Conservation analysis and structure prediction of the protein serine/threonine phosphatases: Sequence similarity with diadenosine tetra-phosphatase from *Escherichia coli* suggests homology to the protein phosphatases

Geoffrey J. Barton[1], *Patricia T. W. Cohen*[2] *and David Barford*[3]

[1]Laboratory of Molecular Biophysics
University of Oxford
Rex Richards Building
South Parks Road
Oxford OX1 3QU, U.K.

[2]Medical Research Council Protein Phosphorylation Unit
Department of Biochemistry
University of Dundee
Dundee DD1 4HN, Scotland, U.K.

[3]W. M. Keck Structural Biology Laboratory
Cold Spring Harbor Laboratory
Cold Spring Harbor
P.O. Box 100
N.Y. 11724, U.S.A.

Subdivision: Protein Chemistry and Structure

[1]Corresponding author: Dr. G. J. Barton
Tel: (+44) 865 275368
Fax: (+44) 865 510454
e-mail: geoff@biop.ox.ac.uk

This paper appeared in: European Journal of Biochemistry, (1994), 220, 225-237

Contents

1 Summary

A multiple sequence alignment of 44 serine/threonine specific protein phosphatases has been performed. This reveals the position of a common conserved catalytic core, the location of invariant residues, insertions and deletions. The multiple alignment has been used to guide and improve a consensus secondary structure prediction for the common catalytic core. The location of insertions and deletions has aided in defining the positions of surface loops and turns. The prediction suggests that the core protein phosphatase structure comprises two domains, the first has a single β–sheet flanked by α–helices, while the second is predominantly α–helical. Knowledge of the core secondary structures provides a guide for the design of site directed mutagenesis experiments that will not disrupt the native phosphatase fold. A sequence similarity between eukaryotic serine/threonine protein phosphatases and the *Escherichia coli* diadenosine tetraphosphatase has been identified. This extends over the N-terminal 100 residues of bacteriophage phosphatases and *Escherichia coli* diadenosine tetra-phosphatase. Residues which are invariant amongst these classes are likely to be important in catalysis and protein folding. These include Arg 92, Asn 138, Asp 59, 88, Gly 58, 62, 87, 93, 137, His 61, 139 and Val 90 and fall into three clusters with the consensus sequences GD[IVTL]HG, GD[LYF]V[DA]RG and GNH, where brackets surround alternative amino acids. The first two consensus sequences are predicted to fall in the β–α–and β–β–loops of a β–α–β–β–secondary structure motif. This places the predicted phosphate-binding site at the N-terminus of the α–helix where phosphate binding may be stabilised by the α–helix dipole.

2 Abbreviations

PP1: The protein phosphatase 1 catalytic subunit and all its isoforms.

PP2A: The protein phosphatase 2A catalytic subunit and all its isoforms (designated PP2 in the human genome nomenclature).

PP2B: The protein phosphatsae 2B catalytic subunit and all its isoforms (also termed calcineurin A; designated PP3 in the human genome nomenclature).

PP1-like: All protein phosphatases that are more similar in sequence to PP1 than to PP2A and PP2B.

PP2A-like: All protein phosphatases that are more similar in sequence to PP2A than to PP1 and PP2B.

PP2B-like: All protein phosphatases that are more similar in sequence to PP2B than to PP1 and PP2A.

3 Introduction

Protein phosphorylation is important in controlling a variety of biological processes: metabolism, cell differentiation and proliferation, gene expression, transport, locomotion and memory, [1, 2]. The activities of proteins controlled by protein phosphorylation are extremely varied and include enzymes, membrane receptors, transport proteins, ion pumps and proteins mediating DNA replication, transcription and translation. In eukaryotes, reversible phosphorylation occurs predominantly on serine, threonine and tyrosine residues, (reviewed by [3]), although recently a histone H4 protein histidine kinase has been reported [4]. Signal transduction in prokaryotes is also mediated by phosphorylation on serine, threonine, histidine and aspartate residues [5], although no instances of tyrosine phosphorylation have been reported [6].

Protein kinases and phosphatases catalyse protein phosphorylation and dephosphorylation, respectively. In eukaryotes most of these enzymes show either specificity for serine and threonine, or tyrosine residues. However, a number of dual specificity protein kinases and phosphatases have recently been reported, (reviewed by [7]). All protein kinases are related in sequence and therefore belong to a single family, that has presumably arisen by gene duplication from a common ancestor.

Serine/threonine protein phosphatases are classified into 4 major classes according to substrate specificity, metal ion dependence and sensitivity to phosphatase inhibitors; PP1, PP2A, PP2B and PP2C, reviewed by [2, 8]. PP1 and PP2A, purified from physiological sources, are active independent of metal ions whereas PP2B is Ca^{2+}-calmodulin dependent and PP2C is Mg^{2+}stimulated.

The isolation and sequencing of cDNA clones of protein phosphatases from a number of tissues of various species has shed light on the structure and evolutionary relationships between these enzymes. Complementary DNA cloning [9, 10] polymerase chain reactions [11] and mutant analysis [12] have allowed the isolation of novel forms of phosphatases belonging to the PP1/PP2A/PP2B family. The primary structures indicate that PP1, PP2A and PP2B share a common catalytic core of approximately 280 residues which has no relationship to PP2C and the more recently discovered protein tyrosine phosphatases, [13, 10].

Techniques for the prediction of protein structure fall into two main classes. Molecular modelling methods (e.g. see [14]) use the known three dimensional structure of a homologue as a scaffold on which to base the prediction. Such predictions can be very reliable and may help explain differing substrate specificities

and suggest the regions of the protein that would best be modified by site-directed mutagenesis (e.g. see [15]). In the absence of a structural homologue, the accuracy of prediction has historically been very low. Methods to predict the secondary structure (α–helix, β–strand or loop) from a single protein sequence give at best 64% accuracy (e.g. see [16]), and few methods assign degrees of confidence to each predicted region. However, significant improvements in the prediction of protein secondary structure have recently been obtained through the use of multiply aligned sequences [17, 18, 19, 20, 21]. Multiply aligned sequences can also allow the accurate identification of residues buried in the protein core [19, 20, 22].

In this paper, the common cores of 44 eukaryotic, two bacteriophage and one bacterial protein phosphatase are analysed by residue conservation and a novel combination of secondary structure prediction methods. The use of multiple sequence data allows a more accurate picture of the protein secondary structure to be determined by accurately defining the positions of insertions and deletions which almost invariably occur between secondary structural elements [23], and by identifying patterns of residue conservation consistent with α–helix and β–strand. In addition, the multiple alignment characterises invariant residues across the eukaryotes and between eukaryotes and bacteriophage. Invariant residues are likely to be important in catalysis through phosphate binding, in forming a putative phosphoryl-enzyme intermediate and also in protein folding. The characterization of invariant residues allows candidate residues which perform these functions to be identified and, when coupled with the secondary structure prediction, provides a basis for designing site-directed mutagenesis experiments.

