

# Molecular Evolution of Type 1 Serine/Threonine Protein Phosphatases

Qing Lin, Edward S. Buckler IV,<sup>1</sup> Spencer V. Muse,<sup>1</sup> and John C. Walker

Division of Biological Sciences, the University of Missouri–Columbia, Columbia, Missouri 65211

Received February 18, 1998; revised May 26, 1998

**Type 1 serine/threonine protein phosphatases (PP1s) play key roles in many cellular processes. To understand the evolutionary relationships among PP1s from various kingdoms and to provide a valid basis to evaluate the structure–function relationships of these phosphatases, 44 PP1 sequences were aligned, revealing a high sequence similarity among PP1 homologs. About one-third of the total amino acids are conserved in all the sequences studied. Most of these conserved amino acids are located within a 270-amino-acid core region. They include most sites critical to the activity and regulation of PP1s based on three-dimensional structural studies of mammalian PP1s. Positional variation analysis using a sliding window approach revealed two variable blocks in the 270-amino-acid core region. The major variable block corresponds to a subdomain composed of three  $\alpha$ -helices ( $\alpha$ G,  $\alpha$ H, and  $\alpha$ I) and three  $\beta$ -sheets ( $\beta$ 7,  $\beta$ 8, and  $\beta$ 9). Phylogenetic analyses suggested that plant and animal PP1s form distinct monophyletic groups. The plant PP1 family contains several subgroups that may be older than the monocot–dicot divergence. In the animal PP1 family, different vertebrate isoforms appear to form distinct subgroups. Relative substitution rate studies indicated that plant PP1s are more diverse than animal PP1s, with an average substitution rate 1.5 times as large as that of animal PP1s. The possible involvement of PP1s in the establishment of multicellularity is discussed.** © 1999 Academic Press

## INTRODUCTION

Protein phosphatases catalyze the dephosphorylation of phosphoproteins and are classified into three groups based on the phosphorylation sites of the substrates: serine/threonine phosphatases, tyrosine phosphatases, and dual-specificity phosphatases (see reviews in Barford, 1996; Cohen, 1997). Serine/threonine phosphatases are encoded by two distinct gene fami-

lies: PPP and PPM. The PPP family contains type 1 protein phosphatases (PP1s), type 2A protein phosphatases (PP2As), and type 2B protein phosphatases (PP2Bs), whereas the PPM family is composed of type 2C protein phosphatases (PP2Cs) and the related mitochondrial pyruvate dehydrogenase phosphatases.

As a major type of serine/threonine protein phosphatase, PP1s have been shown to be present in various eukaryotic organisms. Except in budding yeast *Saccharomyces cerevisiae*, which contains only one *PP1* gene (Feng *et al.*, 1991), PP1s are encoded by multiple genes: two in fission yeast *Schizosaccharomyces pombe* (Ohkura *et al.*, 1989), three in humans and rats (Barker *et al.*, 1994; Sasaki *et al.*, 1990), and four in *Drosophila* (Dombrádi *et al.*, 1993). Based on their sequence similarity, animal PP1s are further classified into three isoforms: PP1 $\alpha$ , PP1 $\beta$  (also called PP1 $\delta$ ), and PP1 $\gamma$ . In plants, the numbers of *PP1* genes present in an individual organism are larger than in animals. At least six *PP1* genes exist in maize (Sheen, 1993), and nine are found in *Arabidopsis* (Smith *et al.*, 1995; Smith and Walker, 1993).

PP1s have been shown to play key roles in various cellular processes. In yeast, mutations in PP1 cause defects in mitosis, meiosis, cell integrity, and glycogen metabolism (see review, Stark, 1996). In addition, the *S. cerevisiae* PP1, GLC7, is implicated in controlling protein synthesis (Wek *et al.*, 1992). In animals, PP1s have also been shown to modulate glycogen metabolism, protein synthesis, and the cell cycle (Axton *et al.*, 1990; Ingebritsen *et al.*, 1983; Redpath and Proud, 1990). In addition, animal PP1s are involved in controlling muscle contraction, RNA splicing, and intracellular transport (Cardinali *et al.*, 1994; Lamb *et al.*, 1988; Pato, 1985). In contrast, the functions of plant PP1s have not been well defined. However, studies using okadaic acid, a specific inhibitor of PP1s and PP2As, suggest that these two types of protein phosphatases may participate in many aspects of plant growth and development, such as ion channel activities, the cell cycle, root morphogenesis, pollination, and signal transduction initiated by pathogen elicitors, sugars, light,

<sup>1</sup> Current address: Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, NC 27695.

and plant hormones (see review in Smith and Walker, 1996).

The wide functional diversity of PP1s is likely a result of the structural complexity of PP1s *in vivo*. Although the numbers of *PP1* genes in yeast and animals are relatively small, PP1s form protein complexes with various regulatory subunits. Rapidly increasing numbers of regulatory subunits (at least 15 in mammals and 11 in *S. cerevisiae*) have been identified (Egloff *et al.*, 1997; Stark, 1996). These regulatory subunits control substrate specificity and subcellular localization of PP1s, and they allow the activity of PP1s to be regulated by extracellular signals. In plants, although native PP1s appear to be associated with other proteins (MacKintosh *et al.*, 1991), no regulatory subunits of PP1s have been isolated.

The three-dimensional structures of mammalian PP1 catalytic subunits have been defined by crystallographic studies (Egloff *et al.*, 1995; Goldberg *et al.*, 1995). The core region of PP1 forms a compact ellipsoidal structure, consisting of a central distorted  $\beta$ -sandwich of 11  $\beta$ -strands surrounded by 7  $\alpha$ -helices on one side and a subdomain composed of 3  $\alpha$ -helices and a 3-strand  $\beta$ -sheet on the other side. In the central region of the distorted  $\beta$ -sandwich, 3  $\beta$ -strands connected by 2  $\alpha$ -helices form a  $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$  motif. This motif is proposed as the active site of the enzyme where two metal ions bind. The presence of metal ions at the active site suggests that PP1 dephosphorylates its substrates through metal ion-mediated hydrolysis (Egloff *et al.*, 1995; Goldberg *et al.*, 1995).

