Identification of mammalian-like purple acid phosphatases in a wide range of plants

G. Schenk *, L.W. Guddat, Y. Ge, L.E. Carrington, D.A. Hume, S. Hamilton, J. de Jersey

Department of Biochemistry, The University of Queensland, Brisbane, Queensland 4072, Australia

Abstract

Purple acid phosphatases (PAPs) comprise a family of binuclear metal-containing hydrolases, members of which have been isolated from plants, mammals and fungi. Polypeptide chains differ in size (animal ~ 35 kDa, plant ~ 55 kDa) and exhibit low sequence homology between kingdoms but all residues involved in co-ordination of the metal ions are invariant. A search of genomic databases was undertaken using a sequence pattern which includes the conserved residues. Several novel potential PAP sequences were detected, including the first known examples from bacterial sources. Ten plant ESTs were also identified which, although possessing the conserved sequence pattern, were not homologous throughout their sequences to previously known plant PAPs. Based on these EST sequences, novel cDNAs from sweet potato, soybean, red kidney bean and *Arabidopsis thaliana* were cloned and sequenced. These sequences are more closely related to mammalian PAP than to previously characterized plant enzymes. Their predicted secondary structure is similar to that of the mammalian enzyme. A model of the sweet potato enzyme was generated based on the coordinates of pig PAP. These observations strongly suggest that the cloned cDNA sequences represent a second group of plant PAPs with properties more similar to the mammalian enzymes than to the high molecular weight plant enzymes. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Binuclear metal centre; Fenton-like reaction; Homology modelling; Phosphate acquisition; Phylogenetics

1. Introduction

PAPs comprise a group of metalloenzymes which contain a binuclear metal ion complex at the active site. They are distinguished from other phosphatases by their acidic pH optima and their characteristic purple colour which is due to the presence of a phenolate to metal charge transfer complex. Several mammalian PAPs (human, cow, pig, mouse and rat) have been isolated and characterised. All are monomeric (~ 35 kDa), contain an Fe(III)–Fe(II) binuclear metal centre and exhibit high sequence identity ($>80\%$) (Klabunde and Krebs, 1997). The most widely studied plant PAPs are from red kidney bean, sweet potato and soybean. They are homodimeric with subunit molecular masses of ~ 55 kDa and are more diverse with respect to metal content than the mammalian enzymes (Beck et al., 1986;

LeBansky et al., 1992; Schenk et al., 1999). In the red kidney bean and soybean PAPs the metal centres are Fe(III)–Zn(II) (Beck et al., 1986; Schenk et al., 1999). Two forms of the enzyme have been isolated from sweet potato. One form contains an Fe–Zn centre similar to the enzymes from red kidney bean and soybean (Durmus et al., 1999), and a second isoform contains Mn instead of Zn (Schenk et al., 1999). Additional cDNAs encoding putative 55 kDa PAPs have been isolated, including a third example from sweet potato and two from *Arabidopsis thaliana*. A comparison of the primary structures of all these plant PAPs shows that they have a high overall sequence similarity ($\geq 70\%$) (Schenk et al., 1999).

The three-dimensional structures of pig and red kidney bean PAPs have been solved to 1.55 Å (Guddat et al., 1999) and 2.65 Å (Klabunde et al., 1996), respectively. Although their sequence identity is low ($<20\%$) (Klabunde et al., 1995) they may be aligned via the metal-ligating residues (Klabunde et al., 1995): Asp135, Asp164, Tyr167, Asn201, His286, His323 and His325 (numbered according to red kidney bean PAP). These

invariant amino acids occur within five blocks of residues which are present in all of the known PAPs. Different numbers of residues separate the blocks. The core structures of pig and red kidney bean enzymes are very similar: 204 Cα atoms may be superimposed with a root mean square deviation (rmsd) of 1.4 Å (Guddat et al., 1999). The plant enzymes have an additional N-terminal domain that does not appear to participate in the catalytic function.

The fungus *Aspergillus ficuum* has a monomeric PAP with a molecular mass of ~85 kDa (Ullah and Cummins, 1988). The nature of the metal centre is unknown. The corresponding gene sequence exhibits high identity (>95%) with sequences from *A. niger* (Mullaney et al., 1995) and *A. nidulans* (Sarkar et al., 1996). No PAP from a bacterial source has been previously identified.

Roles for the mammalian enzyme in bone resorption (Hayman et al., 1996) and iron transport (Nuttleman and Roberts, 1990) have been proposed. The biological function(s) of PAP in plants is not known. In rice and soybean, the enzyme is repressed by inorganic phosphate, suggesting a role in phosphate acquisition by hydrolysis of esters such as phytate (Duff et al., 1994). Recently, a cDNA sequence from an Easter lily pollen-specific clone was reported and shown to have significant similarity with the mammalian PAPs (Kim and Gynheung, 1996). Notably, it encoded a protein of 329 amino acids, similar in length to the mammalian enzymes and significantly smaller than the ~450-residue plant enzymes. It has been speculated that this protein may function as an iron carrier in mature pollen. Its occurrence in Easter lily may reflect a novel PAP structure in a subgroup of plants or the existence of a more widely distributed second family of PAPs in plants. The aim of this study was to distinguish between these two possibilities.

2. Materials and methods

2.1. Materials

Soybeans (cultivar Provar), red kidney beans (cultivar Rufus) and sweet potatoes (cultivar Golden) were purchased locally. *Arabidopsis thaliana* plants (cultivar Columbus) were supplied by Dr. B. J. Carroll. All chemicals were analytical grade unless specified otherwise. Molecular biology reagents and DNA-modifying enzymes were purchased from New England Biolabs, Promega, Progen Industries and Roche Molecular Biochemicals. *Escherichia coli* strain DH5α was obtained from Gibco BRL.

