

# Purple Acid Phosphatases of *Arabidopsis thaliana*

## COMPARATIVE ANALYSIS AND DIFFERENTIAL REGULATION BY PHOSPHATE DEPRIVATION\*

Dongping Li‡, Huifen Zhu‡, Kunfan Liu‡, Xin Liu‡, Georg Leggewie§, Michael Udvardi§, and Daowen Wang‡¶

From the ‡Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China and the §Max Planck Institute of Molecular Plant Physiology, 14424 Potsdam, Germany

Purple acid phosphatases (PAPs) are members of the metallo-phosphoesterase family. They are characterized by the presence of seven conserved amino acid residues involved in coordinating the dimetal nuclear center in their reactive site. We compared the 29 PAPs predicted for *Arabidopsis thaliana* in their varieties of potential metal-ligating residues. Although 24 members possessed sets of metal-ligating residues typical of known PAPs, 1 member lacked four of the seven residues. For the remaining four members, potential metal-ligating residues were generally more similar to those in metal-dependent exonucleases and related proteins. Evidence was obtained for the expression of the majority of the 29 PAPs. To facilitate future investigations, a scheme for naming *Arabidopsis* PAPs and a system for classifying the 29 PAPs are proposed. The cDNA sequences and the responses to phosphate deprivation of seven *Arabidopsis* PAPs (AtPAP7–AtPAP13) were characterized. For some AtPAPs analyzed, there were fully processed transcripts as well as splice variants. The splice variants of AtPAP10 were found to associate with polyribosomes and may be translated into a NH<sub>2</sub>-terminal truncated protein. Phylogenetic investigations showed that AtPAPs 7 and 8, together with similar enzymes from other plant species, formed the low molecular weight plant PAP group. Members of this group were more closely related to PAPs from mammalian cells. AtPAPs 9–13, together with kidney bean PAP, formed the high molecular weight PAP group. In phosphate deprivation experiments, gene transcription of AtPAP11 and AtPAP12 was induced and increased, respectively, whereas that of the remaining five AtPAPs was not affected by phosphate deprivation. The present work demonstrates that structure variation and expression regulation of plant PAPs are more complex than previously described and provides a framework for comprehensive molecular genetic and biochemical studies of all *Arabidopsis* PAPs in the future.

In recent years, there has been considerable interest in purple acid phosphatases (PAPs).<sup>1</sup> Comparative analysis of the structure of PAPs from higher plants and mammals has allowed the identification of conserved sequence and structural motifs in this type of enzymes from many eukaryotic species (1–4). The definition of the conserved sequence motifs has also aided identification of potential PAP coding sequences from bacterial species (4, 5). However, reports on structural, biochemical, and/or functional properties of the bacterial PAPs have not yet appeared. Evolutionarily, PAPs belong to the metallo-phosphoesterase family of proteins, members of which also include phosphoprotein phosphatases, diadenosine tetraphosphatases, exonucleases, 5'-nucleotidases, and other types of phosphomonoesterases (1, 6–12). The structure motif conserved in metallo-phosphoesterases is the  $\beta$ - $\alpha$ - $\beta$ - $\alpha$  fold (1, 13–16). The amino acid residues involved in coordinating the dimetal nuclear center (Fe(III)-M(II)) are located at the carboxyl ends of the parallel  $\beta$ -strands of the  $\beta$ - $\alpha$ - $\beta$ - $\alpha$  fold. The complement of amino acid residues involved in metal ligation in different types of metallo-phosphoesterases has recently been investigated based on the alignment of conserved sequence motifs. For PAPs, a comparison of multiple sequences from both eukaryotic and prokaryotic organisms has delineated seven invariant residues contained in five blocks of conserved amino acid sequences being the ones required for metal coordination (4, 5) (**DXG/GDXXY/GNH(D/E)/VXXH/GHXXH**; bold letters represent metal-ligating residues). For exonucleases and structurally related proteins, there are also seven potential metal-ligating residues that distribute in five blocks of conserved amino acid sequences (**DXH/GDXXX/GNH(D/E)/XX(G/A)H/(G/A)HXXH**) (1). For phosphoprotein phosphatases and structurally related proteins, there are only six metal-coordinating residues (**DXH/GDXXX/GNH(D/E)/XXXH/XXXH**) (1). It is clear that different metallo-phosphoesterases share similarities in both the number and composition of their metal-ligating residues. The purple color in the purified proteins of known PAPs is caused by a charge transfer transition at –560 nm from the metal-coordinating tyrosine residue to the metal ligand Fe(III).

Despite the presence of conserved structural and sequence motifs, PAPs from different, or even the same, species can differ from each other in the composition of their dimetal nuclear center and their overall structures. The composition of the dimetal nuclear center in mammalian PAPs is Fe(III)-Fe(II),

\* This work was supported by Grant J99-A-035 from the Ministry of Science and Technology of China and in part by a grant from the Max Planck Society of Germany. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF492659 (AtPAP7), AF492660 (AtPAP8), AF492661 (AtPAP9), AF492662 (AtPAP10), AF492663 (AtPAP11), AF492664 (AtPAP12), AF492665 (AtPAP13), AY090893 (AtPAP10 splice variant), and AY090894 (AtPAP13 splice variant).

¶ To whom correspondence should be addressed. Tel.: 8610-64889380; Fax: 8610-64854467; E-mail: dwwang@genetics.ac.cn or daowenwang@hotmail.com.

<sup>1</sup> The abbreviations used are: PAP, purple acid phosphatase; AGI, *Arabidopsis* genome initiative; AtPAP, *Arabidopsis* purple acid phosphatase; HMW, high molecular weight; KBPAP, kidney bean purple acid phosphatase; LMW, low molecular weight; RT, reverse transcription; ORF, open reading frame; TIGR, the Institute for Genomic Research.

whereas in plants it is either Fe(III)-Zn(II) (as in KBPAP and PAP from soybean) or Fe(III)-Mn(II) (as in one of the PAPs from sweet potato) (17–20). Structurally, the high molecular weight (HMW) plant PAPs (typified by KBPAP) possess two domains (1, 3). The NH<sub>2</sub>-domain does not have catalytic function. It is composed of two sandwiched  $\beta$ -sheets. By homology modeling, the structure formed by the two sandwiched  $\beta$ -sheets is later found to resemble the fibronectin type III domain commonly seen in animal proteins (21). The COOH-domain has the metal center and is the catalytic domain of the enzyme. It consists of two sandwiched mixed  $\beta$ -sheets that include two  $\beta$ - $\alpha$ - $\beta$ - $\alpha$ - $\beta$  motifs. In contrast, mammalian PAPs (typified by tartrate-resistant acid phosphatase and uteroferrin) have only one domain in their structure with an overall fold similar to that of the catalytic domain of KBPAP (13, 22, 23). The low molecular weight (LMW) plant PAPs have been described only recently, and are substantially smaller in size compared with HMW plant PAPs (4). Homology modeling shows that LMW plant PAPs lack the equivalent of the NH<sub>2</sub>-domain of KBPAP and are hence structurally similar to mammalian PAPs (4).

Biochemically, HMW plant PAPs function as homodimeric proteins with a molecular mass of ~55 kDa/monomer, whereas mammalian PAPs are typically monomeric proteins with a molecular mass of ~35 kDa (1–4, 13, 23). Many PAPs are glycoproteins that are targeted to the secretory pathway (4). One PAP from *Spirodela oligorrhiza* has been found to be glycosylphosphatidylinositol anchored in the cell (24). Another PAP from *Lupinus albus* may contain a third domain (with a structure resembling that of sterol desaturases) at the carboxyl terminus (25, 26). It is not known how common the latter two forms of posttranslational modification are in PAPs from other species.

In *in vitro* reactions, PAPs have been shown to catalyze the hydrolysis of activated phosphoric acid esters and anhydrides at a pH range of 4–7 (1). However, the *in vivo* function of PAPs remains largely unknown. Mammalian PAPs may play a variety of physiological roles, including resorption of bone and cartilage breakdown, iron transport, and generation of reactive oxygen species in a Fenton-like reaction (27–29). In higher plants, PAPs have mostly been studied for their potential involvement in phosphorus nutrition because their hydrolytic activity may aid the release of inorganic P (P<sub>i</sub>) from organic P that is not readily available to plant cells. Two PAPs (secreted PAP, membrane PAP) from *L. albus* have been shown to be highly inducible by phosphate deprivation (25, 26). The induction and secretion of secreted PAP are specific to proteoid roots, a structure that is specifically formed in *L. albus* as one of the adaptations to phosphorus deficiency, indicating that, at least in this species, PAPs may play an important part in phosphorus nutrition (26, 30). AtACP5, a PAP from *Arabidopsis thaliana*, is induced by both phosphate deprivation and oxidative stress (31). This enzyme may be involved in phosphate mobilization as well as the metabolism of reactive oxygen species in stressed or senescent plant tissues (31). The promoter of a separate *Arabidopsis* PAP gene has also been shown to be inducible by phosphate deprivation, although in this case the corresponding PAP enzyme has not been characterized (32). In addition, several *Arabidopsis* PAPs may possess phytase activity because their proteins show high levels of homology to a phytase recently reported from soybean (33). Given the structural and biochemical diversities described above, it is likely that PAPs from higher plants may have yet undiscovered functions, the elucidation of which may contribute significantly to the understanding of important aspects of plant biology.

Genetically, higher plant species often contain multiple genes coding for different PAPs. In sweet potato four PAPs

have been described (4, 18, 19, 34). In *Arabidopsis*, the AGI project has annotated 16 different PAP genes searchable at the Institute for Genomic Research (TIGR) *Arabidopsis* data base. The existence of multiple PAPs in the same species may hinder functional investigations because of potential genetic and functional redundancies. Additionally, studies on PAP may be complicated by the presence in higher plants of other types of acid phosphatases, which often react to environmental stimuli in ways similar to PAPs. For example, the LePS2 acid phosphatase is strongly induced by phosphate deprivation (35). However, molecular investigations show that LePS2 is related to a new class of phosphohydrolases rather than to PAPs (35). To study the function of PAPs in plant biology systematically, it is desirable to employ a plant species for which genetic knowledge on PAPs as well as other types of acid phosphatases is already available.