Although much effort has been focused on the biochemistry and molecular biology of yeast and mammalian PP1s, little is known about the common or distinct functions of individual PP1 isoforms or the molecular evolution of PP1. In this study, we have aligned the PP1 sequences from plants, animals, and unicellular eukaryotes and analyzed their evolutionary relationships. These intersequence relationships provide a basis for assessing the relevance of biochemical and molecular biological studies of the common or distinct functions of individual PP1s and may be useful in predicting the structure, regulation, and function of less well-characterized PP1 proteins, such as the plant PP1s. In addition, this study reveals two regions with relatively high variability, one corresponding to a subdomain identified by X-ray crystallographic studies (Egloff *et al.*, 1995; Goldberg *et al.*, 1995). The identification of these regions may facilitate future studies on the specificity and regulation of different PP1 isoforms *in vivo*.

## MATERIALS AND METHODS

### Sequences

All of the sequences analyzed in this study were obtained from Blast searches of National Center for

Biotechnology Information databases (Altschul *et al.*, 1990). PP1 sequences from four distinct species (TOPP1 from *Arabidopsis thaliana*, PP1 $\alpha$  from rabbit, GLC7 from *S. cerevisiae*, and TbPP1-1 from *Trypanosoma brucei*) were used as query sequences to ensure that no PP1 homologs were overlooked in the search. Only full-length sequences were used, and they are listed in Table 1. In some cases, two sequences from the same

**TABLE 1**  
**PP1 Sequences Used for Alignment and Phylogenetic Analyses**

Taxon	Sequence name	Accession No.
Plants		
<i>Arabidopsis thaliana</i>	TOPP1	M93408
	TOPP2	M93409
	TOPP3	M93410
	TOPP4	M93411
	TOPP5	M93412
	TOPP6	U80920
	TOPP7	U80921
	TOPP8	U80922
Tobacco ( <i>Nicotiana tabacum</i> )	NPP1	Z93768
	NPP2	Z93769
	NPP3	Z94770
<i>Brassicaoleracea</i>	BoPP1	X63558
	PP1Ms	X80788
	PvPP1	Z48221
	OsPP1	U31773
	ZmPP1	M60215
Algae		
<i>Acetabularia cliftonii</i>	AcPP1-1	Z28627
	AcPP1-2	Z28632
<i>Gonyaulax polyedra</i>	GpPP1	U52691
Animals		
Nematode ( <i>Caenorhabditis elegans</i> )	CePP1	P48727
	PP196A	X56438
Fruit fly ( <i>Drosophila melanogaster</i> )	PP187B	X55198
	PP19C	X56439
	PP113C	X69974
	XIPP1 $\gamma$ 1	L17039
Frog ( <i>Xenopus laevis</i> )	XgPP1 $\beta$	D37987
Chicken ( <i>Gallus gallus</i> )	DIS2M1	M27071
	DIS2M2	M27073
Mouse ( <i>Mus musculus</i> )	RnPP1 $\alpha$	D00859
	RnPP1 $\beta$	S78218
	RnPP1 $\gamma$ 1	D90165
Rat ( <i>Rattus norvegicus</i> )	OcPP1 $\alpha$	X07798
Rabbit ( <i>Oryctolagus cuniculus</i> )	OcPP1 $\beta$	X61639
	HsPP1 $\alpha$	X70848
Human ( <i>Homo sapiens</i> )	HsPP1 $\beta$	X80910
	HsPP1 $\gamma$	X74008
Fungi		
Budding yeast ( <i>Saccharomyces cerevisiae</i> )	GLC7	M77175
	DIS2	M27067
Fission yeast ( <i>Schizosaccharomyces pombe</i> )	SDS21	M27069
	BimG	M27067
<i>Emericella nidulans</i>		
Protists		
<i>Paramecium tetraurelia</i>	PtPP1 $\omega$ 1	L26481
	PtPP1 $\omega$ 2	L26482
<i>Trypanosoma brucei</i>	TbPP1-1	M74166
	TbPP1-2	M74167

species, such as human PP1a X70848 vs J04759 and rabbit PP1a X07798 vs X14832, are almost identical, except at one or two nucleotide positions. These are very likely the result of polymorphism or sequencing errors; therefore, only one sequence was included in this study. Similarly, in the cases where the only difference between two sequences was at the amino or carboxyl terminus, only one sequence was analyzed, as all the following analyses used a 270-amino-acid core region of PP1s (see below). The nucleotide sequence of PP1 from *C. elegans* is not available in the databases, and only the amino acid sequence was used.

### Sequence Alignment

Multiple alignments of PP1 amino acid and nucleotide sequences were made using the Clustal method of DNASTAR Megalign program (Higgins and Sharp, 1988). Nucleotide alignments were created by reverse translation of the amino acid alignments. Minor adjustments were performed manually. For amino acid alignments, the PAM250 residue weight table was used. Alignments are available by anonymous ftp from hillbill.biosci.missouri.edu to directory pub/datasets/pp1.

### Phylogenetic Analysis

Phylogenetic analyses using distance methods were carried out with the PHYLIP package (Felsenstein, 1993). Only the sequence region that could be aligned without ambiguity in all homologs was included (see Fig. 1). For such purposes, the amino terminus (up to Gln29 of human HsPP1 $\gamma$ 1) and the carboxyl terminus (from Ala298 of human HsPP1 $\gamma$ 1) were excluded from the analyses. Distance matrices were generated by the PROTDIST program using the PAM model of amino acid substitution (Dayhoff, 1979). These matrices were analyzed by the FITCH program to construct a phylogenetic tree (Fitch and Margoliash, 1967). Confidence in the topology of the tree was evaluated by performing 100 bootstrap replicates using the SEQBOOT program (Felsenstein, 1985).

Phylogenetic analyses using parsimony methods were carried out using the PAUP program (Swofford, 1993). Only first and second bases in codons for the 270 amino acids in the PP1 core region were considered. These two base positions were weighted in the following manner. First, the 270-amino-acid core region was divided into two conserved regions (C1 and C2) and two variable regions (V1 and V2) as described under Results. The number of base substitutions at the first and second codon positions in either conserved or variable regions was inferred using the MacClade program (Maddison and Maddison, 1992), and the inverses of these values were used as weight parameters. The resulting weight parameters for the conserved and variable regions were 0.54 and 0.79 for the first position and 1.10 and 2.22 for the second position. The nucleotide phylogenetic tree was constructed using the heuristic search algorithm.

Branch swapping was performed with the tree-bisection-reconnection procedure and the MULPARS option is in effect. The reliability of the resulting tree was assessed by bootstrapping the data 100 times.