2.2. Database search, sequence alignment and sequence comparisons

Table 1 summarizes the sources of sequences of known PAPs and sequences for which homology with

Table 1
Partial alignment of previously reported purple acid phosphatases and purple acid phosphatase-like sequences. The conserved blocks of residues surrounding the seven metal-ligating residues (in bold) are shown. The numbers of residues between conserved blocks and total number of residues are indicated and the consensus sequence is shown underneath

	Source	Code	1		2		3		4		5	Length	Accession[a]	References[b]
Plants	*A. thaliana*	*Ath*1	G**D**L**G**	25	**G**DLSY	32	G**N**HE	79	VLV**H**	35	G**H**V**H**	469	Q38924	–
	A. thaliana	*Ath*2	G**D**L**G**	24	**G**DLSY	32	G**N**HE	79	VMV**H**	35	G**H**V**H**	466	Z99708-	–
	Sweet potato	*Iba*1	G**D**I**G**	25	**G**DLSY	32	G**N**HE	79	VLV**H**	35	G**H**V**H**	473	AF200825	1
	Sweet potato	*Iba*2	G**D**L**G**	25	**G**DLSY	32	G**N**HE	79	VLM**H**	35	G**H**V**H**	465	AF200826	1,2
	Sweet potato	*Iba*3	G**D**L**G**	25	**G**DLSY	32	G**N**HE	79	VLM**H**	35	G**H**V**H**	427[c]	AJ006870	2
	Soybean	*Gma*	G**D**L**G**	25	**G**DLSY	32	G**N**HE	79	VLM**H**	35	G**H**V**H**	464	AF200824	1
	Red kidney bean	*Pvu*	G**D**L**G**	25	**G**DLSY	32	G**N**HE	79	VLM**H**	35	G**H**V**H**	459	P80366	3
	Easter lily	*Llo*	G**D**W**G**	29	**G**DNFY	33	G**N**HD	89	VGH**H**	33	G**H**D**H**	329	not available	4
Animals	Human	*Hsa*	G**D**W**G**	34	**G**DNFY	34	G**N**HD	87	VAG**H**	33	G**H**D**H**	325	P13686	5
	Mouse	*Mmu*	G**D**W**G**	34	**G**DNFY	34	G**N**HD	89	VAG**H**	33	G**H**D**H**	327	Q05117	6
	Rat	*Rno*	G**D**W**G**	34	**G**DNFY	34	G**N**HD	89	VAG**H**	33	G**H**D**H**	327	P29288	7
	Pig	*Ssc*	G**D**W**G**	34	**G**DNFY	34	G**N**HD	89	VAG**H**	33	G**H**D**H**	338	P09889	8
	Zebrafish[d]	*Dre*	G**D**W**G**	34	**G**DNFY	34	G**N**HD						AI415792	–
Fungi	*A. ficuum*	*Afi*	**N**DMG	25	**G**DLSY	84	G**N**HE	151	VMS**H**	33	G**H**I**H**	614	U18554	9
	A. niger	*Ani*	**N**DMG	25	**G**DISY	84	G**N**HE	151	VMS**H**	33	G**H**I**H**	614	U18553	9
	A. nidulans	*Anid*	**N**DMG	25	**G**DLSY	84	G**N**HE	151	VMS**H**	33	G**H**I**H**	618	Z79750	10
Consensus			**D**X**G**		**G**DXXY		G**N**HE/**D**		VXX**H**		G**H**X**H**			

[a] Database accession codes.
[b] Numbers indicate the following references: (–) unpublished; (1) Schenk et al., 1999; (2) Durmus et al., 1999; (3) Klabunde et al., 1994; (4) Kim and Gynheung, 1996; (5) Ketcham et al., 1989; (6) Cassady et al., 1993; (7) Ek-Rylander et al., 1991; (8) Hunt et al., 1987; (9) Mullaney et al., 1995; (10) Sarkar et al., 1996.
[c] Signal peptide missing.
[d] Only partial sequence available.

known PAPs had been recognised prior to the present work. Alignments of these sequences were performed using ClustalW software (Thompson et al., 1994). Further refinements were introduced by manual analysis, based on previous alignments of PAPs from red kidney bean and pig (Klabunde et al., 1995; Guddat et al., 1999) and from red kidney bean and *A. ficuum* (Mullaney and Ullah, 1998). Gaps were inserted into the alignment to minimise the probability of disrupting secondary structures. Based on the alignment, a consensus sequence for PAPs was derived and used to search the various databases with the GCG algorithm FINDPATTERN (Wisconsin Package, Version 8.1.0, March 1996). Databases were also searched with the Easter lily sequence using the Blast algorithm (Altschul et al., 1997). Unless stated otherwise, per cent sequence identities and similarities (ratio between the number of identical or similar (conserved) residues and the total number of residues, including gaps) were obtained using the program BESTFIT (Wisconsin Package, Version 8.1.0, March 1996).

2.3. RNA extraction, primer design and cDNA cloning

Total RNA from sweet potato tuber and leaf tissues from soybean, red kidney bean and *Arabidopsis thaliana* was extracted using the TRIZOL reagent (Gibco BRL) as recommended by the manufacturer. cDNAs were cloned utilising the 5′- and 3′-RACE methods (Roche Molecular Biochemicals). The design of degenerate primers (forward primers 1—3) was based on an alignment between the Easter lily sequence (Table 1) and several homologous plant ESTs (listed in Table 2; see text). Subsequently, gene-specific reverse primers for 5′-RACE were designed based on the cloned 3′-end sequences. Primers (Table 3) were synthesised by Pacific Oligos, Lismore, Australia. The same conditions were applied to all polymerase chain reaction (PCR) reactions unless stated otherwise: a "hotstart" incubation at 94°C for 2 min was followed by 40 cycles of denaturation

Table 2
Relative location and identity (%) of ESTs homologous to the putative low molecular weight purple acid phosphatase from Easter Lily

EST source	Code	Accession	Location[a]	Identity
Magnolia	sm*Mtr*	AA660270	74–537	68
A. thaliana	sm*Ath*1	AB019230	75–827	68
A. thaliana	sm*Ath*2	T45273	75–562	60
A. thaliana	sm*Ath*3	AC006917	74–576	66
A. thaliana	sm*Ath*4	N96051	76–574	66
A. thaliana	sm*Ath*5	T42810	102–563	67
Rice	sm*Osa*1	C27531	89–356	69
Rice	sm*Osa*2	RICS11290A	285–448	68
Rice	sm*Osa*3	RICS14915A	430–794	72
Rice	sm*Osa*4	AU030890	602–970	70

[a] Numbering according to the Easter lily sequence.

