Molecular cloning and characterization of procirsin, an active aspartic protease precursor from *Cirsium vulgare* (Asteraceae)

Daniela Lufrano [a], Rosário Faro [b], Pedro Castanheira [c], Gustavo Parisi [d], Paula Veríssimo [b,e], Sandra Vairo-Cavalli [a], Isaura Simões [b,c,*], Carlos Faro [b,c,e]

[a] *Laboratorio de Investigación de Proteínas Vegetales (LIPROVE), Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, C.C. 711, 1900 La Plata, Argentina*
[b] *CNC-Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal*
[c] *Biocant, Biotechnology Innovation Center, 3060-197 Cantanhede, Portugal*
[d] *Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Roque Sáenz Peña 352, Bernal, Buenos Aires, B1876BXD, Argentina*
[e] *Department of Life Sciences, University of Coimbra, 3004-517 Coimbra, Portugal*

ABSTRACT

Typical aspartic proteinases from plants of the Astereaceae family like cardosins and cyprosins are well-known milk-clotting enzymes. Their effectiveness in cheesemaking has encouraged several studies on other Astereaceae plant species for identification of new vegetable rennets. Here we report on the cloning, expression and characterization of a novel aspartic proteinase precursor from the flowers of *Cirsium vulgare* (Savi) Ten. The isolated cDNA encoded a protein product with 509 amino acids, termed cirsin, with the characteristic primary structure organization of plant typical aspartic proteinases. The pro form of cirsin was expressed in *Escherichia coli* and shown to be active without autocatalytically cleaving its pro domain. This contrasts with the acid-triggered autoactivation by pro-segment removal described for several recombinant plant typical aspartic proteinases. Recombinant procirsin displayed all typical proteolytic features of aspartic proteinases as optimum acidic pH, inhibition by pepstatin, cleavage between hydrophobic amino acids and strict dependence on two catalytic Asp residues for activity. Procirsin also displayed a high specificity towards κ-casein and milk-clotting activity, suggesting it might be an effective vegetable rennet.

The findings herein described provide additional evidences for the existence of different structural arrangements among plant typical aspartic proteinases.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Aspartic proteinases (APs, EC. 3.4.23) belong to one of the seven catalytic types of proteolytic enzymes and are widely distributed in nature in a variety of organisms such as viruses, bacteria, yeast, fungi, plants and animals (Dunn, 2002; Rawlings and Bateman, 2009; Simões et al., 2011). In plants, APs are distributed among

different clans and families according to MEROPS database (Rawlings et al., 2010) and appear to be the second-largest class of proteases in these organisms (van der Hoorn, 2008). Members of the A1 family are by far the best characterized and constitute the largest group of plant APs. They are found in a variety of tissues (Simões and Faro, 2004; Feijoo-Siota and Villa, 2011; González-Rábade et al., 2011) and their diversity triggered a redefinition of the classification of A1 plant APs into typical, nucellin-like and atypical proteases (Faro and Gal, 2005).

Although plant typical APs display differences in activity and specificity, they share several common features with their non-plant homologues. Typical plant APs are characterized by the presence of two aspartic acid residues organized in two conserved Asp-Thr/Ser-Gly (DT/SG) motifs responsible for catalytic activity, are most active at acidic pH, have a preference for cleavage between hydrophobic amino acids and are sensitive to pepstatin A. The overall three-dimensional structure is also conserved, consisting of two β-barrel-like domains that create a deep cleft where the active site is buried (Simões and Faro, 2004). Typical plant APs,

however, contain a distinctive additional segment of about 100 residues called plant-specific insert (PSI). This insert is folded as an independent domain, shares high topological similarity with the saposin-like proteins (SAPLIPs) and is suggested to be involved in membrane interactions and in the vacuolar sorting of several APs (Kervinen et al., 1999; Egas et al., 2000; Bryksa et al., 2011).

Typical APs are synthesized as single-chain preproenzymes and subsequently processed to single- or two-chain mature forms. In general, the PSI domain is totally or partially removed from the precursor, yet some variations exist on the mechanism and order by which this insert and the pro-segment are excised during conversion towards the active form (Simões and Faro, 2004). Processing of the precursor forms is suggested to involve sequential proteolytic cleavages along the secretory pathway that ultimately generate a complex mixture of heterogeneous mature aspartic proteinases within plant tissues (Duarte et al., 2008). Although no particular enzyme has been identified as a trigger in these processing events, evidence from in vitro processing experiments suggests that complete maturation of proteases may require the action of other(s) proteinase(s), in addition to autoproteolytic processing (White et al., 1999; Domingos et al., 2000; Castanheira et al., 2005; Sampaio et al., 2008). Diversity among plant typical APs may, therefore, be caused by the occurrence of different proteolytic cleavage sites on the same precursor and also by the expression of several genes encoding closely related enzymes, as shown with Cynara cardunculus typical APs multigene family (Pimentel et al., 2007). Although well documented in the literature, the correlation between diversity patterns of plant typical APs and their biological functions is still far from being deciphered.

Interestingly, typical plant APs have been extensively investigated over the years for their commercial potential as coagulants for cheesemaking. Aqueous extracts from flowers of species of the Cardueae tribe (Asteraceae family) like Cynara cardunculus, Cynara humilis and Cynara scolymus have been used in the Iberian Peninsula for centuries to manufacture ovine and/or caprine milk cheeses (Roseiro et al., 2003; Reis and Malcata, 2011). The high quality of artisanal cheeses produced with C. cardunculus plant extracts is recognized in the PDO status of several Portuguese and Spanish ewe milk cheeses such as Serra da Estrela, Serpa, Manchego or La Serena (EEC, 1996). The significance and suitability of cardosins and cyprosins – C. cardunculus active enzyme components – as enzyme coagulants encouraged studies on other species of the Asteraceae family for the identification of new APs with potential to be used in cheesemaking (Tamer and Mavituna, 1996; Llorente et al., 2004; Silva and Malcata, 2005; Vairo Cavalli et al., 2005). Thus far, milk-clotting activity was found in extracts of Centaurea calcitrapa, Onopordum acanthium, O. turcicum and Silybum marianum (Domingos et al., 2000; Vairo Cavalli et al., 2005; Brutti et al., 2012).

In this work, we isolated the cDNA encoding a novel typical AP from flowers of Cirsium vulgare (Savi) Ten., another member of the Asteraceae family also known as bull thistle. The encoded protease was named cirsin and shares with cyprosin a striking sequence similarity (98%). We expressed procirsin in Escherichia coli and, unlike other recombinant plant typical APs, it was active without undergoing any activation process. The herein reported milk-clotting activity and high specificity towards κ-casein also suggests its potential use as a new milk coagulant and strengthens the relevance of further studies on this activity.

2. Results

2.1. cDNA cloning and sequence analysis

The use of flower extracts from plants of the Asteraceae family as milk coagulants is associated with the manufacture of high quality ewe and goat milk cheeses, particularly in Mediterranean countries. In an attempt to identify new milk coagulants, several species of the Asteraceae family found in Argentina were screened for their ability to clot milk. Cirsium vulgare flower extracts were shown to be positive in this screening, thereby suggesting the presence of APs which are known to be responsible for this activity among Asteraceae family members.

To identify and further characterize bull thistle's AP(s), cDNA was synthesized using total RNA isolated from young flower buds and used as PCR template with primers designed for conserved regions of the ORF 5′ and 3′ ends of plant APs. The amplified product of 1530 bp was cloned and several clones were sequenced. This analysis resulted in the identification of at least four different deduced amino acid sequences exhibiting high levels of sequence identity among themselves and with other plant typical APs available in public databases, clearly suggesting the presence of several APs in C. vulgare flowers. One of these cDNA sequences was selected for further characterization and its nucleotide and deduced amino acid sequences are shown in Fig. 1 (GenBank accession number: JN703462). This protein product (named cirsin) encoded a preproenzyme of 509 residues with a putative hydrophobic signal peptide of 24 amino acids predicted by SignalP 4.0 (Petersen et al., 2011), a pro-segment of 44 residues, and a 441 amino acid-long polypeptide interrupted by 105 residues corresponding to the PSI domain. The preproenzyme has a predicted molecular mass of 55 kDa with a theoretical pI of 5.3. The deduced amino acid sequence displays the characteristic primary structure of typical plant APs in addition to two conserved catalytic triads (DTG 103–105 and DTG 290–292) and a tyrosine residue in position 147 (corresponding to DTG 32–34, DTG 215–217 and Y75, respectively, using pepsin numbering). Furthermore, two putative N-glycosylation sites conserved in several typical plant APs (Domingos et al., 2000; Vieira et al., 2001) were also predicted by NetNGlyc (Blom et al., 2004) at residues N139 (NGT motif) and N400 (NET motif), the latter in the PSI domain.

When compared with other plant AP precursors, the amino acid sequence of cirsin showed a high degree of similarity with cyprosin B (CAA57510) from C. cardunculus (identities: 95%, positives: 98%), cenprosin (CAA70340) from C. calcitrapa (identities: 94%, positives: 97%), the AP from Helianthus annuus (BAA76870, identities: 83%, positives: 92%), cardosin A (AJ132884) and cardosin B (AJ237674) from C. cardunculus (identities: 75%, positives: 83% and 85%, respectively) and phytepsin (X56136) from Hordeum vulgare (identities: 71%, positives: 84%) all peptidases belonging to A1 family. The striking sequence similarity with cyprosin is shown in Fig. 2. Most of the amino acid differences are found along the signal peptide, the pro-segment and the PSI domain whereas only 9 residues located in the sequences corresponding to the N- and C-terminal domains of the mature form of cyprosin are different. Remarkably, most of these differences correspond to conservative substitutions suggesting that cirsin and cyprosin are very closely related.

