Plant Molecular Biology **39:** 593–605, 1999.
© 1999 *Kluwer Academic Publishers. Printed in the Netherlands.*

Higher plant tyrosine-specific protein phosphatases (PTPs) contain novel amino-terminal domains: expression during embryogenesis

Anthony P. Fordham-Skelton*, Mark Skipsey, I. Marta Eveans, Robert Edwards and John A. Gatehouse

*Crop Protection Group, Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, UK (*author for correspondence)*

Received 19 June 1998; accepted in revised form 1 October 1998

Key words: higher plant, tyrosine-specific protein phosphatase, embryo

Abstract

Sequences encoding proteins with homology to protein tyrosine phosphatases have been identified in *Arabidopsis*, soybean and pea. Each contains a predicted catalytic domain containing sequence motifs characteristic of tyrosine-specific protein phosphatases (PTPs) which play an important role in signal transduction in other eukaryotes and are distinct from dual-specificity, cdc25 or low-molecular-weight protein tyrosine phosphatases. Their identity as PTPs was confirmed by characterising the soybean PTP expressed as a recombinant His-tagged fusion protein. The enzyme had phosphatase activity towards *p*-nitrophenolphosphate (*p*NPP) and phosphotyrosine, but did not hydrolyse phosphoserine or phosphothreonine at a measureable rate. Phosphotyrosine containing peptides also served as substrates, with K_m values in the micromolar range. Activity was abolished by inhibitors specific for tyrosine phosphatases (vanadate, dephostatin) but was unaffected by inhibitors of serine/threonine protein phosphatases (fluoride, cantharidin, metal-chelating agents). Gel filtration chromatography showed that the recombinant enzyme was a monomer. The *Arabidopsis* PTP sequence was isolated both as a genomic clone and as a partial EST, whereas the pea and soybean sequences were isolated as cDNAs. Southern analysis suggested a single gene in *Arabidopsis* and a small gene family in pea and soybean. In pea, PTP transcripts were present in embryos, and decreased in level with development; transcripts were also detectable in other tissues. The plant PTPs all contain a similar N-terminal domain which shows no similarity to any known protein sequence. This domain may be involved in PTP functions unique to plants.

Introduction

Whilst much progress has been made in understanding the role of serine / threonine protein phosphorylation in plant signalling pathways [28, 30], the role of tyrosine phosphorylation has received less attention. In animals and lower eukaryotes, such as yeasts and slime moulds, the reversible phosphorylation of protein tyrosine residues is involved in the transduction of signals from a wide range of stimuli, including growth factors, cytokines, hormones, extracellular matrix components and cell adhesion molecules. Key regulatory elements of such pathways are tyrosine phosphatases which form a large superfamily that has been classified into a series of families based on functional properties and sequence homology [4, 8, 23]. All members of this superfamily contain a catalytically active cysteine residue within a Cx_5R motif at the active site of the enzyme. Within this superfamily are tyrosine-specific protein phosphatases (PTPs) which are distinct from VH1-like dual-specificity, cdc25 or low-molecular-weight (LMW) tyrosine phosphatases in sharing a conserved catalytic domain of ca. 240 amino acids (aa) which contains diagnostic sequence motifs [8]. PTPs also contain domains either at the

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers AJ006309 (PTP1_ARATH), AJ005589 (PTP1_PEA) and AJ006308 (PTP1_1SOYBN).

N- or C-terminal of the protein which allows the further subclassification of PTPs into receptor-like PTPs which contain a trans-membrane domain and an extracellular domain(s) or non-receptor-like PTPs which have associated targeting or regulatory domains such as phosphotyrosine-binding SH2 domains.

Surprisingly, in view of their importance in signalling in other eukaryotes, our understanding of tyrosine phosphorylation / dephosphorylation in plants is fragmentary, and progress in characterising plant tyrosine phosphatases at the molecular level has, until recently, not been forthcoming [28]. Tyrosine phosphatase activities have been described in wheat seedlings and in pea nuclei [6, 12]. Similarily, tyrosine kinase activities have been described in pea seedlings, maize seedlings and *Catharanthus roseus* roots transformed with *Agrobacterium rhizogenes* [25, 33, 34]. Several dual-specificity kinases with the ability to phosphorylate tyrosine as well as serine and threonine residues have also been described [2, 5, 15, 21]. There is evidence that tyrosine phosphatases may play a role in regulating plant developmental processes. Thus, the *rolB* gene product of *A. rhizogenes* encodes a tyrosine phosphatase and is required for induction of the hairy root phenotype [10]. Similarly, bacterial pathogens of animals produce tyrosine phosphatases which disrupt signalling processes of their hosts [37]. Furthermore, there is some evidence to suggest that tyrosine phosphorylation is involved in phytochrome responses [29, 32].

Tyrosine phosphorylation in plants also plays a role in regulating the activity of mitogen-activated protein (MAP) kinases, which undergo activation by phosphorylation of conserved tyrosine and threonine residues, and are known to be involved in a variety of signal transduction pathways [16]. In tobacco, treatment with elicitors, salicylic acid or by cutting activates MAP kinases. These are tyrosine-phosphorylated and their activity is sensitive to treatment with tyrosine phosphatases [1, 31, 35, 41]. Also, in barley, phenylarsine oxide, a tyrosine phosphatase inhibitor, prevented the activation of a MAP kinase in response to ABA and it appeared that the inhibitory block was upstream of MAP kinase kinase [18]. Taken together, this data indicates that tyrosine phosphatases are involved in the regulation of MAP kinase pathways and other signalling processes in plants. However, the type(s) of tyrosine phosphatase involved are unknown, and, with two exceptions, these important signalling enzymes have not been cloned. Recently, a cDNA (*VH-PTP13*) was identified in the lower plant *Chlamydomonas eugametos* which encodes a phosphatase capable of inactivating MAP kinases from higher plants; however, this protein is most similar to the dual-specificity phosphatases which can hydrolyse both phosphoserine / threonine and phosphotyrosine residues [13]. During the preparation of this manuscript, an *Arabidopsis* cDNA (*AtPTP1*) encoding a PTP was reported [39]. We now report homologous PTPs from pea and soybean and the genomic sequence of *Arabidopsis* PTP1, which shows that the published *AtPTP1* cDNA sequence contains an artefact in the 3′ untranslated region. These PTPs represent a family of proteins in higher plants and differ from other eukaryotic PTPs in their N-terminal domains.

Materials and methods

Database searches and sequence analysis

BLAST searches [3] were carried out using the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/) and the AtDB *Arabidopsis* BLAST server (http://genome-www.stanford.edu/Arabidopsis/). Intron/exon boundaries in genomic DNA were determined using the NetPlant Gene server (http://www.cbs.dtu.dk/netpgene/-cbsnetpgene.html; [14]). Amino acid sequence alignments were performed using the CLUSTAL program within the LASERGENE software package (Oct. 1997 release, DNASTAR, WI) with a PAM250 residue weight table followed by manual manipulation of the output. PTP catalytic domain sequences (Y phosphatase) were retrieved from the Pfam protein domain database at the Sanger Centre, UK (http://www.sanger.ac.uk/Pfam/). Prediction of signal peptide cleavage sites was carried out with the SignalP program (http://www.cbs.dtu.dk/services/SignalP/). Putative promoter region sequences were analysed with a plant *cis*-acting regulatory element database (http://www.dna.affrc.go.jp/htdocs/PLACE).

Arabidopsis *cDNA libraries, ESTs and BAC clone*

The λPRL2 library [24], EST clone AA042465 (CD4-16 cDNA library [17]) and BAC (bacterial artificial chromosome) F18D12 were obtained from the ABRC (Ohio, USA). All clones were derived from the Columbia ecotype. DNA from BAC F18D12 (partial *Eco*RI digest in pBeloBAC Kan) was prepared by alkaline lysis of bacterial cultures (10 ml) followed by purification on a WIZARD miniprep column

(Promega, Madison, WI) with the NaOH/SDS solution heated to 60 °C prior to use to ensure complete cell lysis. A 4.5 kbp *Eco*RI restriction fragment which hybridised to the cDNA PCR product of primers AT2 and AT3 (Figure 1) was subcloned into pUC18 and sequenced by primer walking. Sequence was determined on both strands for the region 420 bp upstream of *PTP1_ARATH* to the *Eco*RI site within the coding sequence and on one strand for the remainder of the clone. Further information regarding the other two predicted genes can be found in the EMBL accession for *PTP1_ARATH*.

DNA manipulation, production of DNA probes and PCR amplification

Standard techniques were used for plasmid DNA propagation and manipulation [26]. DNA was labelled with [α-^{32}P]dCTP either by random priming [9] or by PCR amplification with *Taq* polymerase [11]. All other DNA amplifications were carried out using either native *pfu* DNA polymerase (Stratagene, La Jolla, CA) or EXPAND DNA polymerase mix (Boehringer Mannheim, Lewes, Sussex, UK) to minimise nucleotide misincorporation. Where appropriate, PCR products were cloned into pCR2.1 TOPO (Invitrogen, Leek, Netherlands) and sequenced.

cDNA library screening and 5′ RACE

Pea (*Pisum sativum* genotype JI813) and soybean (*Glycine max* cv. D297) cDNA libraries were constructed in λUNI-ZAP XR (Stratagene) using cDNA derived from 8-day old pea embryos [7] and 14-day old soybean seedling roots. About 8×10^5 and 6×10^5 pfu of the pea and soybean libraries, respectively, were screened with the 284 bp pea PTP PCR product labelled by PCR. Filters were washed to a final stringency of 0.5× SSC / 0.1% SDS at 65 °C. Plasmid DNAs were recovered from positively hybridising clones by *in vivo* excision as described by the manufacturer.