(94°C, 15 s), annealing (55°C, 15 s) and extension (72°C, 45 s). Using the 3′-RACE method the product of first strand cDNA synthesis was used as a template for second strand synthesis with forward primer 1 and reverse anchor primer (supplied in the kit). The 3′-terminal sequence (~650 nucleotides) was amplified from this product using two rounds of semi-nested PCR using reverse anchor primer and successive forward primers 2 and 3. In the 5′-RACE method a poly(A) tail was added to the 5′-end of the first strand cDNA (see 5′/3′-RACE kit manual). Subsequently, second strand synthesis was carried out using an oligo dT anchor primer (supplied in the kit) and a gene-specific reverse primer (reverse primer 1) matched against the respective 3′-end sequences, i.e., from sweet potato, soybean, red kidney bean and *A. thaliana* (SP1, SB1, RKB1 and AT1, respectively). The 5′-terminal sequences were amplified from these initial products in a semi-nested PCR reaction using the anchor primer provided by the kit and a second, nested gene-specific reverse primer (SP2, SB2, RKB2 and AT2, respectively). For the cloning of the *A. thaliana* cDNA a second semi-nested reaction needed to be carried out using reverse primer AT3.

2.4. DNA sequencing and analysis

Automated DNA sequencing applying the dideoxy chain termination method (Sanger et al., 1977) was performed using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit in the Australian Genome Research Facility. Nucleotide and deduced amino acid sequences were aligned as described above.

2.5. Phylogenetic analysis

The phylogenetic relationship of previously known PAPs and the amino acid sequences deduced from the cDNA sequences identified as described earlier was investigated using the MEGA (Kumar et al., 1993) and PHYLIP (Felsenstein, 1989) software packages. Sequence distance matrices were established in pairwise comparisons using a variety of algorithms: BESTFIT (see above), p-distance, Poisson-corrected and Gamma (a = 2) distances (see MEGA manual), and the Dayhoff PAM matrix (see PHYLIP manual). Distance–matrix based phylogenies were derived using the Neighbor–Joining method (Saitou and Nei, 1987). All phylogenies were subjected to bootstrapping (1000 replicates) (Felsenstein, 1985).

2.6. Secondary structure prediction and homology modelling

An initial alignment of the putative low molecular weight PAP from sweet potato and pig PAP was optimised using the predicted secondary structure of the sweet potato enzyme (determined with the algorithm

Table 3
Oligonucleotides used for the cloning of low molecular weight purple acid phosphatases from *A. thaliana* (AT), sweet potato (SP), soybean (SB) and red kidney bean (RKB)

Oligonucleotide	Sequence	Location[a]
Forward		
1	5′-GTGAGYTTTCTCGTNATYGGAGAYTGG-3′	121–148
2	5′-GCTCTTCAGATGGGTRARATHGGRGAG-3′	181–209
3	5′-AGCTTGCAGAARCMATGGTACAGTGTHTTRGGRAA-3′	322–354
Reverse		
AT1	5′-CCTTTCCACGCCTTGAAGCCACCTCC-3′	802–828
AT2	5′-ACTGTTGATGCCGCTTATGTGCTCC-3′	755–780
AT3	5′-ATCCAACGGCAGTCCAAATCCCTGAGT-3′	387–414
SP1	5′-GGAATGGTCGTGGCCGTTCATGTACAGAT-3′	726–755
SP2	5′-CGCATGATGGCCGATGACTATCCT-3′	622–646
SB1	5′-ATGAAGCCTTGGCCGTCGTAGTAAAAC-3′	857–886
SB2	5′-CCTCTCCATGCCTTCGATCCACCACCACTT-3′	797–827
RKB1	5′-TTGTGCTTCCACGTCTCCCCTGTAGTC-3′	301–328
RKB2	5′-CAGGGTCGTCTACACCACTCAATCCA-3′	258–284

[a] Numbering according to the Easter lily sequence (*Llo*).

PEPTIDESTRUCTURE (Jameson and Wolf, 1988) and crystallographic data of the pig enzyme (Guddat et al., 1999). Based on this alignment, a model was constructed for the sweet potato enzyme using the HOMOLOGY package included in INSIGHT 98 (MSI). The model was refined by applying energy minimisation procedures in the program DISCOVER (also part of INSIGHT 98). All buried charged or polar amino acids were involved or could potentially be involved in the formation of either hydrogen bonds or salt bridges. The final model contained no significant cavities, an indication of correct packing. The quality of the model was further validated by statistical analysis with the programs PROCHECK (Laskowski et al., 1993), WHAT IF (Vriend, 1990) and PROFILES-3D (MSI). Superposition of the 222 Cα atoms of the core residues from pig and sweet potato PAP resulted in an overall rmsd of 0.8 Å, while the seven metal-co-ordinating residues superimposed with an rmsd of 0.3 Å. Given the similar arrangements of the metal-co-ordinating residues in the crystal structure of pig PAP and the model of the sweet potato enzyme, two irons could be readily placed into the structure of the model.

3. Results

3.1. Database search

Table 1 shows a partial sequence alignment of all PAPs and PAP-like sequences known at the commencement of this study. The conserved blocks surrounding the metal ligands show kingdom-specific variations. For example, in block 5 the fungal sequence GHIH is GHDH in animals and GHVH in most plants. As noted by Klabunde and Krebs (1997), there are large variations in the number of residues between blocks 2 and 3 and between blocks 3 and 4 when fungal sequences are compared with others. Among all of the plant enzymes, the Easter lily sequence (*Llo*) is the most divergent in terms of both sequence conservation and size of the separating blocks. In both respects this sequence more closely resembles the five mammalian sequences than the other plant sequences. This is also reflected in the number of amino acid residues in the respective proteins (Table 1).