4 Methods

4.1 Source of Sequences and Multiple Alignment

Forty four complete or nearly complete eukaryotic phosphatase sequences were gathered from the literature, NBRF-PIR version 33 databank and unpublished results (see Figure 1[1] for references). The sequences were first compared pairwise using the Needleman and Wunsch algorithm [24], followed by automatic multiple alignment [25, 26]. The resulting alignment clearly showed the common region of the phosphatases, but the alignment contained inconsistencies in the long N- and C- terminal extensions that are only seen in some members of the family.

[1] Figure1.ps

Accordingly, a central segment was identified in PP1 α–(rabbit) starting at residue 21 for 271 residues. This polypeptide was used to locate the common catalytic core in the other 43 phosphatase sequences. The alignment procedure was then repeated on these core sequences. Pairwise standard deviation (S.D.) scores for the core sequences obtained using the Needleman and Wunsch [24] algorithm were all above 25.0, indicating that automatic multiple alignment of the core regions would yield an accurate alignment within common secondary structural regions [26]. After multiple alignment, small numbers of residues were trimmed from the N and C terminii of some of the initially identified core sequences. The resulting multiple alignment is shown in Figure 1[2].

A scan of the NBRF-PIR version 33 databank was performed using the alignment shown in Figure 1[3], and an adaptation of the Smith Waterman [27] local similarity algorithm [28] (Program SCANPS - GJB unpublished). This scan identified a possible match with *E. coli* diadenosine tetra-phosphatase. The bacteriophage phosphatase sequences and diadenosine tetra-phosphatase sequences were then aligned with a subset of the eukaryotic phosphatase sequences as shown in Figure 2[4].

4.2 Conservation Analysis and Secondary Structure Prediction

Conservation analysis was performed with the aid of the AMAS (Analysis of Multiply Aligned Sequences) program [29]. AMAS allows physico chemical properties to be assigned to each amino acid type and the conservation of these properties to be characterised for each position within a sequence alignment. AMAS also allows sub-groups of sequences to be defined so that the similarities and differences between sub-families can be rapidly discerned.

The turn prediction algorithms of Wilmot and Thornton [30] and Rose [31] were applied to each sequence in the alignment. Predictions of α–helix and β–strand were performed using the Robson [32], Lim [33] and Chou & Fasman [34] methods. For N sequences this gives $2N$ predictions for turn/loop, $3N$ predictions for α–helix and $3N$ predictions for β–strand at each aligned position. The total possible number of positive turn predictions at a position is $2N$, similarly, the total possible α–helix or β–strand prediction at a position is $3N$. A "consensus"

[2] Figure1.ps
[3] Figure1.ps
[4] Figure2.ps

secondary structure prediction was obtained by dividing the total number of positive predictions for each state at each position by these maxima. A preliminary secondary structure for the position was then assigned by taking the state with the highest fraction. For example, if we have 10 sequences, at each position there will be 20 possible turn predictions, 30 possible helix predictions and 30 strand predictions. If there are actually 15, 15, 20, then the fractions are turn:15/20 = 0.75, helix: 15/30 = 0.5, strand: 20/30 = 0.67; the position would therefore be assigned to turn. This preliminary prediction was then interpreted in the light of the conservation patterns seen across the complete alignment (Figure 1[5]). All alignment figures were prepared using the ALSCRIPT program [35].

5 Results

5.1 Clustering of sequences

Pairwise sequence comparison of the core phosphatase sequences shows that the sequences cluster into four distinct groups. The sequences within each group show greater similarity to each other than they do to other phosphatase sequences. The groupings are: PP1-like sequences, PP2A-like and PP2B-like. Human PP5 (A.E. McPartlin, H.M. Barker and P.T.W. Cohen, in preparation) and *S.cerevisiae* PPT [36] form a fourth distinct group which shows least similarity to all other phosphatase sequences in this region. Table 1 lists the range of sequence identity seen within and between the classes in the phosphatase domain shown in Figure 1[6]. The three major groups are consistent with known physico-chemical properties of these phosphatases. In addition PP2B-like structures all possess long C-terminal extensions that bind $Ca^{2+}$and calmodulin, whilst all PP2A-like and most PP1-like structures have no long N or C-terminal extensions. The exceptions in the PP1-like group are *S.cerevisiae* PPQ [37] PPZ1 and PPZ2 [38] which have long serine rich N-terminal extensions. PP5 and PPT possess long N-terminal extensions with a repeat structure. The clustering performed here is sufficient to provide a guide for structure prediction. However, more detailed considerations of the evolutionary relationships between the full-length serine/threonine phosphatases will be presented elsewhere.

[5]Figure1.ps
[6]Figure1.ps

Figure 1[7] illustrates a multiple alignment of the common core of the eukaryotic phosphatase sequences; the four groupings are delineated by horizontal lines. The boxing and shading reflects conservation of physico chemical properties within each group and across the entire set of sequences (see legend for details).

5.2 Similarity to Bacteriophage Protein Phosphatases and *E. coli* Tetra-Phosphatase

Scanning the sequence databank with an alignment of sequences is a sensitive method to detect weakly similar sequences belonging to the same family [25, 39]. A scan using the alignment in Figure 1[8] ranked the *E. coli* protein diadenosine tetra-phosphatase [40] within the low scoring "tail" of known phosphatase sequences. A scan with the diadenosine tetra-phosphatase sequence showed the lambda protein phosphatase and *Drosophila melanogaster* PPY to be the most similar proteins in the databank, followed by the *Trypanosoma brucei* PP1.

Figure 2[9] illustrates an alignment of the *E. coli* diadenosine tetra-phosphatase with the protein phosphatases from ϕ80, and lambda, and a representative set of phosphatases from Figure 1[10]. The ϕ80, lambda and tetra-phosphatase sequences share greatest similarity with eukaryotic protein phosphatases in the first 109 positions of the alignment which are shown in Figure 2[11]. The similarity following position 109 is very low and sequence alignments in this region are ambiguous.

The most striking regions of conservation between the sequences in Figure 2[12] start at positions 34, 64 and 96 (58, 87 and 137 on Figure 1[13]). These similarities suggest that the *E. coli* diadenosine tetra-phosphatase, ϕ80 and eukaryotic protein phosphatases have diverged from a common evolutionary ancestor.

[7] Figure1.ps
[8] Figure1.ps
[9] Figure2.ps
[10] Figure1.ps
[11] Figure2.ps
[12] Figure2.ps
[13] Figure1.ps

5.3 Invariant residues, secondary structure and insertions/deletions

The positions of invariant residues within the eukaryotic alignment and between the eukaryotic sequences and the bacteriophage and *E. coli* diadenosine tetraphosphatase sequence are listed in Table 2.

