JOURNAL OF BACTERIOLOGY, Nov. 1995, p. 6510–6517 Vol. 177, No. 22
0021-9193/95/$04.00+0

Isolation and Cloning of a Protein-Serine/Threonine Phosphatase from an Archaeon

JIE LENG,[1] ANGUS J. M. CAMERON,[1] SCOTT BUCKEL,[2] AND PETER J. KENNELLY[1]*

Department of Biochemistry and Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0308,[1] and Pittman-Moore Company, Terre Haute, Indiana 47802[2]

Received 10 July 1995/Accepted 1 September 1995

A divalent metal ion-stimulated protein-serine/threonine phosphatase, PP1-arch, was purified approximately 1,000-fold from the extreme acidothermophilic archaeon *Sulfolobus solfataricus* (ATCC 35091). Purified preparations contained 40 to 70% of total protein as PP1-arch, as determined by assay-ing sodium dodecyl sulfate-polyacrylamide gels for protein phosphatase activity. The first 25 amino acids of the protein's sequence were identified, as well as an internal sequence spanning some 20 amino acids. Using this information, we cloned the gene for PP1-arch via the application of PCR and conventional cloning techniques. The gene for PP1-arch predicted a protein of 293 amino acids that bore striking resemblance to the members of the major family of protein-serine/threonine phosphatases from members of the domain *Eucarya*, the PP1/2A/2B superfamily. The core of the protein, spanning residues 4 to 275, possessed 29 to 31% identity with these eucaryal protein phosphatases. Of the 42 residues found to be absolutely conserved among the known eucaryal members of the PP1/2A/2B superfamily, 33 were present in PP1-arch. If highly conservative substitutions are included, this total reached 37. The great degree of sequence conservation between molecules from two distinct phylogenetic domains implies that the members of this enzyme superfamily had evolved as specialized, dedicated protein phosphatases prior to the divergence of members of the *Archaea* and *Eucarya* from one another.

Protein phosphorylation-dephosphorylation represents a ubiquitous and prominent mechanism for regulating cellular processes. Protein kinases, protein phosphatases, and the phosphoproteins that they modify have been found in virtually every living organism. A priori, such universality might suggest that protein phosphorylation-dephosphorylation constitutes a very ancient means for exerting regulatory control. However, for many years, comparisons between the components of bacterial and eucaryal protein phosphorylation networks showed them to be structurally distinct, suggesting that each arose independently, at different times and places (reviewed in reference 18). For example, while all eucaryal protein kinases could be grouped into a single giant superfamily (14), the protein kinases from bacterial organisms showed no significant resemblance to this group (9). While members of the domain *Eucarya* phosphorylated some proteins on tyrosine residues, members of the domain *Bacteria* apparently did not (13). Therefore, it was and still is widely believed that protein phosphorylation-dephosphorylation constitutes a relatively recent addition to the cell's regulatory repertoire, one that arose separately in each phylogenetic domain after it had diverged from its predecessors.

Recent evidence has eroded the foundation upon which this consensus is based. Tyrosine-phosphorylated proteins have been detected in several members of the *Bacteria* (1, 10, 33, 34, 43). Genes encoding bacterial protein-histidine kinases have been discovered in eucaryal species ranging from rats (31, 32) to plants (6) to yeasts (23, 28). Eucaryal protein kinases (24, 25, 42, 44) and a eucaryal dual-specificity protein phosphatase (33) have been discovered in the genomes of the *Bacteria*. Although fragmentary, this trail of individual domain crossover events raises the possibility that all or part of present-day protein phosphorylation networks originated from common starting points that lay in the unexpectedly distant past. The trail remains fragmentary, however, is severely lacking in sequence information concerning the protein phosphatases of any prokaryote, and is nearly devoid of information concerning the protein phosphorylation networks residing within the members of the third phylogenetic domain, *Archaea.*

In this paper, we present the primary sequence of an archaeal protein-serine/threonine phosphatase, PP1-arch, from the extreme acidothermophile *Sulfolobus solfataricus*. This enzyme, the representative of a broadly distributed archaeal family of divalent metal ion-stimulated protein-serine/threonine phosphatases (17, 29, 30), bears remarkable resemblance to the predominant group of protein-serine/threonine phosphatases in the *Eucarya*, the PP1/2A/2B superfamily (3, 7). This observation suggests that the archaeal and eucaryal members of this enzyme superfamily not only arose from a common ancestor but probably functioned as dedicated protein phosphatases prior to the divergence of these phylogenetic domains one from another.

MATERIALS AND METHODS

Materials. All radiochemicals were purchased from NEN/DuPont (Boston, Mass.). All custom oligonucleotides were synthesized by DNAgency (Aston, Pa.). The DNA Sequenase kit version 2.0 was from U.S. Biochemical (Cleveland, Ohio). The *Taq* DNA polymerase PCR kit was from Perkin-Elmer Cetus (Norwalk, Conn.). Vent DNA polymerase was from New England Biolabs (Beverly, Mass.). Restriction enzymes, T4 DNA ligase, calf intestinal alkaline phosphatase, T4 polynucleotide kinase, plasmids, 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside, isopropyl-β-D-thiogalactopyranoside (IPTG), Magic PCR minipreps, Magic plasmid minipreps, and the Magic DNA cleanup system were from Promega (Madison, Wis.). Fuji (Japan) RX X-ray film was used to detect ^{32}P-labeled materials, while Biomax MR film (Eastman Kodak Co., Rochester, N.Y.) was used to detect ^{35}S-labeled materials. Hydroxylapatite HT was from Bio-Rad (Richmond, Calif.). Centriprep 10 centrifugal ultrafiltration devices were from Amicon (Beverly, Mass.). Unless otherwise noted, general reagents were purchased from Fisher Scientific (Pittsburgh, Pa.) or Sigma (St. Louis, Mo.). All other reagents were from previously listed sources (17).

Growth of *S. solfataricus*. *S. solfataricus* (ATCC 35091) was grown in continuous culture with vigorous aeration at a temperature of 75°C on de Rosa's standard medium (11) with the concentration of yeast extract increased to 2

* Corresponding author. Phone: (540) 231-4317. Fax: (540) 231-9070. Electronic mail address: pjkennel@vt.edu.

g/liter. Cells were harvested by a two-stage procedure involving concentration in a Pellicon Cassette System membrane concentrator (Amicon) followed by centrifugation of the concentrate for 30 min at 4,800 × *g*. The cell pellet was stored at −20°C until needed.

