Plant Molecular Biology **27**: 205–210, 1995.
© 1995 *Kluwer Academic Publishers. Printed in Belgium.*

Short communication

Molecular characterization of cDNAs corresponding to genes expressed during almond (*Prunus amygdalus* Batsch) seed development

Jordi Garcia-Mas [1], Ramon Messeguer [2], Pere Arús [2] and Pere Puigdomènech [1,*]
Unitat Mixta IRTA-CSIC: [1] Departament de Genètica Molecular, CID-CSIC, Jordi Girona 18–26, 08034 Barcelona, Spain; [2] Departament de Genètica Vegetal, IRTA, 08348 Cabrils, Barcelona, Spain (author for correspondence)*

Received 12 April 1994; accepted in revised form 18 October 1994

Key words: almond, cDNA sequence, prunin, oleosin

Abstract

A number of different cDNA clones corresponding to the most abundant mRNAs present in immature seeds have been isolated from an almond (*Prunus amygdalus* cv. Texas) immature seed cDNA library. Those corresponding to proteins involved in storage processes have been further characterized. Two of these cDNAs (PA3BF1 and PA3BE12) code for the almond globulins (prunins), the main family of storage proteins synthesized in seeds during embryogenesis, and another cDNA (PA3BA1) codes for the 15.7 kDa almond oleosin, a protein located on the surface of oil bodies in plant seeds. These cDNAs have been sequenced and their expression during almond fruit development has been studied. Their expression is seed-specific and localized in cotyledons around 100 days after flowering. Both prunin and oleosin genes are present in one or two copies in the almond genome.

Little information is available about processes involved in seed maturation in the Rosaceae family, which contains many fruit crop species. Species from the genus *Prunus* are specially attractive to follow molecular and genetic approaches due to the fact that they have the smallest genomes known among crop species [1]. *Prunus* species are, with those of genus *Malus*, the fruit crops with the highest economic interest. It is therefore interesting to characterize genes expressed in organs such as almond seed that have an economic importance and that have not been studied from a molecular point of view. Biochemical studies in different almond cultivars have reported 20% and 60% of protein and lipid dry weight, respectively, in the mature seed, the carbohydrate fraction being much smaller [10]. Cotyledons developed from the embryo tissue are physically present from around 80–90 days after flowering (DAF) until the maturation of the seed [3] and it is reasonable to suppose that genes coding for seed storage components are those expressed at the highest level during that period.

In dicotyledonous plants, salt-soluble globulins

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers X78118 (ole1), X78119 (pru1) and X78120 (pru2).

are the major family of storage proteins [2]. The most common type of globulins in dicots is the 11–12S globulin which is an oligomeric protein composed of six 50–60 kDa subunits. Each subunit is composed of two polypeptides, linked by a disulfide bond (30–40 kDa α-chain and 20 kDa β-chain); these are processed from a 50–60 kDa precursor [11]. The immature polypeptides contain a signal peptide in the N-terminus which is cleaved during the deposition of the protein in the storage compartment in the cell [8].

Lipid storage in higher plant seeds occurs in the form of triacylglycerols in oil bodies. The surface of these organelles basically consists of a phospholipid monolayer and a group of associated proteins, oleosin being the most abundant of them. Oleosins have been proposed to have a structural role [12] but they also might be involved in the binding of lipase during germination [4]. Oleosins contain in their sequence an amphipathic N-terminal domain followed by a conserved long hydrophobic domain (around 70 amino acids) and an amphipathic α-helical domain in the C-terminus [4].

Here we report the construction of an almond 110 DAF immature seed cDNA library and the cloning of cDNAs corresponding to genes abundantly expressed at this period after screening the library with radioactively labelled cDNA from the same tissue. The accumulation of the corresponding mRNAs during seed development, gene copy number and sequence homologies with related species are described. Immature almond seeds 110 DAF show a well-developed cotyledon tissue. RNA extracted from these immature seeds was used to construct a cDNA library. 384 random clones were screened using first-strand cDNA prepared from the same tissue as a probe. Those clones showing a higher hybridization signal were sequenced in their 5′ end. In particular, clones PA3BF1 and PA3BE12 were identified in the EMBL Database as storage protein cDNAs belonging to the globulin family and clone PA3BA1 had high similarity to plant oleosin reported sequences. The cDNA library was rescreened with those identified cDNAs; 21.6% of clones corresponded to cDNAs with similarity with globulins and 2% hybridized with oleosin.

