

# Characterization and expression analysis of the aspartic protease gene family of *Cynara cardunculus* L.

Catarina Pimentel<sup>1,2,3</sup>, Dominique Van Der Straeten<sup>3</sup>, Euclides Pires<sup>1,4</sup>, Carlos Faro<sup>1,4</sup> and Claudina Rodrigues-Pousada<sup>2</sup>

1 Departamento de Biologia Molecular e Biotecnologia do Centro de Neurociências de Coimbra, Universidade de Coimbra, Portugal

2 Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

3 Unit Plant Hormone Signalling and Bio-imaging, Ghent University, Belgium

4 Departamento de Bioquímica, Faculdade de Ciências e Tecnologia, Universidade de Coimbra, Portugal

## Keywords

aspartic proteases; cardosin; leader intron; pistil; promoter

## Correspondence

C. Rodrigues-Pousada, Instituto de Tecnologia Química e Biológica, Apt. 127, 2781-901 Oeiras, Portugal

Fax: +351 214433644

Tel: +351 214469624

E-mail: claudina@itqb.unl.pt

C. Faro, Departamento de Bioquímica, Faculdade de Ciências e Tecnologia, Universidade de Coimbra, Apt. 3126, 3000 Coimbra, Portugal

Fax: +351 230480208

Tel: +351 239480210

E-mail: cfaro@imagem.ibili.uc.pt

## Database

The nucleotide sequences of *Cynara cardunculus* L. aspartic protease genes have been submitted to the EBI Data Bank under the accession numbers AM286227 (cardosin B) and AM286279 (cyprosin B)

(Received 26 December 2006, revised 13 February 2007, accepted 13 March 2007)

doi:10.1111/j.1742-4658.2007.05787.x

Aspartic proteases (APs) are widely distributed in nature, from simple organisms like the unicellular green algae *Chlamydomonas reinhardtii* and the moss *Physcomitrella patens* [1], to the more complex gymnosperm

and angiosperm plants [2]. In contrast to those of their animal counterparts, the biological functions of plant APs are far from being deciphered. Nevertheless, plant APs have been implicated in a plethora of biological

## Abbreviations

ACS, 1-aminocyclopropane-1-carboxylic acid synthase gene; AP, aspartic protease; GUS,  $\beta$ -glucuronidase; IME, intron mediated enhancement; PR, pathogenesis-related protein; PSI, plant specific insert; SLG, S-locus glycoprotein gene; SLR, S-locus related gene; UTR, untranslated region.

functions, including the degradation and/or proteolytic processing that occur during plant senescence, biotic and abiotic stress responses, programmed cell death, and reproduction [2].

Cardosin A and cardosin B are two floral APs, purified from *Cynara cardunculus* L. pistils, that have been broadly studied and characterized [3–9]. To our knowledge, cardosins A and B represent the best characterized floral APs, together with cyprosins [10,11], two other APs present in the pistils of *C. cardunculus* L. Strikingly, cardosins and cyprosins have never been copurified, and their coexistence in the plant remains elusive.

Like many other plant APs, cardosins are synthesized as inactive zymogens and undergo proteolytic processing, leading to the activation of the enzyme [3,5,9]. Cardosins A and B exhibit distinct enzymatic properties [8], and diverge in terms of tissue localization [3,9]. Cardosin A was mainly found in the protein storage vacuoles of the stigmatic papillae [6], whereas cardosin B accumulates in the extracellular matrix of the floral transmitting tissue [9]. Given that both enzymes share a highly similar primary structure (73%), their distinct biochemical behaviors could be due to the slight differences observed between them [9]. Although the biological functions of cardosins in the flowers of *C. cardunculus* are not completely assigned, their pistil-specific detection in all stages of flower development [6,9] has suggested that they may participate in several flower-specific events, such as flower senescence, defensive mechanisms against insects and/or pathogens, and reproduction [3,9].

Despite the large amount of information gathered in the last decade on plant APs, little is known about AP gene regulation. Indeed, all the data so far available on AP gene expression regulation have been obtained essentially from studies on proteases whose genes are induced upon several environmental stimuli [12–15] or specifically expressed in particular stages of the plant life cycle [16–21].

In this study, the genomic sequences of the cardosin A and B genes and of two new cardosin genes (those encoding cardosins C and D) were isolated and characterized. Our results showed that in cardoon as well as in transgenic *Arabidopsis* plants, cardosin genes exhibit a differential pattern of expression. To gain further understanding of the mechanisms that dictate the flower-specific expression pattern of cardosins, several 5'-deletions of the cardosin A gene promoter region were fused to the  $\beta$ -glucuronidase (GUS) reporter gene and introduced into *Arabidopsis thaliana* plants. This allowed us to delimit a region of 529 bp crucial for cardosin A expression. We also evaluated the relevance

of the leader intron of the cardosin A and B genes on gene expression in *A. thaliana*. Furthermore, the significance of several putative *cis* elements found within the identified regulatory regions of the genes is discussed. Finally, an evolutionary relationship based on sequence comparison of these proteases is presented.

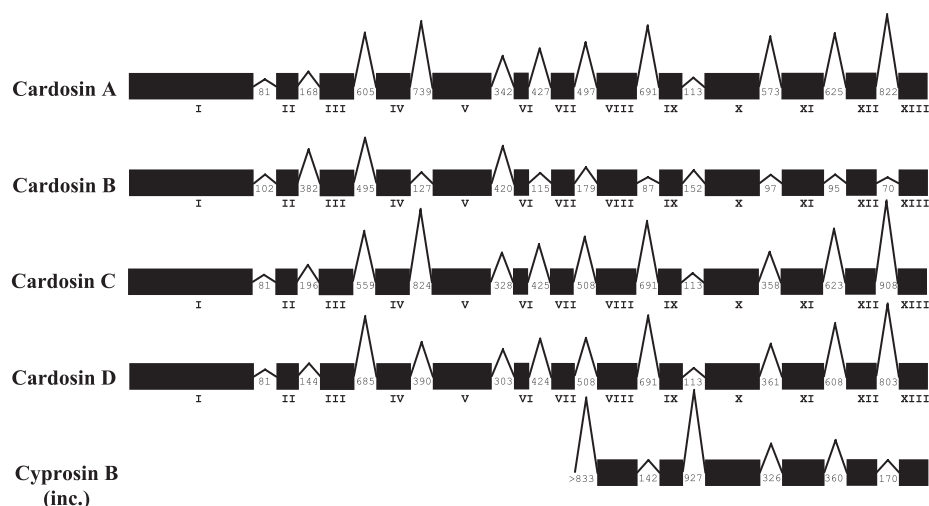
## Results

### Isolation and characterization of cardosin genes

The previously cloned cardosin A full-length cDNA [3] was used to screen a genomic library of *C. cardunculus*. Three phages –  $\lambda$ 5,  $\lambda$ 6, and  $\lambda$ 18 – were isolated and subjected to restriction analysis and subcloning. Phage  $\lambda$ 5 harbored the cardosin A gene, and the remaining phages contained two new cardosin genes, designated cardosins C ( $\lambda$ 6) and D ( $\lambda$ 18). An additional screen with a probe comprising a fragment of the cardosin B gene, including its 3'-UTR, yielded two positive phages,  $\lambda$ 4.1 and  $\lambda$ 4.2. The former harbored the complete sequence of the cardosin B gene, whereas the latter enclosed a partial sequence of the cyprosins gene. Like other plant AP genes, cardosin genes have their coding region interrupted by 12 introns that occur in conserved positions despite their variable sizes (Fig. 1). Both the 5'- and 3'-splice junctions are in good agreement with the exon–intron consensus boundary sequences [22], and the initiation codon is inserted in a well-conserved context (AACATGGG) among plant genes [23].

Comparison of cardosin A, B and D genomic clones with the respective cDNAs (Fig. 2) revealed the presence of an intron in the 5'-UTR of the genes. The nucleotide sequences of the cDNA and genomic clones of cardosins diverge after a perfect match of six bases. At the point of divergence, a consensus splicing acceptor sequence, 5'-AG-3', was found (Fig. 3). The remaining bases of the leader sequence appear in the upstream region of the genomic clone after an intervening sequence of 966 bp (cardosin A), 953 bp (cardosin B) or 1207 bp (cardosin D), with a consensus donor site 5'-GT-3' at the 5'-end, suggesting that this region represents an intron (Fig. 2). To map the transcription initiation site of cardosin genes, primer extension analysis with an antisense oligonucleotide located in the untranslated region determined by 5'-RACE was carried out (data not shown). The 5'-end of cardosin genes identified by primer extension analysis was longer than the one observed by 5'-RACE (Fig. 2).

Although a leader intron seems to be a conserved structural feature among AP genes [16,19], it does not appear in the 5'-UTR of the cardosin C gene. This observation is based on the comparison of the genomic



**Fig. 1.** Schematic representation of the structure of the cardosin and cyprosin B genes. Filled boxes represent exons. Open triangles symbolize introns. The size of each intron is indicated under the triangles in bp. The sequence of cyprosin B isolated was incomplete and encompassed the last six exons of the gene.



**Fig. 2.** Determination of the transcription initiation site of the cardosin A, B and D genes. The alignment of the most extended 5'-RACE products (CardA\_cDNA, CardB\_cDNA, and CardD\_cDNA) against the corresponding genomic sequences (CardA\_gDNA, CardB\_gDNA, and CardC\_gDNA) revealed the presence of an intron within the 5'-UTR of the genes. Primer extension analysis showed that the precise transcription initiation site is located several nucleotides upstream of each gene's longest 5'-RACE product, at the nucleotide indicated by an open arrow. The initiation codon is shaded in black. The leader intron consensus splicing donor and acceptor sequences are boxed. The size of the intron is indicated in bp.

sequences of the cardosin A, C and D gene 5'-flanking regions (Fig. 3). Beyond an initial small match of nucleotides immediately upstream from the initiation codon, the homology among the three genes is interrupted, but it is recovered several nucleotides upstream from the 5'-UTR of the cardosin A and D genes (Fig. 3).