Buffers for purification of PP1-arch. Buffer A consisted of 20 mM Na^+ morpholineethanesulfonic acid (MES; pH 6.5) containing 100 mM NaCl, 1 mM EDTA, 1 mM EGTA [ethylene glycol-bis (β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid], 1 mM dithiothreitol (DTT), 5 μg of leupeptin per ml, 5 μg of soybean trypsin inhibitor per ml, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.5 mM *N*α-*p*-tosyl-L-lysine chloromethyl ketone (TLCK), and 0.5 mM *N*-tosyl-L-phenylalanine chloromethyl ketone (TPCK). Buffer B consisted of 10 mM Na^+ MES (pH 6.5) containing 0.5 mM EDTA, 0.5 μg of leupeptin per ml, and 0.2 mM PMSF. Buffer C consisted of 1 mM sodium phosphate (pH 6.5) containing 0.5 mM EDTA, 0.5 μg of leupeptin per ml, and 0.2 mM PMSF. Buffer D consisted of 20 mM Na^+ MES (pH 6.5) containing 10 mM NaCl, 0.5 mM EDTA, 0.5 μg of leupeptin per ml, and 0.2 mM PMSF.

Standard procedures. Protein phosphatase activity was assayed, using [^{32}P] phosphocasein as a substrate, as described previously (17). Protein concentration was measured by the method of Bradford (5), using premixed reagent and a standardized solution of bovine serum albumin from Pierce (Rockford, Ill.) or, where indicated, by A_{280}. Sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) was performed, unless otherwise stated, as described by Laemmli (21). Gels were stained as described by Fairbanks et al. (12).

Detection of protein phosphatase activity following SDS-PAGE. Samples of the Mono Q fraction, each 20 μl containing 4.2 μg of protein, were resolved in several parallel lanes of an SDS–15% polyacrylamide gel as described by Laemmli (21) except that the temperature used for sample preparation was reduced to 65°C. The resolving gel was 0.1 cm thick and approximately 7.0 cm in length. After electrophoresis, the gel was soaked for 30 min in 50 mM Na^+ MES (pH 6.5) containing 0.5 mM EDTA and then for 30 min in 50 mM Na^+ MES (pH 6.5). Two lanes of the gel were then excised, and each was divided into 34 slices, 0.2 cm long by 1.0 cm across. These were placed in separate microcentrifuge tubes, homogenized, and soaked overnight in 30 μl of 100 mM Na^+ MES (pH 6.5) containing 0.66 mg of bovine serum albumin per ml and either 40 mM $MnCl_2$ or 10 mM EDTA. The mixtures were then assayed for protein phosphatase activity by adding 10 μl of [^{32}P]phosphocasein and following standard procedures (17). Another portion of the gel containing molecular weight markers and a sample of the Mono Q fraction was stained for protein with Coomassie blue.

Amino acid sequence analysis. For N-terminal sequence analysis, the Mono Q fraction (10 to 15 μg) was resolved on an SDS–15% polyacrylamide gel as described above. The protein was electroblotted onto a polyvinylidene difluoride (PVDF) membrane as described by LeGendre et al. (22), using a Bio-Rad semidry electroblotting apparatus operating at an applied potential of 12 V for a period of 40 min, and then visualized by staining with Coomassie blue. The section of the membrane containing PP1-arch was excised with a razor blade and subjected to automated gas-phase Edman sequencing. For the sequences of internal peptides, the Mono Q fraction (10 μl containing 10 μg of protein) was resolved on an SDS-polyacrylamide gel as described previously. The gel was stained with Coomassie blue, and the section of the gel containing PP1-arch was excised. This gel slice was loaded atop a second, SDS–20% polyacrylamide gel. The separating gel used the high-molarity buffer system of Okajima et al. (26), which is optimized for the resolution of small proteins and peptides. The remaining space in the sample well was occupied by 100 μl of sample buffer consisting of 50 mM Tris-HCl (pH 6.8), 20% (vol/vol) glycerol, 0.1% (wt/vol) SDS, 0.2% (wt/vol) bromophenol blue, and 40 ng of *Staphylococcus aureus* V8 protease per μl (16). The gel was electrophoresed until the dye front reached the interface between the stacking and resolving gels, and then electrophoresis was halted for 2 h to permit proteolysis to proceed (16). The resulting peptides were then separated by electrophoresis, electroblotted onto a PVDF membrane, and visualized by staining with Coomassie blue as described above. Sections of the membrane containing the two most visually prominent peptides were then excised and subjected to 20 cycles each of automated gas-phase Edman sequencing at the Protein Sequencing Laboratory at the Virginia Polytechnic Institute and State University.

PCRs. Touchdown PCR was conducted as described by Roux (35) in a volume of 100 μl containing 1 μg of *S. solfataricus* DNA and 250 pmol of each oligonucleotide primer. Initial denaturation at 94°C for 3 min was followed by 55 three-stage cycles consisting of denaturation at 94°C for 1 min, incubation at the annealing temperature for 1 min, and extension for 2 min at 72°C. For the first 45 cycles, the annealing temperature was decreased from an initial value of 55°C to a final value of 41°C in steps of 1°C, with each step lasting three cycles. This was followed by 10 final cycles at an annealing temperature of 40°C.

Standard PCRs were done with 1 μg of DNA and 100 pmol of each primer. Initial denaturation was conducted at 94°C for 2 min. This was followed by 30 three-stage cycles consisting of denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 2 min.

Other cloning procedures. Genomic DNA was isolated from *S. solfataricus* as described by Ausubel et al. (2). Preparation of DNA libraries, identification and isolation of clones, etc., were performed by using standard procedures (37). DNA sequencing was performed by the dideoxy method of Sanger et al. (38), using a commercial kit (Sequenase; U.S. Biochemical).

Expression of the gene for PP1-arch in *E. coli*. The gene for PP1-arch was amplified by PCR from 1 μg of genomic DNA from *S. solfataricus*, using primers that introduced restriction sites for *Nde*I or *Sal*I at each end: 5′-CCACAACA TATGAACATTGAAGAAACG-3′ (nucleotides −10 to +18) and 3′-TTTTTG TCGACTATACTATCTCTTCTAT-5′ (nucleotides +892 to +865). The gene was then cloned into expression vector pT7-7, which had been digested with *Nde*I and *Sal*I, to form plasmid pJL-3. Competent *E. coli* DH5αF′ was transformed with pJL-3 and grown, and the plasmid was isolated in quantity therefrom. Next, *E. coli* BL21(DE3) was transformed with pJL-3 and grown on M9 medium to an A_{600} of 0.6 to 1.0, and a portion of the cells was incubated with 1 mM IPTG for 2 h. Control cells either contained no plasmid or were transformed with vector alone. Cells were harvested by centrifugation, and a portion of the cell pellet was lysed and analyzed for total protein on an SDS–12.5% polyacrylamide gel. A second portion of each pellet was suspended in 50 mM Na^+ MES (pH 6.5) containing 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 1 mM EGTA, 35 μg of PMSF per ml, and 5 μg of leupeptin per ml. The cells were lysed by sonic disruption and centrifuged for 12,000 × *g* for 10 min, and the supernatant was assayed for protein phosphatase activity in the presence and absence of divalent metal ions. For the former, 5 mM EDTA was substituted for the 20 mM Mn^{2+} normally present.

