

Short communication

A cDNA from pea petals with sequence similarity to pollen allergen, cytokinin-induced and genetic tumour-specific genes: identification of a new family of related sequences

Anthony J. Michael

Department of Genetics and Microbiology, Institute of Food Research, Norwich Research Park, Colney, Norwich, NR4 7UA, UK

Received 16 March 1995; accepted in revised form 9 October 1995

Key words: pollen allergen, signal peptide, pea petal, expressed sequence tag

Abstract

A cDNA isolated from pea petals exhibits extensive similarity to pollen allergen genes, a cytokinin-regulated cDNA from soybean suspension cultures, a partial cDNA preferentially expressed in tobacco genetic tumours, four *Arabidopsis* expressed sequence tags (ESTs) and fifteen rice ESTs. This diverse family of pollen allergen-like genes may have a common ancestor or at least share common functional domains. Possession of a putative signal peptide and a presumed extracellular location is a common aspect of this family of sequences.

A cDNA was isolated during the screening of a pea petal-specific cDNA library. The library contained cDNA prepared from mRNA derived from young pea standard and wing petals of the line JI813. Petals had only just begun to develop red pigmentation at the rim of the wing. The cDNAs were cloned unidirectionally in the lambda Uni-zap XR phage vector (Stratagene). After three rounds of screening, the cDNAs were excised in pBluescript from the phage vector using an *in vivo* excision process employing SOLR cells and the ExAssist helper phage (Stratagene). Sequencing of the 5' and 3' ends of the cDNA inserts indicated that one clone (PPA1) exhibited homology to rye grass pollen allergen gene *Lol pI* [2]. The

entire PPA1 cDNA was sequenced on both strands in pBluescript by an oligonucleotide walking strategy employing an ABI 373A automated sequencer (Applied Biosystems). The cDNA insert was 822 bp and contained a potential open reading frame of 258 amino acids with 18 bp of untranslated 5' leader sequence (Fig. 1).

The University of Wisconsin GCG FASTA program revealed that the PPA1 predicted amino acid sequence was 29% identical to that of the rye grass pollen allergen *Lol pI* (Fig. 2), responsible for producing the major rye grass pollen allergen [2, 3]; its *in vivo* function is unknown. A cDNA for *Lol pI* was previously isolated from a pollen-specific library [2]. The protein product may be

```

1  gtctctctctctctcttgcgaatggctgcaattcttctcctcaccattgcatcactaacctca 60
1  .                               M A A I L L L T I A S L T S      20

61  ctttttgcaccaaccacgcctgaatccctggagtttacacggcgcccttggaccagt 120
21  L F A P T T A R I P G V Y T G G P W T S      40

121 gctcatgctactttctacggcggcagtgatgcctccggaacaatgggtgggtgcttgtggg 180
41  A H A T F Y G G S D A S G T M G G A C G      60

181 tacggaaacttatcacagccaaggctacggcggtcaacaccgcagcactaagcacagcacta 240
61  Y G N L Y S Q G Y G V N T A A L S T A L      80

241 ttcaacaatggcttaagttgcgggtgcatgttttgagctcaagtgtgaccaagatccacgg 300
81  F N N G L S C G A C F E L K C D Q D P R      100

301 tgggtgaaccctggaaacccttccatcttaatactgcaacaaacttttgcctccaaat 360
101 W C N P G N P S I L I T A T N F C P P N      120

361 ttcgctgaacctagtgacaatgggtgggtgtaaccacacctcgaccacactttgatctt 420
121 F A E P S D N G G W C N P P R P H F D L      140

421 gccatgcccatgttcttaagatcgctcaataccgcgcgggaatcgccccgttgcttac 480
141 A M P M F L K I A Q Y R A G I V P V A Y      160

481 cgccgtgtgccatgcagaaaggcaggagcatcagattcacaatcaacgggtttccgttac 540
161 R R V P C R K A G G I R F T I N G F R Y      180

541 ttcaacttggttctaatcaccaacgtgcgggtgcaggggatatcggtgcgctgagtgtta 600
181 F N L V L I T N V A G A G D I V R V S V      200

601 aaaggaaccaacacagcgtggatgacaatgagccgtaactgggggcagaactggcaatcc 660
201 K G T N T A W M T M S R N W G Q N W Q S      220

661 aacgcagtttttgtgggccaggtctctcttttagggtcaccggcagtgaccgtcgtaca 720
221 N A V F V G Q A L S F R V T G S D R R T      240

721 tccacgtcatggaacgtggcaccacctcattggcagttcggacaaccttcacaggaaag 780
241 S T S W N V A P P H W Q F G Q T F T G K      260

781 aatttcgcgctctgaaaaaattattataaaaaaaaaaaaaaaaaa 822
261 N F R V *                                     274

