Planta (2009) 230:319–328
DOI 10.1007/s00425-009-0942-2

ORIGINAL ARTICLE

Sequencing and characterization of asclepain f: the first cysteine peptidase cDNA cloned and expressed from *Asclepias fruticosa* latex

Sebastián A. Trejo · Laura M. I. López · Néstor O. Caffini · Claudia L. Natalucci · Francesc Canals · Francesc X. Avilés

Received: 3 February 2009 / Accepted: 29 April 2009 / Published online: 20 May 2009
© Springer-Verlag 2009

Abstract Asclepain f is a papain-like protease previously isolated and characterized from latex of *Asclepias fruticosa*. This enzyme is a member of the C1 family of cysteine proteases that are synthesized as preproenzymes. The enzyme belongs to the alpha + beta class of proteins, with two disulfide bridges (Cys22-Cys63 and Cys56-Cys95) in the alpha domain, and another one (Cys150-Cys201) in the beta domain, as was determined by molecular modeling. A full-length 1,152 bp cDNA was cloned by RT-RACE-PCR from latex mRNA. The sequence was predicted as an open reading frame of 340 amino acid residues, of which 16 residues belong to the signal peptide, 113 to the propeptide and 211 to the mature enzyme. The full-length cDNA was ligated to pPICZα vector and expressed in *Pichia pastoris.* Recombinant asclepain f showed endopeptidase activity on pGlu-Phe-Leu-*p*-nitroanilide and was identified by PMF-MALDI-TOF MS. Asclepain f is the first peptidase cloned and expressed from mRNA isolated from plant latex, confirming the presence of the preprocysteine peptidase in the latex.

S. A. Trejo · L. M. I. López · N. O. Caffini · C. L. Natalucci (✉)
Laboratorio de Investigación de Proteínas Vegetales, Departamento de Ciencias Biológicas, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, C.C. 711, B1900AVW La Plata, Argentina
e-mail: natalucci@biol.unlp.edu.ar; clnatalucci@yahoo.com.ar
URL: http://www.biol.unlp.edu.ar/liprove

S. A. Trejo · F. Canals · F. X. Avilés
Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Campus Universitari, Bellaterra, Cerdanyola del Vallès, 08193 Barcelona, Spain

F. Canals
Medical Oncology Research Program, Proteomics Laboratory, Institut de Recerca Hospital Universitari Vall d'Hebron, Psg. Vall d'Hebron, 119-129, 08035 Barcelona, Spain

Keywords *Asclepias fruticosa* · *Gomphocarpus fruticosus* subsp. *fruticosus* · Plant latex · Cysteine endopeptidase · Cloning · Overexpression in *Pichia pastoris*

Abbreviations

ACN	Acetonitrile
AfCP	Amino acid sequence corresponding to the cysteine protease asclepain f deduced from cDNA
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
PFLNA	pGlu-Phe-Leu-*p*-nitroanilide
PMF	Peptide mass fingerprinting
RACE	Rapid amplification of cDNA ends
rAfCP	Recombinant asclepain f cysteine protease
RMSD	Root mean square deviation
TFA	Trifluoroacetic acid

Introduction

The enzymes known as cysteine proteases, or cysteine peptidases, are often thought of all being related, and sometimes are even referred to the "cysteine protease family", but actually they evolved from a number of quite separate origins (Barrett and Rawlings 2001). This kind of hydrolases has been found in viruses, bacteria, protozoa, fungi, plants, and mammals (Barrett and Rawlings 2004).

Cysteine peptidases can be classified into 58 families (MEROPS release 8.1), which can be grouped into nine clans (CA, CD, CE, CF, CH, CL, CM, CN and CO) and three clans of mixed (C, S, T) catalytic type (PA, PB, and PC), according to a recent classification (Rawlings et al.

2008). The C1A subfamily includes papain and other related plant proteinases, such as chymopapain, caricain, bromelain, actinidin, ficain, and aleurain (Barrett and Rawlings 2004). Most of these enzymes are relatively small proteins with Mr values in the range 20,000–35,000 (Turk et al. 1997).

The enzymes belonging to the C1A subfamily that enter the secretory pathway, intended either for secretion or for the lysosomal (animal) or vacuolar (plant) system, are synthesized as precursors, with N-terminal propeptides as well as the signal peptides. Most of them have propeptides similar to that of papain, but the propeptide of cathepsin B is much shorter and very different in sequence. However, the propeptides of both cathepsin B and papain are thought to act in the same way in the proenzymes, blocking the active site by being bound in the reverse orientation of a substrate (Barrett and Rawlings 2004). The propeptides contain some characteristic elements which are highly conserved in evolution, as the ERFNIN motif, present in the α2 helix of a great number of cysteine protease propeptides belonging to the cathepsin L group (Karrer et al. 1993). Another highly conserved motif is the GXNXFXD heptapeptide, which can also be found in most of the cysteine propeptides, located at the kink of the β-sheet (Wiederanders 2003).

Plants are equipped with large proteolytic machinery that irreversibly regulates the fate of proteins. This machinery has generally been viewed in a housekeeping role, serving to remove non-functional proteins and to release amino acids for recycling. However, proteases also appear to play key roles in the regulation of biological processes in plants, such as the recognition of pathogens and pests and the induction of effective defense responses. Some proteases, such as papain in latex, execute the attack on the invading organism. Other proteases seem to be part of a signaling cascade, as indicated by protease inhibitor studies (Van der Hoorn and Jones 2004).

In the present paper, we describe the cloning and expression of asclepain f, a cysteine protease from *Asclepias fruticosa* (Trejo et al. 2001), which is, as far as we know, the first peptidase cloned and expressed from mRNA isolated from plant latex.

Materials and methods

Plant material

Latex was obtained by superficial incisions of fruits (follicles) of *Asclepias fruticosa* L.—*Apocynaceae*—[syn: *Gomphocarpus fruticosus* (L.) W.T. Aiton subsp. *fruticosus* (L.) W.T. Aiton]. The species is an erect perennial plant, 1–3 m in height; flowers (cream-white with green-purple corona) borne in simple umbels; fruits are inflated bristly green follicles 5–7.5 cm long, tapering to a beak, containing numerous black seeds bearing long silky hairs. The species is native to Africa. Material used in the present work was collected in Villa Elisa, Buenos Aires Province, Argentina.