To obtain further 5′ sequence of *PTP1_PEA* 5′ RACE PCR amplifications were carried out essentially as described in the Boehringer Mannheim 5′ RACE protocol with the following modifications. cDNA was synthesised from 1 μg of poly(A)$^+$ RNA isolated from 10–11-day old pea embryos using primer PS1 (Figure 1) with AMV reverse transcriptase (Promega), the reaction was incubated at 45 °C for 60 min and then 55 °C for 20 min and cDNA was purified on a Boehringer spin column. A homopolymer tail was added with terminal deoxynucleotidyl transferase (Promega) and dATP. This template was amplified with specific primer PS2 (Figure 1) and generic primer 1 (5′-GCGATATCTCGAGGATCC(T)$_{17}$-3′), resultant products were separated by agarose gel electrophoresis and a faint 'smear' of ca. 550 bp was isolated and subject to further rounds of amplification using primer PS2 and generic primer 2 (5′(T)$_{25}$C 3′). PCR products of ca. 550 bp were cloned and those containing an internal *Eco*RI site (Figure 1) were sequenced.

Southern and RNA gel blot analyses

Hybridisation and washing conditions for Southern and RNA gel blots were as described [7]. Total genomic DNA was restricted, separated by electrophoresis through 1% agarose gels and transferred onto nitrocellulose. Filters were washed to a final stringency of 2× SSC / 0.1% SDS at 65 °C. RNA samples were quantified by absorbance at 260 nm and staining of gels with ethidium bromide [7]. RNA gel blots containing poly(A)$^+$ RNA samples (5 μg) from different stages of embryo development were washed to a final stringency of 0.5× SSC / 0.1% SDS at 65 °C.

DNA sequencing

Plasmid template DNAs were sequenced using fluorescently labelled dye terminators and an AmpliTaq cycle sequencing kit (PE ABI, Warrington, Cheshire, UK). Reaction products were analysed on a PE ABI 373 stretch or 377 DNA sequencer. The sequences of primary cDNAs, the genomic clone and genomic PCR products were determined on both strands.

Production and purification of soybean PTP in Escherichia coli

The PTP coding sequence was amplified with primers containing an *Nde*I site (in-frame ATG initiation codon) and a *Xho*I site (fusion to C-terminal 6× His tag of pET24a, Novagen, Madison, WI). The primer sequences were (5′ GC*CATATG*GCCGGCAACCCC-GCCACCAC-3′) and (5′-CG*CTCGAG*CTCCGACT-GCTGCTGAGATAC-3′). After cloning and sequence verification, the construct was excised, ligated into *Nde*I/*Xho*I-restricted pET24a and transformed into *E. coli* BL21 (DE3) with kanamycin selection (50 μg/ml). An overnight culture in L-broth (low-salt, 5 g/l, pH 7.5) was diluted 1:20 into fresh medium (100 ml) and grown at 30 °C to a cell density of

Arabidopsis PTP1_ARATH

```
-420 TTTTTTCTGTGACGTTGCCCATCTTTTTTTTTATAATGATGAGCCCATCTAAAATGTTAT
-360 TCTACAAAACAATAAAAGAAAGACCAGGATTCCTTTGCTTAAAAGACCAGGGTCCTTTGC
-300 TAAAACTATATTATAATGACCAATGAAATCATAATTAATAATGACCAATTAAAATGATGA
-240 CAAAAAAAAAAAAAAACCCAATTAAAATAGTAAATTTATATTATAGTCATCAACTTTAATA
-180 CTTTGTTGCAAAATAAAAGGAAACAGTCGCATAATAATCCACTAATCACAAACCTTTACC
-120 CTTTCTTCTCGGGGCCAATCAAAAACTTCAATTAACCAGAAAAGGTCAGATTCTCCGCCG
 -60 CCTCACACGACGGGCAACTTCAGTTCCTGCCGTCGCGTTAGATCCCGCCGTCAATCGGTC

   1 ATGGCGACCGGTAAAACCTCTTCCGCCGCGAATCTTTTCACTGGCTCGACGCGTTTTGAT
      M  A  T  G  K  T  S  S  A  A  N  L  F  T  G  S  T  R  F  D   20

  61 TTATCATCCGCTGATTCGCCTCCTTCAAAACTCTCTCTCTCCTCCGATCAGCTCAACCAC
      L  S  S  A  D  S  P  P  S  K  L  S  L  S  S  D  Q  L  N  H   40
                                                                <
 121 TGCCACCAAGCTCTCGGCGTTTTCCGGGGAAAGATCCAAAATCCTGACTCGATCGCTCAT
      C  H  Q  A  L  G  V  F  R  G  K  I  Q  N  P  D  S  I  A  H   60
     H10F3T7
 181 GAGTTTACCGGTTTACAGgttatagatccatcctctaaattgagaatattcctcttccta
      E  F  T  G  L  Q                                             66

 241 gtgtttcttccttaaaaacaaattggatcataatacactctcttcgaacgcagGCTAATA
                                                           A  N    68

 301 GGATGTGGCCATCGGAGCTGCTGCTAAACAGTACAGTGGCTATGAACAGTGTCAATGTTG
     R  M  W  P  S  E  L  L  L  N  S  T  V  A  M  N  S  V  N  V    88

 361 AGAAAAACAGATACAGTGATGTTGTTCCATgtaattttcctctctttttttgtgtggatt
     E  K  N  R  Y  S  D  V  V  P                                  98

 421 tgctagttagtggtgtacagttctagatttgattttgctttggatatagTTGACAAGAA
                                                       F  D  K  N 103

 481 CAGGATTGTTCTGAATCCATGTAAAGACTCATCTGCAAAAGGATATGTGAATGCAAGCTT
       R  I  V  L  N  P  C  K  D  S  S  A  K  G  Y  V  N  A  S  L 123

 541 AATTAAGgtacacttcttattagtctgtttatctgtatctttttgtggtgatagatttt
       I  K                                                        125

 601 gaaggttttacaatgtgttaatgttatgtagACGTCGGAGTCTGAGAGTATTTCTCAGTT
                                      T  S  E  S  E  S  I  S  Q  F 135

 661 CATAGCTACGCAAGGTCCTTTACCACACACGATGGAGGACTTCTGGGAGATGGTTATTCA
       I  A  T  Q  G  P  L  P  H  T  M  E  D  F  W  E  M  V  I  Q 155

 721 GCAGCATTGCCCAATCATAGTGATGCTCACTCGATTGGTTGATAATAATAGGgtaaactt
       Q  H  C  P  I  I  V  M  L  T  R  L  V  D  N  N  R           172

 781 ccttgtcttgcttgtgtatatgagtctcttttagacgattggatattaatatgatctgt

                            AT1
                  ___________________\
 841 tgcaaatatactagtacaagcgaagttgttgtgagcatttgtttctcaagttaaatttct
 901 tcgtattaaattttagcaagttttaaatgttccataaacagtgatgaagatagaaactat
 961 ctcgttgatattgattgttgaacttatcttgtttgtttttcatgctaacttatctaatt
                                                   AT2
                                          ______________________\
1021 acttgttacagACTGTTAAATGCGGGGACTATTTTCAAGACGAAGATGGACCTAGAGAAT
                T  V  K  C  G  D  Y  F  Q  D  E  D  G  P  R  E    188

1081 TTGGCAACATATCTCTTACAACAAAGTGGATAAAGACTACTGACACTTCATTGATGTTAC
     F  G  N  I  S  L  T  T  K  W  I  K  T  T  D  T  S  L  M  L   208

1141 GGAATCTTGAGGTTAACTACAAGGAGgtaagaatcttgtatcagttttcttctgtttatt
     R  N  L  E  V  N  Y  K  E                                    217

1201 tttaccagatcaagttatttcttcgagggttaagcatgtttcaattattgatattgtgtt

1261 ttacagACAGAGGATCAGCCCATGTCCGTTTTGCATATTCAGTATCCAGAATGGCCTGAT
           T  E  D  Q  P  M  S  V  L  H  I  Q  Y  P  E  W  P  D   235

1321 CATGGAGTTCCCAAGGATACAGTGGCTGTCCGTGAAATTCTAAAAAGACTATATCAAGTA
      H  G  V  P  K  D  T  V  A  V  R  E  I  L  K  R  L  Y  Q  V   255

1381 CCACCTAGTCTCGGCCCAATCATTGTGCACTGCAGgtacgatgcatgtcaatgagggata
      P  P  S  L  G  P  I  I  V  H  C  S                           267
                                                          / AT3
1441 caagtgttgttcaacgtctcctgagagaattgtttttttctctctcagTGCAGGTATAGG
                                                        A  G  I  G 270
     ____________                              F18D12>
1501 AAGAACTGGAACATACTGTGCGATACATAACACAATCCAAAGAATTCTTGCTGGCGATAT
       R  T  G  T  Y  C  A  I  H  N  T  I  Q  R  I  L  A  G  D  M 290

1561 GTCTGCGTTGGATCTTGCTAAAACCGTGGCACTATTTCGCAAGCAACGCATTGGCATGGT
       S  A  L  D  L  A  K  T  V  A  L  F  R  K  Q  R  I  G  M  V 310

1621 TCAAACCATGgtaaatctttcagcgttacaattttagctaccaaatttctgtatctcac
       Q  T  M                                                     313

1681 cacataatttgttctgaattgttcaattccttactctttttttgaagGATCAATACTTCTT
                                                   D  Q  Y  F  F  318

1741 TTGCTACAATGCTATTGTTGATGAATTAGAAGATCTAACCGCGGGGACAAATGCTGGAAC
       C  Y  N  A  I  V  D  E  L  E  D  L  T  A  G  T  N  A  G  T 338

1801 GAGTTCCTAAAGgtacgcaatttcaacaatcaaatctgagtgagccatacctgttataat
       S  S  *                                                      340
1861 ctatgaatactcttatcttgtctatcttttgatccagCTGAAGGGTTCGTCTGCTTAGAG
1921 AGGGAAAAAGGCTTCTACCCAACTATATCAATTTCGTAATGTGATTCAAGAAGAAACCCG
1981 CGTGAATCTCTACATTCAAGACTTTTCAATTTCTGTAAATTTCCATATCTGATAGGTTTC
2041 ATTGTTACTTTTGTTGTGGAAATGCTGATTAAAAAGCAAATGAAGTCGCCAATGCGAGAT
2101 TCTTCAAAGACTATGAATTCGGTTTGATTTGGTTTGAAGTGAAAAGAACGCG
```