The most accurate secondary structure predictions from an aligned family require sequences that are possible to align with little ambiguity, yet are sufficiently variable to highlight the positions most important to the common fold of the family. Unfortunately, whilst the eukaryotic phosphatase sequences analysed in this study are sufficiently similar to align well, in some regions the similarity is rather too great to enhance dramatically, a conventional single sequence secondary structure prediction. Bearing this limitation in mind, the summary secondary structure prediction is illustrated at the base of Figure 1[14].

All secondary structure predictions are subject to error, however, given an accurate alignment it is possible to assign the secondary structure of some regions with greater confidence than others. As a rough guide, the order of confidence in prediction is: loop (where insertions/deletions occur) $>$ loop (conserved Gly/Pro/Hydrophilic) $>$ surface helix (with clear hydrophobic patterns) $>$ surface strand (with clear hydrophobic patterns) $>$ buried strand (short run of conserved hydrophobic residues). Accordingly, we here describe the arguments supporting the prediction of loop, helix and strand and the reasons for ambiguity where this is shown in Figure 1[15].

5.3.1 Loop Predictions

Insertions and deletions are normally only tolerated at surface non-secondary structure positions of the native protein structure [23]. Accordingly, the location of insertions and deletions in an accurate multiple alignment is a strong indicator of a surface loop, or non-core secondary structure in the native three dimensional structure. There are 16 regions predicted as loop. Of these 10 correspond to the location of insertions/deletions either within the eukaryotic phosphatases, or when the phage and *E. coli* proteins are included (loops at: 6-22, 45-53, 74-82, 87-93, 98-117, 128-131, 151-159, 188-195, 223-269, 298-315). Further support for loop prediction is gained from the observed conservation of Proline and Glycine and

[14] Figure1.ps
[15] Figure1.ps

a predominance of polar residues. For example, between 45-53 there are 4 and 6 residue insertions in PP5 and *S. cerevisiae* PPT, while there is a conserved proline at 53. The 6 predicted loops that do not include insertions/deletions occur at 58-63,136-143, 181-183, 207-215, 275-277 and 284-287. Loop 58-63 includes residues invariant throughout all known phosphatase sequences (Gly 58, Asp 59, His 61 and Gly 62) and may therfore form part of the catalytic site. Similarly, loop 136-143 contains conserved basic and acidic residues (Arg 136, Gly 137, Asn 138, His 139 and Glu 140) and may also contribute to the catalytic site.

5.3.2 Helix Predictions

Five regions are assigned to the helical conformation (23-44, 64-73, 196-205, 278-283, 288-296), with a further 3 possible helical regions where the prediction is ambiguous (see *Ambiguous predictions* below). A helix that has one side exposed to solvent and one that packs against the hydrophobic protein core often has hydrophobic residues on the buried face. Accordingly, conserved hydrophobic residues seen at an $i, i+3, i+4, i+7$ spacing give strong supporting evidence for a predicted helix. The predicted helix at 23-44 shows this type of pattern, with conserved hydrophobics at 24, 27, 28, 31, 34 and 35. The pattern continues with conserved hydrophobics at 41, 42 and 44, but the conserved proline at 39 suggests that the helix may contain a kink. Similar patterns are seen for the predicted helices at 64-73, 160-180 (where a kink may occur at 175), and 278-283.

5.3.3 Beta Strand Predictions

Eight β–strands are predicted for the phosphatase domain. These are at 54-58, 83-86, 94-98, 132-135, 184-187, 270-274, 316-322, 325-329. A surface β–strand with one face exposed to solvent often exhibits a pattern of alternate conserved hydrophobic and hydrophilic residues. Such patterns are seen for strand 54-58 where 54, and 56 are conserved hydrophobic; strand 94-98 where hydrophilic residues are conserved at 94, 96 and 98; and 270-274 with conserved hydrophobics at 270, 272, and 274. Short conserved runs of hydrophobic residues are indicative of buried β–strands. This pattern is shown for the predicted strands at 83-86, 132-135 and 184-187. A further three possible β–strands are described under *Ambiguous predictions* below.

5.3.4 Ambiguous Predictions

The assingnment of secondary structure to either loop, helix or strand was performed by combining the prediction profiles shown in Figure 1[16] with an analysis of conservation patterns. The strands and helices described in the previous sections show reasonable consistency for one predicted state. However, three regions do not show such consistency and are therefore difficult to assign either to β–or α–conformation. The region 118-128 starts with a cluster of conserved hydrophobic residues (118-124) indicative of a buried β–strand, this is continued by an alternating hydrophobic/hydrophilic pattern suggesting that the buried β–strand may extend into a surface strand (125-127). The loop 128-131 conserves proline at 129 and the following region is predicted β–(132-135). Type II (proline) turns often occur between consecutive antiparallel β–strands, so the conservation of proline lends support to the prediction of β–structure on either side of the 132-135 loop. However, this conflicts with the prediction profiles that suggest this region is an α–helix.

At 144-150 the prediction profiles suggest β–strand, but conserved hydrophobics at 144, 147, 148 and 150 indicate a surface α–helix. The tolerance of Gly at 146 may swing the prediction in favour of β–rather than α–structure.

The variable sequence composition at 330-C-terminus make the region difficult to assign to α–or β–. The region might be split into two β–strands (333-339, 347-353), or possibly a single helix if the gap at 342-344 is closed.

6 Discussion

6.1 Invariant Residues Suggest a Common Catalytic Mechanism

The similarities in sequence amongst eukaryotic and bacteriophage phosphatases in Figure 2[17] suggests that they share a common catalytic mechanism. Several PP1, PP2A and PP2B homologues in different species [2] and the novel phosphatases, PP4 (formerly PPX) [41] and PP5 (M. X. Chen, Y.H. Chen and P.T.W. Cohen, in preparation), have been shown to dephosphorylate serine and threonine residues in proteins. The lambda bacteriophage phosphatase ORF221 has been demonstrated

[16] Figure1.ps
[17] Figure2.ps

to possess protein phosphatase activity against glycogen phosphorylase a and PKA phosphorylated casein, [42]. When expressed from a recombinant vector, ORF221 displays phosphatase activity against casein phosphorylated on serine and tyrosine residues [43]. Consequently the 15 residues (Table 1) invariant throughout the eukaryotic and bacteriophage phosphatases are likely to play crucial roles in protein folding and enzyme catalysis.