Nucleotide sequence accession number. The nucleotide sequence for PP1-arch reported in this paper has been submitted to the GenBank data bank, which has assigned it accession number U35278.

RESULTS

Purification of PP1-arch. All operations were performed at 4°C unless otherwise noted. The frozen cell pellet, containing 200 g (wet weight) of *S. solfataricus*, was thawed at room temperature and suspended in 5 volumes of buffer A (see Materials and Methods for details of buffer composition). The cells were then broken by sonic disruption for periods of 1-min duration, using the maximum power setting, 130 W, of a Heat Systems-Ultrasonics, Inc. (Plainview, N.Y.), model 185 sonicator fitted with a large probe. Cell rupture was monitored spectrophotometrically by measuring the release of soluble cell pigments into the supernatant fraction of small aliquots of the sonication mixture that were centrifuged in a microcentrifuge to pellet cells and debris. When two consecutive periods of sonication yielded closely comparable A_{400} values, sonication was discontinued and the entire mixture was clarified by centrifugation at 12,000 × *g* for 30 min to give the soluble extract. Complete rupture generally required six to eight periods of sonic disruption and yielded a final A_{400} of 1.0 to 1.5.

The soluble extract was passed through a precolumn (10 by 4 cm) of carboxymethyl-Trisacryl onto a column (6.25 by 30 cm) of DE-52–cellulose. Both the precolumn and DE-52 column were previously equilibrated in buffer B. The precolumn was discarded, and the DE-52 column washed with buffer B until no protein was detectable in the eluate. Next, the DE-52 column was washed with buffer B containing 150 mM NaCl until no protein could be detected in the eluate. Both washes were discarded, and PP1-arch was eluted from the column by using buffer B containing 400 mM NaCl. That portion of the high-salt batch eluate that contained detectable levels of protein, generally about 1 liter in volume, was pooled as DE-52 fraction I.

DE-52 fraction I was dialyzed against buffer B and applied to a column (2.5 by 40 cm) of DE-52–cellulose that had been equilibrated in the same buffer. The column was washed with buffer B and then with buffer B containing 150 mM NaCl and developed with an 800-ml linear salt gradient of 150 to 400 mM NaCl in buffer B. Fractions of 10 ml were collected and assayed for the presence of protein and protein phosphatase activity. A single peak of protein phosphatase activity that eluted near the midpoint of the gradient was detected. Active fractions were pooled to give DE-52 fraction II.

DE-52 fraction II was dialyzed against buffer C and then loaded onto a column (2.5 by 12 cm) of hydroxylapatite HT equilibrated in the same buffer. The column was washed with

TABLE 1. Summary of purification of PP1-arch from *S. solfataricus*

Fraction	Protein (mg)	Activity (nmol/min)	Sp act (nmol/min/mg)	Recovery (%)	Enrichment (fold)
Soluble extract	1,660	380	0.2	100	1
DE-52 fraction I, batch step	290	580	2.0	152	10
DE-52 fraction II, gradient elution	60	304	5.1	80	26
Hydroxylapatite fraction	7.5	—[a]	—	—	—
G-100 fraction	1.9	65	34.2	17	171
Mono Q fraction	0.5	62	206	17	1,033

[a] —, measurement not possible because of inhibition by P_i in buffer.

buffer C and then developed with an 800-ml linear gradient of 1 to 400 mM sodium phosphate in buffer C. Fractions of 10 ml were assayed for the presence of protein and protein phosphatase activity. Although P_i inhibits PP1-arch, sufficient residual activity could be detected in these enzyme assays to permit the identification of active fractions. These were pooled as the hydroxylapatite fraction and concentrated to a volume of approximately 2 ml via centrifugal ultrafiltration at 3,000 × *g*, using a Centriprep 10 concentrator. This material was then loaded onto a column (5 by 100 cm) of Sephadex G-100 that had been equilibrated in buffer D. The column was eluted with this same buffer. Fractions of 2 ml were collected and assayed for protein and protein phosphatase activity. Active fractions were pooled as the G-100 fraction.

The G-100 fraction was applied, using a fast protein liquid chromatography system, to a column (0.5 by 7 cm) of Mono Q that had been equilibrated in buffer B. The column was washed with 5 ml of this same buffer and then developed with a two-stage gradient. The first stage was a 2-ml linear gradient of 0 to 170 mM NaCl in buffer B. The second stage consisted of a 35-ml linear gradient of 170 to 270 mM NaCl in buffer B. Fractions of 1 ml were collected and assayed for protein and activity. PP1-arch eluted during the second stage of the gradient. Active fractions were pooled to give the Mono Q fraction, which was purified roughly 1,000-fold from the soluble extract (Table 1).

When samples of the Mono Q fraction were prepared for SDS-PAGE by using moderate heating, divalent metal ion-dependent protein phosphatase activity could be recovered from gel slices, permitting unambiguous identification of the polypeptide that was its source (Fig. 1). This analysis indicated that the predominant polypeptide on the gels, representing from 40 up to 70% of the total protein in each preparation, as determined by densitometry, was PP1-arch. The close correlation between the relative molecular mass of the active peptide on SDS-polyacrylamide gels, 32 to 34 kDa, with the molecular mass of the native protein previously estimated from gel filtration chromatography, 28 to 30 kDa (17), indicated that PP1-arch was a monomer and that the other polypeptides observed represented simple contaminants, not species that specifically associate with PP1-arch.

Partial amino acid sequence of PP1-arch. PP1-arch was electroblotted onto PVDF membranes and subjected to automated gas-phase Edman sequence analysis. The results of three such analyses indicated that the N-terminal sequence of the protein was Met-Asn-Ile-Glu-Glu-Thr-Tyr-Glu-Leu-Leu-Glu-Lys. Another sample of electrophoretically purified PP1-arch was subjected to partial digestion with *Staphylococcus aureus* V8 protease. The resulting peptides were separated by SDS-PAGE using a high-molarity buffer system, electroblotted onto PVDF membranes, and visualized with Coomassie blue. Portions of the membranes containing the two most visually prominent bands were excised and subjected to 20 cycles each of automated gas-phase Edman sequence analysis. Two distinct sequences were obtained, Thr-Tyr-Glu-Leu-Leu-Glu-Lys-Ser-Phe-Asp-Ile-Phe-Gly-Gln-Gln-Gly-Pro-Phe-Ile-Gly and Gly-Asp-Arg-Leu-Gly-Ile-Thr-Thr-Val-Glu-Asp-Ile-Ala-Lys-Leu-Lys-Tyr-Pro-Asp-Ile, the first of which overlapped the last seven amino acids of the N-terminal sequence obtained earlier. In net, these analyses yielded the sequence of the first 25 amino acids of the PP1-arch, along with a 20-amino-acid sequence derived from somewhere in its interior (Fig. 2A).