Chemicals

Bovine serum albumin was purchased from Sigma-Aldrich Inc. (St. Louis, MO, USA). pGlu-Phe-Leu-*p*-nitroanilide (PFLNA) was a product of Bachem AG (Torrance, CA, USA). Coomassie Brilliant Blue R-250, acrylamide, and bisacrylamide were obtained from Bio-Rad (Hercules, CA, USA). EcoTaq DNA polymerase was obtained from ECOGEN Barcelona (Barcelona, Spain). Lys-C protease was purchased from Boehringer Ingelheim GmbH Corporate (Ingelheim, Germany). Trypsin Gold and pGEM-T Easy vector were products from Promega Corporation (Madison, WI, USA). *Bst*XI, *Not*I, and *Xho*I restriction enzymes were obtained from Roche (Roche Diagnostics-Roche Applied Science, Indianapolis, IN, U.S.A.). pPICZα expression vector, *Pichia pastoris* KM71H and GS115, and *Escherichia coli* TOP 10F′ cells were purchased from Invitrogen Corporation (Carlsbad, CA, USA). Peptide calibrants and α-cyano-4-hydroxycinnamic acid were purchased from Bruker Daltonics (Bremen, Germany). *Escherichia coli* XL1-Blue was purchased from Stratagene (La Jolla, CA, USA). Pre-stained broad molecular mass standard was purchased from Invitrogen Corporation (Carlsbad, CA, USA). Other reagents used were of the highest grade available.

Protein determination

Protein content was determined by measuring the absorbance at 280 nm of the sample solution, according to the extinction coefficient of asclepain f, or by the Bradford's method (1976) using bovine serum albumin as standard.

Determination of enzyme activity

pGlu-Phe-Leu-*p*-nitroanilide (PFLNA) was the substrate used to determine enzyme activity (Filippova et al. 1984). The reaction mixture contained 1.75 ml of 0.1 M phosphate buffer (pH 6.5) containing 0.1 M KCl, 0.1 mM EDTA and 3 mM DTT, 0.20 ml of substrate solution (2 mM PFLNA in DMSO) and 50 µl of enzyme extract. The reaction rate was determined by measuring the absorbance at 420 nm for 3 min, using an Agilent 8453E UV-visible spectrophotometer (Agilent Technologies, Inc., Santa Clara, CA, USA). An arbitrary enzyme unit ("PFLNA unit", U_{PFLNA}), was defined as the amount of enzyme that produces an increase of 0.001 absorbance unit (1 cm light-path) per minute in the assay conditions.

SDS-PAGE

SDS-polyacrylamide gel electrophoresis was performed in a Miniprotean III Cell (Bio-Rad, Hercules, CA, USA) according to Laemmli (1970). Current was kept constant at 40 mA during stacking and then increased to 60 mA and kept constant for 40 min. Gels (12.5% polyacrylamide) were stained by Coomassie Brilliant Blue R-250. The molecular mass markers used were myosin (250 kDa), phosphorylase (148 kDa), bovine serum albumin (98 kDa), glutamic dehydrogenase (64 kDa), alcohol dehydrogenase (50 kDa), carbonic anhydrase (36 kDa), myoglobin red (22 kDa), lysozyme (16 kDa), aprotinin (6 kDa), and insulin, B chain (4 kDa).

Determination of the partial internal amino acid sequences

The enzyme (ca. 50 μg) was inactived in guanidine chloride 6 M at 100°C, treated with 10 mM DTT in 25 mM ammonium bicarbonate for 1 h at room temperature to make sure that all the cysteine residues were reduced and then alkylated by 50 mM iodoacetamide in 25 mM ammonium bicarbonate for 1 h at 37°C in the dark. The sample was then digested at 37°C with Lys-C protease (from *Lysobacter enzymogenes*) overnight in 0.1 M ammonium bicarbonate (pH 8.0) using a 1:60 enzyme–protein ratio (w/w). The resulting peptides were separated by reverse-phase HPLC using a Vydac C18 column (5 μm, 300 Å, 4.6 mm ID × 250 mm L) of Grace (Deerfield, IL, USA) and a linear gradient of acetonitrile (5–50%) in 0.1% trifluoroacetic acid (TFA) at 0.5 ml/min. Every fraction was analyzed by mass spectrometry (UltraFlex MALDI-TOF mass spectrometer, Bruker Daltonics, Bremen, Germany) and pure samples of peptides were sequenced by Edman's automated degradation using a Beckman LF3000 protein sequencer equipped with a PTH-amino acid analyzer System Gold (Beckman Instruments, Fullerton, CA, USA).

Cloning of asclepain f cDNA

Total RNA was isolated from *Asclepias fruticosa* latex. Ten drops of latex (70 mg) were collected on 350 μl of buffer of the kit (NucleoSpin® RNA Plant Kit, BD Biosciences Clon-Tech, Palo Alto, CA, USA) containing 3.5 μl of β-mercaptoethanol. Total RNA was extracted as recommended by the manufacturers employing spin columns that traps nucleic acids. RNA was eluted with RNase-free water with RNase inhibitor and was concentrated by a Savant Speed-Vac system (TeleChem International, Inc. Sunnyvale, CA, USA).

Single-stranded cDNA was prepared by 3′-RACE (rapid amplification of cDNA ends) using the First Strand cDNA Synthesis Kit for RT-PCR, AMV (Roche Diagnostics-Roche Applied Science, Indianapolis, IN, USA) and an oligo(dT)R_1R_0 primer (5′-CCGGAATTCACTGCAGGGTACCCAATACGACTCACTATAGGGCTTTTTTTTTTTTTTT-3′; R_0 = 5′-CCGGAATTCACTGCAG-3′ and R_1 = GGTACCCAATACGACTCACTATAGGGC-3′).

