Planta (2012) 236:1471–1484
DOI 10.1007/s00425-012-1701-3

ORIGINAL ARTICLE

Characterization of the proteolytic system present in *Vasconcellea quercifolia* latex

María José Torres · Sebastián Alejandro Trejo · Walter David Obregón · Francesc Xavier Avilés · Laura María Isabel López · Claudia Luisa Natalucci

Received: 13 February 2012 / Accepted: 20 June 2012 / Published online: 12 July 2012

Abstract *Vasconcellea quercifolia* (Caricaceae) latex contains several cysteine endopeptidases with high proteolytic activity. Cysteine endopeptidases are the main active compounds used by the plant as a defense mechanism. A proteolytic preparation from *V. quercifolia* ("oak leaved papaya") latex was purified by cation exchange chromatography. From SDS-PAGE and blotting of the selected fractions, the N-terminal amino acid sequences of polypeptides were determined by Edman's degradation. The analysis by peptide mass fingerprinting (PMF) of the enzymes allowed their characterization and confirmed the presence of seven different cysteine proteinases in the latex of *V. quercifolia*. Moreover, the comparison between the tryptic maps with those deposited in databases using the MASCOT tool showed that none of the isolated proteases matched with another plant protease. Notably, a propeptidase was detected in the plant latex, which is being the first report in this sense. Furthermore, the cDNA of one of the cysteine proteases that is expressed in the latex of *V. quercifolia* was cloned and sequenced. The consensus sequence was aligned using the ClustalX web server, which allowed detecting a high degree of identity with cysteine proteases of the Caricaceae family and establishing the evolutionary relationship between them. We also observed a high conservation degree for those amino acid residues which are essential for the catalytic activity and tridimensional structure of the plant proteases belonging to the subfamily C1A. The PMF analysis strongly suggests that the sequence obtained corresponds to the VQ-III peptidase.

Keywords Caricaceae · cDNA · Cysteine endopeptidases · Peptide mass fingerprint · Plant proteases · *Vasconcellea*

Abbreviations

ACN	Acetonitrile
BLAST	Basic local alignment search tool
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
MALDI-TOF MS	Matrix-assisted laser desorption ionization time-of-flight mass spectrometry
PMF	Peptide mass fingerprinting
PTH	Phenylthiohydantoin
PVDF	Polyvinylidene fluoride
RACE	Rapid amplification of cDNA ends
TFA	Trifuoroacetic acid
VqCP-A	Amino acid sequence corresponding to the cysteine protease deduced from cDNA

Introduction

Cysteine proteinases are proteolytic enzymes found in all eukaryotic organisms. The physiological functions attributed

Electronic supplementary material The online version of this article (doi:10.1007/s00425-012-1701-3) contains supplementary material, which is available to authorized users.

M. J. Torres · W. D. Obregón · L. M. I. López · C. L. Natalucci (✉)
Laboratorio de Investigación de Proteínas Vegetales (LIProVe), 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

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

to plant cysteine proteinases include the buildup and breakdown of storage proteins during seed germination, organ senescence, and programmed cell death. Most important perhaps is their involvement in the proteasome proteolytic pathway affecting several metabolic processes, such as hormone signaling, cell cycle, embryogenesis, morphogenesis, flower development, and oxidative stress (Salas et al. 2008). Proteases also appear to play key roles in the recognition of pathogens and pests and in the induction of effective defense responses (Van der Hoorn and Jones 2004; Souza et al. 2011).

An important source of plant proteases is latex, an aqueous suspension or emulsion of various kinds of particles borne within living cells. Alkaloids, starches, sugars, oil, tannins, resins, and gums are found in the complex emulsion/suspension proteins. Laticifers, the latex-bearing structures, vary in origin, anatomy, and distribution. The use of this plant product in traditional medicine and industry is well known (Domsalla and Melzig 2008). When plants are injured, latex exudates transiently until a protein clot forms around the wounded area. The coagulation process is vital in creating a physical barrier against predator attack. Many of these latex changes take place concomitantly with drastic variations in proteolytic activity, suggesting the involvement of proteolytic enzymes during clot formation (Silva et al. 1997; Moutim et al. 1999).

Articulated laticifers, containing milky latex, are present in all the organs of members of Caricaceae (Hegnauer 1964; Badillo 2000, 2001; Van Droogenbroeck et al. 2002, 2004; Kyndt et al. 2005). The most extensively investigated plant cysteine protease is papain (EC 3.4.22.2) from the latex of *Carica papaya* (Otto and Schirmeister 1997), which also contains chymopapain (EC 3.4.22.6, the main proteolytic component), caricain (EC 3.4.22.30), and glycyl endopeptidase (EC 3.4.22.25), also named papaya proteinase IV (Barrett et al. 2004a). Mexicain, the most abundant peptidase from the latex of the fruits of *Jacaratia mexicana*, member of the Caricaceae family, has been isolated and crystallized (Oliver-Salvador et al. 2004; Gavira et al. 2007). Latex from *Carica candamarcensis* (also known as *Vasconcellea cundinamarcensis*), another member of the Caricaceae family common to many areas of South America, contains cysteine proteinases that participate in clot formation and display higher (5- to 7-fold) proteolytic activity than their homologs from *C. papaya* (Baeza et al. 1990; Bravo et al. 1994). Several endopeptidases have been isolated and characterized from the latex of *V. cundinamarcensis* (Walreavens et al. 1993; Walraevens et al. 1999; Jaziri et al. 1994; Pereira et al. 2001; Teixeira et al. 2008; Gomes et al. 2008, 2010). At least two of these proteases also exert proliferative effects when incubated with mammalian cells (Gomes et al. 2005, 2009; Corrêa et al. 2011). Proteolytic activity, measured by BAPNA degradation and expressed as amounts of nkat per milligram protein is found to be 1.25–2.0 times higher in the latex of other Vasconcellea spp. (*V. monoica*, *V. stipulata* and *V.* × *heilbornii*) than in latex of *C. papaya.* Further, cDNA sequences coding for cysteine proteinases in *Vasconcellea* × *heilbornii* and *Vasconcellea stipulata* have been determined (Kyndt et al. 2007). All enzymes of this group, belonging to the C1A subfamily, are synthesized as precursors, with N-terminal propeptides as well as 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. 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. 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 (Trejo et al. 2009).