Cloning of PP1-arch. Given knowledge of the relative orientation of the two segments of amino acid sequence obtained from PP1-arch, PCR was selected as a vehicle for obtaining a partial clone of the enzyme. Three PCR primers were designed: one modeled after a portion of the internal sequence and a pair of nested primers modeled after portions of the N-terminal sequence (Fig. 2A). PCR amplification of *S. solfataricus* genomic DNA by the touchdown method (35), using the internal primer in conjunction with the outermost of the N-

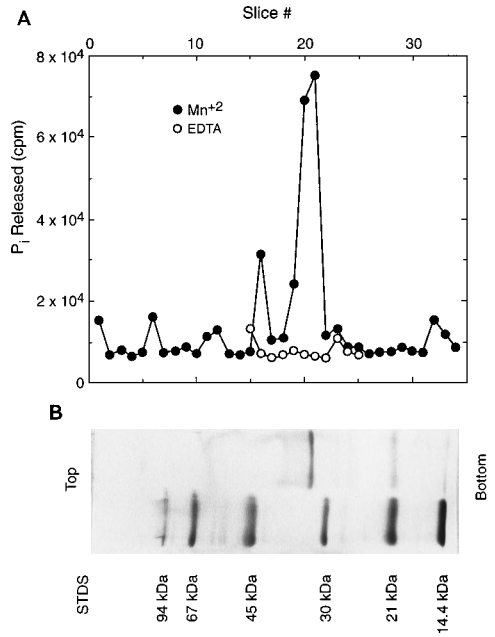

FIG. 1. Identification of PP1-arch on an SDS-polyacrylamide gel. Samples of the Mono Q fraction were resolved in several parallel lanes on an SDS–15% polyacrylamide gel. One portion of the gel was stained for protein, while another was sliced and assayed for protein phosphatase activity as described in Materials and Methods. (A) Graph of protein phosphatase activity in the presence of EDTA or Mn^{2+} in each gel slice; (B) the portion of the gel that was stained with Coomassie blue aligned with the corresponding gel slices. Note that the apparent peak of phosphatase activity in gel slice 16 appears to be artifactual, as it did not appear in two other experiments of this type involving separate preparations of PP1-arch. STDS, size standards.

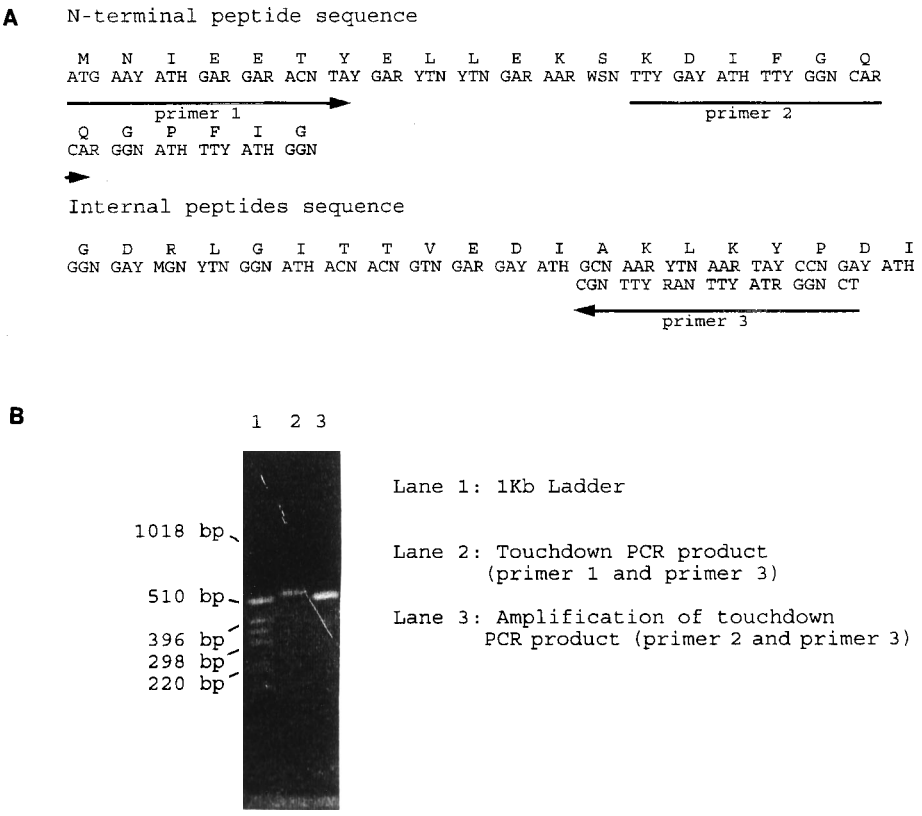

FIG. 2. Isolation of a partial clone of PP1-arch by PCR. (A) N-terminal and internal sequences of PP1-arch as determined by amino acid sequencing, along with the sequences of the oligonucleotide primers modeled thereafter. (B) An ethidium bromide-stained agarose gel showing the results of the PCRs. Lane 1 contains oligonucleotide standards (STDS); lane 2 contains the product of a touchdown PCR using primers 1 and 3 and genomic DNA from *S. solfataricus*; lane 3 contains the product of a standard PCR using primers 2 and 3 and the product of the first touchdown PCR.

terminal primers, generated a single major product more than 500 bp in length (Fig. 2B, lane 2). This PCR product was used as template for a second round of PCR amplification under standard conditions, using the internally derived primer along with the second, nested, N-terminal primer. A second PCR product was generated which, as predicted from the amino acid sequence after which the primers were modeled, was slightly shorter than that initially generated by using the outermost of the nested primers (Fig. 2B, lane 3). Sequence analysis of the larger PCR fragment revealed that it did encode for the portions of PP1-arch that had been analyzed by amino acid sequencing. Comparison of the amino acid sequences obtained from the protein itself with the DNA-derived amino acid sequence showed that they differed in 2 residues out of 45. Direct amino acid sequence analysis had indicated the presence of a phenylalanine at position 14 and a glycine at position 18, while the nucleotide sequence predicted a lysine at position 14 and an arginine at position 18. Since amino acid sequence analysis was performed only once, while analysis of the corresponding portion of the DNA sequence was performed several times on both strands of the PCR product and, later, genomic DNA clones (see below) with consistent results, we attributed these discrepancies to errors in the interpretation of the amino acid sequence data and concluded that the DNA-derived amino acid sequence information at these positions should be regarded as correct.

