G  G  G  N  D  I  Q  W  C  F  S  Q  V  K  G  A  V  D  D  D  V    25

  76 GCAGAAGCAGATATAATTTCTACAGTAGAATTTAATCATTCTGGAGAATTACTAGCAACAGGAGATAAAGGTGGTAGAGTTGTCATCTTT   165
  26 A  E  A  D  I  I  S  T  V  E  F  N  H  S  G  E  L  L  A  T  G  D  K  G  G  R  V  V  I  F    55

 166 CAACAGGAGCAGGAGAACAAAATCCAGTCTCATAGCAGAGGAGAATATAATGTTTACAGCACCTTCCAGAGCCATGAACCAGAGTTTGAC   255
  56 Q  Q  E  Q  E  N  K  I  Q  S  H  S  R  G  E  Y  N  V  Y  S  T  F  Q  S  H  E  P  E  F  D    85

 256 TACTTGAAAAGTTTAGAAATAGAAGAAAAGATCAATAAAATTAGGTGGTTACCCCAGAAAAATGCTGCTCAGTTTTTATTGTCTACCAAT   345
  86 Y  L  K  S  L  E  I  E  E  K  I  N  K  I  R  W  L  P  Q  K  N  A  A  Q  F  L  L  S  T  N   115

 346 GATAAAACAATAAAATTATGGAAAATCAGTGAAAGGGACAAAAGACCAGAAGGGTATAACTTGAAAGAGGAGGATGGAAGGTATAGAGAT   435
 116 D  K  T  I  K  L  W  K  I  S  E  R  D  K  R  P  E  G  Y  N  L  K  E  E  D  G  R  Y  R  D   145

 436 CCTACTACAGTTACTACACTACGAGTGCCAGTCTTTAGGCCTATGGATCTAATGGTTGAGGCCAGTCCACGAAGAATATTTGCCAATGCT   525
 146 P  T  T  V  T  T  L  R  V  P  V  F  R  P  M  D  L  M  V  E  A  S  P  R  R  I  F  A  N  A   175

 526 CATACATATCACATCAACTCAATTTCTATTAATAGTGATTATGAAACATATTTATCTGCAGATGATTTGCGGATTAATCTTTGGCATCTG   615
 176 H  T  Y  H  I  N  S  I  S  I  N  S  D  Y  E  T  Y  L  S  A  D  D  L  R  I  N  L  W  H  L   205

 616 GAAATTACAGACAGGAGTTTTAACATTGTGGATATCAAGCCTGCCAATATGGAAGAGCTAACAGAGGTGATTACAGCAGCAGAATTTCAT   705
 206 E  I  T  D  R  S  F  N  I  V  D  I  K  P  A  N  M  E  E  L  T  E  V  I  T  A  A  E  F  H   235

 706 CCAAACAGCTGTAACACATTTGTATACAGCAGCAGTAAAGGAACTATTCGGCTATGTGACATGAGGGCATCTGCCCTCTGTGATAGACAT   795
 236 P  N  S  C  N  T  F  V  Y  S  S  S  K  G  T  I  R  L  C  D  M  R  A  S  A  L  C  D  R  H   265

 796 TCTAAATTGTTTGAAGAACCTGAAGATCCCAGTAACAGGTCATTTTTTTCCGAAATCATCTCCTCTATTTCGGATGTAAAATTCAGCCAT   885
 266 S  K  L  F  E  E  P  E  D  P  S  N  R  S  F  F  S  E  I  I  S  S  I  S  D  V  K  F  S  H   295

 886 AGTGGTCGATATATGATGACTAGAGACTATTTGTCAGTCAAAATTTGGGACTTAAATATGGAAAACAGGCCTGTGGAAACATACCAGGTG   975
 296 S  G  R  Y  M  M  T  R  D  Y  L  S  V  K  I  W  D  L  N  M  E  N  R  P  V  E  T  Y  Q  V   325

 976 CATGAATACCTCAGAAGTAAACTCTGTTCACTGTATGAAAATGACTGCATATTTGACAAATTTGAATGTTGTTGGAATGGATCTGACAGT  1065
 326 H  E  Y  L  R  S  K  L  C  S  L  Y  E  N  D  C  I  F  D  K  F  E  C  C  W  N  G  S  D  S   355

1066 GTTGTCATGACTGGATCTTACAATAATTTCTTCAGAATGTTTGACAGAAACACAAAGCGAGACATAACCCTAGAAGCATCGCGGGAAAAC  1155
 356 V  V  M  T  G  S  Y  N  N  F  F  R  M  F  D  R  N  T  K  R  D  I  T  L  E  A  S  R  E  N   385

1156 AATAAGCCTCGCACAGTTCTGAAGCCTCGCAAAGTCTGTGCAAGTGGCAAGCGAAAGAAAGATGAAATAAGTGTTGACAGCCTAGACTTC  1245
 386 N  K  P  R  T  V  L  K  P  R  K  V  C  A  S  G  K  R  K  K  D  E  I  S  V  D  S  L  D  F   415

1246 AATAAGAAAATCCTTCACACAGCCTGGCACCCCAAGGAAAATATCATTGCCGTAGCTACTACAAACAATCTGTATATATTTCAAGACAAA  1335
 416 N  K  K  I  L  H  T  A  W  H  P  K  E  N  I  I  A  V  A  T  T  N  N  L  Y  I  F  Q  D  K   445

1336 GTGAATTAGGGTTGGCATTCCTAGCAGAAGAACCCACTTCCTGCTTAGTTGAGATAGTTGAATCTAGCATTCGTTCCTATAAAAGAGAGA  1425
 446 V  N                                                                                        447