Pea PTP1_PEA

```
 -54 GTTGAACTGGGGAACCTCGTCTCTACCCGCTCTCGTCCTCGCCGGAAATCCGTC

   1 ATGGGTGGGAGTTCCGCACCTCCCTCCTCCACGTCTTCTCACAACAATTCAACTTTCTCA
      M  G  G  S  S  A  P  P  S  S  T  S  S  H  N  N  S  T  F  S  20
               :        EcoRI
  61 CCTGATTTCCCTTCCCGAATTCCCCTTACTTCCGATCAGATCAAACACTGCACTGAAGCA
      P  D  F  P  S  R  I  P  L  T  S  D  Q  I  K  H  C  T  E  A  40

 121 CTCGCCCTGTTGAAAAACAAACTTCTCAACCCTCACACCGTCTCCCAAAAGTTCTTTCAC
      L  A  L  L  K  N  K  L  L  N  P  H  T  V  S  Q  K  F  F  H  60

 181 TTGCAGTCGAATAGGATAACGCTGAGTGAGACGACAAAAAAATGCCATGTTGCTCTTAAT
      L  Q  S  N  R  I  T  L  S  E  T  T  K  K  C  H  V  A  L  N  80

 241 TCTGCTAATCTCAGCAAAAATCGATACAGTGATGTTATACCATTTGACAAGAACAGGGTT
      S  A  N  L  S  K  N  R  Y  S  D  V  I  P  F  D  K  N  R  V 100

 301 GTTCTTAAATCGAGTTCTGATTATAGGTCTGAAGCACTAGGGTATATCAATGCAAGCAAA
      V  L  K  S  S  S  D  Y  R  S  E  A  L  G  Y  I  N  A  S  K 120
                                                   <      PS2
 361 ATCTCGACCTCTTCCCCTGGAATTGTGTCCGAGTTTATAGCCACGCAAGGTCCAATGCCA
      I  S  T  S  S  P  G  I  V  S  E  F  I  A  T  Q  G  P  M  P 140

 421 CACACTTTTGAGGATTTTTGGGAAATGATGATACAATATCACTGCCCTGCAATTGTGATG
      H  T  F  E  D  F  W  E  M  M  I  Q  Y  H  C  P  A  I  V  M 160

 481 CTTACTGGGTTGGTCGACAATTACAAGACGGTGAAGTGTGGAGATTATTTTCAATCTGAA
      L  T  G  L  V  D  N  Y  K  T  V  K  C  G  D  Y  F  Q  S  E 180

 541 GATAGACCTAGAGAATTCGGGAATATCTCTCTTACATGTAAATGGACGAAAACTACTAAA
      D  R  P  R  E  F  G  N  I  S  L  T  C  K  W  T  K  T  T  K 200

 601 ACTTCACTTGTGCTGCGACATTTGGAGGTGAACCGTAAAGAGGTAGAAGATACACCGTTA
      T  S  L  V  L  R  H  L  E  V  N  R  K  E  V  E  D  T  P  L 220

 661 TCTGTTTTTCATATTCAGTATCCTGAGTGGCCTGACCATGGAGTTCCAAACAATACTTTG
      S  V  F  H  I  Q  Y  P  E  W  P  D  H  G  V  P  N  N  T  L 240

 721 GCTGTGCGTGCAATTTGGAAAAGATTATACCACTTGCCACCAAATCTTGGCCCAATAGTG
      A  V  R  A  I  W  K  R  L  Y  H  L  P  P  N  L  G  P  I  V 260
       <            PS1
 781 GTGCACTGCAGTGCAGGTATTGGTAGAACTGGAACATACTGCACAATTCATAACACGATT
      V  H  C  S  A  G  I  G  R  T  G  T  Y  C  T  I  H  N  T  I 280

 741 CAGAGAATACTTGCTGGTGATATGTCTGCTATAGACATTGCCAATACAGTATCTGTGTTC
      Q  R  I  L  A  G  D  M  S  A  I  D  I  A  N  T  V  S  V  F 300

 901 AGATCTCAGCGTATTGGAATGGTTCAGACACAGGATCAGTACATTTTTTGCTATGAAGCT
      R  S  Q  R  I  G  M  V  Q  T  Q  D  Q  Y  I  F  C  Y  E  A 320

 961 ATCATAGATGAACTAGAAGACCTTGTATCTCAACAGTAATGAGAGTAGAAGGTGGTGGAC
      I  I  D  E  L  E  D  L  V  S  Q  Q  *

1021 CAAAGAATGTTAGTCTGATTGCCTCACTCTGAAGCCAACGTTCGATTCTTCGATTTTCCA
1081 CTTGGGTCCTACTATGCTCTTGTTTCCCAGCAGCTGATTCTGAAGCATTAGCATTCAGGA
1141 GATTTTATGATTTTCATTGAACTCAGTTTATTCTCAAGCATTCATGTATATTTGTAGTCA
1201 GGTGAAAGCTAACAAATATCAGCATCCTTTTCATTATATTTCTACAATTCTATTGGTGAA
1261 CTGAAAGACACTAATGACCTTT(A)
```

Soybean PTP1_SOYBN

```
                -102 GTTCCACTCTATATCCATCGTCAACAGTGAAGTGCGAATACT
     CAGCAACTGGAGCTGGGGAACCTTGTCTCTACTCGCTCTCGTCCCCGCCGGAGATCCACC

   1 ATGGCCGGCAACCCCGCCACCACTTCCTCCTCCGCATTATCTCCCGAGAAATTTAATTTC
      M  A  G  N  P  A  T  T  S  S  S  A  L  S  P  E  K  F  N  F  20

  61 AATTTCTCCCCCGATAACCCCTCCCGAATCACCCTCACTTCGGATCAGGTCAACCACTGC
      N  F  S  P  D  N  P  S  R  I  T  L  T  S  D  Q  V  N  H  C  40

 121 ACCCAAGCCCTCAACATTCTGAAGGAGAAGCTCCACGCTCCTAACGTGATCACTCAGGAG
      T  Q  A  L  N  I  L  K  E  K  L  H  A  P  N  V  I  T  Q  E  60

 181 TTTGCACACTTGCAGGCGAACAGGATAACGCCGAGCGAGATGAGGAGAAGATGCACCGTC
      F  A  H  L  Q  A  N  R  I  T  P  S  E  M  R  R  R  C  T  V  80

 241 GCGTATGACGACGTCAATTTGAGGAAAAACCGATACACCGACGTTCTGCCATTTGACAAG
      A  Y  D  D  V  N  L  R  K  N  R  Y  T  D  V  L  P  F  D  K 100

 301 AACAGGGTTGTTCTCAAATCTAGTACTGATTATAGACCTGAAGCACAGGGGTATATCAAT
      N  R  V  V  L  K  S  S  T  D  Y  R  P  E  A  Q  G  Y  I  N 120

 361 GCAAGCCTTGTCTCGACCTCCTCAGCTGGAAACGTGTCCCAGTTTATAGCCACGCAAGGT
      A  S  L  V  S  T  S  S  A  G  N  V  S  Q  F  I  A  T  Q  G 140

 421 CCACTTCAACATACTTACGAGGATTTCTGGGAAATGATCATCCAATATCACTGCCCTGCA
      P  L  Q  H  T  Y  E  D  F  W  E  M  I  I  Q  Y  H  C  P  A 160

 481 ATTATCATGCTTACTAGATTGGTGGATAATTACAAGATGGCAAAATGTGGAGATTATTTT
      I  I  M  L  T  R  L  V  D  N  Y  K  M  A  K  C  G  D  Y  F 180

 541 CAAGCTGAGGATAGGCCTAGAGAAGTTGGTAATATTTCTATTATTGGTAAATGGGAGAAC
      Q  A  E  D  R  P  R  E  V  G  N  I  S  I  I  G  K  W  E  N 200

 601 ACGACTGAAACTTCACTCGTGCTGCGCCATTTGGAGGTGAACCATAGAGAGGTAGAAGAT
      T  T  E  T  S  L  V  L  R  H  L  E  V  N  H  R  E  V  E  D 220

 661 GCGCCATTATCTGTTTTGCATATTCAGTATCCTGAATGGCCTGACCATGGAGTTCCCAAA
      A  P  L  S  V  L  H  I  Q  Y  P  E  W  P  D  H  G  V  P  K 240

 721 GATACATTTGCTGTGCGTGAAATTTTGAAAAGATTATACCATTTACCACCAAACTTTGGT
      D  T  F  A  V  R  E  I  L  K  R  L  Y  H  L  P  P  N  F  G 260

 781 CCAATAGTGGTGCACTGCAGTGCAGGTATTGGTAGAACTGGAACATACTGCACAATTCAC
      P  I  V  V  H  C  S  A  G  I  G  R  T  G  T  Y  C  T  I  H 280

 841 AACACAATTCAGAGAATAGTTGCTGGTGATATGTCTGCTGTAGATATTGCTAAGACAATA
      N  T  I  Q  R  I  V  A  G  D  M  S  A  V  D  I  A  K  T  I 300

 901 GCTATGTTCAGGTCCCAGCGTATTGGAATGGTTCAGACCCAGGATCAGTACATTTTCTGC
      A  M  F  R  S  Q  R  I  G  M  V  Q  T  Q  D  Q  Y  I  F  C 320

 961 TATAACGCTATCATAGATGAACTGGAAGACCTTGTATCTCAGCAGCAGTCGGAGTAGAAG
      Y  N  A  I  I  D  E  L  E  D  L  V  S  Q  Q  Q  S  E  *

1021 GTGCTGGAACTTGGAAGGATGTTGAAGCTGCTTGCCTCATTTTGAAGCTAACTATCAATT
1081 CTTCGAACGTTTTTATGTTCCTGCAATGCTCATGTTTTCCAGCAGCTGAAGTCTGAAACA
1141 TTAGCATCTGGGAGATTTTATGAATTTCACCGAATTCTGTTATTCTCATGCTTTTTTGTA
1201 TAATTGCAGGCAGGGAAAGCTGACAAATATTAACATCCTTTTGATTATATTGTTAGAATT
1261 CTATTGGGGAACAGAATGGCATAATGACTACAATAATGCATTGTCCCTCC(A)
```