A phylogenetic analysis (Fig. 3) shows that procirsin shares a cluster with APs from diverse organisms such as C. cardunculus, Nicotiana tabacum, Populus trichocarpa, Ricinus comunis and Glycine max. Most of the proteins retrieved in this analysis are APs from Tracheophyta with very few belonging to Bryophyta. Again, pro-cyprosin B is found as the closest homologue of procirsin.

2.2. Heterologous expression of procirsin in E. coli

To further characterize this new AP from C. vulgare, the cDNA fragment encoding procirsin was amplified and subcloned into the expression vector pET23-d to produce the recombinant pro-enzyme in an E. coli-based expression system (BL21 Star (DE3) strain). A protein of about 50 kDa corresponding to recombinant procirsin was recovered from the insoluble fraction as inclusion bodies (Fig. 4). The inclusion bodies were washed in two steps and, after

```
atgggtacctcaatcaaagcaagcctgcttgccttgttttttgcttttttcttctatcacct
 M  G  T  S  I  K  A  S  L  L  A  L  F  L  L  F  L  L  S  P
accgcaatttcggtctccaatgatggattaattagagttggacttaaaaagagggaaggtg
 T  A  I  S  V  S  N  D  G  L  I  R  V  G  L  K  K  R  K  V
gaccaaattaaccaacttagtggacatggggcgtctatggaaggaaaggctagaaaagat
 D  Q  I  N  Q  L  S  G  H  G  A  S  M  E  G  K  A  R  K  D
tttggctttggtggtactctgagggactcggacagtgacattattgcactaaagaactac
 F  G  F  G  G  T  L  R  D  S  D  S  D  I  I  A  L  K  N  Y
atggatgctcagtactatggtgagattggtattggagctccacctcaaaagttcactgtg
 M  D  A  Q  Y  Y  G  E  I  G  I  G  A  P  P  Q  K  F  T  V
atttttgacaccggaagttctaatctatgggtgccttctgcaaagtgctacttttcagta
 I  F  D  T  G  S  S  N  L  W  V  P  S  A  K  C  Y  F  S  V
gcttgcctttttcactcaaagtacaagtcgagccattcaagtacctacaagaaaaatggc
 A  C  L  F  H  S  K  Y  K  S  S  H  S  S  T  Y  K  K  Ⓝ  G
acatctgctgctattcaatatggaactggatcaatctctggttttgttagccaagactct
 T  S  A  A  I  Q  Y  G  T  G  S  I  S  G  F  V  S  Q  D  S
gtcaaactcggtgatcttgttgttaaagagcaggatttttatagaggcaaccaaagagcct
 V  K  L  G  D  L  V  V  K  E  Q  D  F  I  E  A  T  K  E  P
ggcatcactttcttggcagccaagtttgatggtatacttggccttggatttcaggagatc
 G  I  T  F  L  A  A  K  F  D  G  I  L  G  L  G  F  Q  E  I
tctgttgggaaatctgttcctgtctggtacaacatggttaatcaaggtcttgttcaagaa
 S  V  G  K  S  V  P  V  W  Y  N  M  V  N  Q  G  L  V  Q  E
cctgtgttttcctttggtttaatcgcaatgctaatgagaagaagggggggtgaactcgtg
 P  V  F  S  F  W  F  N  R  N  A  N  E  K  K  G  G  E  L  V
ttgggtggtgttgaccctaatcattttaagggtaagcacacatatgtccctgtgaccgaa
 L  G  G  V  D  P  N  H  F  K  G  K  H  T  Y  V  P  V  T  E
aagggctattggcagtttaatatgggtgacgttcttattgaagataaaactaccggattt
 K  G  Y  W  Q  F  N  M  G  D  V  L  I  E  D  K  T  T  G  F
tgttctgatggttgtgcagcaattgccgactctggaacctctttgttggcaggtccaacg
 C  S  D  G  C  A  A  I  A  D  S  G  T  S  L  L  A  G  P  T
gctattattactgaaatcaatcatgcaattggtgctaagggagtcatgagccagcaatgc
 A  I  I  T  E  I  N  H  A  I  G  A  K  G  *V*  *M*  *S*  *Q*  *Q*  *C*
aaaacattggttagtcagtatggaaaagagtataatcgagatgctcctgtctgaggcgcaa
 *K*  *T*  *L*  *V*  *S*  *Q*  *Y*  *G*  *K*  *S*  *I*  *I*  *E*  *M*  *L*  *L*  *S*  *E*  *A*  *Q*
cctgataaaatctgttctcaaatgaagttatgcacttttgatggtgctcgtgatgttagt
 *P*  *D*  *K*  *I*  *C*  *S*  *Q*  *M*  *K*  *L*  *C*  *T*  *F*  *D*  *G*  *A*  *R*  *D*  *V*  *S*
tcaataattgagagtgtggttgacaagaacaacggcaagtcttctggtggcgcaaacgat
 *S*  *I*  *I*  *E*  *S*  *V*  *V*  *D*  *K*  *N*  *N*  *G*  *K*  *S*  *S*  *G*  *A*  *N*  *D*
gagatgtgtaccttctgtgaaatggcagtcgtttggatgcaaaaccaaatcaaacgaaac
 *E*  *M*  *C*  *T*  *F*  *C*  *E*  *M*  *A*  *V*  *V*  *W*  *M*  *Q*  *N*  *Q*  *I*  *K*  *R*  Ⓝ
gagactgaagataacataatcaactatgtcaacgagctgtgtgatcgcttacccagtcca
 *E*  *T*  *E*  *D*  *N*  *I*  *I*  *N*  *Y*  *V*  *N*  *E*  *L*  *C*  *D*  *R*  *L*  *P*  *S*  P
atgggagaatcagcagtagattgcaacagtctttcctccatgcccaatattgccttcaca
 M  G  E  S  A  V  D  C  N  S  L  S  S  M  P  N  I  A  F  T
attggtggcaaagttttcgagctctgcccagaacagtacatcctcaaaatcggtgaggga
 I  G  G  K  V  F  E  L  C  P  E  Q  Y  I  L  K  I  G  E  G
gaagcagcacaatgcatcagtggattcactgctatggatgtggcccctcctcgtggacct
 E  A  A  Q  C  I  S  G  F  T  A  M  D  V  A  P  P  R  G  P
ctatggatcttgggagatgtctttatgggtcgataccacacggtgttcgattatggcaag
 L  W  I  L  G  D  V  F  M  G  R  Y  H  T  V  F  D  Y  G  K
tcacgagttggatttgcagaagcagcttga
 S  R  V  G  F  A  E  A  A  -
```

Fig. 1. Full-length cDNA and derived-amino acid sequence of cirsin. First strand cDNA was synthesized from total RNA isolated from young flower buds and used as template in PCR amplification with primers specifically designed for the highly conserved 5′ and 3′ ends of APs coding sequences. The regions corresponding to the N-terminal signal peptide (dotted arrow), the pro-segment (dashed arrow) and the plant-specific insert (italic) are indicated. Potential N-linked glycosylation sites are circled and the catalytic triads are boxed. Cirsin cDNA sequence has been submitted to the GenBank under accession no JN703462.

Fig. 2. Amino acid sequence alignment of cirsin and cyprosin (accession number: Q39476). The regions corresponding to the N-terminal signal peptide (◄·····►), the pro-segment (◄- -►) and the plant-specific insert (◄———►) are indicated by arrows over the corresponding sequences. Potential N-glycosylation sites are circled and the catalytic triads are boxed.

denaturing in 8 M urea buffer, the refolding was induced by a rapid dilution step followed by the slow adjustment of the pH to 8.0. Upon refolding, the high molecular mass soluble aggregates were removed by size-exclusion chromatography on a HiLoad 26/60 Superdex 200 gel filtration column (Fig. 5A). Collected fractions were screened for activity towards the typical AP substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) at pH 4.0 and the correspondent digestion products were separated by RP-HPLC. The equivalent chromogenic version of this substrate has been used to assess activity of a wide variety of APs of the typical type (Dunn and Hung, 2000; Castanheira et al., 2005) as pepsin-like enzymes were shown to cleave preferentially peptide bonds comprising aromatic or hydrophobic amino acids in P1 and P1′ positions. Fig. 5D shows a representative RP-HPLC elution profile obtained for recombinant procirsin's active fractions (highlighted with dotted lines in Fig. 5A). The active fractions were then pooled and purified by ion-exchange chromatography on a MonoQ column (Fig. 5B). Proteolytic activity was screened under similar conditions by RP-HPLC and active fractions (Fig. 5D; region highlighted with dotted lines in Fig. 5B) were pooled for SDS–PAGE analysis confirming the presence of a protein with the expected molecular mass ca. 50 kDa (Fig. 5C, Lane 3).

In order to confirm the role of the predicted catalytic aspartate residues for proteolytic activity, the putative catalytic Asp103 residue was replaced by an alanine residue and the mutant pro-enzyme was expressed, purified and tested under similar conditions as described for recombinant wt procirsin. Purified fractions were analyzed by SDS–PAGE and displayed a protein pattern similar to that observed for wild-type (wt) procirsin (Fig. 5E). Activity screening assays confirmed the lack of activity of the active-site mutant towards the typical AP substrate (Fig. 5F), as proposed by the deduced amino acid sequence analysis. The identity of the active-site mutant protein was confirmed by MALDI-TOF MS (Supplementary information, Fig. S1 and Table S1).