1426 GGTCCATTGTGGCGCCCCTTTCCAGTGTTTGACAGTGTGCCATTCGACAACACATTGTTATAGCTACATGGAGAAAGCTCTGTGGATTCA  1515
1516 TCACTGTGGTGTTCTCCATGTCTGCTAGCCATTTAGGTAAGGGTAGGGCACTTTTAATTTAAATGACTTCTTGCACCATCTTGCCTAATG  1605
1606 GACTAGATTGGACTGTATCAACATTGATTTACTCCACTTTTTATGCCTTCCATTGTGATGACGTCAAACACAGTGAAAGCCTTCAGTCAT  1695
1696 GCTATGGGATTTAATTGTGTATCCTCATTACTGTATCATTTGTGGGGTACACCCCTTCCCCCTTTTTTTAAATTAAATACAGCTCATTCT  1785
1786 TACTGTGGCTTGTAGCATTCCTCCTCTTCTGGCCTCCTGGACTGCTCCCCTTCATCTCTTACCCTTGCCCCCTCCACCCGGTCTTGGTGG  1875
1876 TGGTATATTAAAAAAAGAAAGAATGAAAGCACACAAAATGAGTCAGTTTGGGGTCAGTGGTATAAAGGGGGTATATGTTGCAAACAAATG  1965
1966 TTTTAGTAACAGTTGGCTGTAATCACTCCTCGCCGTGTCTGGCACTGAAAATAAGGAAAAG  2026
```

C

```
 -525 GGCCAGGCAAGCCTGAATCCTGTCCCTGCCATCTCGCCACTGCAGCTCGGGTCCAGAAAGGCACCATTTTGTCGCGGCTGCCCGCTCTCC  -436
 -435 CAGGGGGAGGAGGGATCTTTTTTGCATTTTGGAGCGGCTGCCAAGGAGGGGAACCTGTTGGGCATCTCCCCAGACCCGCTTGTGAGCGCC  -346
 -345 TCCGGGGCGGGCGGGCGGGACCAGACCCCTCGGGGCACGGCGTATCTTGGCACCCGGAGGCAGCGGAGGCAGGCGCAGCATCCTCGCTGG  -256
 -255 GAACTGGAGCTGGAGTGAGCGCACCGCGCGGGAGGAGCCGCCGCAGCCTCGCAGAACCCGAGTGGAGGAGGTGACAGCTCCATTGCCGGG  -166
 -165 TTTTTATTTTTTTTCTCTCCGCCTCCCCGTCTCCTCCTCAGGCTCGGACCATGGTGCAGTCCCACTGGCTCCCCTGCCCCCCTCTCCTGT   - 76
                                                  ***
 - 75 GAGACTGGCTGCGGGGAGGGATCATGGATACTTGTCTGCCGGCTTCTGGTTCCCACGCAAGTAAGCCTGCTGTCAATGGAGGAGGACATT    15
                ***                                                                  M  E  E  D  I      5

   16 GATACCCGCAAAATCAACAACAGTTTCCTGCGCGACCACAGCTATGCGACCGAAGCTGACATTATCTCTACGGTAGAATTCAACCACACG   105
    6 D  T  R  K  I  N  N  S  F  L  R  D  H  S  Y  A  T  E  A  D  I  I  S  T  V  E  F  N  H  T    35

  106 GGAGAATTACTAGCGACAGGGGACAAGGGGGGTCGGGTTGTAATATTTCAACGAGAGCAGGAGAGTAAAAATCAGGTTCATCGTAGGGGT   195
   36 G  E  L  L  A  T  G  D  K  G  G  R  V  V  I  F  Q  R  E  Q  E  S  K  N  Q  V  H  R  R  G    65

  196 GAATACAATGTTTACAGCACATTCCAGAGCCATGAACCCGAGTTCGATTACCTGAAGAGTTTAGAAATAGAAGAAAAAATCAATAAAATA   285
   66 E  Y  N  V  Y  S  T  F  Q  S  H  E  P  E  F  D  Y  L  K  S  L  E  I  E  E  K  I  N  K  I    95

  286 AGATGGCTCCCCCAGCAGAATGCAGCTTACTTTCTTCTGTCTACTAATGATAAAACTGTGAAGCTGTGGAAAGTCAGCGAGCGTGATAAG   375
   96 R  W  L  P  Q  Q  N  A  A  Y  F  L  L  S  T  N  D  K  T  V  K  L  W  K  V  S  E  R  D  K   125

  376 AGGCCAGAAGGCTACAATCTGAAAGATGAGGAGGGCCGGCTCCGGGATCCTGCCACCATCACAACCCTGCGGGTGCCTGTCCTGAGACCC   465
  126 R  P  E  G  Y  N  L  K  D  E  E  G  R  L  R  D  P  A  T  I  T  T  L  R  V  P  V  L  R  P   155