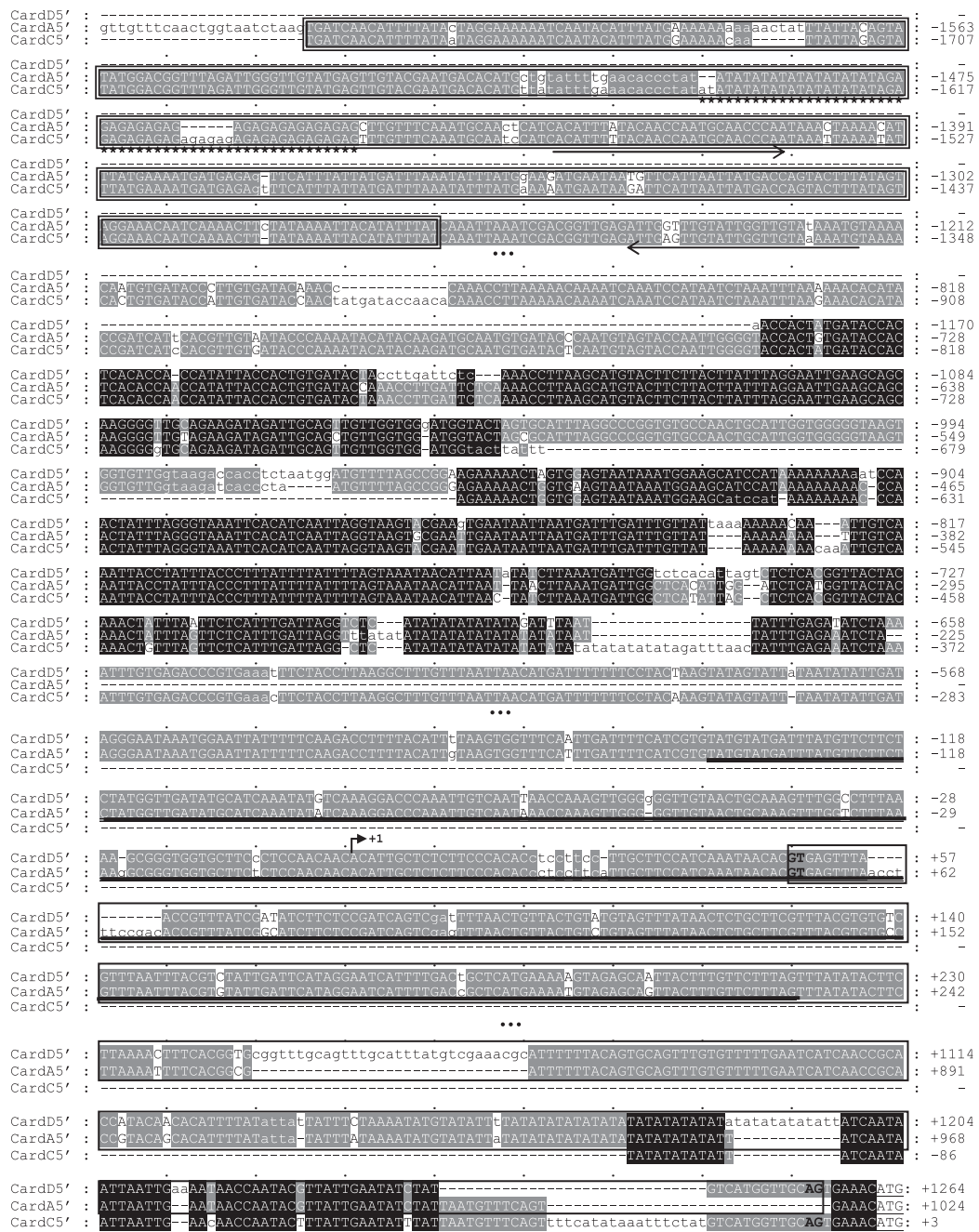
A TATA element [24], TATAAAA, is located 30 bp upstream of the transcription start site of the cardosin B gene, and two 'CAAT' box motifs are found at positions -43 bp and -82 bp. The putative 'TATA' boxes of the cardosins A and D genes (TTTAAAA), located -25 bp upstream of the transcription start site, differ from the consensus sequence found in plant genes (TATAWAWA) [24]. Identical sequences were, however, identified in the rat tropomyosin gene [25] and in the *A. thaliana* phenylalanine ammonia-lyase gene (GenBank accession number X84728). 'CAAT'

motifs are present in positions -71 bp and -74 bp of the cardosin A and D genes, respectively.

The 5'-flanking regions of the cardosin A, C and D genes share a high degree of similarity (Fig. 3). However, the respective region of the cardosin B gene only exhibits a stretch of 388 bp with significant homology to the cardosin A and D genes (Fig. 3).

### Predicted structural features of the new cardoon APs

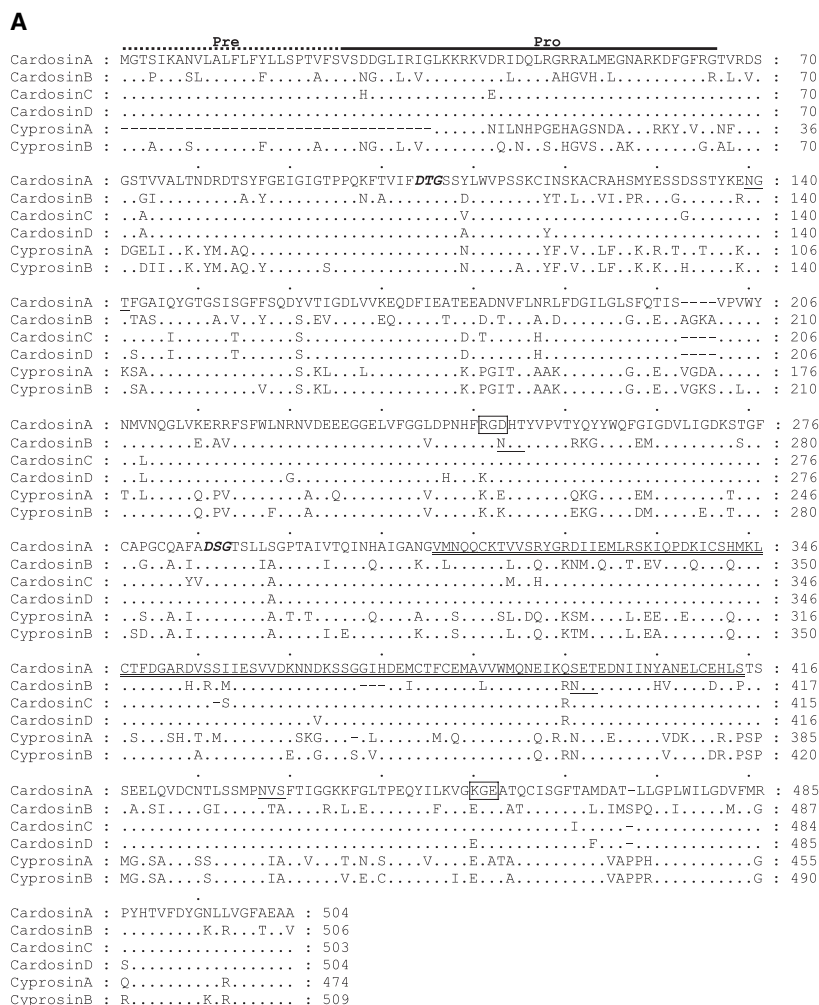
As expected, the deduced amino acid sequences of cardosin C and cardosin D revealed that both enzymes possess the typical structural domain organization of plant APs [2]. Cardosins and cyprosin B share, in terms of primary structure, a high level of similarity, with cardosins A, C and D exhibiting the highest scores. Interestingly, the slight differences



**Fig. 3.** Alignment of the 5'-flanking regions of the cardosin A, C and D genes. Sequences sharing 100% similarity among the genes are shaded in black. The sequences that are 100% identical between two of the genes are in gray. The leader introns of cardosins A and D are boxed. The initiation codon is underlined. The A/T and G/A repeats are indicated by asterisks. The inverted repeat is indicated by arrows. Horizontal lines indicate the absence of a nucleotide in the sequence. Lower-case letters represent unique sequences. The initiation of transcription of the cardosin A and D genes is indicated by a bent arrow. The three dots represent omitted parts of the alignment. The cardosin A sequence that is underlined (from -139 bp to +232 bp) is the only region of the cardosin A 5'-flanking region that shares significant similarity with the corresponding region of the cardosin B gene (from -147 bp to +238 bp). The 529 bp of the promoter region of the cardosin A gene that is relevant for gene expression in *Arabidopsis* and the corresponding region of the cardosin C gene are double boxed.

among cardosins A, C and D comprise the RGD and KGE motifs, which were demonstrated to be important for the interaction of cardosin A with

phospholipase D $\alpha$  [7]. As depicted in Fig. 4, the RGD/KGE motifs found in the primary structure of cardosins A and C are replaced in cardosin D by



**Fig. 4.** Amino acid sequence alignment and homology of cardosins A, B, C and D and cyprosin A and B. (A) The amino acid sequences were deduced from the genomic sequences (this work), with the exception of cyprosin A (X69193) and cyprosin B (X81984). Identical sequences are indicated by dots, and deleted amino acids by horizontal lines. The signal peptide and prosequence are indicated by dashed and continuous lines, respectively. The amino acids forming the catalytic triads in the active site (DTG and DSG) are in bold italic. The RGD and KGE motifs are boxed. Potential N-linked glycosylation sites are marked. (B) Percentage amino acid identity and similarity between *C. cardunculus* APs. The upper and lower parts of the table correspond to similarity and identity percentages, respectively.

