Rapid Identification of Protein Phosphatase 1-binding Proteins by Mixed Peptide Sequencing and Data Base Searching

CHARACTERIZATION OF A NOVEL HOLOENZYMATIC FORM OF PROTEIN PHOSPHATASE 1

(Received for publication, June 10, 1998, and in revised form, July 13, 1998)

Cynthia K. Damer, Jeffrey Partridge, William R. Pearson, and Timothy A. J. Haystead‡

From the Departments of Pharmacology, Markey Center for Cell Signaling and Biochemistry, University of Virginia, Charlottesville, Virginia 22908

Microcystin-affinity chromatography was used to purify 15 protein phosphatase 1 (PP1)-binding proteins from the myofibrillar fraction of rabbit skeletal muscle. To reduce the time and amount of material required to identify these proteins, proteome analysis by mixed peptide sequencing was developed. Proteins are resolved by SDS-polyacrylamide gel electrophoresis, electroblotted to polyvinylidene fluoride membrane, and stained. Bands are sliced from the membrane, cleaved briefly with CnBr, and applied without further purification to an automated Edman sequencer. The mixed peptide sequences generated are sorted and matched against the GenBank using two new programs, FASTF and TFASTF. This technology offers a simple alternative to mass spectrometry for the subpicomolar identification of proteins in polyacrylamide gels. Using this technology, all 15 proteins recovered in PP-1C affinity chromatography were sequenced. One of the proteins, PP-1bp55, was homologous to human myosin phosphatase, MYPT2. A second, PP-1bp80, identified in the EST data bases, contained a putative PP-1C binding site and a nucleotide binding motif. Further affinity purification over ATP-Sepharose isolated PP-1bp80 in a quaternary complex with PP-1C and two other proteins, PP-1bp29 and human p20. Recombinant PP-1bp80 also bound PP-1C and suppressed its activity toward a variety of substrates, suggesting that the protein is a novel regulatory subunit of PP-1.

Protein phosphatase 1 (PP-1, EC 3.13.16) dephosphorylates serine and threonine residues on proteins that control a diverse range of cellular processes from metabolism and muscle contraction to the cell cycle and gene expression (1). The involvement of PP-1 in all of these events raises the question as to how all of these processes can be regulated independently of one another? This question is even more apparent when one considers that the catalytic subunit of PP-1 (PP-1C) is expressed at micromolar concentrations in cells and shares 49% sequence homology within its catalytic core with at least 3 other highly expressed phosphatases, PP-2A, PP-2B, and PP-5 (2–5). The key to this paradox is the finding that the functions of PP-1 are closely linked to its subcellular localization with regulatory targeting subunits that confer substrate specificity (1). At the time of writing, 10 regulatory subunits have been sequenced at the protein and DNA level in mammalian tissues, and at least 10 others identified in Saccharomyces cerevisiae (6). Five of the mammalian subunits potently inhibit the enzyme’s activity toward all substrates, whereas the others target the activity of PP-1C toward myosin, glycogen synthase, p53, or the nucleus (6). The small number of PP-1 regulatory proteins discovered thus far is unlikely to account for all the intracellular actions of the phosphatase, suggesting that many others must exist.

One of the mechanisms by which PP-1C recognizes its regulatory subunits was recently characterized (6). Amino acid sequence alignments of nine of the mammalian regulatory subunits identified the motif (K/R)(V/I)F as common in these proteins. This alignment of 10 PP-1C-binding proteins identified a similar arrangement, (K/R)X(V/I)XF. Co-crystallization studies involving synthetic peptides containing the sequence RRVSF with PP-1C showed that the peptide bound to a hydrophobic channel on the protein surface at the interface of two β sheets opposite to the catalytic cleft. Although (R/K)VF appears to specify a PP-1C binding site, a data base search reveals this motif is in over 10% of all proteins present in the GenBank. Various explanations can be put forward to reduce the possible number of proteins that might bind PP-1C in vivo. First, not all amino acids in the X position of the (R/K)V/IXF sequence or immediately N- or C-terminal to the motif are likely to be tolerated. Second, the (R/K)V/IXF motif may be buried within the folded protein. However, because of its high frequency of occurrence in protein sequences, a data base search using the (R/K)V/IXF motif alone is unlikely to be fruitful in identifying additional PP-1C targeting subunits.