Figure 1. Nucleotide and deduced amino acid sequences of PTPs from *Arabidopsis*, pea and soybean. Nucleotide numbering is from the first base of the predicted methionine initiation codon, shown on the left, and numbering of amino acid residues is shown on the right of each sequence. Conserved residues within the predicted PTP active site are bold underlined and asterisks denote stop codons. Annealing sites for PCR primers are overlined. For *PTP1_ARATH*, predicted intron sequences are in lower case, the start of EST H10F3T7 is marked by (<) and the end of the genomic sequence from BAC F18D12 is marked (>) at an *Eco*RI site. The sequence in the 3′-untranslated region which diverges from that presented for the corresponding *AtPTP1* cDNA [39] is underlined. For *PTP1_PEA*, (:) at nucleotide position 69 indicates the longest cDNA clone obtained and the underlined residue at position 1238 is the site of alternate polyadenylation for one cDNA clone. For *PTP1_SOYBN*, an in frame stop codon 5′ to the predicted methionine initiation codon is underlined.

$OD_{600} = 1.0$. The culture was induced with the addition of IPTG to a final concentration of 1.0 mM and growth was continued for a further 5.5 h. Recombinant protein was purified essentially as described by Shi *et al.* [27]. Cells were pelleted, resuspended in 2.0 ml 20.0 mM Tris-HCl, 0.5 M NaCl, 5.0 mM imidazole pH 7.8 and lysed by sonication followed by centrifugation at 15 000 × *g* for 15 min. The supernatant was loaded onto a 1.0 ml HITRAP chelating Sepharose column (Pharmacia, St Albans, Herts, UK) pre-washed with 0.1 M $NiSO_4$ (according to the manufacturers' instructions) followed by 20.0 mM Tris-HCl, 0.5 M NaCl, 5.0 mM imidazole, pH 7.8. After washing with 20.0 mM Tris-HCl, 0.5 M NaCl, 20.0 mM imidazole pH 7.8, recombinant PTP was eluted with 20.0 mM Tris-HCl, 0.5 M NaCl, 0.3 M imidazole pH 7.8. Imidazole was removed by dialysis against 20.0 mM Tris-HCl, 0.1 M NaCl, 0.1 mM DTT pH 7.8 overnight at 4 °C. Protein concentrations were determined with the BioRad assay reagent using BSA in the same buffer as the reference standard. Protein samples were analysed using denaturing SDS-PAGE with 12.5% gels run under reducing conditions [19].

Phosphatase assays

Phosphatase assays with the generic phosphatase substrate *p*-nitrophenolphosphate (*p*NPP) were carried out at room temperature in microtitre plates. Standard assays (total volume 100 μl) were carried out in 85.0 mM MES buffer, pH 6.25 containing 0.5 mM DTT, 7.6 mM *p*NPP with 0.5 ng/μl enzyme. Assays were initially carried out over the pH range 5–7 using MES and bis-tris propane buffers to determine the pH optimum. Reactions were terminated with the addition of an equal volume of 0.4 M NaOH and the OD_{410} determined using a Dynatech MR5000 microtitre plate reader. Release of phosphate from tyrosine-phosphorylated peptides was measured using a tyrosine phosphatase assay system as described by the manufacturer (Promega). Phosphorylated peptides were END(pY)INASL (peptide 1) and DADE(pY)LIPQQG (peptide 2). pH optimisation was carried out using sodium maleate and bis-tris buffers; activity was consistently higher in the latter buffer, which was used for standard assays. These were carried out as above in 85.0 mM bis-tris buffer, pH 5.8, containing 0.8 mM DTT, 100 μM peptide with 0.2 ng/μl enzyme. Release of phosphate from free phosphoamino acids was measured as described above with a final substrate concentration of 200 μM. Stock cantharidin (Sigma) and dephostatin (Calbiochem, Beeston, Nottingham, UK) were dissolved in DMSO and appropriate solvent controls were used. For inhibition with dephostatin, inhibitor was incubated with the enzyme for 3 min prior to the addition of substrate. Kinetic parameters (K_m and V_{max}) were estimated by varying substrate concentrations in the standard assays, and using Michaelis-Menten double reciprocal plots of rate against substrate concentration at fixed enzyme amounts.

Gel filtration chromatography

Gel filtration chromatography of purified recombinant soybean PTP was carried out on a column of Sephacryl S-200 (1.6 cm dia. × 60 cm) in 50 mM Tris-HCl buffer, pH 7.8, containing 100 mM NaCl, at a flow rate of 0.3 ml/min. Absorbance was monitored continuously at 280 nm, and 3 ml fractions were collected. Fractions were assayed for the presence of recombinant protein by SDS-PAGE followed by western blotting. 15 μl aliquots of selected fractions were mixed with an equal volume of 2× SDS sample buffer prior to electrophoresis. Blots were probed with a mouse monoclonal antibody specific for the 6× His tag on the recombinant protein (Invitrogen) as primary antibody, and anti-mouse IgG peroxidase-linked secondary antibody (BioRad), before visualising using the enhanced chemiluminescence (ECL) method (Amersham) using the protocol provided with the kit.

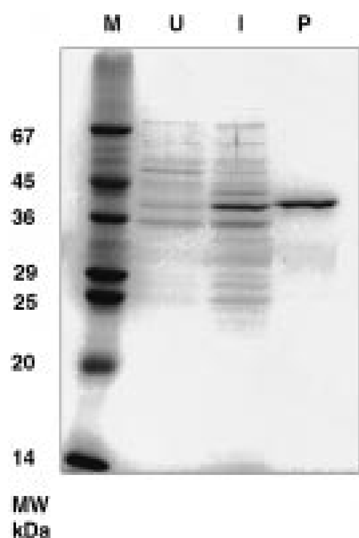
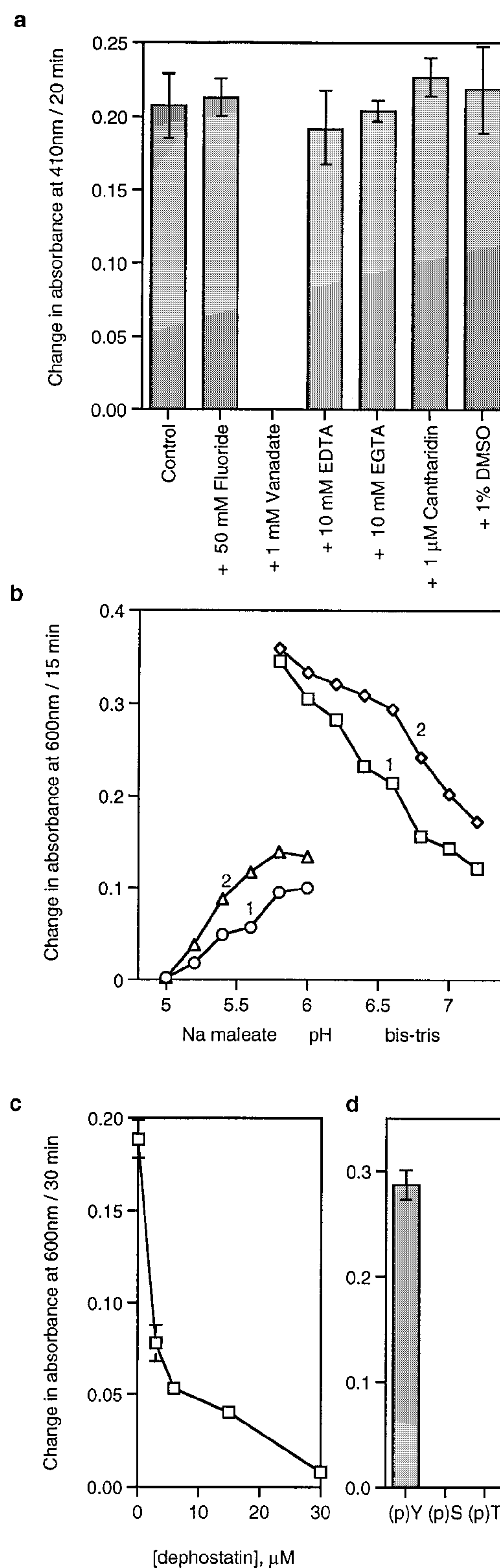


Figure 2. Purification of recombinant soybean PTP. SDS-PAGE analysis of PTP expressed in *E. coli* as a 6× His-tagged fusion protein. Samples are soluble extracts of uninduced (U) and induced (I) bacterial cultures and (P) is the purified fraction from a nickel affinity column. Molecular markers are SDS 7 (BioRad).



