Plant Molecular Biology **24**: 733–741, 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

Maria C. Cordeiro[1,4], Zhong-Tian Xue[2], Marciej Pietrzak[3], M. Salomé Pais[1] and Peter E. Brodelius[2,*]
[1] *Departamento Biologia Vegetal FCL, Bloco C2, Campo Grande, 1700 Lisboa, Portugal;* [2] *Department of Plant Biochemistry, University of Lund, POB 7007, S-22007 Lund, Sweden (* author for correspondence);* [3] *Department of Botany, University of Basel, Hebelstrasse 1, CH-4056 Basel, Switzerland;* [4] *Present address: Departmento de Biologia, Universidade de Evora, Apartado 94, P-7001 Evora Codex, Portugal*

Received 12 August 1993; accepted in revised form 11 January 1994

Key words: amino acid sequence, aspartic proteinase, cDNA sequence, *Cynara cardunculus*, cyprosin, flower development, organ specific expression

Abstract

Poly(A)$^+$ RNA isolated from flower buds of *Cynara cardunculus* has been used to prepare a cDNA library. Screening of the cDNA after expression of cloned DNA with antibodies raised against the large subunit of cyprosin 3 resulted in the isolation of six positive clones. One of these clones (*cypro1s*; a 1.7 kb *Eco* RI fragment) codes for cyprosin. The nucleotide sequence contain a 1419 bp open reading frame coding for 473 amino acids (aa) including a putative full-length mature protein (440 aa) and a partial prosequence (33 aa). *Cypro1s* contains a 162 bp 3′ non-coding region followed by a poly(A) tail. The deduced amino acid sequence shows high homology to other plant aspartic proteinases. The homology to mammalian and microbial aspartic proteinases is somewhat lower. Plant aspartic proteinases contain an insert of around 100 aa. We are modelling where this plant-specific insert will appear in the structure of cyprosin. Using *cypro1s* as a probe in northern blot analysis, the expression of cyprosin in developing flowers and other tissues has been studied. The signal on the northern blot increased for RNA samples from early (flower buds 6 mm in length) to later stages of floral development (flower buds up to 40 mm in length). In late stages of floral development (open flowers 50 mm in length and styles from such flowers) no hybridization signal was visualized showing that the synthesis of mRNA encoding the cyprosin starts in early stages of floral development and switches off at maturation of the flower. Southern blot analysis of genomic DNA showed 4–5 strong hybridizing bands and several minor bands indicating that the cyprosin genes are organized as a multi-gene family in *C. cardunculus*.

Introduction

Water extracts of dried flowers of cardoon or curdle thistle (*Cynara cardunculus*) are traditionally used for cheese making in Portugal. The Serpa and Serra cheeses are highly appreciated for their taste and quality. The clotting activity of the extract from flowers is due to proteolytic enzymes

The nucleotide sequence data reported will appear in EMBL, GenBank and DDBJ Nucleotide Sequences Databases under the accession number X69193.

present in the violet part of the flowers. Three heterodimeric aspartic proteinases (cyprosin 1, 2 and 3, previously called cynarase) have been purified from flowers of *C. cardunculus* and partly characterized [12]. They are glycosylated endopeptidases with maximum activity at pH 5.1 using casein as substrate.

Isoelectric focusing and reversed-phase HPLC have revealed a microheterogeneity of the apparently pure cyprosins [4]. Cyprosin 3 could be separated into three isozymes with close isoelectric points (i.e. 3.85, 4.00 and 4.15). Peptide mapping of cyprosins 2 and 3 by trypsin or BrCN cleavage indicate that they are derived from common procyprosin(s) [4].

The organ-specific expression of cyprosins has been investigated [4]. Immunostained western blots revealed the presence of cyprosins in low amounts in very young flowers. The amount of enzyme increased towards later stages of flower development. The enzyme could not be detected in leaves or seeds. Furthermore, the cyprosins are not expressed in cell suspension cultures of *C. cardunculus* obtained from hypocotyl-derived calli [5].

The milk clotting and proteolytic activities of the cyprosins have been compared to those of chymosin [3]. Cyprosin 3 presented the highest proteolytic activity among the three cyprosins and when compared to chymosin, similar clotting activities were obtained.

We now report on the isolation and characterization of a cDNA clone encoding cyprosin. The deduced amino acid sequence is used to model the structure of cyprosin using the structure of bovine chymosin b as template.