To determine whether the recombinant cirsin precursor was able to undergo in vitro autoactivation processing towards intermediate and/or mature forms, as described for other plant typical APs (Glathe et al., 1998; Domingos et al., 2000; Castanheira et al., 2005), purified recombinant wt procirsin was incubated with different buffers over a pH range of pH 3.0–7.0. As shown in Fig. 6, no defined pattern of conversion was observed in SDS–PAGE analysis at any pH value tested and also no increase in proteolytic activity was detected as a result of this incubation. Hence, no clear intermediate or mature forms could be assigned to recombinant procirsin from these activation assays.

To further evaluate the significance of the PSI domain in the proteolytic processing and activity of procirsin, a third construct was designed where the sequence encoding the entire PSI domain (region between Val315 and Pro420) was deleted. This construct (procirsinΔPSI) was expressed in E. coli in the form of inclusion bodies, refolded and purified according to a similar procedure as that described for recombinant wt procirsin. However, the purified protein product was shown to be inactive against the synthetic substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP).

2.3. Biochemical characterization

Given the results obtained with the different cirsin constructs tested, all characterization studies were focused exclusively on wt procirsin. Purified recombinant wt procirsin active fractions were pooled and tested for ability to cleave several model synthetic substrates. Only the typical AP fluorogenic substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) was significantly cleaved in these assays. Substrates designed for proteases which were described as atypical APs (MCA)Lys-Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu-Lys(DNP) (Simões et al., 2007) and (MCA)Lys-Lys-Leu-Ala-Asp-Val-Val-Asn-Ala-Leu-Glu-Lys-Lys(DNP) were only poorly or not cleaved, respectively. No hydrolysis was observed
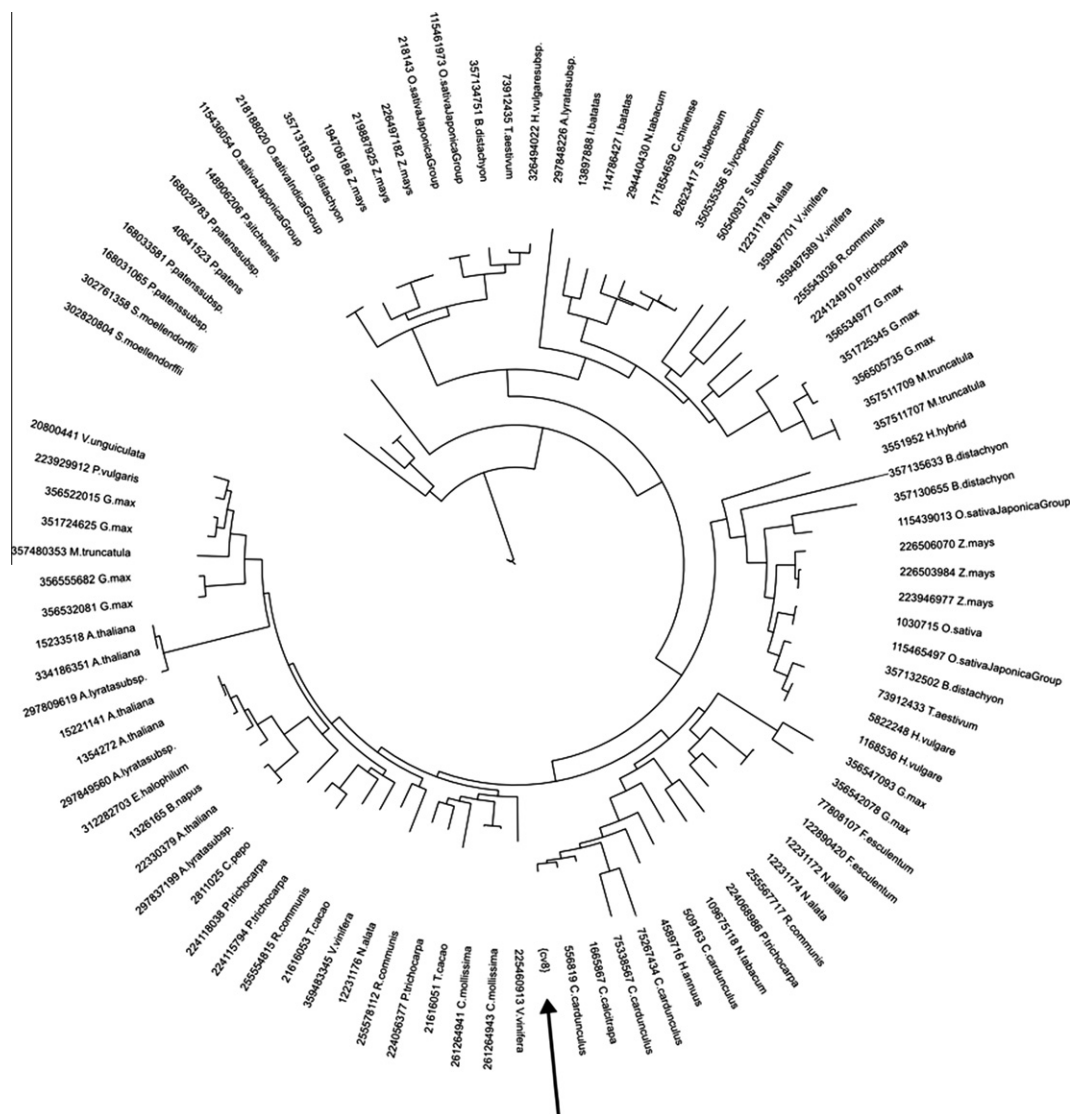
Fig. 3. Phylogenetic analysis of the homologous proteins of procirsin. The tree was obtained by maximum likelihood calculation using the program Phyml as explained in Experimental Procedures. Different proteins in the tree are indicated with NCBI ID followed by genera and species names. Procirsin is indicated with an arrow.

with the fluorogenic BACE 1 substrate (MCA)Lys-Ser-Glu-Val-Asn-Leu-Asp-Ala-Glu-Phe-Lys(DNP), renin substrate 1 Arg-Glu(E-DANS)Ile-His-Pro-Phe-His-Leu-Val-Ile-His-Thr-Lys(DABCYL)Arg nor the HIV protease substrate 1 Arg-Glu(EDANS)Ser-Gln-Asn-Tyr-Pro-Ile-Val-Asn-Lys(DABCYL)Arg. This cleavage specificity pattern is consistent with the high similarity found between cirsin amino acid sequence and those of typical plant APs. Oxidized insulin β-chain was also used as a substrate to evaluate recombinant procirsin specificity due to its high content of hydrophobic amino acids (Kervinen et al., 1993; Bleukx et al., 1998; Park et al., 2000; Payie et al., 2003; Athauda et al., 2004; Guevara et al., 2004; Simões et al., 2007). Incubation of oxidized insulin β-chain with recombinant wt procirsin was followed by RP-HPLC separation of digestion products. Analysis by MS of the predominant products revealed that the most susceptible cleavage site was at Leu15-Tyr16.

Hydrolysis of the substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) was studied for determination of kinetic parameters. Kinetic data was well described by the classic hyperbolic curve consistent with the Michaelis–Menten model. The values determined for cleavage at pH 4.0 were K_m: 7.13 ± 1.30 μM and

k_{cat}: 4.24 ± 0.05 s^{-1}. The catalytic efficiency determined for recombinant procirsin was 594 mM^{-1} s^{-1}.

The effect of pH on recombinant procirsin proteolytic activity was studied using the same substrate – (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) – in a range of pH values between pH 2.25 and pH 7.0 (Fig. 7A). The maximum activity was observed at pH 4.0 which is consistent with optimum acidic pHs reported for other typical plant APs (Kervinen et al., 1993; Ramalho-Santos et al., 1998; White et al., 1999; Castanheira et al., 2005; Mazorra-Manzano and Yada, 2008; Sampaio et al., 2008) and the pH values for half maximal activity were 3.19 and 4.45. The temperature dependence of the cleavage was also determined. As shown in Fig. 7B, the highest activity was detected between 30 °C and 37 °C at pH 4.0. For temperatures above 37 °C there was a decrease in enzyme activity and at 65 °C recombinant procirsin retained only 10% of activity.

As expected, incubation with the specific aspartic protease inhibitor pepstatin A produced a strong inhibition (98%) of recombinant procirsin activity (Table 1) and is in agreement with the lack of activity observed for the active-site mutant procirsinD103A. The effect of specific inhibitors of other classes of proteinases was also
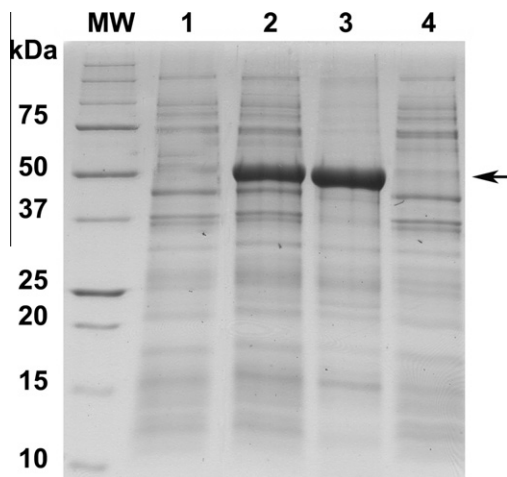
Fig. 4. Analysis of recombinant procirsin expression in *E. coli*. Procirsin-pET23-d plasmid construct containing the cDNA encoding procirsin without the putative signal peptide (first 25 amino acid residues) was expressed in *E. coli* BL21Star (DE3) strain in the form of inclusion bodies. Samples collected throughout expression were analyzed by SDS–PAGE: (Lane 1) *E. coli* protein extracts collected immediately before induction; (Lane 2) *E. coli* protein extracts collected after 3 h of expression; (Lane 3) insoluble protein fraction; (Lane 4) soluble protein fraction. Gel was stained with Coomassie Brilliant Blue; MW, molecular weight marker.

evaluated and, somehow unexpectedly, pefabloc and EDTA were able to partially inhibit recombinant wt procirsin activity. Incubation with DTT as well as with E-64 and amastatin did not have a major effect on activity.