  466 ATGGACCTGATGGTGGAGGCCACCCCACGAAGAGTATTTGCCAACGCACACACATATCACATCAACTCCATATCTGTCAACAGCGACTAT   555
  156 M  D  L  M  V  E  A  T  P  R  R  V  F  A  N  A  H  T  Y  H  I  N  S  I  S  V  N  S  D  Y   185
```

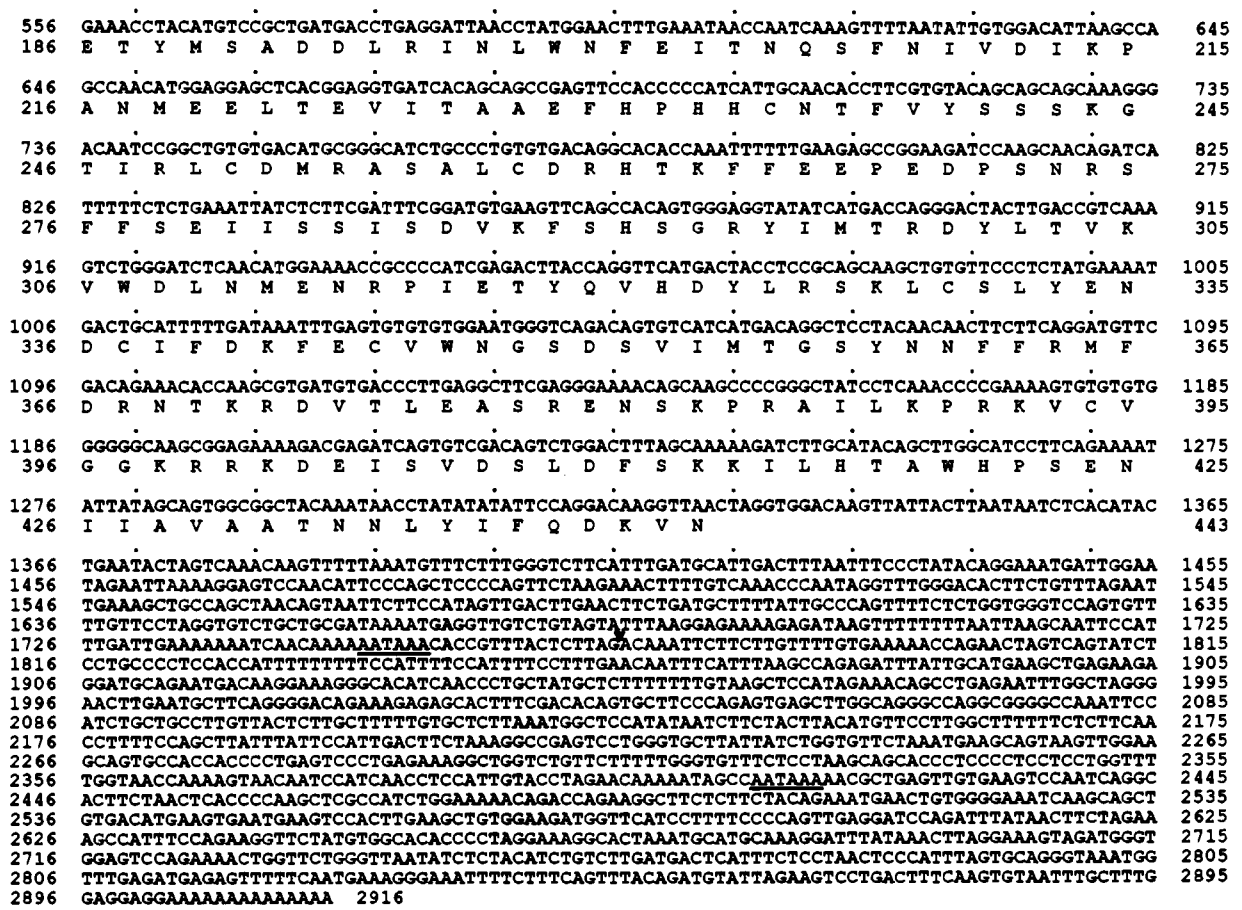


FIGURE 2: Structure and sequence of human PR55 α and β cDNAs. (A) cDNAs encoding the PR55 α and β isoforms. The cDNAs were isolated from a WI38 human lung fibroblast cell line cDNA library (WZ) and a human fetal brain cDNA library (BB). The coding region of the mRNA is indicated by a solid line and the noncoding sequences by a thin line. Internal *Eco*RI (E) restriction sites are indicated. (B) Nucleotide and deduced amino acid sequence for the α isoform of the PR55 subunit (λWZPR55-62α). The nucleotides are numbered starting with the first nucleotide of the initiator codon and nucleotides extending 5′ of base 1 are designated with negative numbers. The deduced amino acid sequence is shown below the nucleotide sequence. (C) Nucleotide and deduced amino acid sequence for the β isoform of the PR55 subunit. Nucleotides are numbered as in (B). The sequence was assembled from the overlapping clones λBBPR55-8 from nt −525 to +99, λBBPR55-12 from +100 to +1770, and λBBPR55-1 from +1770 onward. The locations of the two upstream initiator codons are indicated by a dotted underline and putative poly(A) addition signals by a solid underline. The location of the poly(A) tail of the shorter mRNA (λBBPR55-12) is marked by an inverted triangle.

phoresis of the in vitro transcription and translation product gave a distinct band of about 55 kDa that migrated at the same position as the PR55 from the isolated trimeric PP2A from rabbit skeletal muscle. The nucleotide sequence around the putative initiator codon conforms well with the consensus sequence for translation initiation in vertebrates (Kozak, 1989). The cDNA clone contained no poly(A) addition signal or poly(A) tail, which agrees with the larger size of the detected mRNAs (see below).

The mRNA sequence encoding PR55β was encompassed by several overlapping clones (λBBPR55-8, λBBPR55-12, and λBBPR55-1) (Figure 2A,C). They spanned an open reading frame of 1329 nucleotides encoding a protein of 51 710 Da (see below). The nucleotide sequence of the PR55 α and β isoforms showed 75% identity in the coding region but only 40% in the 5′- and 3′-noncoding regions. λBBPR55-8 contained an unusually long 5′-nontranslated sequence of 525 nucleotides with a stop codon 240 bp upstream of the predicted translation start site. Two potential ATG initiator codons were found 5′ of the putative initiator codon at positions −115 to −113 and −52 to −50. They initiated short open reading frames of 13 and 22 amino acids with the longer peptide overlapping the major open reading frame. Both upstream start codons matched perfectly the consensus sequence by Kozak (1989), whereas the predicted initiator codon had a less favorable context since it lacked a purine at position −3. To confirm that the predicted start codon initiates the correct open reading frame, we performed in vitro transcription and translation analysis using as templates a full-length PR55β cDNA and a cDNA fragment lacking the upstream ATGs (data not shown). The full-length construct gave poor translation from several weak initiation sites, whereas the truncated form gave rise to a major band of about 52 kDa, with a slightly faster electrophoretic mobility than the PR55α in vitro translation product and the purified protein. These data suggest that the third ATG is the major initiation codon and that the upstream ATGs might play a role in modulating the translation of the β mRNA.

Two classes of PR55β cDNAs were isolated that appear to be due to the use of different polyadenylation sites (Figure 2A). λBBPR55-12 contained 441 bases in the 3′-noncoding region and a poly(A) addition signal (AATAAA) (Fitgerald & Shenk, 1981) followed by a 17-nucleotide poly(A) tail. λBBPR55-1 extended 1587 bases into the 3′-noncoding region and was followed by a 14-nucleotide poly(A) tail that was not preceded by a consensus poly(A) addition signal. An additional putative poly(A) addition signal is located at position +2412 to +2417; no cDNA clone containing a poly(A) tail at this position was isolated.