Results

Identification of a PTP gene in Arabidopsis, PTP1_ARATH

A sequence corresponding to part of the active site of yeast PTP1 (PIIVHCSAGVGRTGF; Genbank 131557) was used to search the Genbank *Arabidopsis* DNA accessions using the 'TBLASTN' program. A potential match to a translated end sequence from BAC (bacterial artificial chromosome) F18D12 (chromosome 1, Genbank B28373) was found. This sequence could be extended if it was assumed that an intron interrupted the coding sequence after the predicted active site cysteine residue (Figure 1) and extension of the sequence showed further similarity to tyrosine phosphatases. Primers AT2 and AT3, corresponding to sequences in F18D12 5′ and 3′ of the region homologous to PTPs (Figure 1), were used to amplify across the predicted intron using an *Arabidopsis* cDNA library (λPRL-2) as template. This produced a PCR product (284 bp) corresponding to the size predicted by the genomic clone after removal of the putative intron. The PCR product had an identical sequence to the corresponding region of F18D12. The genomic clone thus contains a transcribed gene encoding a tyrosine phosphatase-like protein and was designated *PTP1_ARATH* (Figures 1 and 5a). Further sequencing of a 4.5 kb *Eco*RI restriction fragment

Figure 3. Characterisation of recombinant soybean PTP. a. Effects of phosphatase inhibitors and metal-chelating agents on PTP activity using *p*-nitrophenolphosphate as substrate. Enzyme activity was monitored by absorbance at 410 nm; under the conditions used, 1 nmole of released *p*NP would give an absorbance of 0.34. b. pH dependence of PTP activity in sodium maleate and bis-tris buffers using either peptides 1 or 2 as substrates. Enzyme activity was estimated by measuring released phosphate colorimetrically; under the conditions used, 1 nmole of released phosphate would give an absorbance of 0.40 at 600 nm. c. Inhibition of PTP activity by dephostatin using peptide 1 as substrate. Enzyme activity was measured as in b. d. PTP activity using phosphotyrosine [(p)Y], phosphoserine [(p)S] or phosphothreonine [(p)T] as substrate. Enzyme activity was measured as in b. For a, c and d, values are the mean of three determinations and bars indicate the mean ± the standard error.

of F18D12 confirmed that the predicted product of this gene contained a domain with similarity to PTP catalytic domains (Figure 5a; see below).

Although the *PTP1_ARATH* genomic sequence was not full length, as it terminates at an *Eco*RI site in the coding sequence 3′ to the active site, an expressed sequence tag (EST) H10F3T7 corresponded to the predicted transcript of *PTP1_ARATH* (Figure 1). The sequence of H10F3T7 contained the 3′ end of the *PTP1_ARATH* coding sequence and the 3′-untranslated region (UTR) (Figure 1). However, this clone was a chimera, containing an unrelated cDNA corresponding to part of a predicted gene (T7N9.5; Genbank 2213585). A region corresponding to H10F3T7, used to complete the composite sequence of *PTP1_ARATH* shown in Figure 1, was amplified from *Arabidopsis* genomic DNA. Sequences were checked against each other to ensure that no errors had arisen from PCR and that no introns had been overlooked. There is a single base discrepancy at position 5062 between the genomic sequence obtained by PCR and that of H10F3T7 (see corresponding accession number). The *AtPTP1* cDNA sequence recently published by Xu *et al.* [39] corresponds to *PTP1_ARATH* over the coding region but diverges completely in the 3′ UTR after base 2105 (Figure 1, underlined), and is thus a chimaeric clone (see Discussion).

The sequence of the 4.5 kbp *Eco*RI restriction fragment of *Arabidopsis* genomic clone F18D12 showed that a second coding sequence was present 578 bp upstream of the putative initation codon of *PTP1_ARATH* (Figure 7b). This gene, which was in the opposite orientation to *PTP1_ARATH*, encodes a polypeptide with some similarity to a *Nicotiana* membrane associated salt-inducible protein (Genbank U08285). The region between the two coding sequences, which presumably contains the promoter regions of both genes, was A+T-rich but no previously characterised plant *cis*-acting elements were identified when the sequence was searched against the PLACE database (June 1998).

Isolation of cDNAs encoding PTPs from pea and soybean

Primers AT2 and AT3, used to amplify a cDNA fragment from the *Arabidopsis* library, also amplified a sequence from a pea embryo cDNA library which was highly similar to the corresponding region of *PTP1_ARATH* (not shown). The PCR product was used to screen a cDNA library prepared from pea embryos undergoing the early stages of embryogenesis [7]. Eleven positively hybridising clones were recovered and sequencing showed them to be identical in overlapping regions, with the exception of one which was polyadenylated at an alternate site within the 3′ UTR (*PTP1_PEA*; Figure 1). None of the clones were full length, and thus the sequence was extended by 5′ RACE (rapid amplification of cDNA ends) to produce a complete composite sequence (Figure 1). In the longest RACE products an ATG putative start codon was present, but there were no in-frame stop codons in the 5′ UTR to allow its unambiguous designation as the translation initiation codon. Other evidence suggested that the coding sequence derived from cDNA and 5′ RACE cloning was full-length. Firstly, the transcript size is ca. 1.5 kb (Figure 6; see below) and the sequence of *PTP1_PEA* is 1.336 kb excluding a poly(A) tail. Secondly, a PTP encoding sequence, *PTP1_SOYBN*, was also isolated from a soybean root cDNA library using the pea PCR product as a probe (Figure 1). Upstream of the predicted ATG initiation codon in *PTP1_SOYBN* was an in-frame stop codon. This suggested that the coding sequence deduced for *PTP1_PEA* was full-length, as the N-terminal regions of soybean and pea PTPs are homologous (Figure 5b) and this homology extends into the 5′ UTR with 80% similarity over a 54 bp region (not shown). As in the case of *PTP1_ARATH*, the deduced pea and soybean proteins contain PTP-like catalytic domains (Figure 5a; see below).

Properties of soybean PTP expressed in E. coli

In order to characterise the functional properties of plant PTPs, the soybean *PTP1* cDNA was chosen as it was the only complete coding sequence isolated during these studies. This was expressed in an expression

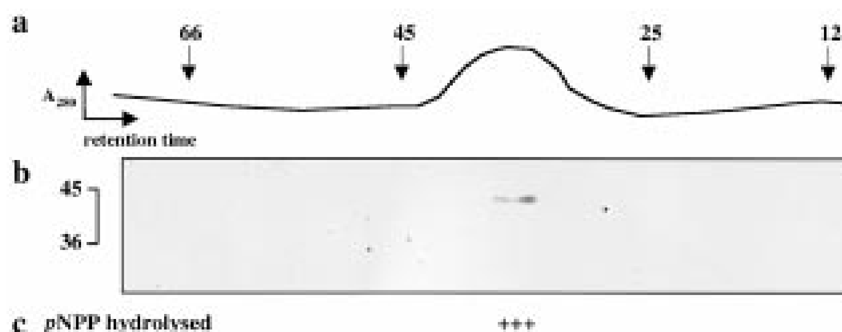


Figure 4. Gel filtration chromatography of recombinant soybean PTP on a column of Sephacryl S-200. a. UV absorbance profile for gel filtration column. Arrows indicate elution positions of standard proteins of indicated molecular mass (kDa). b. Western blot of fractions from gel filtration column probed with antibody directed towards 6× His tag on recombinant protein. c. Phosphatase activity of fractions from gel filtration column against *p*NPP.

vector (pET24a) with a C-terminal histidine tag. Following IPTG induction of transformed bacteria, a protein of ca. 39 kDa was produced (Figure 2). The PTP protein was found in both the soluble and insoluble fractions from the bacterial cells; with a greater proportion of PTP being soluble when the bacteria were grown and induced at 30 °C rather than 37 °C (results not presented). The protein was purified from lysed bacteria using a nickel-chelate affinity column, with a typical yield of approximately 1 mg per 100 ml of starting culture. The identity of the soybean PTP was confirmed by N-terminal sequencing which showed that the initial methionine residue has been removed. The purified PTP hydrolysed the generic phosphatase substrate *p*-nitrophenolphosphate (*p*NPP) (Figure 3a) with a pH optimum of pH 6.2 with activity declining to <50% of the maximum over a range of ± 0.5 pH units either side of the maximum. Enzyme activity was dependent on the presence of reducing agents and declined rapidly if DTT (dithiothreitol) or glutathione were not present. All assays were thus performed in the presence of DTT. With *p*NPP as substrate, a K_m value of 8.5 mM was determined. Figure 3a also shows the effects of various phosphatase inhibitors on PTP activity. Vanadate, a tyrosine phosphatase inhibitor, abolished the enzyme activity completely, whereas fluoride, EDTA, EGTA or cantharidin had no significant effect, demonstrating that this enzyme is only sensitive to inhibitors diagnostic for tyrosine phosphatases and not those of serine / threonine phosphatases [20]. In addition, the enzyme could dephosphorylate free phosphotyrosine but hydrolysis of either free phosphoserine or free phosphothreonine was not detected under the assay conditions used (Figure 3d).