We have selected the model plant species *A. thaliana* in our studies on higher plant PAPs. A search of the annotated genome data base of *Arabidopsis* reveals that, in addition to PAP genes, this species also contains genes encoding several other types of phosphatases that may share some similarities with PAPs in reacting to physiological and environmental cues. These enzymes include histidine acid phosphatase (1 gene), vegetative storage protein type of acid phosphatases (10 genes), and phosphatidic acid phosphatase (4 genes). However, PAPs form by far the largest group of acid phosphatases in *Arabidopsis* with 16 different members in the TIGR data base. In this paper we report the finding of more *Arabidopsis* PAP members from data base searching, comparative analysis of conserved metal-ligating amino acid residues in 29 PAPs, and propose schemes for naming and classifying these PAPs. We also describe cDNA cloning and amino acid sequence analysis for seven PAPs encoded by chromosome 2. The patterns of primary structure variation in both LMW and HMW *Arabidopsis* PAPs are also reported. As an entry point to functional dissection, the response of the seven PAPs to phosphate deprivation in suspension cells is examined. Based on the work presented here, effective strategies for further studies of *Arabidopsis* PAPs are discussed.

#### EXPERIMENTAL PROCEDURES

**Data Base Search, Nomenclature, and Clustering Analysis of *Arabidopsis* PAPs**—Initially, a search using the phrase “purple acid phosphatase” was conducted at the TIGR web site ([www.tigr.org/tdb/e2k1/ath1/GeneNameSearch.shtml](http://www.tigr.org/tdb/e2k1/ath1/GeneNameSearch.shtml)). This search resulted in the retrieval of the predicted amino acid sequences of 16 different PAPs. In the second stage, the predicted protein sequences were each used as query sequences for Blastp searches at the TAIR *A. thaliana* genomic blast web page ([arabidopsis.org/Blast/index.html](http://arabidopsis.org/Blast/index.html)) using the default setting. Identification of additional predicted PAPs in the Blastp search was based on two criteria: at least 20% identical to the query sequence and the possession of amino acid sequence elements that could be related to the conserved sequence motifs (DXG/GDXXY/GNH(E/D)/VX<sub>2</sub>H/GHXH) defined previously for PAPs. Thirteen additional predicted PAPs were found in this step. In the third stage, the predicted protein sequences of the newly retrieved 13 PAPs were each used as query sequences for another round of Blastp search at the TAIR web site. However, no more predicted PAPs were found, suggesting that the total number of predicted PAPs in the *Arabidopsis* genome was likely to be 29. To identify expressed sequence tags and/or cDNA sequences for the 29 PAPs, several nucleic acid and protein data bases, including the TIGR *Arabidopsis* Gene Index ([www.tigr.org/tdb/agil](http://www.tigr.org/tdb/agil), this site also contains full-length cDNA sequences for 5000 *Arabidopsis* genes), MIPS ([mips.gsf.de/proj/thal/db/index.html](http://mips.gsf.de/proj/thal/db/index.html)), and GenBank<sup>TM</sup> ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), were searched. For naming the 29 *Arabidopsis* PAPs according to their positions on the different chromosomes, the prefix “AtPAP” (for *A. thaliana* PAP) was used.

To establish a classification scheme for *Arabidopsis* PAPs, clustering analysis of amino acid sequences was conducted. It is known from previous work and from our study (see “Results”) that the predicted protein sequences of some *Arabidopsis* genes contained localized errors



because of imprecise prediction of intron and exon boundaries during the annotation process. To improve the accuracy of the clustering analysis, two measures were taken. First, for the seven PAPs (AtPAP7–AtPAP13; Table III) investigated in this study and the three PAPs (AtPAP3, AtPAP17, AtPAP18; Table III) studied by previous researchers, amino acid sequences derived from cDNAs were used. Consequently, in only 19 of the 29 *Arabidopsis* PAPs, the predicted amino acid sequences were employed for the clustering analysis. Second, during the analysis, the “complete deletion option” was chosen with respect to sites involved in alignment gaps. The incorrectly predicted amino acid sites would cause gaps in multiple alignment of sequences. The adoption of the complete deletion option may minimize the impact of the prediction errors on the clustering analysis (37). The protein sequences of 29 PAPs were aligned using the program ClustalW 1.8 (searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) (36). The alignment result was converted to MEGA format, which was then subject to clustering analysis using several different programs (neighbor joining, minimum evolution, and parsimony) at the MEGA 2 web site (www.megasoftware.net/) (37).

**Plant Materials, Phosphate Deprivation Treatment, and Detection of AtPAP cDNAs**—The Col-1 ecotype of *A. thaliana* was used throughout this study. General conditions for plant growth and initiation and maintenance of suspension cell cultures were the same as those described previously (38, 39). For phosphate deprivation of suspension cells, cells grown in phosphate-sufficient liquid medium (1 mM  $P_i$ , in the form of  $NaH_2PO_4$ ) were collected by vacuum, washed three times using sterile distilled water, and dispersed into the low phosphate liquid medium (0.01 mM  $P_i$ ). A cell sample was immediately taken as the day 0 control. Further samples were collected at 1, 3, and 5 days later. Control cell samples were also taken at identical points from suspension cells grown in the phosphate-sufficient medium. For phosphate deprivation of *Arabidopsis* seedlings, the method described by Haran *et al.* (32) was adopted. All cell and tissue samples were stored at  $-70^\circ C$  prior to RNA extraction.

Total RNA was extracted from cell and tissue samples using the RNeasy Plant Mini Kit (Qiagen). RNA concentrations were determined with the aid of a spectrophotometer (BioPhotometer, Eppendorf AG). Using nucleotide sequence information of the predicted PAPs, PCR primers were synthesized for amplifying cDNAs corresponding to the coding regions of 20 of the 29 predicted PAPs. A complete list of the PCR primers used in this study is available at the web site www.mpimgolm.mpg.de/udvardi/(under section “Supplementary Materials for Publications”). The primers used for amplifying coding region cDNAs of the seven PAPs encoded by chromosome 2 are listed in Table I. In each reverse transcription reaction using the Moloney murine leukemia virus reverse transcriptase (Invitrogen) and random hexamers (New England Biolabs), 10  $\mu g$  of total RNA was used. For amplifying cDNAs of individual PAPs, varying amounts of the reverse transcription mix (1–3  $\mu l$ ) were used in 50- $\mu l$  PCR reactions containing PAP-specific primers and the high fidelity *Taq* polymerase ExTaq (TaKaRa). The annealing temperature (48–55  $^\circ C$ ) and the extension time (1–3 min) were adjusted to suit the amplification of the coding region of individual PAPs. After 35 cycles of amplification, the PCR products were examined by agarose gel electrophoresis with ethidium bromide staining. The PCR fragments of expected size were purified from the gel, followed by cloning into the plasmid vector pGEMT-easy (Promega). The inserts in selected positive clones were sequenced commercially (TaKaRa). The resulting cDNA sequences were compared with the predicted ORFs and previously reported cDNAs of *Arabidopsis* PAPs using various programs at the NCBI network (www.ncbi.nlm.nih.gov/).

**Analysis of Splice Variants**—For each of the seven PAPs (AtPAP7–AtPAP13) that were analyzed in more detail in this study, cDNA sequences were determined using several clones, some of which contained inserts of different sizes. This led to the identification of splice variants (based on the presence of incompletely removed intron sequences in the cDNAs) for two of the seven PAPs (AtPAP10 and AtPAP13). To investigate whether the splice variant of AtPAP10 was present in the cytoplasm and could access cellular translation apparatus, polyribosomes were prepared from suspension cells grown in either phosphate-sufficient or phosphate-limiting medium (with 1 mM and 0.01 mM  $P_i$ , respectively) using sucrose density gradient centrifugation detailed in published protocols (40, 41). In total, 12 fractions were collected from each sucrose gradient. Total RNA was extracted from each fraction and was further treated with RNase-free DNase (TaKaRa) to eliminate potential contamination by genomic DNA. In analogy to the study by Petracek *et al.* (41), monoribosomes were associated with fractions 4 and 5, whereas polyribosomes were contained in fractions 6–11. To detect the distribution of wild type transcript and the splice variant of AtPAP10 in frac-

TABLE I

## A list of PCR primers

The underlined nucleotides form *Bam*HI (GGATCC), *Kpn*I (GGTACC), *Nde*I (CATATG), or *Sal*I (GTCGAC) restriction site.

| AGI code                    | Primer           | For cDNA amplification  |
|-----------------------------|------------------|---|
| At2g01880                   | 01880F<br>01880R | 5' CGACCATGTTCTTCTTATGGATCT 3'<br>5' GGTTCAGGTAATTATGCCCATGCCT 3'                 |
| At2g01890                   | 01890F<br>01890R | 5' CTCCAAAGTCTCTCAATGGATAGC 3'<br>5' GTCTGACCATCGATTAGATATCGG 3'                  |
| At2g03450                   | 03450F<br>03450R | 5' <u>ACGTCGAC</u> TTTTTGACTAACCAACAT 3'<br>5' <u>GGTCGAC</u> AAATCTGTGTGGCAAC 3' |
| At2g16430                   | 16430F<br>16430R | 5' TCAGGATCCAAAAATGGGTCGTG 3'<br>5' <u>TCCGGTAC</u> CTCATAGATTCTCAAG 3'           |
| At2g18130                   | 18130F<br>18130R | 5' <u>AACCATAT</u> GGAGTTGAGTCAC 3'<br>5' <u>ACTGGATC</u> CTAAGCAGTCTCTCC 3'      |
| At2g27190                   | 27190F<br>27190R | 5' <u>GCAGGATC</u> CTCGTCTCGGATTCAG 3'<br>5' <u>CCAGGTAC</u> CACAGTCACTAATCGA 3'  |
| At2g32770                   | 32770F<br>32770R | 5' <u>CTTCTCG</u> CTGCAACAAACAT 3'<br>5' CTACTTGGTTAGTTGCAACCGCA 3'               |
| For splice variant analysis |                  |   |
| WTF                         |                  | 5' ATACTTCTGGTTTATCCATC 3'  |
| SVF                         |                  | 5' TTGTGTAGCTGATGTTGTG 3'   |
| Tubulin                     |                  |   |
| TuF                         |                  | 5' GAGGGAATATGGCCGTTTAGG 3'   |
| TuR                         |                  | 5' CACTTCACCGACCAATCAATGG 3'  |

tions 4–11 by RT-PCR, 5  $\mu g$  of total RNA from each fraction was reverse-transcribed as described above. The resultant cDNA mixtures were used for PCR amplification of wild type transcripts (using the primers WTF and 16430R; Table I, Fig. 3A) and splice variants (using the primers SVF and 16430R; Table I, Fig. 3A). The amplified products were checked by agarose gel electrophoresis with ethidium bromide staining.