Degenerate oligonucleotide primers were synthesized (Table 1) on the basis of the partial amino acid sequences determined at the protein level (N-terminal: LPDSVDWREKGVVFPIRNQGK and one of the internal peptide: SYEGHCSIAMQ). These oligos were used together with oligo(dT)R_1R_0 complementary primers to obtain the complete cDNA. PCR was performed using EcoTaq DNA polymerase and a thermal cycler EZ Cycler (Ericomp Inc., San Diego, CA, USA). Amplification was carried out by a first cycle of 95°C for 5 min and 72°C for 15 min, 30 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min and completed by a 15-min extension at 72°C; finally the block temperature was held at 8°C.

The single-stranded cDNA was used as a template for PCR amplification of cDNAs using primers Nf_{1-8} and R_0. The amplified fragments were further subjected to Nested PCR in two different reactions using primers Nf_{3-11} and R_1 and Nf_{3-11} and Pr_{3-10}, respectively.

Amplified products were gel-purified (QIAEX II Agarose Gel Extraction Kit, QIAGEN GmbH, Hilden, Germany), ligated into the pGEM-T Easy (Promega Corporation, Madison, WI, USA) vector and transformed into competent *E. coli* XL1-Blue cells. The cloning products were checked by PCR followed by gel electrophoresis, selected and sequenced in both directions (Sambrook et al. 1989). Nucleotide and deduced amino acid sequences were analyzed with the Translate Tool program (http://www.expasy.ch/tools/dna.html) and CLUSTAL W (Thompson et al. 1994), a multiple alignment program. Deduced amino acid sequences were used as inputs for database searching using the BLAST algorithm (Altschul et al. 1997).

Prepropeptide sequence determination

5′-RACE was performed using BD SMART™ RACE cDNA Amplification Kit (BD Biosciences, San Jose, CA, USA) according to the manufacturer's instructions. PCR

Table 1 Primers used for cDNA cloning of AfCP protease

Primer	Orientation	Sequence (5′ → 3′)
Nf_{1-8}	Forward	yTkCCdGATTCCGATGTTTGGmG
Nf_{3-11}	Forward	GATTCnGTwGATTGGmGrGAAAAAGG
Pr_{3-10}	Reverse	CATTGCwATnGAACAATGnCCTTC

Primers were designed on the basis of the partial amino acid sequences determined at the protein level. Letters n, y, k, d, m, w, and r indicate degenerate sites (n = A/T/C or G; y = C or T; k = G or T; d = A/G or T; m = A or C; w = A or T, and r = A or G

was performed with a reverse primer (5′-CCGCAAGCTC CACTAAATATACCCC-3′) corresponding to 422–446 location of AfCP cDNA, and as forward primers those supplied by the kit. The obtained 5′-end cDNA was sequenced and then was analyzed as explained before.

Molecular modeling

The tertiary structure of AfCP (amino acid sequence corresponding to the cysteine protease asclepain f deduced from cDNA) was predicted by the homology modeling method described by Núñez et al. (2002) with the program MODELLER (Sali and Blundell 1993), using the crystal structures of papain, caricain, actinidain, zingipain, and chymopapain as templates. The alignment of the amino acid sequence of the target enzyme (AfCP) and the templates was performed by the programs ClustalW and SEAVIEW (Galtier et al. 1996).

All preliminary models were evaluated by VERIFY3D (Bowie et al. 1991; Lüthy et al. 1992) and PROCHECK (Laskowski et al. 1993) programs to identify well- and poorly folded fragments. Preliminary models with similar 3D folds were superimposed, poorly scoring regions were edited and the "average model" structure was created by merging the remaining segments with the programs COMPARER (Sali and Blundell 1990) and JOY (Mizuguchi et al. 1998).

Additional information, such as secondary structure, placement of insertions and deletions in loops, conservation of putative catalytic residues and the disulphide bonds were taken into account to obtain a compact and well-folded structure. For each alternative alignment, new models were built and evaluated.

The graphic image was produced by PyMOL™ (DeLano Scientific LLC, Palo Alto, CA, USA). The obtained model for AfCP was compared with the selected templates by overlapping them, using the COMPARER program. At the same time, the root mean square deviation (RMSD) was determined.

Expression of proasclepain f

To obtain the expression construct for the production of recombinant asclepain f cysteine protease (rAfCP) in *Pichia pastoris* the cDNA of pro AfCP from RNA latex was synthesized by RT-PCR. A forward (Ex1) and a reverse (Ex2) primer were used: Ex1 (5′-TCTCTC**GAGA AAAGA**ATCACTTGTATTAGTACCAATTGGCG-3′) contains *Xho*I cleavage site (underlined) as well as Kex2 signal cleavage (bold) and Ex2 (5′-CCTTCGCGGCCGCT-CAAACAGGATAAGAGGGCTGCATGG-3′) contains *Not*I cleavage site (underlined). The PCR products were purified, ligated into the pGEM-T Easy vector and transformed into a competent *E. coli* XL1-Blue strain. The resulting DNA plasmid, previously checked by sequencing, was digested with *Xho*I and *Not*I restriction enzymes and ligated to pPICZα expression vector, which was previously digested with the same enzymes. Competent *E. coli* TOP10F′ cells were transformed with the ligation products. Correct insertion and sequence was confirmed by sequencing. The constructed vector was linearized with *Bst*XI restriction enzyme and then 3 μg of this construct was used to obtain transformed *Pichia pastoris* KM71H cells. The positive clones were selected from a solid medium YPDS [2% (w/v) peptone, 1% (w/v) yeast extract, 18.2% (w/v) sorbitol, and 2% (w/v) agar] containing 100 μg/ml Zeocin™, after growing 3 days at 30°C, and transferred to a liquid medium YPD [2% (w/v) peptone, 1% (w/v) yeast extract, and 2% (w/v) glucose] for its conservation.