Recently, we have reported a high proteolytic activity in *Vasconcellea quercifolia* (Caricaceae) latex, in comparison with the *C. papaya* latex, its ability to hydrolyze casein and to clotting milk. In addition, we purified a new cysteine protease with higher specificity than papain and characterized it (Torres et al. 2010). In the present paper, the proteolytic system contained in *Vasconcellea quercifolia* latex was analyzed and characterized using proteomic and molecular biology tools. Knowledge of these new proteases with individual characteristics is important both to learn about their biological function as well as to infer their potential pharmacological or industrial applications.

Materials and methods

Plant material

Unripe fruits of *Vasconcellea quercifolia* A.St. -Hil. (voucher specimen accession code: LPAG 5647) were collected at the beginning of summer of 2009 in the Guasayán hills, province of Santiago del Estero, Argentina. This fruit is native to South America and grows in Argentina, Bolivia, Brazil, Ecuador, Paraguay, Peru, and Uruguay (Scheldeman et al. 2007).

V. quercifolia, also called "ñacaratiá", "mamon of the mount", or "oak leaved papaya" is a dioecious tree growing to 3–6 m high. Flowers are on axillary inflorescences and the female flowers are solitary or in sparse racemes. Fruits are small and ovoid berries, 3–5 cm long, bright orange, and juicy (Dimitri 1979).

Partially purified preparation (VQ preparation)

A partially purified preparation, named VQ, was prepared by centrifugation of latex obtained by superficial incisions of fruits and collected in buffer (pH 5.6, containing sodium tetrathionate to avoid autodigestion and EDTA for phenoloxidase inactivation) in an ice-water bath, as previously described (Torres et al. 2010).

Purification of the *Vasconcellea quercifolia* preparation

The VQ preparation (6 ml) was applied onto a column (Pharmacia XK 16/40, with AK16 adaptors) packed with SP-Sepharose High Performance (20 ml) equilibrated with 50 mM Tris–HCl buffer (pH 7.5) containing 1 mM sodium tetrathionate. Cation exchange chromatography (FPLC system, Pharmacia Biotech, Uppsala, Sweden) was developed by adding 20 ml of the starting buffer (50 mM Tris–HCl, pH 7.5, containing 1 mM sodium tetrathionate and 0.15 M NaCl), followed by 80 ml of a sodium chloride linear gradient (0.15–0.45 M) and 20 ml of a sodium chloride linear gradient (0.45–1.0 M), both prepared in the same buffer. In the chromatographic fractions, proteins were detected by measuring absorbance of eluates at 280 nm and the caseinolytic activity of all fractions was measured. In an additional experiment, the fractions were collected directly from the chromatographic column in 1 mM iodoacetic acid.

Proteolytic activity on casein

The reaction mixture contained 1.1 ml of 1 % casein solution and 0.1 ml of enzyme solution, both in 0.1 M Tris–HCl buffer (pH 8.5) containing 20 mM cysteine. The reaction was carried out at 37 °C and stopped by the addition of 5 % trichloroacetic acid (1.8 ml). Then, each test tube was centrifuged at 4,000×*g* for 20 min and the absorbance of the supernatant was read at 280 nm (Natalucci et al. 1996).

Proenzyme activation

In order to obtain the active protease from the putative proenzyme, different treatments were tested. Fraction III was incubated at 60 °C for 10 min in buffer pH 4.0 either with or without 20 mM cysteine, with 100 μM pepsin or in buffer Tris–HCl pH 8.0. Then, each sample was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and its caseinolytic activity was determined. For this treatment the fraction III was selected because of its evidently shown 31 kDa band by SDS-PAGE.

SDS-PAGE

SDS–polyacrylamide gel electrophoresis was performed in a Miniprotean III Cell (Bio-Rad, Hercules, CA, USA) according to Laemmli (1970). Iodoacetic acid (1 mM) was added to sample buffer to avoid autodigestion of the protease. The current was kept constant at 40 mA during stacking and then increased to 60 mA, and kept constant for 45 min. Gels (12.5 % polyacrylamide) were stained with colloidal Coomassie (17 % w/v ammonium sulfate, 34 % methanol, 0.5 % acetic acid, and 0.1 % w/v Coomassie G-25) and then destained using distilled water.

N-terminal sequence

Samples were subjected to SDS-PAGE and transferred to a PVDF membrane. The sample bands, corresponding to an approximate weight of 26 kDa, were cut and subjected to a decoloration process. Then, the Cys sulphydryls were reduced with 10 mM DTT in NH_4HCO_3 buffer (50 mM, pH 8.0) for 30 min at 37 °C and carbamidomethylated with 20 mM iodoacetamide in 50 mM NH_4HCO_3 pH 8.0 for 30 min at 37 °C.

Each sample was introduced into the reaction chamber of the sequencer and the N-terminal sequence was determined by Edman's automated degradation using an Procise 492 (Applied Biosystems, Carlsbad, CA, USA) peptide sequencer equipped with a 140C Microgradient System HPLC (Applied Biosystems), a Series 200 UV/VIS detector (PerkinElmer, Waltham, MA, USA) and a Spheri5-PTH column (Applied Biosystems) for the amino acid analysis.

Protein homology searches were performed using the BLAST network service (Altschul et al. 1997).

Peptide mass fingerprinting (PMF)

Tryptic digestion

In-gel protein digestion was carried out using the In-Gel $Digest_{ZP}$ Kit (Millipore, Billerica, MA, USA). Protein bands of SDS-PAGE, corresponding to an approximate weight of 26 kDa, 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 sulphydryls 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 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 by applying vacuum. Finally, the tryptic peptides were eluted with 50 % ACN/0.1 % TFA.