**Phylogenetic Investigations and Primary Structure Comparisons**—Using amino acid sequences deduced from cDNA clones, phylogenetic relationships of AtPAP7–AtPAP13 to PAPs isolated from other eukaryotic organisms were investigated. PAPs from bacterial species were not included in this investigation because of some uncertainties in potential metal-ligating residues (4). Multiple sequence alignment was carried out using ClustalW 1.8. Phylogenetic trees were constructed using several tree building methods (neighbor joining, minimum evolution, parsimony) available at the MEGA 2 web site (37). For comparisons of primary structure, the amino acid sequences of LMW PAPs (AtPAP7, AtPAP8) were compared with that of uteroferrin (as the representative of known LMW PAPs), whereas the amino acid sequences of HMW PAPs (AtPAP9–AtPAP13) were compared with that of KBPAP (as the representative of known HMW PAPs).

**Evaluation of Transcriptional Responses of AtPAPs to Phosphate Deprivation Treatment using Semiquantitative RT-PCR**—The transcriptional responses of AtPAP7–AtPAP13 to phosphate deprivation treatment were evaluated using suspension cells. During the experiments, suspension cells were subject to phosphate deprivation as described above, followed by the extraction of total RNA from both treated and control samples. For all samples to be compared, the same amount of total RNA (10  $\mu g$ ) was used in reverse transcription reactions. Prior to PCR amplification of *Arabidopsis* PAPs, the cDNA content of all reverse transcription reactions was normalized by amplifying the transcript of tubulin using primers TuF and TuR (Table I). In amplifying the transcripts of AtPAP7–AtPAP12 using normalized cDNA samples, the same set of cycling parameters (annealing temperature, 52  $^\circ C$ ; extension time, 2 min; number of amplification cycles; Ref. 28) was employed. In amplifying the transcripts of AtPAP13 using the same batch of normalized cDNA samples (and with annealing temperature of 52  $^\circ C$  and extension time of 2 min), the number of amplification cycles had to be increased to 32 to obtain consistent results. The kinetics of the PCR amplifications was checked by amplifying tubulin transcripts using the same annealing temperature and extension time but with different numbers of amplification cycles (24, 28, 32) (Fig. 6). For comparing the levels of AtPAP transcripts across the samples, the same amount of reaction mixture was taken from all PCR amplifications, followed by agarose gel electrophoresis and ethidium bromide staining. The results were imaged using the gel documentation system AlphaImager 2200 (Alpha Innotech Corp.).

## RESULTS

**Comparative Analysis of Potential Metal-ligating Residues in *Arabidopsis* PAPs**—Repeated Blastp searches of the protein

TABLE II  
Conserved sequence motifs containing potential metal-ligating residues in Arabidopsis PAPs

The symbols plus (+) and minus (−) in this table indicate positive and negative scores, respectively. EST, expressed sequence tag.

| PAP <sup>a</sup> ,<br>Chromosome<br>location       | BAC locus <sup>b</sup> | GenBank<br>accession | Motif             |       |          |      |      |
|--|------------------------|----------------------|-------------------|-------|----------|------|------|
|  |                        |                      | I                 | II    | III      | IV   | V    |
| At1g13750, 1                                       | F21F23.18              | AC027656             | GDMG              | GDICY | GNHE     | FLAH | GHAH |
| At1g13900, 1                                       | F16A14.11              | AC068197             | GDMG              | GDISY | GNHE     | VQGH | GHVH |
| At1g14700, 1                                       | F10B6.10               | AC006917             | GDWG              | GDNFY | GNHD     | VIGH | GHDH |
| At1g25230, 1                                       | F4F7.38                | AC079374             | GDWG              | GDNFY | GNHD     | VVGH | GHDH |
| At1g52940, 1                                       | F14G24.21              | AC019018             | GDLG              | GDLSY | GNHE     | VLVH | GHVH |
| At1g56360, 1                                       | F14G9.2                | AC058785             | GDLG              | GDLSY | GNHE     | VLVH | GHVH |
| At2g01880 <sup>c</sup> , 2                         | T23K3.7                | AC007069             | GDWG              | GDNFY | GNHD     | VVGH | GHDH |
| At2g01890 <sup>c</sup> , 2                         | T23K3.8                | AC007069             | GDWG              | GDNFY | GNHD     | VVGH | GHDH |
| At2g03450, 2                                       | T4M8.12                | AC006284             | GDMG              | GDISY | GNHE     | VQGH | GHVH |
| At2g16430, 2                                       | F16F14.7               | AC079374             | GDLG              | GDISY | GNHE     | VLMH | GHVH |
| At2g18130, 2                                       | F8D23.9                | AC007212             | SDLG              | GDLSY | GNHE     | VLVH | GHVH |
| At2g27190, 2                                       | T22O13.4               | AC007290             | GDLG              | GDLSY | GNHE     | VLVH | GHVH |
| At2g32770 <sup>c</sup> , 2                         | F24L7.9                | AC003974             | GDLG              | GGESY | GEHE     | ATWS | SHVD |
| At2g46880 <sup>c</sup> , 2                         | F14M4.29               | AC004411             | ADMH              | GDNFY | GNHD     | VYLH | GHDH |
| At3g07130, 3                                       | T1B9.21                | AC012395             | GDLG              | GDVSY | GNHE     | ASWH | GHVH |
| At3g10150 <sup>c</sup> , 3                         | T22K18.2               | AC010927             | ADMH              | GDVSY | GNHD     | JEWH | GHDH |
| At3g17790, 3                                       | MEB5.10                | AB019230             | GDWG              | GDNFY | GNHD     | VVGH | GHDH |
| At3g20500, 3                                       | K10D20.4               | AB000410             | GDLG              | GDLSY | GNHE     | VLFH | GHVH |
| At3g46120, 3                                       | F12M12.90              | AL355775             | GDLG <sup>d</sup> | GDLSY | GNHE     | VLVH | GHVH |
| At3g52780, 3                                       | F3C22.180              | AL353912             | GDLG              | GDLSY | GNHE     | AVVH | GHVH |
| At3g52810, 3                                       | F3C22.210              | AL353912             | GDLG              | GDLSY | GNHE     | VVMH | GHVH |
| At3g52820, 3                                       | F3C22.220              | AL353912             | GDLG              | GDLSY | GNHE     | VLLH | GHVH |
| At4g13700, 4                                       | F18A5.90               | AL035528             | GDLG              | GDLTY | GNHE     | ATMH | GHVH |
| At4g24890, 4                                       | F13M23.30              | AL035523             | GDMG              | GDLTY | GNHE     | FLAH | GHVH |
| At4g36350, 4                                       | F23E13.190             | AL022141             | GDLG              | GDLSY | GNHE     | VMVH | GHVH |
| At5g34850, 5                                       | T5E15.10               | AC019013             | GDMG              | GDLSY | GNHE     | VLMH | GHVH |
| At5g50400, 5                                       | MXI22.12               | NC003076             | GDMG              | GDITY | GNHE     | FLAH | GHVH |
| At5g57140 <sup>c</sup> , 5                         | MUL3.9                 | NC003076             | ADMH              | GDNIE | GNHD     | AFEH | GHDH |
| At5g63140 <sup>c</sup> , 5                         | MDG12.10               | NC003076             | ADMH              | GDNIE | GNHD     | AYFH | GHDH |
| Consensus defined previously for PAPs <sup>e</sup> |                        |                      | DxG               | GDXSY | GNH(E,D) | VxxH | GHyH |

<sup>a</sup> Represented by the AGI code.

<sup>b</sup> Bacterial artificial chromosome (BAC) locus number is obtained from the TIGR data base.

<sup>c</sup> The predicted protein sequence differ from the one deduced from cDNA analysis.

<sup>d</sup> This element is not present in AGI predicted protein sequence owing to error in coding sequence prediction. Correction of the prediction error allows the identification of the GDLG element in the derived amino acid sequence.

<sup>e</sup> Defined by Schenk *et al.* (4). The bold letters represent metal-ligating residues.

<sup>f</sup> The spectrum of potential metal ligating residues in these PAPs differ to some extent from that of typical PAPs.

data base of *Arabidopsis* led to the finding of 29 potential PAPs (Table II). In 24 *Arabidopsis* PAPs, the complete set of seven invariant amino acid residues involved in the ligation of the dimetal nuclear center in known PAPs was found (Table II). However, in 5 of the 29 PAPs (Table II, shaded lines), the varieties of potential metal-ligating residues differed from those in typical PAPs. In At2g32770 (AtPAP13 in Table III), the lack of four of the seven invariant residues was caused by drastic changes of amino acid compositions in three of the five sequence elements (Table II), which were deduced to hold positions for metal ligation residues by amino acid sequence comparisons. In At2g46880, At3g10150, At5g57140, and At5g63140 (AtPAP14, AtPAP16, AtPAP28, and AtPAP29, respectively, in Table III), the varieties of potential metal-ligating residues were generally more similar to those found in several other types of metallo-phosphoesterases from yeast, bacterium, and human (Table II, Fig. 1) (1), some of which had been found to possess exonuclease or phosphodiesterase activities (42–44). However, they were still considered as potential PAPs because their overall amino acid sequences exhibited significant levels of homology to known PAPs. As shown in Fig. 1, proteins homologous to At2g46880, At3g10150, At5g57140,

and At5g63140 had also been found in two other plant species (chickpea and rice).

**Evidence for Expression, Nomenclature, and Classification of Arabidopsis PAPs**—The above data base search results raised the question of whether the multiple *Arabidopsis* PAPs were expressed. Several lines of evidence presented in Table III indicated that the transcripts for the majority of the 29 PAPs were present in *Arabidopsis* cells. First, in various expressed sequence tag projects, expressed sequence tags were discovered for 13 PAPs (Table III). Second, in public data bases, full-length cDNA clones were isolated and sequenced for five PAP genes (Table III). Third, in our study, we had detected the cDNAs for 20 PAPs (Table III). From Table III, it could also be seen that previous investigators had tentatively named *Arabidopsis* PAPs in different manners. Based on the finding that the great majority of the 29 predicted PAPs were expressed, and the need in future to carry out more systematic studies on individual PAPs, a system for naming *Arabidopsis* PAPs is presented in Table III. This nomenclature is consistent with the current system for naming members of multigene families in *A. thaliana*.