**B**

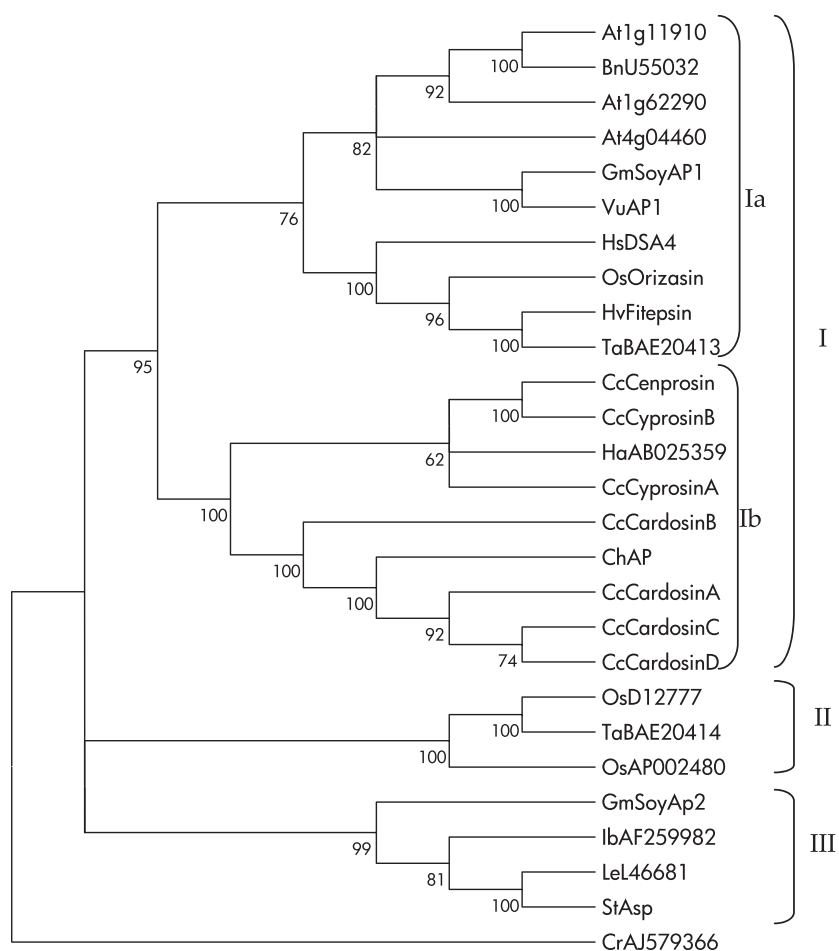
	Cardosin A	Cardosin B	Cardosin C	Cardosin D	Cyprosin A	Cyprosin B
Cardosin A	76	80	97	97	74	78
Cardosin B	76	80	80	79	75	80
Cardosin C	96	75	97	97	74	78
Cardosin D	96	75	96	97	74	78
Cyprosin A	70	77	69	70	86	86
Cyprosin B	75	77	74	75	82	86

KGD/EGE motifs. These differences may have relevant functional implications, as cardosin B harbors a RGN/EGE motif and does not interact with phospholipase D $\alpha$  [7].

**Evolutionary relationships of cardosins and their plant counterparts**

The amino acid sequences of *C. cardunculus* APs were compared with those of several other plant APs, by means of the phylogenetic analysis program MEGA version 3.0 [26], using the neighbor-joining method. On the basis of the resulting phylogenetic tree, three distinct

groups within the typical plant AP family can be defined (Fig. 5). Group I comprises the best studied APs, and may further be divided into two smaller groups. Group Ia includes the APs of the Brassicaceae and Fabaceae families, as well as those found in monocotyledonous plants. These APs have been implicated in the proteolytic processing and/or degradation of storage proteins (*A. thaliana* and *Brassica napus* APs, orizasin and fitepsin), in leaf senescence (At4g0446, BnU55032, and VuAP1), and in programmed cell death events (SoyAP1 and fitepsin) [18,19,21,27–32]. Although the wheat AP (BAE20413) has not yet been biochemically or molecularly characterized, its inclusion



**Fig. 5.** Phylogenetic relationship between several APs. The phylogenetic analysis was carried out by the neighbor-joining method using MEGA version 3.0. One thousand bootstrap replicates were calculated, and bootstrap values are shown at each node. Nodes were collapsed to a single horizontal line whenever statistical support was less than 60%. On the basis of the AP family tree, it is possible to divide typical plant APs into three groups (I, II, and III). Group I may be subdivided into two smaller groups: Ia and Ib. The two first letters of the sequence name indicate the plant species. At, *A. thaliana*; Bn, *B. napus*; Cca, *Ceutarea calcitrapa*; Cc, *C. cardunculus*; Ch, *Cy. humilis*; Cr, *Ch. reiinhardtii*; Gm, *Glycine max* (soy); Ha, *Helianthus annuus* (sunflower); Hs, *Hemerocallis* sp. (lily); Hv, *Hordeum vulgare* (barley); Ib, *Ipomoea batatas* (sweet potato); Le, *Lycopersicon esculentum* (tomato); Os, *Oryza sativa* (rice); St, *Solanum tuberosum* (potato); Ta, *Triticum aestivum* (wheat); Vu, *Vigna unguiculata* (cowpea). The following characters indicate the sequence accession number (or the AGI code, in the particular case of *Arabidopsis* APs) or the name of the enzyme: orizasin, D32165; fitepsin, X56136; cenprosin, Y09123; cyprosin A, X69193; cyprosin B, X81984; VuAP1, AF287258; DSA4, AF082029; SoyAP1, AB069959; and SoyAP2, AB070857. Cardosin amino acid sequences were deduced from the genomic sequences (this study).

in this group suggests that it might be involved in similar biological functions. Within group I, the APs At1g11910 and BnU55032, GmSoyAP1 and VuAP1, as well as fitepsin and BAE20413, form a clade and appear to be potential orthologs (Fig. 5).

Group Ib includes the APs from the Asteraceae family, which have mostly been found in flowers and therefore have been proposed to participate in flower-specific events [3,9,33]. The topology of group Ib suggests that, at some time during the evolution of *C. cardunculus*, an AP ancestor gene has duplicated

and given rise to the branches comprising cyprosin and cardosin. Subsequent duplications within both branches should have occurred originating the group actual configuration (Fig. 5).

The APs of group II have never been studied; however, as they are evolutionarily related, it is possible that they share similar or complementary biological functions. Interestingly no dicotyledonous plants were found within this group (Fig. 5).

Finally, group III contains the tomato (L46681) and potato (StAsp) APs, whose genes are induced upon

biotic stress challenge [14,15]. The group also includes one of the soy Aps (SoyAP2), which is expressed in several tissues and may be involved in seed germination [32], and the sweet potato AP (AF259982).

### Cardosin genes exhibit distinct expression patterns in *C. cardunculus*

Given the overall similarity among the cardosin A, C and D genes, it became evident that our previous work did not allow discrimination of these genes [3,6]. Within this context, we had designed primer pairs specific for each cardosin gene (Fig. 6A) and evaluated gene expression by RT-PCR in three stages of pistil development and in several other organs of *C. cardunculus* (Fig. 6B). Our results showed that: (a) with the exception of stems, the cardosin A and D genes share a similar pattern of expression, being ubiquitously expressed; (b) cardosin B gene expression is pistil-specific; and (c) cardosin C expression is flower-specific and restricted to the pollen and to the pistils of partially opened capitula (Fig. 6B).

### Cardosin promoter regions are functional in *A. thaliana*

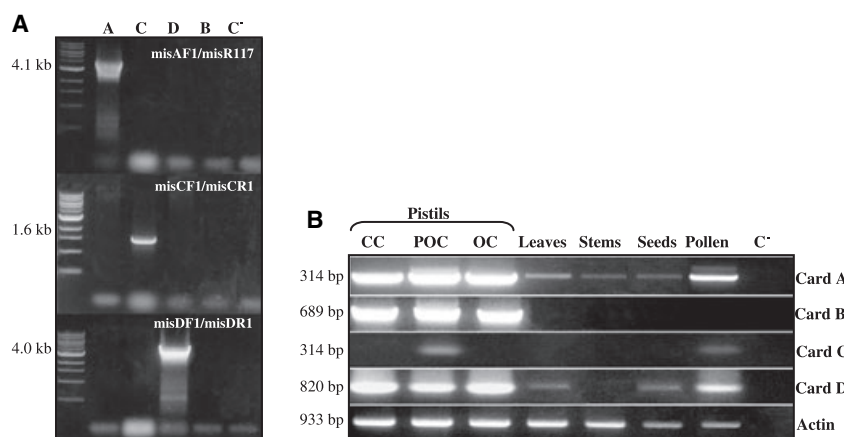
To further investigate the spatial and temporal expression patterns of cardosin genes, each of their 5'-flanking regions (promoter and leader intron) was fused to the GUS reporter gene in order to generate the con-

structs –2912pA::GUS (cardosin A), –3459pB::GUS (cardosin B), –2040pC::GUS (cardosin C), and –1186pD::GUS (cardosin D).

