

FOR THE RECORD

The protein phosphatase 2C (PP2C) superfamily: Detection of bacterial homologues

PEER BORK, NIGEL P. BROWN, HEDVIG HEGYI, AND JÖRG SCHULTZ

European Molecular Biology Laboratory, Heidelberg, and Max-Delbrück-Center for Molecular Medicine,
Berlin-Buch, Germany

(RECEIVED February 14, 1996; ACCEPTED April 5, 1996)

Abstract: A thorough sequence analysis of the various members of the eukaryotic protein serine/threonine phosphatase 2C (PP2C) family revealed the conservation of 11 motifs. These motifs could be identified in numerous other sequences, including fungal adenylate cyclases that are predicted to contain a functionally active PP2C domain, and a family of prokaryotic serine/threonine phosphatases including SpoIIE. Phylogenetic analysis of all the proteins indicates a widespread sequence family for which a considerable number of isoenzymes can be inferred.

Keywords: alignment; bacterial serine/threonine phosphatases; homology search; PP2C; SpoIIE.

Whereas the basic metabolic pathways and many other essential cellular processes are clearly identified in all three kingdoms (eubacteria, eukaryotes, and archaeobacteria), the evolution of many eukaryotic regulatory processes remains to be elucidated. Protein phosphatases mediate major signaling steps, but so far, only a few bacterial or archaeobacterial relatives have been reported. For example, a protein tyrosine phosphatase has been identified in the cyanobacterium *Nostoc commune* (Potts et al., 1993) and a serine/threonine phosphatase was found in the archaeobacterium *Sulfolobus solfataricus* (Leng et al., 1995). The *Escherichia coli* ApaH gene encoding a diadenosine tetraphosphatase, a pleiotropic regulator, was shown to be related to a family of serine/threonine phosphatases (Koonin, 1993).

Two distinct serine/threonine phosphatase sequence families exist (Klumpp et al., 1994): PP2C and PPP (Cohen, 1994). The latter comprise PP1, PP2A, and PP2B, which have distinct biochemical properties (Cohen, 1989).

The antagonistic roles of PPPs with respect to protein kinase-triggered regulation pathways are well established (for review see Cohen, 1992), but experiments aiming at the determination of the cellular functions for PP2C family members have started only very recently (Leung et al., 1994; Meyer et al., 1994; Rob-

inson et al., 1994; Shiozaki & Russell, 1995). Nevertheless, there is a detailed biochemical knowledge showing the requirement for Mg^{2+} , Mn^{2+} -binding in vitro, and the stimulation by Ca^{2+} ions (Cohen, 1989). Recently, numerous PP2C-like genes from various species have been cloned and characterized. Here we summarize all the data on the PP2C family and characterize conserved sequence features of the family that have enabled us to identify new members and to establish a link to prokaryotic phosphatases.

Iterative collection of the PP2C family members: An alignment of the PP2C members annotated in PROSITE (Bairoch & Bucher, 1994; entry PS01032) was the starting point for iterative sequence homology searches using several pattern search methods (for details of the strategy see Bork & Gibson, 1996). After each iteration, newly detected sequences were added to the alignment using CLUSTALW (Thompson et al., 1994). A total of 11 conserved motifs were identified that mostly seem to correspond to structural elements and exposed functional residues that are scattered along the sequence (Fig. 1). Mitochondrial pyruvate lipamide phosphatase was included into the initial alignment because it has been shown to belong to the family (Lawson et al., 1993) and as it has been annotated in PROSITE. Whereas the first iteration only collected members that are already known to belong to the family, but that are too divergent or too recently sequenced to be annotated in PROSITE, subsequent searches revealed the presence of all motifs in a hypothetical yeast protein (YCW9; e.g., 7.4 standard deviations above the expected using profile searches; Gribskov et al., 1987). Another yeast protein was identified as a putative phosphatase by motif searches (MoST; Tatusov et al., 1994) using the conserved motifs 2 and 11 (YHN6 in Fig. 1; in both cases, the stringent threshold $r = 0.01$ was used, whereby r is the ratio of observed versus expected occurrences; Tatusov et al., 1994).