Mass spectrometry

MALDI-TOF MS was used for protein identification by peptide mass fingerprinting (PMF). Analysis was performed using an UltrafleXtreme MALDI-TOF mass spectrometer (Bruker Daltonics, Bremen, Germany). The sample was spotted on a GroundSteel 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*. The PMF spectra of the samples analyzed were acquired and compared. Comprehensive peak assignments were accomplished using the BioTools software package (Bruker Daltonics). MASCOT search tool (URL "http://www.matrixscience.com") was used for the identification of tryptic maps.

Glycoprotein detection

ECL™ Glycoprotein Detection Module (Amersham, GE Healthcare Biosciences, Uppsala, Sweden) was used to determine if the proteins were glycosylated. Samples were separated by SDS-PAGE and the bands were transferred to a nitrocellulose membrane. The carbohydrate portions of a target protein are oxidized with sodium metaperiodate to form aldehydes which can spontaneously react with hydrazides. Biotin-X-hydrazide is used to attach biotin onto the oxidized carbohydrates, and biotinylated proteins are detected by addition of streptavidin conjugated to horseradish peroxidase for luminol-based detection using electrochemiluminescence (ECL) reagents.

Cloning of a cysteine protease cDNA

Total RNA was isolated from *V. quercifolia* latex. Drops of latex (100 mg) were collected on 350 µl of buffer of the kit (Cytoplasmic RNA Reagent, Invitrogen, Carlsbad, CA, USA), containing 3.5 µl of β-mercaptoethanol and 1 µl of RNase inhibitor. Total RNA was extracted as recommended by the manufacturers, using spin columns that trap nucleic acids, contaminating DNA was removed by a DNase solution, and RNA was eluted with 20 µl of RNase-free water with RNase inhibitor.

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)R1R0 primer (5′-CCGGAATTCACTGCAGGG TACCCAATACGACTCACTATAGGGCTTTTTTTTTTT TTTTTT-3′).

A degenerate oligonucleotide primer Nt3-11Af (5′-GA TTCnGTwGATTGGmGrGAAAAAGG-3′) for cysteine protease of latex (Trejo et al. 2009) was used together with oligo(dT)R1 (5′-GGTACCCAATACGACTCACTATAGG GC-3′) to obtain the complete cDNA. PCR was performed using Eco Taq 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, 25 cycles of 94 °C for 30 s, 44–56 °C for 30 s, and 72 °C for 1 min and completed by a 10-min extension at 72 °C; finally the block temperature was held at 8 °C.

The amplified products were gel-purified (QIAEX II Agarose Gel Extraction Kit, Qiagen GmbH, Hilden, Germany), ligated into the pGEM-T Easy (Promega, Madison, WI, USA) vector and transformed into competent *Escherichia 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 obtained with the ExPASy Translate tool program ("http://www.expasy.ch/tools/dna.html"). Deduced amino acid sequences were used as input for database searching using the BLAST (basic local alignment search tool) algorithm (Altschul et al. 1997) and analyzed by ClustalX (Thompson et al. 1997).

Phylogenetic tree construction

The evolutionary history was inferred using the Maximum Parsimony method. The MP tree was obtained using the Close-Neighbor-Interchange algorithm with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 20 amino acid sequences. There were a total of 229 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura et al. 2011).

Theoretical properties and molecular modeling

The theoretical physicochemical properties (pI, Mw and molar extinction coefficient) of the amino acid sequence corresponding to the putative cysteine protease deduced from cDNA were predicted using the software available on the ExPASy (ProtParam and ScanSite). Tryptic digestion in silico was performed using the Sequence Editor 3.1 (Bruker Daltonics) to obtain the theoretical PMF.

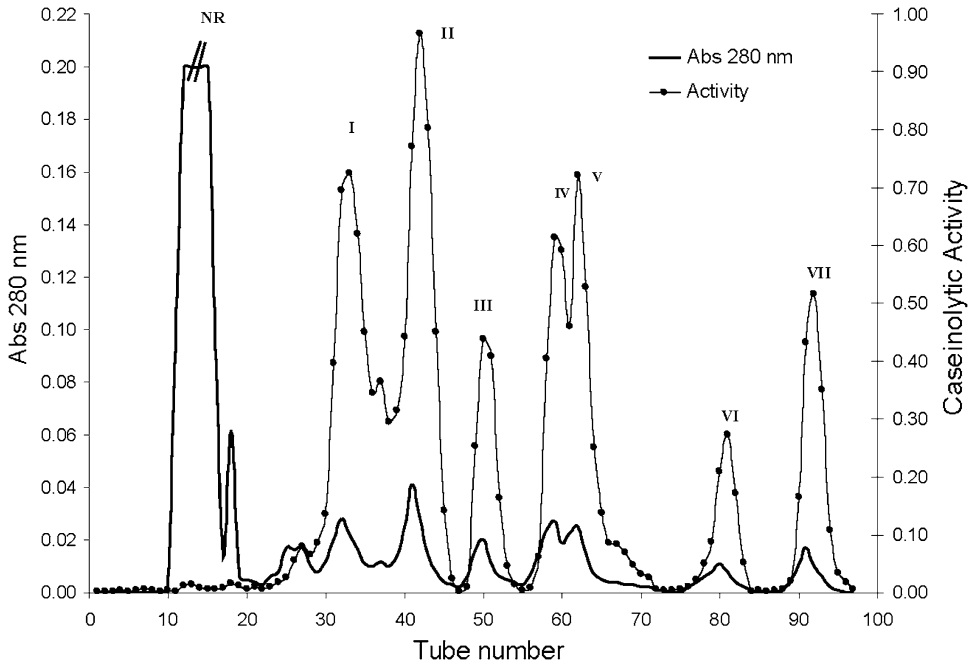


Fig. 1 Cation exchange chromatography of VQ preparation. Column: SP-Sepharose HP; starting buffer: Tris–HCl pH 7.5 containing 1 mM sodium tetrathionate and 0.15 M NaCl; linear gradients: 0.15–0.45 M NaCl (80 ml) and 0.45–1.0 M NaCl (20 ml); flow rate: 1 ml/min