Expression of PR55 Isoforms in Human Cell Lines. The complexity of the transcripts encoding the PR55 isoforms in several human cell lines was investigated by Northern blot

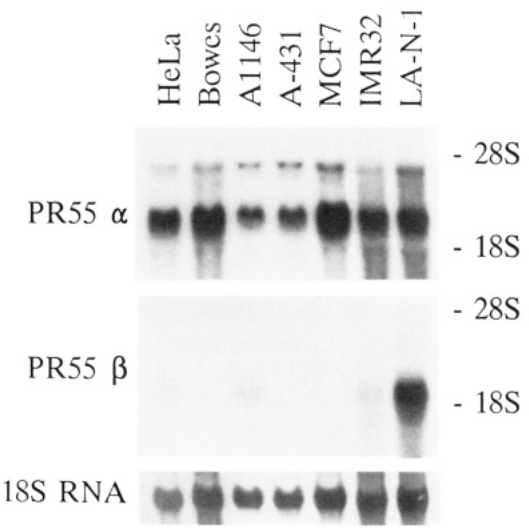


FIGURE 3: Analysis of PR55 α and β mRNA levels in human cell lines. Total RNA (20 μg) from the indicated cell lines was analyzed with PR55 α- and β-specific probes, derived from the 3′ noncoding regions. The α-specific probe was a 800-bp *Hin*cII fragment (nucleotides 1230–2026) from λWZPR55-62α, and the β probe was a 550-bp *Acc*I fragment (nucleotides 1219–1770) from λBBPR55-12β, labeled to specific activities of 4.2×10^9 and 6.4×10^9 cpm/μg, respectively. The filters were washed with 1× SSC/0.1% SDS at 65 °C and exposed at −70 °C with two intensifying screens for 1 and 3 days for the α and β probes, respectively.

analysis using gene-specific probes (Figure 3). Three transcripts encoding the PR55α of about 2.3 and 2.5 kb and a less abundant one of 4.4 kb were detected in all cell lines examined. (Lower exposures and scanning of autoradiograms revealed two α transcripts in the 2.3–2.5-kb hybridization signal.) The PR55β mRNA was highly expressed in the neuroblastoma-derived cell line LA-N-1 at levels approaching the α transcript. In all other cell lines analyzed, including two neuroblastoma-derived cell lines, IMR32 and SK-N-SH (data not shown), as well as a medulloblastoma-derived cell line, TE671 (data not shown), the PR55β transcript was present at much lower levels. Preliminary analysis of the β transcript in RNA isolated from porcine tissues (brain, heart, muscle, kidney, liver, and ovaries) showed only high levels in brain (Mayer et al., manuscript in preparation).

Discussion

In this paper we present the molecular cloning of cDNAs encoding two isoforms of the 55-kDa regulatory subunit of human protein phosphatase 2A. Nucleotide sequence analysis indicated that they were encoded by different genes that show about 75% identity in the coding region but diverge in the 5′- and 3′-noncoding regions. The predicted amino acid sequence was 86% identical between the two isoforms, with the most pronounced difference in the amino terminal region (Figure 4A). Comparison of the predicted amino acid sequence from the human PR55 isoforms with the partial rabbit cDNA clone and the peptide sequences obtained from the rabbit skeletal muscle protein (Figure 4A) indicated a high degree of evolutionary conservation that was similar to that previously reported for the catalytic and PR65 subunits of PP2A (Stone et al., 1987b; Hemmings et al., 1990). This conservation is not restricted to type 2A protein phosphatases but has also been observed for the other serine/threonine-specific protein phosphatases as well as the phosphotyrosine phosphatases (Cohen & Dombradi, 1989; Krueger et al., 1990).

Northern blot analysis revealed that the PR55α transcript was expressed at moderate levels in all human cell lines analyzed (Figure 3). The major 2.3–2.5-kb transcripts probably corresponded to the 2.2-kb human cDNA isolated, while the less abundant 4.4-kb transcript might have been produced by the use of a second polyadenylation site. Alternatively, the longer transcript could represent a partially processed mRNA. Several PR55α cDNA clones containing putative intron sequences, intervening at position +1260 and +1526 (Figure 2B), were isolated. Similarly, partially processed cDNAs containing intron sequences were isolated from a porcine liver cDNA library (unpublished data). This could indicate that regulation of PR55α expression might also occur at the level of mRNA processing. Only the neuroblastoma-derived cell line LA-N-1 showed significant amounts of the PR55β transcript. Analysis of PR55 mRNA expression in different porcine tissues revealed that the β isoform was only expressed in brain (Mayer et al., manuscript in preparation). In contrast, the PR55α transcript was expressed in all tissues examined.