Proposed PP2C activity in fungal adenylate cyclases: Blastp database searches (Altschul et al., 1990) with members of the eukaryotic PP2C family significantly detect (probability of chance matches $p < 10^{-15}$) the presence of a complete PP2C domain in fungal cyclases; a similarity that has already been noted (Tamura et al., 1989; Robinson et al., 1994). Remarkably, the

Reprint requests to: Peer Bork, EMBL, Meyerhofstr.1, 69117 Heidelberg, Germany; e-mail: bork@embl-heidelberg.de.

Abbreviations: LRR, leucine-rich repeat; PP2C, protein phosphatase II C; PPP, phosphoprotein phosphatase.

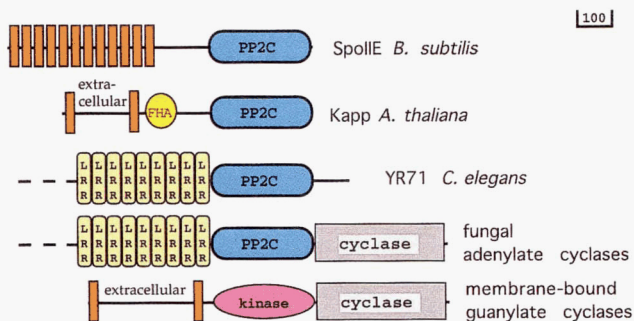


Fig. 2. PP2C as module in diverse proteins. Only PP2Cs are shown for which neighboring domains with a distinct function are known. Vertical orange bars denote transmembrane regions. Only the C-terminus of the cyclases, including some of the LRRs, are displayed. The location of PP2C in adenylate cyclases is striking compared to the functional protein kinase domain in several guanylate cyclases (the catalytic domains of adenylate and guanylate cyclases are homologous).

LRRs also precede the PP2C domain. Thus, this subfamily of PP2Cs, which is most characterized by a modification of motif 8 (e.g., the lack of an otherwise conserved aspartate-glycine dipeptide; Fig. 1), appears to be more widespread than in fungal adenylate cyclases.

Identification of prokaryotic relatives with serine/threonine phosphatase activity: Two prokaryotic proteins from *Mycoplasma genitalium* and *Synechocystis sp.* have a relationship to the group of characterized eukaryotic PP2Cs (Fig. 1) that is detectable by Blastp ($p < 10^{-9}$). Both putative gene products have been obtained by genome sequencing projects (Fraser et al., 1995; Kaneko et al., 1995). The respective gene in *M. genitalium* succeeds a gene coding for a putative protein kinase (Fraser et al., 1995). Because PP2Cs have not been and protein kinases

have very rarely been observed in prokaryotes, horizontal gene transfer would be a reasonable working hypothesis. We were able to identify, however, a more remote group of bacterial PP2Cs, including SpoIIE from *Bacillus subtilis* (running MoST on motifs 4, 8, and 11 with the threshold $r = 0.01$). The alignment (Fig. 1) shows that the prokaryotic phosphatases share all but two motifs (5a, 5b) with the eukaryotic PP2Cs.

In support of our finding, SpoIIE has been characterized recently as a serine/threonine protein phosphatase involved in asymmetric cell division and sporulation (Duncan et al., 1995). Several other *Bacillus* proteins, as well as an ORF from *Synechocystis sp.*, indicate a more widespread occurrence of this type of phosphatase in eubacteria. This group of prokaryotic serine/threonine phosphatases differs from the eukaryotic ones in some of the conserved motifs (Fig. 1). Because SpoIIE is involved in fundamental processes of bacterial cell division (Arigoni et al., 1995; Duncan et al., 1995), the roots of the PP2C family indeed seem to lie in the eubacterial kingdom, although additional prokaryotic phosphatases supporting this rooting need to be identified. Although Mn^{2+} -binding has been demonstrated for SpoIIE (Duncan et al., 1995), we are not aware of data about the Mg^{2+} -dependency of the prokaryotic members; Mg^{2+} -binding might be one reason for the presence of additional conserved motifs within the eukaryotic PP2C members compared to the prokaryotic ones (Fig. 1).