```

Fig. 1. Nucleotide sequence and deduced amino acid sequence of the pea PPA1 cDNA. The single-letter amino acid code is used to represent the protein sequence. An asterisk represents the TGA termination codon.

present on the surface of the pollen, thereby exerting its allergenic effect [2]. A specific region of the protein was determined to be responsible for the allergenicity of *Lol pI* [2]. Two other closely related pollen allergen genes (*PhlpI* from *Phleum pratense* and *Hol-I* from *Holcus lanatus*) exhibit similarity to PPA1 (see legend to Fig. 2 for accession numbers). An additional partial sequence of a pollen allergen homologue from maize, *Zea mI*, also exhibits similarity to PPA1 [4] but because the sequence is partial, it is not included in Fig. 2.

The PPA1 sequence is also 30% identical to the cytokinin-regulated soybean suspension cul-

ture *cim1* protein sequence [5]. The sequence of *cim1* contains a putative signal sequence followed by relatively hydrophilic residues. Cytokinin application caused rapid induction of *cim1* mRNA accumulation in soybean suspension cultures, leading Crowell to postulate [5] that *cim1*, as a putatively extracellular protein induced by cytokinin, may be involved in the deposition of cell wall material and in the formation of the cell division plate during cytokinesis. A partial cDNA clone, TID301, representing a gene preferentially expressed during the development of tobacco genetic tumours [6], also exhibits similarity to PPA1 (Fig. 2). The genetic tumours from which TID301