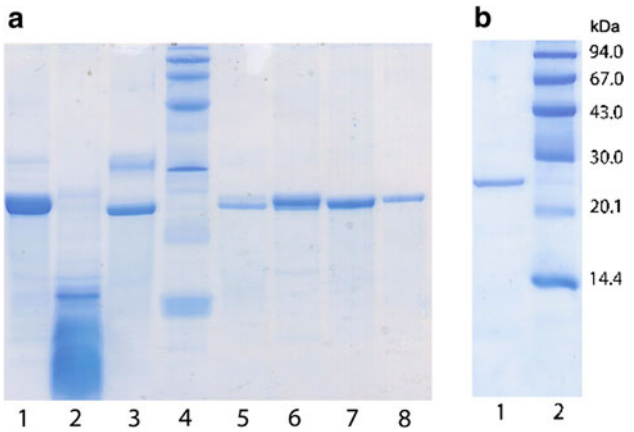


Fig. 2 SDS-PAGE of the fractions obtained from cation exchange chromatography. **a** *Lane 1* VQ-I fraction, *lane 2* VQ-II fraction, *lane 3* VQ-III fraction, *lane 4* molecular weight markers, *lane 5* VQ-IV fraction, *lane 6* VQ-V fraction, *lane 7* VQ-VI fraction, *lane 8* VQ-VII fraction. **b** *Lane 1* inhibited VQ-II fraction, *lane 2* molecular weight markers

The tertiary structure of the putative protease was predicted using the 3D-JIGSAW network service ("http://bmm.cancerresearchuk.org/~3djigsaw/"), a homology modeling program (Bates and Sternberg 1999; Bates et al. 2001; Contreras-Moreira and Bates 2002), using papain (PDB ID: 1PPN) of latex from *Carica papaya* as a suitable template. The graphic image was produced by PyMOL (DeLano Scientific LLC, Palo Alto, CA, USA).

Results and discussion

A partially purified preparation, named VQ devoid of gums and other insoluble materials, was obtained from latex of unripe fruits of *Vasconcellea quercifolia*. VQ showed several active fractions, all of which were basic proteins, when a sample of this preparation was submitted to iso-electrofocusing–zymogram (3–10 pH gradient) and 2D-PAGE using immobiline 7–11, as previously reported (Torres et al. 2010). The presence of several basic cysteine proteases in the lattice of one plant species is a frequent event (Dubey and Jagannadham 2003; Domsalla and Melzig 2008), which has also been found in other species belonging to the Caricaceae family (*Carica papaya*, *Vasconcellea cundinamarcensis*, *Vasconcellea* × *heilbornii* and *Vasconcellea stipulata*).

Purification of the VQ preparation proteases

According to the basic pI values of the proteolytic components, the VQ preparation was purified by cation exchange chromatography (SP-Sepharose High Performance). The main achievement in the chromatographic procedure was to separate all fractions in a single chromatographic step. The elution profile revealed the presence of seven caseinolytic active fractions (VQ-I, VQ-II, VQ-III, VQ-IV, VQ-V, VQ-VI and VQ-VII) which represented about 30 % of total protein, while no proteolytic activity (Fig. 1) was detected in the unretained fraction (NR). The SDS-PAGE analysis of the active fractions (Fig. 2a)

showed in all cases the presence of a band with molecular weight ca. 26 kDa, characteristic of plant cysteine proteases (Sequeiros et al. 2005). As shown in Fig. 2a, the most active fraction (II) presented an inconspicuous band of ca. 26 kDa and multiple bands with smaller molecular mass, suggesting a typical self-degradation process; this behavior could be due to the highest activity of this fraction compared with the other fractions, as could be seen in the chromatographic profile (Fig. 1). In a subsequent experiment, fraction II was collected directly from the chromatographic column in 1 mM iodoacetic acid. This treatment allowed the inhibition of its autodigestion (Fig. 2b).

Structural characterization of the proteases

N-terminal sequence

The N-terminal sequence of each protease was determined by Edman's automated degradation and the sequences obtained were compared among them (Table 1). These results indicate that at least five out of the seven peptides isolated from *V. quercifolia* latex are structurally different, because the eight residues determined for the VQ-II sequence matched with the first eight residues of the VQ-I sequence and the 15 residues of the VQ-VI N-terminal sequence matched with the first 15 residues of VQ-VII.

The seven sequences showed the presence of highly conserved residues and motifs of "papain-like" cysteine proteases: the Pro2 residue and the DWR and GAV motifs. Additionally, the Ser4 residue (position 5 for VQ-V) characteristic of cysteine proteases of the Caricaceae family was conserved. The Gln9 residue (Q) and a basic amino acid located at position 10 (R or K) appeared in all sequences (except for VQ-II, which is shorter).

We also determined the degree of identity of the N-terminal sequence of each protease of *V. quercifolia* with all the reported cysteine proteases of the Caricaceae family; a high degree of identity was found with those present in the latex from other species of the genera *Vasconcellea*, *Jacaratia* and *Carica* (Table 2).

The proteases VQ-I, VQ-II, and VQ-V presented a very similar N-terminal sequence, particulary characterized by the Tyr residue at the N-terminus (Table 1). We named this set of proteases as "group A". When the N-terminals of this group were compared with those of other proteases of the same family (Caricaceae), VQ-I had the highest identity with VS-B (*V. stipulata*) and VXH-C (*V.* × *heilbornii*), while VQ-II showed the highest identity with VXH-C, and VQ-V showed the highest identity with VS-B. On the other hand, the proteases VQ-III, VQ-IV, VQ-VI, and VQ-VII ("group B") had Ile as the N-terminal amino acid (Table 1) and presented a high degree of identity between them as well as with VS-A (*V. stipulata*), VXH-A, and VXH-B (*V.* × *heilbornii*).