The novel phosphatases PPQ, PPT, PPV, PPY, PPZ, PPH3, SIT4 have not yet been demonstrated to be active phosphatases. However, virtually all residues that are invariant between PP1, PP2A , PP2B, PP4 and PP5 are also invariant between PPQ, PPT, PPV, PPY, PPZ, PPH3 and SIT4. The exception, serine, is invariant at position 192 for all sequences except PPT where it is a proline, but since serine is not conserved in the bacteriophage sequences it is unlikely to be an essential residue. The *E. coli* tetra-phosphatase removes pyrophosphate from diadenosine tetra-phosphate. Since 14 of the 15 residues, that are invariant throughout eukaryote and bacteriophage phosphatases, are also invariant in *E. coli* tetra-phosphatase (Table 2), it is likely that the *E. coli* phosphatase shares a common catalytic mechanism with the eukaryotic and bacteriophage enzymes. It is not known whether *E. coli* diadenosine tetraphosphatase possesses protein phosphatase activity.

The importance of the invariant residues, Asp 59 and 88 and His 61 and 139 in catalysis is supported recent site directed mutagenesis study of lambda bacteriophage phosphatase ORF221, (J. E. Dixon and S. Zhou, unpublished results). Substitution of Asp 59 for Asn, His 61 for Asp, Asp 88 for Asn and His 139 for Asn abolishes phosphatase activity. In contrast substitution of either Asp 91 for Asn and Glu 140 for Gln reduces activity by 29% and 82%, respectively. Although Asp 91 and Glu 140 are invariant throughout eukaryotic and bacteriophage sequences, Asp 91 is an Ala and Glu 140 is an Asp in the *E. coli* diadenosine tetra-phosphatase sequence, consistent with the notion that these residues are not essential for phosphatase activity.

There are no invariant Arg, Lys, His, Asp or Glu residues after position 296 or before position 58 of the eukaryotic phosphatases (Figure 1[18]). Moreover, before position 58, the sequences of the phage and *E. coli* phosphatases show little similarity to the eukaryotic proteins (Figure 2[19]). Together, this suggests that neither of these regions are likely to be involved directly in either phosphate

[18] Figure1.ps
[19] Figure2.ps

binding or metal ion binding.

6.2 Change in charge sign

The invariant residues discussed above clearly signal positions that are important to the common fold or function of the phosphatases. However, more subtle patterns of residue conservation can point to positions that are important to the specificity of different members of the protein family. For example, in the well known serine proteinases, the physico-chemical properties of the residue at the base of the S1 binding pocket contribute to the differing specificity of trypsin and chymotrypsin for peptide substrates. The residue is an Asp in trypsin-like enzymes which preferentially bind peptides at Lys or Arg, while this position is a Ser in chymotrypsin-like enzymes where the smaller side-chain allows binding of more bulky aromatic residues (e.g. see [44]). Within the protein phosphatase sequences shown in Figure 1[20], there are many positions at which amino-acids showing similar physico-chemical properties are conserved within one of the four sub-families, yet different residues are seen in other sub-families. The most striking positions are those where a charged residue is conserved within a sub-family of sequences (e.g. the PP1s and PP2As), yet is of different charge in other sub-families. Since charged residues are normally seen on the surface of a protein or performing important structural roles in the hydrophobic core, a change in conserved charge at a position suggests that the residue may be important for defining the functional specificity of the protein sub-family. Such positions are summarised in Table 3.

At position 4, a negative charge is predominant in all families except the PP5/PPT (Lys/Arg). At 142, Lys/Arg is conserved in the PP2B and PP2A sub-families, but Asp in the PP5/PPT group; PP1s conserve Ala/Ser at this position. At 151, Lys is conserved in the PP2Bs, but Glu in PP5/PPT, all other sequences have Tyr/Phe at this location. At 165 Glu/Asp is conserved in PP2Bs, but Lys/Arg/His predominate in PP2As and PP1s. At 213, His is conserved in PP2As, Asp/Glu in PP1s, whilst at 303 Arg is conserved in PP2Bs with Glu conserved in the PP1 family.

It is currently not possible to predict functions for these residues, however they may represent good targets for mutagenesis experiments aimed at probing the functional specificity of individual protein phosphatases.

[20] Figure1.ps

6.3 Phosphate binding site

Important structural requirements of the catalytic site of protein serine/threonine phosphatases are to bind serine-phosphate and threonine-phosphate. The results of a survey of the structures of 30 phosphate binding sites in 18 different proteins were reported by Johnson in 1984 [45]. It was shown that phosphate binding sites can be separated into those involved solely in binding and those present at the catalytic site. The latter sites are generally less well defined. Arginine is frequently involved in binding phosphate with lysine less often. The phosphate binding sites are commonly present at the N-terminus of an alpha-helix to benefit from positive helix dipole interactions. Phosphate groups also interact with the side chains of serine, threonine, asparagine and glutamine and with the NH peptide groups of the main-chain. Glycines residues are common at these positions.

Amongst the eukaryotic protein serine/threonine phosphatase sequences (Figure 1[21]), Arg is invariant at positions 92, 136, 207, 268 and 293. However, when the bacteriophage sequences are included only positions 92 and 136 are invariant, and on including the diadenosine tetra-phosphatase sequence only position 92 is invariant. This strongly suggests a role for Arg 92 in phosphate binding and less so for Arg 136. Arg 92 and 136 occur in highly conserved regions of the sequence predicted as loops. Asn 138, 324, 331; Gln 63; Ser 96 and 321 and Lys 125 are invariant amongst the eukaryotic sequences, although only asparagine 138 is conserved throughout all the phosphatase sequences. Asn 138 is therefore a candidate for a phosphate binding residue. There are 11 invariant glycine residues in the eukaryotic sequences, many of which are located close in primary structure to invariant arginine, aspartate and histidine residues, suggesting a role in determining the position of these residues on surface turns and loops. Amongst eukaryotic and bacteriophage phosphatases and the diadenosine tetra-phosphatase sequences, glycine is invariant at positions 58, 62, 87, 93 and 137 and it is possible that one or more of these glycines is involved in phosphate binding.

The secondary structure prediction (Figure 1[22]) suggests that the region between 54 and 98 has the pattern $\beta-\alpha-\beta-\beta-$. This is similar but not identical, to the phosphate-binding structure of the dehydrogenase protein family. In the dehydrogenases, the phosphatate binds near the N-terminus of the first α–helix in a $\beta-\alpha-\beta-\alpha-\beta$–motif where all β–strands are parallel, and phosphate binding is stabilised by the α–-helix dipole [46]. In the phosphatases, if the region 54-

[21] Figure1.ps
[22] Figure1.ps

98 were to fold as a parallel $\beta-\alpha-\beta$–unit with an additional antiparallel strand at the end of the sheet, then the $\beta-\alpha$–loop (58-63) that contains the invariant residue pattern GD[IVLTL]HG, and the $\beta-\beta$–loop which contains the pattern GD[LYF]V[DA]RG would be suitably positioned to form a phosphate binding site with stabilisation from the dipole of the intervening helix (64-73).