Using a consensus motif derived from Table 1 (DxG/GDx$_2$Y/GNH(E,D)/Vx$_2$H/GHxH), we retrieved

Table 4
Partial alignment of new putative purple acid phosphatases. The conserved blocks of residues surrounding the seven metal-ligating residues (in bold) are shown. The numbers of residues between conserved blocks are indicated

Source	Code	1		2		3		4		5	Length	Accession
C. elegans	Cel	GDTG	33	G**N**IY	33	**GNHD**	98	IS**GH**	32	**GHDH**	382	Z70782
Synechocystis	Syn	GDIA	38	GD**N**QY	27	**GNHE**	72	AYW**H**	32	**GHDH**	326	D90900
M. tuberculosis	Mtu	GDQS	40	GDLCY	31	**GNHE**	89	VC**MH**	35	**GHEH**	529	Z77724
M. leprae[a]	Mle	GDYS	39	GDLC**Y**	31	**GNHE**	89	VC**MH**	35	GHEH	527	U00011

[a] Sequence contains several potential point mutations and frame shifts. Amino acids underlined are residues where there is some uncertainty.

Table 5
Partial alignment of cloned putative low molecular weight plant purple acid phosphatases[a]

Source	Code	1		2		3		4		5	Length	Accession
A. thaliana	Sm*Ath*	G**D**W**G**	29	**G**D**N**F**Y**	33	**G**N**HD**	89	VV**GH**	33	**GHDH**	315	AF200827
Sweetpotato	Sm*Iba*	G**D**W**G**	29	**G**D**N**F**Y**	33	**G**N**HD**	89	VI**GH**	33	**GHDH**	313	AF236107
Soybeans	M*Gma*	G**D**W**G**	29	**G**D**N**F**Y**	33	**G**N**HD**	89	VI**GH**	33	**GHDH**	333	AF236108
Red kidney bean	Sm*Pvu*	G**D**W**G**	29	**G**D**N**F**Y**	33	**G**N**HD**	89	VV**GH**	33	**GHDH**	332	AF236109

[a] The conserved blocks of residues surrounding the seven metal-ligating residues (in bold) are shown. The numbers of residues between conserved blocks and total number of residues (Length) are indicated.

full-length sequences for four new PAP-like proteins from genomic databases (Table 4). One sequence, from *Caenorhabditis elegans* (*Cel*), encodes a protein of 382 amino acids. An acid phosphatase which is glycosylated and has a pH optimum of ~3.5 has been purified and partially characterised from this organism (Beh et al., 1991). However, insufficient enzyme was isolated to determine whether the enzyme was purple or whether it contained metal ions. Our search has also revealed for the first time the occurrence of PAP-like sequences in bacteria: *Mycobacterium tuberculosis* (*Mtu*), *M. leprae* (*Mle*) and *Synechocystis* sp. (*Syn*). A search of complete genome sequences for *Saccharomyces cerevisiae* and *Escherichia coli* failed to reveal any PAP-like sequences.

Incomplete sequences and ESTs may not contain all five conserved sequence blocks. Hence, PAP-like sequences may be overlooked when the full consensus sequence (Table 1) is used as a search tool. We have therefore searched databases for the occurrence of the motif GDx$_2$Y (block 2 of the pattern in Table 1) since the invariant tyrosine is a minimum requirement for the purple colour. Within the EST databases we recovered a range of partial PAP-like sequences from three plants including magnolia (*Medicago truncatula*, sm*Mtr*), *Arabidopsis thaliana* (sm*Ath*1 — sm*Ath*5) and rice (*Oryza sativum*, sm*Osa*1 — sm*Osa*2, Table 2). All of these sequences are more similar to the Easter lily sequence than to any of the known 55 kDa plant PAPs. Two additional rice ESTs (sm*Osa*3, sm*Osa*4) were retrieved by searching with the complete Easter lily sequence. Comparison of the multiple ESTs for *A. thaliana* with each other gave a maximum identity of 77%. This suggests that there are at least five genes coding for the small form of the plant PAP in this plant.

3.2. Cloning and sequencing

The database search described above indicated that a low molecular weight (mammalian-like) PAP may be widespread in plants. To assess the conservation of this enzyme across a range of plants, we sought to clone full-length cDNAs from sweet potato, soybean, red kidney bean and *Arabidopsis thaliana* as described in Materials and Methods. Only one sequence was obtained for each of the four plant species (Table 5), despite sequencing at least six clones for each species. Comparison of the *A. thaliana* sequence (sm*Ath*) with the *A. thaliana* ESTs (Table 2) shows that it is identical to one of them (sm*Ath*4). All of the deduced amino acid sequences display high similarity to the Easter lily sequence (~78%; Table 6). All five motifs characteristic of PAPs (Klabunde and Krebs, 1997) are conserved. In Tables 6 and 7 the low molecular weight and high molecular weight plant forms are compared with each other and with the pig enzyme. The low molecular weight plant enzymes are clearly more closely related to mammalian PAPs than to the 55 kDa plant enzymes.

Table 6
Sequence similarity[a] between low molecular weight forms of plant purple acid phosphatases and pig purple acid phosphatase. Explanation of the sequence codes is given in Tables 1 and 5

	sm*Iba*	sm*Gma*	sm*Pvu*	*Llo*	*Ssc*[b]
Sm*Ath*	78.0	77.8	79.0	78.6	52.7 (33.2)
Sm*Iba*		77.8	76.0	80.7	52.6 (29.7)
sm*Gma*			77.8	78.9	53.7 (26.7)
sm*Pvu*				78.4	50.0 (28.9)
Llo					54.9 (29.2)

[a] Sequence identities and similarities were determined with the program BESTFIT using its default settings.
[b] Numbers in brackets indicate % identity. All other numbers show % sequence similarity.

Table 7
Sequence similarity[a] between high and low molecular weight forms of plant purple acid phosphatase and pig purple acid phosphatase. Explanation of the sequence codes is given in Tables 1 and 5

	sm*Ath*	sm*Iba*	sm*Gma*	sm*Pvu*	*Llo*	*Ssc*[b]
*Ath*1	47.3	44.0	43.5	42.4	45.0	43.0 (18.9)
*Ath*2	43.6	38.7	42.1	49.1	39.8	43.5 (20.8)
*Iba*1	46.0	44.5	45.7	47.1	42.0	42.7 (18.7)
*Iba*2	46.4	45.7	45.5	38.5	43.0	43.5 (18.2)
*Iba*3	45.1	43.4	44.1	38.9	39.7	44.4 (19.4)
Gma	48.8	45.8	43.3	41.8	39.7	42.7 (18.2)
Pvu	41.2	44.5	42.3	45.7	38.5	44.6 (19.6)