Using amino acid sequences of 19 predicted PAPs and those of 10 PAPs (AtPAP3, AtPAP7–AtPAP13, AtPAP17, AtPAP18) derived from cDNA analysis, clustering analysis was conducted with the aim to establish a classification scheme for *Arabidopsis* PAPs. The patterns in the clustering of the 29 PAPs using different methods were essentially similar. The result obtained with the minimum evolution method (Fig. 2) was adopted as the basis for a classification scheme of *Arabidopsis* PAPs. The 29 PAPs could be classified into three major groups (groups I, II, and III), each with more than 95% bootstrap support (Fig. 2). A further division of the three major groups yielded eight subgroups (Ia-1, Ia-2, Ib-1, Ib-2, IIa, IIb, IIIa, and IIIb; Fig. 2). Except for Ia-2, statistical support for all other subgroups was equal or greater than 95% (Fig. 2). There was a general correlation between the classification of the groups and the number of amino acid residues in the 29 PAPs (Fig. 2). Thus, most of proteins in group I contain more than 400 amino acid residues, all proteins in group II are composed of more than 500 residues, and all proteins in group III comprise fewer than 400 residues (Fig. 2). AtPAP13, for which four of the seven potential metal-ligating residues could not be identified, clustered with AtPAP15 and AtPAP23, creating the subgroup Ib-1 (Fig. 2). AtPAPs 14, 16, 28, and 29 formed the subgroup IIIa (Fig. 2). Proteins in the remaining subgroups all contained the seven invariant amino acid residues typical of known PAPs.

**Characterization of Coding Region and Amino Acid Sequences of AtPAP7–AtPAP13**—In our long term effort to address the function of PAPs in the biology of *Arabidopsis*, we focused initially on seven PAPs (AtPAP7–AtPAP13) encoded by PAP genes on chromosome 2. With four members in group I, one member in group II, and two members in group III, the seven PAPs represent many of the structural and functional diversities in *Arabidopsis* PAPs. In RT-PCR experiments, the cDNAs for AtPAP7, AtPAP8, AtPAP9, AtPAP10, AtPAP12, and AtPAP13 were readily amplified using total RNA samples prepared from *Arabidopsis* cells or tissues grown under four different conditions ([www.mpimp-golm.mpg.de/udvardi/zusaetzlich/index-e.html](http://www.mpimp-golm.mpg.de/udvardi/zusaetzlich/index-e.html)). In contrast, the cDNA for AtPAP11 was only obtained from using RNA samples extracted from phosphate-deprived cells or root tissues, suggesting that the expression of this PAP was regulated by phosphate availability. In cDNA sequencing experiments, clones containing fully processed, wild type ORFs were identified for each of the seven PAPs. In addition, cDNA clones harboring variant ORFs retaining one or more introns sequences were also obtained for AtPAP10 and AtPAP13 (see below).

TABLE III  
Evidence for expression and nomenclature of Arabidopsis PAPs

| AGI code  | Expressed sequence tag | Full-length cDNA sequence in public data base |                      | Detection of cDNA in this work <sup>b</sup> | Previous name                         | Proposed name |
|-----------|------------------------|---|----------------------|---|---------------------------------------|---------------|
|           |                        | Ceres, Inc.                                   | GenBank <sup>a</sup> |   |                                       |               |
| Atlg13750 | +                      | —   | —                    | +   |                                       | AtPAP1        |
| Atlg13900 | —                      | —   | —                    | —   |                                       | AtPAP2        |
| Atlg14700 | +                      | —   | + (AY070051)         | +   |                                       | AtPAP3        |
| Atlg25230 | +                      | —   | —                    | +   |                                       | AtPAP4        |
| Atlg52940 | —                      | —   | —                    | —   | At9 <sup>c</sup>                      | AtPAP5        |
| Atlg56360 | —                      | —   | —                    | —   |                                       | AtPAP6        |
| At2g01880 | +                      | —   | —                    | +   |                                       | AtPAP7        |
| At2g01890 | +                      | —   | + (AY065434)         | +   |                                       | AtPAP8        |
| At2g03450 | —                      | —   | —                    | +   |                                       | AtPAP9        |
| At2g16430 | +                      | —   | —                    | +   | At11 <sup>c</sup>                     | AtPAP10       |
| At2g18130 | —                      | —   | —                    | +   | At8 <sup>c</sup>                      | AtPAP11       |
| At2g27190 | +                      | —   | +                    | +   | Ath1 <sup>d</sup> , At10 <sup>c</sup> | AtPAP12       |
| At2g32770 | —                      | —   | —                    | +   | At2 <sup>c</sup>                      | AtPAP13       |
| At2g46880 | —                      | —   | —                    | —   |                                       | AtPAP14       |
| At3g07130 | —                      | —   | —                    | +   | At1 <sup>c</sup>                      | AtPAP15       |
| At3g10150 | —                      | —   | —                    | —   |                                       | AtPAP16       |
| At3g17790 | +                      | +   | + (AJ133747)         | +   | AtACP5 <sup>e</sup>                   | AtPAP17       |
| At3g20500 | +                      | —   | + (AY062488)         | +   |                                       | AtPAP18       |
| At3g46120 | —                      | —   | —                    | —   |                                       | AtPAP19       |
| At3g52780 | +                      | —   | —                    | +   | At4 <sup>c</sup>                      | AtPAP20       |
| At3g52810 | —                      | —   | —                    | +   | At5 <sup>c</sup>                      | AtPAP21       |
| At3g52820 | +                      | —   | —                    | +   | At6 <sup>c</sup>                      | AtPAP22       |
| At4g13700 | —                      | —   | —                    | +   | At3 <sup>c</sup>                      | AtPAP23       |
| At4g24890 | —                      | —   | —                    | —   |                                       | AtPAP24       |
| At4g36350 | —                      | —   | —                    | +   | Ath2 <sup>d</sup> , At7 <sup>c</sup>  | AtPAP25       |
| At5g34850 | +                      | —   | —                    | +   |                                       | AtPAP26       |
| At5g50400 | —                      | —   | —                    | —   |                                       | AtPAP27       |
| At5g57140 | +                      | —   | —                    | +   |                                       | AtPAP28       |
| At5g63140 | —                      | —   | —                    | —   |                                       | AtPAP29       |

<sup>a</sup> GenBank<sup>TM</sup> accession number of the cDNA clone is in parentheses.

<sup>b</sup> More details given under "Experimental Procedures" and "Results."

<sup>c</sup> Described by Hegeman and Grabau (33).

<sup>d</sup> Described by Schenk *et al.* (4).

<sup>e</sup> Investigated by del Pozo *et al.* (31).

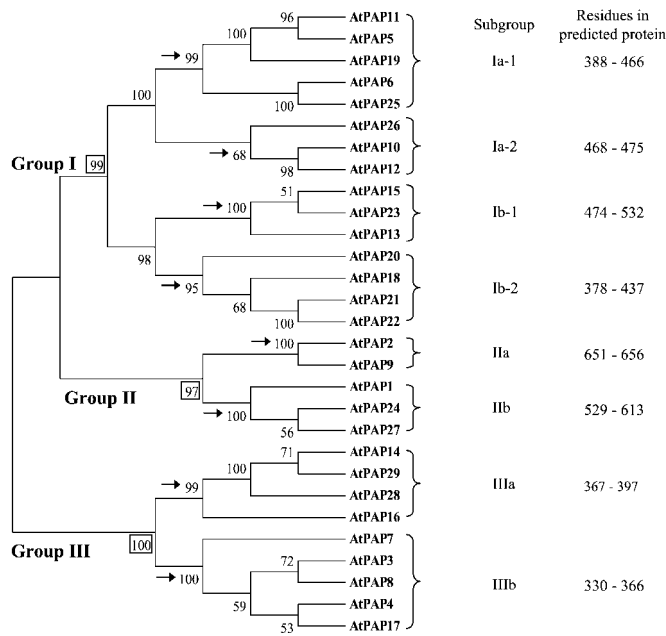
|                    |     |              |    |                |    |               |     |              |    |               |
|--------------------|-----|--------------|----|----------------|----|---------------|-----|--------------|----|---------------|
| <b>At2g46880</b>   | 48  | ILQVSDMHYGF  | 36 | IVFSGDNVYGLCE  | 22 | VAILGNHDDQESD | 106 | APGLVYLHIMP  | 39 | GVFSGHDHVNDF  |
| <b>At3g10150</b>   | 37  | IAIFADLHFGED | 28 | VVYLGDVVTANNI  | 23 | ATLFGNHDDASF  | 115 | IPELIFWHIPSK | 44 | AVFVGHNHGLDW  |
| <b>At5g57140</b>   | 60  | ILQVADMHFGMG | 36 | IAFTGDNIFGSST  | 21 | AAVLGNHDDHST  | 105 | PPALAFFHPIPL | 38 | AAFVGHDHVNDF  |
| <b>At5g63140</b>   | 49  | ILQVADMHFANG | 36 | IVFTGDNIFGFVD  | 21 | VAILGNHDDQEST | 104 | APGLAYFHPIPL | 39 | SVFVGHDHVNDF  |
| <b>CaPAP</b>       | 59  | ILQVADMHYADG | 36 | IVFTGDNIFGFDS  | 21 | VAVLGNHDDQEGS | 103 | APGLAYFHPIPL | 40 | AVFTGHDHVNDF  |
| <b>OsPAP</b>       | 53  | VALFADLHYGEN | 28 | VVYLGDLVTANNL  | 23 | ATVFGNHDDMPF  | 114 | IPEIIFWHIPST | 46 | AIFVGHNHGLDW  |
|                    |     |              |    |                |    |               |     |              |    |               |
| <b>YLR361c</b>     | 250 | IVQLADLHLGVG | 35 | VVFTGDIQIMGDRS | 22 | AWVWGNHDDDEGS | 93  | GLSMAFFHPIPL | 44 | VVSCGHDHNCNDY |
| <b>SPCC1020.05</b> | 210 | ILQLSDLHYSNS | 37 | VLLTGDLINGDTS  | 21 | AVNFGNHDDLGD  | 78  | PIQMAVLHIPLK | 35 | LAIAGHDHVNDF  |
|                    |     |              |    |                |    |               |     |              |    |               |
| <b>EcSbcd</b>      | 3   | ILHTSDWHLGQN | 28 | IIVAGDVFDTGSP  | 22 | VVLGNHDSVAT   | 87  | LPPIATGHLTTV | 28 | YIALGHIHRAQI  |
| <b>CaPase1</b>     | 46  | IVQFADLHQND  | 20 | VILTGDNIDGRYC  | 23 | AAVLGNHDTESL  | 83  | VPAFMFHPIPII | 39 | AIFVGHDHNTNF  |
| <b>EsIcc</b>       | 17  | ILQITDTHLFAQ | 30 | IVATGDLAQDQSS  | 17 | VWLPGNHDFQPA  | 56  | RHTLLLLHHHPL | 29 | YLLCGHIHQELD  |
|                    |     |              |    |                |    |               |     |              |    |               |
| <b>Asm</b>         | 201 | ILFLTDLHWDHD | 60 | VYWTGDIPIAHVDV | 27 | YPAVGNHESIPV  | 93  | DKVHIIGHIPPG | 22 | AQFFGHTHVDEF  |