2.4. κ-Casein cleavage and milk-clotting activity

A good milk-clotting enzyme is characterized by a high specificity for cleaving the κ-casein macropeptide which triggers the destabilization of the casein micelles and, therefore, induces milk-clotting (Hyslop, 2003; Jacob et al., 2011). As *C. vulgare* flower extracts have shown milk-clotting activity and given the similarity between cirsin and other typical plant APs recognized as milk coagulants, in particular with cyprosin from *C. cardunculus*, we have tested recombinant wt procirsin for its milk-clotting potential. The cleavage of purified bovine κ-casein by the recombinant enzyme was evaluated by SDS–PAGE analysis. A well-defined cleavage product of 15 kDa was observed, demonstrating the high specificity of recombinant procirsin towards κ-casein (Fig. 8A). A similar digestion profile has been obtained by action of chymosin as well as other plant APs over κ-casein (Timotijević et al., 2006) and corresponds to the cleavage of the peptide bond Phe105-Met106 which is responsible for triggering milk-clotting.

To further confirm the milk-clotting activity of procirsin, the enzyme was tested using a skim milk solution through overnight incubation at 37 °C and curd formation was determined by visual observation. As illustrated in Fig. 8B, recombinant procirsin was able to coagulate milk and this activity was totally inhibited when the incubation was performed in the presence of pepstatin A.

3. Discussion

Typical plant APs are the group of enzymes that include the saposin-like domain known as plant-specific insert (PSI) in their precursor sequences. Although *in silico* analysis has demonstrated this group constitutes the minority of the AP genes in *Arabidopsis thaliana* (Faro and Gal, 2005), it is still the best characterized among plant APs. In this work, we report the identification of a novel typical AP – cirsin – expressed in flowers of *C. vulgare*. cDNA amplification with conserved primers resulted in the isolation of

other three typical APs coding sequences clearly suggesting that in *C. vulgare*, as in *C. cardunculus* (Pimentel et al., 2007), several closely related enzymes of this group are present. To start characterizing these new members of plant typical APs, we have cloned, expressed, purified and characterized procirsin, demonstrating also its ability to coagulate milk. The cDNA derived-amino acid sequence of cirsin displays the usual organization of plant typical APs precursors. The overall similarity between cirsin amino acid sequence and those of other plant typical APs, together with the presence of the hallmark sequence motifs of APs, clearly assigns cirsin as a new member of this family of enzymes.

Rather surprising is the striking similarity with cyprosin B from *C. cardunculus* (Cordeiro et al., 1994) which ranges 95% of sequence identity. This suggests that cirsin and cyprosin are very closely related anticipating a high sequence/structure relationship between these enzymes. When compared with cyprosin B in more detail, the similarity is lower in the PSI domain (93%) than in the remaining sequence. This observation is in agreement with a higher diversity among PSI sequences that might be the result of evolutionary adaptations (Cordeiro et al., 1994) as suggested by the role of the PSI in plant APs either in vacuolar sorting or in correct folding (or both), depending on the enzyme (Tormakangas, 2001; Brodelius et al., 2005; Terauchi et al., 2006; Duarte et al., 2008).

In contrast with other plant typical APs which were shown to be inactive when expressed as precursors in prokaryotic systems (Asakura et al., 2000; Castanheira et al., 2005), the herein reported expression of procirsin in *E. coli* resulted in the purification of a single-chain protein that was active towards the substrate typically used to study APs. Recombinant procirsin was shown to share the same enzymatic features of other typical APs: maximum activity at acidic pH, inhibition by pepstatin A, strict dependence on two catalytic Asp residues for activity, a preference towards cleavage of peptide bonds between hydrophobic amino acids (Simões and Faro, 2004) and milk-clotting activity. A pro-enzyme with catalytic activity is not totally unexpected as some precursors of mammalian APs from A1 family, like BACE and renin, show significant protease activity (Pitarresi et al., 1992; Shi et al., 2001; Ermolieff et al., 2000). However, this phenomenon was not yet clearly described for typical plant APs. According to our data, procirsin pro-region appears to adopt different conformations as proposed for BACE (Ermolieff et al., 2000), rendering the active site in equilibrium between an open vs. a closed state and, thereby, enabling the substrate to access the catalytic cleft. Although displaying an acidic optimum pH, recombinant procirsin pH dependence activity does not show the typical bell-shaped profile characteristic of mature APs, also suggesting that the anchorage of the pro-region in the active site may likely depend on a network of electrostatic interactions that are disturbed only under very limited pH intervals. Further studies are required to assign residues controlling the reported procirsin activity. However, these results provide additional evidences that plant typical APs precursors may occur in diverse structural arrangements – from different inactivation mechanisms (prophytepsin vs. procardosin) (Kervinen et al., 1999; Castanheira et al., 2005) to active pro forms (procirsin) – and this may be correlated with the diversity patterns of plant typical APs and their biological functions.

Remarkably, when the autocatalytic activation of recombinant procirsin was assessed *in vitro* at different pH values no conversion was observed nor an increment in proteolytic activity, clearly suggesting that although catalytically active, the precursor form of cirsin is unable to trigger autoactivation or processing. This clearly contrasts with recombinant plant typical APs as prophytepsin (Glathe et al., 1998), procardosin A (Castanheira et al., 2005) and procenprosin (Domingos et al., 2000) to which multi-step activation upon acidification has been clearly assigned. Although one cannot exclude that procirsin may require very specific conditions to onset an intramolecular proteolytic event, these results appear
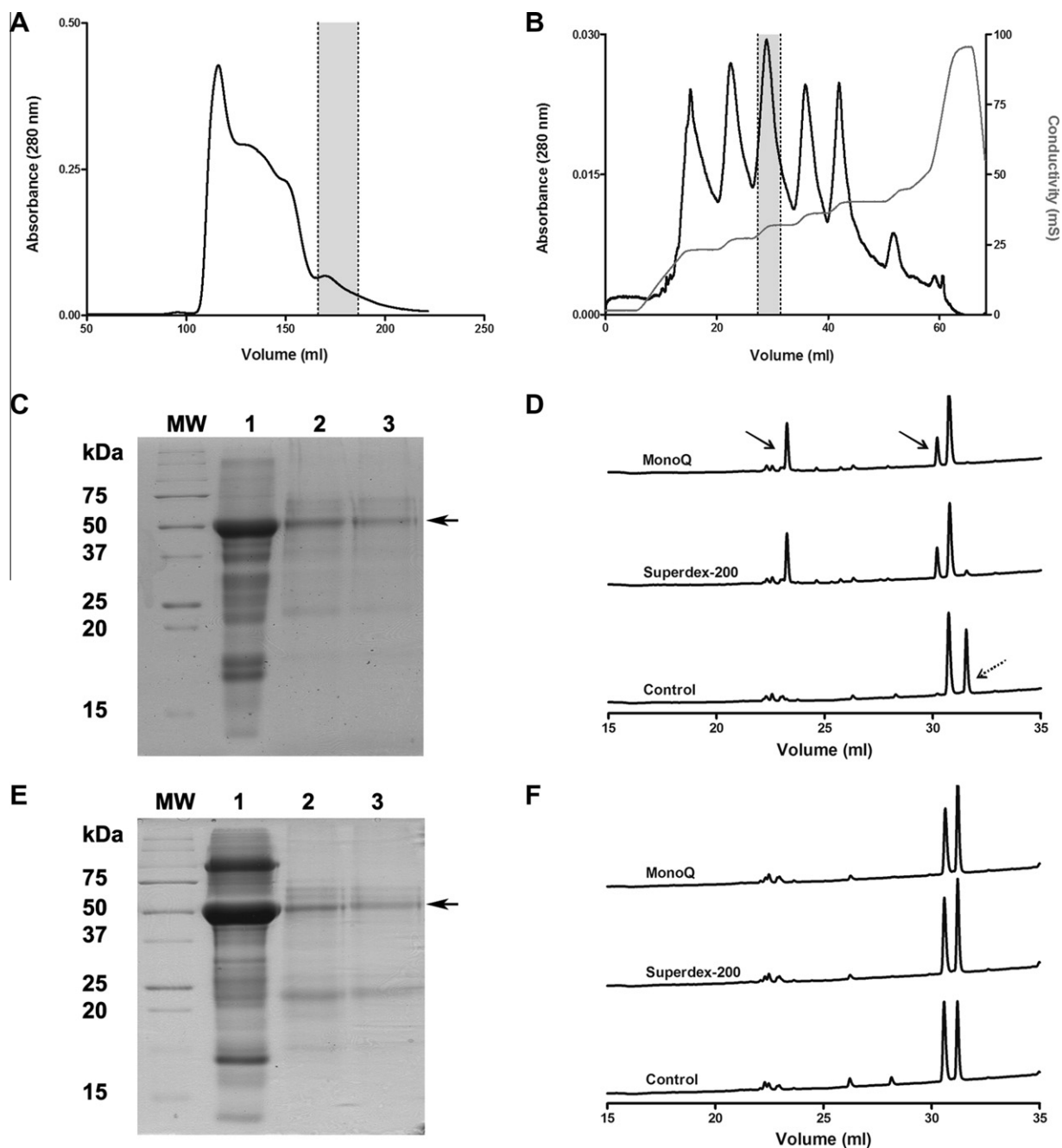
Fig. 5. Purification of recombinant procirsin. Wild-type procirsin and the active-site mutant procirsin (D103A) were expressed as inclusion bodies in *E. coli* and refolded by rapid dilution. The recombinant proteins were purified by size-exclusion chromatography followed by ion-exchange chromatography as described in Experimental Procedures. Fractions with activity towards the typical AP fluorogenic substrate are highlighted with dotted lines. (A) Size-exclusion chromatography of wt procirsin on a HiLoad 26/60 Superdex 200 column. (B) Superdex 200 (S-200) eluate was pooled (active fractions) and further purified by ion-exchange chromatography in a Mono Q column. (C) Collected fractions from the different steps of the wt procirsin purification procedure were analyzed by SDS–PAGE and stained with Coomassie Brilliant Blue: (Lane 1) Refolded recombinant wt procirsin applied to size-exclusion chromatography; (Lane 2) Superdex 200 eluate applied to Mono Q column (pooled fractions highlighted with dotted lines); (Lane 3) Mono Q eluate (pooled fractions highlighted with dotted lines). (D) RP-HPLC profiles of activity towards typical AP substrate. Eluates' proteolytic activity of wt procirsin (Superdex 200 and MonoQ pooled fractions) was assayed by digestion of the substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) and the resulting fragments were separated by RP-HPLC. Digestion products are marked with a solid arrow. The control corresponds to the peptide incubated in the absence of enzyme. (E) SDS–PAGE analysis of collected fractions from the different steps of the purification of the active-site mutant procirsin(D103A). The gel was stained with Coomassie Brilliant Blue: (Lane 1) Refolded recombinant procirsin(D103A) applied to size-exclusion chromatography; (Lane 2) Superdex 200 eluate applied to Mono Q column; (Lane 3) Mono Q eluate. (*F*) RP-HPLC elution profiles obtained in activity screening assays using recombinant procirsin(D103A) purified fractions. Eluates' proteolytic activity (Superdex 200 and MonoQ) was assayed by digestion of the typical AP substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) and the resulting fragments were separated by RP-HPLC. The control corresponds to the peptide incubated in the absence of enzyme.