Selected transformed clones and a control clone (*Pichia pastoris* GS115 Albumin, phenotype Mut^S) were grown at 30°C and 250 rpm to a final density of 4 OD_{600} in 50 ml tubes containing 15 ml of liquid medium BMGY [2% (w/v) peptone, 1% (w/v) yeast extract, 1.34% (w/v) yeast nitrogen base, 1% (w/v) glycerol, 4×10^{-5}% (w/v) biotin, and 100 mM potassium phosphate, pH 6.0]. Cells were harvested by centrifugation at 3,000*g* for 5 min at room temperature and resuspended in BMMY medium (similar to BMGY but with 0.5% methanol instead of glycerol) in 1/15 of the original volume. To maintain induction, cultures were supplemented every 24 h with methanol to a final concentration of 0.5% (v/v) and continued for 3 days. Culture supernatants were sampled at different times to monitor the rAfCP production. The supernatants of each sample were dialysed (cut off 6,000–8,000 Da) overnight against distilled water at 4°C. The proenzyme autocatalytic activation was assayed in 0.1 M acetic/sodium acetate buffer (pH 4.0) containing 20 mM cysteine at different temperatures (20, 37, 45 and 60°C) and incubation times (10, 30 and 60 min). The expression of the active enzyme was quantified by determination of specific activity on PFLNA and evaluated by SDS-PAGE.

Peptide mass fingerprinting (PMF)

In-gel protein digestion was realized using the In-Gel $Digest_{ZP}$ Kit (MILLIPORE, Billerica, MA, USA). Protein bands were excised with a scalpel, diced into 1-mm^3 pieces and introduced into the wells, treated with 25 mM NH_4HCO_3/5% acetonitrile (ACN) and then with 25 mM NH_4HCO_3/50% ACN (washing/destaining solutions) and finally dehydrated with pure ACN. Proteins were then reduced with 10 mM DTT in buffer 25 mM NH_4HCO_3/5% ACN for 1 h at room temperature. Cys sulphydrils were alkylated with 50 mM iodoacetamide in 25 mM NH_4HCO_3/5% ACN for 1 h at 37°C. Washing/destaining solutions

were used for additional washing and finally the gel pieces were dehydrated with pure ACN, and then dried. Trypsin (0.02 μg) in 25 mM NH_4HCO_3/5% ACN was added to rehydrate the gel pieces in each well and the digestion was carried out for 3 h at 37°C. The tryptic peptides obtained were extracted from the gel with 0.2% TFA and captured on the C18 resin applying vacuum. Finally, the tryptic peptides were eluted with 50% ACN/0.1% TFA.

MALDI-TOF-MS was used for protein identification by peptide mass fingerprint (PMF). Analysis was performed using an UltraFlex MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). The sample was spotted on an AnchorChip target 160 and mixed with freshly prepared matrix solution (10 mg/ml of α-cyano-4-hydroxycinnamic acid in aqueous solution containing 30% ACN and 0.1% TFA). External calibration was performed using peptide calibrants. Peptide masses were acquired with a range of ca. 800–4,000 *m/z*. PMF spectra of samples analyzed were acquired and compared. Comprehensive peak assignments were accomplished using the BioTools software package (Bruker Daltonics).

Results and discussion

Partial amino acid sequences

Asclepain f was isolated and purified from latex of *Asclepias fruticosa* following the procedure previously described. After digestion of the enzyme with Lys-C protease several peptides could be isolated by HPLC and sequenced: YEGHCGIAMQ (P1), LPDSVDWREK (P2), and YPYIFRQGQCYQP (P3), according to HPLC elution order. P2 peptide corresponds to the N-terminus of the mature protein that had been previously reported (Trejo et al. 2001).

Cloning and sequence analysis of asclepain f

Total RNA was extracted from *Asclepias fruticosa* latex and used to synthesize the cDNA encoding asclepain f by RT-PCR. The N-terminal and the P1 peptide sequences were used to design degenerate oligonucleotide primers (Table 1), on the basis of a previous analysis of the codon usage (Countcodon program, version 4 http://www.kazusa.or.jp/codon/countcodon.html) in proteins from species belonging to the *Asclepiadeae* tribe (*Apocynaceae*).

The consensus sequence was obtained from four selected clones by using the CLUSTAL W multiple sequence alignment program. The sequence was composed of 759 nucleotides, the former 627 ones encoding for the asclepain f sequence (excluded the first two amino acids, as it comes from a Nested-PCR using the Nf_{3-11} primer), followed by the tga stop codon, the polyadenylation signal (aataaa) and the polyadenine tail composed of 24 residues. The proposed sequence for the mature enzyme, shown in Fig. 1, contains 211 amino acids, included Leu and Pro residues coming from the N-terminus sequence. Correspondence of the asclepain f deduced sequence (AfCP) with amino acid sequences determined at the protein level is underlined.

The calculated molecular mass, monoisotopic mass, and molar extinction coefficient were 23,406, 23,391 Da, and

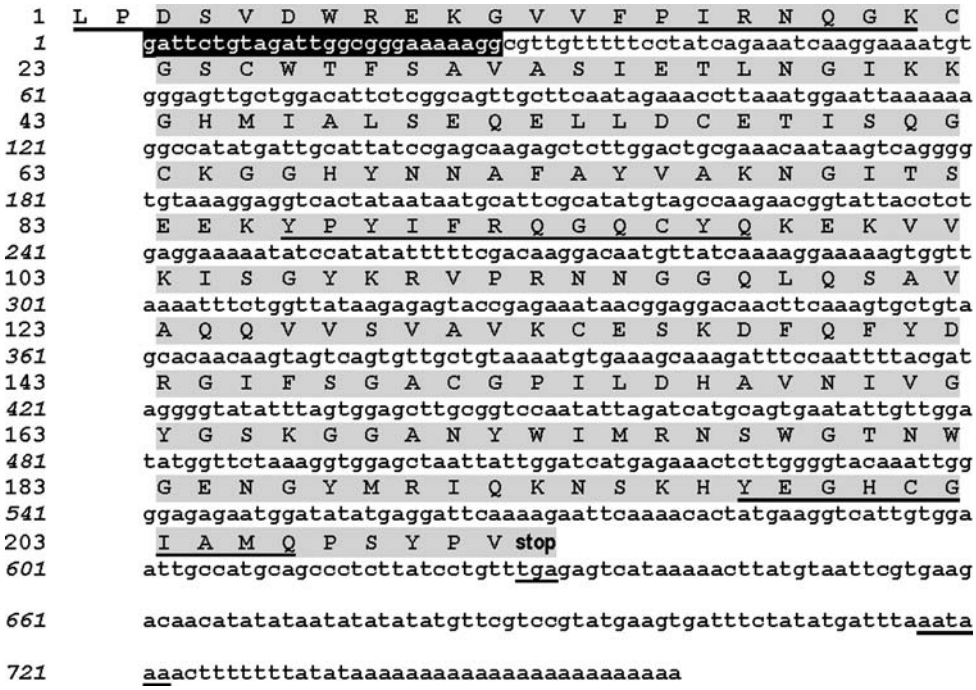


Fig. 1 Nucleotide and deduced amino acid sequences of AfCP. Coincident amino acid sequences with those determined at the protein level are *underlined*. In the nucleotidic sequence the stop signal (tga) and the polyadenylation signal sequence (aataaa) are also *underlined*. *White* on a *black background* sequence corresponds to Nf_{3-10} primer