Materials and methods

Isolation of poly(A)$^+$ from C. cardunculus *flower buds for construction of a cDNA library*

Developing flower buds collected in May in the field near Lisbon, Portugal were used as starting material for isolation of total RNA. Total RNA was isolated using the guanidinium thiocyanate procedure [2], extracted twice with phenol-chloroform and precipitated with ethanol. RNA pellet was dissolved in 30 mM sodium acetate buffer, pH 6.0, containing 5 mM EDTA and 0.5% (w/v) sodium sarcosinate. Proteinase K was added to this solution (0.5 mg/ml) and the mixture was incubated for 1 h at 68 °C. Finally, 1/10 volume of 5.7 M CsCl was added to this mixture and subsequently the mixture (26 ml portions) was layered on top of a 5.7 M CsCl cushion (10 ml) in each SW27 ultracentrifuge tube. Tubes were centrifuged at 25000 rpm at 20 °C for 18 h. The pellet obtained contained pure RNA.

Poly(A)$^+$ RNA was isolated using an oligo(dT)-cellulose column [1]. Isolated poly(A)$^+$ RNA was used as a template for synthesis of cDNA using a commercial kit (cDNA Synthesis System Plus, Amersham). The cDNA library in a λgt11 cloning vector was made following the procedure of Young and Davis [27]. This library was screened, after expression of cloned cDNA, with antibodies produced against the large subunit of cyprosin 3. Six positive clones (*cypro1* to *cypro6*) were isolated.

Subcloning

Cypro1 was shown to contain two *Eco* RI fragments (2.0 and 1.7 kb). The 1.7 insert was isolated from an agarose gel using a DEAE membrane (NA 45 DEAE, Schleicher & Schuell) binding method. The isolated DNA was ligated to pBluescript plasmid transformed into *Escherichia coli* XL 1 Blue competent cells. The colourless transformants containing the recombinant plasmids were selected by X-gal containing medium and confirmed by the insert size and restriction pattern of the cloned DNA. The subclone obtained was designated *cypro1s*.

Restriction mapping

Cloned cDNA, *cypro1s*, was digested with *Eco* RI and double-digested with restriction enzymes:

Eco RI plus *Hind* III, *Bam* HI or *Pst* I (all supplied by Boehringer Mannheim). After digestion the DNAs were loaded on a 1% agarose gel, separated by electrophoresis and visualized by EtBr staining.

Preparation of hybridization probes

The insert of the cDNA clones *cypro1s* was labelled with α-^{32}P-dCTP (3000 Ci/mmol; Amersham) according to the instructions of the Oligo-labelling Kit (Pharmacia). Labelled DNA was collected after gel filtration over a Sephadex G-100 column.

Southern blot analysis

DNAs were prepared from cDNA clones according to a mini-preparation protocol [14] and were isolated from young flower tissue as described [19]. Isolated DNAs were digested with restriction enzymes at 37 °C for 2 h for plasmid DNA and 6 h for plant DNA and applied to a 0.7–1.0% (w/v) agarose gel. After electrophoresis the DNAs were denatured and transferred onto GeneScreen membranes (New England Nuclear). Hybridization was performed according to the instruction of the manufacturer.

DNA sequencing

DNA sequencing of *cypro1s* cDNA was performed by the dideoxy chain termination method of Sanger [21]. Purification of single-stranded template DNA from pBluescript SK-cDNA clone and sequencing were carried out as described in the Sequenase 2.0 Kit (United States Biochemicals) protocol. Initial DNA sequencing was carried out using reverse primer or M13–40 primer. In addition, oligonucleotides (15-mers) were synthesized according to sequence information obtained and used directly as primers for further sequencing.

Northern analysis

Total RNA was isolated from flower buds and flowers at different stages of development, bracts, leaves and midribs essentially according to the LiCl precipitation method [15]. Total RNAs (15 μg) were separated under denaturing conditions on 1.5% (w/v) agarose gels containing 9% (v/v) formaldehyde. After electrophoresis the RNAs were transferred onto a GeneScreen membrane. Hybridization was performed according to the instructions of the manufacturer. The membrane was analysed by autoradiography.