FIG. 1. Comparison of potential metal-ligating residues in At2g46880, At3g10150, At5g57140, and At5g63140 and related proteins from plant, yeast, bacterium, and human. The nucleotides conserved in all sequences are boxed. The potential metal-ligating residues are marked by asterisks. The number at the beginning of each sequence indicates amino acid position; the number in between the sequence elements represents the sum of residues separating two adjacent blocks. Accession numbers for related proteins from plant, yeast, bacterium, and human are AJ276267 (*CaPAP*, putative PTS protein of *Cicer arietinum*), AP004031 (*OsPAP*, hypothetical protein of *Oryza sativa*), NC\_001144 (*YLR361c*, probable membrane protein of *Saccharomyces cerevisiae*), AL023518 (*SPCC1020.05*, hypothetical protein of *Schizosaccharomyces pombe*), P13457 (*EcSbcd*, exonuclease SBCD of *Escherichia coli*), NC\_003030 (*CaPase1*, predicted Icc-like phosphohydrolase of *Clostridium acetobutylicum*), D16557 (*EsIcc*, cAMP phosphodiesterase of *E. coli*), and P17405 (*Asm*, sphingomyelin phosphodiesterase of *Homo sapiens*).

Comparisons of the coding sequences and derived amino acid sequences obtained for AtPAP7–AtPAP13 in this study with those predicted by AGI and those deposited in public data bases by previous investigators on PAP cDNAs isolated from Col-1 ecotype gave the following results. First, for AtPAPs 9–13, the coding sequences and derived amino acid sequences from our study were identical to those predicted by AGI. The coding sequence and derived amino acid sequence of AtPAP12 from

this study differed from those by Cheuk and colleagues (GenBank<sup>TM</sup> accession no. AY065067) for the same PAP by a single nucleotide and amino acid change, respectively. This minor difference may have been caused by PCR or sequencing error. Second, for AtPAPs 7 and 8, the coding sequences determined by us and those predicted by AGI differed substantially. Comparisons among the coding sequences determined from cDNA analysis, the predicted coding sequences and their correspond-

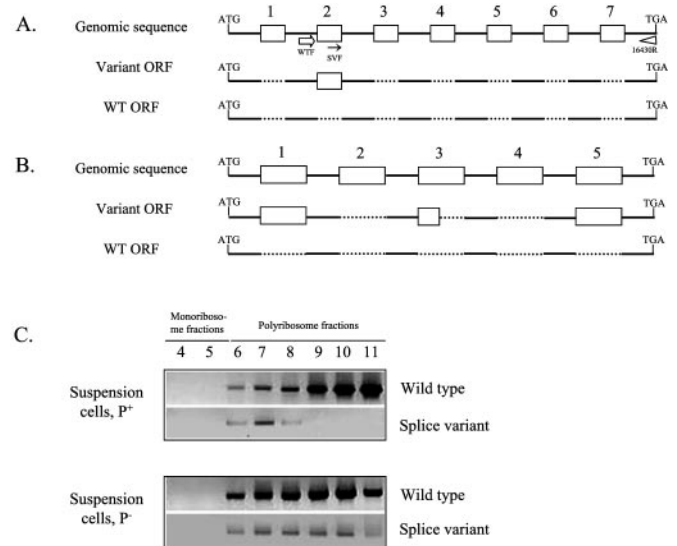




**FIG. 2. A classification scheme for *Arabidopsis* PAPs based on clustering analysis of amino acid sequences.** The clustering analysis used amino acid sequences of 19 predicted PAPs and those of 10 PAPs (AtPAP3, AtPAP7–AtPAP13, AtPAP17, AtPAP18) derived from cDNA analysis. The analysis, from which the classification scheme was derived, was conducted using the minimum evolution method (with PC distance and complete deletion options). Bootstrap values are percentages of 500 replications. The main groups (groups I, II, and III) are further divided to yield the eight subgroups (second column). The bootstrap values for the three main groups are boxed, whereas those for the eight subgroups are indicated by arrows. The ranges of amino acid residues in the predicted proteins of the different subgroups are listed in the third column.

ing genomic sequences showed that the difference between predicted and cDNA-derived coding regions may have been caused by error in predicting intron-exon boundaries in genomic sequences (data not shown). The cDNA-derived coding sequence of AtPAP8 also varied from the one by Yamada and colleagues (GenBank<sup>TM</sup> accession no. AY065434) and the one by Schenk *et al.* (4) for the same PAP (www.mpimp-golm.mpg.de/udvardi/zusaetzlich/index-e.html). Schenk *et al.* (4) determined the cDNA sequence for an *Arabidopsis* PAP named as SmAth. In nucleotide sequence comparison, the first 780 nucleotides of SmAth coding sequence were identical to those of the AtPAP8 coding region determined by us. Intriguingly, however, at the 3' end of the coding region sequence, SmAth differed in multiple ways (deletions involving single or more nucleotides, and insertions involving single nucleotides) from AtPAP8 coding region determined by us or Yamada and colleagues.

**Association of Splice Variants of AtPAP10 with Polyribosomes**—The intron-containing, variant ORF of AtPAP10 and AtPAP13 was analyzed in more detail in terms of nucleotide sequence and structure. The genomic DNA sequence encoding AtPAP10 consisted of eight exons and seven introns (Fig. 3A). The complete sequence of the variant ORF of AtPAP10 showed that the second intron was retained in the cDNA (Fig. 3A). In contrast, all intron sequences were removed from the cDNA containing the fully processed, wild type ORF of AtPAP10 (Fig. 3A). Similar analysis revealed that the variant cDNA of AtPAP13 was also caused by the presence of intron sequences. In this case, two intact introns (the first and fifth intron) and the 5' half of the third intron were retained in the cDNA (Fig. 3B). Conceptual translation of the variant ORF of AtPAP10 yielded a hypothetical protein of 348 amino acids. The amino acid



**FIG. 3. Analysis of splice variants of AtPAP10 and AtPAP13.** A, a diagram illustrating the structure of the genomic sequence of AtPAP10 ORF, the cDNA sequence of AtPAP10 variant ORF, and the cDNA sequence of AtPAP10 wild type (WT) ORF. Introns are represented by boxes, exons by filled horizontal lines. The dashed lines indicate the removal of intron sequences. ATG is start codon, TGA the stop codon. The primer combination of WTF and 16430R was used for detecting wild type transcript, whereas that of SVF and 16430R was used for detecting splice variant. Diagrams are not drawn to scale. B, a diagram illustrating the structure of the genomic sequence of AtPAP13 ORF, the cDNA sequence of AtPAP13 variant ORF, and the cDNA sequence of AtPAP13 wild type ORF. The representation of different sequence elements was the same as described in A. C, association of wild type transcripts and splice variants of AtPAP10 with polyribosomes present in the different fractions of a sucrose gradient. The polyribosomes were prepared from suspension cells grown in phosphate-sufficient (top panel) or phosphate-limiting (bottom panel) medium.

sequence of this hypothetical protein was completely homologous to that of the catalytic domain of the putative wild type AtPAP10 protein, but lacked the amino-terminal 120 amino acid residues of the wild type protein. An attempted translation of the variant ORF of AtPAP13 produced a hypothetical protein of 428 amino acids. It differed from the putative wild type AtPAP13 protein by one deletion (of 81 amino acid residues) at the amino-terminal region, one deletion (of 37 amino acid residues) at the carboxyl-terminal region, and one insertion (29 amino acid residues) approximately in the middle region of the protein.

The above results indicated that the transcription of the genes coding for AtPAP10 and AtPAP13 can produce wild type (fully processed) transcripts as well as splice variants in which intronic sequences were not completely removed. The splice variants may be resided in either nucleus or cytoplasm. If nuclear in origin, the splice variants may have been intermediates of the splice reaction. If cytoplasmic, the splice variants may serve important functions. To distinguish the two possibilities, experiments were undertaken to determine whether splice variants were associated with polyribosomes located in the cytoplasm. Following well established protocols that entailed the separation of polyribosomes from monoribosomes through the use of sucrose density gradient centrifugation (40, 41), polyribosome fractions were prepared from suspension cells grown in either phosphate-sufficient or phosphate-limiting medium. The distribution of wild type transcripts and splice variants of AtPAP10 in the different fractions of the sucrose gradients was detected by RT-PCR. The oligonucleotide primers for the PCR had been designed for selective amplification of either wild type transcripts or splice variants (Fig. 3A).