Peptide mass fingerprinting (PMF)

Using a useful proteomic tool (PMF), adequate to differentiate isoenzymes present in latex (Obregón et al. 2009), we were able to differentiate all the proteases isolated from the latex of *V. quercifolia*, including those with sequences N-terminal partially matching (VQ-I/VQ-II and VQ-VI/VQ-VII). Electrophoretic spots ca. 26 kDa (SDS-PAGE, Fig. 2) were selected to perform tryptic digestion and the peptides were analyzed by MALDI-TOF MS (Fig. 3). The comparison of the peptide maps revealed the presence of seven different proteases. Furthermore, comparison of the tryptic maps with those deposited in databases by using the

Table 1 Comparison of N-terminal sequences of the proteases present in *V. quercifolia* latex

Protease	N-terminal sequence	Identity %						
		VQ-I	VQ-II	VQ-III	VQ-IV	VQ-V	VQ-VI	VQ-VII
VQ-I [*V. quercifolia*]	YPE–SVDWRQKGAVTPVKNQNRCGSCWA	**100**						
VQ-II [*V. quercifolia*]	YPE–SVDWR	**100**	**100**					
VQ-III [*V. quercifolia*]	IPA–SIDWRQRGAVTPVRH	67	62	**100**				
VQ-IV [*V. quercifolia*]	IPA–SIDWRQRGAVVPD	62	62	**88**	**100**			
VQ-V [*V. quercifolia*]	YPEDSIDWRQK	**82**	78	64	64	**100**		
VQ-VI [*V. quercifolia*]	IPA–SIDWRQKGAVTP	80	62	**93**	**87**	73	**100**	
VQ-VII [*V. quercifolia*]	IPA–SIDWRQKGAVTPIRLQGQ	62	62	**83**	**81**	73	**100**	**100**

The conserved residues are shaded. The identity % was calculated between pairs of sequences according to the sequence with the lowest number of residues determined in each case. Numbers in bold correspond to the enzyme sequences with percent of identity higher than 80 %

Table 2 Comparison of N-terminal sequences of the proteases present in *V. quercifolia* latex with other proteases of the family Caricaceae

Protease	N-terminal sequence	Reference	Identity %						
			VQ-I	VQ-II	VQ-III	VQ-IV	VQ-V	VQ-VI	VQ-VII
VS-A [*V. stipulata*]	IPA-SIDWRQKGAVTPVRNQGSCGSCWT	Kyndt et al. (2007)	74	62	89	81	73	100	86
VS-B [*V. stipulata*]	YPE-SIDWRQKGAVTPVKNQNPCGSCWA	Kyndt et al. (2007)	93	88	72	69	91	87	67
VXH-A [*V.* × *heilbornii*]	IPA-SIDWRQKGAVTPVRNQGSCGSCWT	Kyndt et al. (2007)	74	62	89	81	73	100	86
VXH-B [*V.* × *heilbornii*]	IPA-SIDWRQKGAVTPVRHQGSCGSCWT	Kyndt et al. (2007)	70	62	94	81	73	100	86
VXH-C [*V.* × *heilbornii*]	YPE-SVDWRQKGAVTPVKDQNPCGSCWA	Kyndt et al. (2007)	93	100	67	62	82	80	62
VXH-D [*V.* × *heilbornii*]	IPT-SIDWRQKGAVTPVRNQGGCGSCWT	Kyndt et al. (2007)	74	62	83	75	73	93	81
CC-I [*V. candamarcensis*]	IVA-SIDWRQKGAVTPVRNQQSCGSCWT	Walreavens et al. (1993)	70	62	83	75	64	93	81
CC-II [*V. candamarcensis*]	YPG-SVDWRQKGAVTPVKDQNPCGSCWA	Walreavens et al. (1993)	89	88	67	62	73	80	62
CC-III [*V. candamarcensis*]	YPE-SIDWRKKGAVTPVKNQGSCGSCWA	Walreavens et al. (1993)	85	88	67	62	82	80	67
CC-IV [*V. candamarcensis*]	YPE-SIDWRKKGAVTPVKNQGSCGSCWA	Walreavens et al. (1993)	85	88	67	62	82	80	67
CMS1MS2 [*V. candamarcensis*]	IPT-SIDWRQKGAVTPVRNQGGCGSCWT	Gomes et al. (2008)	74	62	83	75	73	93	81
Mexicain [*J. mexicana*]	YPE-SIDWREKGAVTPVKNQNPCGSCWA	Oliver-Salvador et al. (2004)	89	88	67	62	82	80	62
Chymomexicain [*J. mexicana*]	YPE-SIDWRDKGAVTPVKNQNPCGSCWA	Lian (1999)	89	88	67	62	82	80	62
Papain [*C. papaya*]	IPE-YVDWRQKGAVTPVKNQGSCGSCWA	Mitchel et al. (1970)	85	75	67	62	64	80	67
Chymopapain [*C. papaya*]	YPQ-SIDWRAKGAVTPVKNQGACGSCWA	Watson et al. (1990)	81	75	67	62	73	80	67
Caricain [*C. papaya*]	LPE-NVDWRKKGAVTPVRHQGSCGSCWA	Dubois et al. (1988)	74	75	67	50	54	67	62
Glycylendopeptidase [*C. papaya*]	LPE-SVDWRAKGAVTPVKHQGYCESCWA	Ritonja et al. (1989)	78	88	67	56	64	73	62

The identity % were calculated in pairs according to the sequence with the lowest number of residues determined in each case