[a] Sequence identities and similarities were determined with the program BESTFIT using its default settings.
[b] Numbers in brackets indicate % identity. All other numbers show % sequence similarity.
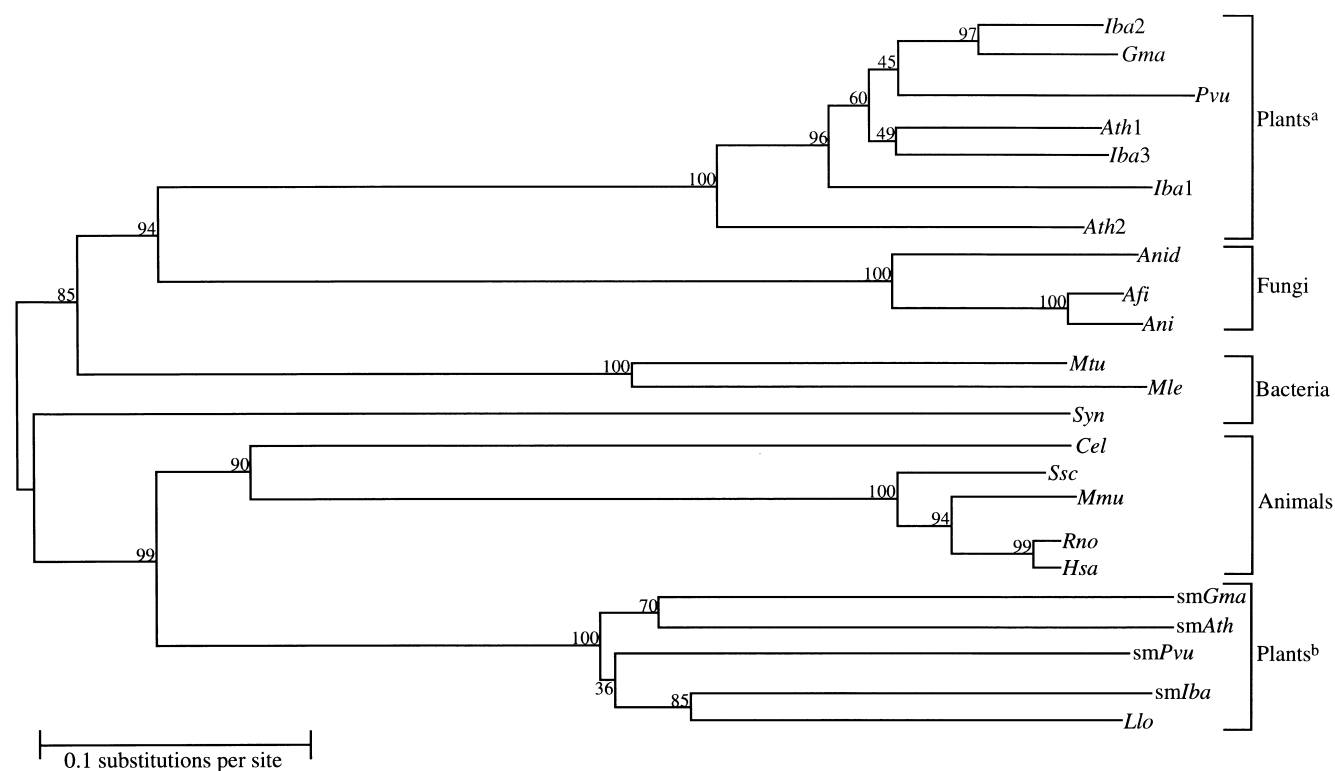
Fig. 1. Phylogenetic tree showing the relationship between all full-length purple acid phosphatases and purple acid phosphatase-like sequences listed in Tables 1, 4 and 5. Bootstrap numbers (1000 replicates) are indicated at internal nodes. The scale bar indicates the number of substitutions per site. Explanation of the sequence codes is given in Tables 1, 4 and 5. Plants[a] and plants[b] are represented by high and low molecular weight plant purple acid phosphatases, respectively.

3.3. Phylogenetic analysis

A phylogenetic comparison of all full-length sequences listed in Tables 1, 4 and 5 was undertaken. Since compositional heterogeneity was observed (data not shown) a number of phylogeny inferring algorithms were applied. The tree topologies obtained by all methods were essentially the same. One such tree is shown in Fig. 1. In general, the predicted phylogeny reflects taxonomic relationships except that the low molecular weight plant PAPs cluster with the mammalian sequences (99% bootstrap support). An interesting observation is the branching of the bacterial sequences. While *Syn* appears to share a common ancestral gene with animal and low molecular weight plant PAPs the two mycobacterial sequences are more closely related to the fungal and high molecular weight plant enzymes (85% bootstrap support).

3.4. Secondary structure prediction and homology model

Since the four cloned cDNAs encoding putative low molecular weight plant PAPs are homologous (Table 6), a representative sequence, that of sweet potato, was chosen for secondary structure prediction and construction of a homology model using the structure of pig

PAP (Guddat et al., 1999) (*Ssc*, Table 1). A sequence alignment between the sweet potato and pig sequences is shown in Fig. 2. The sequence of the 55 kDa red kidney bean enzyme is also included for comparison. An homology model for sm*Iba* was constructed using the alignment together with the coordinates for the pig enzyme (Guddat et al., 1999). In Fig. 3 ribbon diagrams of both structures are shown. High molecular weight red kidney bean PAP consists of two domains, an N-terminal domain and a catalytic C-terminal domain (Klabunde et al., 1996). As in the pig enzyme (Guddat et al., 1999), the N-terminal domain is absent from the low molecular weight plant enzyme.