The cDNA insert in clone PA3BF1 was 1905 nucleotides long and encoded a protein formed by 551 amino acids. It contained a 20 residue long peptide in the N-terminus with the features of a signal peptide. The consensus pattern for cleavage of the protein into two subunits [13] was also identified (Fig. 1). It also showed two conserved cysteine residues which are thought to be involved in disulfide bonds between the two subunits [2] and another two cysteines probably involved in intra-subunit disulfide bonds. A remarkable feature of this protein is the existence of a domain rich in glutamine residues (from Gln-96 to Leu-167) which is not conserved when comparing this sequence to that of other plant globulins. Upon searching the Databases the similarity to other globulin protein sequences appeared, the proteins of the legumin family being the most similar ones. In accordance to the name given to other storage protein sequences, these proteins could be termed as prunins, the protein encoded by clone PA3BF1 prunin Pru1 and the one encoded by clone PA3BE12 prunin Pru2.

The cDNA insert in clone PA3BE12 (prunin Pru2) was 1787 nucleotides long and encoded a polypeptide containing 504 residues (Fig. 1). This clone was not full-length and it showed an incomplete signal peptide but a complete mature protein. The prunin sequences (Fig. 1) show a 63% identity in their mature sequence. The region with less similarity is located in the glutamine-rich region of prunin Pru1, and corresponds to the region with the most variable region in all the reported plant legumin-like sequences.

The proteins coded by prunin Pru1 and Pru2 cDNAs correspond to mature polypeptides of predicted molecular weights of 61.0 and 55.9 kDa respectively that upon processing in the consensus cleavage site would give rise to two polypeptide subunits. Prunin Pru1 would be cleaved in an acid α subunit of 40.1 kDa and pI 5.4 and a basic β one of 20.9 kDa and pI 9.6. The predicted subunits of prunin Pru2 have 34.5 kDa and pI 4.6 for the α subunit and 21.4 kD and pI 9.5 for the β subunit. The pattern of proteins accumulated in