The touchdown PCR product was then used as a probe to isolate a larger clone from an *Eco*RI library of *S. solfataricus* genomic DNA that had been ligated into plasmid vector pGEM-3Z. Sequence analysis of the clones thus identified indicated that the gene for PP1-arch contained an internal *Eco*RI site that caused the gene sequence to terminate after the codon for amino acid 223 of the protein. Subsequently, the complete structural gene for PP1-arch was obtained from a second library prepared by using *Hin*dIII. The complete nucleotide sequence for the structural gene for PP1-arch is presented in Fig. 3. The DNA-derived amino acid sequence of PP1-arch predicts a protein product of 293 amino acids with a calculated molecular mass of 33,765 Da. This figure is in good agreement with those determined by gel filtration chromatography (17) and SDS-PAGE (Fig. 1). Transformation of *E. coli* with this gene and subsequent induction of its expression with IPTG led to the appearance of an appropriately sized polypeptide on SDS-polyacrylamide gels along with divalent metal ion-stimulated phosphocasein phosphatase activity, strong evidence that it does encode PP1-arch (Fig. 4).

DISCUSSION

We have isolated, cloned, and analyzed in molecular detail PP1-arch, the divalent metal ion-stimulated protein-serine/threonine phosphatase from the extreme acidophilic archaeon *S. solfataricus*. Comparison with the sequences of other protein phosphatases clearly indicates that PP1-arch is closely related to the PP1/2A/2B superfamily of protein-serine/threonine phosphatases from the *Eucarya* (Fig. 5). It is more distantly, but recognizably, related to the protein phosphatases found in bacteriophages lambda and ϕ80. Included in its sequence are

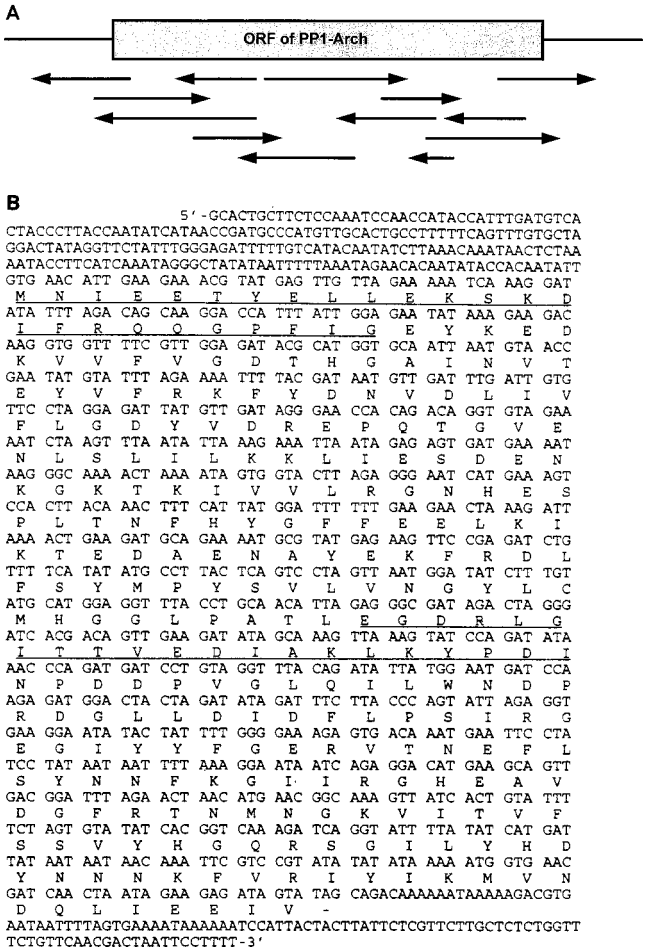

FIG. 3. Complete oligonucleotide and inferred amino acid sequence of PP1-arch. (A) Schematic describing the strategy used to sequence the gene for PP1-arch on both strands; (B) oligonucleotide and predicted amino acid sequences of PP1-arch. The underlined areas indicate areas of the protein that were previously analyzed by amino acid sequencing.

the three major signature motifs, Gly_{36}-Asp-Xaa-His-Gly_{40}, Gly_{63}-Asp-Xaa-Val-Asp-Arg_{68}, and Arg_{100}-Gly-Asn-His-Glu_{104}, thought to be characteristic of this superfamily of protein-serine/threonine phosphatases (3, 46) as well as other phosphohydrolases such as diadenosine tetraphosphatase from *E. coli* (19).

The region of visible resemblance between PP1-arch and known members of the PP1/2A/2B superfamily extends over a considerable portion of its structure, residues 4 to 275 of its 293 component amino acids (Fig. 5). Within this area, it shares 31% identity with a eucaryal PP1 from the rabbit (85 identical residues), 29% identity with a eucaryal PP2A from *Saccharomyces cerevisiae* (78 identical residues), and 30% identity with a eucaryal PP2B from the rat (81 identical residues). Given the extraordinarily high degree of amino acid sequence conservation within the *Eucarya* (3), these figures can be regarded as representative for its resemblance to virtually any eucaryal member of this superfamily. The level of sequence identity with a bacteriophage protein phosphatase from phage lambda, whose ultimate origins remain cryptic at this time, is considerably less—about 17% (39 identical residues over a 236-residue span). PP1-arch possesses 33 of the 42, or 4 in 5, of the amino acid residues conserved among all known members of the eucaryal PP1/2A/2B superfamily. When conservative substitutions such as threonine for serine and lysine for arginine are included, the degree of conservation rises to 37 of 42 residues, almost 9 in 10. Of these 42 residues, 15 are also conserved between the eucaryal and bacteriophage enzymes. Fourteen of these fifteen are also found in PP1-arch.

Barton et al. (3) have established sequence criteria for classifying the individual types within the PP1/2A/2B superfamily. On the basis of these criteria, PP1-arch would be categorized as a type 1 protein-serine/threonine phosphatase. The proline at position 151 and the lysine at position 165 are indicative of members of both the PP1 and PP2A subfamilies, while features indicative of a PP1 are the serine at position 142 and the aspartate at position 213. However, classical biochemical data corroborating this sequence-based assignment, such as inhibition by okadaic acid, substrate preferences, or sensitivity to heat-stable inhibitors (7), have proved negative, in the case of the first named (17), or inconclusive.