As the PR55β isoform was expressed to a significant extent only in brain, it seems possible that this subunit forms part of a tissue-specific holoenzyme with distinct properties and/or substrate specificities. Patterson and Flavin (1986) purified a MAP-2-specific protein phosphatase from bovine brain with a similar subunit composition as $PP2A_1$ but an apparently different substrate specificity. In this connection, it should be noted that Tung et al. (1985) demonstrated by peptide mapping that two distinct regulatory subunits of respectively 54 and 55 kDa are present in the purified $PP2A_0$ and $PP2A_1$ holoenzymes from rabbit skeletal muscle. However, the predicted peptide maps of the human PR55 α and β isoforms are similar, and the restricted tissue distribution of PR55β makes it unlikely that the 54-kDa subunit of the $PP2A_0$ holoenzyme corresponds to the β isoform. Thus, it is possible that there is another isoform of PR55.

The PR55β mRNA contained two short open reading frames upstream of the predicted initiator codon. Interestingly, two other subunits of PP2A also contain an upstream open reading frame. The human PR65α cDNA (Hemmings et al., 1990) has an ATG at position −62 to −60 with an open reading frame extending into the coding region. Furthermore, an open reading frame that starts near the transcription start site and extends into the coding region was identified in the human catalytic subunit β gene (Khew-Goodall et al., 1991). Other examples of upstream open reading frames have been described for some protoonocogenes and other proteins involved in signal transduction [reviewed by Kozak (1986, 1989)]. The fact that the upstream initiator codons decreased the translation of the correct open reading frame suggests that they might be involved in translational regulation. In higher eucaryotes no direct regulation by upstream ATGs has been reported; however, the *GCN4* gene from yeast, a transcriptional activator of amino acid biosynthesis genes, is negatively regulated at the translational level by upstream open reading frames (Mueller & Hinnebusch, 1986). It remains to be established how upstream open reading frames might contribute to the regulation of the expression of the different subunits of PP2A.

The PR55 subunits showed no extensive similarity to any known protein in the EMBL and NBRF databases (releases 24 and 36, respectively). However, some similarity to the c-Abelson protein tyrosine kinase (Groffen et al., 1983) was found in a restricted region (Figure 4B); over a stretch of 122 amino acids PR55α showed 40% similarity (16% identity) to the c-Abelson kinase sequence, with five of the identical residues being tyrosines. This homology was considered significant because the conserved residues have a high score in a weighted scale based on structural resemblence and genetic

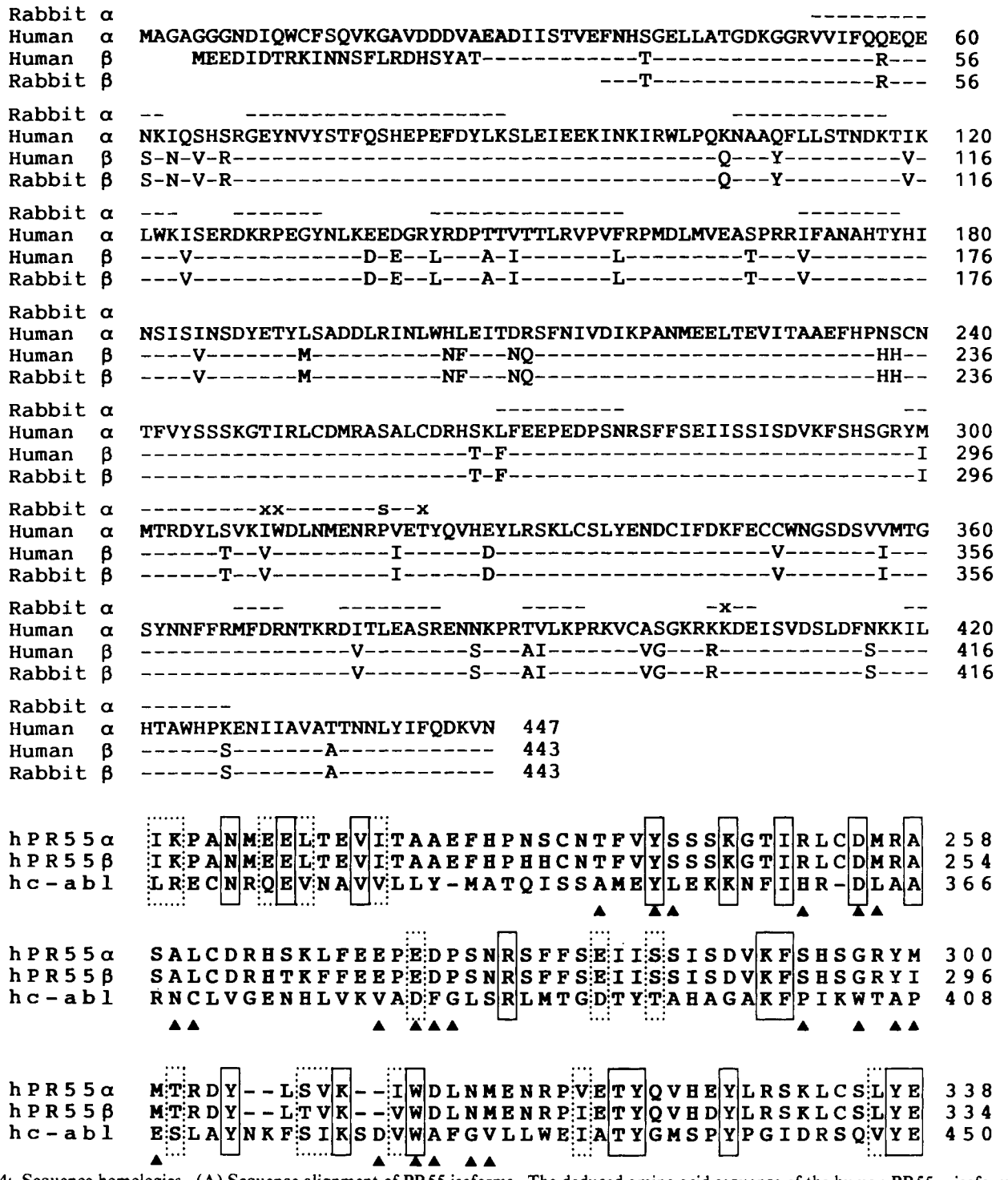


FIGURE 4: Sequence homologies. (A) Sequence alignment of PR55 isoforms. The deduced amino acid sequence of the human PR55 α isoform is presented with the human and rabbit β sequences aligned below and the peptide sequence obtained from the rabbit skeletal muscle protein above. Identities are shown as (-) and differences in the single letter code. Unidentified residues are designated with x. (B) Sequence homology between the PR55 subunit isoforms and the human c-*abl* protein tyrosine kinase. The amino acid sequences were aligned to obtain optimal matches and the numbers indicate the amino acid position in the protein. Identical amino acids are marked with a solid box and structurally related residues with a dotted box. Conserved amino acid residues between tyrosine protein kinase catalytic domains VI–IX are indicated by a triangle.