### Positional Variation Analysis

Amino acid sequence divergence was evaluated over the 270-amino-acid core region used in the phylogenetic analyses with a sliding window of 20 residues. First, the amino acid distance tree (see above) was used to determine the number of unambiguous character changes at each amino acid position using the MacClade program (Maddison and Maddison, 1992). Then the average number of character changes was calculated over the sliding window across the sequence and plotted against amino acid positions. A plot based on an amino acid parsimony tree using the PAUP program produced a similar profile.

### Relative Rate Test

Comparison of average substitution rates between plant and animal PP1s was carried out using Li and Bousquet's (1992) relative-rate tests with *S. cerevisiae* GLC7 as an outgroup. The Li and Bousquet test was chosen over standard relative rate tests (e.g., Wu and Li, 1985; Muse and Weir, 1992) for its ability to compare average substitution rates between two groups of taxa. The number of nucleotide substitutions was measured with Nei and Gojobori's (1986) method. The null hypothesis of rate homogeneity between the two PP1 groups was tested by calculating *Z* statistics.

## RESULTS

### Sequence Alignment

Pairwise comparisons of 44 PP1 sequences from various organisms (Table 1) were carried out, and the resulting similarity matrix is shown in Table 2. The minimum amino acid similarity is about 55%, obtained by comparison between PP1s from *T. brucei* (TbPP1-1 and TbPP1-2) and those from other species over their entire length. This indicates that TbPP1s are the most distantly related PP1s compared with the others. When TbPP1s are excluded, the similarity of PP1 amino acid sequences increases to over 60%, indicating the high conservation of PP1s through evolution. This is espe-

TABLE 2

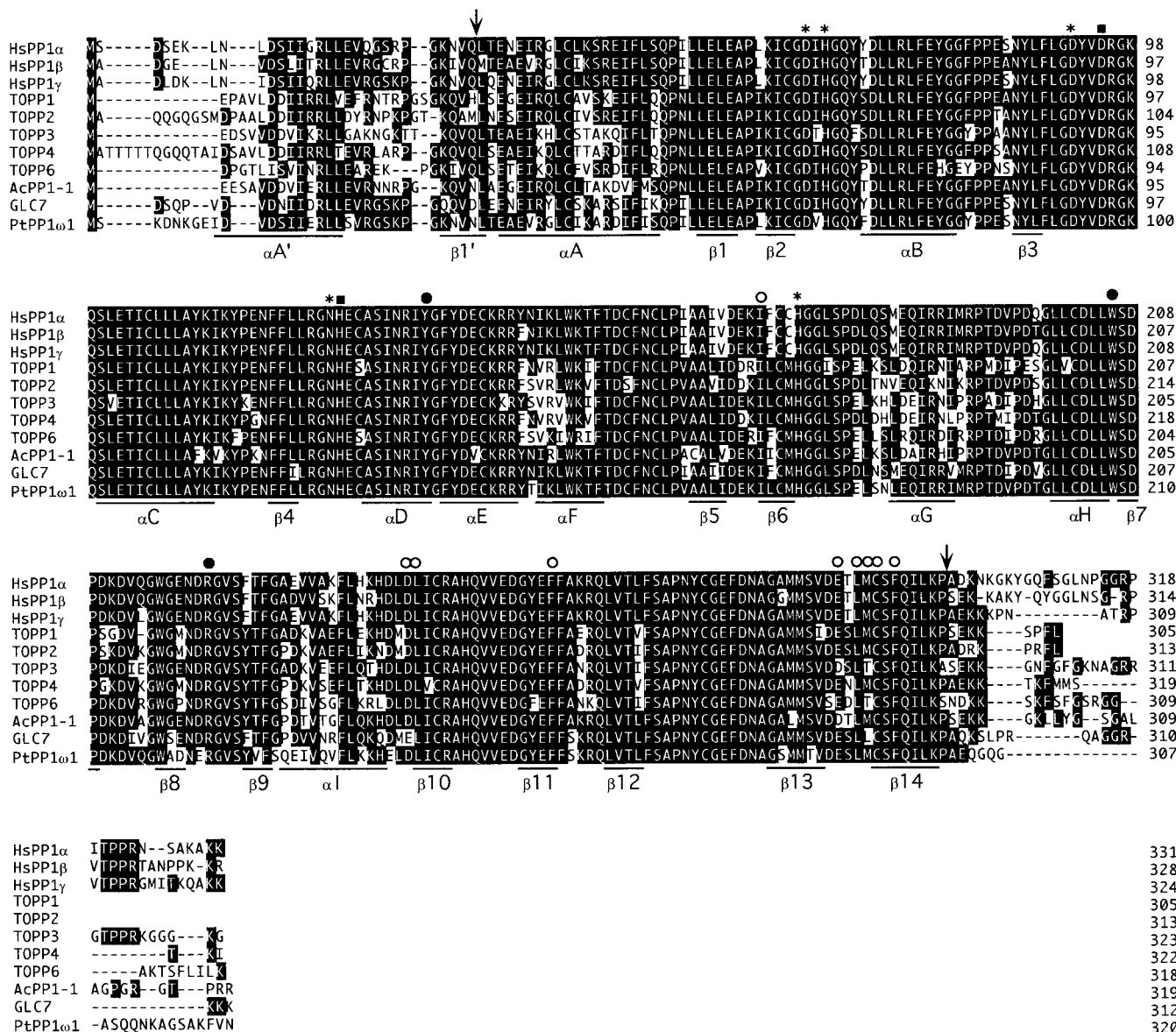
#### Percentage Range of PP1 Amino Acid Similarity

	Plants	Algae	Animals	Fungi	Protists
Higher plants	66.0–91.8	65.9–78.7	63.9–79.1	65.5–77.0	54.5–74.1
Algae		74.8–88.7	75.2–81.8	71.8–77.7	58.0–76.5
Animals			79.8–100	72.4–84.5	56.5–80.5
Fungi				76.3–89.8	55.4–77.9
Protists					58.7–99.7

cially obvious for animal PP1s, whose minimum similarity approaches 80%. One striking example of the high conservation is that the  $\beta$  isoforms of PP1 from chickens, rats, mice, rabbits, and humans are identical at the amino acid level.