Results and discussion

Library screening

A cDNA library in a λgt11 cloning vector was prepared using poly(A)$^+$ mRNA isolated from young flowers of *C. cardunculus*. This library was screened, after expression of cloned cDNA, with antibodies produced against the large subunit of cyprosin 3 [12]. Six positive plaques were picked up and the recombinants were multiplied. These six positive clones were named *cypro1* to *cypro6*. DNA was produced from each lambda recombinant and *Eco* RI fragments were isolated for subcloning. The *Eco* RI fragments were cloned into a pUC9 plasmid cloning vector. The resulting recombinant clones were designated pCYP1 to pCYP6 and corresponded to λgt11 clones *cypro1* to *cypro6*. Analysis of these clones showed that the *cypro1* clone actually contained two *Eco* RI fragments (2.0 and 1.7 kb). These fragments were subcloned into the pBluescript phagemid. The subclone containing the 1.7 kb insert was called *cypro1s* and analysed further.

Identification of a E. coli *1.7 kb insert of* cypro1s *clone encoding the cyprosin*

The cDNA clones, 2.0 kb (*cypro4*) and 1.7 kb (*cypro1s*), were sequenced. The first clone to be sequenced, i.e. the 2.0 kb insert, did not show any

homology to aspartic proteinases [17]. Therefore, the 1.7 kb insert was sequenced and found to be highly homologous to such enzyme. The nucleotide homology between *cypro1s* and the barley aspartic proteinase (HvAP) and rice aspartic proteinase (OsAP) cDNAs is 70.1% and 63.4%, respectively. The nucleotide sequence of *cypro1s* contains a 1281 bp open reading frame coding for 427 amino acids. Upstream of the first ATG codon another 46 amino acids may be included in this open reading frame as part of a prosequence frequently observed in aspartic acid proteinases. However, a highly hydrophobic region for the characteristics of a signal peptide could not be identified in the 5′ region of *cypro1s*.

Northern analysis using the ^{32}P-labelled *Eco* RI 1.7 kb insert as probe revealed that the size of the mRNA encoding cyprosins in *C. cardunculus* is 1.8 kb and therefore it is highly likely that the *cypro1s* cDNA is near full-length and that the translation starts at a ATG codon about 100 bp upstream of the 5′ end of this cDNA. This assumption is supported by comparisons to the nucleotide and amino acid sequences of other aspartic proteinases as outlined below.

Cypro1s contains a 162 bp 3′ non-coding region followed by a poly(A) tail. This cDNA has a typical eukaryotic poly(A) addition signal AATAAA at nucleotides 1503 to 1508. The sequence data confirm a restriction mapping that was done for *cypro1s*, i.e. that there is no *Bam* HI or *Hin*d III site but a *Pst* I site at nucleotides 1165 to 1170.

The deduced amino acid sequence of cyprosin

The deduced amino acid sequence of *cypro1s* is shown in Fig. 1. The identity of the cloned cyprosin was confirmed by an internal amino acid sequence (**bold** in Fig. 1; aa 145–153), obtained by N-terminal microsequencing of a peptide isolated after BrCN hydrolysis of the large subunit of cyprosin 2 [4]. The putative N-terminal end of the mature protein is indicated. Furthermore, two putative active-site aspartic acid residues (Asp-36 and Asp-223) were been identified (Fig. 1). Two putative glycosylation sites (Asn-331 and Asn-378) were also found (Fig. 1). It is well established that the cyprosins are glycoproteins [12].

Comparison of deduced amino acid sequence of cyprosin to other aspartic proteinases

The deduced amino acid sequence of cyprosin is compared to the sequences of some other aspartic proteinases in Fig. 1. Aspartic proteinases are most often synthesized as preproenzymes and the sequences in Fig. 1 contain a signal sequence (except for cyprosin and rhizopuspepsin) and a prosequence (only in part for cyprosin). The start of the mature proteins is indicated. The size of the mature aspartic proteinases of plant origin is somewhat larger (440 (putative), 442 and 439 (putative) aa for cyprosin, HvAP [20] and OsAP [11], respectively) than that of rhizopuspepsin (325 aa [24]) and that of the two mature mammalian enzymes (348 aa for human cathepsin D [8], and 323 aa bovine chymosin b [10]). This difference in size is due to an internal additional sequence of around 100 aa in the plant enzymes.

A more detailed comparison between cyprosin and the other enzymes listed in Fig. 1 was carried out. In this comparison the sequence of the mature plant enzymes was divided into three regions, a N-terminal region (250 aa), a plant-specific region (around 100 aa) and a C-terminal region (88 aa), while the other three mature enzymes were divided into the corresponding N- and C-terminal regions.