Recently, we utilized a combination of microcystin and PP-1C-affinity chromatography to purify 36 proteins that specifically bound to PP-1C in the cytosolic and particulate fractions of rabbit skeletal muscle (7). Although the large number of proteins recovered by this method supported the hypothesis that the major mechanism of regulation of PP-1 in vivo involves localization of phosphatase catalytic subunit with regulatory proteins that target the enzyme to specific substrates, the identity of most of these proteins was unknown. In this report, we have used a novel sequencing technology (mixed peptide sequencing) to rapidly characterize 15 proteins that were recovered from the particulate fraction of rabbit skeletal muscle. One of the proteins identified, PP-1bp55, is related to the recently cloned human myosin phosphatase, MYPT2 (11). Analysis of the full-length sequence of a second protein, PP-1bp80, identified a PP-1C binding site and a nucleotide binding motif. Further affinity purification over γ-phosphate-linked ATP-Sepharose isolated PP-1bp80 in a quaternary complex...
Identification of Protein Phosphatase-binding Proteins by Mixed Peptide Sequencing and the FASTF and TFASTF Algorithms—Previously, we had shown that microcystin affinity chromatography can be used to isolate fractions from crude cell extracts that are highly enriched in PP-1C and PP-2A binding proteins, but not their catalytic subunits, which remain largely associated with the covalently linked microcystin (7). Using this procedure, over 100 distinct proteins were recovered from the particulate and cytosolic fractions of rabbit skeletal muscle. Additional purification of these complex mixtures over a PP-1C affinity column isolated 20 distinct proteins between residues 101 and 484 and between 205 and 376, respectively.

For insertion into GST fusion vectors, the plasmid DNA from each clone was incubated with the appropriate restriction enzymes in order to remove the insert DNA from the phagemid. Clones 232705 and 31829, as well as the GST fusion vectors pGEX 4T-1 and 4T-3 (Amerham Pharmacia Biotech) were digested with restriction enzymes EcoRI and NotI. Clone 933493 and pGEX 4T-1 were digested overnight at 37 °C with restriction enzymes EcoRI and NotI. The clones were separated on a 1% agarose gel via electrophoresis, excised, and purified. The restriction-digested GST fusion vectors were purified via phenol chloroform extraction. The clones were ligated into the pGEX 4T-1 GST fusion vector, which yielded direct reading frames using T4 DNA ligase (Promega). The ligation reactions were performed at 25 °C for 4 h using a 3:1 ratio of insert to vector DNA, 2 μl of ligation buffer (10×), 1 μl of T4 DNA ligase, and water to bring the final volume to 20 μl. Following ligation, the clones transformed into competent cells. Water (20 μl) was added to the ligation reactions, and 20 μl of this mixture combined with 100 μl of Escherichia coli strain BL21 and incubated on ice for 20 min. The mixture was heat-shocked at 50 °C for 90 s and placed back on ice for 2 min. The cells were added to 1 ml of LB in a 15 ml culture tube and incubated for 1 h at 37 °C. The BL21 cells containing the GST fusion protein plasmid DNA were plated onto LB Amp plates in 10-, 50-, and 100-μl aliquots and incubated at 37 °C overnight. Single colonies were isolated from these plates, which overgrew on the plates, and streaking onto new plates, grew and lyed with 2× crack buffer (100 mM NaOH, 10 mM EDTA pH 8.0, 1% SDS, 10% glycerol, 5 mg of brom cresol green). The plasmid DNA was visualized by electrophoresis on a 1% agarose gel (Fig. 1B).Fig. 1B shows that clone 232705 was ~2000 base pairs, clone 31829 was ~1600 base pairs, and clone 933493 ~1200 base pairs. This analysis demonstrated that each of the clones contained the GST fusion vector and were of the predicted insert size.