A multiple alignment of PP1 sequences was carried out as described under Materials and Methods. Because of space limitation, only the alignment of representative sequences is presented in Fig. 1. Alignment of six of the PP1 sequences from *Arabidopsis* is included to illustrate the variability among the plant genes. The alignment reveals the high sequence similarity among

PP1 homologs except at the amino and carboxyl termini and is consistent with a previous analysis of the protein serine/threonine phosphatases (Barton *et al.*, 1994). About one-third of the total amino acids are identical in all the sequences studied. Most conserved amino acids are clustered in a 270-amino-acid core region and include most sites critical to the functions of PP1s based on three-dimensional structure studies of mammalian PP1s, such as amino acids involved in catalysis or in binding to metal ions, substrates, and regulatory subunits (Egloff *et al.*, 1995, 1997; Goldberg *et al.*, 1995). However, a few exceptions exist. Tobacco NPP3



**FIG. 1.** Multiple alignments of PP1 amino acid sequences. Alignments were made as described under Materials and Methods. Shaded areas: residues identical to the consensus sequence. Asterisks: metal-binding residues. Filled squares: catalytic sites. Filled circles: substrate (phosphoserine or phosphothreonine)-binding sites. Open circles: regulatory subunit-binding sites. Arrows: the border of the region used in subsequent phylogenetic analyses. Dashes: gaps. Lines and letters under the alignments: the secondary structures according to crystallographic studies (Goldberg *et al.*, 1995).

has an Arg instead of a Trp at position 206 (the numeration of amino acid positions is based on HsPP1 $\gamma$ 1 sequence). Trp206 was proposed as one of the phosphoserine or threonine-binding sites (Egloff *et al.*, 1995). In addition, deviations at three putative regulatory subunit-binding sites were observed. *Drosophila* PP1 13C has an Asn substitution for Asp or Glu287. *Gonyaulax polyedra* has Phe instead of Leu at position 289. The most variable residue that may be involved in binding of PP1 catalytic subunits to regulatory subunits is at position 290. Instead of Met, *Arabidopsis* TOPP3, TOPP6, and TOPP7 and alfalfa PP1Ms have Thr; fungal PP1s and tobacco NPP3 have Leu; and *G. polyedra* PP1 has Cys.

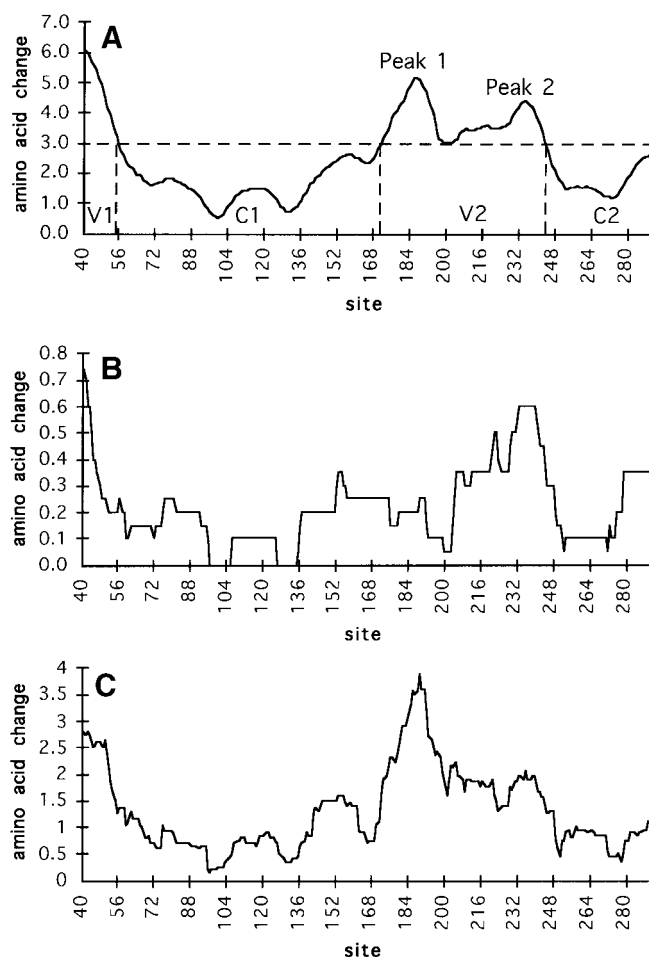
#### Positional Variation Analysis

To further define amino acid divergence of PP1 sequences, positional variation analysis was carried out over the 270-amino-acid core region with a sliding window of 20 amino acids (Fig. 2A). The result shows that the levels of divergence vary among different regions of PP1 sequences. Two major peaks were observed: peak 1 at site 185, and peak 2 at site 234. High variation was also observed at the amino end of the region studied. If 50% of the maximum character change per window (3.0) is used as a threshold, the 270-amino-acid region can be divided into four blocks: variable block 1 (V1, from position 30 to 54 of HsPP1 $\gamma$ 1), conserved block 1 (C1, from position 55 to 171 of HsPP1 $\gamma$ 1), variable block 2 (V2, from position 172 to 243 of HsPP1 $\gamma$ 1), and conserved block 2 (C2, from position 244 to 298 of HsPP1 $\gamma$ 1). Supposing that the main 3-dimensional structures of PP1s are retained in evolution, the major conserved region (C1) includes the  $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$  motif at the active site, three  $\alpha$ -helices ( $\alpha$ D,  $\alpha$ E, and  $\alpha$ F), and two  $\beta$ -sheets ( $\beta$ 5 and  $\beta$ 6). The major variable region (V2) corresponds to a subdomain formed by three  $\alpha$ -helices ( $\alpha$ G,  $\alpha$ H, and  $\alpha$ I) and three  $\beta$ -sheets ( $\beta$ 7,  $\beta$ 8, and  $\beta$ 9).

The amino acid positional variation analyses among the animal PP1 family (Fig. 2B) and the plant PP1 family (Fig. 2C) indicate that their overall divergence patterns are similar to that when all the PP1 sequences were included. However, small differences exist. For instance, the animal PP1 family lacks peak 1, whereas plant peak 1 is pronounced.