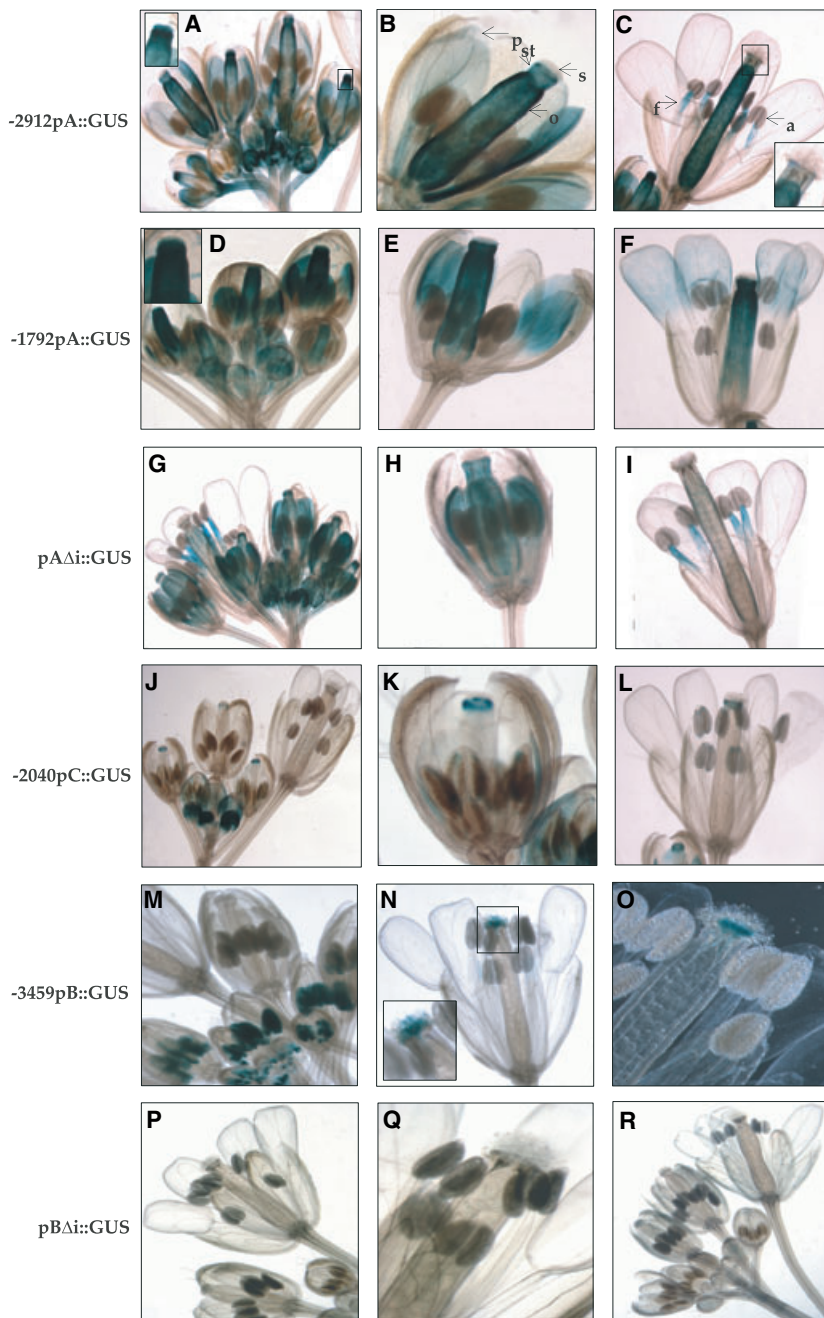
The –2912pA::GUS construct drives GUS expression in the pistils, petals and filaments in the early stages of *A. thaliana* flower development in six of the independent transformed plant lines analyzed (Fig. 7A–C). The expression is mainly restricted to the flowers, although staining can also be observed in young stems. At the initial stages of pistil development, intense staining is observed in the stigma, style and ovary. However, the stigma staining tends to disappear at the later stages of flower development (Fig. 7A–C).

The 5'-flanking region of the cardosin B gene (–3459pB::GUS) induced GUS expression in the anthers, at the initial stages of flower development (Fig. 7M), and in the stigmatic papillae of mature flowers (Fig. 7N,O). Within six independently transformed *Arabidopsis* lines, GUS activity was not detected in other plant organs, being confined to floral tissues.

In seven of eight plant lines transformed with the cardosin C promoter region (–2040pC::GUS), the transgene expression was confined to undifferentiated flowers and styles (Fig. 7J–L), whereas construct –1186pD::GUS (containing the cardosin D 5'-flanking region) was not able to drive GUS expression in the eight independent plant lines analyzed (data not shown). In addition, none of the negative controls showed GUS staining (data not shown).



**Fig. 6.** Expression of cardosin genes during flower development and in several organs of *C. cardunculus* (A) Control analysis of the specificity of the PCR amplification of each cardosin gene. Phage DNA including each cardosin gene – cardosin A (A), cardosin C (C), cardosin D (D), and cardosin B (B) – was used as template in these experiments. The gene-specific primers used were misAF1/misR1117 (cardosin A), misCF1/misCR1 (cardosin C), and misDF1/misDR1 (cardosin D). The primer pairs only amplified the corresponding gene, confirming their specificity. (B) RT-PCR analysis of cardosin genes, using the corresponding gene-specific primer pairs. The actin 2 gene of *A. thaliana* was used as an amplification positive control. CC, pistils of closed capitulum; POC, pistils of partially open capitulum; OF, pistils of open capitulum; C, negative control.



**Fig. 7.** Histochemical analysis of GUS activity in transgenic *A. thaliana* plants transformed with cardosin A, B and D constructs, containing the 5'-flanking regions of the genes fused to the reporter. Each row of panels represents independent flowers of plant lines transformed with the same construct, at different stages of development. The names of the constructs are indicated on the right of each row of the respective panels. o, ovary; p, petal; s, stigmatic papillae; st, style; f, filament; a, anther.

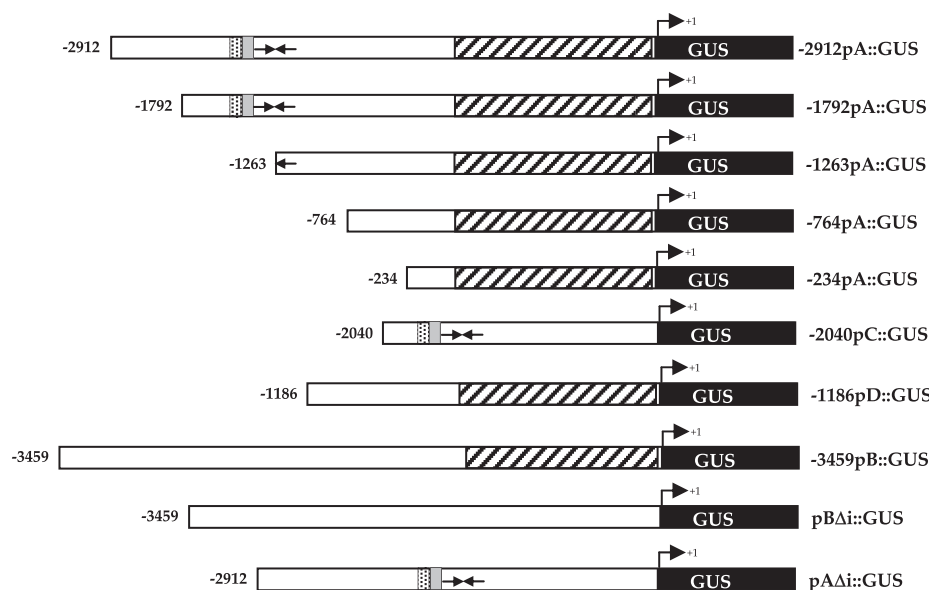
### A 529 bp region is crucial for cardosin A expression in *A. thaliana*

As a first approach to the identification of *cis* regulatory elements involved in the control of cardosin gene expression, we analyzed several 5'-deletions of the cardosin A promoter (Fig. 8) and examined their effect on gene expression in transgenic plants. Our results clearly show that the removal of 1 kb of the cardo-

sin A promoter region (–1792pA::GUS) did not greatly affect the transgene expression (Fig. 7D–F) in eight independent lines tested.

A subsequent 529 bp deletion of the promoter region from position –1792 to position –1263 (Fig. 8) completely abrogated transgene expression in all plant lines (data not shown). As successive 500 bp deletions to position –234 (Fig. 8) did not restore the transgene expression, the presence of a negative regulator was





**Fig. 8.** Structure of cardosin 5'-flanking region–GUS fusion constructs. The striped boxes represent the leader intron. The stippled and gray boxes indicate the (A/T) and (G/A) repeats, respectively. The inverted repeat found in the promoter regions of cardosins A and C is indicated by opposing arrows. The initiation codon of each construct is indicated by a bent arrow.

ruled out, and we assumed that important regulatory elements were present within the 529 bp region from –1792 bp to –1263 bp.

*In silico* analysis of this region (Fig. 3) revealed the presence of three putative regulatory elements: a long repetition ( $n = 10$ ) of the dinucleotide A/T, followed by a long repetition ( $n = 12$ ) of the dinucleotide G/A and an inverted repeat. All of these sequences were also found in the cardosin C promoter region, but were absent from the corresponding region of the cardosin D gene analyzed (Fig. 8).

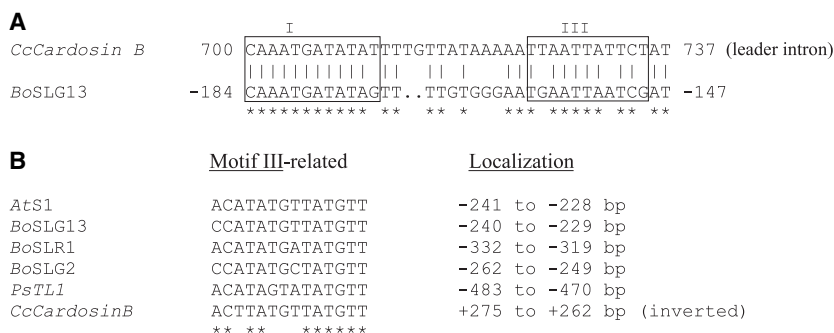
### The cardosin B but not the cardosin A leader intron is essential for gene expression

It is known that introns may participate in gene regulation, by modulating the level of expression and/or determining the specific pattern of expression of a gene [34–39]. To evaluate the relevance of the leader intron in cardosin expression, we deleted it from the 5'-flanking region of the genes (Fig. 8). The deletion of the cardosin A leader intron (construct pAΔi::GUS) did not affect the staining pattern of GUS (Fig. 7G–I), which was essentially similar to the one obtained when *A. thaliana* plants were transformed with the construct –2913pA::GUS (Fig. 7A–C), in six of the eight lines considered. Conversely, the deletion of the respective region from the cardosin B gene (construct pBΔi::GUS; Fig. 8) completely abolished the transgene

expression (Fig. 7P–R) in all plant lines, highlighting its important role in the regulation of cardosin B gene expression.