The great degree of amino acid sequence conservation between PP1-arch and eucaryal protein phosphatases, especially compared with the bacteriophage enzymes, is consistent with our current understanding of the evolutionary relationships among these phylogenetic domains, in which the *Archaea* and *Eucarya* reside on a common branch of the rooted phylogenetic tree with the *Bacteria* situated alone on the other (27). Even so, the degree of identity observed between PP1-arch and its eucaryal counterparts appears astonishingly high, nearly

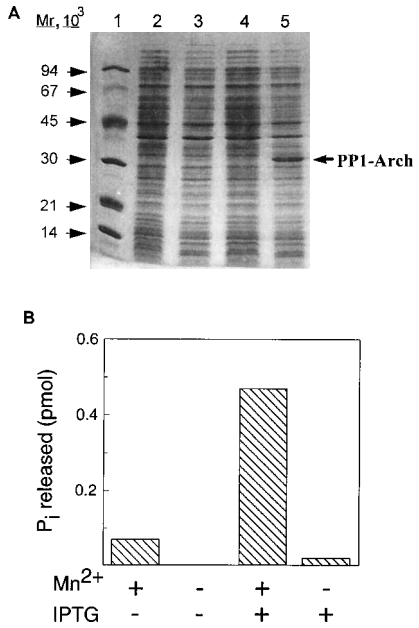

FIG. 4. Expression of the gene for PP1-arch in *E. coli*. The gene for PP1-arch was cloned into expression vector pT7-7 to form plasmid pJL-3 as described in Materials and Methods. *E. coli* BL21(DE3) was transformed with plasmid pJL-3. Control cells either contained no plasmid or were transformed with vector alone. Cultures were grown until the optical density at 600 nm was in the range of 0.6 to 1.0, the cultures were divided in half, and one portion of the cells was incubated with IPTG for 2 h. Cells were then harvested, and one portion was subjected to SDS-PAGE while another was lysed and assayed for phosphocasein protein phosphatase activity. (A) SDS-polyacrylamide gel of total *E. coli* proteins. Lanes: 1, molecular weight standards; 2 and 3, cells transformed with vector alone incubated in the absence and presence of IPTG, respectively; 4 and 5, cells transformed with plasmid pJL-3 and incubated in the absence and presence of IPTG, respectively. The position of the putative PP1-arch protein is indicated at the right. (B) Assay for protein phosphatase activity in the presence or absence of Mn^{2+} in extracts of cells transformed with pJL-3 and incubated in the presence or absence of IPTG.

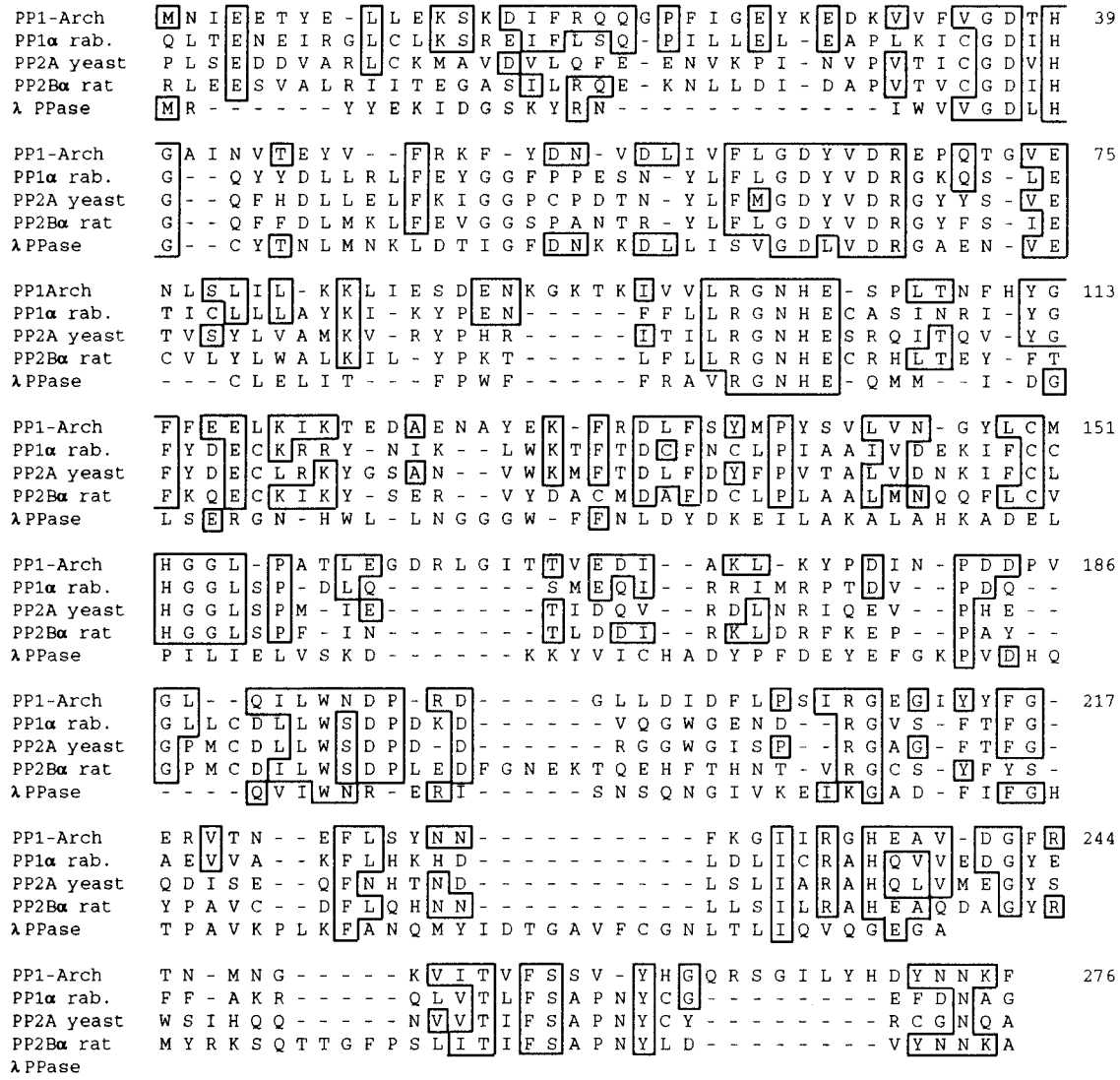

FIG. 5. Comparison of PP1-arch with members of the PP1/2A/2B superfamily. Shown is the derived amino acid sequence of residues 1 to 276 of PP1-arch aligned with the corresponding regions of PP1α from the rabbit (PP1α rab.) (4), PP2A from *Saccharomyces cerevisiae* PPH 22 (PP2A yeast) (41), PP2Bα1 from the rat (PP2Bα rat) (20), and the open reading frame 221 gene product from bacteriophage lambda (λ PPase) (8). Identities with PP1-arch are enclosed with boxes. Dashes indicate where gaps were introduced in order to align sequences. Sequence alignment was done by eye, referring heavily to the prior analyses by Barton et al. (3) and Zhou et al. (46) of the homologies within the other members of the PP1/2A/2B superfamily. Essentially, areas of high sequence identity were aligned first, followed by the regions linking them, maximizing match-ups with amino acid residues found to be highly conserved in the eucaryal enzymes or between the eucaryal and bacteriophage enzymes.