The N- and C-terminal regions of cyprosin show a relatively high homology to the other aspartic proteinases, i.e. over 90% for the two plant enzymes and over 80% for the other three aspartic proteinases. All enzymes are highly conserved around the two active site aspartyl residues (Asp-36 and Asp-223). It is, however, interesting to note that the normal Asp-Thr-Gly sequence of catalytic aspartyl residues has been changed in one case to Asp-Ser-Gly (Asp-223 in cyprosin) in all three plant enzymes.

The overall homology between mature cyprosin (omitting the plant-specific region) and the other

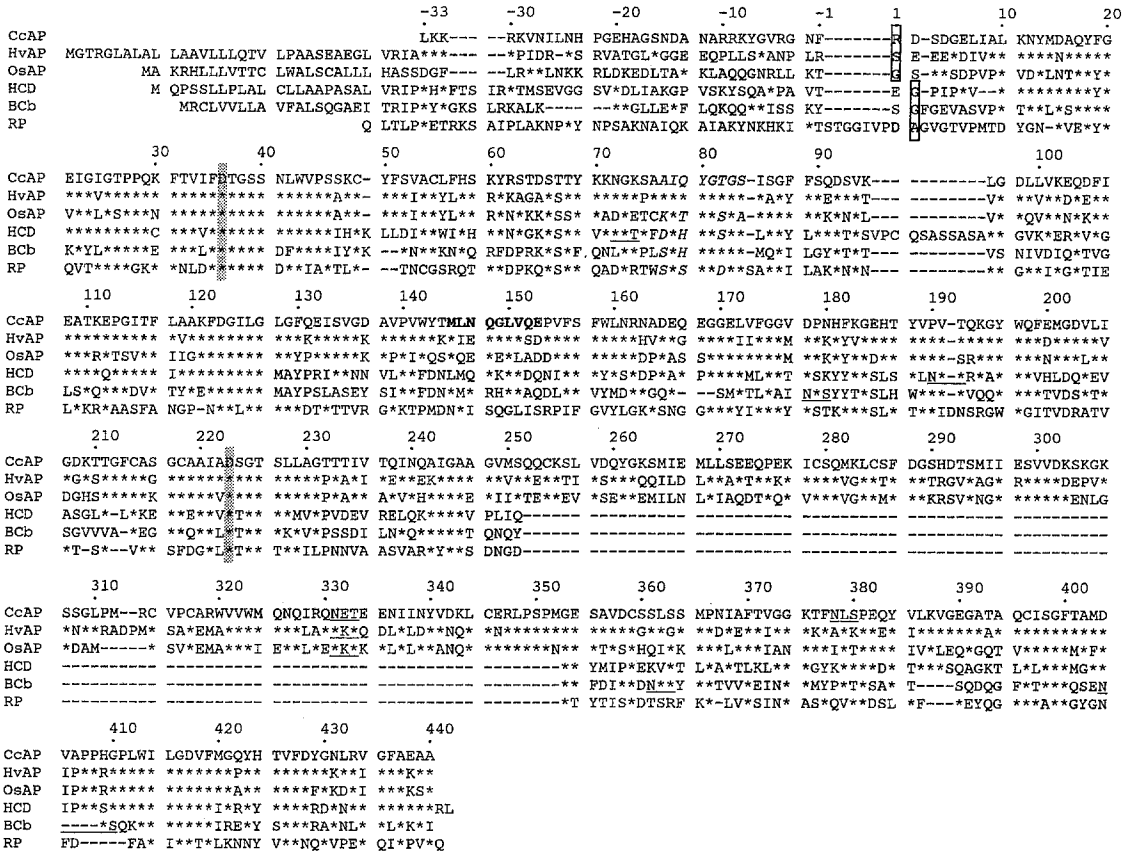

Fig. 1. Alignment of amino acid sequences of cyprosin (CcAP), barley aspartic proteinase (HvAP) [20], rice aspartic proteinase (OsAP) [11], human cathepsin D (HCD) [8], bovine chymosin b (BCb) [10] and rhizopuspepsin (RP) [7]. Stars in the lower five sequences indicate identity with the cyprosin sequence. Dashes indicate insertions of gaps at positions at which there are no homologous residues. The N-terminal amino acid of the mature proteins are boxed. The active site aspartyl residues are hashed. Potential glycosylation sites are underlined. An amino acid sequence obtained from a BrCN fragment is **bold** (amino acids 178 to 186). The putative flap region (Ala-110 to Ser-117) is in *italics*. The numbering of the amino acids starts from the putative N-terminus of the mature cyprosin. The prosequences are indicated by negative numbering.

five mature aspartic proteinases ranges from 34% (rhizopuspepsin) to 78% (HvAP) identity and from 85% (bovine chymosin B) to 95% (HvAP) similarity. Aspartic acid proteinases appear to be highly conserved proteins.