Table 1. K_m and V_{max} of soybean PTP with phosphopeptide substrates. Values are the mean of two independent determinations.

Substrate	K_m (μM)	V_{max} (pmol phosphate released per minute per μg protein)
Peptide 1	36 ± 2	$1.7 \times 10^4 \pm 0.4 \times 10^4$
Peptide 2	32 ± 1	$1.9 \times 10^4 \pm 0.7 \times 10^4$

The purified enzyme was also used in assays measuring the release of phosphate from peptides containing phosphotyrosine. The pH optimum for hydrolysis of these substrates (pH 5.8) was slightly lower than for *p*NPP (Figure 3b). The K_m values for these peptide substrates were ca. 30–40 μM (see Table 1), two orders of magnitude lower than for *p*NPP. The effects of the specific tyrosine phosphatase inhibitor, dephostatin, on release of phosphate from the phosphopeptide END(pY)INASL were also assayed. As shown in Figure 3c, this compound inhibited phosphate release by >50% when present at 2.5 μM, and at 30 μM could inhibit >95% of total enzyme activity.

Gel filtration chromatography was used to investigate the subunit composition of the recombinant soybean PTP expressed in *E. coli*. When separated on a calibrated column of Sephacryl S-200, the PTP ran as a single peak, corresponding to a molecular weight of ca. 36 000 (Figure 4a), agreeing within experimental error to the molecular weight of the PTP polypeptide determined by SDS-PAGE. Western blotting of fractions from the column showed that the distribution of PTP subunits over the fractions corresponded to the peak, with no PTP being detectable in fractions other than those of the peak of absorbance

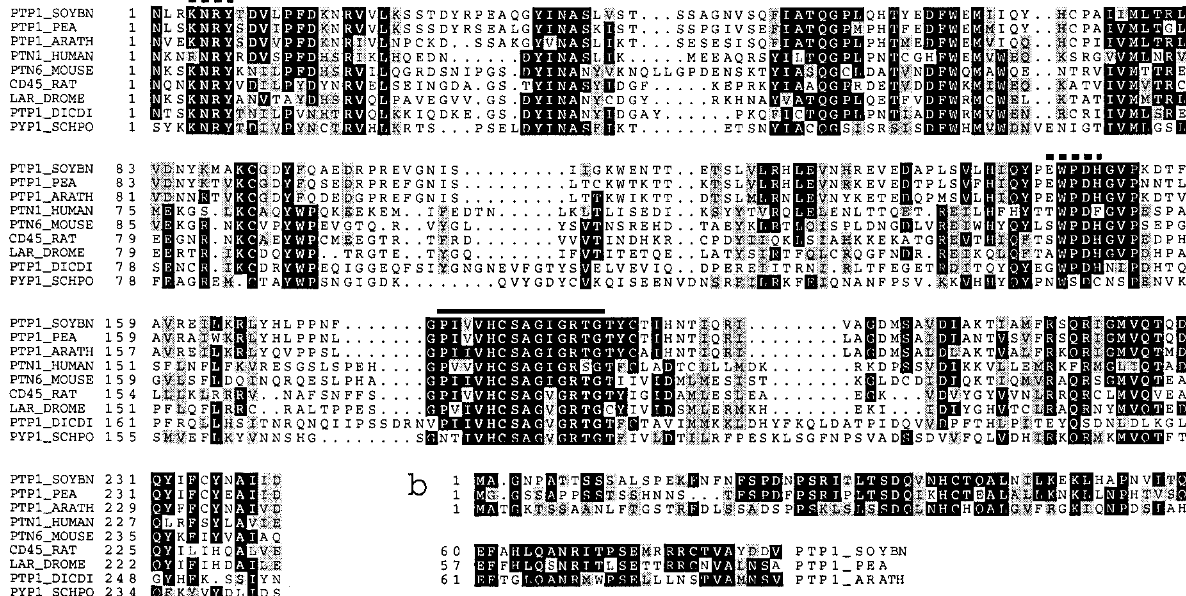


Figure 5. a. Sequence alignment of plant PTP catalytic domains with representative PTPs. Swissprot accessions are from human (PTN1_HUMAN/40-276), mouse (PTN6_MOUSE/270-514), rat (CD45_RAT/628-862), *Drosophila* (LAR_DROME/1497-1728), *Dictyostelium* (PTP1_DICDI/121-377) and yeast (PYP1_SCHPO/295-538); numbering defines the catalytic domain region. The region of each plant PTP sequence used was 86–326, soybean; 83–323, pea; and 87–325, *Arabidopsis*. Identical residues (5 or more) at a particular position are shown on a black background and conservative changes are on a grey background. A full stop indicates spacing introduced to maximise the alignment. Motifs specific for PTPs are also shown. The active site motif is overlined in black, the general acid motif containing an aspartate residue is overlined with a dashed line [8] and the phosphotyrosine-recognition subdomain (KNRY) is also overlined with a dashed line [40]. b. Alignment of the N-terminal regions of soybean (1–85), pea (1–82) and *Arabidopsis* (1–86) PTPs. Residues shown on a black background are conserved in two or more of the sequences, grey shading indicates conservative changes.

(Figure 4b). Finally, the peak fractions were shown to be the only fractions containing phosphatase activity against *p*NPP (Figure 4c). Taken together, these data show that active PTP is a monomeric protein.

Plant PTPs contain a conserved catalytic domain and a novel amino-terminal domain

The soybean, pea and *Arabidopsis* PTP-like proteins are predicted to contain 338, 332 and 340 amino acid residues respectively. As none of the sequences appeared to contain a predicted cleavable N-terminal signal peptide the mature proteins should be similar in size to the predicted values, ca. 38 kDa, i.e. similar to the value determined for the recombinant soybean PTP. The proteins contained a region of ca. 240 aa which is homologous to the catalytic domains of PTPs from other species (Figure 5a). The plant PTPs also contained 2 conserved motifs (sWPDh and PIVVHCSAGvGRTG; Figure 5a) specific to the PTP family of proteins rather than VH1-like, cdc25 or LMW phosphatase families [8], and a motif (KNRY) which permits the access of phosphotyrosine rather than shorter phosphoserine / phosphothreonine residues to the active site. The presence of these motifs clearly places the plant phosphatases within the PTP family of proteins and is consistent with the properties of the recombinant soybean PTP. Interestingly, the first 82–86 N-terminal residues of the plant PTPs (Figure 5b) do not show significant similarity to either targeting, regulatory or receptor sequences of other PTPs, nor to any predicted proteins within the publicly available databases.

Expression of PTP1_PEA

Expression of *PTP1_PEA* in developing pea embryos was monitored by probing an RNA gel blot of poly(A)$^+$ RNA isolated at different developmental stages (8–22 days after flowering; daf) with the pea *PTP1* cDNA sequence. Results for two pea genotypes [38] are shown in Figure 6. The probe hybridised to a mRNA species of ca. 1.5 kb and more weakly to a larger mRNA species. The 1.5 kb mRNA was most

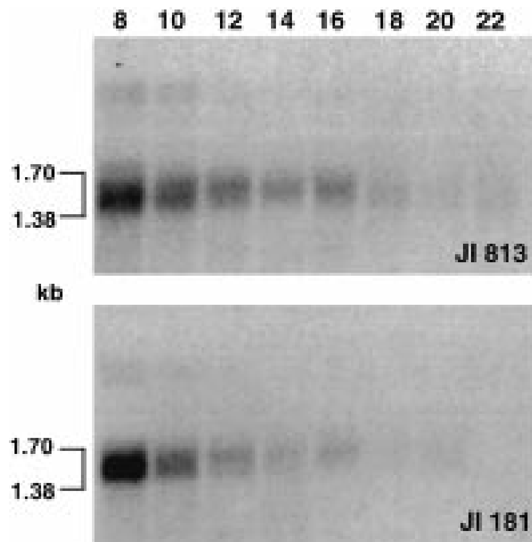


Figure 6. RNA gel blot analysis of PTP expression in developing embryos of two pea genotypes (JI 813 and JI 181). Poly(A)$^{+}$ RNA samples isolated from pea embryos 8–22 d.a.f. were probed with the longest *PTP1_PEA* cDNA. Filters were washed to a final stringency of 0.5× SSC 0.1% SDS at 65 °C and exposed to X-ray film for 3 weeks; size markers were glyoxylated restriction fragments of λ DNA.