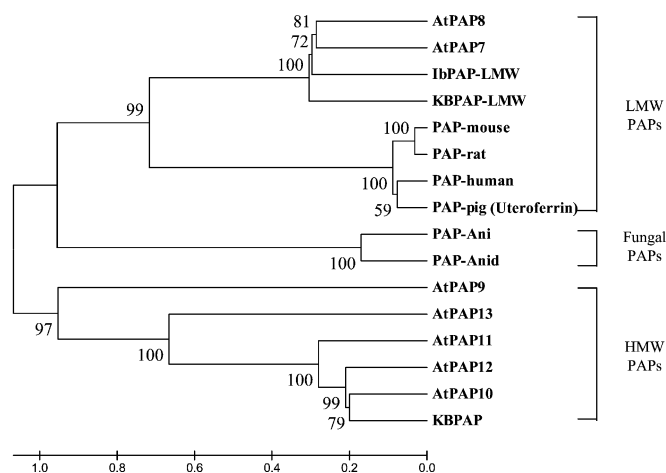


FIG. 4. Phylogenetic relationships of AtPAP7-AtPAP13 to homologous proteins from eukaryotic organisms. The phylogenetic tree presented was constructed using the neighbor joining method (with PC distance and pairwise deletion options). Bootstrap values are percentages of 500 replications. The rate of amino acid substitution is shown below the phylogenetic tree. The GenBank™ accession nos. for the homologous proteins are AF236107 (*IbPAP-LMW*), AF236109 (*KBPAP-LMW*), Q05117 (*PAP-mouse*), P29288 (*PAP-rat*), P13686 (*PAP-human*), P09889 (*PAP-pig*, uteroferrin), U18553 (*PAP-Ani*), Z79750 (*PAP-Anid*), and P80366 (*KBPAP*).

The wild type transcripts were associated exclusively with polyribosome fractions in the cells grown with either phosphate-sufficient or phosphate-limiting medium (Fig. 3C). There was a clear, but limited, association of the splice variant with polyribosomes in the cells grown with sufficient phosphate supply (Fig. 3C). However, in the cells cultured with limited phosphate supply, the association of the splice variants with polyribosomes became more extensive (Fig. 3C). During this experiment, it was interesting to note the pattern for the association of the wild type transcripts with polyribosomes was also affected by phosphate deprivation (Fig. 3C).

**Phylogenetic and Structural Relationships of Arabidopsis PAPs to Homologous Proteins from Other Eukaryotes**—The availability of amino acid sequences of AtPAP7-AtPAP13 derived from cDNA analysis permitted a more reliable assessment of phylogenetic relationships of Arabidopsis PAPs to homologous proteins from other eukaryotes. In the phylogenetic tree constructed by the neighbor joining method, there were three clades (LMW PAPs, fungal PAPs, and HMW PAPs), each with more than 95% bootstrap support (Fig. 4). AtPAP9-AtPAP13 clustered with KBPAP, forming the HMW PAP clade (Fig. 4). In contrast, AtPAP7 and AtPAP8 clustered with LMW PAPs from other plant species, forming one of two groups in the LMW PAP clade (Fig. 4). In parallel with the LMW plant PAP group was the group comprising PAPs from mammalian species. This suggested that LMW PAPs from plants and mammals might have shared a common ancestral gene in their evolutionary history.

The amino acid sequences of AtPAP7 and AtPAP8 could be aligned with that of uteroferrin without the need to introduce frequent and major gaps (Fig. 5A). Furthermore, strong conservations were observed in the amino acid sequence elements containing the seven invariant metal-ligating residues (Fig. 5A, boxed sequence elements), and in the asparagine residue that may undergo posttranslational glycosylation modification (Fig. 5A, indicated by empty triangle). However, the three residues implicated in substrate catalysis in mammalian PAPs were not strictly conserved in AtPAP7 and AtPAP8 (Fig. 5A, marked by asterisks). In contrast, extensive gaps were seen in the alignment of the amino acid sequences of AtPAP9-

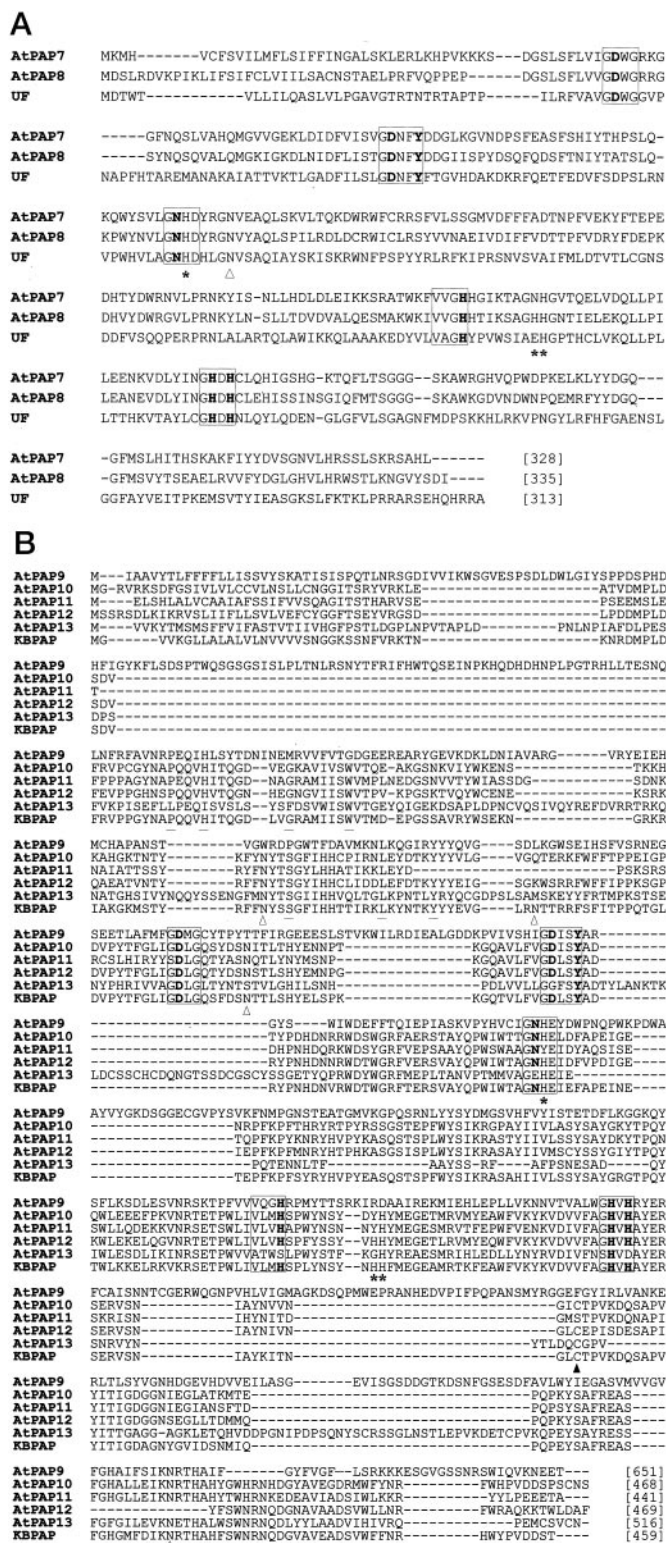


FIG. 5. Comparison of the primary structure of Arabidopsis PAPs with that of uteroferrin and KBPAP. A, comparison among AtPAP7, AtPAP8, and uteroferrin (UF). B, comparison among AtPAP9, AtPAP10, AtPAP11, AtPAP12, AtPAP13, and KBPAP. The conserved sequence motifs are boxed, and the metal-ligating residues are represented by bold letters. Asterisks indicate residues participating in substrate catalysis in uteroferrin or KBPAP. The empty triangle marks the site of glycosylation in uteroferrin or KBPAP. The filled triangle indicates the cysteine residue involved in disulfide bond formation in KBPAP. The underlined residues in B represent residues conserved between the NH<sub>2</sub>-domain of KBPAP and the fibronectin type III domain of human fibronectin.



AtPAP13 to that of KBPAP (Fig. 5B). Substantial variation was found in the sequence elements containing the potential metal-ligating residues (Fig. 5B, boxed sequence elements), and, most obviously, four of the seven potential metal-ligating residues could not be recognized at the anticipated positions of the amino acid sequence of AtPAP13 (Fig. 5B). In addition, strict conservation was not found for the asparagine residues (Fig. 5B, indicated by *empty triangles*) previously identified as sites of glycosylation in KBPAP, the cysteine residue (Fig. 5B, indicated by *filled triangle*) that had been involved in the formation of the disulfide bridge required for the dimerization of KBPAP, and the three residues participating in substrate catalysis by KBPAP (Fig. 5B, marked by *asterisks*). In a previous investigation, Tsyguelnai and Doolittle (21) found that the  $\text{NH}_2$ -domain of KBPAP shared seven conserved residues with the fibronectin type III domain of human fibronectin. A complete preservation of the seven residues was seen only in two of the five *Arabidopsis* PAPs (AtPAP10 and AtPAP12; Fig. 5B, underlined residues).

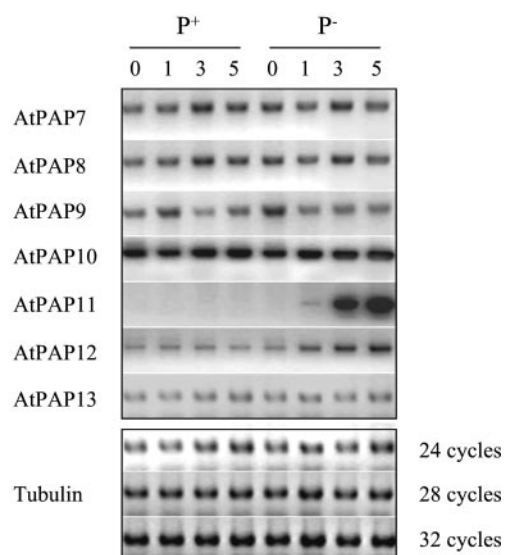
**Differential Transcriptional Responses of Arabidopsis PAPs to Phosphate Deprivation**—Because suspension cultures of different plant species had been employed in past studies on the regulation of many types of genes by phosphate deprivation treatment (31, 45–53), the transcriptional responses of AtPAP7–AtPAP13 to phosphate deprivation treatment were investigated using *Arabidopsis* suspension cells. Total RNA was prepared from both control and phosphate-deprived cell samples. After reverse transcription, the cDNA content of different samples was normalized (as described under “Experimental Procedures”). Care was also taken to ensure that amplification of the different AtPAPs entered, but did not exceed, the exponential phase of the PCR, and to maintain equal loading for all PCR samples to be checked on agarose gels. With these optimizations, three patterns of transcriptional responses were found for AtPAP7–AtPAP13 (Fig. 6) in relation to phosphate deprivation treatment. For AtPAP7, AtPAP8, AtPAP9, AtPAP10, and AtPAP13, their transcript levels were not obviously affected by low phosphate treatment (Fig. 6). The transcript level of AtPAP12 increased upon low phosphate treatment (Fig. 6). In stark contrast, a dramatic *de novo* induction of transcription was observed for AtPAP11 in response to phosphate deprivation (Fig. 6).