to indicate that procirsin may undergo a different processing mechanism, most likely requiring the action of host cell proteinases to achieve the complete conversion towards the final active form of the enzyme. In fact, a similar mechanism was proposed

for BACE, also shown to be active in the precursor form but unable of autocatalytic activation (Ermolieff et al., 2000).

Expression of procirsin lacking the PSI domain led to an inactive protein, suggesting a critical role of this internal segment for
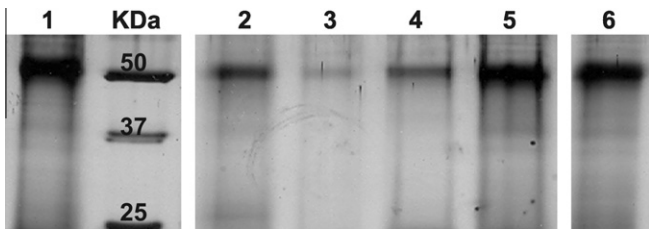
Fig. 6. Effect of pH on the proteolytic processing of recombinant wt procirsin. Purified recombinant wt procirsin was incubated overnight at 37 °C over a pH range of pH 3 to pH 7.0 (0.1 M sodium citrate pH 3 and pH 4; 0.1 M sodium acetate pH 5 and pH 6; 0.05 M Tris–HCl pH 7.0). Incubated samples were analyzed by SDS–PAGE and the gel was silver stained: Lane 1, recombinant wt procirsin (control without incubation); Lanes 2–6, recombinant wt procirsin incubated at pH 3, 4, 5, 6 and 7, respectively. Wt procirsin showed some autodegradation when incubated under acidic conditions but no defined pattern of conversion was observed.
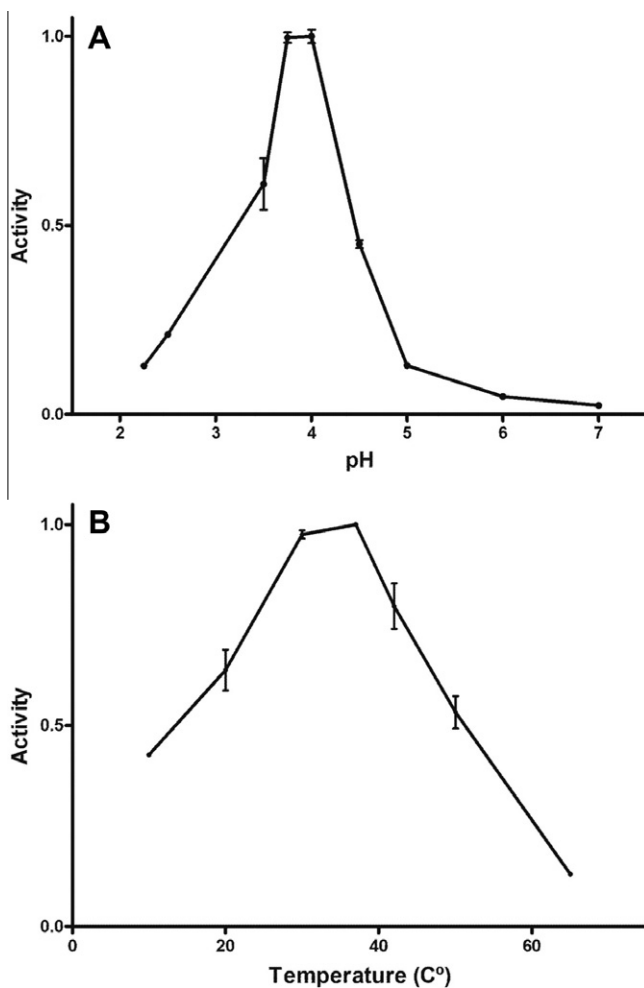
Table 1
Activity of recombinant wt procirsin after incubation with protease inhibitors.

Inhibitor	Concentration (mM)	Residual activity (%)[a]
Pepstatin A	0.001	2
EDTA	0.005	61
E-64	0.01	88
Pefabloc	1	47
Amastatin	0.01	80
Dithiothreitol	1	74

[a] The reactions were carried out with the synthetic substrate (MCA)Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) (192 μM) in 0.05 M sodium acetate buffer, pH 4.0 at 37 °C after 6 min of incubation at room temperature of the enzyme with each inhibitor.

between both enzymes. Again, this contrasts with the results described for the single-chain form of procardosin A without PSI that was shown to undergo autoactivation and to display similar properties when compared with the natural enzyme (Castanheira et al., 2005). All together, these results strengthen the idea that plant typical APs precursors assume diverse structural arrangements that most likely contribute to regulate enzyme activity, stressing the relevance of further dissecting their processing/activation mechanisms.

4. Conclusions

The cDNA encoding a typical aspartic proteinase precursor was isolated from *C. vulgare* flowers. The encoded protein product – cirsin – was produced in *E. coli* in the zymogen form. Recombinant procirsin was shown to be active towards the substrate typically used to study APs and to share the same enzymatic features of other typical APs: maximum activity at acidic pH, inhibition by pepstatin A and strict dependence on two catalytic Asp residues for activity. Recombinant procirsin also displayed milk-clotting activity and a high specificity towards κ-casein. Although active in the precursor form, recombinant procirsin was unable to trigger autoactivation under acidic conditions, suggesting that procirsin processing requires the action of other proteinases.

5. Experimental

5.1. Plant material

Developing flower buds of *C. vulgare* (Savi) Ten. were collected in fields of surrounding areas of La Plata, Argentina during November and kept at −80 °C until use. A voucher specimen (LPE 1160) has been deposited at the LPE herbarium (Faculty of Exact Sciences, National University of La Plata, Argentina).

5.2. cDNA cloning and sequence analysis

Total RNA was isolated from flower buds of *C. vulgare* using Invisorb Spin Plant RNA Mini kit (Invitek, Life Science) according to the manufacturers' instructions, eluted in RNase-free water and its quality was checked by agarose gel electrophoresis. A volume equivalent to 1.2 μg of total RNA was used in the synthesis of single-stranded cDNA with the First Strand cDNA Synthesis Kit for Reverse Transcription-PCR (Roche Applied Science) and oligo (dT)$_{15}$ primer. The resulting first strand cDNA was then used as template in PCR amplification with primers specifically designed for the highly conserved 5′ and 3′ ends of aspartic proteinases' coding sequences (forward primer: 5′-ATGGGTACCTCAATCAAAG-CA-3′; reverse primer: 5′-TCAAGCTGCTTCTGCAAATCC-3′). The amplified cDNA was cloned and both strands were sequenced by automated DNA sequencing. The obtained sequences were

Fig. 7. Effect of pH and temperature on the activity of recombinant wt procirsin. Purified recombinant wt procirsin was tested for activity using as substrate the synthetic fluorogenic peptide (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP). (A) Activity studies at different pH values were performed by incubating wt procirsin at 37 °C with buffers between pH 2.25 and pH 7.0 containing 0.1 M NaCl (0.05 M sodium citrate, pH 2.25–2.50; 0.05 M sodium acetate, pH 3.50–6.0; 0.05 M Tris–HCl, pH 7.0. (B) Activity studies at different temperatures. Purified recombinant wt procirsin was incubated in 0.05 M sodium acetate, pH 4.0 with 0.1 M NaCl buffer at temperatures between 10 °C and 65 °C.