Comparison of the leader intron of the cardosin B gene with pistil-specific genes revealed the presence of putative regulatory elements. A region of SLG<sub>13</sub>, a gene involved in the prevention of self-pollination in *Brassica*, encompassing three boxes (I, II, and III), located 400 bp upstream of the initiation codon, is required for pistil-specific gene expression in transgenic tobacco [40]. A sequence sharing 77% similarity with that mentioned above and spanning 34 bp was identified in the leader intron of the cardosin B gene (Fig. 9). In addition, another element (motif III-related) was identified 438 bp downstream of the SLG<sub>13</sub>-like sequence [50] (Fig. 9). A similar motif is potentially implicated in pistil-specific expression of a pathogenesis-related protein gene from *Pyrus serotina* in transgenic tobacco (Fig. 9) [41,42]. Moreover, a motif III-related element also appears in the *Arabidopsis AtS1* gene (a ‘*Brassica*-like’ *S* gene; Fig. 9) that is expressed specifically in papillar cells and may function in pollination [43].

We have made two extra constructs harboring only the leader intron of the cardosin A and B genes. When tested under physiologic conditions, these constructs were not able to drive GUS expression in *Arabidopsis* (data not shown), revealing that they cannot act as alternative promoters.



**Fig. 9.** Conserved sequences among pistil-specific genes are also present in the leader intron of the cardosin B gene. (A) Sequences similar to those identified by Dzelzkalns *et al.* (boxes I and III) within the promoter region of the SLG<sub>13</sub> gene [40] also appear in the leader intron of the cardosin B gene. (B) A motif found in the S1, SLG, SLR1 and PR5 genes [41,42] is also present within the cardosin B leader intron, but in an inverted position. Asterisks denote identical nucleotides. At, *A. thaliana* L.; Cc, *C. cardunculus*; Bo, *B. oleracea*; Ps, *Py. serotina*.

## Discussion

Cloning by library screening of four full-length genes encoding cardosins A, B, C and D precursors, together with the cloning of a partial sequence of the cyprosin B gene and the isolation of the cyprosin A cDNA [11], indicates that *C. cardunculus* APs are encoded by a multigene family composed of at least six members, and reveal the coexistence of cardosins and cyprosin within the same plant.

The gene structure of cardosins basically reflects the same genomic organization as that of the few other typical AP genes that have been analyzed [16,19,29,30]. Given that monocotyledon and dicotyledon APs display the same pattern of exon–intron arrangement, the insertion of introns within the coding region possibly occurred before the divergence of both classes of plants [30]. Regarding the introns, the loss or gain of sequences may have taken place after monocotyledon and dicotyledon divergence, a fact that may explain: (a) the variable length of introns among different species and between gene family members of the same plant; and (b) the absence of one intron in the *A. thaliana* genes *AtPaspA2* and *AtPaspA3* [29].

Cardosins and cyprosin share a similar structural domain organization and display a high degree of identity in terms of primary structure. Interestingly, the slight differences among cardosins and between cardosins and cyprosin comprise the motifs RGD and KGE (Fig. 4). These motifs are known to mediate the cardosin A–phospholipase D $\alpha$  interaction, which may play an important physiologic role [7]. Cardosin B, which harbors an EGE instead of a KGE motif, does not bind to phospholipase D $\alpha$  [7]. Within *C. cardunculus* APs, only cardosin A and cardosin C possess the RGD and KGE motifs (Fig. 4), and therefore the for-

mation of a complex *in planta* with phospholipase D $\alpha$  is possibly restricted to these proteases.

In contrast to cyprosin, cardosins do not contain the residues Lys11 and Tyr13 (phytepsin amino acids numbering) in the N-terminal domain. These residues are well conserved among plant APs, and are involved in the inactivation mechanism of the precursor form of the enzymes [2,44]. Cardosins and the *Cy. humilis* AP are the only plant APs known to date whose Lys11/Tyr13 residues are absent from the primary structure, a feature that may explain the enzymatic activity exhibited by recombinant procardosins (Vieira *et al.*, unpublished results). From the scenario of plant AP evolution, it becomes evident that in *C. cardunculus*, the loss of the inactivation mechanism of the precursor forms occurred after the duplication of an ancestral gene common to cardosins and cyprosin (Fig. 5).

Comparison of protein data [3,9] with the results of gene expression studies (Fig. 6B) clearly indicates that cardosins are specifically expressed in the flowers of cardoon, although minor levels of cardosin A and D transcripts could also be detected in other plant organs (Fig. 6).

To further analyze the expression of cardosins, we fused their promoter region with the reporter gene GUS and assayed its activity in transgenic *A. thaliana* (Fig. 7). *A. thaliana* possesses three AP genes whose promoter regions do not exhibit any significant homology with the corresponding regions of cardosin genes (data not shown), which is in agreement with the different pattern of expression displayed by the APs of both species [6,9,29]. Nevertheless, the lack of sequence data on other plant AP promoter regions, in addition to the evolutionary proximity of groups Ia and Ib (Fig. 5), support our use of the model plant *A. thaliana* in our studies.

In *A. thaliana*, cardosins exhibit a flower-specific pattern of expression (Fig. 7), mirroring what was observed in the cardoon. The tissue-specific pattern, with some overlap of cardosins (cardosins A and C), suggests that they may have specific and/or complementary functions within the flower. The presence of cardosins A and C in the initial stages of flower development is in agreement with the presence of cyprosin B in flower meristems, as the antibody used in the immunocytochemical study could not discriminate among cardosins and cyprosin B [45]. The presence of cardosins in flower meristems, in addition to the lower amount of processed protein detected in immature cardoon flowers [3], led us to hypothesize that at least cardosins A and C may play important physiologic roles in the early stages of flower development, before being processed into the two-chain active enzymes.

We did not detect GUS activity driven by the 5'-flanking region of the cardosin A and C genes in pollen grains. This observation is not in agreement with the detection in pollen of cardosin A and C transcripts in *C. cardunculus* L (Fig. 6). However, the absence of other noncoding regions of the genes (introns and 3'-UTR), as well as the different sensitivities of the techniques (RT-PCR/GUS histochemical detection), as previously observed by others [46], may explain the discrepancies observed.

As a first attempt to determine the molecular basis for the flower-specific pattern of cardosin expression, several 5'-deletions of cardosin A promoter regions were generated, fused to the reporter GUS (Fig. 8), and then analyzed by means of GUS histochemical detection. Therefore, we were able to delimit a region of 529 bp, located between positions -1792 bp and -1263 bp, that is crucial for transgene expression in *A. thaliana*. An *in silico* search of this region against several plant *cis* element databases revealed the presence of at least three motifs that may also be relevant in cardosin A gene regulation. These motifs comprise an A/T repeat, followed by a G/A repeat, preceded by an inverted repeat (Figs 3 and 8). The A/T repeats enhance gene expression [47–49] in a copy number-dependent way [49]. Regions with five or more A/T repeats are recognized by the high-mobility group protein in several plant species [50]. Repeats of the dinucleotide G/A were reported to bind to the basic pentacysteine protein, and may be involved in the regulation of the expression of a diversity of genes [51]. Moreover, several inverted repeats are known to regulate gene expression [52]. Further studies should be performed to establish whether the 529 bp region is involved in the enhancement of gene expression or in the determination of its flower-specific pattern. Muta-

tional analysis will determine the relevance of the identified motifs.

The high similarity among cardosins A, C and D, which extends beyond the coding region (Fig. 3), and the relevance of the region of 529 bp in cardosin A expression may explain the ineffectiveness of the cardosin D flanking region in driving GUS expression. The cardosin C gene construct (-2040pC::GUS) also contains a highly similar region (Fig. 3). However, the sequence of the cardosin D promoter region included in the analysis does not contain the equivalent 529 bp region of cardosin A (Figs 3 and 8). It is therefore possible that its absence in the cardosin D promoter region could explain the lack of GUS activity in the transgenic *A. thaliana* plants transformed with the construct -1186 pD::GUS.

Although the presence of a leader intron is a well-conserved feature in plant AP genes [16,19], its biological role has not yet been clarified. To address this subject, we deleted the corresponding regions from the cardosin A and B genes (Fig. 8) and assayed GUS activity. The removal of the cardosin A leader intron did not affect GUS histochemical localization (Fig. 7G–I), whereas its deletion from the cardosin B gene completely abrogated gene expression (Fig. 7P–R). However, we cannot rule out the hypothesis that GUS levels varied when we deleted the cardosin A leader intron.

Two mechanisms may explain the lack of GUS expression driven by the cardosin B construct depleted of the leader intron: (a) positive regulatory elements lay within the leader intron; or (b) the leader intron may be increasing the steady state of mRNA levels, without significantly affecting the rate of transcription (intron-mediated enhancement [53]). Nuclear run-on transcription assays, in addition to the design of constructs in which the leader intron is inverted or partially deleted, will give clues to indicate the precise mechanism.

The similarity between the leader introns of cardosins A and B is restricted to the first 182 bp (Fig. 3). If the leader intron of cardosin B harbors important regulatory elements, they should be confined to the remaining sequence. We therefore compared this region of the cardosin B gene with the promoter regions of plant genes that are specifically expressed in pistils, namely genes involved in self-pollination avoidance (Fig. 9). The presence of a motif III-related sequence [41] within the leader intron of the cardosin B gene is particularly interesting, as its removal from a SLG<sub>13</sub> promoter::GUS construct abolished reporter expression in pistils of transgenic tobacco [40]. Furthermore, the deletion of an SLR1 gene region harboring one motif III-related sequence also eliminated its expression in the stigma and styles of transgenic

tobacco, and it was shown that as yet unidentified pistil transcription factor binds that region [54]. Therefore, the motif III-related sequence is a good candidate for a positive regulatory element of cardosin B gene pistil expression. Future work on transgene expression driven by a motif III-mutated leader intron of the cardosin B gene fused with a minimal promoter will certainly provide further insights into the relevance of this sequence.

## Experimental procedures

### Plant material and growth conditions

*C. cardunculus* tissue samples were collected from field-grown plants, frozen immediately in liquid nitrogen, and kept at  $-80^{\circ}\text{C}$  until use. *A. thaliana* (Columbia ecotype) surface-sterilized seeds were sowed in Murashige and Skoog medium (Duchefa Biochemie, Haarlem, the Netherlands), pH 5.8, containing 0.7% w/v of plant agar (Duchefa Biochemie). After 48 h of  $4^{\circ}\text{C}$  stratification treatment in the dark, the seeds were germinated at  $22^{\circ}\text{C}$  under a 16 : 8 h light/dark cycle. Two-week-old seedlings were subsequently transferred to soil until completion of the plant life cycle.