All known PAPs are glycosylated. A potential glycosylation site for the low molecular weight sweet potato enzyme is located at residue N25 (indicated by an asterisk in Fig. 2). The most obvious difference between the published pig and 55 kDa red kidney bean structures in the immediate vicinity of the binuclear centres is the presence of two loops in the pig enzyme (Fig. 3) (Guddat et al., 1999). One of those (loop I, comprising residues 142–159 in the pig enzyme) is also present in the sweet potato model (residues 142–155; Fig. 3). However, in the pig enzyme there is a disulphide bridge between Cys 142 and Cys 200, which is absent from the sweet potato enzyme. The other loop (loop II; residues 18–24 in the
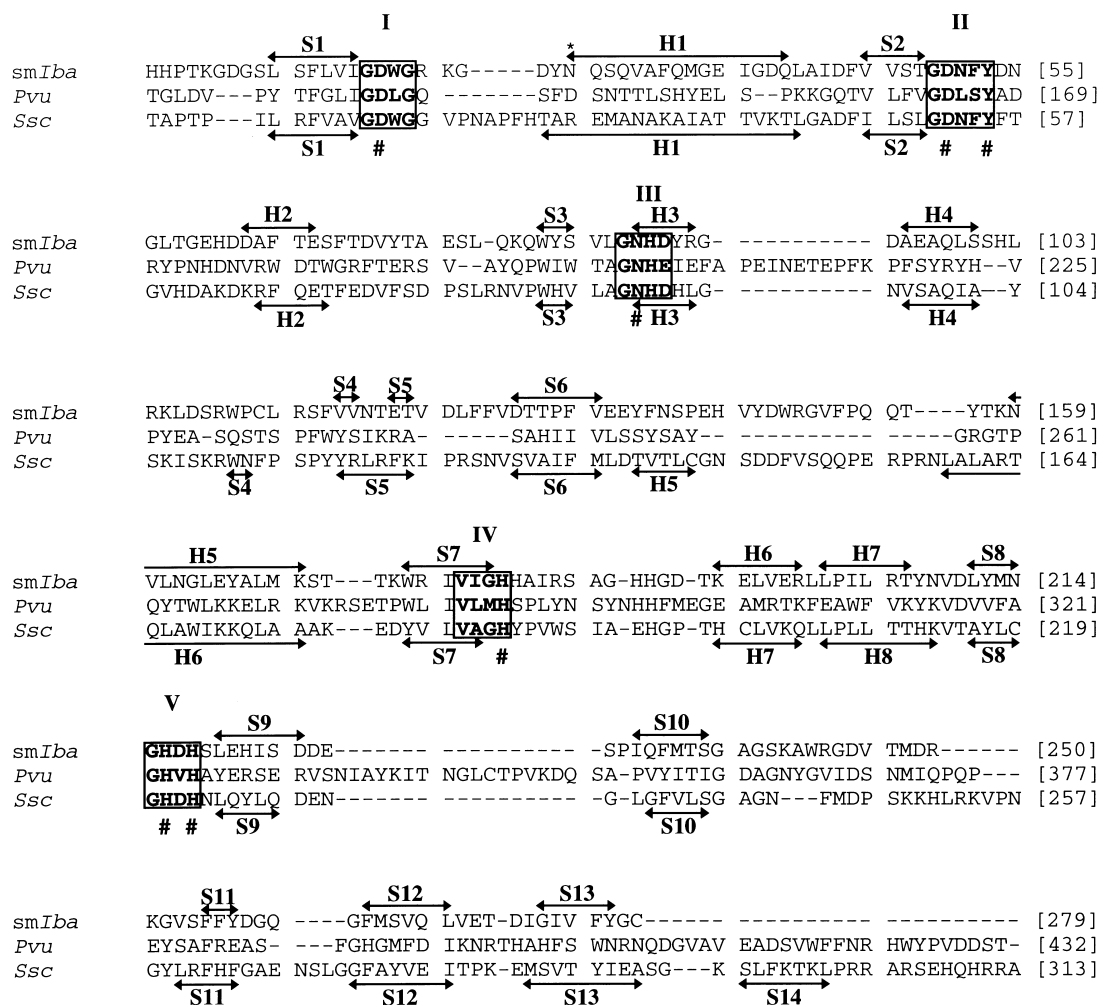
```
                              I                                                          II
                   S1                        H1                              S2
smIba    HHPTKGDGSL SFLVIGDWGR KG-----DYN QSQVAFQMGE IGDQLAIDFV VSTGDNFYDN [55]
Pvu      TGLDV---PY TFGLIGDLGQ -------SFD SNTTLSHYEL S--PKKGQTV LFVGDLSYAD [169]
Ssc      TAPTP---IL RFVAVGDWGG VPNAPFHTAR EMANAKAIAT TVKTLGADFI LSLGDNFYFT [57]
                   S1       #                 H1                        S2    #   #

              H2                         S3        III                         H4
                                                   H3
smIba    GLTGEHDDAF TESFTDVYTA ESL-QKQWYS VLGNHDYRG- ---------- DAEAQLSSHL [103]
Pvu      RYPNHDNVRW DTWGRFTERS V--AYQPWIW TAGNHEIEFA PEINETEPFK PFSYRYH--V [225]
Ssc      GVHDAKDKRF QETFEDVFSD PSLRNVPWHV LAGNHDHLG- ---------- NVSAQIA--Y [104]
              H2                         S3       # H3                         H4

              S4  S5               S6
smIba    RKLDSRWPCL RSFVVNTETV DLFFVDTTPF VEEYFNSPEH VYDWRGVFPQ QT----YTKN [159]
Pvu      PYEA-SQSTS PFWYSIKRA- -----SAHII VLSSYSAY-- ---------- -----GRGTP [261]
Ssc      SKISKRWNFP SPYYRLRFKI PRSNVSVAIF MLDTVTLCGN SDDFVSQQPE RPRNLALART [164]
              S4             S5               S6               H5

              H5                S7        IV                H6         H7        S8
smIba    VLNGLEYALM KST---TKWR IVIGHHAIRS AG-HHGD-TK ELVERLLPIL RTYNVDLYMN [214]
Pvu      QYTWLKKELR KVKRSETPWL IVLMHSPLYN SYNHHFMEGE AMRTKFEAWF VKYKVDVVFA [321]
Ssc      QLAWIKKQLA AAK---EDYV IVAGHYPVWS IA-EHGP-TH CLVKQLLPLL TTHKVTAYLC [219]
              H6                S7       #              H7         H8        S8

         V                 S9                          S10
smIba    GHDHSLEHIS DDE------- ---------- -SPIQFMTSG AGSKAWRGDV TMDR------ [250]
Pvu      GHVHAYERSE RVSNIAYKIT NGLCTPVKDQ SA-PVYITIG DAGNYGVIDS NMIQPQP--- [377]
Ssc      GHDHNLQYLQ DEN------- ---------- -G-LGFVLSG AGN---FMDP SKKHLRKVPN [257]
         # #            S9                          S10

              S11            S12              S13
smIba    KGVSFFYDGQ ----GFMSVQ LVET-DIGIV FYGC------ ---------- ---------- [279]
Pvu      EYSAFREAS- ---FGHGMFD IKNRTHAHFS WNRNQDGVAV EADSVWFFNR HWYPVDDST- [432]
Ssc      GYLRFHFGAE NSLGGFAYVE ITPK-EMSVT YIEASG---K SLFKTKLPRR ARSEHQHRRA [313]
              S11            S12              S13              S14
```