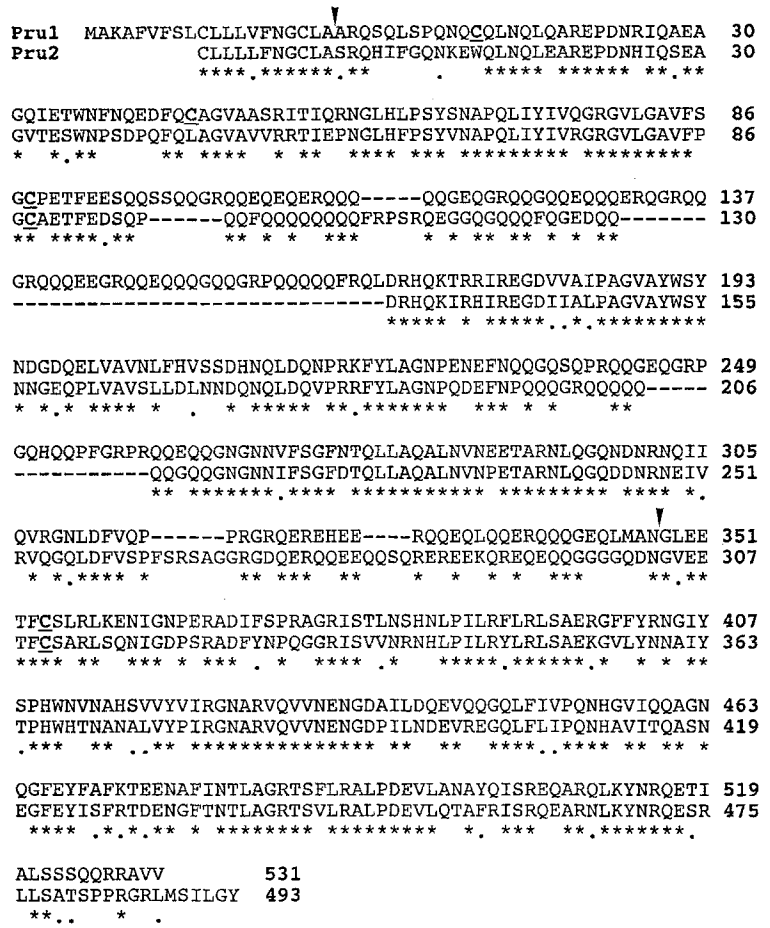

Fig. 1. Alignment of the prunin Pru1 and prunin Pru2 protein sequences deduced from the cDNA sequence. Arrows indicate the predicted cleavage sites of both the signal peptide and the two α and β subunits. Cysteine residues are underlined. Sequence numbers are counted from the start of the mature protein. The amino acid residues that are completely conserved in the two sequences are indicated with an asterisc and those amino acids with similarity (according to the PCGENE software) are shown with a dot.

mature almond seed approximately corresponds to these figures (Fig. 3A). In the protein analysis two main groups of proteins at 41.8 and 38.6 kDa and 22.8 and 20.4 kDa may be observed. This pattern is similar to that found in other species [5]. These two groups of proteins may correspond to the α and β subunits of prunins as the measured molecular weight values fit with those predicted for the polypeptides processed from the prunin sequences here described.

The cDNA library was screened with a radish napin probe [9] in order to identify storage proteins corresponding to the albumin fraction. The results were negative suggesting that in almond the legumin-like prunins are the most abundant components, in accordance with the results obtained by the protein extraction. Another possible explanation would be that the napin probe is not similar enough to the almond albumin cDNAs to form stable hybrids.

In the same cDNA library another highly expressed clone was identified which encoded a protein having high similarity to the reported plant oleosin sequences. It was 789 nucleotides long and encoded a peptide having 149 residues with a predicted molecular mass weight of 15.7 kDa When aligned to the sequences of the other reported oleosins the highest similarity was located on the hydrophobic region while the N and C-termini were less conserved (Fig. 2). It has been