#### Molecular Phylogeny

The phylogenetic analyses based on both nucleotide and amino acid sequences (Figs. 3A and 3B) show that plant and animal PP1s may form distinct groups, although the bootstrapping support varies by the kind of sequences used. The bootstrap values for plant and animal PP1s, respectively, are 63 and 58% for nucleotide sequences and 53 and 90% for amino acid sequences. Most unicellular eukaryotes included in this study have their two PP1 genes in an exclusive clade, suggesting multiple independent gene duplications. An

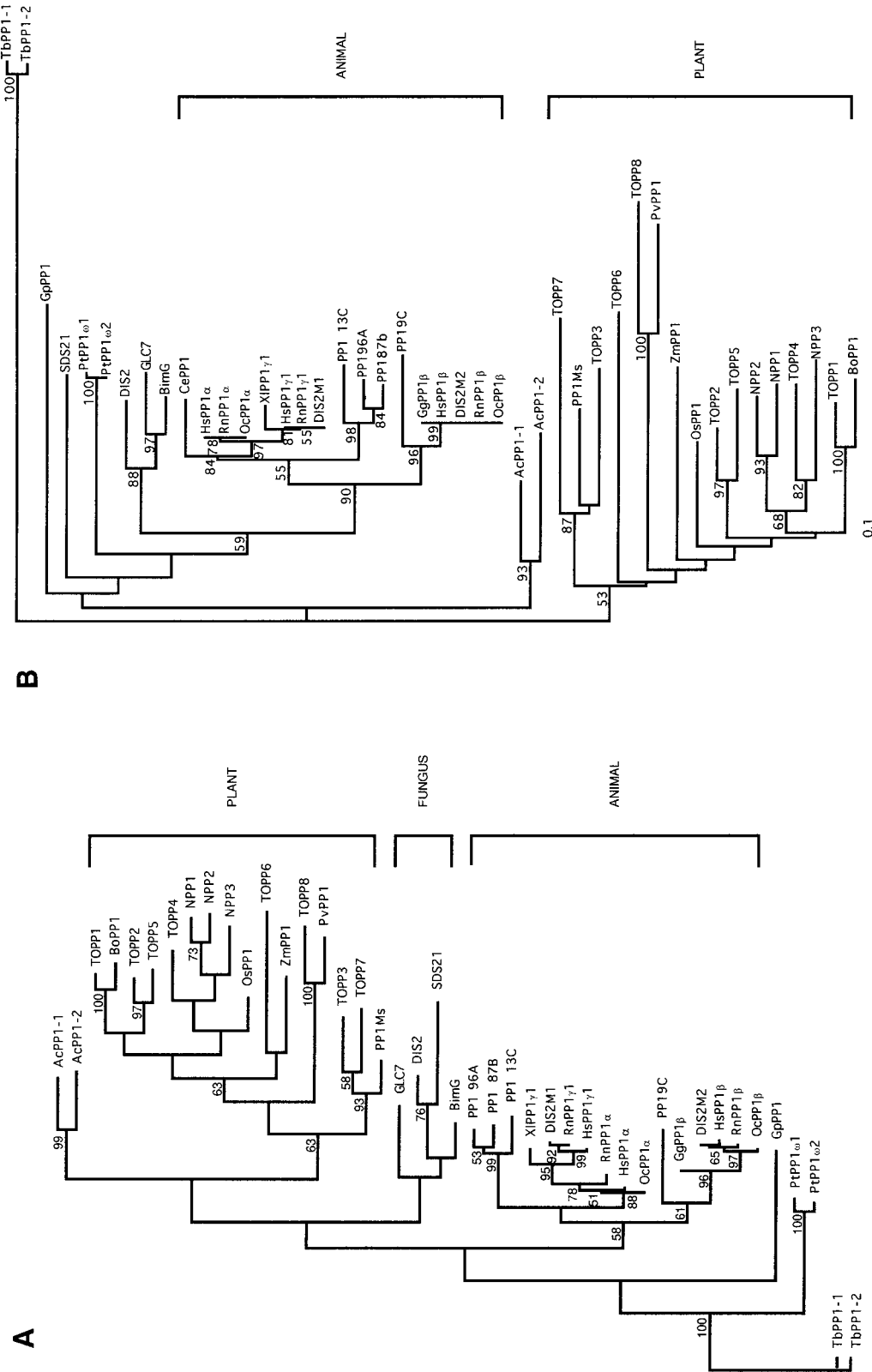


**FIG. 2.** Positional variation analysis. The average amino acid change over the sliding window of 20 residues across the 270-amino-acid core region was plotted against amino acid positions. (A) Amino acid changes among all PP1 proteins per 20-amino-acid block. C1: conserved region 1. C2: conserved region 2. V1: variable region 1. V2: variable region 2. (B) Amino acid changes among animal PP1 proteins per 20-amino-acid block. (C) Amino acid changes among plant PP1 proteins per 20-amino-acid block.

exception for the above notion is that the two PP1s from *S. pombe* (SDS21 and DIS2) are not clustered using amino acid sequences. The intersequence relationships of PP1s among plants, animals, and different species of algae, protists, and fungi are not resolved in this study.

The family of plant PP1s contains several small subgroups of two or three members. The phylogenetic relationships of these subgroups do not agree with the phylogenetic relationships of the plant taxa. The trees significantly reject monophyly for *Arabidopsis* PP1 isoforms. There is no support for monocot and dicot forms being monophyletic. It is possible that these subgroups resulted from gene duplications prior to the divergence of monocots and dicots.

Similarly, animal PP1s also form distinct subgroups. Both vertebrate PP1 $\gamma$  and PP1 $\beta$  isoforms are monophyletic, whereas PP1 $\alpha$  isoforms are monophyletic based



**FIG. 3.** Phylogenetic analysis of PP1 genes. (A) A phylogenetic tree based on PP1 nucleotide sequences constructed using maximum parsimony. (B) A phylogenetic tree based on PP1 amino acid sequences constructed by FITCH clustering of Dayhoff distances. The scale bar indicates the scale of sequence divergence. For both trees, *TbPP1-1* and *TbPP1-2* were used as outgroups. The lengths of branches in the tree shown in A are proportional to the number of nucleotide changes between sequences. The numbers above or below branches are the percentage of bootstrap trials that support the clade. Branches without bootstrap percentages were found to be less than 50%.

on amino acid sequences but not on nucleotide sequences. The vertebrate PP1 $\alpha$  and PP1 $\gamma$  isoforms are closer to each other than to PP1 $\beta$  isoforms. The three *Drosophila* PP1 $\alpha$  isoforms are grouped together, but their relationships with the mammalian PP1 $\alpha$ –PP1 $\gamma$  subgroup and the PP1 $\beta$  subgroup is not well supported. The only *Drosophila* PP1 $\beta$  is grouped with the vertebrate PP1 $\beta$  subgroup, suggesting that the PP1 $\beta$  isoforms originated before the divergence of invertebrates and vertebrates.