interconvertibility (Feng et al., 1985) and occurred with the same spacing. The sequence homology region was within the catalytic domains VI–IX of the protein kinase family but does not include all of the conserved residues found in protein kinases (Hanks et al., 1988). This could mean that the PR55 has acquired a domain that is shared with protein kinases or, more likely, that the two proteins have common structural properties involved in the binding of substrates or other proteins. It is interesting to note that tyrosine phosphatase activity of the dimeric PP2A, but not the trimeric forms, can be upregulated by a specific protein termed the phosphotyrosyl phosphatase activator (Cayla et al., 1990a,b); thus it appears that PR55 and the 72-kDa regulatory subunits are inhibitory in this reaction.

The molecular cloning of the different subunits has increased our understanding of the molecular complexity of PP2A holoenzymes. Both the catalytic and PR65 subunits, found in all holoenzyme forms, are encoded by two genes (Stone et al.,

1987a; Hemmings et al., 1990), and we report here the identification of two genes encoding the PR55 subunit. This increases the possibility of association of different subunits to form specific holoenzymes. From protein sequence analysis of the PR65 subunit of both the dimeric (Hemmings et al., 1990) and trimeric form of PP2A (this study), it appears that the PR65α isoform is common to both holoenzymes. Similarly protein sequences derived from the catalytic subunit (Stone et al., 1987b) and the PR55 subunit (this study) correspond to the α isoform of each protein. Thus it remains to be elucidated whether all combinations of subunit isoforms are found in vivo and what the consequence of subunit heterogeneity is on the catalytic activity, substrate specificity, and subcellular localization of the holoenzymes. In this respect, it would be of interest to investigate the subunit composition of PP2A holoenzymes purified from brain because in this tissue the β isoform mRNAs are more abundant than observed in other tissues.

ACKNOWLEDGMENTS

We thank Pam Jones, Roland Ball, and George Thomas for critically reading the manuscript. We are grateful to Marie-Rose Verbiest and Fabienne Withof for protein isolation, Francis Maurer for initial work isolating cDNA clones, and Franz Fischer and Natalie Corman for oligonucleotide synthesis.

REFERENCES

Agostinis, P., Goris, J., Pinna, L. A., & Merlevede, W. (1987) *Biochem. J. 248*, 785–789.

Anderson, N. G., Maller, J. L., Tonks, N. K., & Sturgill, T. W. (1990) *Nature 343*, 651–653.

Ballou, L. M., & Fischer, E. H. (1986) *Enzymes (3rd Ed.) 17*, 311–361.

Ballou, L. M., Siegmann, M., & Thomas, G. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7154–7158.

Bauw, G., Van Damme, J., Puype, M., Vanderkerckhove, J., Gesser, B., Ratz, G. P., Lauridsen, J. B., & Celis, J. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7701–7705.

Bradford, M. M. (1976) *Anal. Biochem. 72*, 248–254.

Cayla, X., Goris, J., Hermann, J., Hendrix, P., Ozon, R., & Merlevede, W. (1990a) *Biochemistry 29*, 658–666.

Cayla, X., Goris, J., Hermann, J., Jessus, C., Hendrix, P., & Merlevede, W. (1990b) *Adv. Enzyme Regul. 30*, 265–285.

Chen, S.-C., Kramer, G., & Hardesty, B. (1989) *J. Biol. Chem. 264*, 7267–7275.

Cohen, P. (1989) *Annu. Rev. Biochem. 58*, 453–508.

Cohen, P. T. W., & Dombradi, V. (1989) *Adv. Protein Phosphatases 5*, 447–463.

Cohen, P., Holmes, C. F. B., & Tsukitani, Y. (1990) *Trends Biochem. Sci. 15*, 98–102.

Crouch, D., & Safer, B. (1980) *J. Biol. Chem. 255*, 7918–7924.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Felix, M. A., Cohen, P., & Karsenti, E. (1990) *EMBO J. 9*, 675–684.

Feng, D. F., Johnson, M. S., & Doolittle, R. F. (1985) *J. Mol. Evol. 21*, 112–125.

Fitzgerald, M., & Shenk, T. (1981) *Cell 24*, 251–260.

Goris, J., Hermann, J., Hendrix, P., Ozon, R., & Merlevede, W. (1989) *FEBS Lett. 245*, 91–94.

Groffen, J., Heisterkamp, N., Reynolds, F. H., & Stephenson, J. R. (1983) *Nature 304*, 167–169.

Haccard, O., Jessus, C., Cayla, X., Goris, J., Merlevede, W., & Ozon, R. (1990) *Eur. J. Biochem. 192*, 633–642.

Hanks, S. K., Quinn, A. M., & Hunter, T. (1988) *Science 241*, 42–52.

Hemmings, B. A., Adams-Pearson, C., Maurer, F., Müller, P., Goris, J., Merlevede, W., Hofsteenge, J., & Stone, S. R. (1990) *Biochemistry 29*, 3166–3173.