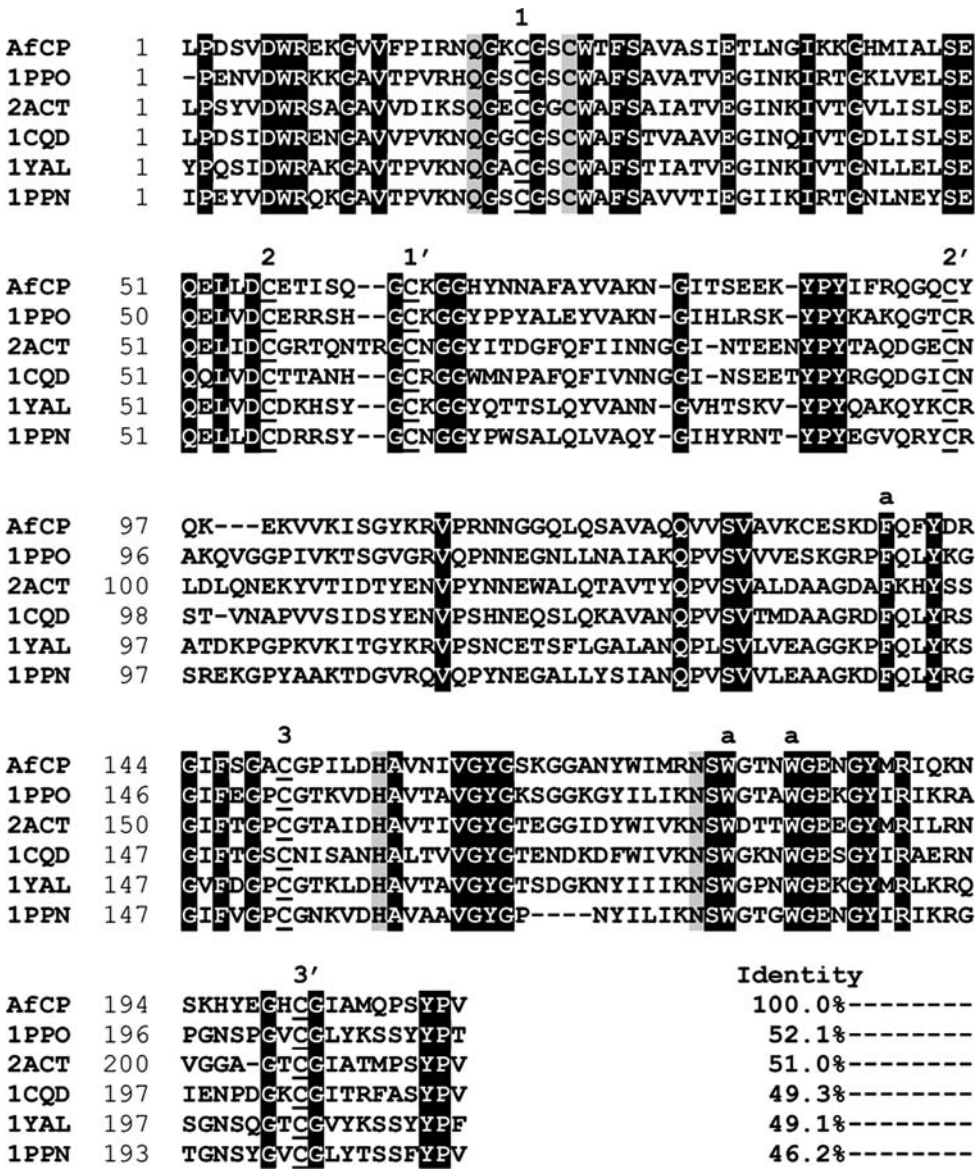


Fig. 2 Sequence alignment of AfCP with plant proteases belonging to the subfamily C1A. AfCP sequence was aligned with the sequences of caricain (1PPO), actinidain (2ACT), zingipain (1CQD), chymopapain (1YAL), and papain (1PPN). Active-site residues are *black* on a *gray background*. Conserved motifs and conserved amino acid residues are *white* on a *black background*. Cysteines involved in disulfide bridges are *underlined* (1-1′, 2-2′and 3-3′). *Dashes* represent gaps introduced to maximize alignment. **a** indicate the amino acids involved in the hydrophobic pocket

44,290 M^{-1} cm^{-1}, respectively, when the AfCP sequence was analyzed by GPMAW v6.0 (Lighthouse data, DK-5230 Odense M, Denmark). The calculated pI value was 9.7, in agreement with the experimental value, higher than 9.3 (Trejo et al. 2001).

The deduced sequence was analyzed by BLAST, finding 46–52% identity with 13 cysteine peptidases belonging to the C1A family, with several of them with high-resolution crystal structure already deposited at PDB.

The alignments of AfCP with the four plant peptidases with higher identity values are shown in Fig. 2. Caricain (Pickersgill et al. 1991) and actinidain (Baker and Dodson 1980) are the most similar peptidases, followed by zingipain (Choi et al. 1999) and chymopapain (Maes et al. 1996): 52.1, 51.0, 49.3, and 49.1% identity, respectively. Papain (Pickersgill et al. 1992), the "canonical" member belonging to the subfamily C1A type peptidase, was also included for comparative purposes.