Fig. 2. Sequence alignment between the low molecular weight sweet potato purple acid phosphatase (sm*Iba*), the high molecular weight (55 kDa) red kidney bean enzyme (*Pvu*) and pig purple acid phosphatase (*Ssc*). Secondary structure elements (S: β-sheets; H: α-helix), as predicted (sm*Iba*) or deduced from the structure (*Ssc*) are indicated by bars and numbered consecutively. Numbers in brackets at the end of each line indicate the number of residues. The five conserved regions (boxed) which contain the seven invariant metal-ligating amino acids (#) are also shown. The signal peptides are not included.

pig enzyme), which may be involved in substrate binding (Guddat et al., 1999), is absent from both plant enzymes.

4. Discussion

We have discovered four new full-length cDNA sequences and 10 ESTs encoding putative PAPs using a consensus sequence derived from previously known PAPs. Homologous full-length cDNA sequences have been cloned from tuber tissue of sweet potato (sm*Iba*) and leaf tissue from *Arabidopsis thaliana* (sm*Ath*), soybean (sm*Gma*) and red kidney bean (sm*Pvu*) (Table 5). The full-length cDNAs and all of the ESTs (Table 2) resemble the Easter lily sequence and also therefore the mammalian sequence. The low molecular weight form is therefore present in a variety of plants, and co-exists with the high molecular weight form in some plants

including *A. thaliana*, sweet potato, soybean and red kidney bean. The observation of multiple EST sequences with high similarity to the low molecular weight form in *A. thaliana* and rice databases (Table 2) suggests the existence of several genes in each species. Similar observations of multiple forms of the high molecular weight enzymes have been made in sweet potato and *A. thaliana* (see Table 1). Only one low molecular weight cDNA sequence was, however, cloned from *A. thaliana* leaf tissue suggesting that the expected isoforms may be expressed in a tissue-specific manner, as is the Easter lily gene which is pollen-specific (Kim and Gynheung, 1996).

Despite significant sequence similarities between mammalian and putative low molecular weight plant PAPs (Tables 1 and 6), it remains to be shown that the latter encode proteins which contain a binuclear centre and exhibit phosphatase activity, since none of the
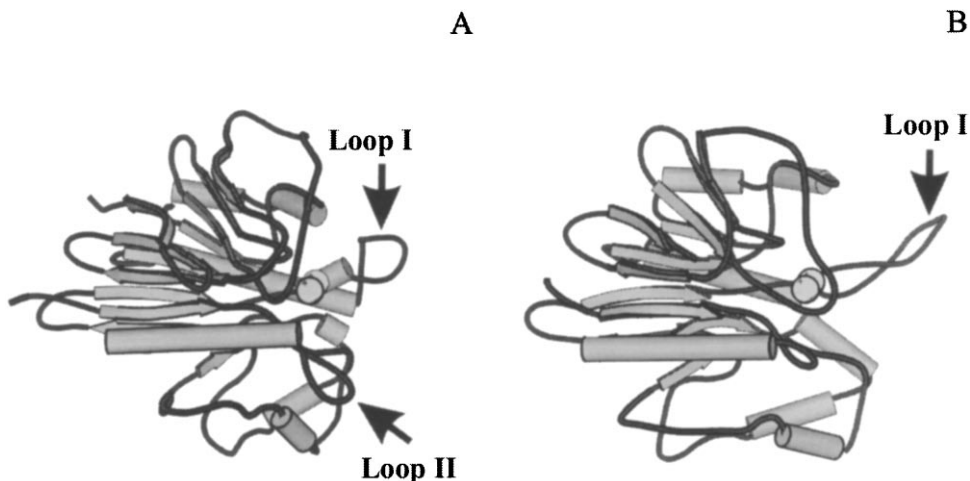
Fig. 3. Ribbon diagrams of the three-dimensional structure of pig purple acid phosphatase (A) and the model of sweet potato enzyme (B). The structures were superimposed with an rmsd of 0.8 Å. The metals of the binuclear centre are drawn as spheres. Loops I and II are described in the text. The illustration was generated using MOLSCRIPT version 2.0.1 (Kraulis, 1991).

proteins has been isolated or expressed. Prediction of the secondary structure of the sweet potato enzyme (Fig. 2) supports the presence of the same core structure as in pig and red kidney bean PAPs. Since all seven amino acids involved in metal binding are not only totally conserved, but also appear in the same order (Tables 1, 4 and 5; Fig. 2), it is highly likely that the low molecular weight enzymes fold in a manner similar to known PAPs.

In all known PAPs the chromophoric site is occupied by FeIII. All mammalian enzymes also have a redox-active FeII in the divalent site. The 55 kDa plant PAPs have been shown to contain Zn or Mn (Beck et al., 1986; Durmus et al., 1999; Schenk et al., 1999) as the divalent metal ion. Hence, the redox potential of these enzymes will differ significantly from that of mammalian PAPs. In addition to their phosphatase activity, the mammalian enzymes may function as iron carriers (Nuttleman and Roberts, 1990) or generators of reactive oxygen species in a Fenton-like reaction (Sibille et al., 1987). Based on the sequence similarities, we speculate that the low molecular weight plant enzymes may contain an FeIII–FeII centre, and that their role(s) may therefore differ from that of the 55 kDa enzymes. The additional N-terminal domain is likely to contribute additional functions to the 55 kDa subgroup of plant PAPs. It is now important to express and purify the low molecular weight protein in sufficient quantities to allow physical and chemical analysis.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Beck, J.L., McConaghie, L.A., Summors, A.C., Arnold, W.N., de Jersey, J., Zerner, B., 1986. Properties of a purple acid phosphatase from red kidney bean: a zinc–iron metalloenzyme. Biochim. Biophys. Acta 869, 61–68.