### Construction of a *C. cardunculus* library and cardosin gene isolation

DNA was extracted from young leaves of *C. cardunculus* subspecies *flavescens*, according to the method of Jofuku & Goldberg [55] and partially digested with *Sau3AI* (New England Biolabs, Beverly, MA, USA). The resulting fragments, ranging from 15 to 25 kb, were cloned into the  $\lambda$ Dash II vector (Stratagene, Beverly, MA, USA). Primary recombinants were amplified to create a stable library. A representative aliquot (380 000 plaque-forming units) of the library was independently screened by plaque hybridization with two different probes, CA-5'-and CB-3',  $^{32}\text{P}$ -labeled (Megaprime DNA Labelling System; Amersham, Uppsala, Sweden), using standard protocols [56]. CA-5'-consisted of a fragment obtained by standard PCR amplification of the cDNA of cardosin A (GeneBank accession number AJ132884 [3], using the primer pair S#1/R333 (Table 1). CB-3' comprised a 300 bp DNA fragment including the 3'-UTR region of the cardosin B gene obtained by 3'-RACE and amplified by PCR with the primers CardBS and FLCBR (Table 1). After three rounds of purification with probe CA-5', three positive plaques, named  $\lambda 5$ ,  $\lambda 6$ , and  $\lambda 18$  were isolated. The library screening with the CB-3' probe retrieved two positive plaques,  $\lambda 4.1$  and  $\lambda 4.2$ . The phage DNA was extracted [55], hydrolyzed with several restriction enzymes, and analyzed by Southern blotting. Fragments of phage DNA positively hybridizing with probes CA-5' or CB-3' were cloned into the pZER01 plasmid (Invitrogen,

Carlsbad, CA, USA) and sequenced using the GenomeLab DTCS-Quick Start Kit (Beckman Coulter, Krefeld, Germany) and the automatic sequencer CEQ 8000 Genetic Analysis System (Beckman Coulter).

As the cardosin A gene included in phage  $\lambda 5$  was not complete, we specifically amplified its 3'-region by PCR. Therefore, 300 ng of the nuclear DNA of *C. cardunculus* L was isolated (Invisorb Spin Plant Genomic DNA purification kit; Invitex, Berlin, Germany), and the region containing the three last exons of the cardosin A gene was PCR amplified with the primer pair Int10aF/R#2 (Table 1).

### RACE experiments

Total RNA from the pistils of *C. cardunculus* flower buds was isolated with the RNeasy Plant mini kit (Qiagen, Valencia, CA, USA) and used to generate an adaptor-ligated double-stranded cDNA RACE library with the Marathon cDNA Amplification kit (Clontech, Palo Alto, CA, USA). The 5'-UTR regions of the cardosin genes were amplified by PCR under standard conditions [56], using primer R333 (Table 1) combined with the kit-provided adaptor primers. R333 hybridizes with exon I of cardosin genes (Fig. 8) at a position 333 bp downstream of the ATG. The 3'-UTR of the cardosin B gene was amplified by means of a similar strategy, but using the CardBS primer as specific oligonucleotide (Table 1). The PCR products were cloned with the TA cloning kit (Invitrogen), and sequenced by automated DNA sequencing as described above.

### RT-PCR analysis of cardosin gene expression

Total RNA from *C. cardunculus* was isolated, as described above, from several tissues and from pistils of closed, partially opened and opened capitulum (the globular inflorescence grouped in a common receptacle surrounded by bracts). The RNA integrity and possible nuclear DNA contamination were evaluated in a 1% agarose gel, and equal amounts of RNA (1  $\mu\text{g}$ ) were used for cDNA synthesis (1st Strand cDNA Synthesis Kit for RT-PCR; Roche, Basel, Switzerland). Four sets of primers, specific for each of the cardosin genes, were used for RT-PCR amplification (Table 1): misAF1/misR1117 ( $63^{\circ}\text{C}$ , cardosin A); misCF1/misCR1 ( $65^{\circ}\text{C}$ , cardosin C); misDF1/misDR1 ( $60^{\circ}\text{C}$ , cardosin D); and CardBS/CardBR ( $58^{\circ}\text{C}$ , cardosin B) [9]. Owing to the high level of similarity among cardosins A, C and D, specific amplification of each gene was only possible after the introduction of an artificial mismatch in the antepenultimate or penultimate base of the specific primers (Table 1). As a negative control, the RNA from seeds was amplified through a RT-PCR reaction, with each set of primers, without the prior addition of avian myeloblastosis virus-reverse transcriptase (AMV-RT). The

**Table 1.** Sequences of the oligonucleotides used in this work.

Name	Sequence (5'– to 3')	Usage
S#1	ATGGGTACCTCAATCAAAGCAA	Probe CA-5'
R333	AGAACAGAACTTCCGGTATCG	Probe CA-5'-and 5'-RACE
Int10aF	GTGTGACACCGTAATAAGCAG	PCR
R#2	TCAAGCTGCTTCTGCAAATCC	PCR
5'PextA	CACACCCTCCTTCATTGCTTCCATCAAATAACAC	5' Primer extension
misAF1	ATGGGACATTTGGCGCTATCCA <sup>a</sup>	RT-PCR (cardosin A)
misR1117	GGTGCACATCTCATCATGTCTG <sup>a</sup>	RT-PCR (cardosin A)
misCF1	CCCCTGGTTGTCAAGCATATCTA <sup>a</sup>	RT-PCR (cardosin C)
misCR1	AGACTTGTGCTTATTCTTGTGC <sup>a</sup>	RT-PCR (cardosin C)
misDF1	GGGTGCCTTCTTCAAAGTGGTA <sup>a</sup>	RT-PCR (cardosin D)
misDR1	TGTATGCCACCAGAAGACTTCA <sup>a</sup>	RT-PCR (cardosin D)
CardBS	GATCTCGGCTGGGAAAGCG	RT-PCR (cardosin B) Probe CB-3'
CardBR	ATACCATTGCAGTCTACTATC	RT-PCR (cardosin B)
FLCBR	TTTATTGGACCATTTTATTCCGG	Probe CB-3'
ActF	GATATGGAAAAGATCTGGCATCAC	RT-PCR (Atactin 2)
ActR	TCATACTCGGCCTTGAGATCC	RT-PCR (Atactin 2)
– 2912AF	<u>AAAAAGCAGGCTATGAATTGCTAGAGTTGGTTAATGC</u> <sup>b</sup>	– 2912pA::GUS and pAΔi::GUS
– 1792AF	<u>AAAAAGCAGGCTTGCTGTTCTAAGTGTACTAGCTGGA</u> <sup>b</sup>	– 1792pA::GUS
– 1263AF	<u>AAAAAGCAGGCTCAAATTAATCGACGGTTGAG</u> <sup>b</sup>	– 1263pA::GUS
– 764AF	<u>AAAAAGCAGGCTCAATGTAGTACCAATTGGGGTACC</u> <sup>b</sup>	– 764pA::GUS
– 234AF	<u>AAAAAGCAGGCTGAGAAATCTATGGAATAAATAAAAAATTAGGG</u> <sup>b</sup>	– 234pA::GUS
PromAR	<u>AGAAAGCTGGGTCGATGTTTCACTGAAACATTAATAGATATTC</u> <sup>b,c</sup>	All chimeric cardosin A constructs except pAΔi::GUS
PromARΔi	<u>AGAAAGCTGGGTCGATGTTTCACTGAAACATTAATAGATATTC</u> <sup>b,c,d</sup>	pAΔi::GUS
– 2040CF	<u>AAAAAGCAGGCTTTAATAAGTTGTGCTTACACACAGTT</u> <sup>b</sup>	– 2040pC::GUS
– 2040CR	<u>AGAAAGCTGGGTCGATGTTTCACTGCAACCATGAC</u> <sup>b,c</sup>	– 2040pC::GUS
– 1186DF	<u>AAAAAGCAGGCTAACCCTATGATACCCTCACACCA</u> <sup>b</sup>	– 1186 pD::GUS
– 1186DR	<u>AGAAAGCTGGGTCGATGTTTCACTGCAACCATGAC</u> <sup>b,c</sup>	– 1186 pD::GUS
– 3459BF	<u>AAAAAGCAGGCTCTTAACGACTGCGTATATCCTC</u>	– 3459pB::GUS and pBΔi::GUS
PromBR	<u>AGAAAGCTGGGTCGATGTTTCACTGCAACCATGAC</u> <sup>b,c</sup>	– 3459pB::GUS
PromBRΔi	<u>AGAAAGCTGGGTCGATGTTTCACTGCAACCATGAAAGG</u> <sup>b,c,d</sup>	pBΔi::GUS

<sup>a</sup> The artificial mismatch introduced in the primer is underlined. <sup>b</sup> The *att* sequences are underlined. <sup>c</sup> The ATG within the reverse primer, mutated to ATC, is in Italic. <sup>d</sup> The six bases preceding the ATG were introduced in primers and are in Bold.

*A. thaliana* actin 2 gene (AGI code: At3g18780) was amplified by RT-PCR with the specific primer pair ActF/ActR (Table 1) and used as a positive control. At least two replicas were carried out for each RT-PCR reaction, and the products were sequenced to further confirm the specificity of the amplification.