As could be expected, a high conservation degree was observed particularly for those amino acid residues which are essential for catalytic activity and tridimensional structure: Cys25 and His159 (papain numbering), which constitute the catalytic dyad in all cysteine peptidases, as well as active site residues Gln19 and Asn175 (papain numbering), characteristic of Clan CA (Barrett and Rawlings 2004). Also, Phe141, Trp177 and Trp181 (papain numbering), involved in the hydrophobic pocket of the catalytic site, are also conserved in AfCP. In addition to the Cys25 residue of the active site, seven Cys residues are located in AfCP: Cys22, Cys56, Cys63, Cys95, Cys150, and Cys201, probably involved in disulfide bridges (as suggested by modeling), as well as a free thiol group provided by Cys



Fig. 3 Tertiary structure of AfCP predicted by molecular modeling. **a** Ribbon model of AfCP; catalytic site residues (*orange*) and disulfide bridges (*yellow*) are represented with *sticks*. **b** Backbone overlapped structure of AfCP with papain, caricain, actinidain, zingipain, and chymopapain. **c** Overlapped stick structure of catalytic site residues of AfCP with papain, caricain, actinidain, zingipain, and chymopapain

133. A free thiol group is also present in chymopapain, *Carica candamarcensis* endopeptidase, and ficain, but is absent in papain (Barrett and Rawlings 2004). On the basis of the above information, asclepain f should be included in the cysteine peptidase subfamily C1A.

Tertiary structure of AfCP predicted by molecular modeling

The predicted molecular model (Fig. 3a), analyzed with the VERIFY3D program, revealed that more than the 84% of the residues showed a score greater than 0.2 (1D–3D compatibility), with no negative values in any case. On the other hand, the PROCHEK program showed that 99.3% of the binding distances and 93.3% of the binding angles are within the allowed limits and that all the residues fall into the stable conformational areas of the Ramachandran plot.

Since the RMSD of α-carbons is approximately 1 Å for proteins with 50% of identity, the proposed model for AfCP (Fig. 3b) agrees with the crystal structures used as templates (caricain, actinidain, zingipain, chymopapain and papain). Moreover, the overlapping of the catalytic center residues shows a particular disposition highly concordant with those of the templates (Fig. 3c).

The enzyme structure shows the typical papain-like fold composed of two domains separated by a groove containing the active site. The α-helix-rich (L) domain contains the catalytic residues Gln19 and Cys25 and it is stabilized by two disulfide bridges (Cys22-Cys63 and Cys56-Cys95). The β-barrel-like (R) domain contains the catalytic residues His156 and Asn176 and it is stabilized by another disulfide bridge (Cys150-Cys201).

The substrate preferences shown by the members of the papain family are generally believed to be dictated by the S2 subsite (Drenth et al. 1976). In papain, the S2 subsite is essentially composed of residues Tyr67, Pro68, Trp69, Val133, Val157, and Phe207 (Maes et al. 1996), while in asclepain f the S2 pocket (based on a model in which the Cα trace of asclepain f was superimposed on that of papain) is composed by His67, Tyr68, Asn69, Ala130, Leu154, and Ser208. The comparison of S2 subsites for papain and asclepain f revealed important differences in molecular geometry and in polarity, specially notable in the 68, 69, and 207 residues (papain numbering), from which substrate specificity differences can be inferred (in fact, benzoyl-DL-arginine *p*-nitroanilide is hydrolyzed by papain but not by asclepain f, data not shown). Asclepain f should be considered as a distinct enzyme entity on the basis of these considerations, as well as on the basis of low-percentage identity to papain and other members of the papain family (Fig. 2).

Cloning and sequence analysis of asclepain f prepropeptide

Proteases are synthesized as inactive or less-active precursor molecules in order to prevent inappropriate proteolysis; then they are activated by limited intra- or intermolecular proteolysis cleaving off an inhibitory peptide. In the case of papain family cysteine peptidases, proforms are also involved in other important functions, such as the participation in correct intracellular targeting and assistance in folding of the mature enzyme (Wiederanders 2003).

With the purpose of finding out the complete sequence of asclepain f, total RNA was extracted from latex and used for obtaining the corresponding cDNA by RT-5′RACE. The consensus sequence, obtained from four selected clones, was composed of 839 nucleotides, of which the former 48 ones encode the signal sequence (16 amino acids) as predicted using the SignalP 3.0 algorithm (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen et al. 1997), the following 339 ones for the propeptide (113 amino acids), and the remaining nucleotides matched with part of AfCP cDNA previously obtained.

Asclepain f prepropeptide (pAfCP) sequence contained the ERFNIN and GNFD motifs, highly conserved in cysteine protease propeptides. The presence of the ERFNIN motif in the α2 helix allows its inclusion in the cathepsin L group (Karrer et al. 1993). Both motifs are also conserved