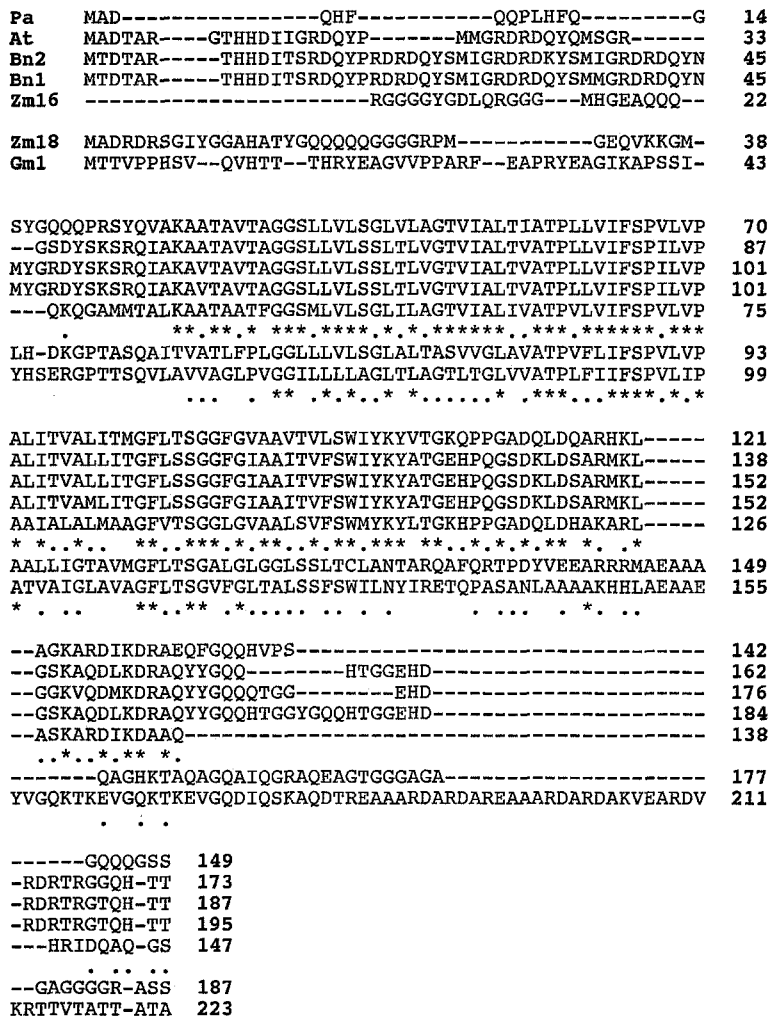

Fig. 2. Alignment of the almond oleosin protein sequence deduced from the cDNA sequence with homologous proteins from different plant species. The upper group includes low-molecular-mass oleosins Pa (*Prunus amygdalus*), At (*Arabidopsis thaliana*), Bn2 and Bn1 (*Brassica napus*) and Zm 16 kDa (*Zea mays*). The two lower sequences Zm 18 kDa (*Zea mays*) and Gm1 (*Glycine max*) belong to the high-molecular-mass oleosin group. The completely conserved amino acid residues are indicated with an asterisc and those amino acids with similarity (according to the PCGENE software) are shown with a dot.

described that some species have two oleosin genes which encode different isoforms [4]. From the alignments made with the almond oleosin (Fig. 2) it can be observed that there is a group of plant oleosins which share more identity with almond oleosin (*Arabidosis thaliana* (62.2%), *Brassica napus* 2 (62.2%), *B. napus* 1 (62.2%) and *Zea mays* 16 kDa (59.2%)) and another group with less identity (*Z. mays* 18 kDa (37.2%) and *Glycine max* 1 (35.8%)). This result confirms the existence of at least two families of oleosin genes in the plant kingdom, the high-molecular-mass and the low-molecular-mass isoforms.

P. amygdalus (cv. Texas) genomic DNA was cut with *Eco* RI, *Hind* III and *Bam* HI, gel-fractionated and blotted on nylon filters as previously described [6]. The filters were hybridized with different cDNA probes. A 1.6 kb *Eco* RI-*Xho* I conserved probe from the prunin Pru1 clone PA3BF1 coding region was hybridized with the DNA blot filters and gave one main band in each digestion and other faint bands (Fig. 3B, I). When