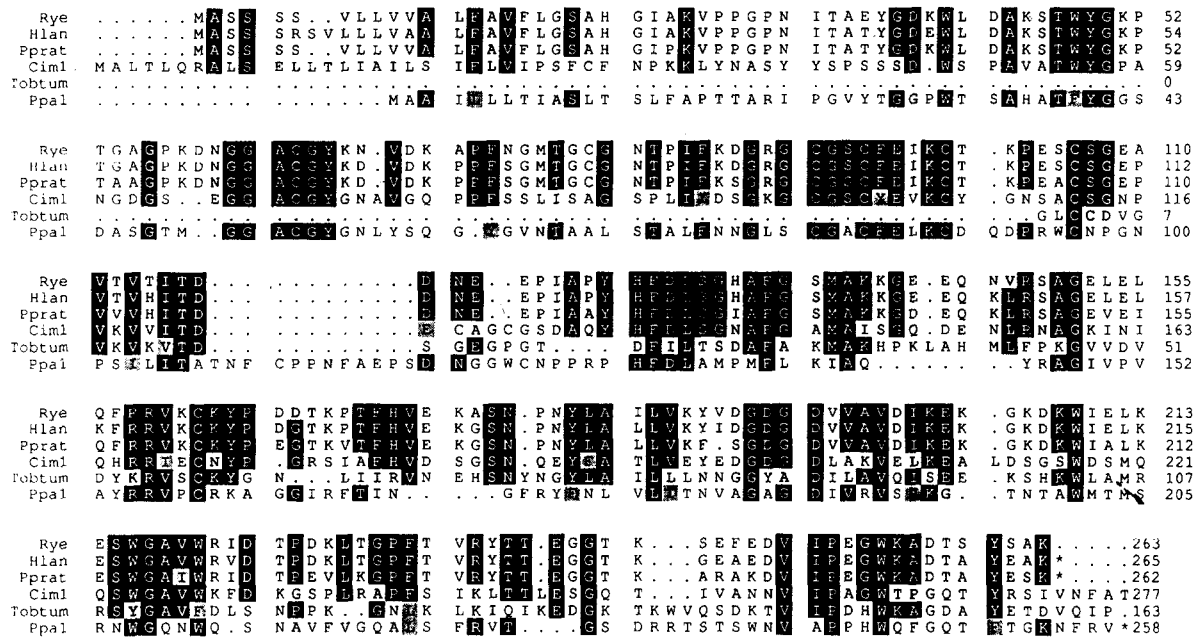


Fig. 2. Comparison of PPA1 with pollen allergen related sequences. The seven sequences are aligned using the PILEUP and PRETTYBOX programs [20]. Solid black boxes represent identical amino acids, dark hatched boxes indicate similar amino acids and lightly hatched boxes related amino acids. The Rye sequence represents *Lol p1* from rye (accession no. M57474), Hlan represents *Hol-II* from *H. lanatus* (accession number Z27084), Pprat represents *PhlpI* from *P. pratense* (accession number Z27090), Cim1 represents *cim1* from soybean (accession number U03860), Tobtum represents TID301 (accession number D26459), Ppa1 represents the pea PPA1 sequence and Prpro represents PR4b from tobacco (accession number PIR3: S23800). The sequence of TID301 is partial.

was isolated arose on the progeny of interspecific crosses between *Nicotiana glauca* and *Nicotiana langsdorffii*. TID301 is not detectably expressed in the leaves or stems of either *N. glauca* or *N. langsdorffii* but is expressed in genetic tumours. It has been suggested that cytokinin is involved in the formation of genetic tumours [6]. Expression of TID301 may be subject to regulation by cytokinin as there is evidence that the expression of the endogenous *rolB* and *C* genes of the *N. glauca* genome is closely correlated with the extent of the tumorous state [7]. Recently the product of the *rolC* gene was shown to exhibit a cytokinin-N-glucosidase activity thereby releasing free cytokinin from apparently inert conjugates [8]. The relatedness of the amino acid sequences in Fig. 2 is greater than the displayed identities because of the conserved hydrophobic and aromatic residues, which are not indicated by the PRETTYBOX display.

PPA1 displays a more limited similarity to several pathogenesis-related (PR) and wound-induced proteins such as the deduced amino acid sequence of the PR-4b clone (accession number PIR3: S23800) [9] over the first 60% of the PPA1 sequence; (not shown). The limited similarity to the PR proteins was detected by using the GCG program PROFILESEARCH employing a profile of *Lol p1*, *Hol-II*, *PhlpI*, *cim1*, TID301 and PPA1 to search the PIR databases. After detecting the pollen allergen proteins, the next most similar amino acid sequences were of the PR4 group of proteins. Three other PR protein sequences also display limited similarity to PPA1 (accession nos. PIR3: S23801, PIR2: S18598 and PIR3: S23799) belonging to the PR-4 family of proteins [9, 10], for which the *in vivo* function is not known. A soybean protein (accession number PIR2: S28028) displaying homology [11] with potato wound-induced proteins [12] also exhib-