### Chimeric plasmid constructs

The cardosin A, B, C and D 5'-flanking regions were cloned, using GATEWAY BP and LR reactions (Invitrogen), into the binary vector pKGWFS7 [57]. In this vector, an in-frame fusion between the regions coding for *EgfpER* and β-glucu-

ronidase (*gus*) is cloned downstream of the GATEWAY cassette. The initiation codon is therefore provided by the vector, and is the same for all the constructs. The 5'-flanking region of each cardosin gene was amplified by PCR from the respective genomic clone ( $\lambda 5$ ,  $\lambda 6$ ,  $\lambda 18$  or  $\lambda 4.1$ ), using the Platinum Pfx DNA Polymerase (Invitrogen). At the 5'-end, the primers contained *attB* sites, allowing subsequent fragment cloning by homologous recombination. The resulting PCR products with terminal *attB1* and *attB2* sequences were purified, and incubated with pDONR221 vector (Invitrogen) containing the *attP1* and *attP2* recombination sites, and the BP CLONASE enzyme (Invitrogen). This mixture was used to transform DH5 $\alpha$ -competent cells (Invitrogen), and the recombinant clones were selected on kanamycin-containing LB plates (50  $\mu\text{g}\cdot\text{mL}^{-1}$ ; Fluka Biochemika, Buchs, Switzerland). Positive recombinant pDONR plasmids were incubated with the binary vector pKGWFS7 [57] in the presence of LR CLONASE enzyme (Invitrogen). The pKGWFS7 vector has *attR1* and *attR2* recombination sites positioned upstream of the GUS reporter gene. After kanamycin selection, the resulting constructs were analyzed by restriction enzyme hydrolysis and sequenced as described above. Four plasmids containing the 5'-flanking regions of cardosin A (-2912pA::GUS), cardosin B (-3459pB::GUS), cardosin C (-2040pC::GUS) and cardosin D (-1186pD::GUS) were produced in this way (Fig. 8). As we did not perform an analysis of green fluorescent protein fluorescence, and for simplicity, the constructs were named, for example, 2912pA::GUS and not 2912pA::GFP::GUS. The primer pairs used in all chimeric constructs are listed in Table 1. To avoid the use of cardosin ATG as the initiation codon, all the reverse primers used to construct each GATEWAY cassette harbor a mutation in ATG (mutated to ATC) (Table 1). A similar strategy was used to make the constructs pA $\Delta$ i::GUS and pB $\Delta$ i::GUS, which enclosed the same 5'-flanking region included in constructs -2912pA::GUS and -3459pB::GUS, respectively, but without the leader intron. The leader intron was deleted by PCR amplification with the primer pairs -2912AF/PromAR $\Delta$ i (cardosin A gene) and -3459BF/PromBR $\Delta$ i (cardosin B gene) (Table 1, Fig. 8). A nested set of 5'-deletions in the cardosin A gene promoter region was also generated. Fragments differing from 500 bp at their 5'-end were amplified by PCR and cloned into pKGWFS7, generating constructs -1792pA::GUS, -1263pA::GUS, -764pA::GUS and -234pA::GUS (Fig. 8). In addition, a fragment of the 3'-UTR sequence of the rice *Os-ACS5* gene (GenBank accession no. X9706) [58,59] was cloned by homologous recombination into vector pKGWFS7 and used as a negative control.

### Plant transformation

Constructs were introduced into *Agrobacterium tumefaciens* strain LBA4404 by electroporation [60]. Bacteria harboring

the plasmid with the desired cardosin gene 5'-flanking region were grown to saturation in LB medium, and used to transform wild-type *A. thaliana* plants ( $T_0$  plants) by the floral dip method [61]. Transformants ( $T_1$  plants) were selected on Murashige and Skoog medium containing 50  $\text{mg}\cdot\text{L}^{-1}$  kanamycin and 0.7% w/v plant agar. Kanamycin-resistant plants were grown to the next generation ( $T_2$ ) and analyzed for GUS staining. *A. thaliana* infected with untransformed *Ag. tumefaciens* LBA4404 was also used as a negative control.

### Histochemical analysis of GUS activity

Histochemical GUS staining was performed for  $T_2$  vegetative tissues (leaves, roots, stems, siliques, inflorescences) and pollen with 5-bromo-4-chloro-3-indoxyl- $\beta$ -D-glucuronic acid (X-Gluc; ImmunoSource, Zoersel-Halle, Belgium) as substrate [62]. Samples were stained for 16 or 24 h at 37 °C. The stained organs were washed and incubated with 70% ethanol for 2 h, and clarified by incubation with a CLP solution [50 g of chloral hydrate (Riedel – de Haën, Seelze, Germany)] dissolved in 20 mL of lactic acid (Fluka Biochemika, Germany) and 25 mL of melted phenol crystals (Merck, Darmstadt, Germany)]. A stereo microscope (HQ Leica Microsystems, Wetzlar, Germany) attached to an image acquisition system was used to obtain the photographs. Eight independent *A. thaliana* independently transformed lines were analyzed per construct.

### Sequence analysis

WISE 2 (<http://www.ebi.ac.uk/Wise2>) and NEURAL NETWORK PROMOTER PREDICTION ([http://www.fruitfly.org/seq\\_tools/promoter.html](http://www.fruitfly.org/seq_tools/promoter.html)) were used in gene structure and promoter prediction, respectively. The multiple sequence alignments were constructed using CLUSTALW (<http://www.ebi.ac.uk/clustalw>), and edited and shaded in the program GENEDOC version 2.6 (<http://www.psc.edu/biomed/genedoc/gddl.htm>). Phylogenetic analysis was conducted with MEGA version 3.0 [26], using the neighbor-joining method with Poisson correction.

### Acknowledgements

We gratefully acknowledge financial support provided to C. P. by Fundação para a Ciência e Tecnologia (PRAXIS XXI/BD/21655/99), FEBS (FEBS short-term fellowship) and Fundação Calouste Gulbenkian (short-term fellowship).

### References

- 1 Schaaf A, Reski R & Decker EL (2004) A novel aspartic proteinase is targeted to the secretory pathway and

- to the vacuole in the moss *Physcomitrella patens*. *Eur J Cell Biol* **83**, 145–152.
- 2 Simões I & Faro C (2004) Structure and function of plant aspartic proteinases. *Eur J Biochem* **271**, 2067–2075.
  - 3 Faro C, Ramalho-Santos M, Vieira M, Mendes A, Simoes I, Andrade R, Verissimo P, Lin XL, Tang J & Pires E (1999) Cloning and characterization of cDNA encoding cardosin A, an RGD-containing plant aspartic proteinase. *J Biol Chem* **274**, 28724–28729.
  - 4 Frazão C, Bento I, Costa J, Soares CM, Verissimo P, Faro C, Pires E, Cooper J & Carrondo MA (1999) Crystal structure of cardosin A, a glycosylated and Arg-Gly-Asp-containing aspartic proteinase from the flowers of *Cynara cardunculus* L. *J Biol Chem* **274**, 27694–27701.
  - 5 Ramalho-Santos M, Verissimo P, Cortes L, Samyn B, Van Beeumen J, Pires E & Faro C (1998) Identification and proteolytic processing of procardosin A. *Eur J Biochem* **255**, 133–138.
  - 6 Ramalho-Santos M, Pissarra J, Verissimo P, Pereira S, Salema R, Pires E & Faro CJ (1997) Cardosin A, an abundant aspartic proteinase, accumulates in protein storage vacuoles in the stigmatic papillae of *Cynara cardunculus* L. *Planta* **203**, 204–212.
  - 7 Simões I, Mueller EC, Otto A, Bur D, Cheung AY, Faro C & Pires E (2005) Molecular analysis of the interaction between cardosin A and phospholipase D $\alpha$ . *FEBS J* **272**, 5786–5798.
  - 8 Verissimo P, Faro C, Moir AJG, Lin YZ, Tang J & Pires E (1996) Purification, characterization and partial amino acid sequencing of two new aspartic proteinases from fresh flowers of *Cynara cardunculus* L. *Eur J Biochem* **235**, 762–768.
  - 9 Vieira M, Pissarra J, Verissimo P, Castanheira P, Costa Y, Pires E & Faro C (2001) Molecular cloning and characterization of cDNA encoding cardosin B, an aspartic proteinase accumulating extracellularly in the transmitting tissue of *Cynara cardunculus* L. *Plant Mol Biol* **45**, 529–539.
  - 10 Brodelius PE, Cordeiro M, Mercke P, Domingos A, Clemente A & Pais MS (1998) Molecular cloning of aspartic proteinases from flowers of *Cynara cardunculus* subsp. *flavescens* cv. cardoon and *Centaurea calcitrapa*. *Aspartic Proteinases* **436**, 435–439.
  - 11 Cordeiro MC, Xue ZT, Pietrzak M, Pais MS & Brodelius PE (1994) Isolation and characterization of a cDNA from flowers of *Cynara cardunculus* encoding cyprosin (an aspartic proteinase) and its use to study the organ-specific expression of cyprosin. *Plant Mol Biol* **24**, 733–741.
  - 12 Cruz de Carvalho MH, d'Arcy-Lameta A, Roy-Macauley H, Gareil M, El Maarouf H, Pham-Thi AT & Zuily-Fodil Y (2001) Aspartic protease in leaves of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* L. Walp): enzymatic activity, gene expression and relation to drought susceptibility. *FEBS Lett* **492**, 242–246.
  - 13 Guevara MG, Oliva CR, Machinandiarena M & Daleo GR (1999) Purification and properties of an aspartic protease from potato tuber that is inhibited by a basic chitinase. *Physiol Plant* **106**, 164–169.
  - 14 Guevara MG, Almeida C, Mendieta JR, Faro CJ, Verissimo P, Pires EV & Daleo GR (2005) Molecular cloning of a potato leaf cDNA encoding an aspartic protease (StAsp) and its expression after *P. infestans* infection. *Plant Physiol Biochem* **43**, 882–889.
  - 15 Schaller A & Ryan CA (1996) Molecular cloning of a tomato leaf cDNA encoding an aspartic protease, a systemic wound response protein. *Plant Mol Biol* **31**, 1073–1077.
  - 16 Asakura T, Watanabe H, Abe K & Arai S (1995) Rice aspartic proteinase, oryzasin, expressed during seed ripening and germination, has a gene organization distinct from those of animal and microbial aspartic proteinases. *Eur J Biochem* **232**, 77–83.
  - 17 Bhalariao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J *et al.* (2003) Gene expression in autumn leaves. *Plant Physiol* **131**, 430–442.
  - 18 Buchanan-Wollaston V, Page T, Harrison E, Breeze E, Lim PO, Nam HG, Lin JF, Wu SH, Swidzinski J, Ishizaki K *et al.* (2005) Comparative transcriptome analysis reveals significant differences in gene expression and signalling pathways between developmental and dark/starvation-induced senescence in Arabidopsis. *Plant J* **42**, 567–585.
  - 19 Cruz de Carvalho MH, Pham-Thi AT, Gareil M, d'Arcy-Lameta A & Fodil YZ (2004) Isolation and characterization of an aspartic proteinase gene from cowpea (*Vigna unguiculata* L. Walp.). *J Plant Physiol* **161**, 971–976.
  - 20 Page T, Griffiths G & Buchanan-Wollaston V (2001) Molecular and biochemical characterization of post-harvest senescence in broccoli. *Plant Physiol* **125**, 718–727.
  - 21 Panavas T, Pikula A, Reid PD, Rubinstein B & Walker EL (1999) Identification of senescence-associated genes from daylily petals. *Plant Mol Biol* **40**, 237–248.
  - 22 Brown J & Simpson C (1998) Splice site selection plant pre-mRNA splicing. *Annu Rev Plant Physiol Plant Mol Biol* **49**, 77–95.
  - 23 Joshi CP, Zhou H, Huang X & Chiang VL (1997) Context sequences of translation initiation codon in plants. *Plant Mol Biol* **35**, 993–1001.
  - 24 Joshi CP (1987) An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res* **15**, 6643–6653.
  - 25 Helfman DM, Cheley S, Kuismanen E, Finn LA & Yamawaki-Kataoka Y (1986) Nonmuscle and muscle tropomyosin isoforms are expressed from a single gene