For the expression and purification of the recombinant proteins, individual colonies of each clone were grown to an optical density of 0.8 (600 nm) at 30 °C in LB containing ampicillin (100 μg/ml). Isopropyl-1-thio-β-D-galactopyranoside (0.1 mM final) was added to induce expression of the cloned subunit. After 16 h, cells were harvested by centrifugation at 13,000 × g for 10 min. Following lysis of bacterial cells were thawed and lysed in 20 ml of PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.5) containing 0.5 mg/ml lysozyme. Following complete lysis, Nonidet P-40 (40 μl/liter) was added and the lysate centrifuged at 30,000 × g for 30 min. The supernatant was mixed with 1.0 ml of a 1 slurry of GST-Sepharose (Amerham Pharmacia Biotech) for 30 min. The Sepharose was collected by centrifugation for 10 min, and was washed with 1 ml PBS (150 mM NaCl) and 1 ml of PBS with 1 mg/ml of Nonidet P-40. The resin was packed into a glass column (1 × 5 cm) and equilibrated into PBS. One column volume of this buffer containing 1 unit of thrombin was allowed to flow into the column and the flow stopped for 60 min. Flow was resumed and column fractions collected (1 ml). In order to remove the thrombin, the eluted proteins were diluted 20-fold into column buffer A (25 mM Tris-HCl, pH 8.0, 1 mM dithiothreitol, 1 mM benzamidine, 1 mM phenylmethylsulfonyl fluoride), the isothiocyanate eluate was first passed over an avidin-Sepharose column (5 × 5 cm). This step ensured that any proteins that were biotinylated endogenously were removed. The eluate was then passed over a second avidin-Sepharose column (5 × 10 cm) that had been previously saturated with purified PP-1C bound to MC-biotin (8). To remove proteins that bound nonspecifically either through ionic or hydrophobic interactions, the column was washed first with 40 volumes of buffer A containing 1 M NaCl, and then with 40 volumes of buffer A containing no NaCl. The column was re-equilibrated into buffer A, and then one column volume of buffer A containing 10 mM biotin applied and the flow stopped. After 60 min, the flow was resumed. The eluted proteins were characterized by SDS-PAGE and silver staining.

**Expression and Purification of Recombinant PP-1bp80—**A data base search with the full-length sequence of PP-1bp80 identified 13 overlapping human EST sequences covering approximately two thirds of the entire molecule from residue 1 to 484 (Fig. 1A). No EST sequences were identified that matched PP-1bp80 beyond residue 484 in the translated protein. The exact insert size of the reported EST sequences was unknown; therefore, three of the largest reported clones (clones 232705, 31829, and 933493) were obtained from the ATCC for characterization. Bacterial cultures of the clones where plated out onto LB Amp plates and grown overnight at 37 °C. Single colonies were isolated from these plates and 5-ml overnight cultures grown to optical density of 0.8 (600 nm) with water followed by methanol. The pieces were washed once with 5% acetic acid, 10% methanol, and then with several changes of water. The membrane was air-dried and individual stained with sequencer-grade methanol. The transferred proteins were stained with Amido Black (1 mg/ml in 5% acetic acid, 10% methanol) for 1 min, washed once with 5% acetic acid, 10% methanol, and then with several changes of water. The membrane was air-dried and individual stained bands cut out and placed in Eppendorf tubes. The pieces were washed three times with water (1 ml) and then methanol. Carno solution (100 μl, 500 mg/ml in 70% formic acid) was added and digestion carried out for 90 min at room temperature. The reactions were stopped by removal of the membrane and three alternate washes with 1 ml of water followed by methanol. The pieces were placed without further purification into an Applied Biosystems 494 automated sequenator. A single run of Edman sequencing (8–18 cycles of pulsed liquid chemistry) was carried out, and the mixed peptide sequence data generated were sorted and matched to the public data bases with the FASTF or TFASTF algorithms. The FASTF and TFASTF programs can be obtained via FTP (ftp://ftp.virginia.edu/pub/fasta) or by contacting the authors.

**RESULTS**

Identification of Protein Phosphatase-binding Proteins by Mixed Peptide Sequencing and the FASTF and TFASTF Algorithms—Previously, we had showed that microcystin affinity chromatography can be used to isolate fractions from crude cell extracts that are highly enriched in PP-1C and PP-2A binding proteins, but not their catalytic subunits, which remain largely associated with the covalently linked microcystin (7). Using this procedure, over 100 distinct proteins were recovered from the particulate and cytosolic fractions of rabbit skeletal muscle. Additional purification of these complex mixtures over a PP-1C affinity column isolated 20 distinct proteins.
from the cytosolic fraction and another 16 from the particulate fraction. Although the recovery of so many proteins by these methods supported the hypothesis that the catalytic subunits of PP-1 and PP-2A are regulated by multiple and distinct regulatory subunits that target the phosphatases toward specific substrates, the identity of most of these proteins was unknown.

In this present study, we report the development of a novel sequencing technology, mixed peptide sequencing, and the FASTF and TFASTF algorithms, which enabled the rapid identification of proteins in the MC-Sepharose and PP-1C affinity column eluates from the particulate fraction. Initially, the affinity column eluates (1/50th) were characterized by SDS-PAGE and silver staining (Fig. 2, A and B). For mixed peptide sequencing, the eluates (1/50th) were first separated by SDS-PAGE, then electroblotted to PVM. Following staining with Amido Black, the stained proteins were sliced from the membrane, and the slices placed in Eppendorf tubes for treatment with CNBr. After washing, the slices were placed directly into an automated Edman sequencer. Mixed peptide sequences are defined as PTH amino acids that are simultaneously recovered during Edman degradation chemistry of N-terminally blocked proteins that have been cleaved internally at multiple sites. Fig. 3 shows an example of a mixed peptide sequencing run performed on a 16-kDa silver staining protein present in the MC-Sepharose eluate (Fig. 2A). In the first Edman cycle, three major PTH amino acids are recovered: threonine (T), isoleucine (I), and lysine (K). These are written in tabular form in order of picomolar amount recovered. In the second cycle, two major PTH amino acids appeared, glutamic acid (D) and aspartic acid (N). The amount of aspartic acid recovered in this cycle was approximately 1.5-fold greater than the glutamic acid residue.