#### DISCUSSION

Considerable insights have been gained in past x-ray crystallography studies on the structure of PAPs (1, 3, 13, 23). However, the progress in structural studies has so far not been matched by improved understanding of the biological functions of PAPs. To advance investigations into higher plant PAPs, we have performed comparative analysis of multiple PAPs from *A. thaliana*.

Blastp searches of *Arabidopsis* protein data base identified 29 PAPs encoded by this species, 24 of which possessed the seven invariant amino acid residues involved in the coordination of the dimetal nuclear center of known PAPs. Transcripts for the great majority of the 24 PAPs were found in *Arabidopsis* cells (Table III). Furthermore, AtPAP17 (AtACP5, Table III), 1 of the 24 AtPAPs, has been previously shown to resemble PAPs from mammalian cells in primary structure and biochemical properties, and to be involved in phosphate mobilization and the metabolism of reactive oxygen species *in vivo* (31). On the basis of these results, we suggest that the 24 *Arabidopsis* PAPs are all likely to be active metallo-phosphoesterases, although they may differ from each other in aspects of their *in vivo* function.

In contrast to the above situation, AtPAP13 did not contain the whole complement of the seven invariant metal-ligating



**FIG. 6. Differential transcriptional responses of AtPAP7–AtPAP13 to phosphate deprivation treatment in suspension cells as assessed by semiquantitative RT-PCR.** Samples were taken from phosphate-deprived ( $P^-$ ) as well as the control, phosphate-sufficient ( $P^+$ ) suspension cultures at selected time points. Semiquantitative PCR assay was conducted as described under “Experimental Procedures.” The transcript level of AtPAPs 7–10 and 13 was not significantly affected, whereas that of AtPAP11 and AtPAP12 was induced and increased, respectively, by phosphate deprivation treatment. In these experiments, the amplification of tubulin transcripts was used to normalize the cDNA content of different reverse transcription reactions, and to monitor the kinetics of PCR (bottom panel).

residues typical of known PAPs, nor did it possess varieties of potential metal-ligating residues resembling those of other types of metallo-phosphoesterases. This indicates that AtPAP13 may not be a biochemically active PAP. However, an alternative physiological function for AtPAP13 in *Arabidopsis* cells may still exist because its coding gene was transcribed. In AtPAPs 14, 16, 28, and 29, the varieties of potential metal-ligating residues resembled those of exonucleases and phosphodiesterases (1, 42–44). In the light of this finding, it will be interesting to study further the structure and biochemical properties of the four unusual AtPAPs to reveal their differences to typical PAPs. It is also worthwhile to note that proteins homologous to AtPAPs 14, 16, 28, and 29 are present in both eukaryotic and prokaryotic organisms and that the six plant PAPs listed in Fig. 1 appear to be more similar to their counterparts from bacterial cells in the size (and probably primary structure as well) of their polypeptide chains. This suggests that the bacterial and plant PAPs of this type may have evolved with similar structural constraints and, consequently, may share similarities in properties.

Based on clustering analysis, the 29 AtPAPs could be classified into three main groups, a further division of which yielded eight subgroups (Fig. 2). Although detailed studies have yet to be performed for the majority of the 29 AtPAPs, the following lines of evidence indicate that the classification scheme we proposed may reflect structural and biochemical differences as well as phylogenetic relationships among the compared enzymes. First, although the classification scheme we proposed was derived essentially from alignment of amino acid sequence elements conserved in 29 *Arabidopsis* PAPs (because of the employment of the complete deletion option during the clustering analysis), there was a general correlation between the division of the groups and the size of the predicted proteins of AtPAPs (Fig. 2). This indicates that our classification scheme may reflect structure differences among the different types of AtPAPs. Second, at the subgroup level, there was a clustering



of AtPAPs that were likely to have similar biochemical properties. For example, most members in subgroups Ib-1 and Ib-2 have been shown to exhibit high percentages of identity to a phytase purified from soybean seedlings. Third, in our phylogenetic analysis using the amino acid sequences derived from cDNA analysis of AtPAP7–AtPAP13, it was found that the seven PAPs partitioned in both the HMW and LMW PAP groups (Fig. 4). AtPAP9–AtPAP13, which belonged to the HMW PAP group (Fig. 4), were located in groups I and II of our classification scheme. In contrast, AtPAPs 7 and 8, which were phylogenetically close to LMW PAPs (Fig. 4), resided in group III. This suggests that our classification of *Arabidopsis* PAPs was in accord with their evolutionary differences. It is interesting to note that group II AtPAPs in our classification system (Fig. 2) are proteins the size of which is generally larger than that of group I members. However, based on the clustering of AtPAP9 (a representative of group II) with AtPAP10–AtPAP12 and KBPAP in phylogenetic analysis (Fig. 4), we propose that group II AtPAPs may still be related to the HMW PAPs in phylogeny.

Because of phylogenetic relatedness (Fig. 4), the primary structure of AtPAPs 7 and 8 was compared with that of uteroferrin (as the representative of LMW PAPs), and the primary structure of AtPAP9–AtPAP13 compared with that of KBPAP (as the representative of HMW PAPs). These comparisons were conducted with the aim to find out the patterns in the conservation (or variation) of the primary structure among PAPs functioning in plant or animal cells. Despite the fact that plants and animals diverged approximately one billion years ago, AtPAPs 7 and 8 and uteroferrin were remarkably similar in terms of protein size, variety of metal-ligating residues, residue for glycosylation, and some of the residues involved in substrate catalysis. In contrast, there was extensive variation in the primary structure among AtPAP9–AtPAP13 and in between the five AtPAPs and KBPAP. This variation included changes in the size of the protein (as already noted in Fig. 2), the variety of potential metal-ligating residues, potential site for glycosylation or disulfide bond formation, the residues involved substrate catalysis, and the residues that may play a role in the structure formation of the NH<sub>2</sub>-terminal domain. Clearly, less variation was seen in the LMW type of AtPAPs (as represented by AtPAPs 7 and 8) when compared with the HMW type of AtPAPs (as represented by AtPAPs 9–13) in terms of protein size and variety of potential metal-ligating residues. This would imply that the LMW and HMW type of AtPAPs may have evolved in different manners, although the underlying reasons are presently not known. From a more practical point of view, the identification of the conservation (or variation) pattern through primary structure comparisons may yield information on biochemical, catalytic and/or functional properties of the AtPAPs under investigation, *e.g.* from the alignment shown in Fig. 5B, it may be predicted that AtPAPs 10 and 12 would resemble KBPAP in many aspects of their enzymatic properties.

While characterizing the cDNAs of AtPAP7–AtPAP13, a surprising finding was that the genes encoding AtPAP10 and AtPAP13 produced both wild type, fully processed transcripts and splice variants. The splice variants of AtPAP10 were found to associate with polyribosomes, and the extent of this association could be increased by phosphate deprivation (Fig. 3C). The putative product of AtPAP10 splice variant may be translated and be enzymatically active, and if it is, it may behave more like a LMW plant PAPs because of the lack of the amino-terminal domain that would be present in the wild type enzyme. The putative product of AtPAP13 splice variant was also smaller because of deletions in both amino- and carboxyl-ter-

минаl regions, and like the wild type protein, did not possess the whole complement of the seven invariant metal-ligating residues typical of PAPs. In animal cells, the production and function of NH<sub>2</sub>-terminal splice variants have been demonstrated for a number of genes (54–59). Our results indicate that the function of some plant genes may also involve more than one form of a protein translated from either wild type transcripts or NH<sub>2</sub>-terminal splice variants. Further experiments are needed to investigate the biochemical, catalytic, and functional properties of the two putative forms of AtPAP10 in *Arabidopsis* cells. In contrast to AtPAP10, functional investigations in AtPAP13 may be complicated by the fact that both forms of putative proteins (wild type protein and the product of the splice variant) did not contain the whole complement of metal-ligating residues.

Because of their phosphoesterase activity, it is understandable that past studies have found evidence for the involvement of PAPs in phosphorus nutrition of plant cells. However, previous investigators have focused mainly on PAPs inducible by phosphate deprivation. In our study on AtPAP7–AtPAP13, it was found that there were three different patterns of transcriptional responses of PAPs in relation to phosphate deprivation treatment in suspension cells. The transcript level of five AtPAPs (AtPAPs 7–10 and 13) was not affected, whereas that of AtPAP11 and AtPAP12 was induced and increased, respectively (Fig. 6). These results suggest that transcriptional responses of PAPs to phosphate deprivation are more complex than previously reported. From the comparison of primary structure shown in Fig. 5B, it could be seen that one of the differences among AtPAP10, AtPAP11, and AtPAP12 (three AtPAPs possessing the three different patterns of transcriptional responses to phosphate deprivation treatment) was that the cysteine residue involved in dimerization of KBPAP was conserved in AtPAP10 and AtPAP12, but not in AtPAP11 (Fig. 5B). Further research is needed to study whether or not the biochemical function of AtPAP11 involves the formation of homodimers. Although phosphate deprivation did not appear to change the transcript levels of AtPAPs 7–10, and 13, it affected the association of AtPAP10 transcripts with polyribosomes (Fig. 3C) and caused an increased level of AtPAP10 splice variants that have the potential to direct the production of an enzyme with potentially new properties. This indicates that, in future studies on the function of PAPs in phosphorus nutrition of plant cells, attention ought to be given to both inducibly and constitutively expressed members of these enzymes.

Judging from the comparative analysis presented, there appear to be high levels of variations in the structure, transcription regulation, and responses to phosphate deprivation in *Arabidopsis* PAPs. Consequently, functional dissection for 29 AtPAPs will be a complex task, which will require the implementation of multiple strategies in research. The use of a mutant for functional investigation of an acid phosphatase in *Arabidopsis* has been reported, although it is not known whether the affected acid phosphatase is actually a PAP (60). As the number of knockout mutants for *Arabidopsis* genes increases in various functional genomics projects, mutants missing one or more PAPs will become important tools for functional studies. Another important strategy will be to identify regulatory loci affecting the expression of multiple PAPs. A MYB transcription factor required for the increased expression of several genes under phosphate deprivation conditions has been identified (61), one of which is that for AtACP5 (AtPAP17, Table II) (31, 61). However, the transcription of other AtPAPs was not reported in that work. Recently, a *pho3* locus was reported to cause a 30% reduction of root acid phosphatase activity in *Arabidopsis* (62). It will be interesting to test if the

reduction of root acid phosphatase activity by *pho3* involves reduced expression of any of the 29 AtPAPs. Finally, in addition to the above genetic strategies, a biochemical approach aimed at comparing the biochemical properties of different AtPAPs using *in vitro* expressed proteins would also generate information useful in the interpretation of the results from the genetic experiments. Only by combining the different research strategies discussed above may a more complete understanding of the function of PAPs in *Arabidopsis* biology be attainable.