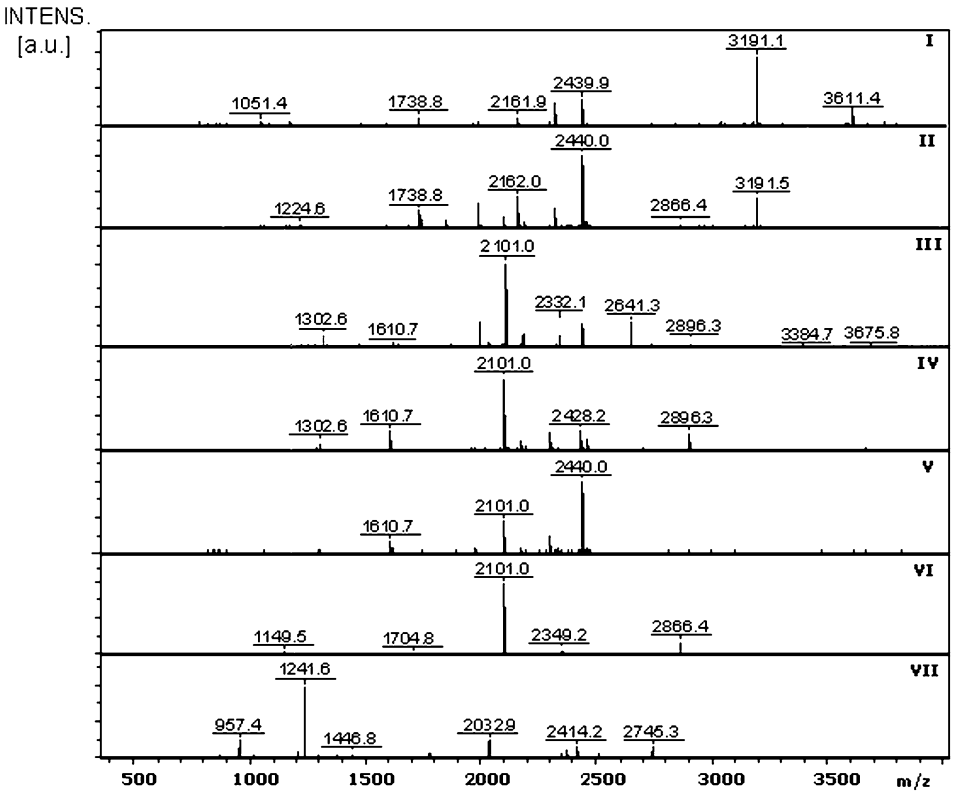


Fig. 3 Mass spectra corresponding to the PMF of the proteases isolated by SP-Sepharose HR chromatography

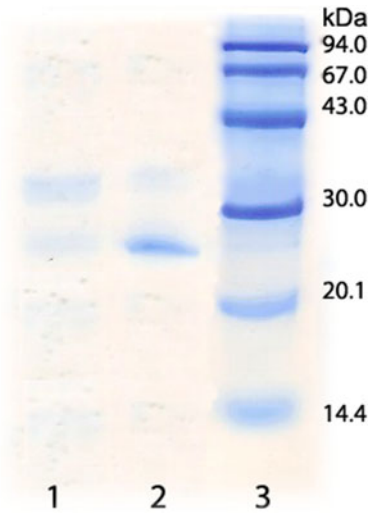


Fig. 4 SDS-PAGE of proenzyme activation. *Lane 1* fraction III, *lane 2* fraction III after incubation at 60 °C for 10 min with 100 μM pepsin in buffer pH 4.0, *lane 3* molecular weight markers

MASCOT tool ("http://www.matrixscience.com") showed that none of the isolated proteases matched with other plant protease. It is worth mentioning that we have previously purified and characterized a protease named quercifoliain I from *V. quercifolia* latex; we can infer that this enzyme could correspond to VQ-II according to some experimental data: the ionic strength which is eluted using cation exchange chromatography (SP-Sepharose), the N-terminal sequence, the high caseinolytic activity, and the comparison of PMFs (Torres et al. 2010).

Additionally, we determined that the proteases present in the latex of *V. quercifolia* are not glycoproteins like those of *Carica papaya* endopeptidases, CC III (*Carica candamarcensis*) being the only glycosylated peptidase reported from latex of Caricaceae family up to date (Jaziri et al. 1994).

Proenzyme detection

Since endopeptidases constitute a potential danger for the plant, they are stored in the laticifers as inactive proforms that rapidly convert into active mature enzymes after the release of latex from the plant. This explains why these proforms have never been isolated from papaya latex (Azarkan et al. 2003). To generate the proteolytically active endopeptidases, the region of the propeptide must be cleaved through an enzymatic digestion process (Brömme et al. 2004). In the case of *V. quercifolia* latex, fraction III obtained from the chromatography of VQ preparation (Fig. 1) when analyzed by SDS-PAGE (Fig. 4, lane 1) revealed that it was composed of two polypeptides (26 and 31 kDa). The treatment of fraction III with pepsin increased its caseinolytic activity (40 %) and the band of 26 kDa was intensified (Fig. 4, lane 2), suggesting the release of the

mature enzyme from its proenzyme. PMF analysis from the SDS-PAGE band allowed strengthening the idea that the 31-kDa polypeptide corresponds to the proenzyme. When comparing the PMF of the band of 31 kDa with the PMF obtained by tryptic in silico digestion of the propapain, the presence of a peptide (2,492 Da) corresponding to the papain propeptide (KNNSYWLGLNVFADMSNDEFK) was identified. This peptide includes the highly conserved heptapeptide GxNxFxD (GNFD in short form) characteristic of the subfamily C1A proenzymes. Further, the polypeptide with an *m*/*z* value of 2,508 was identified as the same sequence, but with the methionine oxided (MSO). Besides, when the PMF of the 31-kDa polypeptide was compared with the PMFs corresponding to all the isolated peptidases, six peptides matched with VQ-III (Table 1, Supplementary material). The results obtained constitute the first report of a propeptidase isolated from plant latex.

Molecular cloning of a putative protease

cDNA encoding cysteine proteases was synthesized by RT-PCR from total RNA extracted from *Vasconcellea quercifolia* latex and cloned in *E. coli*. A degenerate oligonucleotide primer Nt3-11Af (5′-GATTCnGTwGATTG GmGrGAAAAAGG-3′) for cysteine protease of latex (Trejo et al. 2009) was used. This primer contains the sequence coding for the DWR motif, characteristic of papain-like cysteine proteases. The sequences obtained from the selected clones were analyzed and the consensus sequence was obtained using the CLUSTAL X multiple sequence alignment program (Fig. 5). This sequence was composed of 819 nucleotides, of which the former 633 (including 26 nucleotides corresponding to the used primer) encoded for a putative cysteine protease sequence, followed by the "TAA" stop codon, the polyadenylation signal "aataaa", and the polyadenine tail composed of 40 residues.