Common features of the enlarged PP2C superfamily: In spite of the high divergence of the PP2C superfamily, the signature motifs identified leave little doubt about the enlarged PP2C superfamily. The similarities are strongly supported by secondary structure predictions that have been conducted separately for each subfamily (consensus shown in Fig. 1).

The consecutive β -strands separated by α -helices suggest a $\beta+\alpha$ structural type in which antiparallel β -sheets are surrounded by α -helices. Interestingly, the secondary structure ar-

Fig. 1 (facing page). Multiple alignment of conserved regions of the PP2C superfamily showing 11 sequence motifs. It was constructed using CLUSTALW (Thompson et al., 1994) by merging six pre-aligned subfamilies based on phylogenetic analysis (Fig. 3). Only very few regions were modified, as suggested by the secondary structure predictions (Rost et al., 1994) conducted independently for each subfamily. First and last columns show subfamily number (1–6) and SWISS-PROT identifier and database accession number, respectively. Intervening columns give the alignment as 11 motifs numbered along the top, with gaps indicated by hyphens, and numbers in angle brackets to indicate residues that are not displayed. Organism codes: ARATH, *Arabidopsis thaliana*; BACME, *Bacillus megaterium*; BACSU, *Bacillus subtilis*; BOVIN, *Bos taurus*; CAEEL, *Caenorhabditis elegans*; GIALA, *Giardia lamblia*; HUMAN, *Homo sapiens*; LEICH, *Leishmania chagasi*; MYCGE, *Mycoplasma genitalium*; NEUCR, *Neurospora crassa*; PARTE, *Paramecium tetraurelia*; RAT, *Rattus norvegicus*; SACKL, *Saccharomyces kluyveri*; SCHPO, *Schizosaccharomyces pombe*; SYNY,SYNY3, *Synechocystis sp.*; USTMA, *Ustilago maydis*; YEAST, *Saccharomyces cerevisiae*. Functional assignment codes: P2CA_RAT, P2CB_RAT, P2CA/ARATH, P2C_LEICH, protein phosphatases 2C and isoforms; P2C_YEAST, P2C1_SCHPO, P2C2_SCHPO, P2C3_SCHPO, P2C/GIALA, YBF6_YEAST, YER9_YEAST, YBX5_YEAST, putative protein phosphatase 2C homologues; F42G9_6/CAEEL, P2C/PARTE, KAPP_ARATH, ABI1/ARATH, protein phosphatases; PPH1/ARATH, T23F11_1/CAEEL, putative protein phosphatase homologues; PDP_BOVIN, PDP/YEAST, pyruvate dehydrogenase phosphatase, catalytic subunit; SP2E_BACSU, SP2E/BACME, stage II sporulation protein E; ICFG_SYNY3, ICFG protein; RSBU_BACSU, Sigma factor SIBG regulation protein RSBU; RSBU_BACSU, Sigma-B negative effector; CYAA_SACKL, CYAA_YEAST, CYAA_NEUCR, CYAA_SCHPO, UAC1/USTMA, adenylate cyclases; YCW9_YEAST, Y108_MYCGE, YHN6_YEAST, YR71_CAEEL, SLRD_24/SYNY, RSC454_1/HUMAN, FEM2/CAEEL, hypothetical proteins and/or unassigned function. Residues conserved between subfamilies are shown in bold; those with properties conserved in 90% of all sequences are colored: green, hydrophobic; blue, similarly charged or polar. The consensus line immediately beneath the alignment indicates residues or amino acid properties conserved in 80% of all sequences: h, hydrophobic; a, aromatic; l, aliphatic; o, S/T; p, polar; c, charged; +, positive; -, negative; s, small; u, tiny; t, turn-like (polar or tiny). Consensus PHD secondary structure predictions: H/h, helix; E/e, strand; L/dot, loop with >80%/70% expected accuracy, respectively (Rost et al., 1994), are then shown for the entire alignment (all) and for the six numbered subfamilies. An additional helix lying between motifs 3 and 4 is indicated in angle brackets, but not shown in the alignment. Assuming two β -strands N-terminal of motif 2, and a helix-strand arrangement in motif 6, the unit $\beta\beta\beta\alpha\alpha\beta$ (in brackets in the graphical interpretation of the predictions) seems triplicated (arrows, β -strands; cylinders, α -helices; grey, not verified by the automatic prediction).

rangement $\beta\beta\alpha\alpha\beta$ seems to occur three times successively (Fig. 1), and might indicate a triplication of this unit. Assuming an antiparallel β -sheet, all conserved residues located in the predicted β -hairpins would be on one side of the sheet, probably forming the active site.