Hermann, J., Cayla, X., Dumortier, K., Goris, J., Ozon, R., & Merlevede, W. (1988) *Eur. J. Biochem. 173*, 17–25.

Jessus, C., Goris, J., Staquet, S., Cayla, X., Ozon, R., & Merlevede, W. (1989) *Biochem. J. 260*, 45–51.

Imaoka, T., Imazu, M., Usui, H., Kinohara, N., & Takeda, M. (1983) *J. Biol. Chem. 258*, 1526–1535.

Khew-Goodall, Y., & Hemmings, B. A. (1988) *FEBS Lett. 238*, 265–268.

Khew-Goodall, Y., Mayer, R. E., Maurer, F., Stone, S. R. & Hemmings, B. A. (1991) *Biochemistry 30*, 89–97.

Kozak, M. (1986) *Cell 47*, 481–483.

Kozak, M. (1989) *J. Cell Biol. 108*, 229–241.

Krueger, N. X., Streuli, M., & Saito, H. (1990) *EMBO J. 9*, 3241–3252.

Lathe, R. (1985) *J. Mol. Biol. 183*, 1–12.

Lewin, B. (1990) *Cell 61*, 743–752.

Li, H.-C. (1981) in *Protein Phosphorylation* (Rosen, O. M., & Krebs, E. G., Eds.) Cold Spring Harbor Conference on Cell Proliferation, Vol. 8, Parts A and B, pp 441–457, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mueller, P. P., & Hinnebusch, A. G. (1986) *Cell 45*, 201–207.

Mumby, M. C., Russel, K. L., Garrard, L. J., & Green, D. D. (1987) *J. Biol. Chem. 262*, 6257–6265.

Pallas, D. C., Cherington, V., Morgan, W., DeAnda, J., Kaplan, D., Schaffhausen, B., & Roberts, T. M. (1988) *J. Virol. 62*, 3934–3940.

Pallas, D. C., Shahrik, L. K., Martin, B. L., Jaspers, S., Miller, T. B., Brautigan, D. L., & Roberts, T. M. (1990) *Cell 60*, 167–176.

Paris, H., Ganapathi, M. K., Siberman, S. R., Aylward, J. H., & Lee, E. Y. C. (1984) *J. Biol. Chem. 259*, 7510–7518.

Pato, M. D., & Adelstein, R. S. (1980) *J. Biol. Chem. 255*, 6535–6538.

Pato, M. D., & Adelstein, R. S. (1983) *J. Biol. Chem. 258*, 7047–7054.

Patterson, C. L., Jr., & Flavin, M. (1986) *J. Biol. Chem. 261*, 7791–7796.

Pearson, W. R., & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 2444–2448.

Prives, C. (1990) *Cell 61*, 735–738.

Ramachandran, C., Goris, J., Waelkens, E., Merlevede, W., & Walsh, D. A. (1987) *J. Biol. Chem. 262*, 3210–3218.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Stone, S. R., Hofsteenge, J., & Hemmings, B. A. (1987a) *Biochemistry 26*, 7215–7220.

Stone, S. R., Hofsteenge, J., & Hemmings, B. A. (1987b) *Adv. Protein Phosphatases 4*, 375–389.

Sturgill, T. W., Ray, L. B., Erikson, E., & Maller, J. L. (1988) *Nature 334*, 715–718.

Takeda, M., Usui, H., Imazu, M., Imaoka, T., Kinohara, N., Tsukamoto, H., & Maeta, K. (1985) *Adv. Protein Phosphatases 1*, 215–233.

Tamura, S., & Tsuiki, S. (1980) *Eur. J. Biochem. 111*, 217–224.

Tamura, S., Kikuchi, H., Kikuchi, K., Hiraga, A., & Tsuiki, S. (1980) *Eur. J. Biochem. 104*, 347–355.

Tung, H. Y. L., Alemany, S., & Cohen, P. (1985) *Eur. J. Biochem. 148*, 253–263.

Usui, H., Imazu, M., Maeta, K., Tsukamoto, H., Azuma, K., & Takeda, M. (1988) *J. Biol. Chem. 263*, 3752–3761.
Virshup, D. M., Kauffman, M. G., & Kelly, T. J. (1989) *EMBO J. 8*, 3891–3898.
Waelkens, E., Goris, J., & Merlevede, W. (1987) *J. Biol. Chem. 262*, 1049–1059.
Walter, G., Carbone-Wiley, A., Joshi, B., & Rundell, K. (1988) *J. Virol. 62*, 4760–4762.
Walter, G., Ruediger, R., Slaughter, C., & Mumby, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2521–2525.
Yang, S.-D., Vandenheede, J. R., Goris, J., & Merlevede, W. (1980) *J. Biol. Chem. 255*, 11759–11767.

Structure of the (+)-CC-1065–DNA Adduct: Critical Role of Ordered Water Molecules and Implications for Involvement of Phosphate Catalysis in the Covalent Reaction†

Chin Hsiung Lin,‡ John M. Beale,§ and Laurence H. Hurley*,‡,§

Drug Dynamics Institute and Division of Medicinal Chemistry, College of Pharmacy, and Department of Chemistry, University of Texas at Austin, Austin, Texas 78712

Received January 25, 1991; Revised Manuscript Received February 22, 1991