**Fig. 1. Characterization of PP-1bp80 clones and purification of recombinant proteins.** A, identification of overlapping THC sequences identified by BLAST search and their alignment against the full-length sequence of PP-1bp80. B, crack screen showing the insert size of clones 232705, 318291, and 933493. C, recombinant GST fusion proteins were produced and isolated from glutathione-Sepharose following treatment with thrombin. The cleaved fusion proteins were purified by anion-exchange chromatography and peak fractions analyzed by SDS-PAGE and Coomassie Blue staining. Each protein was verified by N-terminal sequencing on PVM membrane.
indicating that it was derived from two separate peptides. Thus, D, N, and N is recorded for the second cycle. One then proceeds to the third cycle recording amino acids that are recovered and so on until the end of the sequencing run is reached. An X is recorded in cases where an amino acid is not identified. Methionine is added to the beginning of the list of each set of amino acids, as CnBr was used to cleave the protein. The list of amino acids identified in Fig. 3 is written as shown, using a semicolon to separate each set: Call 1, mkdldded; Call 2, minegeae; Call 3, mtntversnl.

The data derived from Fig. 3 suggests that CnBr digestion cleaved the protein internally in three places, as each cycle showed the appearance of three new PTH amino acids. Inasmuch as the membrane slice was derived from a portion of the gel containing a protein that was relatively pure ( > 90%), one would therefore predict that the mixed sequences were derived from at least three separate internal peptides. The list of amino acids in the format written above are entered into the FASTF algorithm and identifies the 16-kDa protein unambiguously in the PVM membrane.

To determine the sensitivity of the methodology, three other proteins of varying staining intensity at 80 kDa (PP-1bp80), 29 kDa (PP-1bp29), and 20 kDa (MLC20) were selected from the gels in Fig. 2 (A and B) for mixed peptide sequencing. Table I shows that TFASTF identified the 80- and 29-kDa proteins as novel proteins that were reconstructed from expressed sequence tags (ESTs) in the DNA data base (see below and Fig. 5). FASTF identified the 20-kDa protein as myosin light chain 20. Recovery of individual PTH amino acids (in pmol) after each cycle was used to determine both the amount of protein present on the membrane following CnBr treatment, as well as the repetitive yield during sequencing. The data in Table I shows that mixed peptide sequencing can be used to identify proteins on PVM over a range from 40 to 0.1 pmol of total protein. Repetitive yield, using the aligned sequences shown in Table I was 95.2 ± 3.2% (S.D.) for the 4 proteins. The average background noise due to nonspecific contaminating amino acids varied from 3 ± 2 fmol (S.D.; n = 12) for the 29-kDa protein to 20 ± 5 fmol (S.D.; n = 6) for MLC20. These data demonstrate that mixed peptide sequencing can unambiguously identify proteins in polyacrylamide gels with a sensitivity that rivals state of the art MALDI time of flight or nano-electrospray mass spectrometry.

To further characterize proteins in the affinity column eluates, mixed peptide sequencing was used to rapidly identify 18 protein identified by FASTF (Fig. 4B). Consistent with a significantly lower expectation score (e = 0.001), and in contrast to calmodulin, presenilin shows poor alignment with any combination of the sorted amino acids. Furthermore, the predicted molecular mass of 44.7 kDa for presenilin is significantly greater than that of the 16-kDa protein excised from the PVM membrane.

