

**Plant Gene Register**

# The Extensin from *Prunus amygdalus*<sup>1</sup>

Jordi Garcia-Mas, Ramon Messeguer, Pere Arús, and Pere Puigdomènech\*

Unitat Mixta Institut de Recerca i Tecnològic Agroalimentària Centre d'Investigació Desenvolupament-Consejo Superior de Investigaciones Científicas, Departament de Genètica Molecular, Centre d'Investigació Desenvolupament-Consejo Superior de Investigaciones Científicas, Jordi Girona, 18, 08034 Barcelona, Spain (J.G.-M., P.P.); and Departament de Genètica Vegetal, IRTA, 08348 Cabriels, Spain (R.M., P.A.)

HRGP<sup>2</sup> are among the best-known components of the plant cell wall (2). In dicotyledonous species, where they have been called extensins, these proteins are formed by highly repeated elements including the sequence Ser-Pro<sub>4</sub> (7). The best-known HRGP from monocots is the one from maize that has been characterized at protein, cDNA, and genomic levels (8). The genes encoding HRGP are interesting systems in particular because their expression is developmentally regulated and they are induced by pathogen attack and wounding (3, 6).

Little information is available on the genes coding for cell-wall components of woody species. In particular, no information is available on the structure and expression of extensins in these species. The genus *Prunus* is an interesting group of species for several reasons: some of them comprise an important group of cultivated crops (stone fruits) and they contain an exceptionally low amount of DNA per haploid cell, i.e. *P. persica* (peach) is only twice as large as that of *Arabidopsis* (1). Extensins may be interesting probes for the study of seed development in these species. These probes are currently being used in restriction fragment-length polymorphism mapping of *Prunus*.

The extensin gene from *Prunus amygdalus* has been cloned by screening an almond tree root cDNA library with a tobacco extensin cDNA clone (4). The longest cDNA insert was sequenced and it contains an open reading frame with a methionine residue followed by a protein fragment having the typical features of a signal peptide and a repetitive proline-rich polypeptide. Therefore, the cDNA appears to contain the complete protein sequence. Six cDNAs have been sequenced; three of them are identical to the one presented here but they are shorter in their 5' end, indicating that they are incomplete cDNAs. In one case, a gap in the coding sequence has been observed. It does not change the reading frame and it is flanked by repeated sequences, indicating that it may be the result of recombination either in the plant or in the bacteria during the cloning procedure. The other three cDNAs are identical among themselves and have a small number of nucleotide substitutions (4 in 600 bp) when com-

pared with the other ones, reflecting either allelic differences or the presence of two genes in the haploid genome.

The protein sequence of the almond extensin is formed by three main repeating units: PYHYK, SPPPP, and SPSPPKH. These elements are well conserved along the sequence both at amino acid and nucleotide levels, indicating the possibility of homogenization events, as has been proposed for maize (5). The use of codons is not random, and in particular the G content is very low (7.2% of total nucleotides in the cDNA sequence). This nucleotide is mainly used in the third position for the lysine AAG codon. A single band is observed by northern analysis, and by Southern analysis bands corresponding to a single gene appear, although a small number of faintly hybridizing bands can also be observed. This indicates that the complexity of the extensin gene family is low in almond, but other related sequences may exist in its genome.

```
1 ACAA AAAAG
10 ATGGGAAAATGGGATCTTCATCAGTGACTTCTCTGTGGTCACTCTCTAGTGGCTATAGTGTCTCAGTTGGCCATCAGAAACCTCA
   N G K M G S S S V T S L V V T L L V A I V S L S L P S E T S   30
100 GCAAACTATGCTTACTGATCTCTCCACCACGATTTCTCTCCCTACCAATACAGTCCACACCTCTCTCTCTCTCTCTCTCTCTCTCTCT
   A N Y P Y S S P P P P V S P P V H Y K S P P P P S P T P V   60
190 CACTACTCACCACCTAAGCATCCCTACCATACAAAGTCAGTCTCTCTCCAGTGACTCTCACCCTACAGCATCCCTACCATACAAAG
   H Y S P P K N P Y H Y K S P P P P V H Y S P P K H P Y H Y K   90
280 TCTCTCTCTCCAGTGTACTCACCACCTAAGCATCCCTCTCCAGTCAAGTCTCCACCACCTCACCATCACCCTCAAGCACCC
   S P P P P V Y S P P K N P Y H Y K S P P P P S P S P P K H P   120
370 TACCCTACAGTCTCCACCACCTACCTTACCCTAAGCACCCCTACCCTACAGTCTCCACCACCTACCATCACCATCACC
   Y H Y K S P P P P S P S P P K N P Y H Y K S P P P P S P S P   150
460 CCTAAGCACCTACCACTACAGTCTCCACCACCTACCTTCCAGCTAAGCAGCCATACCCTACAGTCTCCACCACCTACCA
   P K N P Y H Y K S P P P P S P S P P K N P Y H Y K S P P P P   180
550 TCACCACCAGAACGCTTACCATACAGTCTCCACCACCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
   S P P P K P Y H Y K S P P P P S P T P P V Y S P P K H P Y   210
640 CACTACAAGTCTCCACCACCTACCAACCAAGAGCTTACCCTACAGTCTCCACCACCTACCAACCAAGAGCTTACCCTACAGT
   H Y K S P P P P S P P K P Y H Y K S P P P P T P V Y K P P   240
730 GTTACTCACCACCTCACCACCTAAGAACCATACAGCACCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
   V Y S P P P P P K P Y K P P T P P V H T A P P H P Y I Y S   270
820 TCACCACCCTCTCTCACCCTACAGTAAAGAGAGCTAGCTGATGATGAGTGGTAAAGTGAAGCAGCTCAATAAAGAGGGGG
   S P P P P H H Y Y   278
910 TGTCGAAGCAAGCCCTTATCCCTACCAAAAGAGAGAGTGGAGTCTCCGAGTTTGTATGACGTTTATTTATTCAGCAGTGTGTA
1000 TTTTATTAGACAGTGTAGTGCTCTGTTTATTATTATTTCTCAAAAGAGATGCCATGATGTTGTGCGTTATTTCACTCTCTCT
1090 CCTAATGGAGAGATCATATAATATAAATAATAGCTTTCCCAn
```