6.4 Phosphoryl-Enyzme Intermediate

Little is known about the catalytic mechanism of serine/threonine specific protein phosphatases. Trans-phosphorylation reactions were unable to establish the presence of a phosphoryl-enyzme intermediate, [47]. Alkaline phosphatase catalyses dephosphorylation through a phosphoryl-serine enzyme intermediate, [48], whereas protein tyrosine phosphatases [49] proceed via a phosphoryl-cysteine intermediate. Low molecular mass phosphatases proceed via phosphoryl-cysteine [50] or phosphohistidine [51] intermediates. Recent evidence suggests that purple acid phosphatase may also function through a covalent phosphoryl-enzyme intermediate involving a histidine residue, [52]. However, the absence of invariant cysteines and serines in the protein serine/threonine phosphatases indicates that phosphoryl-cysteine and phosphoryl-serine intermediates do not occur in this family of phosphatases.

The less closely related reaction catalysed by phosphoglucomutase proceeds through a phosphoryl-histidine intermediate, and ATPases proceed via a phosphoryl-aspartate intermediate. Likely candidates for a putative intermediate in the protein serine/threonine phosphatases are therefore the invariant Asp 59 and 88 and His 61 and 139.

6.5 Are metal ions present at the catalytic centre?

Metal ions are involved in the catalytic mechanisms of a number of different phosphatases. Vincent and Averill have shown that the mammalian purple acid phosphatases have one $\mathrm{Fe}^{2+}$and one Fe^{3+}at the active centre, while plant purple acid phosphatases have one Fe^{2+}ion and one Zn^{2+}ion. [52]. Two Zn^{2+}ions and a Mg^{2+}ion are bound at the catalytic site of *E. coli* alkaline phosphatase with the inorganic phosphate substrate complexed by the Zn^{2+}ions [53]. The situation with the protein serine/threonine phosphatases is less clear. Only low levels of iron and zinc (0.2 and 0.1 mol/mol protein, respectively) were found in a preparation which was probably a mixture of PP1 and PP2A [54] , whereas 0.6-0.9 mol/mol of both

iron and zinc were found in a complex of PP2B catalytic and regulatory subunits [55, 56] , data used to suggest that these serine/threonine protein phosphatases are metalloenzymes, with sequence similarities to the purple acid phosphatases. [57].

The property of Mn^{2+}stimulation of some forms of PP1 and PP2A and the Mn^{2+}dependence of the lambda bacteriophage phosphatase [42, 43] suggests a Mn^{2+}binding site in serine/threonine protein phosphatases. However, there is no evidence that metal ions are located at the catalytic site of protein phosphatases or participate in catalysis.

Recent studies have shown that inhibitor 2 acts as a molecular chaperone in the refolding of PP1 catalytic subunit and in carrying out this function can change its Mn^{2+}dependency. Bacterially expressed PP1 alpha and PP1 beta are totally dependent on $Mn^{2+}$for activity. After incubation with inhibitor 2 and reactivation by phosphorylation of inhibitor 2 with glycogen synthase kinase 3 in the presence of Mg-ATP, the expressed PP1 isoforms becomes Mn^{2+}independent [58]. It was further demonstrated that PP1 could be denatured in 6M GuCl and refolded to the fully active state in the presence of 2 mM EGTA, inhibitor 2, glycogen synthase kinase 3 and Mg-ATP. Since the refolding to the fully active Mn^{2+}independent enzymes was performed in the presence of EGTA, Mn^{2+}cannot be an essential feature of the catalytic site of PP1. Instead Mn^{2+}may stimulate the expressed PP1 by maintaining the tertiary structure of the enzyme in a similar conformation to that of the native enzyme. These studies also show that if iron and zinc ions are present at the active centre, they must be bound covalently and thus it is unlikely that they would be lost during purification. They also demonstrate that although Mn^{2+}has been shown to stimulate the activity of certain forms of PP1 and PP2A [59, 60, 61], it is not essential for full activity and indeed there is no evidence for the presence of Mn^{2+}in the native enzyme [54, 55, 56].

Although the exact role of the metal binding site in serine/threonine phosphatases in either catalysis or in defining the protein conformation is unclear, there are several conserved residues that potentially could bind metal ions. Residues known to form ligands at metal ion binding sites include aspartate, asparagine, cysteine, glutamate, glutamine and histidine. The crystal structure of alkaline phosphatase shows that one Zn^{2+}ion is ligated by two His and one Asp while the second $Zn^{2+}$is ligated by one His and two Asps [53]. NMR studies suggests that the 2 histidines, a tyrosine and an acidic residue ligate the 2 Fe^{2+}ions bound to the catalytic site of purple acid phosphatases, [62].

Possible metal binding residues which are conserved throughout the eukaryotic protein serine/threonine phosphatase sequences are Asp 59, 66, 88 and 91; Asn

138, 324 and 331, Glu 140 (Asp in diadenosine tetra-phosphatase) and 153, and His 61, 139, 188 and 295. These residues occur in regions of the sequence predicted to form loops. When the bacteriophage and diadenosine tetra-phosphatase sequences are included only Asp 59 and 88; Asn 138; Glu 140 and His 61 and 139 are conserved.

6.6 Secondary Structure Prediction and Domain Structure

The secondary structure prediction shown in Figure 1[23] suggests that the Ser/Thr phosphatases are made up of both both α–helix and β–strand. The most highly conserved region between eukaryotic and phage phosphatases, and *E. coli* diadenosine tetra-phosphatase comprises the first 143 positions in the alignment (Figure 1[24]). It seems reasonable to suggest that this region makes up a single structural domain in the common phosphatase tertiary structure. The overall predicted secondary structure for the predicted domain is: α–β–*α–β–*β–(β–/α–)β–β–*α–, with "*" indicating the putative phosphate binding loops and parentheses showing uncertainty in the predicted state. This prediction is consistent with a single β–sheet with α–helices packing on either face. Positions 144-356 probably comprise a further structural domain but with a higher proportion of α–helix and the summary secondary structure (β–/α–)α–β–α–(α–)β–α–α–.

In common with all predictions, our prediction for the Ser/Thr phosphatases is likely to contain errors. Accordingly, we have presented the evidence for and against the assigned state of each predicted region. These data will provide a valuable guide for those planning site-directed mutagenesis experiments to probe the functional specificity of the phosphatases. Consideration of our results should avoid uneccesary effort performing mutations that are likely to disrupt the hydrophobic core of the protein. For example, mutations that alter the length of the protein chain should be restricted to regions strongly predicted as loop, while point mutations of conserved hydrophobic residues in predicted secondary structures should be avoided.