**Fig. 2. Identification of phosphatase-binding proteins in SDS gels by mixed peptide sequencing.** Gels A and B are silver stains of proteins present in the isothiocyanate (A) and biotin (B) eluates following MC-affinity chromatography. Gel C shows the results of Far Western analysis of gel B with PP-1C bound to MC-fluorescein (8). The number assignments in B and C indicate proteins that were recovered in PP-1C affinity chromatography and also bound PP-1C MC-fluorescein in Far Western analysis. Proteins were identified in A and B by mixed peptide sequencing. The picomole yield of the highest recovered PTH-amino acid in the first Edman cycle is given for each protein sequenced. The expectation values (e) for identified proteins are compared with the expectation values for the next highest scoring unrelated protein. Generally, data bases searched in the above examples were SwissProt, NBRF, human EST, C. elegans, and yeast.
Identification of PP-1 Subunits by Proteome Analysis

prominently staining proteins in the MC-Sepharose eluate and 16 proteins recovered from PP-1C affinity chromatography. Mixed peptide sequencing data derived from these proteins indicated that on average between 6 and 12 cycles of Edman sequencing was sufficient to unambiguously identify any given protein, provided it was present in the GenBank. The expectation scores for the 36 proteins identified in Fig. 2 ranged from 1.1e-6 to 1.5e-95. This contrasts with expectation scores for the next highest scoring unrelated proteins, which ranged from 3.3 to 9.8e-6 (Fig. 2). In each case the amount of protein that was sequenced was determined by the highest recovered PTH amino in the first cycle and ranged from 60 pmol (PP-1C) to 0.15 pmol (tubulin) (Fig. 2). In most cases, CnBr digestion yielded between three and five simultaneous sequences. This finding is consistent with the frequency of occurrence of methionines in any given protein. In a sample of over 30,000 proteins in the Swiss Prot and PIR1 protein data bases, we determined that 4% of all proteins contain no methionines, 9% of all proteins contain 1 methionine, and 50% of all proteins contain 6 methionines. This suggests that approximately 1:20 proteins are unlikely to cut with CnBr.

The 18 proteins identified in the MC-Sepharose eluate included the catalytic subunits of PP-1C, PP-2AC, and PP-5; several of their regulatory subunits; an uncharacterized EST; as well as a several potential substrate and/or chaperone proteins (Fig. 2A). From previous work (7), the proteins identified in this fraction represent approximately 15–20% of actual number known to be present in the eluate. Therefore, to isolate proteins that bound specifically to phosphatase 1, the MC-Sepharose eluate was repurified by PP-1C-MC-biotin affinity chromatography. Fig. 2B shows that mixed peptide sequencing of the PP-1C affinity column eluate unambiguously identified 9 of the 16 proteins visible in the gel. Two of these proteins were identified as a potential PP-1 regulatory subunit based on their primary amino acid sequence. FASTF aligned mixed peptide sequence data from PP-1bp55 with the N terminus of the myosin phosphatase, MYPT2, which was recently cloned from a human brain library (11). MYPT2 shows approximately 80% homology with the M110 regulatory subunit of smooth muscle myosin phosphatase, SMPP-1M (11). In addition, Northern analysis demonstrated that MYPT2 is highly expressed in skeletal and heart muscle (11). These findings therefore suggest that PP-1bp55 is the skeletal muscle isoform of M110 regulatory subunit of PP-1. The 55-kDa molecular mass of PP-1bp55 is somewhat smaller than the reported 110-kDa mass for full-length sequence of MYPT2, suggesting it is a proteolytic fragment. This finding is reminiscent of the smooth muscle M110 subunit, which shows spontaneous degradation to a stable 58-kDa N-terminal fragment upon purification (12–15). PP-1bp80 was identified by TFFASTF as several uncharacterized EST sequences in the mammalian DNA data bases. Because EST sequences are generally derived from short sequences of cDNA (300–400 base pairs), mixed peptide sequences obtained for these proteins enabled three full-length sequences to be constructed from overlapping EST sequences in the human, mouse, and Caenorhabditis elegans data bases (Figs. 1 and 5).

Inspection of the derived full-length sequence of PP-1bp80 revealed the presence of a conserved PP-1C binding domain in the C-terminal domain and a nucleotide binding motif in its N-terminal half (Fig. 5). Consistent with the presence of a PP-1C binding motif, PP-1bp80 also showed cross-reactivity with the catalytic subunit in Far Western analysis (Fig. 2C). Taken together with its recovery in PP-1C affinity chromatography, these results tentatively identify the protein as a new regulatory subunit of PP-1.