one in every three amino acids. This level of sequence identity approaches that observed among the eucaryal members of the PP1/2A/2B superfamily, which bottoms out at about 35% (3). The homology is also quite general, extending throughout all but the most extreme N- and C-terminal portions of the protein. In our opinion, this indicates that the common progenitor of these enzymes had acquired and refined virtually all of the basic structural attributes necessary to function as a specialized, dedicated protein phosphatase well before the *Archaea* and *Eucarya* diverged from one another. This, in turn, implies that the covalent modification of proteins by phosphorylation-dephosphorylation was operative in this common ancestor and had been for a substantial period. If so, it would be expected that this archaeal protein phosphatase would possess as a functional partner a eucaryal protein kinase(s). As this article was in its last stages of preparation, reports suggesting the existence of two types of protein kinases in the *Archaea* appeared. In one, searches of DNA sequence databases revealed the existence of open reading frames whose predicted gene products bore significant homology to the predominant family of eucaryal protein kinases, those whose catalytic domains resemble the C subunit of the cyclic AMP-dependent protein kinase, in three different strains of methanogenic archaeons, *Methanococcus voltae*, *M. vannielii*, and *M. thermolithotrophicus* (40). The second reported that a homolog of the other major family of protein kinases, the so-called histidine kinases, was required for the chemo- and phototaxis by the halophilic archaeon *Halobacterium salinarium* (36).

PP1-arch from *S. solfataricus* differs from the eucaryal members of the PP1/2A/2B superfamily in several potentially significant ways. The eucaryal members of the PP1/2A/2B superfamily can be readily inactivated by sulfhydryl modifying

reagents (45), while neither PP1-arch nor its presumed homologs from halophilic or methanogenic archaeons show sensitivity to any of a wide array of such compounds (17, 29, 30). PP1-arch from *S. solfataricus* also is insensitive to the classic PP1/2A/2B inhibitors okadaic acid, calyculin A, and microcystin LR (17). These structurally diverse compounds inhibit virtually every eucaryal member of this enzyme superfamily (7) but not the bacteriophage lambda enzyme (8). Intriguingly, in earlier studies on the functionally similar protein phosphatases from two other archaeons, the activity from the methanogen *Methanosarcina thermophila* TM-1 proved moderately sensitive to inhibition by these compounds (30), while that from the halophile *Haloferax volcanii* was not (29). At this point, it is difficult to rationalize why either sensitivity or insensitivity to these compounds is not a general phenomenon among archaeal protein phosphatases.

Another significant difference between PP1-arch and its eucaryal homologs resides in its subunit composition. While the eucaryal members of the PP1/2A/2B superfamily are thought to exist as oligomers in vivo, many of whose additional polypeptide components serve as modulators or targeting moieties (15, 39), the only form of active PP1-arch that we have detected in cell homogenates is a soluble, monomeric one—the equivalent of a naked catalytic subunit. In this it mirrors the structure of its more distantly related bacteriophage-encoded counterparts. Although the activity of PP1-arch is so greatly stimulated by divalent metal ions in vitro (17) as to render it essentially divalent metal ion dependent, the identities of the stimulatory metal ions—Mn^{2+}, Ni^{2+}, and Co^{2+}—render it highly unlikely that they serve as physiologically operative regulators. Our findings therefore raise questions as to whether the activity of PP1-arch is regulated in vivo at all and, if so, how. Does PP1-arch act as a constitutive source of broad-specificity dephosphorylating activity to balance the action of regulated protein kinases? Is it regulated at the level of gene expression? Does an archaeal equivalent of an inhibitor protein such as heat-stable inhibitor 1 exist in the *Archaea*? Future studies on archaeal protein phosphorylation networks hold great promise as tools for tracing the origins and evolution of the protein phosphorylation networks that govern a wide spectrum of cellular processes.

ACKNOWLEDGMENTS

This work was supported by grant 3502 from the Council for Tobacco Research, U.S.A.

We thank Bassam T. Wakim of Loyola University Medical Center, Chicago, Ill., for his generous assistance in the N-terminal sequence analysis of PP1-arch. We also thank Patrick Bender, Timothy Larson, Stephen Boyle, and the members of their laboratory groups for many helpful suggestions.