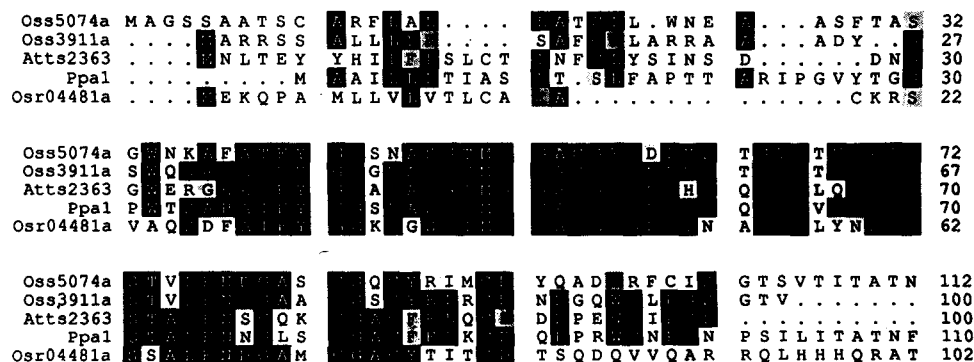


Fig. 3. Comparison of the N-terminus of PPA1 with the deduced amino acid sequences of one *Arabidopsis* and three rice ESTs. The sequences were aligned with the PILEUP and PRETTYBOX programs. The *Arabidopsis* Atts2363 clone has the accession number Z29931 and the three rice clones Oss5074a, Oss3911a and Osr04481a have the accession numbers D42004, D41419 and D23861 respectively.

ited limited similarity to PPA1 (not shown). There is an interesting parallel with the homology detected between the pollen allergen protein of birch *BetvI* [13], the PR1 group of pathogenesis-related proteins [14] from bean and ABA-responsive proteins from pea [15].

The UW GCG package TFASTA program was used to search the GenBank/EMBL databases and uncovered 15 rice and 4 *Arabidopsis* ESTs with hypothetical protein sequences exhibiting a much higher degree of homology with PPA1 than either *LolpI* or the pollen allergen genes, *cim1* or TID301 (Fig. 3), suggesting that the genes represented by the *Arabidopsis* and rice ESTs are closer in function to the pea PPA1 gene than PPA1 is to the legume-derived *cim1*. The ESTs are randomly chosen cDNA clones sequenced by one pass representing ca. 350 bp, usually at the 5' end. By reference to PPA1, one of the *Arabidopsis* ESTs appears to be complete at the N-terminus (Atts2363) and there are three different but related rice sequences also complete at the N-terminus (Oss5074a, Oss3911a and Osr04481a). With respect to the PPA1 sequence, the other *Arabidopsis* and rice ESTs are probably partial clones with no area of overlap with the N-terminus complete EST sequences so that it was not possible to determine if these partial clones represented additional incomplete copies of the full-length EST sequences or different genes.

A hydropathy plot of PPA1 [16] indicated that the N-terminus possesses a putative signal sequence terminating at residue 22 so that the mature PPA1 peptide would be 236 residues long (Fig. 4). This suggests that PPA1 may be exported to the apoplast as has been suggested may also be the case for the soybean *cim1* hypothetical protein [5]; the pollen allergen proteins may be present on the surface of pollen as *LolpI* also possesses a putative signal sequence [3]. The PR4 group of pathogenesis-related proteins possesses a hydrophobic N-terminus domain reminiscent of a signal peptide, agreeing with the extracellular locations of PR4 [10]. Each of the *Arabidopsis* and rice ESTs shown in Fig. 3 possesses a putative signal sequence as defined by the SIGCLEAVE

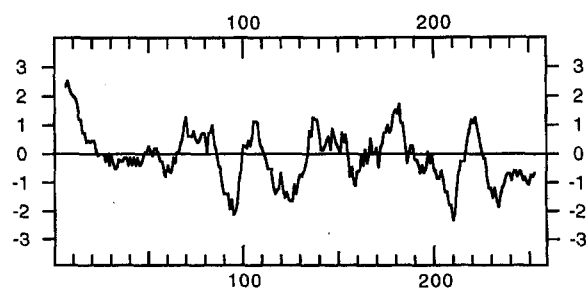


Fig. 4. Hydropathy plot of the pea PPA1 protein sequence. The amino acid sequence of PPA1 shown in Fig. 1 was analysed by the method of Kyte and Doolittle [14] to derive the hydropathic profile of the PPA1 protein product. Hydrophobic regions are represented by positive values.

program of the UW GCG package. The hypothetical mature protein of Oss04481a starts at residue 19, that of Oss3911a also at residue 19, that of Oss5074a at residue 29 and that of Atts2363 at residue 24 on their respective sequences shown in Fig. 3. Neither PPA1, *cim1*, TID301 or PR4b possess a region with homology to the peptide fragment of *Lol pI* implicated in its allergenic activity [2].