The plant-specific region

As outlined above a plant-specific region is well identified in the three plant enzymes (Fig. 1; aa 251 to 352). The size of these inserts is 102, 104 and 99 aa for cyprosin, HvAP and OsAP, respectively. The difference in size of the three plant specific regions is due to a deletion of two and five amino acids in cyprosin and rice aspartic proteinase found at the same position in the HvAP.

When comparing cyprosin to the other two plant enzymes the homology in the plant-specific region is considerably lower (91.2 and 87.7% for HvAP and OsAP, respectively) than for the N- and C-terminal regions (98.8 and 94.1% for HvAP and OsAP, respectively). Apparently, the amino acid sequence of the plant-specific regions have diverted during evolution to a greater extent

than the rest of these plant enzymes. In fact, the homology of the plant-specific regions is in the same order as that between plant (omitting the plant-specific region) and mammalian enzymes (85–92%). Consequently, this loop appears to be less important for enzyme structure and function.

A consensus sequence for the three plant enzymes shows that the plant-specific region exhibits 43% conserved residues indicating that this region has similar characteristics. In this respect it is interesting to note that the six cysteine residues present in the plant-specific region are conserved in all three enzymes indicating a possible conserved tertiary structure of this region through disulphide bridges. Furthermore, a putative glycosylation site (Asp-331; underlined in Fig. 1) is conserved in the plant-specific region in all three enzymes.

It has been suggested that all aspartic proteinases are derived from the same ancestral protein by a divergent evolutionary process [26]. This assumption was based on similarities between various mammalian, fungal and yeast enzymes. Not until recently the amino acid sequences of some plant enzymes have been deduced [11, 20, this paper]. Due to the high homology observed between these plant enzymes and other aspartic proteinases it is highly likely that also the plant enzymes have evolved from the common ancestral protein. However, the plant enzymes differ from other aspartic proteinases in that they contain a relatively large plant-specific region.

It may be interesting to speculate how the plant-specific region was introduced into the plant enzymes during evolution. It is most likely that this has occurred early in plant evolution. It could have been introduced into the plant proteins by an intron-exon exchange process or by insertion of a new exon at some stage during evolution. Some genes coding for aspartic proteinases have been isolated and sequenced; examples are calf prochymosin [13], human pepsinogen [22] and human cathepsin D [18]. All genes contain nine exons and eight introns and all exon-intron junction points coincide completely even in the nucleotide sequence showing a very conserved structure of these genes. Based on deduced amino acid sequences it may be concluded that intron 7 of these mammalian enzymes is located at a position corresponding to the plant-specific region. Thus, it may be possible that part of intron 7 of the mammalian genes corresponds to part of an exon of the plant genes encoding the plant-specific sequence. The isolation of a plant gene encoding an aspartic proteinase to investigate the number and positions of exons and introns will give an answer to this hypothesis. The isolation and characterization of cyprosin gene(s) is in progress in our laboratories.

Modelling of the localization of the plant-specific region in the cyprosin molecule

The crystal structure of a number of aspartic proteinases have been solved at high resolution [6]. Some examples are porcine pepsin, calf chymosin, rhizopuspepsin, penicillopepsin and endothiapepsin. The crystal structures of these aspartic proteinases are closely related in overall shapes and the tracing of chain foldings. As an example, the structure of bovine chymosin b is shown in