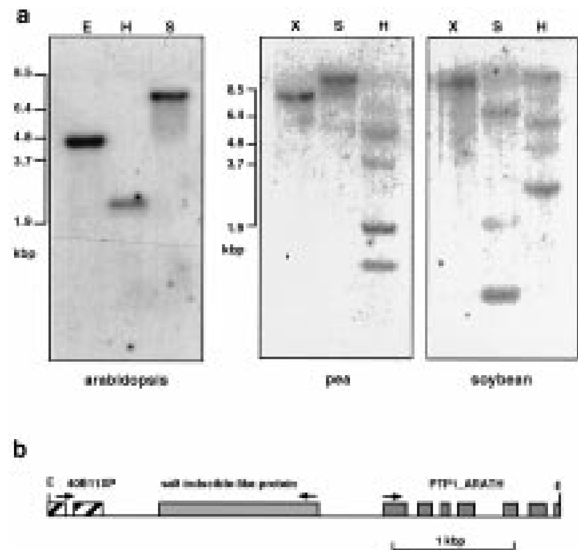


Figure 7. Genomic analysis of sequences encoding PTPs. a. Genomic DNA of *Arabidopsis*, pea and soybean was probed with the corresponding PTP sequence. The *Arabidopsis* probe, labelled by PCR, corresponded to the genomic sequence between primers AT1 and AT3 and the cDNA sequence between primers AT2 and AT3 (Figure 1). Probes for pea and soybean were made by random priming of the corresponding cDNAs. Restriction enzymes used were: E, *Eco*RI; H, *Hin*dIII; X, *Xba*I; S, *Sac*I. Hybridising fragments were visualised by autoradiography and size markers were λDNA digested with *Bst*EII. b. Schematic representation of the 4.5 kbp genomic region derived from *Arabidopsis* BAC F18D12. Predicted exons are shown as boxes, arrows indicate the direction of transcription. The gene furthest 5′ to *PTP1_ARATH* corresponds to EST 40B11XP and contains a WD40 domain. The gene immediately 5′ to *PTP1_ARATH* bears some similarity to numerous membrane-associated salt-inducible proteins.

abundant in embryos 8 daf, and then declined, so that by 18 daf it was almost undetectable. A similar pattern of expression was observed for the larger mRNA suggesting that it may be a precursor of the 1.5 kb mRNA. Transcript abundance declined more rapidly in line JI 181 than in JI 813. Hybridisations with protein phosphatase 2A catalytic and regulatory subunit probes have previously been published [7], in both cases, transcripts were detected from 8–22 daf, with only slight variations, and can be considered a suitable 'control' for loading. The steady state level of the major *PTP1* mRNA species appears to be low, since signals were undetectable in RNA gel blots of total RNA preparations from embryos, leaf and root tissues. However, RT-PCR experiments suggest that expression of pea *PTP*-like genes is not restricted to embryos as transcripts could be detected in pods, leaves, stems and tendrils (data not shown).

Genomic analysis

Southern blotting of genomic DNA from *Arabidopsis*, pea and soybean was used to estimate copy numbers for *PTP* genes. The results (Figure 7a) suggests that only a single gene was present in *Arabidopsis*, even after a relatively low stringency wash (2× SSC at 65 °C). The 4.5 kbp *Eco*RI fragment observed on the blot corresponds to the 4.5 kbp fragment in the genomic clone F18D12 which contained part of *PTP1_ARATH* (Figure 7b, see above). In genomic DNA from both pea and soybean, the Southern blots suggested the presence of a small family of PTP-encoding genes.

Discussion

The present paper reports the first molecular characterisation of PTPs in plant species other than *Arabidopsis*, and confirms and extends the characterisation of the *Arabidopsis* PTP recently reported by Xu *et al.* [39]. The predicted plant PTPs are polypeptides of M_r ca. 38 000 and contain a domain with greatest similarity to the PTP family of tyrosine phosphatases. When expressed as a recombinant protein, the soybean PTP showed specificity towards phosphotyrosine

containing substrates. Hydrolysis of phosphotyrosine but not phosphoserine or phosphothreonine was consistent with *Yersinia* and rat PTPs which hydrolyse phosphoserine or phosphothreonine several orders of magnitude more slowly than phosphotyrosine [42]. The plant PTP was inhibited by vanadate and dephostatin but not by cantharidin or metal-chelating agents. The requirement for reducing agents suggests that the active site cysteine residue was very susceptible to oxidation, a feature common to tyrosine phosphatases [37]. Enzymes with the properties of PTPs, which dephosphorylate phophotyrosine-containing peptides, but not phosphoserine or phosphothreonine residues, have been partially purified from pea nuclei and wheat seedlings [6, 12]. The properties of the pea nuclei enzyme are similar to those determined for the recombinant soybean PTP, in that it had a pH optimum of ca. 5.5 (with a protein substrate) and K_m values in the μM range for peptide substrates, was insensitive to inhibition by fluoride or okadaic acid, but was strongly inhibited by vanadate. However, a M_r of ca. 90 000 was estimated for the pea enzyme, much larger than the predicted molecular weight for the PTPs described here. The wheat enzyme also had similar functional properties and an estimated M_r of 37 000. However, interestingly, the wheat tyrosine phosphatase was a glycoprotein, and although the PTP1s have multiple potential N-linked glycosylation sites, they have no predicted signal peptide and thus would not be expected to enter the endoplasmic reticulum and be glycosylated. It is thus at present not clear that either of the enzyme activities previously reported correspond to the PTP proteins described here. The PTP from *Arabidopsis* was expressed in *E. coli* as a recombinant protein by Xu *et al.* [39], and showed similar specific hydrolytic activities to phosphotyrosine- and phosphoserine/threonine-containing protein substrates, and sensitivities to inhibition by vanadate and okadaic acid as the soybean PTP expressed in the present paper. However, Xu *et al.* did not report pH optima or K_m/V_{max} values for the *Arabidopsis* enzyme, and carried out all PTP assays at pH 7.0.

The *AtPTP1* cDNA sequence reported by Xu *et al.* [39] shows a significant discrepancy with the composite sequence from *Arabidopsis* cDNA and genomic clones reported in the present paper. The previously reported cDNA diverges completely from the present sequence after base 2105 (Figure 1), and continues to encode a 'complete' cDNA 1973 bp in length. The size determined for the mRNA corresponding to *PTP1_PEA*, 1.5 kb, is thus not in agreement with the size of the *Arabidopsis AtPTP1* cDNA, 1973 bp [39], but does agree with both the pea and soybean cDNAs and the predicted *Arabidopsis* PTP1 mRNA size derived from the sequences described here. Unfortunately, no size calibration is given for the mRNA corresponding to *AtPTP1* in the RNA gel blots shown by Xu *et al.* [39]. When the diverged 3′-flanking sequence in the *AtPTP1* cDNA was used separately to search the sequence databases, it was found to be identical to 6 previously identified *Arabidopsis* ESTs, and to contain a reading frame (on the oppposite strand) encoding a predicted polypeptide with homology to lipid transfer proteins. We therefore concluded that the previously published *AtPTP1* cDNA sequence was chimeric. A correction to the original report has been published elsewhere (Plant Cell, October issue 1998). The occurrence of chimeric clones in the widely available *Arabidopsis* cDNA libraries has been noted, and is emphasised by our analyses.

RNA gel blot analysis showed that pea *PTP1* transcripts are not abundant in pea embryos, as signals could only be obtained using poly(A)$^+$ RNA preparations with long exposure times and were not detected in total RNA prepared from embryos, other organs or soybean cell cultures. The data for mRNA levels of *PTP1* in developing pea embryos suggests that expression is developmentally regulated; not only is a decrease in PTP mRNA level seen as embryo development passes from cell division to cell expansion phases, but also the two lines differ in both their developmental patterns [38] and in the decrease in levels of PTP mRNA with development, with expression in line JI 813 continuing at higher levels for a longer period. Line JI 813 does not show such a clear transition from cell division to cell expansion as line JI 181 [38]. Clearly, further studies are required to understand the significance, if any, of this difference.

Xu *et al.* [39] concluded that expression of *AtPTP1* was stress-responsive, on the basis of changes in mRNA levels under salt and low temperature stress conditions, and thus deduced that PTP had a role in stress responses in plants. PTP genes may be both stress-responsive and developmentally regulated in embryos; other stress-responsive genes are known to be expressed during embryogenesis. PTP expression is not confined to embryos since transcripts were detected in other pea tissues in the present paper, and were detected in roots, leaves, flowers and stems in *Arabidopsis* [39]. The finding that only one EST sequence has been registered which corresponded to the transcript predicted from the *Arabidopsis PTP1* gene

is consistent with relatively low transcript levels estimated in the present paper, at least in the libraries used for the EST sequencing programmes. In contrast to our results, Xu *et al*. [39] could detect *AtPTP1* transcripts in total RNA preparations from all *Arabidopsis* tissues examined, apparently as a relatively abundant message. The reason for this discrepancy in the estimates of the quantitative levels of expression of PTP genes is not clear at present.

In *Arabidopsis* there appears to be only one *PTP1*-like gene, whilst in pea and soybean evidence suggested small gene families. However, it is likely that there are other genes encoding members of the tyrosine phosphatase superfamily in higher plants. These may only be weakly similar to the PTP encoding sequences and would not therefore be detected by hybridisation analyses presented here. In support of this assumption, the PTPs described here are divergent to the *Chlamydomonas VH-PTP13* dual-specificity phosphatase, outside of the active site region, and homologues of this gene which may exist in higher plants [13] would not be detected by the probes used here. Furthermore, there is a predicted *Arabidopsis* gene (mul8.6; Genbank AB009054) which contains a motif (IAVVHCMAGKGRTG) similar to the active site regions of tyrosine phosphatases and is most similar to P-TEN, a mammalian dual-specificity phosphatase which acts as a tumour suppressor [22].