Southern analysis [17] (Fig. 5) of pea genomic DNA digested with enzymes that do not cut in the PPA1 sequence, using the PPA1 cDNA as a probe, revealed single bands after high stringency washes (2×30 min in $0.1 \times$ SSC and 0.1% SDS at 65°C). This indicates that the PPA1 gene may be present as a single-copy gene in pea although the results of the Southern analysis do not rule out closely spaced copies. Formaldehyde agarose RNA gel blots [18] were used to analyse the expression of the PPA1 gene (Fig. 6) in immature and mature petals and young leaves of pea plants. PPA1 transcript accumulated more in young unpigmented petals than in mature pigmented petals and was expressed at a low level in young leaves.

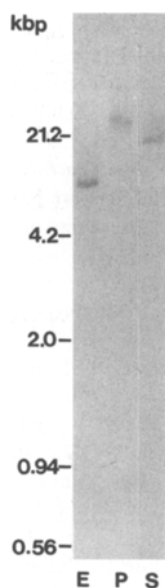


Fig. 5. Southern analysis of the PPA1 gene. $30 \mu\text{g}$ of pea genomic DNA was loaded per track on a 1.0% agarose gel. The DNA was digested with *Eco* RI (E), *Pst* I (P) and *Sal* I (S).

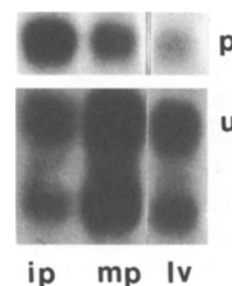


Fig. 6. RNA gel blot of the PPA1 transcript. Each track was loaded with $0.5 \mu\text{g}$ of poly(A)⁺ RNA produced by two passages of total RNA over an oligo d(T) cellulose column. The lane marked ip is loaded with mRNA from immature pea petals before pigmentation commenced (derived from the pea line JI813), lane mp is loaded with mRNA from mature red petals of the line JI813, lane lv is mRNA from young leaves of the same line. Panel p is the nitrocellulose filter probed with the PPA1 cDNA and panel u is the same filter stripped and then hybridised with the *Antirrhinum* ubiquitin cDNA. Washes were at high stringency (2×30 min in $0.1 \times$ SSC and 0.1% SDS at 65°C) and the temperature was reduced to 48°C the ubiquitin probe.

The expression pattern of PPA1 is similar to that of *Lol pI*, which is expressed only in pollen and inflorescences, with expression being greater in inflorescences than in pollen [3]. The *Zea mI* gene is expressed in pollen but not in roots, leaves or etiolated seedlings; however mRNA accumulation was not analysed in the different floral organs [4]. The expression pattern of the *cim1* gene has only been analysed in soybean suspension cultures [5] and the TID301 gene is expressed in genetic tumours but not in stems, leaves or callus, however its expression was not analysed in flowers [6]. It is interesting to note that PR proteins in general, besides being induced by pathogen attack are developmentally regulated in tobacco flowers [19] and probably function in the normal process of flowering and reproductive physiology.

In conclusion, sequence analysis of a pea petal cDNA has uncovered a family of pollen allergen-like genes from diverse species expressed in flowers and other tissues. Each member of this group possesses a putative signal sequence (the TID301 clone is not complete at the N-terminus) and may be an extracellular protein. The *cim1* clone is regulated, and TID301 may be regulated, by cytokinin. It is not known if cytokinin regulates the

other members of this family. Several *Arabidopsis* and rice ESTs exhibit a much higher homology to the pea PPA1 sequence than to the other pollen allergen-like family members. The region responsible for the allergenicity of the *bona fide* pollen allergen genes is not conserved in the other sequences. It is possible that the pollen allergen-like gene family may have a common ancestor or at least share common functional domains that may be shuffled between genes during the course of plant evolution. In this context it will be interesting to know if PPA1 is regulated by cytokinin and whether PPA1 expression is associated with genetic tumours in pea. Experiments are now being initiated to examine these possibilities.

Acknowledgements

I would like to thank Julie Hofer, Mike Rhodes, Nick Walton and Noel Ellis for critical reading of the manuscript. I would also like to thank Julie Hofer for constructing the cDNA library and John Payne for operating the ABI373A. The *An-tirrhinum* ubiquitin cDNA was a kind gift of Cathie Martin.