ABSTRACT: (+)-CC-1065 is an extremely potent antitumor agent produced by *Streptomyces zelensis*. The potent effects of (+)-CC-1065 and its alkylating analogues are thought to be due to the formation of a covalent adduct through N3 of adenine in DNA. It has been previously postulated, on the basis of modeling studies, that a phosphate may be involved in stabilization of the adduct and in acid catalysis of this reaction. In this study, using ^{1}H NMR in combination with ^{17}O-labeled water and phosphate, we demonstrate the involvement of a bridging water molecule between a phenolic proton on the alkylating subunit of (+)-CC-1065 and an anionic oxygen in the phosphate on the noncovalently modified strand of DNA. In addition, a second ordered water molecule associated with one of the protons on N6 of the covalently modified adenine is also identified. This structure has important implications for catalytic activation of the covalent reaction between (+)-CC-1065 and DNA and, consequently, the molecular basis for sequence-selective recognition of DNA by the alkylating subunit of (+)-CC-1065. On the basis of the example described here, the use of ^{1}H NMR in ^{17}O-labeled water may be a powerful probe to examine other structures and catalytic processes for water-mediated hydrogen-bonded bridges that occur between small molecules and DNA or enzymes.

(+)-CC-1065 is an antitumor antibiotic with a unique structure (Hanka et al., 1984; Chidester et al., 1981) (Figure 1) and mechanism of action (Hurley et al., 1984; Warpehoski & Hurley, 1988). Previous studies have demonstrated that this antibiotic is extraordinary for both its base and DNA sequence specificity (Hurley et al., 1984, 1988, 1990; Reynolds et al., 1985). An analogue of (+)-CC-1065 designed and synthesized by Upjohn scientists was recently introduced into phase I clinical trials (J. P. McGovren, The Upjohn Company, personal communication, 1990). Structurally, (+)-CC-1065 consists of three repeated pyrroloindole subunits (A, B, and C in Figure 1) attached via amide linkages that are approximately 15° out of plane, providing the drug molecule with a right-hand twisted banana shape (Hanka et al., 1984; Chidester et al., 1981). Subunit A contains the DNA-reactive cyclopropane ring that alkylates N3 of adenine when it binds within certain reactive sequences (Lin & Hurley, 1990; Scahill et al., 1990) (Figure 1). Since only adenines in certain sequence contexts react with (+)-CC-1065, this drug has sequence selectivity (Hurley et al., 1988). Surprisingly, the A subunit alone contains sufficient structural information to encode the primary molecular basis for sequence selectivity (Hurley et al., 1988), and this subunit is also essential for antitumor activity (Warpehoski et al., 1988). However, as we have previously demonstrated, the noncovalent binding interactions of the B and C subunits with DNA can modulate or fine-tune this sequence selectivity (Hurley et al., 1988). We have previously suggested that the primary basis for sequence selectivity is through *a sequence-dependent catalytic activation and/or a sequence-dependent conformational flexibility* (Warpehoski & Hurley, 1988). In this report we provide structural evidence that is in accord with bifunctional catalysis of adduct formation in a mechanism that involves two critically positioned hydrogen-bonded water molecules at opposite ends of the covalent reaction site.

Materials and Methods

Chemicals. (+)-CC-1065 was obtained from The Upjohn Company and used without further purification. [^{17}O]Water (60 and 45 atom % ^{17}O) was purchased from Cambridge Isotope Laboratories. Reagents used to prepare the NMR buffer, sodium phosphate (99.99%), and sodium chloride (99.99%) were purchased from Aldrich. HPLC water and methanol were purchased from Baxter Scientific and Fisher, respectively. Hydroxylapatite used to purify the 12-mer duplex and the 12-mer adduct was purchased from Calbiochem. Sephadex G-25 (superfine) was purchased from Pharmacia.

† This research was supported by grants from the Public Health Service (CA-49751), the Welch Foundation, and the Burroughs Wellcome Scholars Program.
* To whom correspondence should be addressed.
‡ Department of Chemistry.
§ Drug Dynamics Institute and Division of Medicinal Chemistry, College of Pharmacy.

0006-2960/91/0430-3597$02.50/0