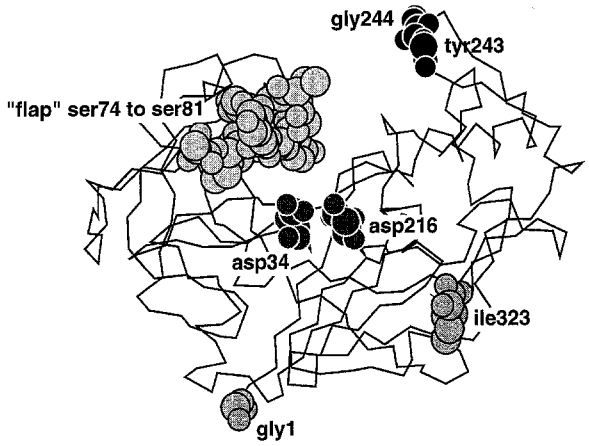

Fig. 2. The structure of bovine chymosin b ([9]; Brockhaven database entry 1CMS). The active site aspartyl residues (Asp-216), the flap region (Ser-74 to Ser-81), the N-terminal (gly-1) and C-terminal (Ile-323) amino acids, and the amino acids between which the plant specific region is inserted (Tyr-243 and Gly-244) are marked. The figure was made using the program MacImdad 3.06.

Fig. 2 [9]. The secondary structure consists almost entirely of pleated sheet with very little α-helix. The molecule is bilobal with two domains of similar structure containing one catalytic aspartic acid residue each related by a pseudodyad rotation axis. In virus proteinases this structure is maintained through two subunits each containing one catalytic aspartic acid. A large cleft, about 4 nm long, runs across the molecule and separates the two domains. In the middle of this extensive cleft, i.e. the active site, the structure of all enzymes are very similar. The molecule also contain a hairpin loop known as the 'flap' (cf. Fig. 2), which projects over the cleft at the active site and which participate in binding the substrate to the enzyme.

It may be predicted, due to high homology in primary structure, that the tertiary structures of aspartic proteinases in general are similar and that they are derived from the same ancestral protein [26]. because of the similarity in the primary and tertiary structures, it has been possible to model the tertiary structures of aspartic proteinases by fitting the amino acid sequences onto existing presumably homologous crystal structures [23].

Based on this discussion it should thus be possible to model at least in part the structure of plant aspartic proteinases using known structures. It will be interesting to study the influence of the extra loop (plant-specific region) on the three dimensional structure of plant aspartic proteinases. If the homologous sequence of cyprosin (i.e. the N- and C-terminal regions) is superimposed on the structures of other aspartic proteinases it is possible with a high probability to determine where the plant-specific loop will appear in the structure. Using the primary structure of cyprosin and the crystal of chymosin b [9], we can predict that the loop will be inserted at a position corresponding to between Tyr-243 and Gly-244 as indicated in Fig. 2. Similar results are obtained using the three dimensional structure of rhizopuspepsin [24]. The fact that the plant-specific region appears on the surface of the know structure is an indication that the structure of the homologoues regions is conserved.

The function of the plant-specific region is not at present known but the location of the loop would allow for the formation of a second substrate binding 'flap' over the active site cleft in analogy to dimeric aspartic proteinases from retroviruses (e.g. HIV-1 [16]). To evaluate this possibility it will be necessary to determine the three-dimensional structure of a plant proteinase in the presence of a peptide inhibitor (e.g. pepstatin). Attempts to crystallize cyprosin for such investigations are in progress in our laboratories.

Organ-specific expression and developmental regulation of the cyprosin gene

We have previously studied the organ-specific accumulation of cyprosin in *C. cardunculus* [4]. Immunostained western blots revealed the presence of cyprosin in young flower buds in small amounts. The amount of enzyme increased towards later stages of development and it was mostly present in the violet parts of styles and corollas of mature flowers. The enzyme could not be detected in leaves or seeds.

To evaluate if this specific expression of cyprosin is exerted at mRNA level and gene expression is developmentally regulated, equal amounts of total RNA, isolated from different organs including flower buds, bracts and leaves, were separated by agarose gel electrophoresis and blotted to GeneScreen membranes. Hybridization was carried out with the ^{32}P-labelled *Eco* RI 1.7 kb insert of *cypro1s*. The cDNA clone hybridized to a 1.8 kb mRNA from flower and bract tissues but there was no detectable hybridization to mRNA from leaves as shown in Fig. 3.