- by alternative RNA splicing and polyadenylation. *Mol Cell Biol* **6**, 3582–3595.
- 26 Kumar S, Tamura K & Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.
  - 27 Asakura T, Watanabe H, Abe K & Arai S (1997) Oryzasin as an aspartic proteinase occurring in rice seeds: purification, characterization, and application to milk clotting. *J Agric Food Chem* **45**, 1070–1075.
  - 28 Buchanan-Wollaston V & Ainsworth C (1997) Leaf senescence in *Brassica napus*: cloning of senescence related genes by subtractive hybridisation. *Plant Mol Biol* **33**, 821–834.
  - 29 Chen X, Pfeil JE & Gal S (2002) The three typical aspartic proteinase genes of *Arabidopsis thaliana* are differentially expressed. *Eur J Biochem* **269**, 4675–4684.
  - 30 D'Hondt K, Stack S, Gutteridge S, Vandekerckhove J, Krebbers E & Gal S (1997) Aspartic proteinase genes in the Brassicaceae *Arabidopsis thaliana* and *Brassica napus*. *Plant Mol Biol* **33**, 187–192.
  - 31 Runeberg-Roos P & Saarma M (1998) Phytpepsin, a barley vacuolar aspartic proteinase, is highly expressed during autolysis of developing tracheary elements and sieve cells. *Plant J* **15**, 139–145.
  - 32 Terauchi K, Asakura T, Nishizawa NK, Matsumoto I & Abe K (2004) Characterization of the genes for two soybean aspartic proteinases and analysis of their different tissue-dependent expression. *Planta* **218**, 947–957.
  - 33 Cordeiro MC, Lowther T, Dunn BM, Guruprasad K, Blundell T, Pais MS & Brodelius PE (1998) Substrate specificity and molecular modelling of aspartic proteinases (Cyprosin) from flowers of *Cynara cardunculus* subsp. *flavescens* cv. cardoon. *Aspartic Proteinases* **436**, 473–479.
  - 34 Bolle C, Herrmann RG & Oelmuller R (1996) Intron sequences are involved in the plastid- and light-dependent expression of the spinach PsdD gene. *Plant J* **10**, 919–924.
  - 35 Casas-Mollano JA, Lao NT & Kavanagh TA (2006) Intron-regulated expression of SUVH3, an *Arabidopsis* Su (var) 3–9 homologue. *J Exp Bot* **57**, 3301–3311.
  - 36 Chung BY, Simons C, Firth AE, Brown CM & Hellens RP (2006) Effect of 5'UTR introns on gene expression in *Arabidopsis thaliana*. *BMC Genomics* **7**, 120–133.
  - 37 Curi GC, Chan RL & Gonzalez DH (2005) The leader intron of *Arabidopsis thaliana* genes encoding cytochrome c oxidase subunit 5c promotes high-level expression by increasing transcript abundance and translation efficiency. *J Exp Bot* **56**, 2563–2571.
  - 38 Rose AB (2004) The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*. *Plant J* **40**, 744–751.
  - 39 Jeon JS, Lee S, Jung KH, Jun SH, Kim C & An G (2000) Tissue-preferential expression of a rice alpha-tubulin gene, OsTubA1, mediated by the first intron. *Plant Physiol* **123**, 1005–1014.
  - 40 Dzelzkalns VA, Thorsness MK, Dwyer KG, Baxter JS, Balent MA, Nasrallah ME & Nasrallah JB (1993) Distinct cis-acting elements direct pistil-specific and pollen-specific activity of the brassica-S locus glycoprotein gene promoter. *Plant Cell* **5**, 855–863.
  - 41 Ficker M, Kirch HH, Eijlander R, Jacobsen E & Thompson RD (1998) Multiple elements of the S-2-RNase promoter from potato (*Solanum tuberosum* L.) are required for cell type-specific expression in transgenic potato and tobacco. *Mol Gen Genet* **257**, 132–142.
  - 42 Sassa H, Ushijima K & Hirano H (2002) A pistil-specific thaumatin/PR5-like protein gene of Japanese pear (*Pyrus serotina*): sequence and promoter activity of the 5' region in transgenic tobacco. *Plant Mol Biol* **50**, 371–377.
  - 43 Dwyer KG, Kandasamy MK, Mahosky DI, Acciai J, Kudish BI, Miller JE, Nasrallah ME & Nasrallah JB (1994) A superfamily of S locus-related sequences in *Arabidopsis*: diverse structures and expression patterns. *Plant Cell* **6**, 1829–1843.
  - 44 Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A & Zdanov A (1999) Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. *EMBO J* **18**, 3947–3955.
  - 45 Brodelius M, Hiraiwa M, Marttila S, Al Karadaghi S, Picaud S & Brodelius P (2005) Immunolocalization of the saposin-like insert of plant aspartic proteinases exhibiting saposin C activity. Expression in young flower tissues and in barley seeds. *Physiol Plant* **125**, 405–418.
  - 46 Takahashi N, Kuroda H, Kuromori T, Hirayama T, Seki M, Shinozaki K, Shimada H & Matsui M (2004) Expression and interaction analysis of *Arabidopsis* Skp1-related genes. *Plant Cell Physiol* **45**, 83–91.
  - 47 Bustos MM, Guiltinan MJ, Jordano J, Begum D, Kalkan FA & Hall TC (1989) Regulation of beta-glucuronidase expression in transgenic tobacco plants by an A/T-rich, cis-acting sequence found upstream of a French bean beta-phaseolin gene. *Plant Cell* **1**, 839–853.
  - 48 Czarnecka E, Ingersoll JC & Gurley WB (1992) AT-rich promoter elements of soybean heat shock gene Gmhsp17.5E bind two distinct sets of nuclear proteins in vitro. *Plant Mol Biol* **19**, 985–1000.
  - 49 Sandhu JS, Webster CI & Gray JC (1998) A/T-rich sequences act as quantitative enhancers of gene expression in transgenic tobacco and potato plants. *Plant Mol Biol* **37**, 885–896.
  - 50 Pedersen TJ, Arwood LJ, Spiker S, Guiltinan MJ & Thompson WF (1991) High mobility group chromosomal proteins bind to AT-rich tracts flanking plant genes. *Plant Mol Biol* **16**, 95–104.
  - 51 Meister RJ, Williams LA, Monfared MM, Gallagher TL, Kraft EA, Nelson CG & Gasser CS (2004) Definition and interactions of a positive regulatory element of



- the Arabidopsis INNER NO OUTER promoter. *Plant J* **37**, 426–438.
- 52 Sidorenko LV, Li X, Cocciolone SM, Chopra S, Tagliani L, Bowen B, Daniels M & Peterson T (2000) Complex structure of a maize Myb gene promoter: functional analysis in transgenic plants. *Plant J* **22**, 471–482.
- 53 Mascarenhas D, Mettler IJ, Pierce DA & Lowe HW (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* **15**, 913–920.
- 54 Hackett RM, Cadwallader G & Franklin FC (1996) Functional analysis of a *Brassica oleracea* SLR1 gene promoter. *Plant Physiol* **112**, 1601–1607.
- 55 Jofuku KD & Goldberg RB (1988) Analysis of plant genome structure. In *Plant Molecular Biology: a Practical Approach* (Shaw CH, ed.), pp. 37–66. IRL Press, Oxford.
- 56 Ausubel F, Brent R, Moore D, Seidman J, Smith J & Struhl K (1995) *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience, New York.
- 57 Karimi M, Inze D & Depicker A (2002) GATEWAY vectors for Agrobacterium-mediated plant transformation. *Trends Plant Sci* **7**, 193–195.
- 58 Van Der Straeten D, Zhou Z, Prinsen E, Van Onckelen HA & Van Montagu MC (2001) A comparative molecular-physiological study of submergence response in lowland and deepwater rice. *Plant Physiol* **125**, 955–968.
- 59 Zhou Z, de Almeida EJ, Rouan D, Michiels F, Van Montagu M & Van Der SD (2002) Tissue localization of a submergence-induced 1-aminocyclopropane-1-carboxylic acid synthase in rice. *Plant Physiol* **129**, 72–84.
- 60 Wen-Jun S & Forde BG (1989) Efficient transformation of *Agrobacterium* spp. by high voltage electroporation. *Nucleic Acids Res* **17**, 8385.
- 61 Clough SJ & Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* **16**, 735–743.
- 62 Jefferson RA, Kavanagh TA & Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6**, 3901–3907.