cirsin's folding. Similar data was reported for the expression of PSI-deleted procyprosin in *Pichia pastoris* (White et al., 1999) being consistent with a high sequence/structure similarity relationship
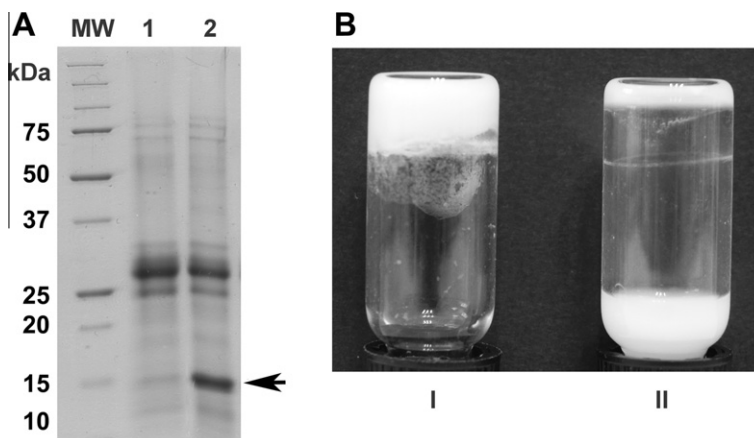
Fig. 8. SDS–PAGE analysis of bovine κ-casein hydrolysis and milk-clotting activity of purified recombinant wt procirsin. (A) κ-Casein in 0.05 M sodium acetate with 0.1 M NaCl buffer pH 5.5 was incubated with purified recombinant wt procirsin at 37 °C, overnight. Digestion fragments were analyzed by SDS–PAGE and the gel was stained with Coomassie Brilliant Blue: (Lane 1) κ-casein control incubation without enzyme; (Lane 2) κ-casein digestion profile in the presence of purified recombinant wt procirsin. Major digestion product is marked with an arrow. (B) Milk-clotting activity of recombinant wt procirsin was tested by incubation with skim milk reconstituted in 10 mM CaCl2, pH 6.50 (12% w/v) at 37 °C. Curd formation was evaluated by visual observation (*tube I*). Milk-clotting activity was inhibited by the presence of pepstatin A (*tube II*) with no visible curd formation.

subjected to multiple sequence alignment with Basic Local Alignment from the National Center for Biotechnology Information (BLAST) and conserved domains (CD) were identified using NCBI's CD-Search service (Marchler-Bauer et al., 2011), confirming the nature of the cDNA-deduced amino acid sequences as AP homologues. One of these sequences was selected for further characterization (termed cirsin) and deposited in GenBank under the accession number JN703462.

5.3. Expression in E. coli, refolding and purification of recombinant cirsin

cDNA encoding procirsin without the putative signal peptide (first 25 amino acids residues) was amplified with specific primers that include restriction sites for NheI and XhoI (forward primer: 5′-gctagcTCCAATGATGGATTAATTAGAG-3′; reverse primer: 5′-ctcga gTCAAGCTGCTTCTGCAAATCC-3′). The resulting product of approximately 1500 bp was cloned into pGEM-T Easy vector (Promega) and plasmid DNA from positive clones was analyzed by NheI/XhoI double-digestion. The purified digestion product (Invirsob Spin DNA Extraction Kit, Invitek) was then subcloned into pET23-d expression vector (Novagen) to obtain wild-type procirsin_pET-23d construct (hereby named procirsin). The positive clones selected by restriction analysis were confirmed by automated DNA sequencing.

Quick Change Site-Directed Mutagenesis kit (Invitrogen) was used to generate procirsin active-site mutant (D103A) in pET23-d (procirsin(D103A)-pET-23d construct) using the primers 5′-GTTCAC TGTGATTTTTGCGACC**GGATCC**TCTAATCTATGGGTG-3′ (forward primer) and 5′-CACCCATAGATTAGA**GGATCC**GGTCGCAAAAATCACAGT GAAC-3′ (reverse primer) where mutation is underlined with a solid line; a restriction site for BamHI was introduced by a silent mutation (bold underline) in order to select positive mutant clones prior confirmation by DNA sequencing.

A third construct was generated consisting of procirsin without the PSI domain (procirsinΔPSI_pET-23d): the construct procirsin-pET-23d was used as template in a deletion mutagenesis using the primers 5′-CCCATGCCACCGCCACCTCCCTTAGCACCAATTGCAT G-3′ (forward primer) and 5′-CTAAGGGAGGTGGCGGTGGCATGG GAGAATCAGCAGTAGATTG-3′ (reverse primer). Positive clones were confirmed by automated DNA sequencing.

Escherichia coli BL21 star (DE3) strain (Novagen) was transformed with each plasmid construct, wt procirsin, procirsin(D103A) and

procirsinΔPSI. A single-colony transformant was inoculated into 20 ml Luria Bertani (LB) medium containing 100 μg/ml ampicilin and grown overnight at 37 °C. The cultures were then transferred to 1 L of fresh LB medium with 30 μg/ml ampicilin and allowed to grow at 37 °C until an $OD_{600 nm}$ of 0.7. Isopropyl-beta-D-thiogalacto-pyranoside (IPTG) was then added to a final concentration of 0.5 mM to induce protein expression at 37 °C. After 3 h, cells were harvested by centrifugation at 9000*g* for 15 min and resuspended in 50 mM Tris–HCl, pH 7.5, containing 50 mM NaCl. Lysozyme (7000 U/ml) (Sigma) was added to the suspension, and a freeze/thawing cycle was then performed for further cell disruption. The lysate was subsequently incubated with DNase (12 μg/ml, Sigma) and $MgCl_2$ (20 mM) for 3 h on ice. The cell lysate was then submitted to two sequential washing steps: the lysate was diluted into 1 L of 50 mM Tris–HCl, pH 7.5, containing 50 mM NaCl and kept overnight at 4 °C under agitation, this material was then centrifuged at 10,000*g* and washed again overnight with 1 L of 50 mM Tris–HCl, pH 7.5, 50 mM NaCl containing 0.1% (v/v) Triton X-100.

Inclusion bodies were recovered by centrifugation at 9000 *g*, 4 °C for 20 min and finally denatured in 0.1 M Tris, 1 mM glycine buffer (pH 10.5) containing 8 M urea, 1 mM EDTA, and 100 mM β-mercaptoethanol at 4 °C during 24 h. Denatured inclusion bodies were ultracentrifuged at 10,000 rpm, 20 min and 4 °C and the resulting supernatant was transferred to 0.1 M Tris, 1 mM glycine buffer (pH 10.5) containing 8 M urea, 1 mM EDTA, 10 mM DTT, 1 mM reduced glutathione, 0.1 mM oxidized glutathione and 100 mM β-mercaptoethanol. Afterwards, protein refolding was carried out by rapid dilution (20-fold) into 20 mM Tris–HCl, and the pH was adjusted to pH 8.0. The recombinant proteins were concentrated 42 days after refolding in a tangential flow ultrafiltration system (Pellicon 2, Millipore) and ultracentrifuged at 144,028*g* for 20 min at 4 °C to remove high molecular weight aggregates and insoluble impurities. The supernatant was applied to a HiLoad 26/60 Superdex 200 gel filtration column (GE Healthcare Life Sciences) connected to an FPLC system (DuoFlow-BioRad), equilibrated in 20 mM Tris–HCl buffer (pH 8.0) containing 0.4 M urea at a flow rate of 2.0 ml/min. Fractions were collected and proteolytic activity was screened towards the typical AP substrate (7-methoxycoumarin-4-acetic acid) (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys-(2,4-dinitrophenyl) (DNP), at pH 4.0. Substrate cleavage was monitored by reverse phase high-performance liquid chromatography (RP-HPLC), as described below.

The active fractions corresponding to non-aggregated forms of the recombinant wt procirsin were pooled and further purified by ion-exchange chromatography with a MonoQ (GE Healthcare Life Sciences) column connected to a FPLC system (Äkta FPLC, GE Healthcare Life Sciences) using the same buffer as for size-exclusion chromatography. Elution was carried out with a linear gradient of NaCl (0–0.5 M) at a flow rate of 0.75 ml/min. Protein concentration was estimated by quantification at 280 nm using NanoDrop (Thermo Scientific).

Recombinant procirsin(D103A) and procirsinΔPSI were purified according to a similar procedure and equivalent fractions were pooled and assayed for proteolytic activity.

5.4. Polyacrylamide gel electrophoresis

Samples collected along expression and purification of all procirsin constructs were analyzed by SDS–PAGE using 12% polyacrylamide gels in a Bio Rad Mini-Protean 3 Cell (Bio Rad). Samples were treated with loading buffer (0.0625 M-Tris–0.1 M-HCl pH 6.8, 20 g SDS/l, 0.02 g bromophenol blue/l and 100 g glycerol/l) with 5% β-mercaptoethanol at 95 °C for 10 min before loading. Gels were stained with Coomassie Brilliant Blue R-250 (Sigma).

5.5. Proteolytic activity screening assays

Recombinant wt procirsin, procirsin(D103A) and procirsinΔPSI proteolytic activity was screened throughout purification by incubation with the typical AP substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) using a protease:substrate volume ratio of 1:10 in 50 mM sodium acetate buffer (pH 4.0) with 0.1 M NaCl, overnight, at 37 °C. Substrate cleavage was monitored by RP-HPLC on a KROMASIL 100 C18 250, 4.6 mm column, using a Prominence system (Shimadzu Corporation, Tokyo, Japan). Elution was performed with a linear gradient (0–80%) of acetonitrile in 0.1% (v/v) trifluoroacetic acid at a flow rate of 1 ml/min for 50 min, monitoring the absorbance at 220 nm.