**Figure 1.** Nucleotide sequence and deduced amino acid sequence of a cDNA from *P. amygdalus* encoding extensin. The nucleotide sequence is numbered from the first nucleotide of the insert and the amino acid sequence from the first methionine residue. Putative polyadenylation signals are underlined and possible locations for introns in the 3' untranslated region are shown in bold characters.

<sup>1</sup> J.G.M. is recipient of a fellowship from Plan Nacional. The work has been carried out under a grant from Plan Nacional de Investigación Científica y Técnica (BIO90–0885).

<sup>2</sup> Abbreviation: HRGP, hydroxyproline-rich glycoproteins.

**Table I.** Characteristics of an Extensin cDNA from *P. amygdalus*


---

**Organism:**  
*Prunus amygdalus* (cv Texas), Rosaceae.

**Locus:**  
Unknown.

**Function:**  
Encodes a protein probably involved in cell wall structure.

**Clone type:**  
cDNA.

**Source:**  
cDNA library constructed from poly(A)<sup>+</sup> mRNA isolated from 45-d-old root.

**Identification:**  
cDNA library in λ-ZAP (Stratagene) screened with a tobacco extensin cDNA probe (4).

**Sequencing strategy:**  
Restriction fragment subcloning; double-stranded plasmid sequencing in pBluescript of both strands using automatic sequencing with fluorescent-labeled primers (A.L.F., Pharmacia).

**Homologies:**  
57% similarity with carrot extensin protein sequence.

**cDNA structure:**  
1146 nucleotides (nt), 834 nt open reading frame and 303 nt 3'-trailer plus poly(A)<sup>+</sup> tail. Putative polyadenylation sites and consensus splicing sequences (AGGT) in the 3' untranslated region.

**Deduced protein structure:**  
278 aa (M, 31,000) including 31 aa signal peptide; peptides SPPPP, PYHYK, and SPSPPKH repeated along the sequence.

**EMBL accession number:**  
X65718 PAEXTS.

---

**ACKNOWLEDGMENTS**

The authors are indebted to Anna Pons and Pedro Mañá for their help in DNA sequencing.

**LITERATURE CITED**

1. **Arumuganathan K, Earle ED** (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* **9**: 208–218
2. **Cassab GI, Varner JE** (1989) Cell wall proteins. *Annu Rev Plant Physiol* **39**: 321–353
3. **Ludevid MD, Ruiz-Avila L, Vallés MP, Stiefel V, Torrent M, Torné JM, Puigdomènech P** (1990) Expression of genes for cell-wall proteins in dividing and wounded tissues of *Zea mays* L. *Planta* **180**: 524–529
4. **Memelink J, Hoge JHC, Schilperoort RA** (1987) Cytokinin stress changes the developmental regulation of several defense-related genes in tobacco. *EMBO J* **6**: 3579–3583
5. **Raz R, José M, Moya A, Martínez-Izquierdo JA, Puigdomènech P** (1992) Different mechanisms generating sequence variability are revealed in distinct regions of the hydroxyproline-rich glycoprotein gene from maize and related species. *Mol Gen Genet* **233**: 252–259
6. **Showalter AM, Bell JN, Cramer CL, Bailey JA, Varner JE, Lamb CJ** (1985) Accumulation of hydroxyproline-rich glycoprotein mRNAs in response to fungal elicitor and infection. *Proc Natl Acad Sci USA* **82**: 6551–6555
7. **Smith JJ, Muldoon EP, Willard JJ, Lampion DTA** (1986) Tomato extensin precursors P1 and P2 are highly periodic structures. *Phytochemistry* **25**: 1021–1030
8. **Stiefel V, Ruiz-Avila L, Raz R, Vallés MP, Gómez J, Pagès M, Martínez-Izquierdo JA, Ludevid MD, Langdale JA, Nelson T, Puigdomènech P** (1990) Expression of maize cell wall hydroxyproline-rich glycoprotein gene in early leaf and root vascular differentiation. *Plant Cell* **2**: 785–793