The corresponding polypeptide sequence contains seven cysteine residues characteristic of cysteine phytopeptidases (Napper et al. 1994). The amino acid residues characteristic of the active site of this type of enzymes: Cys25, His159, Asn175, and Gln 19 (papain numbering) can be observed in the sequence (Fig. 5). The amino acid sequence deduced for the putative protease was subjected to tryptic digestion

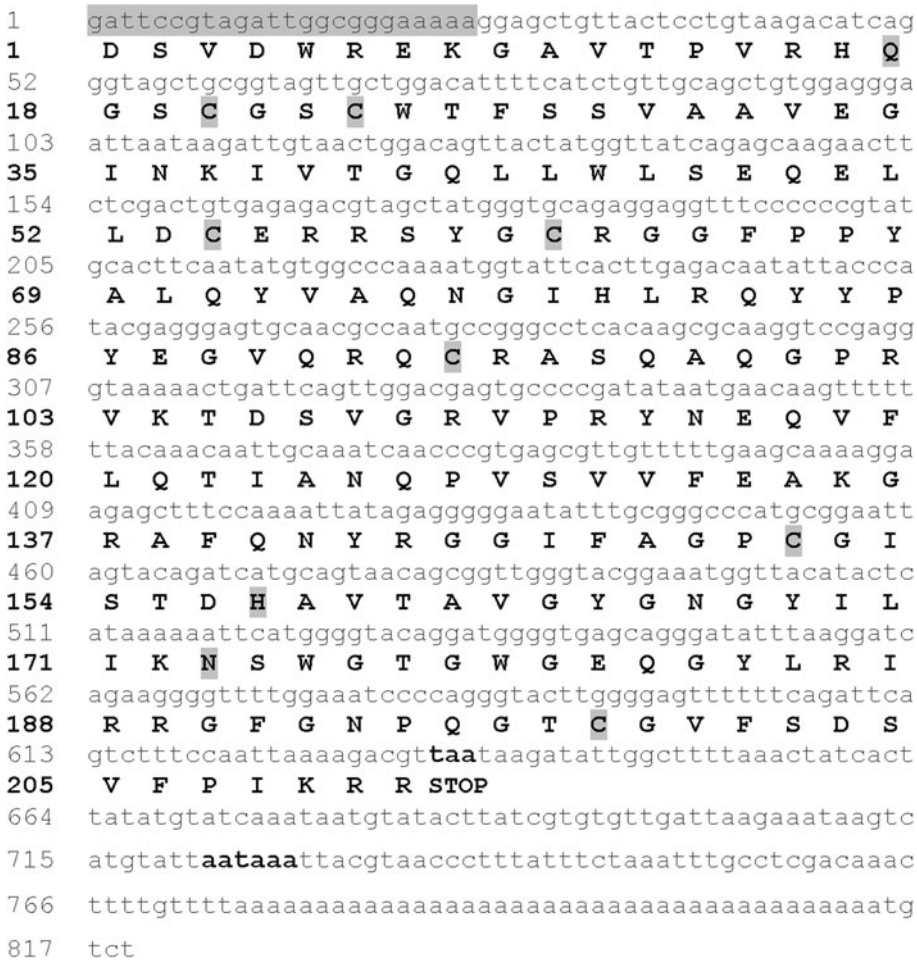
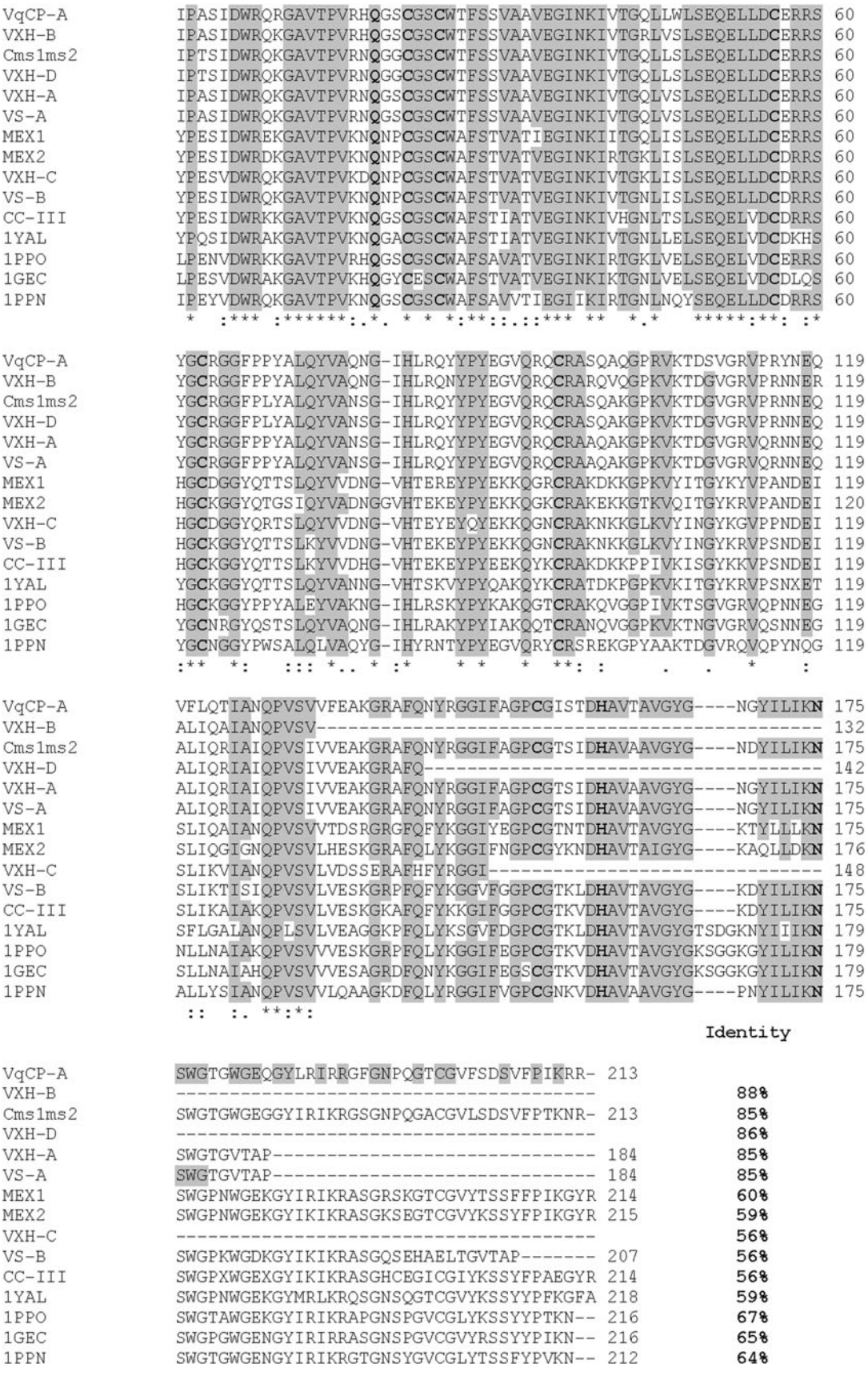