Provided that all proteins identified in this study are indeed protein phosphatases (this has been shown experimentally for divergent members such as SpoIIE; see Fig. 3), polar and charged residues conserved in all the divergent subfamilies should be crucial for catalytic activity. Only two residues, an aspartate in motif 2 and a glycine in motif 5, remain invariant. Thus, the motif conservation within the PP2C family has to be reevaluated. We predict that the most conserved region in the classical eukaryotic PP2Cs (between motif 5 and 6 in Fig. 1) that is absent in the bacteria reflects more substrate specificity or

Mg^{2+}/Ca^{2+} -binding than a catalytic requirement. The conserved polar and charged residues in motifs 1, 2, 5, and 11 seem essential for the phosphatase activity. Most of the other conserved polar residues are absent from one or another subfamily, pointing to different regulation or substrates.

The variety of substrates can be estimated by the number of paralogues that occur in a single organism and also by the arrangement of the sequenced members in a phylogenetic tree (Fig. 3). As in many branches, the grouping of species is different from what one would expect; the existence of many more PP2C-like paralogues can be assumed. The public databases searched contained about 80% of all yeast proteins. Thus, extrapolating from the eight yeast PP2Cs we have recognized (Figs. 1, 3), there might be at total of about 10 paralogues in yeast, and we cannot exclude the presence of additional remote