Beh, C.T., Ferrari, D.C., Chung, M.A., McGhee, J.D., 1991. An acid phosphatase as a biochemical marker for intestinal development in the nematode Caenorhabditis elegans. Dev. Biol 147, 133–143.

Cassady, A.I., King, A.G., Cross, N.C.P., Hume, D.A., 1993. Isolation and characterization of the genes encoding mouse and human type-5 acid phosphatase. Gene 130, 201–207.

Duff, S.M., Sarath, G., Plaxton, W.C., 1994. The role of acid phosphatases in plant phosphorus metabolism. Physiol. Plant 90, 791–800.

Durmus, A., Eicken, C., Sift, B.H., Kratel, A., Kappl, R., Hüttermann, J., Krebs, B., 1999. The active site of purple acid phosphatase from sweet potatoes (Ipomoea batatas). Eur. J. Biochem. 260, 709–716.

Ek-Rylander, B., Bill, P., Norgård, M., Nilsson, S., Andersson, G., 1991. Cloning, sequence and developmental expression of a type 5 tartrate-resistant acid phosphatase of rat bone. J. Biol. Chem. 266, 24684–24689.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Felsenstein, J., 1989. PHYLIP version 3.4, University of Washington, Seattle, WA, USA.

Guddat, L.W., McAlpine, A.S., Hume, D.A., Hamilton, S., de Jersey, J., Martin, J.L., 1999. Crystal structure of mammalian purple acid phosphatase. Structure 7, 757–767.

Hayman, A.R., Jones, S.J., Boyde, A., Foster, D., Colledge, W.H., Carlton, M.B., Evans, M.J., Cox, T.M., 1996. Mice lacking

tartrate-resistant acid phosphatase (Acp5) have disrupted endohondral ossification and mild osteoporosis. Development 122, 3151–3162.

Hunt, D.F., Yates III, J.R., Shabanowitz, J., Zhu, N.Z., Zirino, T., Averill, B.A., Daurat-Larroque, S.T., Shewale, J.G., Roberts, R.M., Brew, K., 1987. Sequence homology in the metalloproteins, purple acid phosphatase from beef spleen and uteroferrin from porcine uterus. Biochem. Biophys. Res. Commun. 144, 1154–1160.

Jameson, B.A., Wolf, H., 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosci. 4, 181–186.

Ketcham, C.M., Roberts, R.M., Simmen, R.C.M., Nick, H.S., 1989. Molecular cloning of the type 5, iron-containing, tartrate-resistant acid phosphatase from human placenta. J. Biol. Chem. 264, 557–563.

Kim, S.R., Gynheung, A., 1996. Isolation and characterisation of a pollen-specific cDNA clone from Easter lily. J. Plant Biol. 39, 197–202.

Klabunde, T., Stahl, B., Suerbaum, H., Hahner, S., Karas, M., Hillenkamp, F., Krebs, B., Witzel, H., 1994. The amino acid sequence of the red kidney bean Fe(III)–Zn(II) purple acid phosphatase – determination of the amino acid sequence by a combination of matrix-assisted laser desorption ionization mass spectrometry and automated Edman sequencing. Eur. J. Biochem. 226, 369–375.

Klabunde, T., Sträter, N., Krebs, B., Witzel, H., 1995. Structural relationship between the mammalian Fe(III)–Fe(II) and the Fe(II)–Zn(II) plant purple acid phosphatases. FEBS Lett. 367, 56–60.

Klabunde, T., Sträter, N., Fröhlich, R., Witzel, H., Krebs, B., 1996. Mechanism of Fe(III)–Zn(II) purple acid phosphatase based on crystal structures. J. Mol. Biol. 259, 737–748.

Klabunde, T., Krebs, B., 1997. The dimetal center in purple acid phosphatases. Struct. Bond. 89, 177–198.

Kraulis, P.J., 1991. MOLSCRIPT – a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.

Kumar, S., Tamura, K., Nei, M., 1993. Molecular Evolutionary Genetics Analysis, version 1.0. Pennsylvania State University, University Park, PA, USA.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

LeBansky, B.R., McKnight, T., Griffing, L.R., 1992. Purification and characterization of a secreted purple phosphatase from soybean suspension culture. Plant Physiol. 99, 391–395.

Mullaney, E.J., Daly, C.B., Ehrlich, K.C., Ullah, A.H.J., 1995. The *Aspergillus niger* (*ficuum*) aphA gene encodes a pH 6.0-optimum acid phosphatase. Gene 162, 117–121.

Mullaney, E.J., Ullah, A.H.J., 1998. Conservation of the active site motif in *Aspergillus niger* (*ficuum*) pH 6.0 optimum acid phosphatase and kidney bean purple acid phosphatase. Biochem. Biophys. Res. Commun. 243, 471–473.

Nuttleman, P.R., Roberts, R.M., 1990. Transfer of iron from uteroferrin (purple acid phosphatase) to transferrin related to acid phosphatase activity. J. Biol. Chem. 265, 12192–12199.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.

Sarkar, A., Caddick, M.X., Bignell, E., Tiburn, J., Arst, H.N.J., 1996. Regulation of gene expression by ambient pH in *Aspergillus*: genes expressed at acid pH. Biochem. Soc. Trans. 24, 360–363.

Schenk, G., Ge, Y., Carrington, L.E., Wynne, C.J., Searle, I.R., Carroll, B.J., Hamilton, S., de Jersey, J., 1999. Binuclear metal centres in plant purple acid phosphatases: Fe–Mn in sweet potato and Fe–Zn in soybean. Arch. Biochem. Biophys. 370, 183–189.

Sibille, J.C., Doi, K., Aisen, P., 1987. Hydroxyl radical formation and iron-binding proteins. J. Biol. Chem. 262, 59–62.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Ullah, A.H.J., Cummins, B.J., 1988. *Aspergillus ficuum* extracellular pH 6.0 optimum acid phosphatase: purification, N-terminal amino acid sequence, and biochemical characterisation. Prep. Biochem. 18, 37–65.

Vriend, G., 1990. WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8, 52–56.