Protein Ser/Thr phosphatases are under investigation by X-ray crystallography in several laboratories. The prediction presented here may be helpful in guiding the interpretation of low resolution electron density maps, particularly when fitting

[23] Figure1.ps
[24] Figure1.ps

the protein sequence to the observed density.

Recently, a number of blind tests of secondary structure prediction methods have been performed (for a recent review see [63]). Our prediction for the phosphatases is another in this series. Accordingly, when the first structure of a protein phosphatase is solved, comparison of prediction with experiment will provide further guidance for the development of improved secondary structure prediction methods.

7 Acknowledgements

GJB thanks the Royal Society and Prof. L. N. Johnson for support, and Mr C. D. Livingstone for his assistance with the AMAS program. PTWC thanks the Medical Research Council, U.K, the Cancer Research Campaign, U.K. and the Human Frontier Science Program. DB thanks the American Cancer Society (grant number IRG-153) for support.

8 Figure Legends

8.1 Legend to Figure 1

Multiple alignment of the catalytic core of 44 Protein Phosphatase sequences: H, human; B, bovine; Rb, rabbit; Rt, rat; M, mouse; P, pig; D.m, *Drosophila melanogaster*; S.c, *Saccharomyces cerevisiae*; S.p, *Schizosaccharomyces pombe.*. Reference numbers for the first publication of each complete sequence are shown in square brackets. Numbers in parentheses refer to 1. (A. E. McPartlin, H. M. Barker and P. T. W. Cohen, in preparation) and 2. (M. X. Chen, Y.H. Chen and P. T. W. Cohen, in preparation).

Secondary structure prediction histograms and summary secondary structure prediction obtained by combining the prediction histograms with an analysis of residue conservation patterns (see text). The sequences are divided into four groups or sub-families according to similarity. These sub-families are delineated by horizontal lines. The shading is based upon a physico-chemical property scoring scheme [17]. Shading was calculated allowing up to two gaps at each position, and ignoring the 10% least frequently represented amino acids at each position.

White on black positions are identical residues in all sequences,or where residues are identical within a sub-family (e.g. at position 28, C is conserved in the PP2A sub-family, but not within other sub-families). Light shading shows positions where residue physico-chemical properties are conserved within a sub-family. Italic text indicates positions where a sub-family does not have highly conserved physico-chemical properties. The line marked conservation shows positions that share physico-chemical properties across all sequences on a scale of 0-10 (10 = +) calculated according to Zvelebil *et al.* [17]. The helix, beta and loop histograms plot the frequencies of each state predicted by the three secondary structure and two turn prediction algorithms (see text). The helix, strand and loop summary lines show the positions of predicted secondary structures obtained by combining the information from the histograms with a consideration of residue conservation patterns (see text). Black shading on these summaries indicates the positions of the conserved hydrophobic residues.

8.2 Legend to Figure 2

Alignment of protein phosphatases from $\phi 80$, lambda [64], with *E. coli* diadenosine tetra-phosphatase and representative sequences from Figure 1[25]. The shading protocol is identical to that described in the legend to Figure 1[26]. The line Figure 1[27] Numbers cross-references positions in this alignment with Figure 1[28]. This alignment has been optimised for the sequences that are shown, as a consequence, the alignment between positions 50 and 60 differs slightly from that shown in Figure 1[29]. The Leu at position 43 (position 67 in Figure 1[30]) is shown as an identity since the only sequence without Leu at this position is PPT (Val). Similarly, the Leu at position 80 (position 121 in Figure 1[31]) is also shown as an identity, since the only sequence withou Leu at this position is PPT (Phe).

[25] Figure1.ps
[26] Figure1.ps
[27] Figure1.ps
[28] Figure1.ps
[29] Figure1.ps
[30] Figure1.ps
[31] Figure1.ps

9 Tables

9.1 Table 1

Range of sequence identity seen within and between each class of phosphatase.

	PP2B	PP2A	PP1	PP5/PPT
PP2B	100/59			
PP2A	44.6/36.9	100/53		
PP1	43.9/37.3	51.9/38.4	100/54	
PP5/PPT	41.6/34.6	41.8/35.2	43.4/34.9	44.9/44.9

9.2 Table 2

Positions of conserved (identical) residues within eukaryotic sequence and among eukaryotic, bacteriophage and E. coli diadenosine tetra-phosphatase. Residue 140 is an Asp in diadenosine tetra-phosphatase. () residue is conserved between eukaryotic sequences and diadenosine tetra-phosphatase. Bacteriophage and diadenosine tetra-phosphatase show sequence similarities between positions 58 to 140. The regions of highest conservation are : 58-63, 85-93, 136-140, 188-191, 321-325.

Residue Type	Eukaryotes	Eukaryotes and Bacteriophage	Eukaryotes, bacteriophage diadenosine tetra-phosphatase
Alanine	322		
Arginine	92,136,207,268,293	92,136	92
Asparagine	138,324,331	138	138
Aspartate	59,66,88,91	59,88,91	59,88
Glutamate	140,153	140	140=D*
Glutamine	63		
Glycine	58,62,75,87,93,137	58,62,87,93	58,62,87,93
	189,190,215,269,301	137	137
Histidine	61,139,188,295	61,139	61,139
Lysine	125		(125)*
Phenyalanine	85,150,171,282		
Proline	212,323		
Serine	96,321		(96)*
Tryptophan	222		
Valine	90	90	90
TOTAL	42	15	14

9.3 Table 3 - Changes in sign of potential charge

Position	Predominant Charge			
	PP2B	PP2A	PP1	PP5/PPT
4	-	-	-	+
142	+	+		-
151	+			-
165	-	+	+	
303	+		-	

References

[1] Fischer, E. H. and Krebs, E. G. 1990, *Biochim. Biophys. Acta* **1000**.

[2] Cohen, P. 1989, *Annu. Rev. Biochem.* **58**, 453–508.

[3] Krebs, E.G. 1986, *The Enzymes, (Boyer, P.D. And Krebs, E.G., Eds., 3rd Edit.)* **17**, 3–20.

[4] Huang, J.M., Wei, Y.F., Kim, Y.H., Osterberg, L. and Matthews, H.R. 1991, *J. Biol. Chem.* **266**, 9023–9031.

[5] Stock, J.B., Stock, J.B. and Mottonen, J.M. 1990, *Nature (London)* **344**, 395–400.

[6] Cozzone, A.J. 1988, *Ann. Rev. Microbiol* **42**, 97–125.

[7] Lindberg, R.A., Quinn, A.M. and Hunter, T. 1992, *Trends Biochem. Sci.* **17**, 114–119.

[8] Bollen, M. and Stalmans, W. 1992, *Crit. Rev. Biochem. Mol. Biol.* **27**, 227–281.

[9] Cohen, P.T.W, Brewis, N.D., Hughes, V. and Mann, D.J. 1990, *FEBS Lett.* **268**, 355–359.