REFERENCES

1. **Atkinson, M., C. Allen, and L. Sequeira.** 1992. Tyrosine phosphorylation of a membrane protein from *Pseudomonas solanacearum*. J. Bacteriol. **174:** 4356–4360.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, New York.
3. **Barton, G. J., P. T. W. Cohen, and D. Barford.** 1994. Conservation analysis and structure prediction of the protein-serine/threonine phosphatases: sequence similarity with diadenosine tetraphosphatase from *Escherichia coli* suggests homology to the protein phosphatases. Eur. J. Biochem. **220:**225–237.
4. **Berndt, N., D. G. Campbell, F. B. Caudwell, P. Cohen, E. F. da Cruz e Silva, O. B. da Cruz e Silva, and P. T. W. Cohen.** 1987. Isolation and sequence analysis of a cDNA clone encoding a type-1 protein phosphatase catalytic subunit: homology with protein phosphatase 2A. FEBS Lett. **223:**340–346.
5. **Bradford, M. M.** 1976. A rapid and simple method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72:**248–254.
6. **Chang, C., S. F. Kwok, A. B. Bleecker, and E. M. Meyerowitz.** 1993. *Arabidopsis* ethylene-response gene *ETR-1*: similarity of product to two-component regulators. Science **262:**539–544.
7. **Cohen, P.** 1991. Classification of protein-serine/threonine phosphatases: identification and quantitation in cell extracts. Methods Enzymol. **201:**389–398.
8. **Cohen, P. T. W., and P. Cohen.** 1989. Discovery of a protein phosphatase activity encoded in the genome of bacteriophage λ. Probable identity with open reading frame 221. Biochem. J. **260:**931–934.
9. **Cozzone, A. J.** 1988. Protein phosphorylation in prokaryotes. Annu. Rev. Microbiol. **42:**97–125.
10. **Dadssi, M., and A. J. Cozzone.** 1990. Evidence for protein-tyrosine kinase activity in the bacterium *Acinobacter calcoaceticus*. J. Biol. Chem. **265:**20996–20999.
11. **de Rosa, M., A. Gambacorta, and J. D. Bullock.** 1975. Extremely thermophilic acidophilic bacteria convergent with *Sulfolobus acidocaldarius*. J. Gen. Microbiol. **86:**156–164.
12. **Fairbanks, G., T. L. Steck, and D. F. H. Wallach.** 1971. Electrophoretic analysis of the major polypeptides of the human erythrocyte membrane. Biochemistry **10:**2606–2617.
13. **Foster, R., J. Thorner, and G. S. Martin.** 1989. Nucleotidylation, not phosphorylation, is the major source of the phosphotyrosine detected in enteric bacteria. J. Bacteriol. **171:**272–279.
14. **Hanks, S. K., and A. M. Quinn.** 1991. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol. **200:**38–62.
15. **Hubbard, M. J., and P. Cohen.** 1993. On target with a new mechanism for the regulation of protein phosphorylation. Trends Biochem. Sci. **18:**172–177.
16. **Kennedy, T. E., M. A. Gawinowicz, A. Barzilai, E. R. Kandel, and J. D. Sweatt.** 1988. Sequencing of protein from two-dimensional gels by using *in situ* digestion and transfer of peptides to polyvinylidene difluoride membranes: application to proteins associated with sensitization in *Aplysia*. Proc. Natl. Acad. Sci. USA **85:**7008–7012.
17. **Kennelly, P. J., K. A. Oxenrider, J. Leng, J. S. Cantwell, and N. Zhao.** 1993. Identification of a serine/threonine-specific protein phosphatase from the archaebacterium *Sulfolobus solfataricus*. J. Biol. Chem. **268:**6505–6510.
18. **Kennelly, P. J., and M. Potts.** 1994. Prokaryotic protein phosphatases: reflections of the past. Windows to the future? Adv. Protein Phosphatases **8:** 53–68.
19. **Koonin, E. V.** 1993. Bacterial and bacteriophage protein phosphatases. Mol. Microbiol. **8:**785–786.
20. **Kuno, T., T. Takeda, M. Hirai, A. Ito, H. Mukai, and C. Tanaka.** 1989. Evidence for a second isoform of the catalytic subunit of calmodulin-dependent protein phosphatase (calcineurin A). Biochem. Biophys. Res. Commun. **165:**1352–1358.
21. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
22. **LeGendre, N., M. Mansfield, A. Weiss, and P. Matsudaira.** 1993. Purification of proteins and peptides by SDS-PAGE, p. 74–100. *In* P. Matsudaira (ed.), A practical guide to protein and peptide purification for microsequencing, 2nd ed. Academic Press, San Diego, Calif.
23. **Maeda, T., S. M. Wurgler-Murphy, and H. Saito.** 1994. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. Nature (London) **369:**242–245.
24. **Matsumoto, A., S.-K. Hong, H. Ishizuka, S. Horinouchi, and T. Beppu.** 1994. Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase. Gene **146:**47–56.
25. **Munoz-Dorado, J., S. Inouye, and M. Inouye.** 1993. Eukaryotic-like protein serine/threonine kinases in *Myxococcus xanthus*, a developmental bacterium exhibiting social behavior. J. Cell. Biochem. **51:**29–33.
26. **Okajima, T., T. Tanabe, and T. Yasuda.** 1993. Nonurea SDS-polyacrylamide gel electrophoresis with high-molarity buffers for the separation of proteins and peptides. Anal. Biochem. **211:**293–300.
27. **Olsen, G. J., and C. R. Woese.** 1994. Ribosomal RNA: a key to phylogeny. FASEB J. **7:**113–123.
28. **Ota, I. M., and A. Varshavsky.** 1993. A yeast protein similar to bacterial two-component regulators. Science **262:**566–569.
29. **Oxenrider, K. A., and P. J. Kennelly.** 1993. A protein-serine phosphatase from the halophilic archaeon *Haloferax volcanii*. Biochem. Biophys. Res. Commun. **194:**1330–1335.
30. **Oxenrider, K. A., M. E. Rasche, M. Thorsteinsson, and P. J. Kennelly.** 1993. An okadaic acid-sensitive protein phosphatase from the archaeon *Methanosarcina thermophila* TM-1. FEBS Lett. **331:**291–295.
31. **Popov, K. M., N. Y. Kedishvili, Y. Zhao, Y. Shimomura, D. W. Crabb, and R. A. Harris.** 1993. Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases. J. Biol. Chem. **268:** 26602–26606.
32. **Popov, K. M., Y. Zhao, Y. Shimomura, M. J. Kuntz, and R. A. Harris.** 1992. Branched-chain α-ketoacid dehydrogenase kinase. Molecular cloning, expression, and sequence similarity with histidine protein kinases. J. Biol. Chem. **267:**13127–13130.
33. **Potts, M., H. Sun, K. Mockaitis, P. J. Kennelly, D. Reed, and N. K. Tonks.**

1993. A protein-tyrosine/serine phosphatase encoded by the genome of the cyanobacterium *Nostoc commune* UTEX 584. J. Biol. Chem. **268:**7632–7635.
34. **Ray, M. K., G. S. Kumar, and S. Shivaji.** 1994. Tyrosine phosphorylation of a cytoplasmic protein from the antarctic psychrotrophic bacterium *Pseudomonas syringae*. FEMS Microbiol. Lett. **122:**49–54.
35. **Roux, K. H.** 1994. Using mismatched primer-template pairs in touchdown PCR. BioTechniques **16:**812–814.
36. **Rudolph, J., and D. Oesterhelt.** 1995. Chemotaxis and phototaxis require CheA histidine kinase in the archaeon *Halobacterium salinarium*. EMBO J. **14:**667–673.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
39. **Shenolikar, S., and A. C. Nairn.** 1991. Protein phosphatases: recent progress. Adv. Second Messenger Phosphoprotein Res. **23:**1–121.
40. **Smith, R. F., and K. Y. King.** 1995. Identification of a eukaryotic-like protein kinase gene in the Archaebacteria. Protein Sci. **4:**126–129.
41. **Sneddon, A. A., P. T. W. Cohen, and M. J. R. Starke.** 1990. *Saccharomyces cerevisiae* protein phosphatase 2A performs an essential cellular function and is encoded by two genes. EMBO J. **9:**4339–4346.
42. **Urabe, H., and H. Ogawara.** 1995. Cloning, sequencing, and expression of serine/threonine kinase-encoding genes from *Streptomyces coelicolor* A3(2). Gene **153:**99–104.
43. **Warner, K. M., and G. S. Bullerjahn.** 1994. Light-dependent tyrosine phosphorylation in the cyanobacterium *Prochlorothrix hollandica*. Plant Physiol. **105:**629–633.
44. **Zhang, C.-C.** 1993. A gene encoding a protein related to eukaryotic protein kinases from the filamentous heterocystous cyanobacterium *Anabaena* PCC 7120. Proc. Natl. Acad. Sci. USA **90:**11840–11844.
45. **Zhang, Z., S. Zhao, S. Deans-Zirattu, G. Bai, and E. Y. C. Lee.** 1993. Mutagenesis of the catalytic subunit of rabbit muscle protein phosphatase-1. Mol. Cell. Biochem. **127/128:**113–119.
46. **Zhou, G., J. M. Denu, L. Wu, and J. E. Dixon.** 1994. Mutational analysis of a ser/thr phosphatase: identification of residues important in substrate binding and catalysis. J. Biol. Chem. **269:**28084–28090.