5.6. Enzyme activity assays

The proteolytic activity of purified recombinant wt procirsin was tested towards several model substrates at pH 4.0, 37 °C and the rate of hydrolysis was followed by the increase in fluorescence intensity (excitation/emission wavelengths of 328/393 nm for MCA/DNP fluorogenic peptides and excitation/emission wavelengths of 335/490 nm for EDANS/DABCYL fluorogenic peptides) using a Horiba Jobin Yvon FluoroMax-3 Spectrofluorometer. The fluorogenic subtrates assayed were: 2 μM typical AP substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) in 50 mM sodium acetate buffer with 0.1 M NaCl; 2 μM CDR1 substrate (Simões et al., 2007) (MCA)Lys-Leu-His-Pro-Glu-Val-Leu-Phe-Val-Leu-Glu-Lys(DNP) and 2 μM atypical AP substrate 1 (MCA)Lys-Lys-Leu-Ala-Asp-Val-Val-Asn-Ala-Leu-Glu-Lys-Lys(DNP); 2 μM BACE 1 substrate (MCA)Lys-Ser-Glu-Val-Asn-Leu-Asp-Ala-Glu-Phe-Lys(DNP) in 50 mM sodium acetate buffer with 0.1 M NaCl and 8% (v/v) Me$_2$SO; 2 μM HIV protease substrate 1 Arg-Glu(EDANS)-Ser-Gln-Asn-Tyr-Pro-IleVal-Gln-Lys(DABCYL)-Arg and 2 μM Renin substrate 1 Arg-Glu(EDANS)-Ile-His-Pro-Phe-His-Leu-Val-Ile-His-Thr-Lys(DABCYL)-Arg in 50 mM sodium acetate buffer, 0.1 M NaCl with 9.6 % (v/v) Me$_2$SO.

Since purified recombinant procirsin effectively cleaved only the typical AP substrate, the enzyme was further characterized towards this substrate. The dependence of activity with the pH was assayed between pH 2.25 and pH 7.0 at 37 °C in the following buffers: 0.05 M sodium citrate, pH 2.25–2.5; 0.05 M sodium acetate, pH 3.5–6.0; 0.05 M Tris–HCl, pH 7.0 containing 0.1 M NaCl. The effect of temperature on recombinant procirsin proteolytic activity was measured by pre-incubation of the enzyme in 0.1 M NaCl, 50 mM sodium acetate, pH 4.0 at temperatures ranging from 10 °C to 65 °C during 10 min. To determine the effect of a set of inhibitors, recombinant procirsin (2.1 μg) was pre-incubated with each of the following compound, pepstatin A (0.001 mM), EDTA (0.005 mM), pefabloc (1 mM), E-64 (0.01 mM), amastatin (0.01 mM) and dithiothreitol (1 mM) for 6 min at room temperature in 50 mM sodium acetate buffer (pH 4.0) with 0.1 M NaCl before determination of proteolytic activity. Kinetics studies were performed at 37 °C using concentrations between 0.5 μM and 5 μM of the substrate (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP) in 50 mM sodium acetate buffer pH 4.0 with 0.1 M NaCl. The enzyme concentration was determined by active-site titration of the purified recombinant procirsin with pepstatin A. The relationship between fluorescence change and peptide concentration was determined by measuring the total fluorescence change that occurred upon complete hydrolysis of the peptide with pepsin. The kinetic parameters were calculated from the Lineweaver–Burk plot using appropriate software.

5.7. Activation assays

The proteolytic processing of purified recombinant wt procirsin was evaluated by incubation with activation buffers in a ratio of 1:1 (v/v). The samples were incubated overnight at 37 °C with the buffers: 0.1 M sodium citrate pH 3 and pH 4; 0.1 M sodium acetate pH 5 and pH 6; 0.05 M Tris–HCl pH 7.0. Incubated samples were then analyzed by SDS–PAGE and assayed for activity towards (MCA)Lys-Lys-Pro-Ala-Glu-Phe-Phe-Ala-Leu-Lys(DNP).

5.8. Specificity against oxidized insulin -chain

Oxidized insulin β-chain (1 mg/ml) was incubated with purified recombinant procirsin (substrate: enzyme mass ratio of 100:1) in 0.1 mM sodium acetate buffer (pH 4.0). After an overnight incubation at 37 °C the reaction mixture was centrifuged at 20,000g during 6 min and the digestion fragments were separated by RP-HPLC on a C18 column (KROMASIL 100 C18 250, 4.6 mm), using a Prominence system (Shimadzu Corporation, Tokyo, Japan). Elution was carried out with a linear gradient (0–80%) of acetonitrile in 0.1% v/v trifluoroacetic acid for 30 min at a flow rate of 1 ml/min. Absorbance was monitored at 220 nm. Two major isolated peptides were collected, freeze-dried, and submitted to identification with a 4000 QTRAPsystem (Proteomics Unit of the Center for Neuroscience and Cell Biology, University of Coimbra, Portugal).

5.9. Digestion of κ-casein

κ-Casein (0.3 mg/ml) was incubated with purified recombinant wt procirsin in 50 mM sodium acetate buffer (pH 5.5) at a substrate: enzyme ratio of 250:1 (w/w). The reaction was allowed to proceed at 37 °C for 18 h and then analyzed by SDS–PAGE. κ-Casein incubated under similar conditions in the absence of enzyme was used as a negative control.

5.10. Milk-clotting activity

Milk-clotting activity was determined by adding purified recombinant procirsin to skim milk powder reconstituted (12% w/v) in 10 mM CaCl$_2$, pH 6.5. An enzyme:milk ratio of 1:5 (v/v) was used and the mixture was then incubated at 37 °C until curd formation was observed. Controls without enzyme and with enzyme pre-incubated 10 min with pepstatin A (0.001 mM) were also performed.

5.11. PMF by MALDI-TOF MS

Purified recombinant procirsin(D103A) was analyzed by peptide mass fingerprinting. *In situ* tryptic digestion of an electrophoretically homogeneous band was performed following the protocol of (Obregón et al., 2009). Peptides were dissolved in 5 μl 0.1% TFA (v/v) and analyzed by MALDI-TOF MS using a matrix of acid α-cyano-4-hydroxycinnamic (HCCA). Peptide masses were acquired with Flex Control Software in a range of ca. 1000–3500 m/z. External calibration was performed using a peptide calibrants mixture composed of bradykinin [1–7 (757.39916)], angiotensin II (1046.5418), angiotensin (1296.6848), substance P (1347.7354), bombesin (1619.8223), renin substrate (1758.93261), adrenocorticotropic hormone: ACTH clip [1–17 (2093.0862)], adrenocorticotropic hormone: ACTH clip [18–39 (2465.1983)] and somatostatin 28 (3147.4710). The results were processed using the MASCOT search engine (http://www.matrixscience.com). Search parameters were (1) type of search, peptide mass fingerprint; (2) enzyme, trypsin; (3) database, SwissProt 55.2; (4) taxonomy, Viridiplantae; (5) Variable modifications, carbamidomethyl (C), oxidation (M); (6) mass values, monoisotopic; (7) peptide mass tolerance: ±100 ppm; (8) peptide charge state, 1+. Additionally, the sequence deduced from mutant cDNA was subjected to theoretical tryptic digestion by means of SequenceEditor 3.1 software (Bruker Daltonics, Biotools 3.1) and compared with empirical PMF of recombinant procirsin (D103).

5.12. Bioinformatic characterization of procirsin

Sequence similarity searches were performed with BLAST to obtain close homologous proteins to procirsin. We retrieved all the sequences with E-values below 1×10^{-150} (120 proteins). These sequences were aligned with ClustalX and a phylogenetic estimation were obtained using a maximum likelihood approach using Phyml (Guindon et al., 2009). This estimation was done using JTT + F model of protein evolution with a discrete gamma distribution. Branching support was obtained using non-parametric bootstrapping with 500 replicants followed by a majority rule consensus tree.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.phytochem.2012.05.028.

References

Asakura, T., Matsumoto, I., Funaki, J., Arai, S., Abe, K., 2000. The plant aspartic proteinase-specific polypeptide insert is not directly related to the activity of oryzasin 1. FEBS J. 267, 5115–5122.

Athauda, S.B.P., Matsumoto, K., Rajapakshe, S., Kuribayashi, M., Kubomura-Yoshida, N., Kojima, M., Iwamatsu, A., Shibata, C., Inoue, H., Takahashi, K., 2004. Enzymic

and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases. Biochem. J. 306, 295–306.

Bleukx, W., Brijs, K., Torrekens, S., Van Leuven, F., Delcour, J.A., 1998. Specificity of a wheat gluten aspartic proteinase. BBA 1387, 317–324.

Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., Brunak, S., 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 4, 1633–1649.

Brodelius, M., Hiraiwa, M., Martilla, S., Al Karadaghi, S., Picaud, S., Brodelius, P., 2005. Immunolocalization of 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.

Brutti, C.B., Pardo, M.F., Caffini, N.O., Natalucci, C.L., 2012. *Onopordum acanthium* L. (Asteraceae) flowers as coagulating agent for cheesemaking. LWT – Food Sci. Technol. 45, 172–179.

Bryksa, B.C., Bhaumik, P., Magracheva, E., De Moura, D.C., Kurylowicz, M., Zdanov, A., Dutcher, J.R., Wlodawer, A., Yada, R.Y., 2011. Structure and mechanism of the saposin-like domain of a plant aspartic protease. J. Biol. Chem. 286, 28265–28275.

Castanheira, P., Samyn, B., Sergeant, K., Clemente, J.C., Dunn, B.M., Pires, E., Van Beeumen, J., Faro, C., 2005. Activation, proteolytic processing, and peptide specificity of recombinant cardosin A. J. Biol. Chem. 280, 13047–13054.

Cordeiro, M.C., Xue, Z.T., Pietrzak, M., Pais, M.S., Brodelius, P.E., 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.

Domingos, A., Cardoso, P.C., Xue, Z.-tian, Clemente, A., Brodelius, P.E., Pais, M.S., 2000. Purification, cloning and autoproteolytic processing of an aspartic proteinase from *Centaurea calcitrapa*. FEBS J. 267, 6824–6831.