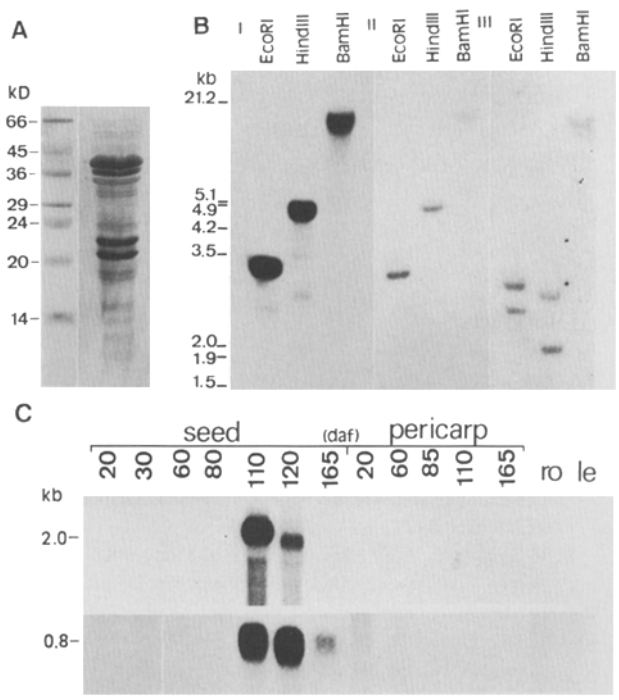

Fig. 3. A. Gel electrophoresis of total proteins extracted from mature almond seed. The protein was extracted in the presence of 4% SDS and run on a 15% acrylamide gel. Major bands appear around 40 and 20 kDa which fit with the deduced size of prunin acid and basic mature peptides. Molecular weight marker is Dalton Mark VII-L (Sigma, St. Louis, MO). B. DNA blot analysis for *Prunus amygdalus* DNA using prunin and oleosin probes. I. The filter was hybridized with a probe corresponding to the 5′ coding region of prunin Pru1 cDNA. II. The filter was hybridized with a probe corresponding to the 3′ non-coding region of prunin Pru1 cDNA. III. The filter was hybridized with a probe corresponding to the 3′ non-coding region of prunin Pru2 cDNA. Each lane contains 5 μg of genomic almond leaf DNA restricted with *Eco* RI, *Hin*d III and *Bam* HI. Molecular weight marker is λ DNA restricted with *Eco* RI and *Hin*d III. C. mRNA accumulation of prunin and oleosin in different developing almond seed tissues. Each lane contains 10 μg of total RNA from different tissues: seed, pericarp, root (Ro) from 30-day germinating plantlet and young leaf (Le). Numbers indicate DAF. Upper panel: RNA blot analysis of *P. amygdalus* prunin Pru1 using the 5′ coding region as a probe. Lower panel: RNA blot analysis of *P. amygdalus* oleosin using the PA3BA1 cDNA clone as a probe.

a 300 bp *Xho* I-*Xho* I probe from the 3′ non-coding region from clone PA3BF1 was used, the same main bands were observed (Fig. 3B, II), but the faint bands disappeared. Even if the filter was overexposed it was not possible to observe other bands (not shown). The faint bands appeared again when a polymerase chain reaction (PCR) amplified 3′ non-coding region probe from prunin Pru2 clone PA3BE12 was used (Fig. 3B, III). These results indicate that the almond genome contains only two genes each one coding for one of the prunin proteins (Pru1 and Pru2) and corresponding to the cDNA clones PA3BF1 and PA3BE12. This is in contrast with the results obtained in other plant species such as soybean, where at least five genes coding for globulins have been described [7], but we should consider the possibility that the Pru clones do not hybridize with other legumin-like genes present in almond. Almond appears to be a simple system in which storage proteins are mainly legumin-like and consist of two main pairs of processed polypeptides encoded each one by a single gene. However,it cannot be excluded that genes coding for other storage protein fractions expressed at other periods of seed development or at lower levels of expression than prunins are present in the almond seed. When the oleosin cDNA was used as a probe in DNA blot experiments a pattern consistent with the existence of one or two genes coding for the almond oleosin was observed (not shown).

10 μg of total RNA extracted from different tissues and different times of fruit and seed development were gel-fractionated and blotted onto nylon filters as previously described [6]. When hybridization was performed with a 5′ probe corresponding to the coding region from prunin Pru1 (Fig. 3C, upper panel) a unique 2 kb message was displayed in the RNA sample corresponding to 110 DAF seed. Due to the fact that almond seed development requires about six months, it is difficult to collect RNA samples at the precise periods that would allow to distinguish the induction of the prunin and oleosin mRNAs. However, it is clear that the expression of the gene coding for prunin Pru1 is tightly regulated and dependent on the organ (it is not expressed in pericarp, root and leaves) and on the developmental stage. When using the 3′ probe corresponding to the untranslated region from prunin Pru2 its accu-

mulation pattern was identical to the one observed for prunin Pru1 (not shown), indicating that both prunin genes are abundantly expressed and specifically in the same stage of cotyledon development.