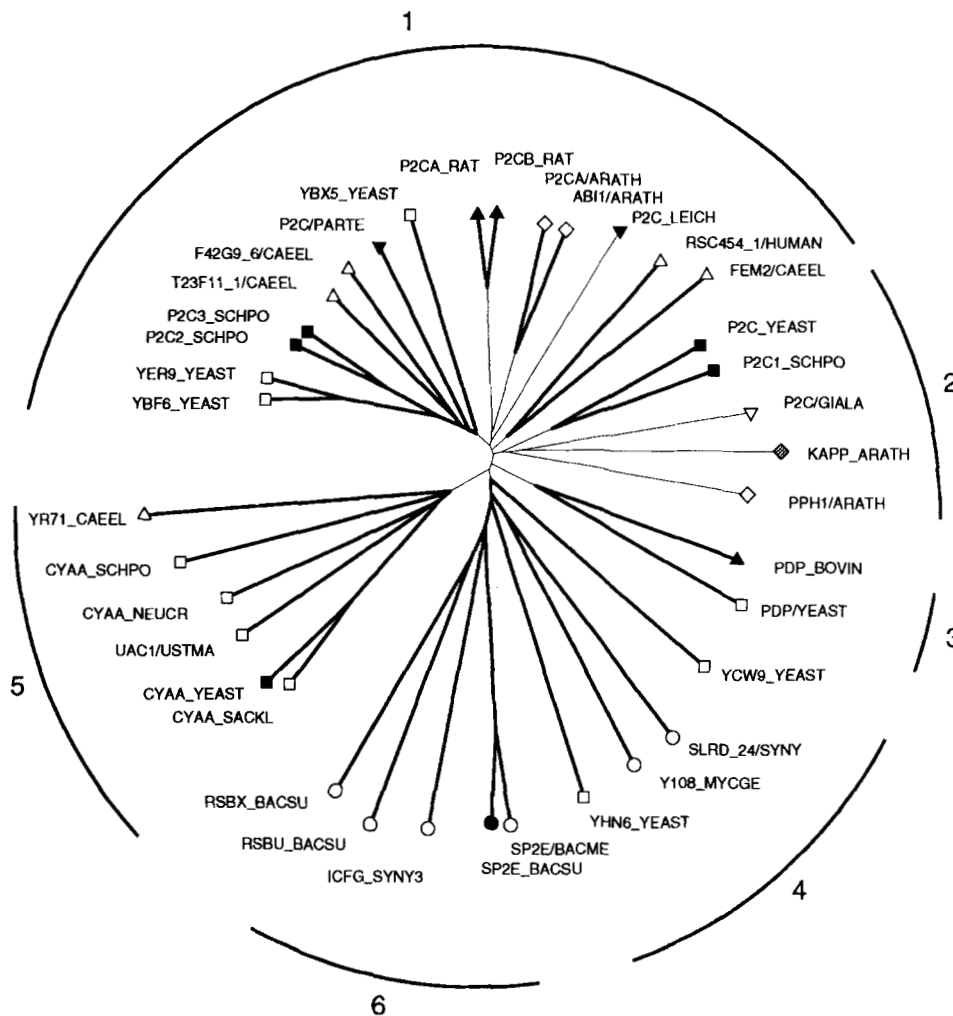


Fig. 3. Unrooted phylogenetic tree of selected PP2C members. The neighbor-joining method in CLUSTALW (Thompson et al., 1994) was used with bootstrapping to construct the tree based on the conserved sequence motifs alone. Thick lines indicate sub-trees stable in at least 500/1,000 bootstraps. Sequences are labeled with their SWISS-PROT identifiers and symbols denoting major organism groups: circle, bacteria; square, fungi; diamond, higher plants; downward triangle, protozoa; upward triangle, animals. Experimentally determined phosphatase activity is shown by a filled symbol (gray or black), with black reserved for verified Mg^{2+}/Mn^{2+} -binding. The location of the experimentally verified phosphatases in the tree suggests that most, if not all, members are catalytically active. Based on the tree, six subfamilies numbered around the edge and shown in Figure 1 have been defined.

members that might have escaped our attention. In animals, regulatory cascades are usually much more complex and the number of paralogues therein will probably be much higher.