The intensity of hybridizing transcripts increased from early stages of floral development (flower buds 6–10 mm in length) to later stages of floral development (flower buds up to 40 mm in length) (Fig. 3). In the later stages of floral development (open flowers 50 mm in length and styles from such flowers) no hybridization signal was visualized (data not shown) indicating that the synthesis of mRNA encoding the cyprosin starts in early stages of floral development and switches

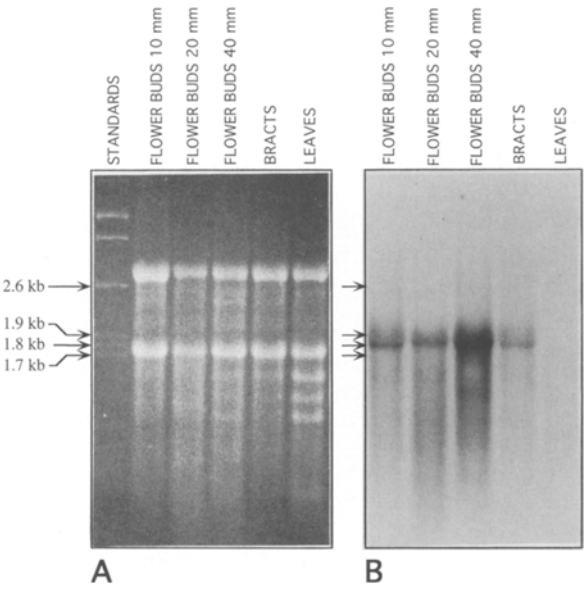

Fig. 3. Analysis of organ-specific expression of cyprosin in *Cynara cardunculus*. A. Total RNA isolated from various tissues of *C. carduncus*. B. Northern blot analysis of samples shown in panel A. The blot was probed with ^{32}P-labelled *cypro1s* insert.

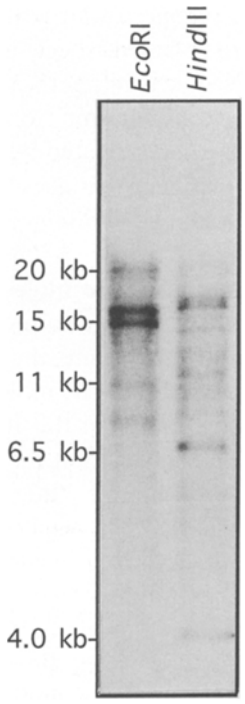

Fig. 4. Southern hybridization of *Cynara* genomic DNA. Flower DNA was digested with *Eco* RI (lane 1) or *Hin*d III (lane 2). The blot was probed with ^{32}P-labelled *cypro1s* insert.

off at maturation of the flower. These findings are in good agreement with enzyme activity (proteolytic and milk clotting) measured in extracts made from flowers in different stages of development [4].

Genomic organization of the cyprosin genes

Genomic DNA from young flower tissue of *C. cardunculus* was digested with restriction endonucleases and analysed by Southern hybridization (Fig. 4). The *Eco* RI and *Hin*d III digests each showed 4–5 strong hybridizing bands and several minor bands when the entire 1.7 kb insert of *cypro1s* was used as probe. From these hybridization patterns it may be suggested that the cyprosin genes are organized as a multi-gene family. The simultaneous expression of a number of cyprosin genes can explain the heterogeneity observed in the gene product [4]. Further studies are in progress to isolate cyprosin genes and to characterize this multi-gene family.

Acknowledgements

M.C. was supported by a doctoral fellowship from Junta Nacional de Investigacao Cientifica e Técnica (JNICT). The authors would like to thank Ms. F. Caeiro for isolation of RNA from flowers of *Cynara cardunculus* for construction of the cDNA library. The financial support to P.E.B. from the Swedish Council for Forestry and Agricultural Research is gratefully acknowledged.

References

1. Aviv J, Leder P: Purification of biologically active globin mRNA by chromatography on oligothymdilic acid cellulose. Proc Natl Acad Sci USA 69: 1406–1412 (1972).
2. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ: Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18: 5294–5299 (1979).
3. Cordeiro M, Jacob E, Puhan Z, Pais MS, Brodelius PE:

Milk clotting and proteolytic activities of purified cynarases from *Cynara cardunculus* L.: a comparison to chymosin. Milchwissenschaft 47: 683–687 (1992).
4. Cordeiro M, Pais MS, Brodelius PE: Tissue specific expression of multiple forms of cyprosin (aspartic proteinase) in flowers of *Cynara cardunculus*. *Physiol Plant*, in press (1993).
5. Cordeiro M, Xue ZT, Pais MS, Brodelius PE: Proteases from cell suspension cultures of *Cynara cardunculus*. Phytochemistry 33: 1323–1326 (1993).
6. Davies DR: The structure and function of the aspartic proteinases. Annu Rev Biophys Biophys Chem 19: 189–215 (1990).
7. Delaney R, Wong RNS, Meng GZ, Wu NH, Tang J: Amino acid sequence of rhizopuspepsin isozyme pl 5. J Biol Chem 262: 1461–1467 (1987).
8. Faust PL, Kornfeld S, Chirgwin JM: Cloning and sequence analysis of cDNA for human cathepsin D. Proc Natl Acad Sci USA 82: 4910–4914 (1985).
9. Gilliland GL, Winborne EL, Nachman J, Wlodawer A: The three-dimensional structure of recombinant bovine chymosin at 2.3 Å resolution. Proteins Struct Funct Genet 8: 82–101 (1990).
10. Harris TJR, Lowe PA, Lyons A, Thomas PG, Eaton MAW, Millican TA, Patel TP, Bose CC, Carey NH, Doel MT: Molecular cloning and nucleotide sequence of cDNA coding for calf preprochymosin. Nucl Acids Res 10: 2177–2187 (1982).
11. Hashimoto H, Nishi R, Uchimiya H, Kato A: Nucleotide sequence of a cDNA encoding aspartic proteinase in rice. EMBO Data Library D12777 (1992).
12. Heimgartner U, Pietrzak M, Geertsen R, Brodelius P, da Silva Figueiredo AC, Pais MSS: Purification and partial characterization of milk clotting proteases from flowers of *Cynara cardunculus* L. (cardoon). Phytochemistry 29: 1405–1410 (1990).
13. Hidaka M, Sasaki K, Uozumi T, Beppu T: Cloning and structural analysis of the calf prochymosin gene. Gene 43: 197–203 (1986).
14. Maniatis T, Fritsch EF, Sambrook J: Molecular Cloning; A laboratory Manual, 2nd Edition. Cold Spring Harbor Laboratories, Cold Spring Harbor, NY (1989).
15. Naito S, Dube DH, Beachy RN: Differential expression of conglycinin α and β subunit genes in transgenic plants. Plant Mol Biol 11: 109–120 (1988).
16. Navia MA, Fitzgerald PMD, McKeever BM, Leu C-T, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP: Three-dimensional structure of aspartyl proteinase from human immunodeficiency virus HIV-1. Nature 337: 615–620 (1989).
17. Pietzak M, Brodelius P, Xue Z-T, Cordeiro M, Pais MS: *Cynara cardunculus* mRNA sequence of *cypro4*. EMBO Data Library X69672 (1992).
18. Redecker B, Heckendorf B, Grosch H-W, Mersmann G, Haslik A: Molecular organization of the human cathepsin D gene. DNA Cell Biol 10: 423–431 (1991).
19. Rogers SO, Bendich AJ: Extracttion of DNA from plant tissue. In: Gelvin SB, Schilperoort RA, Verma DPS (eds) Plant Molecular Biology Manual, pp. A6/1–A1/10. Kluwer Academic Publishers, Dordrecht (1988).
20. Runeberg-Roos P, Törmäkangas K, Östman A: Primary structure of barley-grain aspartic proteinase. A plant aspartic proteinase resembling mammalian cathepsin D. Eur J Biochem 202: 1021–1027 (1991).
21. Sanger F: Determination of nucleotide sequences in DNA. Science 214: 1205–1210 (1981).
22. Sogawa K, Fujii-Kuriyama Y, Mizukami Y, Ichihara Y, Takahashi K: Primary structure of human pepsinogen gene. J Biol Chem 258: 5306–5311 (1983).
23. Srinivasan N, Blundell TL: An evaluation of the performance of an automated procedure for comparative modelling of protein teriary structure. Protein Engng 6: 501–512 (1993).
24. Suguna K, Bott RR, Padlan FA, Subramanian E, Sheriff S, Cohen GH, Davies DR: Structure and refinement at 1.8 Å resolution of the aspartic proteinase from *Rhizopus chinensis*. J Mol Biol 196: 877–900 (1987).
25. Tabor S, Richardson CC: DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc Natl Acad Sci USA, 84: 4767–4771 (1987).
26. Tang J, Wong RNS: Evolution in the structure and function of aspartic proteases. J Cell Biochem 33: 53–63 (1987).
27. Young RA, Davis RW: Efficient isolation of genes by using antibody probes. Proc Natl Acad Sci USA 80: 1194–1198 (1983).