The pattern of accumulation of oleosin mRNA when the entire cDNA was used as a probe is similar to that shown for prunin cDNAs with a maximum around 110 DAF and a still detectable expression in late developmental stages around 160 DAF (Fig. 3C, lower panel). This result indicates that oleosin expression is seed specific and mainly concentrated during cotyledon maturation; the induction of the oleosin mRNA expression appears to occur between 80 and 110 DAF, when the cotyledons develop from the embryo.

Two groups of proteins present in the almond seed have been characterized at cDNA level and the sequences of the proteins have been obtained. One of these groups is what appears to be the main storage protein fraction in *Prunus* named prunin. It consists of two main pairs of polypeptides of the legumin-like type of globulins. These proteins are each encoded by single genes. A cDNA belonging to the low-molecular-mass family of oleosins has also been described. These oleosins appear to be encoded in almond by a small gene family. Both oleosin and prunin mRNAs are specifically accumulated in the developing cotyledon. It appears that the complexity of the genes encoding these storage proteins is low in almond. This fact could be in accordance with the low complexity of the *Prunus* genome.

Acknowledgements

The authors acknowledge the invaluable help of Prof. Joseph E. Varner (Dept. of Biology, Washington University, St. Louis) for critically reading the manuscript and Dr Francisco Vargas and Dr Ignasi Batlle (IRTA, Mas Bové, Reus) for their help in collecting the biological material. J. G. M. is recipient of a fellowship from Plan Nacional de Formación de Personal Investigador (Ministerio de Educación y Ciencia). The work has been carried out under grant Bio93-0901 (Plan Nacional de Investigación Científica y Técnica).

References

1. Arumuganathan K, Earle ED: Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9: 208–218 (1991).
2. Borroto K, Dure L: The globulin seed storage proteins of flowering plants are derived from two ancestral genes. Plant Mol Biol 8: 113–131 (1987).
3. Grigorian V: L'embryogenèse chez l'amandier (*Prunus amygdalus* Batsch): étude comparée de la dormance des graines et de la dormance des bourgeons végétatifs. Ph.D. thesis, University of Bordeaux I, Bordeaux (1972).
4. Huang AHC: Oil bodies and oleosins in seeds. Annu Rev Plant Physiol Plant Mol Biol 43: 177–200 (1992).
5. Luthe DS: Electrophoretic analysis of seed proteins in the Dicotyledoneae. Plant Mol Biol Rep 10: 254–262 (1992).
6. Montoliu L, Rigau J and Puigdomènech P: A tandem of alpha-tubulin genes preferentially expressed in radicular tissues from *Zea mays*. Plant Mol Biol 14: 1–15 (1989).
7. Nielsen NC, Dickinson CD, Cho TJ, Thanh VH, Scallon BJ, Fisher RL, Sims TL, Drews GN, Goldberg RB: Characterization of the glycinin gene family in soybean. Plant Cell 1: 313–328 (1989).
8. Pang PP, Pruitt RE, Meyerowitz EM: Molecular cloning, genomic organization, expression and evolution of 12S seed storage genes of *Arabidopsis thaliana*. Plant Mol Biol 11: 805–820 (1988).
9. Raynal M, Depigny D, Grellet F, Delseny M: Characterization and evolution of napin-encoding genes in radish and related crucifers. Gene 99: 77–86 (1991).
10. Saura F, Cañellas J, Soler L: La almendra. Composición, variedades, desarrollo y maduración. Instituto Nacional de Investigaciones Agrarias, Madrid (1988).
11. Shotwell MA, Larkins BA: The biochemistry and molecular biology of seed storage proteins. In: Stumpf PK, Conn EE (eds) The Biochemistry of Plants, vol. 15, pp. 297–345. Academic Press, New York (1989).
12. Vance VB, Huang AHC: The major protein from lipid bodies of maize. J Biol Chem 262: 11275–11279 (1987).
13. Vonder Haar RA, Allen RD, Cohen EA, Nessler CL, Thomas TL: Organization of the sunflower 11S storage protein gene family. Gene 74: 433–443 (1988).