**Acknowledgments**—We thank Dr. Jiayang Li for helpful discussions on our *Arabidopsis* research and Dr. Hongqing Ling for critical comments on the manuscript.

## REFERENCES

- Klabunde, T., Strater, N., Frohlich, R., Witzel, H., and Krebs, B. (1996) *J. Mol. Biol.* **259**, 737–748
- Klabunde, T., Strater, N., Krebs, B., and Witzel, H. (1995) *FEBS Lett.* **367**, 56–60
- Strater, N., Klabunde, T., Tucker, P., Witzel, H., and Krebs, B. (1995) *Science* **268**, 1489–1492
- Schenk, G., Guddat, L. W., Ge, Y., Carrington, L. E., Hume, D. A., Hamilton, S., and de Jersey, J. (2000) *Gene (Amst.)* **250**, 117–125
- Schenk, G., Korsinczyk, M. L., Hume, D. A., Hamilton, S., and de Jersey, J. (2000) *Gene (Amst.)* **255**, 419–424
- Goldberg, J., Huang, H. B., Kwon, Y. G., Greengard, P., Nairn, A. C., and Kuriyan, J. (1995) *Nature* **376**, 745–753
- Kissinger, C. R., Parge, H. E., Knighton, D. R., Lewis, C. T., Pelletier, L. A., Tempczyk, A., Kalish, V. J., Tucker, K. D., Showalter, R. E., Moomaw, E. W., Gastinel, L. N., Habuka, N., Chen, X., Maldonado, F., Baker, J. E., Bacquet, R., and Villafranca, J. E. (1995) *Nature* **378**, 641–644
- Griffith, J. P., Kim, J. L., Kim, E. E., Sintchak, M. D., Thomson, J. A., Fitzgibbon, M. J., Fleming, M. A., Caron, P. R., Hsiao, K., and Navia, M. A. (1995) *Cell* **82**, 507–522
- Koonin, E. V. (1994) *Protein Sci.* **3**, 356–358
- Zhuo, S., Clemens, J. C., Stone, R. L., and Dixon, J. E. (1994) *J. Biol. Chem.* **269**, 26234–26238
- Lohse, D. L., Denu, J. M., and Dixon, J. E. (1995) *Structure* **3**, 987–990
- Vincent, J. B., and Averill, B. A. (1990) *FEBS Lett.* **263**, 265–268
- Uppenberg, J., Lindqvist, F., Svensson, C., Ek-Rylander, B., and Andersson, G. (1999) *J. Mol. Biol.* **290**, 201–211
- Zhang, J., Zhang, Z., Brew, K., and Lee, E. Y. (1996) *Biochemistry* **35**, 6276–6282
- Huang, H. B., Horiuchi, A., Goldberg, J., Greengard, P., and Nairn, A. C. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 3530–3535
- Mertz, P., Yu, L., Sikkink, R., and Rusnak, F. (1997) *J. Biol. Chem.* **272**, 21296–21302
- Klabunde, T., and Krebs, B. (1997) *Bond* **89**, 177–198
- Schenk, G., Carrington, L. E., Hamilton, S. E., de Jersey, J., and Guddat, L. W. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 2051–2052
- Durmus, A., Eicken, C., Sift, B. H., Kratel, A., Kappl, R., Huttermann, J., and Krebs, B. (1999) *Eur. J. Biochem.* **260**, 709–716
- Beck, J. L., McConaghie, L. A., Summors, A. C., Arnold, N. W., de Jersey, J., and Zerner, B. (1986) *Biochim. Biophys. Acta* **869**, 61–68
- Tsyguelnai, I., and Doolittle, R. F. (1998) *J. Mol. Evol.* **46**, 612–614
- Lindqvist, Y., Johansson, E., Kaija, H., Vihko, P., and Schneider, G. (1999) *J. Mol. Biol.* **291**, 135–147
- Guddat, L. W., McAlpine, A. S., Hume, D., Hamilton, S., de Jersey, J., and Martin, J. L. (1999) *Struct. Fold. Des.* **7**, 757–767
- Nakazato, H., Okamoto, T., Nishikoori, M., Washio, K., Morita, N., Haraguchi, K., Thompson, G. A., Jr., and Okuyama, H. (1998) *Plant Physiol.* **118**, 1015–1020
- Wasaki, J., Omura, M., Osaki, M., Ito, H., Matsui, H., Shinano, T., and Tadano, T. (1999) *Soil Sci. Plant Nutr.* **45**, 439–449
- Miller, S. S., Liu, J., Allan, D. L., Menzhuber, C. J., Fedorova, M., and Vance, C. P. (2001) *Plant Physiol.* **127**, 594–606
- Hayman, A. R., Jones, S. J., Boyde, A., Foster, D., Colledge, W. H., Carlton, M. B., Evans, M. J., and Cox, T. M. (1996) *Development* **122**, 3151–3162
- Nuttleman, P. R., and Roberts, R. M. (1990) *J. Biol. Chem.* **265**, 12192–12199
- Sibille, J. C., Doi, K., and Aisen, P. (1987) *J. Biol. Chem.* **262**, 59–62
- Johnson, J. F., Vance, C. P., and Allan, D. L. (1996) *Plant Physiol.* **112**, 31–41
- del Pozo, J. C., Allona, I., Rubio, V., Leyva, A., de la Pena, A., Aragoncillo, C., and Paz-Ares, J. (1999) *Plant J.* **19**, 579–589
- Haran, S., Logendra, S., Seskar, M., Bratanova, M., and Raskin, I. (2000) *Plant Physiol.* **124**, 615–626
- Hegeman, C. E., and Grabau, E. A. (2001) *Plant Physiol.* **126**, 1598–1608
- Durmus, A., Eicken, C., Spener, F., and Krebs, B. (1999) *Biochim. Biophys. Acta* **1434**, 202–209
- Baldwin, J. C., Karthikeyan, A. S., and Raghothama, K. G. (2001) *Plant Physiol.* **125**, 728–737
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Nei, N., and Kumar, S. (2000) *Molecular Evolution and Phylogenetics*, Oxford University Press, Oxford, United Kingdom
- Hu, Y., Bao, F., and Li, J. (2000) *Plant J.* **24**, 693–701
- Mou, Z., He, Y., Dai, Y., Liu, X., and Li, J. (2000) *Plant Cell* **12**, 405–418
- Davies, E., and Abe, S. (1995) in *Methods in Cell Biology* (Galbraith, D. W., Bourque, D. P., and Bohnert, H. J., eds) pp. 209–222, Academic Press, San Diego, CA
- Petracek, M. E., Dickey, L. F., Huber, S. C., and Thompson, W. F. (1997) *Plant Cell* **9**, 2291–2300
- Takahashi, T., Suchi, M., Desnick, R. J., Takada, G., and Schuchman, E. H. (1992) *J. Biol. Chem.* **267**, 12552–12558
- Imamura, R., Yamanaka, K., Ogura, T., Hiraga, S., Fujita, N., Ishihama, A., and Niki, H. (1996) *J. Biol. Chem.* **271**, 25423–25429
- Schuchman, E. H., Suchi, M., Takahashi, T., Sandhoff, K., and Desnick, R. J. (1991) *J. Biol. Chem.* **266**, 8531–8539
- Abel, S., Nurnberger, T., Ahnert, V., Krauss, G. J., and Glund, K. (2000) *Plant Physiol.* **122**, 543–552
- Jost, W., Bak, H., Glund, K., Terpstra, P., and Beintema, J. J. (1991) *Eur. J. Biochem.* **198**, 1–6
- Kock, M., Löffler, A., Abel, S., and Glund, K. (1995) *Plant Mol. Biol.* **27**, 477–485
- Löffler, A., Glund, K., and Irie, M. (1993) *Eur. J. Biochem.* **214**, 627–633
- Burleigh, S. M., and Harrison, M. J. (1998) *Gene (Amst.)* **216**, 47–53
- Duff, S. M., Plaxton, W. C., and Lefebvre, D. D. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9538–9542
- Malboobi, M. A., and Lefebvre, D. D. (1995) *Plant Mol. Biol.* **28**, 859–870
- Palma, D. A., Blumwald, E., and Plaxton, W. C. (2000) *FEBS Lett.* **486**, 155–158
- Theodorou, M. E., Cornel, F. A., Duff, S. M., and Plaxton, W. C. (1992) *J. Biol. Chem.* **267**, 21901–21905
- Bamberger, C., and Schmale, H. (2001) *FEBS Lett.* **501**, 121–126
- Chraibi, A., Verdumo, C., Merillat, A. M., Rossier, B. C., Horisberger, J. D., and Hummler, E. (2001) *Cell Physiol. Biochem.* **11**, 115–122
- Kone, B. C., and Higham, S. C. (1998) *J. Biol. Chem.* **273**, 2543–2552
- Nakajima, T., Ichii, S., Furuyama, J., Tamaoki, T., and Hashimoto, T. (1996) *Life Sci.* **58**, 761–768
- Schumacher, M. A., Moff, I., Sudanagunta, S. P., and Levine, J. D. (2000) *J. Biol. Chem.* **275**, 2756–2762
- Sunada, Y., Bernier, S. M., Utani, A., Yamada, Y., and Campbell, K. P. (1995) *Hum. Mol. Genet.* **4**, 1055–1061
- Trull, M. C., and Deikman, J. (1998) *Planta* **206**, 544–550
- Rubio, V., Linhares, F., Solano, R., Martin, A. C., Iglesias, J., Leyva, A., and Paz-Ares, J. (2001) *Genes Dev.* **15**, 2122–2133
- Zakhleniuk, O. V., Raines, C. A., and Lloyd, J. C. (2001) *Planta* **212**, 529–534