The remaining 7 of the 9 proteins identified in the PP-1C affinity column eluate had strong similarity with protein sequences in the NBRF, SwissProt, or EST data bases and included calsequestrin (sp)cacs_human), tubulin (spptb2_caeel), PP-1Cδ (sp)pp1b_human), a Ca^{2+}-activated ATPase (spptb2_caeel), PP-1bp32 (sppt21_stypl), PP-1bp29 (spAA108607_MOUSE), and human p20 (pir)B53814). The Ca^{2+}-activated ATPase partial sequence also showed strong similarity with the human EST sequence (AA462167, expectation score (e) <10^-25), which clearly encodes a Ca^{2+} ATPase. Interestingly, PP-1bp32 showed significant similarity to the protocadherin Styela plicata protein-tyrosine phosphatase PTP21 (16). PP-1bp32 is of larger molecular mass than PTP21 (32 kDa), suggesting it may be a previously undescribed mammalian tyrosine phosphatase that associates with PP-1C. PP-1bp29 is present as several uncharacterized EST sequences in

Fig. 3. Mixed peptide sequencing of calmodulin. A 16-kDa silver staining protein from Fig. 2A was cut from PVM and digested with CnBr. The PVM piece was placed in the protein sequencer and seven cycles of Edman sequencing carried out. The figure shows the reverse phase high performance liquid chromatograms of the PTH amino acids recovered for each cycle.
the mouse DNA data base (Fig. 5). The protein shows no known homologies with any other protein in the data base. Examination of the primary sequences of the 7 proteins does not, with the exception of PP-1Cε, readily explain their recovery from the PP-1C affinity column. None of them contain the putative PP-1 binding motif IK/RVILXF (6). Furthermore, as shown previously (7), none of these proteins showed cross-reactivity with PP-1C in Far Western analysis (Fig. 2C). Because of the stringent column washing conditions that were employed during affinity purification, their recovery is therefore more likely to be due to a strong interaction with one or more of the other proteins identified in Fig. 2B, rather than PP-1C itself. This finding suggests that these proteins may be physiological targets of PP-1. Ca"²⁺-dependent ATPase, calsequestrin, and tubulin have all been identified as phosphoproteins in skeletal muscle (17–20). Human p20 may also be a target substrate, as it is related to the crystallin family of phosphoproteins (21).

The identity and physiological function of 7 proteins (PP-1b75, PP-1b61, PP-1b60, PP-1b42, PP-1b40, PP-1b34, and PP-1b28) identified in Fig. 2B remains unknown. All 7 proteins produced strong mixed peptide sequence data, but at the time of writing neither FASTF nor TFASTF were able to match them to any known proteins in the GenBank. As shown previously (7), Far Western analysis of the PP-1C affinity column eluate suggests all 7 proteins may bind the catalytic subunit directly (Fig. 2C). This finding, coupled with their recovery from the PP-1C-MC-biotin affinity column, suggests that they may be novel regulatory subunits of PP-1. This hypothesis is currently being investigated. Importantly, the finding that none of the proteins sequenced in the biotin eluate were known PP-2A regulatory subunits, despite the relative abundance of these proteins in the MC-Sepharose eluate, demonstrates that PP-1C-MC-biotin affinity chromatography specifically recovers proteins that interact with PP-1C or its regulatory subunits.

**PP-1b80 Is an ATP-binding Protein That Forms a Holoenzymic Complex with PP-1C, PP-1b29, and Human p20**—Inspection of the predicted amino acid sequence of PP-1b80 identifies, in addition to a PP-1C binding site, several motifs that tentatively suggest its is either a member of the protein kinase family or an ATP-binding protein (Fig. 5). Alignment of the full-length sequence of PP-1b80 with cyclic AMP-depend-
Identification of PP-1 Subunits by Proteome Analysis

Mixed peptide sequence data and cycle yield for four proteins in the microcystin column affinity eluates ranging from 40 pmol to 0.15 pmol.

Data shown are representative examples of four typical proteins that were identified in Fig. 2 by mixed peptide sequencing. Each protein was excised from PVM, treated with CNBr as described under “Materials and Methods,” and placed in an Applied Biosystems 494 automated Edman sequencer linked to an on-line microbore HPLC system for the separation of PTH amino acids. Column 2 shows the major PTH amino acids in pmols that were detected after each cycle of Edman degradation. PTH amino acids that did not change during each cycle were not reported and generally represented a background of <1–20 pmol. Individual PTH amino acids were identified by their retention times relative to purified standards. Column 3 shows the resulting translated sequence using the FASTF or TFASTF algorithms. Column 4 shows the identification of the protein and the data base used. Expectation values for each example are given in Fig. 1. Methionine is included in the translated sequences because CNBr was used to cleave the protein.