References

1. Licht JD, Grossel MJ, Figge J, Hansen UM: *Drosophila Krüppel* protein is a transcriptional repressor. *Nature* 346: 76–79 (1990).
2. Perez M, Ishioka GY, Walker LE, Chesnut RW: cDNA cloning and immunological characterization of the rye grass allergen *LolpI*. *J Biol Chem* 265: 16210–16215 (1990).
3. Griffith IJ, Smith PM, Pollock J, Theerakulpisut P, Avjoglou A, Davies S, Hough T, Singh MB, Simpson RJ, Ward LD, Knox B: Cloning and sequencing of *LolpI* the major allergenic protein of rye-grass pollen. *FEBS Lett* 279: 210–215 (1991).
4. Bradwater AH, Rubinstein AL, Chay CH, Klapper DG, Bedinger PA: *Zea mI*, the maize homolog of the allergen-encoding *LolpI* gene of rye grass. *Gene* 131: 227–230 (1993).
5. Crowell DN: Cytokinin regulation of a soybean pollen allergen gene. *Plant Mol Biol* 25: 829–835 (1994).
6. Fujita T, Kouchi H, Ichikawa T, Syono K: Cloning of cDNAs for genes that are specifically or preferentially expressed during the development of tobacco genetic tumors. *Plant J* 5: 645–654 (1994).
7. Ichikawa T, Ozeki Y, Syono K: Evidence for the expression of the *rol* genes of *Nicotiana glauca* in genetic tumors of *N. glauca* × *N. langsdorffii*. *Mol Gen Genet* 220: 177–180 (1990).
8. Estruch J, Chriqui D, Grossman K, Schell J, Spena A: The plant oncogene *rolC* is responsible for the release of cytokinins from glucoside conjugates. *EMBO J* 10: 2889–2895 (1991).
9. Friedrich L, Moyer M, Ward E, Ryals J: Pathogenesis-related protein 4 is structurally homologous to the carboxy-terminus domains of hevein, Win-1 and Win-2. *Mol Gen Genet* 230: 113–119 (1991).
10. Linthorst HJM, Danhash N, Brederode FT, van Kan JAL, de Wit PJGM, Bol JF: Tobacco and tomato PR proteins homologous to *win* and pro-hevein lack the 'hevein' domain. *Mol Plant-Microb Interact* 4: 586–592 (1991).
11. Covello PS, Gray MW: Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome *c* oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. *EMBO J* 11: 3815–3820 (1992).
12. Stanford A, Bevan M, Northcote D: Differential expression within a family of novel wound-induced genes in potato. *Mol Gen Genet* 215: 200–208 (1989).
13. Breiteneder H, Pettenburger K, Bito A, Valenta R, Kraft D, Rumpold H, Scheiner O, Breitenbach M: The gene coding for the major birch pollen allergen *BetvI*, is highly homologous to a pea disease resistance response gene. *EMBO J* 8: 1935–1938 (1989).
14. Walter MH, Liu JW, Grand C, Lamb CJ, Hess D: Bean pathogenesis-related (PR) proteins deduced from elicitor-induced transcripts are members of a ubiquitous new class of conserved PR proteins including pollen allergens. *Mol Gen Genet* 222: 353–360 (1990).
15. Iturriaga EA, Leech MJ, Barratt DHP, Wang TL: Two ABA-responsive proteins from pea (*Pisum sativum* L.) are closely related to intracellular pathogenesis-related proteins. *Plant Mol Biol* 24: 235–240 (1994).
16. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132 (1982).
17. Southern EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98: 503–517 (1975).
18. Ausabel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (eds): *Current protocols in molecular biology*. Wiley Interscience, New York (1987).
19. Lotan T, Ori N, Fluhr R: Pathogenesis-related proteins are developmentally regulated in tobacco flowers. *Plant Cell* 1: 881–887 (1989).
20. Devereux J, Haeblerli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* 12: 387–395 (1984).