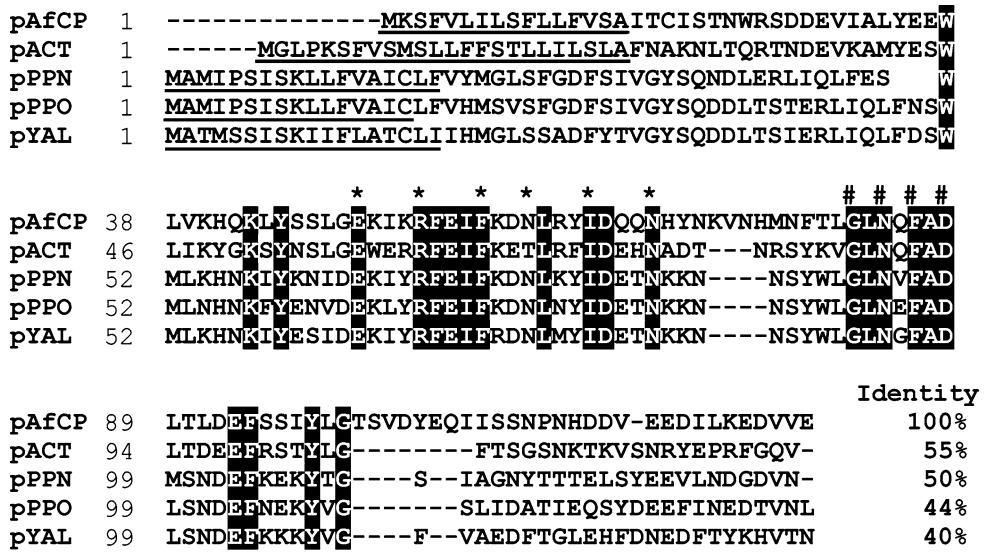


Fig. 4 Sequence alignment of asclepain f prepropeptide with those of some plant proteases belonging to the subfamily C1A. Asclepain f prepropeptide sequence was aligned with those corresponding to actinidain (pACT), papain (pPPN), caricain (pPPO), and chymopapain (pYAL). Conserved amino acid residues are *white* on a *black background*. *Asterisks* (*) and *numerals* (#) indicate the ERFNIN and GNFD motifs, which characterize the prepropeptide sequences of plant proteases belonging to the subfamily C1A. *Dashes* represent gaps introduced to maximize alignment. Signal peptide sequences are *underlined*

in papain (Cohen et al. 1986), caricain (Taylor et al. 1995), and chymopapain (Maes et al. 1996) prepropeptide sequences (Fig. 4), which show the higher identity degree with pAfCP. In the case of actinidain (Baker and Dodson 1980), the ERFNIN motif is only partially conserved (the former N is replaced by T).

On the basis of the sequences determined for the prepropeptide region and the mature enzyme, the full sequence of the preproenzyme was derived and, subsequently, submitted and deposited in EMBL nucleotide sequence database (accession number FM201283).

Heterologous expression of proasclepain f

Pichia pastoris based expression system was used for the production of recombinant asclepain f cysteine protease (rAfCP). This system allows post-translational modifications such as disulfide bond formation and proteolytic processing (Brömme et al. 2004).

A construct encoding the prodomain and the mature domain of AfCP was amplified by RACE-PCR using the Ex1 and Ex2 primers to express the recombinant proenzyme. In this construct different sequence elements were included, from 5′ to 3′ ends: the yeast α-factor secretion signal sequence, the *Xho*I cleavage site, the Kex2 signal cleavage and the full-length proasclepain f. The *Not*I cleavage site was placed at this end, preceding the rest of the vector sequence (pPICZα).

Escherichia coli TOP10F′clones containing the correct in-frame sequence were selected for transformation into *P. pastoris* strain KM71H. From more than 30 transformed zeocin-resistant clones three (A, B and C) were selected among the most productive regarding the expression level of the recombinant enzyme during the induction period, using an albumin expressing clone as control. Most heterologous expression systems for cysteine proteases produce proteolytically inactive proenzymes which are stable at slightly acid to slightly alkaline pH values, but the proregion can be removed autocatalytically by incubating the precursor proteins in acetate buffer at pH 4–4.5 (Brömme et al. 2004). In order to obtain rAfCP in its active form, the best results were obtained incubating at pH 4.0 for 60 min at 60°C the culture supernatants of the selected clones. As can be seen in Fig. 5, the highest specific activity in activated culture supernatants was detected for clone B after 3 days induction, while protein content is almost equal in the three clones (about 220 μg/ml, day 3).

Figure 6 shows the corresponding SDS-PAGE, where a band that could be ascribed to the mature enzyme was detected. The identity of this band was confirmed by trypsin digestion and PMF-MALDI-TOF-MS comparisons with the mature enzyme isolated from plant latex (75.2 and 71.6% intensity coverage and 45.5 and 45.0% sequence coverage, respectively). Additionally, the detection of cysteine peptidase activity on the PFLNA substrate confirmed the expression of the functional enzyme.

It is known that about 110 latex-containing plant families contain at least one proteolytic enzyme (Domsalla and Melzig 2008). Most of them are cysteine or serine endopeptidases. Proteolytic enzymes of *Carica papaya* (papain,

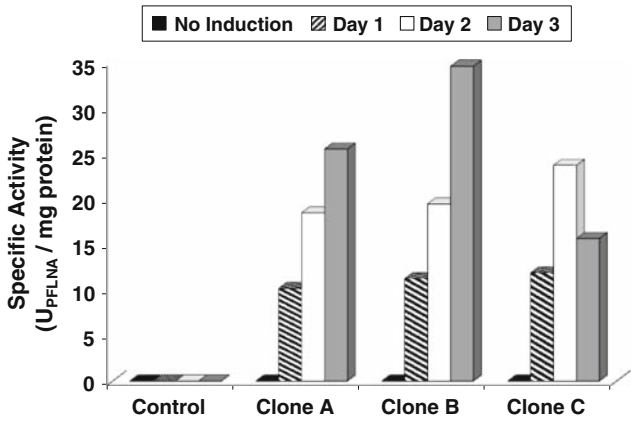


Fig. 5 Specific activity on PFLNA of different *Pichia pastoris* clones. Samples were previously activated in 0.1 M acetic/sodium acetate buffer (pH 4.0) containing 20 mM cysteine for 60 min at 60°C. Data represent the mean value of five determinations and each experiment was repeated twice

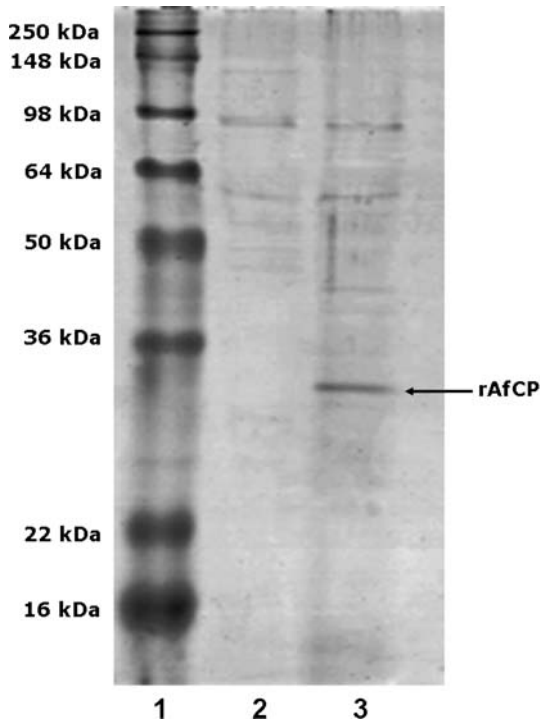


Fig. 6 SDS-PAGE (12.5%) of culture supernatant of clone B after 3 days of induction. Molecular mass standards (*Lane 1*); culture supernatant of control clone: *Pichia pastoris* GS115 Albumin, phenotype Mut^S (*Lane 2*); culture supernatant of clone B after 3 days of induction (*Lane 3*). Supernatants of control clone and clone B were previously activated in 0.1 M acetic/sodium acetate buffer (pH 4.0) containing 20 mM cysteine for 60 min at 60°C