One common feature between the nucleotide tree and the amino acid tree is that the branch lengths of plant PP1s are larger than those of animal PP1s. This suggests that plant PP1s have diversified more during evolution than animal PP1s have. To further compare the level of divergence among plant and animal PP1s, we performed relative substitution rate tests.

#### *Relative Substitution Rate*

The results of relative substitution rate tests are consistent with those of the phylogenetic analyses. Plant PP1s have diversified 1.5 times as fast as animal PP1s have. The ratio of nucleotide substitution rates between plant and animal PP1s is 1.2 for region V1, 2.1 for region C1, 1.3 for region V2, and 1.4 for region C2. Except for the small region V1, the rate heterogeneity between plant and animal PP1s is significant in all the other regions studied ( $P < 0.05$ ).

## DISCUSSION

Phylogenetic analysis of a gene family not only generates information about its evolutionary history, but also provides a rational basis for assessing its structure–function relationships. In this study, we carried out phylogenetic analysis of type 1 serine/threonine protein phosphatases to investigate the evolutionary mechanisms and to facilitate our understanding of their structure–function relationships.

#### *Multiple Alignment*

Multiple alignments of PP1s from various kingdoms indicate that PP1 sequences are highly conserved. About one-third of the amino acid residues are unchanged among all the sequences studied. These residues include most of those corresponding to sites critical to the activity and regulation of mammalian PP1s. This fact suggests that the major features of PP1 structures may be conserved in evolution. The high degree of conservation in PP1 sequences also suggests that PP1s may perform similar essential functions in different kingdoms. One such function may be the control of the cell cycle: PP1s have been implicated in cell cycle regulation in fungi, animals, and plants (Axton *et al.*, 1990; Booher and Beach, 1989; Brautigan *et al.*, 1990; Doonan and Morris, 1989; Hasezawa and Nagata, 1992; Ohkura *et al.*, 1989; Stark *et al.*, 1994).

Despite the high degree of overall sequence conservation among PP1s, variations do exist, especially at the amino and carboxyl termini. Even in the relatively conserved 270-amino-acid core region, positional variation analysis reveals a major variable block (V2) that corresponds to a subdomain composed of three  $\alpha$ -helices ( $\alpha$ G,  $\alpha$ H, and  $\alpha$ I) and three  $\beta$ -sheets ( $\beta$ 7,  $\beta$ 8, and  $\beta$ 9) of rabbit PP1 $\alpha$  (Goldberg *et al.*, 1995). Based on crystallographic studies (Egloff *et al.*, 1995, 1997; Goldberg *et al.*, 1995), this subdomain is located at the surface of PP1, which may explain its relatively large number of changes compared to interior regions. However, it is interesting to note that some secondary elements in the major conserved block (C1), such as  $\alpha$ -helices  $\alpha$ D,  $\alpha$ E, and  $\alpha$ F and  $\beta$ -sheets  $\beta$ 5 and  $\beta$ 6, are also located on the surface of PP1s. The reasons for greater diversity in some exterior regions of PP1s than in others are unknown. Further biochemical and molecular studies should provide insights into the functions of these regions.

In addition to the identification of variable regions, several deviations at the conserved sites important to the activity or regulation of PP1s were also observed. Tobacco NPP3 has an Arg instead of a Trp at one of the phosphoserine or threonine-binding sites (Egloff *et al.*, 1995). It would be interesting to examine the specificity and activity of NPP3. In addition, deviations at three putative regulatory subunit-binding sites were observed. These sites would be attractive for site-directed mutagenesis studies to assess their roles in the interaction between PP1 catalytic subunits and regulatory subunits.

#### *Phylogenetic Analyses*

Phylogenetic analyses show that the evolution of PP1 genes includes both extensive gene duplication and divergence. At least two duplication events have occurred in animals, whereas several duplications have occurred in plants. These independent events in plants and animals suggest that the diversification of PP1s may have been important to the evolution and development of these taxa. The gene duplications appear to have occurred early during evolution of both plants (before the monocot and dicot split) and animals (before the flatworm, insect, and vertebrate split). This notion is also consistent with the different chromosomal positions of PP1 genes from rats, humans, *Drosophila*, and *Arabidopsis* (Baker *et al.*, 1994; Dombrádi *et al.*, 1993; Muramatsu *et al.*, 1994, 1995; Yamada *et al.*, 1995; Lin and Walker, unpublished data). Because the early stages of plant and animal evolution may have involved developing careful control of cell division, multicellular organization, and tissue differentiation, PP1's duplications and divergence may have contributed to these changes.

The ancient, duplicated PP1 genes have been well conserved through evolution. The high degree of conser-

vation is most obvious within vertebrate PP1 subgroups. One striking example is the PP1 $\beta$  subgroup with identical amino acid sequences among members from chickens, rats, mice, rabbits, and humans. Considering that birds and mammals probably diverged from each other over 300 Myr ago (Carroll, 1988), the extreme evolutionary constraint on the PP1 $\beta$  subgroup suggests that they perform essential functions distinct from other isoforms. In addition, the phylogenetic analyses also suggest that  $\beta$  isoforms should be contrasted with  $\alpha$  and  $\gamma$  isoforms in functional studies. However, these hypotheses do not exclude the possibility that PP1 $\beta$  isoforms share some common functions with other isoforms. PP1 $\beta$  has been shown to be coexpressed with PP1 $\alpha$  in many different rat tissues (Shima *et al.*, 1993). In addition, both PP1 $\alpha$  and PP1 $\beta$  bind to a glycogen-binding subunit in rat liver, presumably performing similar functions in regulating glycogen metabolism (Moorhead *et al.*, 1995).

Similar to animal PP1s, different PP1 subgroups across plant species have also been well maintained, suggesting that they perform distinct functions in plants. However, the intersequence relationships of plant subgroups are not reliably resolved in this study. This may be a result of the relatively high divergence among plant PP1s and the limited numbers of plant PP1 sequences. Only two full-length PP1 sequences from monocots (maize ZmPP1 and rice OsPP1) were available for this study, although at least six *PP1* genes exist in the maize genome (Sheen, 1993). The lack of monocot sequences may prevent orthologous PP1 relationships from being recognized. The identification of additional *PP1* genes from monocots will enable us to characterize the evolutionary relationships of plant PP1s and make informative functional comparisons among plant PP1 sequences.