Despite the molecular and functional characterisation of PTPs, their role(s) in the plant are not fully characterised. The lack of predicted signal peptides or transmembrane spanning regions suggests they should be classified as non-receptor type, but their location(s) within the cell are at present unknown. By analogy with other PTPs, a role in signal transduction processes can be predicted. As could be anticipated from their designation as PTPs, these proteins do not contain additional motifs specific for either dual-specificity or cdc25 phosphatases and thus it would seem unlikely that they, respectively, function as MAP kinase phosphatases, nor are they cdc25 homologues. However, at least in yeast, PTPs can dephosphorylate some MAP kinases (reviewed in [36]) and a similar activity in plants cannot be ruled out. Whilst the catalytic domains of plant PTPs contain the consensus motifs of PTPs, and the *Arabidopsis* and soybean PTPs have the enzymatic characterisitics of PTPs, a novel feature of these proteins lies within their N-terminal regions. Previously described non-receptor-type PTPs contain additional domains important in regulating activity by targeting to specific substrates or cellular locations. The N-terminal regions of plant PTPs described here may also represent such a domain, but one which is different to those previously described in other eukaryotes. Identifying the substrates of plant PTPs and the function of their N-terminal domains should allow insight into the signalling pathways in which they operate.

Acknowledgements

The supply of material from the ABRC, Ohio, US is acknowledged with thanks. A.F.S. gratefully acknowledges present funding from the Scottish Office which enabled this work to be published and advice from Ewan Birney (Sanger Centre, Hinxton, UK) on database analyses.

References

1. Adam AL, Pike S, Hoyos ME, Stone JM, Walker JC, Novacky A: Rapid and transient activation of a myelin basic protein kinase in tobacco leaves treated with harpin from *Erwinia amylovora*. Plant Physiol 115: 853–861 (1997).
2. Ali N, Halfter U, Chua NH: Cloning and biochemical characterization of a plant protein kinase that phosphorylates serine, threonine, and tyrosine. J Biol Chem 269: 31626–31629 (1994).
3. Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucl Acids Res 25: 3389–3402 (1997).
4. Barford D, Jia ZC, Tonks NK: Protein tyrosine phosphatases take off. Nature Struct Biol 2: 1043–1053 (1995).
5. Bianchi MW, Guivarch D, Thomas M, Woodgett JR, Kreiss M: *Arabidopsis* homologs of the *shaggy* and *Gsk-3* protein kinases: Molecular cloning and functional expression in *Escherichia coli*. Mol Gen Genet 242: 337–345 (1994).
6. Cheng H-F, Tao M: Purification and characterization of a phosphotyrosyl-protein phosphatase from wheat seedlings. Biochim Biophys Acta 998: 271–276 (1989).
7. Evans IM, Fawcett T, Boulter D, Fordham-Skelton AP: A homolog of the 65 kDa regulatory subunit of protein phosphatase 2A in early pea (*Pisum sativum* L.) embryos. Plant Mol Biol 24: 689–695 (1994).
8. Fauman EB, Saper MA: Structure and function of the protein tyrosine phosphatases. Trends Biochem Sci 21: 413–417 (1996).
9. Feinberg AP, Vogelstein B: A technique for radiolabelling DNA restriction fragments to high specific activity. Anal Biochem 132: 6–13 (1984).
10. Fillipini F, Rossi V, Marin O, Trovato M, Costantino P, Downey PM, Sciavo FL, Terzi M: A plant oncogene as a phosphatase. Nature 379: 499–500 (1996).
11. Fordham-Skelton AP, Yarwood A, Croy RRD: Synthesis of saporin gene probes from partial protein sequence data – use of inosine oligonucleotides, genomic DNA and the polymerase chain reaction. Mol Gen Genet 221: 134–138 (1990).

12. Guo YL, Roux SJ: Partial purification and characterization of an enzyme from pea nuclei with protein tyrosine phosphatase activity. Plant Physiol 107: 167–175 (1995).
13. Haring MA, Siderius M, Jonak C, Hirt H, Walton KM, Musgrave A: Tyrosine phosphatase signaling in a lower plant – cell-cycle and oxidative stress-regulated expression of the *Chlamydomonas eugametos Vh-Ptp13* gene. Plant J 7: 981–988 (1995).
14. Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S: Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. Nucl Acids Res 24: 3439–3452 (1996).
15. Hirayama T, Oka A: Novel protein kinase of *Arabidopsis thaliana* (*Apk1*) that phosphorylates tyrosine, serine and threonine. Plant Mol Biol 20: 653–662 (1992).
16. Hirt H: Multiple roles of MAP kinases in plant signal transduction. Trends Biochem Sci 2: 11–15 (1997).
17. Kieber JJ, Rothenberg M, Roman G, Feldman KA, Ecker JR: *CTR1*, a negative regulator of the ethylene response pathway in *Arabidopsis*, encodes a member of the RAF family of protein-kinases. Cell 72: 427–441 (1993).
18. Knetsch MLW, Wang M, Snaarjagalska BE, Heimovaara-Dijkstra S: Abscisic acid induces mitogen-activated protein kinase activation in barley aleurone protoplasts. Plant Cell 8: 1061–1067 (1996).
19. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227: 680–685 (1970).
20. MacKintosh C, MacKintosh RW: Inhibitors of protein kinases and phosphatases. Trends Biochem Sci 19: 444–448 (1994).
21. Mu JH, Lee HS, Kao TH: Characterization of a pollen-expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. Plant Cell 6: 709–721 (1994).
22. Myers MP, Stolarov JP, Eng C, Li J, Wang SI, Wigler MH, Parsons R, Tonks NK: P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. Proc Natl Acad Sci USA 94: 9052–9057 (1997).
23. Neel BG, Tonks NK: Protein tyrosine phosphatases in signal transduction. Curr Opin Cell Biol 9: 193–204 (1997).
24. Newman T, Debruijn FJ, Green P, Keegstra K, Kende H, McIntosh L, Ohlrogge J, Raikhel N, Somerville S, Thomashow M, Retzel E, Somerville C: Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 106: 1241–1255 (1994).
25. Rodriguez-Zapata LC, Hernandez-Sotomayor SMT: Evidence of protein-tyrosine kinase activity in *Catharanthus roseus* roots transformed by *Agrobacterium rhizogenes*. Planta 204: 70–77 (1998).
26. Sambrook J, Fritsch EF, Maniatis T: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).
27. Shi P-Y, Maizels N, Weiner AM: Recovery of soluble, active recombinant protein from inclusion bodies. Biotechniques 23: 38–39 (1998).
28. Smith RD, Walker JC: Plant protein phosphatases. Annu Rev Plant Physiol Plant Mol Biol 47: 101–125 (1996).
29. Sommer D, Wells TA, Song PS: A possible tyrosine phosphorylation of phytochrome. FEBS Lett 393: 161–166 (1996).
30. Stone JM, Walker JC: Plant protein kinase families and signal transduction. Plant Physiol 108: 451–457 (1995).
31. Suzuki K, Shinshi H: Transient activation and tyrosine phosphorylation of a protein kinase in tobacco cells treated with a fungal elicitor. Plant Cell 7: 639–647 (1995).
32. Tong CG, Kendrick RE, Roux SJ: Red light-induced appearance of phosphotyrosine-like epitopes on nuclear proteins from pea (*Pisum sativum* L.) plumules. Photochem Photobiol 64: 863–866 (1996).
33. Torruella M, Casano LM, Vallejos RH: Evidence of the activity of tyrosine kinase(s) and of the presence of phosphotyrosine proteins in pea plantlets. J Biol Chem 261: 6651–6653 (1986).
34. Trojanek J, Ek P, Scoble J, Muszynska G, Engstrom L: Phosphorylation of plant proteins and the identification of protein-tyrosine kinase activity in maize seedlings. Eur J Biochem 235: 338–344 (1996).
35. Usami S, Banno H, Ito Y, Nishihama R, Machida Y: Cutting activates a 46-kilodalton protein kinase in plants. Proc Natl Acad Sci USA 92: 8660–8664 (1995).
36. van Vactor D, O'Reilly AM, Neel BG: Genetic analysis of protein tyrosine phosphatases. Curr Opin Genet Dev 8: 112–126 (1998).
37. Walton KM, Dixon JE: Protein tyrosine phosphatases. Annu Rev Biochem 62: 101–120 (1993).
38. Wang TL, Hedley CL: Genetic and developmental analysis of the seed. In: Casey R, Davies DR (eds) Peas: Genetics, Molecular Biology and Biotechnology, pp. 83–120. CAB International, Wallingford, Oxfordshire, UK (1993).
39. Xu Q, Fu H-F, Gupta R, Luan S: Molecular cloning and characterisation of a tyrosine-specific protein phosphatase encoded by a stress-responsive gene in *Arabidopsis*. Plant Cell 10: 849–857 (1998).
40. Yuvaniyama J, Denu JM, Dixon JE, Saper MA: Crystal structure of the dual-specificity protein phosphatase Vhr. Science 272: 1328–1331 (1996).
41. Zhang SQ, Klessig DF: Salicylic acid activates a 48-kD MAP kinase in tobacco. Plant Cell 9: 809–824 (1997).
42. Zhang ZY, Wang Y, Dixon JE: Dissecting the catalytic mechanism of protein tyrosine phosphatases. Proc Natl Acad Sci USA 91: 1624–1628 (1994).