chymopapain, caricain, and glycyl endopeptidase) have been the most extensively studied latex peptidases. Although proforms (proenzymes) of proteinases have not been isolated from papaya latex, they have been successfully expressed in heterologous hosts: the proform of papain has been expressed in insect cells, *E. coli* and *Saccharomyces cerevisiae*, and the proforms of the other three related proteinases have also been produced by *E. coli* cells (El Moussaoui et al. 2001). No information has been reported about the expression of other latex peptidases. Hence, asclepain f, the cysteine protease from *Asclepias fruticosa* here characterized is, as far as we know, the first peptidase cloned and expressed from mRNA isolated from plant latex, confirming the presence of the preproenzyme mRNA in it.

Acknowledgments The present work was supported by grants from ANPCyT, CONICET, and CIC (Argentina), CYTED (Project IV.22) and BIO2007-68046-C02 (MEC, Spain). S.A. Trejo was CONICET Fellow, L.M.I. López belongs to CONICET Researcher Career, and N.O. Caffini and C.L. Natalucci belong to CIC Researcher Career. The IBB-UAB is a member of ProteoRed, funded by Genoma Spain and follows the quality criteria set up by ProteoRed standards.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Baker EN, Dodson EJ (1980) Crystallographic refinement of the structure of actinidin at 1.7 Angstroms resolution by fast Fourier least-squares. Acta Crystallogr A 36:559–572

Barrett AJ, Rawlings ND (2001) Evolutionary lines of cysteine peptidases. Biol Chem 382:727–733

Barrett AJ, Rawlings ND (2004) Introduction: the clans and families of cysteine peptidases. In: Barrett AJ, Rawlings ND, Woessner JF (eds) Handbook of proteolytic enzymes, 2nd edn. Elsevier, London, pp 1051–1071

Bowie JU, Lüthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. Science 253:164–170

Bradford MB (1976) A rapid and sensitive method for the quantitation of micrograms quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248–254

Brömme D, Nallaseth FS, Turk B (2004) Production and activation of recombinant papain-like cysteine proteases. Methods 32:199–206

Choi KH, Laursen RA, Allen KN (1999) The 2.1 Angstrom structure of a cysteine protease with proline specificity from *Ginger Rhizome*, *Zingiber officinale*. Biochemistry 38:11624–11633

Cohen LW, Coghlan VM, Dihel LC (1986) Cloning and sequencing of papain-encoding cDNA. Gene 48:219–227

Domsalla A, Melzig MF (2008) Occurrence and properties of proteases in plant lattices. Planta Med 74:699–711

Drenth J, Kalk KH, Swen HM (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain. Biochemistry 15:3731–3738

El Moussaoui A, Nijs M, Paul C, Wintjens R, Vincentelli J, Azarkan M, van Looze Y (2001) Revisiting the enzymes stored in the laticifers of *Carica papaya* in the context of their possible participation in the plant defence mechanism. Cell Mol Life Sci 58:556–570

Filippova IY, Lysogorskaya EN, Oksenoit ES, Rudenskaya GN, Stepanov VM (1984) L-Pyroglutamyl-L-phenylalanyl-L-leucine-*p*-nitroanilide—a chromogenic substrate for thiol proteinase assay. Anal Biochem 143:293–297

Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548

Karrer KM, Pfeifer SL, Di Tomas ME (1993) Two distinct gene subfamily within the family of cysteine protease genes. Proc Natl Acad Sci USA 90:3063–3067

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291

Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356:83–85

Maes DJ, Bouckaert F, Poortmans WL, Looze Y (1996) Structure of chymopapain at 1.7 A resolution. Biochemistry 35:16292–16298

Mizuguchi K, Deane CM, Blundell TL, Johnson MS, Overington JP (1998) JOY: protein sequence-structure representation and analysis. Bioinformatics 14:617–623

Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1–6

Núñez MR, Jiye SM, Mizuguchi K (2002) Protein fold recognition and comparative modeling using HOMSTRAD, JOY and FUGUE. In: Tsigelny I (ed) Protein structure prediction: bioinformatic approach. International University Line, La Jolla, pp 143–169

Pickersgill RW, Rizkallah P, Harris GW, Goodenough PW (1991) Determination of the structure of papaya protease omega. Acta Crystallogr B 47:766–771

Pickersgill RW, Harris GW, Garman E (1992) Structure of monoclinic papain at 1.60 Angstroms resolution. Acta Crystallogr B 48:59–67

Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ (2008) MEROPS: the peptidase database. Nucleic Acids Res 36:D320–D325

Sali A, Blundell TL (1990) Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. J Mol Biol 212:403–428

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Taylor MA, Baker KC, Briggs GS, Connerton IF, Cummings NJ, Pratt KA, Revell DF, Freedman RB, Goodenough PW (1995) Recombinant pro-regions from papain and papaya proteinase IV-are selective high affinity inhibitors of the mature papaya enzymes. Protein Eng 8:59–62

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Trejo SA, López LMI, Cimino CV, Caffini NO, Natalucci CL (2001) Purification and characterization of a new plant endopeptidase isolated from latex of *Asclepias fruticosa* L. (*Asclepiadaceae*). J Protein Chem 20:445–453

Turk B, Turk V, Turk D (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. Biol Chem 378:141–150

Van der Hoorn RA, Jones JD (2004) The plant proteolytic machinery and its role in defence. Curr Opin Plant Biol 7:400–407

Wiederanders B (2003) Structure–function relationships in class CA1 cysteine peptidase propeptides. Acta Biochim Polon 50:691–713