Duarte, P., Pissarra, J., Moore, I., 2008. Processing and trafficking of a single isoform of the aspartic proteinase cardosin A on the vacuolar pathway. Planta 227, 1255–1268.

Dunn, B.M., 2002. Structure and mechanism of the pepsin-like family of aspartic peptidases. Chem. Rev. 102, 4431–4458.

Dunn, B.M., Hung, S.H., 2000. The two sides of enzyme-substrate specificity: Lessons from the aspartic proteinases. BBA – Protein Struct. Mol. Enzymol. 1477, 231–240.

EEC, 1996. Foodstuffs. Protected geographical indications and designations of origin. E.E.C. Official Journal L 163, 1263.

Egas, C., Lavoura, N., Resende, R., Brito, R.M.M., Pires, E., Pedroso de Lima, M.C., Faro, C., 2000. The saposin-like domain of the plant aspartic proteinase precursor is a potent inducer of vesicle leakage. J. Biol. Chem. 275, 38190–38196.

Ermolieff, J., Loy, J.A., Koelsch, G., Tang, J., 2000. Proteolytic activation of recombinant pro-memapsin 2 (Pro-beta-secretase) studied with new fluorogenic substrates. Biochemistry 39, 16263.

Faro, C., Gal, S., 2005. Aspartic proteinase content of the arabidopsis genome. Curr. Protein Pept. Sci. 6, 493–500.

Feijoo-Siota, L., Villa, T.G., 2011. Native and biotechnologically engineered plant proteases with industrial applications. Food Bioprocess Technol. 4, 1066–1088.

Glathe, S., Kervinen, J., Nimtz, M., Li, G.H., Tobin, G.J., Copeland, T.D., Ashford, D.A., Wlodawer, A., Costa, J., 1998. Transport and activation of the vacuolar aspartic proteinase phytepsin in barley (*Hordeum vulgare* L.). J. Biol. Chem. 273, 31230–31236.

González-Rábade, N., Badillo-Corona, J.A., Aranda-Barradas, J.S., Oliver-Salvador, M.D.C., 2011. Production of plant proteases *in vivo* and *in vitro* – A review. Biotechnol. Adv. 29, 983–996.

Guevara, M.G., Veríssimo, P., Pires, E., Faro, C., Daleo, G.R., 2004. Potato aspartic proteases: Induction, antimicrobial activity and substrate specificity. J. Plant Pathol. 86, 233–238.

Guindon, S., Delsuc, F., Dufayard, J.F., Gascuel, O., 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol. Biol. 537, 113–137.

van der Hoorn, R.A.L., 2008. Plant proteases: From phenotypes to molecular mechanisms. Annu. Rev. Plant Biol. 59, 191–223.

Hyslop, D., 2003. Enzymatic coagulation of milk. In: Fox, P.F., Mcsweeney, P.L.H. (Eds.), Advanced Dairy Chemistry. Volumen 1: Proteins. Kluwer Academic/Plenum Publishers, Berlin, pp. 839–878.

Jacob, M., Jaros, D., Rohm, H., 2011. Recent advances in milk clotting enzymes. Int. J. Dairy Technol. 64, 14–33.

Kervinen, J., Sarkkinen, P., Kalkkinen, N., Mikola, L., Saarma, M., 1993. Hydrolytic specificity of the barley grain aspartic proteinase. Phytochemistry 32, 799–803.

Kervinen, J., Tobin, G.J., Costa, J., Waugh, D.S., Wlodawer, A., Zdanov, A., 1999. Crystal structure of plant aspartic proteinase prophytepsin: Inactivation and vacuolar targeting. EMBO J. 18, 3947–3955.

Llorente, B.E., Brutti, C.B., Caffini, N.O., 2004. Purification and characterization of a milk-clotting aspartic proteinase from globe artichoke (*Cynara scolymus* L.). J. Agric. Food Chem. 52, 8182–8189.

Marchler-Bauer, A., Lu, S., Anderson, J., Chitsaz, F., Derbyshire, M., DeWeese-Scott, C., Fong, J., Geer, L., Geer, R., Gonzales, N., Gwadz, M., Hurwitz, D., Jackson, J., Ke, Z., Lanczycki, C., Lu, F., GH, M., Mullokandov, M., Omelchenko, M., Robertson, C., Song, J., Thanki, N., Yamashita, R., Zhang, D., Zhang, N., Zheng, C., Bryant, S., 2011. CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, 225–229.

Mazorra-Manzano, M.A., Yada, R.Y., 2008. Expression and characterization of the recombinant aspartic proteinase A1 from *Arabidopsis thaliana*. Phytochemistry 69, 2439–2448.

Obregón, W.D., Liggieri, C.S., Trejo, S.A., Avilés, F.X., Vairo-Cavalli, S.E., Priolo, N.S., 2009. Characterization of papain-like isoenzymes from latex of *Asclepias*

curassavica by molecular biology validated by proteomic approach. Biochimie 91, 1457–1464.

Park, H., Yamanaka, N., Mikkonen, A., Kusakabe, I., Kobayashi, H., 2000. Purification and characterization of aspartic proteinase from sunflower seeds. Biosci. Biotechnol. Biochem. 64, 931–939.

Payie, K.G., Tanaka, T., Gal, S., Yada, R.Y., 2003. Construction, expression and characterization of a chimaeric mammalian-plant aspartic proteinase. Biochem. J. 372, 671–678.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786.

Pimentel, C., Van Der Straeten, D., Pires, E., Faro, C., Rodrigues-Pousada, C., 2007. Characterization and expression analysis of the aspartic protease gene family of *Cynara cardunculus* L. FEBS J. 274, 2523–2539.

Pitarresi, T.M., Rubattu, S., Heinrikson, R., Sealey, J.E., 1992. Reversible cryoactivation of recombinant human prorenin. J. Biol. Chem. 267, 11753–11759.

Ramalho-Santos, M., Veríssimo, P., Cortes, L., Samyn, B., Van Beeumen, J., Pires, E., Faro, C., 1998. Identification and proteolytic processing of procardosin A. FEBS J. 255, 133–138.

Rawlings, N.D., Barrett, A.J., Bateman, A., 2010. MEROPS: The peptidase database. Nucleic Acids Res. 38, D227–D233.

Rawlings, N.D., Bateman, A., 2009. Pepsin homologues in bacteria. BMC Genomics 10, 437.

Reis, P.J.M., Malcata, F.X., 2011. Current state of Portuguese dairy products from ovine and caprine milks. Small Ruminant Res. 101, 122–133.

Roseiro, L.B., Barbosa, M., Ames, J.M., Wilbey, R.A., 2003. Cheesemaking with vegetable coagulants-the use of *Cynara* L. for the production of ovine milk cheeses. Int. J. Dairy Technol. 56, 76–85.

Sampaio, P.N., Fortes, A.M., Cabral, J.M.S., Pais, M.S., Fonseca, L.P., 2008. Production and characterization of recombinant cyprosin B in *Saccharomyces cerevisiae* (W303-1A) strain. J. Biosci. Bioeng. 105, 305–312.

Shi, X.-ping, Chen, E., Yin, K.-chang, Na, S., Garsky, V.M., Lai, M.-tain, Li, Y.-ming, Platchek, M., Register, R.B., Sardana, M.K., Tang, M.-J., Thiebeau, J., Wood, T., Shafer, Gardell, S.J., 2001. The pro domain of β-secretase does not confer strict zymogen-like properties but does assist proper folding of the protease domain. Biochemistry 276, 10366–10373.

Silva, S.V., Malcata, F.X., 2005. Food chemistry studies pertaining to coagulant and proteolytic activities of plant proteases from *Cynara cardunculus*. Food Chem. 89, 19–26.

Simões, I., Faro, C., 2004. Structure and function of plant aspartic proteinases. FEBS J. 271, 2067–2075.

Simões, I., Faro, R., Bur, D., Faro, C., 2007. Characterization of recombinant CDR1, an Arabidopsis aspartic proteinase involved in disease resistance. J. Biol. Chem. 282, 31358–31365.

Simões, I., Faro, R., Bur, D., Kay, J., Faro, C., 2011. Shewasin A, an active pepsin homolog from the bacterium *Shewanella amazonensis*. FEBS J. 278, 3177–3186.

Tamer, M., Mavituna, F., 1996. Protease from callus and cell suspension cultures of *Onopordum turcicum* (Compositae). Biotechnol. Lett. 18, 361–366.

Terauchi, K., Asakura, T., Ueda, H., Tamura, T., Tamura, K., Matsumoto, I., Misaka, T., Hara-Nishimura, I., Abe, K., 2006. Plant-specific insertions in the soybean aspartic proteinases, soyAP1 and soyAP2, perform different functions of vacuolar targeting. J. Plant Physiol. 163, 856–862.

Timotijević, G.S., Radović, S.R., Maksimović, V.R., 2006. Aspartic proteinases from buckweath (*Fagopyrum esculetum* Moench) seeds – Purification and properties of the 47 kDa enzyme. Arch. Biol. Sci. 58, 171–177.

Tormakangas, K., 2001. A vacuolar sorting domain may also influence the way in which proteins leave the endoplasmic reticulum. Plant Cell 13, 2021–2032.

Vairo Cavalli, S., Claver, S., Priolo, N., Natalucci, C., 2005. Extraction and partial characterization of a coagulant preparation from *Silybum marianum* flowers. Its action on bovine caseinate. J. Dairy Res. 72, 271–275.

Vieira, M., Pissarra, J., Veríssimo, 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.

White, P.C., Cordeiro, M.C., Arnold, D., Brodelius, P.E., Kay, J., 1999. Processing, activity, and inhibition of recombinant cyprosin, an aspartic proteinase from cardoon (*Cynara cardunculus*). J. Biol. Chem. 274, 16685–16693.