#### *Comparison of Plant and Animal PP1s*

One objective of this study was to provide a rational basis for determining to what degree the information about animal PP1s can be used to predict the structure, regulation, and function of less well-characterized plant PP1s. As discussed above, multiple alignment of PP1s suggests that plant PP1s may have a basic structure similar to that of animal PP1s, may also be regulated by interacting with regulatory subunits, and may share some essential functions with animal PP1s. In addition, the overall profile of positional amino acid change of plant PP1s is similar to that of animal PP1s. It is likely that the evolution of PP1 was directed primarily by stringent structural and functional constraints imposed by itself.

However, the effects of biological and ecological environments in which selection and random drift operate also shape the evolution of PP1s. Positional variation analyses not only show an overall similarity between

plant and animal PP1s, but also reveal differences in the major variable block (V2). In addition, more PP1 genes exist in individual plant species (at least six in maize and at least nine in *Arabidopsis*) (Sheen, 1993; Smith and Walker, 1993; Smith *et al.*, 1995) than in animals (three in mammals and four in *Drosophila*) (Baker *et al.*, 1994; Dombrádi *et al.*, 1993; Sasaki *et al.*, 1990), enabling plant PP1s to be more diverse than animal PP1. Considering that the functional diversity of PP1s in yeast and mammals largely results from the ability of the catalytic subunits to interact with various regulatory subunits, a small increase in *PP1* gene number would greatly enhance the potential combinations between PP1 catalytic subunits and regulatory subunits.

Plants and animal PP1s differ in their evolutionary rate dynamics, at both high and low taxonomic levels. Plant PP1s evolve roughly 1.5 times faster than animal PP1s, slightly higher than the average difference between plant and animal genes. This difference reflects typical differences between plant and animal molecular evolution, most notably the ubiquitous presence of multiple gene copies in animal nuclear genes, but might also reflect adaptive differences. Unlike plant PP1s, animal PP1s often show remarkable conservation of subgroups. Most striking is the identity of the PP1- $\beta$  subgroup in which vertebrates that diverged over 300 million years ago share identical amino acid sequences. The most reasonable explanation for such striking evolutionary behavior is the establishment of unique selective constraints in concert with the emergence of vertebrates. This contrasts sharply with evolutionary behavior found in plants, where no such conservation is observed. The difference in substitution rates may partially result from the extreme conservation within vertebrate PP1 subgroups. It appears that once these PP1 subgroups evolved, there was strong selection to maintain their amino acid sequences. This may be a consequence of their interactions with multiple regulatory subunits. Unlike animal PP1s, plant PP1s do not show extreme conservation within subgroups. Little is known about how the PP1 enzyme is regulated in plants, but perhaps the larger *PP1* gene number reduces the constraints on their divergence. The relatively high diversity of plant PP1s may suggest that some subgroups retain functions essential to all eukaryotes, whereas the others develop new functions specific for plant growth, development, or responses to external stimuli.

#### ACKNOWLEDGMENTS

We thank Dr. Julie M. Stone for critical comments on the manuscript. This research was supported by U. S. Department of Agriculture Grant 9602552 and a University of Missouri Food for the 21st Century Program grant to J.C.W.



## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- Axton, J. M., Dombrádi, V., Cohen, P., and Glover, D. M. (1990). One of the protein phosphatase 1 isoenzymes in *Drosophila* is essential for mitosis. *Cell* **63**: 33–46.
- Barker, H. M., Brewis, N. D., Street, A. J., Spurr, N. K., and Cohen, P. T. W. (1994). Three genes for protein phosphatase 1 map to different human chromosomes: Sequence, expression and gene localization of protein serine/threonine phosphatase 1 beta (PPP1CB). *Biochim. Biophys. Acta* **1220**: 212–218.
- Barford, D. (1996). Molecular mechanisms of the protein serine/threonine phosphatases. *Trends Biochem. Sci.* **21**: 407–412.
- Barton, G. J., Cohen, P. T. W., and Barford, D. (1994). Conservation analysis and structure prediction of the protein serine/threonine phosphatases. *Eur. J. Biochem.* **220**: 225–237.
- Booher, R., and Beach, D. (1989). Involvement of a type 1 protein phosphatase encoded by *bws1<sup>+</sup>* in fission yeast mitotic control. *Cell* **57**: 1009–1016.
- Brautigan, D. L., Sunwoo, J., Labbé, J.-C., Ferbadéz, A., and Lamb, N. J. C. (1990). Cell cycle oscillation of phosphatase inhibitor-2 in rat fibroblasts coincident with p34<sup>cdc2</sup> restriction. *Nature* **344**: 74–78.
- Cardinali, B., Cohen, P., and Lamond, A. I. (1994). Protein phosphatase 1 can modulate alternative 5' splice site selection in a HeLa splicing extract. *FEBS Lett.* **352**: 276–280.
- Carroll, R. L. (1988). "Vertebrate Paleontology and Evolution," Freeman, New York.
- Cohen, P. T. W. (1997). Novel protein serine/threonine phosphatases: Variety is the spice of life. *Trends Biochem. Sci.* **22**: 245–251.
- Dayhoff, M. O. (1979). "Atlas of Protein Sequences and Structure," National Biomedical Research Foundation, Washington, DC.
- Dombrádi, V., Mann, D. J., Saunders, R. D., and Cohen, P. T. W. (1993). Cloning of the fourth functional gene for protein phosphatase 1 in *Drosophila melanogaster* from its chromosomal location. *Eur. J. Biochem.* **212**: 177–183.
- Doonan, J. H., and Morris, N. R. (1989). The *bimG* gene of *Aspergillus nidulans* required for completion of anaphase, encodes a homolog of mammalian phosphoprotein phosphatase 1. *Cell* **57**: 987–996.
- Egloff, M.-P., Cohen, P. T. W., Reinemer, P., and Barford, D. (1995). Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J. Mol. Biol.* **254**: 942–959.
- Egloff, M.-P., Deborah, F. J., Moorhead, G., Cohen, P. T. W., Cohen, P., and Barford, D. (1997). Structural basis for the recognition of regulatory subunits by the catalytic subunit of protein phosphatase 1. *EMBO J.* **16**: 1876–1887.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.
- Felsenstein, J. (1993). "PHYLIP (Phylogeny Inference Package)," Department of Genetics, University of Washington, Seattle.
- Feng, Z. H., Wilson, S. E., Peng, Z. Y., Schlender, K. K., Reimann, E. M., and Trumbly, R. J. (1991). The yeast *GLC7* gene required for glycogen accumulation encodes a type-1 protein phosphatase. *J. Biol. Chem.* **266**: 23796–23801.
- Fitch, W. M., and Margoliash, E. (1967). Construction of phylogenetic trees. *Science* **155**: 279–284.
- Goldberg, J., Huang, H., Kwon, Y., Greengard, P., Nairn, A. C., and Kuriyan, J. (1995). Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* **376**: 745–753.
- Hasezawa, S., and Nagata, T. (1992). Okadaic acid as a probe to analyze the cell cycle progression in plant cells. *Bot. Acta* **105**: 63–69.
- Higgins, D. G., and Sharp, P. M. (1988). Clustal: A package for performing multiple sequence alignment on a microcomputer. *Gene* **73**: 237–244.
- Ingebritsen, T. S., Stewart, A. A., and Cohen, P. (1983). The protein phosphatases involved in cellular regulation. 6. Measurement of type-1 and type-2 protein phosphatases in extracts of mammalian tissues; an assessment of their physiological roles. *Eur. J. Biochem.* **132**: 297–307.
- Lamb, N. J. C., Fernandez, A., Conti, M.-A., Adelstein, R. D., Glass, D. B., Welch, W. J., and Feramisco, J. R. (1988). Regulation of actin microfilament integrity in living nonmuscle cells by the cAMP-dependent protein kinase and the myosin light chain kinase. *J. Cell Biol.* **106**: 1955–1971.
- Li, P., and Bousquet, J. (1992). Relative-rate test for nucleotide substitutions between two lineages. *Mol. Biol. Evol.* **9**: 1185–1189.
- MacKintosh, C., Coggins, J., and Cohen, P. (1991). Plant protein phosphatases: Subcellular distribution, detection of protein phosphatase 2C and identification of protein phosphatase 2A as the major quinate dehydrogenase phosphatase. *Biochem. J.* **273**: 733–738.
- Maddison, W. P., and Maddison, D. R. (1992). "MacClade: Analysis of Phylogeny and Character Evolution," Version 3.0, Sinauer Associates, Sunderland, MA.
- Moorhead, G., Mackintosh, C., Morrice, N., and Cohen, P. (1995). Purification of the hepatic glycogen-associated form of protein phosphatase-1 by microcystin-Sepharose affinity chromatography. *FEBS Lett.* **362**: 101–105.
- Muramatsu, Y., Yamada, T., Moralejo, D. H., Szpirer, C., and Matsumoto, K. (1995). Mapping of the gene for rat protein phosphatase 1-delta (Ppp1cd) to chromosome 6. *Mamm. Genome* **6**: 307–308.
- Muramatsu, Y., Yamada, T., Yamada, J., Serikawa, T., and Matsumoto, K. (1994). Chromosomal assignment of the gene for rat protein phosphatase 1-gamma-1 (Ppp1lc). *Cytogenet. Cell. Genet.* **67**: 58–60.
- Muse, S. V., and Weir, B. S. (1992). Testing for equality of evolutionary rates. *Genetics* **132**: 269–276.
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- Ohkura, H., Kinoshita, N., Miyatani, S., Toda, T., and Yanagida, M. (1989). The fission yeast *dis2<sup>+</sup>* gene required for chromosome disjoining encodes one of two putative type 1 protein phosphatases. *Cell* **57**: 997–1007.
- Pato, M. D. (1985). Properties of the smooth muscle phosphatases from turkey gizzards. *Adv. Prot. Phosphatases* **1**: 367–382.
- Redpath, N. T., and Proud, C. G. (1990). Activity of protein phosphatases against initiation factor-2 and elongation factor-2. *Biochem. J.* **272**: 175–180.
- Sasaki, K., Shima, H., Kitagawa, Y., Irino, S., Sugimura, T., and Nagao, M. (1990). Identification of members of the protein phosphatase 1 gene family in the rat and enhanced expression of protein phosphatase 1-alpha gene in rat hepatocellular carcinomas. *Jpn. J. Cancer Res.* **81**: 1272–1280.
- Sheen, J. (1993). Protein phosphatase activity is required for light-inducible gene expression in maize. *EMBO J.* **12**: 3497–3505.
- Shima, H., Hatano, Y., Chun, Y.-S., Sugimura, T., Zhang, Z., Lee, E. Y. C., and Nagao, M. (1993). Identification of PP1 catalytic subunit isotypes PP1 $\gamma$ 1, PP1 $\delta$  and PP1 $\alpha$  in various rat tissue. *Biochem. Biophys. Res. Commun.* **192**: 1289–1296.
- Smith, R. D., Lin, Q., Cannon, J. F., and Walker, J. C. (1995). Type-1 and type-2C protein phosphatases of higher plants. *Adv. Protein Phosphatases* **9**: 105–120.
- Smith, R. D., and Walker, J. C. (1993). Expression of multiple type 1

- phosphoprotein phosphatases in *Arabidopsis thaliana*. *Plant Mol. Biol.* **21**: 307–316.
- Smith, R. D., and Walker, J. C. (1996). Plant protein phosphatases. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**: 101–125.
- Stark, M. (1996). Yeast protein serine/threonine phosphatases: Multiple roles and diverse regulation. *Yeast* **12**: 1647–1675.
- Stark, M., Black, S., Sneddon, A., and Andres, P. D. (1994). Genetic analyses of yeast protein serine/threonine phosphatases. *FEMS Microbiol. Lett.* **117**: 121–130.
- Swofford, D. (1993). "Phylogenetic Analysis Using Parsimony," Illinois Natural History Survey, Champaign, IL.
- Wek, R. C., Cannon, J. F., Dever, T. E., and Hinnebusch, A. G. (1992). Truncated protein phosphatase *GLC7* restores translational activation of *GCN4*-expression in yeast mutants defective for the eIF-2 $\alpha$  kinase *GCN2*. *Mol. Cell. Biol.* **12**: 5700–5710.
- Wu, C-I., and Li, W-H. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* **82**: 1741–1745.
- Yamada, T., Muramatsu, Y., Moralejo, D. H., Szpirer, J., Szpirer, C., and Matsumoto, K. (1995). Mapping of the gene for rat protein phosphatase 1-alpha (*Ppp1ca*) to chromosome 1. *Mamm. Genome* **6**: 308.