[10] Cohen, P.T.W. 1991, *Methods Enzymol.* **201**, 398–408.

[11] Chen, M.X., Chen, Y.H. and Cohen, P.T.W. 1992, *FEBS Lett.* **306**, 54–58.

[12] Arndt, K.T., Styles, C.A. and Fink, G.R. 1989, *Cell* **56**, 527–537.

[13] Fischer, E.H., Charbonneau, H. and Tonks, N.K. 1991, *Science* **253**, 401–406.

[14] Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. and Thornton, J. 1987, *Nature (London)* **326**, 347–352.

[15] Johnson, D. A. and Barton, G. J. 1992, *Protein Science* **1**, 370–377.

[16] Holley, H. L. and Karplus, M. 1989, *Proc. Natl. Acad. Sci. USA* **86**, 152–156.

[17] Zvelebil, M. J. J. M., Barton, G. J., Taylor, W. R. and Sternberg, M. J. E. 1987, *J. Mol. Biol.* **195**, 957–961.

[18] Crawford, I. P., Niermann, T. and Kirchner, K. 1987, *Proteins: Struct., Funct., Genet.* **2**, 118–129.

[19] Barton, G. J., Newman, R. H., Freemont, P. F. and Crumpton, M. J. 1991, *Eur. J. Biochem.* **198**, 749–760.

[20] Russell, R. B., Breed, J. and Barton, G. J. 1992, *FEBS Lett.* **304**, 15–20.

[21] Benner, S. A. and Gerloff, D. 1990, *Adv. Enz. Regul.* **31**, 121–181.

[22] Barton, G. J. and Russell, R. B. 1993, *Nature (London)* **361**, 505–506.

[23] Pascarella, S. and Argos, P. 1992, *J. Mol. Biol.* **224**, 461–471.

[24] Needleman, S. B. and Wunsch, C. D. 1970, *J. Mol. Biol.* **48**, 443–453.

[25] Barton, G. J. 1990, *Methods Enzymol.* **183**, 403–428.

[26] Barton, G. J. and Sternberg, M. J. E. 1987, *J. Mol. Biol.* **198**, 327–337.

[27] Smith, T. F. and Waterman, M. S. 1981, *J. Mol. Biol.* **147**, 195–197.

[28] Barton, G. J. 1993, *Comput. Appl. Biosci.* **9**, 729–734.

[29] Livingstone, C. D. and Barton, G. J. 1993, *Comput. Appl. Biosci.* **9**, 745–756.

[30] Wilmot, A. C. M. and Thornton, J. M. 1988, *J. Mol. Biol.* **203**, 221–232.

[31] Rose, G. D. 1978, *Nature (London)* **272**, 586–591.

[32] Garnier, J., Osguthorpe, D. J. and Robson, B. 1978, *J. Mol. Biol.* **120**, 97–120.

[33] Lim, V.I. 1974, *J. Mol. Biol.* **88**, 873–894.

[34] Chou, P. Y. and Fasman, G. D. 1978, *Adv. Enzymol.* **47**, 45–148.

[35] Barton, G. J. 1993, *Protein Eng.* **6**, 37–40.

[36] Chen, M. X., Chen, Y. H. and Cohen, P. T. W. 1993, *Eur. J. Biochem.*

[37] Chen, M. X., Chen, Y. H. and Cohen, P. T. W. 1993, *Eur. J. Biochem.* **216**, 269–279.

[38] Hughes, V., Muller, A., Stark, M. J. R. and W., Cohen P. T. 1993, *Eur. J. Biochem.* **216**, 269–278.

[39] Gribskov, M., McLachlan, A. D. and D., Eisenberg. 1987, *Proc. Natl. Acad. Sci. USA* **84**.

[40] Blanchin-Roland, S., Blanquet, S., Schmitter, J. M. and Fayat, G. 1986, *Mol. Gen. Genet* **205**, 515–522.

[41] Brewis, N.D., Street, A. J., Prescott, A.R. and Cohen, P.T.W. 1993, *EMBO J.* **12**, 987–996.

[42] Cohen, P.T.W. and Cohen, P. 1989, *Biochem. J.* **260**, 931–934.

[43] Zhou, S., Clemens, J. C., Hakes, D. J., Barford, D. and Dixon, J.E. 1993, *J. Biol. Chem., In press*.

[44] Fersht, A. *Enzyme Structure and Mechanism - Second Edition*, pages 27–29. W. H. Freeman and Company, 1985.

[45] Johnson, L.N. 1984, *Inclusion Compounds, (Atwood, J.L. And Davies, J.E.D. And Macnicol , D. Eds) Academic Press, London* pages 509–569.

[46] ad P. T. van Duijnen, W. G. Hol and Berendsen, H. J. 1978, *Nature (London)* **273**, 443–446.

[47] Martin, B. L. and Graves, D. J. 1993, *Archives Biochemistry Biophysics - Submitted*.

[48] Schwartz, J.H. and Lipmann, F. 1961, *Proc. Natl. Acad. Sci. USA* **47**, 1996–2005.

[49] Guan, K.L. and Dixon, J.E. 1991, *J. Biol. Chem.* **266**, 17026–17030.

[50] Zhang, Z.-Y., Davis, J.P. and Van Etten, R.L. 1992, *Biochemistry* **31**, 1701–1711.

[51] Ostanin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.-M. and Van Etten, R.L. 1992, *J. Biol. Chem.* **267**, 22830–22836.

[52] Vincent, J.B., Crowder, M.W. and Averill, B.A. 1991, *J. Biol. Chem.* **266**, 17737–17740.

[53] Kim, E.E. and Wyckoff, H.W. 1991, *J. Mol. Biol.* **218**, 449–464.

[54] Yan, S.C.B. and Graves, D.J. 1982, *Mol. Cell Biol.* **42**, 21–29.

[55] King, M.M. and Huang, C.Y. 1984, *J. Biol. Chem.* **259**, 8847–8856.

[56] Rao, J. and Wang, J.H. 1989, *J. Biol. Chem* **264**, 1058–1061.

[57] Vincent, J.B. and Averill, B.A. 1990, *FEBS Lett.* **263**, 265–268.

[58] Alessi, D.R., Street, A. J., Cohen, P. and Cohen, P.T.W. 1993, *Eur. J. Biochem (in press)*.

[59] Zhang, Z., Bai, G., Deans-Zirattu, S., Browner, M.F. and Lee, E.Y.C. 1992, *J. Biol. Chem.* **267**, 1484–1490.

[60] Berndt, N. and Cohen, P.T.W. 1990, *Eur. J. Biochem.* **190**, 291–297.

[61] Ingebritsen, T.S., Stewart, A.A. and Cohen, P. 1983, *Eur. J. Biochem.* **132**, 297–307.