### TABLE I

<table>
<thead>
<tr>
<th>Cycle no.</th>
<th>Called amino acid</th>
<th>Translation and alignments</th>
<th>Protein and data base</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>K 37.1</td>
<td>MINEVDAD</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>L 24.6</td>
<td>(27) MINEVDAD (34)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>D 14.2</td>
<td>NKGTDSEE</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D 16.0</td>
<td>(52) MKGTDSEE (58)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>E 18.8</td>
<td>MNTLFGENL</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>D 15.1</td>
<td>(85) MNTLFKEL(92)</td>
<td></td>
</tr>
</tbody>
</table>

 PP-1bp80 is a regulatory subunit of PP-1C—Far Western analysis of the PP-1C-affinity column eluate with PP-1C bound to microcystin fluorescein indicated that PP-1bp80 binds the catalytic subunit of PP-1 directly (Fig. 2C). To test this hypothesis further, recombinant fragments of PP-1bp80 were passed separately over avidin-Sepharose (100 ml) that had previously been saturated with PP-1C bound to MC-biotin. The Sepharose was washed extensively with buffer containing 1 M NaCl, and then the bound proteins eluted with 10 mM biotin. Fig. 6B. Recovery of PP-1bp80 from ATP-Sepharose demonstrates that the protein is an ATP-binding protein. This hypothesis is supported by the finding that elution of the ATP-Sepharose with 10 mM NADH, AMP, or ADP failed to elute any of the proteins identified in Fig. 6. PP-1bp80 is therefore the first example of a PP-1 regulatory subunit that is also a nucleotide-binding protein. Recovery of PP-1bp29 and human p20 in the complex suggest that these proteins bind PP-1bp80 directly, since neither protein interacts with PP-1C in Far Western analysis (Fig. 2C). In addition, neither protein contain motifs that would characterize them as nucleotide-binding proteins.

PP-1bp80 is a Regulatory Subunit of PP-1C—Far Western analysis of the PP-1C-affinity column eluate with PP-1C bound to microcystin fluorescein indicated that PP-1bp80 binds the catalytic subunit of PP-1 directly (Fig. 2C). To test this hypothesis further, recombinant fragments of PP-1bp80 were passed separately over avidin-Sepharose (100 ml) that had previously been saturated with PP-1C bound to MC-biotin. The Sepharose was washed extensively with buffer containing 1 M NaCl, and then the bound proteins eluted with 10 mM biotin. Fig. 6C shows that recombinant protein derived from clone 232705 was recovered in the biotin eluate associated with PP-1C. In contrast, none of the proteins expressed by clones 318291 or 933493 was recovered. These data suggest that recombinant protein expressed by clone 232705 contains a domain that binds PP-1C that is absent in the proteins expressed by 318291 and 933493. Inspection of the encoded sequences of each of the clones indicates that the putative PP-1C binding domain in 232705 lies within the first 101 amino acids of the N terminus of PP-1bp80. Significantly,
Identification of PP-1 Subunits by Proteome Analysis

Amino acid sequence of PP-1bp80, PP-1bp55, and PP-1bp29. The sequences shown were identified after a database search with the TFASTF algorithm from peptide sequences shown in Table I. For PP-1bp80 the full-length sequence (AA1763991) was derived from a 2.2-Mb contiguous nucleotide sequence from chromosome III of C. elegans (23). The human and mouse alignments were derived from overlaps in the GenBank from human EST sequences AA308403, AA324868, AA278895, and AA144032 and mouse EST sequences AA544724, W45835, W171374, and AA409546. For PP-1bp29 the full-length sequence was derived from chicken EST sequence U46756. Overlapping sequences for human and mouse were derived from human ESTs AA109802, AA108607, and AA114777.

The mixed peptide sequence from PP-1bp55 is matched to the N-terminal domain of human MYPT2 (AB003062) (11). Periods indicate amino acid identities; gaps are denoted by substituted amino acids.

Fig. 5. Amino acid sequence of PP-1bp80, PP-1bp55, and PP-1bp29. The sequences shown were identified after a database search with the TFASTF algorithm from peptide sequences shown in Table I. For PP-1bp80 the full-length sequence (AA1763991) was derived from a 2.2-Mb contiguous nucleotide sequence from chromosome III of C. elegans (23). The human and mouse alignments were derived from overlaps in the GenBank from human EST sequences AA308403, AA324868, AA278895, and AA144032 and mouse EST sequences AA544724, W45835, W171374, and AA409546. For PP-1bp29 the full-length sequence was derived from chicken EST sequence U46756. Overlapping sequences for human and mouse were derived from human ESTs AA308403, AA324868, AA278895, and AA144032 and mouse ESTs AA109802, AA108607, and AA114777. The mixed peptide sequence from PP-1bp55 is matched to the N-terminal domain of human MYPT2 (AB003062) (11). Periods denote amino acid identities; gaps are denoted by substituted amino acids.
as the C-terminal portion of PP-1bp80 containing the (K/R)xF/V motif is missing in 232705, its ability to bind PP-1C suggests a second PP-1 binding site in the protein.