The 11 motifs identified should therefore prove useful in the identification of additional members remaining to be sequenced. The alignment in Figure 1 should guide the characterization of specific features of the identified subfamilies when more functional information can be attached to particular members.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Arigoni F, Pogliano K, Webb CD, Stragier P, Losick R. 1995. Localization of protein implicated in establishment of cell type to sites of asymmetric division. *Science* 270:637–40.
- Bairoch A, Bucher P. 1994. PROSITE: Recent developments. *Nucleic Acids Res* 22:3583–3589.
- Bork P, Gibson T. 1996. Applying motif and profile searches. *Methods Enzymol* 266:162–184.
- Chern Y, Chiou JY, Lai HL, Tsai MH. 1995. Regulation of adenylyl cyclase type VI activity during desensitization of the A2a adenosine receptor-mediated cyclic AMP response: Role for protein phosphatase 2A. *Mol Pharmacol* 48:1–8.
- Cohen P. 1989. The structure and regulation of protein phosphatases. *Annu Rev Biochem* 58:453–508.
- Cohen P. 1992. Signal integration at the level of protein kinases, protein phosphatases and their substrates. *Trends Biochem Sci* 17:408–413.
- Cohen PTW. 1994. Nomenclature and chromosomal localization of human protein serine/threonine phosphatase genes. *Adv Protein Phosphat* 8:371–376.
- Duncan L, Alper S, Arigoni F, Losick R, Stragier P. 1995. Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science* 270:641–644.
- Feger G, De-Vendittis E, Vitelli A, Masturzo P, Zahn R, Verrotti AC et al. 1991. Identification of regulatory residues of the yeast adenylyl cyclase. *EMBO J* 10:349–359.
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397–403.
- Gribskov M, McLachlan AD, Eisenberg D. 1987. Profile analysis: Detection of distantly related proteins. *Proc Natl Acad Sci USA* 84:4355–4358.
- Jewett JR, Koller KJ, Goeddel DV, Lowe DG. 1993. Hormonal induction of low affinity receptor guanylyl cyclase. *EMBO J* 12:769–777.
- Kaneko T, Tanaka A, Sato S, Kotani H, Suzuki T, Miyajima N et al. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1Mb region map position: 64%–92% of the genome. *DNA Res* 2:153–166.
- Klumpp S, Hanke C, Donella-Deana A, Beyer A, Kellner R, Pinna LA et al. 1994. A membrane-bound protein phosphatase type 2C from *Paramecium tetraurelia*. *J Biol Chem* 269:32774–32780.
- Koller KJ, de-Sauvage FJ, Lowe DG, Goeddel DV. 1992. Conservation of the kinaselike regulatory domain is essential for activation of the natriuretic peptide receptor guanylyl cyclases. *Mol Cell Biol* 12:2581–2590.
- Koonin EV. 1993. Bacterial and bacteriophage protein phosphatases. *Mol Microbiol* 8:785–786.
- Lacombe ML, Eberentz-Lhomme C. 1983. Protoporphyrin IX activates the Mg dependent guanylate cyclase from rat liver plasma membranes. *Biochem Biophys Res Commun* 116:47–53.
- Lawson JE, Niu XD, Browning KS, Trong HL, Yan J, Reed LJ. 1993. Molecular cloning and expression of the catalytic subunit of bovine pyruvate dehydrogenase phosphatase and sequence similarity with protein phosphatase 2C. *Biochemistry* 32:8987–8993.
- Leng J, Cameron AJM, Buckel S, Kennelly PJ. 1995. Isolation and cloning of a protein-serine/threonine phosphatase from an archaeon. *J Bacteriol* 177:6510–6517.
- Leung J, Bouvier-Durand M, Morris PC, Guerrier D, Chefedor F, Giraudat J. 1994. *Arabidopsis* ABA response gene AB11: Features of a calcium-modulated protein phosphatase. *Science* 264:1448–1452.
- Meyer K, Leube MP, Grill E. 1994. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* 264:1452–1455.
- Potts M, Sun H, Mockaitis K, Kennelly PJ, Reed D, Tonks NK. 1993. A protein-tyrosine/serine phosphatase encoded by the genome of the cyanobacterium *Nostoc commune* UTEX 584. *J Biol Chem* 268:7632–7635.
- Robinson MK, van-Zyl WH, Phizicky EM, Broach JR. 1994. TPD1 of *Saccharomyces cerevisiae* encodes a protein phosphatase 2C-like activity implicated in tRNA splicing and cell separation. *Mol Cell Biol* 14:3634–3645.
- Rost B, Sander C, Schneider R. 1994. PHD—An automatic mailserver for protein secondary structure prediction. *Comput Appl Biosci* 10:53–60.
- Rudner XL, Mandal KK, de-Sauvage FJ, Kindman LA, Almenoff JS. 1995. Regulation of cell signaling by the cytoplasmic domains of the heat-stable enterotoxin receptor: Identification of autoinhibitory and activating motifs. *Proc Natl Acad Sci USA* 92:5169–5173.
- Shiozaki K, Russell P. 1995. Counteractive roles of protein phosphatase 2C (PP2C) and a MAP kinase kinase homolog in the osmoregulation of fission yeast. *EMBO J* 14:492–502.
- Singh S, Lowe DG, Thorpe DS, Rodriguez H, Kuang WJ, Dangott LJ et al. 1988. Membrane guanylate cyclase is a cell-surface receptor with homology to protein kinases. *Nature* 334:708–712.
- Suzuki N, Choe HR, Nishida Y, Yamawaki-Kataoka Y, Ohnishi S, Tamaoki T et al. 1990. Leucine-rich repeats and carboxyl terminus are required for interaction of yeast adenylyl cyclase with RAS proteins. *Proc Natl Acad Sci USA* 87:8711–8715.
- Tamura S, Lynch KR, Lerner J, Fox J, Yasui A, Kikuchi K et al. 1989. Molecular cloning of rat type 2C (IA) protein phosphatase mRNA. *Proc Natl Acad Sci USA* 86:1796–1800.
- Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc Natl Acad Sci USA* 91:12091–12095.
- Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.