[62] Wang, Z., Ming, L.-J. Que, L. Jr., Vincent, J.B., Crowder, M. W. and Averill, B.A. 1992, *Biochemistry* **31**, 5263–5268.

[63] Rost, B., Schneider, R. and Sander, C. 1993, *Trends Biochem. Sci.* pages 120–123.

[64] Kroger, M. and Hobom, G. 1982, *Gene* **20**, 25–38.

[65] Liu, Y. S., Ishii, M., Tokai, H., Tsutsumi, O., Ohki, R., Akada, K., Tanaka, E., Tsuchiya, S., Fukui and Miyakawa, T. 1991, *Mol. Gen. Genet* **227**, 52–59.

[66] Cyert, M. S., Kunisawa, R., Kaim, D. and Thorner, J. 1991, *Proc. Nat Acad Sci.* **88**, 7376–7380.

[67] Guerini, D. and Klee, C.B. 1989, *Proc. Natl. Acad. Sci. USA* **1989**, 9183–9187.

[68] McPartlin, A.E., Barker, H.M. and Cohen, P.T.W. 1991, *Biochim. Biophys. Acta.* **1088**, 308–310.

[69] Kuno, T., Takeda, T., Hirai, M., Ito, A., Mukai, H. and Tanaka, C. 1989, *Biochem. Biophys. Res Com.* **165**, 1352–1358.

[70] Sneddon, A.A., Cohen, P.T.W. and Stark, M.J.R. 1990, *EMBO J.* **9**, 4339–4346.

[71] Kinoshita, N., Ohkura, H. and Yanagida, M. 1990, *Cell* **63**, 405–415.

[72] da Cruz e Silva, O.B. and Cohen, P.T.W. 1987, *FEBS Lett.* **226**, 176–178.

[73] Stone, S. R., Hofsteenge and Hemmings, B. A. 1987, *Biochemistry* **26**, 7215–7220.

[74] da Cruz e Silva, O.B., Alemany, S., Cambell, D.G. and Cohen, P. T.W. 1987, *FEBS Lett.* **221**, 415–422.

[75] Green, D. D, Yang, S.I. and Mumby, M. C. 1987, *Proc. Natl. Acad. Sci. USA* **84**, 4880–4884.

[76] Orgad, S., Brewis, N.D., Alphey, L., Axton, J.M., Dudai, Y. and Cohen, P.T.W. 1990, *FEBS Lett.* **275**, 44–48.

[77] MacKintosh, R.W., Haycox, G., Hardie, D.G. and Cohen, P.T. W. 1990, *FEBS Lett.* **276**, 156–160.

[78] Brewis, N.D. and Cohen, P.T.W. 1992, *Biochim. Biophys. Acta* **1171**, 231–233.

[79] Brewis, N. D. PhD thesis, University of Dundee, Scotland, UK., 1993.

[80] Ronne, H., Carlberg, M., Hu, G-Z. and Nehlin, J.O. 1991, *Mol. Cell. Biol.* **11**, 4876–4884.

[81] Ohkura, H., Kinoshita, N., Miyatani, S., Toda, T. and Yanagida, M. 1989, *Cell* **57**, 997–1007.

[82] Sasaki, K., Shima, H., Kitagawa, Y., Irino, S., Sugimura, T. and Nagao, M. 1990, *Jpn. J. Cancer Res.* **81**, 1272–1280.

[83] Doonan, J.H and Morris, N.R. 1989, *Cell* **57**, 987–996.

[84] Dombradi, V., Mann, D.J, Saunders, R.D. and Cohen, P.T.W. 1993, *Eur. J. Biochem* **212**, 177–183.

[85] Berndt, N., Campbell, D.G., Caudwell, F.B., Cohen, P., da Cruz e Silva, E.F., da Cruz e Silva, O.B. and Cohen, P.T.W. 1987, *FEBS Lett.* **223**, 340–346.

[86] Nitschke, K., Flieg, U., Schell, J. and Palme, K. 1992, *EMBO J.* **11**.

[87] Dombradi, V., Axton, J.M., Brewis, N.D., da Cruz e Silva, E.F., Alphey, L. and P.T.W., Cohen. 1990, *Eur. J. Biochem* **194**, 739–745.

[88] Dombradi, V., Axton, J.M., Glover, D.M. and Cohen, P.T.W. 1989a, *Eur. J. Biochem.* **183**, 603–610.

[89] Smith, R.D. and Walker, J.C. 1991, *Plant Physiol.* **97**, 677–683.

[90] Dombradi, V., Axton, J.M., Glover, D.M. and Cohen, P.T.W. 1989b, *FEBS Lett.* **247**, 391–395.

[91] Evers, R. and Cornelissen, W. C. A. 1990, *Nucl. Acid. Res.* **18**, 5089–5090.

[92] Cohen, P.T.W. 1990, *Genetics And Human Nutrition, (Ed. Randle, P.J. Bell, J.I. And Scott, J.) John Libby* pages 27–40.

PP2B S.c CMP2 [65, 66],
PP2B S.c CMP1 [65, 66],
PP2B beta1 H [67],
PP2B beta3 [68],
PP2B beta2 H [67],
PP2B beta2 Rt [69],
PP2B alpha1 Rt [69],
PP2A S.c PPH22 [70],
PP2A S.c PPH21 [70],
PP2A S.p $ppa2^{+}$ [71],
PP2A S.p $ppa1^{+}$ [71],
PP2A beta Rb [72],
PP2A beta P [73],
PP2A alpha B Rb P [74, 75, 73],
PP2A 28d D.m [76],
PP2A B.napus [77],
SIT4 S.c [12],
PP4 (PPX) H [78],
PP4 (PPX) Rb [41],
PP4 D.m [79],
PPV D.m [9],
PPH3 S.c [80],
PP1 S.c DIS2 [81],
PP1 gamma Rt [82], PP1 m1 M [81],
PP1 A.nidulans [83],
PP1 S.p $dis2^{+}$ [81],
PP1 13C D.m [84],
PP1 alpha Rb [85],
PP1 A.thialana [86],
PP1 96A D.m [87],
PP1 S.p $sds21^{+}$ [81],
PP1 87B D.m [88],
PP1 beta Rb Rt [87, 82] PP1 m2 M [81],
PP1 9C D.m [87],
PP1 Z.mays [89],
PPY 55A D.m [90],
PP1 T.brucei [91],

PP1 B.napus [77],
PPZ1 S.c [92],
PPQ S.c [36],
PPZ2 S.c [9, 38],
PP5 H (A.E. McPartlin, H.M. Barker and P.T.W. Cohen, in preparation) (1)
PPT S.c (M. X. Chen, Y.H. Chen and P.T.W. Cohen, in preparation) (2)