To test the effects of the recombinant proteins on PP-1C activity, each was mixed in a 2-fold molar excess with purified PP-1C. The mixtures were then applied to a Waters SW300 gel filtration column equilibrated in buffer A containing 150 mM NaCl. Column fractions were then assayed for nonspecific phosphatase activity using phosphorylase a, myelin basic protein (MBP), and myosin light chain (MLC20) as the substrates (Fig. 6A). The ATP eluate was applied to a G75 superfine column equilibrated in buffer A. Gel B shows a silver stain of the eluted protein from a single peak fraction. The proteins shown in A and B were identified by mixed peptide amino acid sequencing. C, recombinant protein purified from clone 232705 was passed over an avidin-Sepharose column that had previously been saturated with PP-1C bound to MC-biotin. The column was washed extensively and eluted with biotin. Peak column fractions were analyzed by SDS-PAGE and Coomassie Blue staining. D, PP-1C was mixed in the presence (●) and absence (○) of a 2-fold (mol/mol) excess of recombinant PP-1bp80 clone 232705. The mixture was applied to a Waters SW300 gel filtration column and column fractions assayed for phosphorylase phosphatase activity.

**DISCUSSION**

In this present study, we have utilized a novel sequencing method, mixed peptide sequencing, and two new algorithms, FASTF and TFASTF, to rapidly unambiguously identify 36 proteins recovered from microcystin affinity chromatography. The approach identified at least two bona fide PP-1 regulatory subunits that were recovered in PP-1C affinity chromatography. One of these proteins, PP-1bp55, is homologous to human MYPT2, indicating it is the skeletal muscle isoform of M110 subunit of SMPP-1M. A second protein, PP-1bp80, is a previously undescribed PP-1 regulatory subunit that was identified in the DNA data bases as several uncharacterized EST sequences. Inspection of the full-length reconstructed sequence of PP-1bp80 identified the (K/R)xF/V motif, which is conserved in 10 other mammalian and yeast PP-1 regulatory subunits (6). Two additional lines of evidence suggest PP-1bp80 is a PP-1 regulatory subunit. First, PP-1bp80 was recovered as a quaternary complex with PP-1C, PP-1bp29, and human p20 following ATP-Sepharose affinity chromatography. Densitometry of the purified complex indicated that all 4 proteins were recovered in the complex with a stoichiometry of 1:1:1:1. Second, a truncated form of recombinant PP-1bp80 bound PP-1C and altered its specific activity toward three phosphoprotein substrates. PP-1bp80 is therefore the second example of a PP-1 regulatory subunit that alters the specific activity of PP-1C toward nonspecific substrates. This phenomenon was first observed for the M110 subunit that targets PP-1C toward smooth muscle myosin (12–14). In this case, the M110 subunit functions to reduce nonspecific phosphatase activity while enhancing smooth muscle myosin phosphatase activity 10–20-fold (12–14). Although the physiological substrate for the PP-1bp80 complex has yet to be identified, by analogy with smooth muscle M110, one might predict that the 80-kDa subunit will greatly enhance the ability of PP-1C to bring about its dephosphorylation. Human p20 and PP-1bp29 are currently being investigated as candidate substrates for the PP-1bp80 complex. Human p20 is related to the crystallin family of phosphoproteins (21). The crystallins are ubiquitously expressed proteins, which, in addition to forming the lens material of the eye (21), have also been implicated as stress response proteins. The crystallins are also unique in that they can convert from a liquid to gel like state whenrophospho-
In addition to identifying two of chaperone proteins like the crystallins in stress situations, the function for PP-1bp80 in controlling the phosphorylation state p20 with PP-1bp80, PP-1bp29, and PP-1C points to a possible mechanism of regulation in which phosphatase activity could be controlled in response to changes in intracellular ATP concentrations. Cellular stress is one scenario in which intracellular ATP concentrations could be altered. The co-purification of human p20 with PP-1bp80, PP-1bp29, and PP-1C in controlling the phosphorylation state of chaperone proteins like the crystallins in stress situations. In addition to identifying two bona fide PP-1 regulatory subunits, seven others were tentatively identified in this present study by their recovery in PP-1C affinity chromatography and far Western analysis. All of these proteins appear to be novel, because at the time of writing neither FASTF or TFASTF were described is not limited only to the identification of proteins per se, but, as shown in the case of PP-1, enables meaningful functional assignments to be made when combined with Far Western analysis or screening for binding partners with a purified protein.

REFERENCES