Fig. 5 Putative cysteine protease consensus sequence. The nucleotide sequence shaded in gray corresponds to the sequence of Nt3-11Af forward primer used. In the nucleotide sequence highlighted the triplet "taa" corresponded to the stop signal for protein synthesis and the sequence "AATAAA" corresponded to the polyadenylation signal sequence. The amino acid residues of catalytic site and the cysteines are shaded in *gray*


VqCP-A IPASIDWRQRGAVTPVRHQGSCGSCWTFSSVAAVEGINKIVTGQLLWLSEQELLDCERRS 60
VXH-B IPASIDWRQKGAVTPVRHQGSCGSCWTFSSVAAVEGINKIVTGRLVSLSEQELLDCERRS 60
Cms1ms2 IPTSIDWRQKGAVTPVRNQGGCGSCWTFSSVAAVEGINKIVTGQLLSLSEQELLDCERRS 60
VXH-D IPTSIDWRQKGAVTPVRNQGGCGSCWTFSSVAAVEGINKIVTGQLLSLSEQELLDCERRS 60
VXH-A IPASIDWRQKGAVTPVRNQGSCGSCWTFSSVAAVEGINKIVTGQLVSLSEQELLDCERRS 60
VS-A IPASIDWRQKGAVTPVRNQGSCGSCWTFSSVAAVEGINKIVTGQLVSLSEQELLDCERRS 60
MEX1 YPESIDWREKGAVTPVKNQNPCGSCWAFSTVATIEGINKIITGQLISLSEQELLDCERRS 60
MEX2 YPESIDWRDKGAVTPVKNQNPCGSCWAFSTVATVEGINKIRTGKLISLSEQELLDCDRRS 60
VXH-C YPESVDWRQKGAVTPVKDQNPCGSCWAFSTVATVEGINKIVTGKLISLSEQELLDCDRRS 60
VS-B YPESIDWRQKGAVTPVKNQNPCGSCWAFSTVATVEGINKIVTGNLISLSEQELLDCDRRS 60
CC-III YPESIDWRKKGAVTPVKNQGSCGSCWAFSTIATVEGINKIVHGNLTSLSEQELVDCDRRS 60
1YAL YPQSIDWRAKGAVTPVKNQGACGSCWAFSTIATVEGINKIVTGNLLELSEQELVDCDKHS 60
1PPO LPENVDWRKKGAVTPVRHQGSCGSCWAFSAVATVEGINKIRTGKLVELSEQELVDCERRS 60
1GEC LPESVDWRAKGAVTPVKHQGYCESCWAFSTVATVEGINKIKTGNLVELSEQELVDCDLQS 60
1PPN IPEYVDWRQKGAVTPVKNQGSCGSCWAFSAVVTIEGIIKIRTGNLNQYSEQELLDCDRRS 60
VqCP-A YGCRGGFPPYALQYVAQNG-IHLRQYYPYEGVQRQCRASQAQGPRVKTDSVGRVPRYNEQ 119
VXH-B YGCRGGFPPYALQYVAQNG-IHLRQNYPYEGVQRQCRARQVQGPKVKTDGVGRVPRNNER 119
Cms1ms2 YGCRGGFPLYALQYVANSG-IHLRQYYPYEGVQRQCRASQAKGPKVKTDGVGRVPRNNEQ 119
VXH-D YGCRGGFPLYALQYVANSG-IHLRQYYPYEGVQRQCRASQAKGPKVKTDGVGRVPRNNEQ 119
VXH-A YGCRGGFPPYALQYVANSG-IHLRQYYPYEGVQRQCRAAQAKGPKVKTDGVGRVQRNNEQ 119
VS-A YGCRGGFPPYALQYVANSG-IHLRQYYPYEGVQRQCRAAQAKGPKVKTDGVGRVQRNNEQ 119
MEX1 HGCDGGYQTTSLQYVVDNG-VHTEREYPYEKKQGRCRAKDKKGPKVYITGYKYVPANDEI 119
MEX2 HGCKGGYQTGSIQYVADNGGVHTEKEYPYEKKQGKCRAKEKKGTKVQITGYKRVPANDEI 120
VXH-C HGCDGGYQRTSLQYVVDNG-VHTEYEYQYEKKQGNCRAKNKKGLKVYINGYKGVPPNDEI 119
VS-B HGCKGGYQTTSLKYVVDNG-VHTEKEYPYEKKQGNCRAKNKKGLKVYINGYKRVPSNDEI 119
CC-III HGCKGGYQTTSLKYVVDHG-VHTEKEYPYEEKQYKCRAKDKKPPIVKISGYKKVPSNDEI 119
1YAL YGCKGGYQTTSLQYVANNG-VHTSKVYPYQAKQYKCRATDKPGPKVKITGYKRVPSNXET 119
1PPO HGCKGGYPPYALEYVAKNG-IHLRSKYPYKAKQGTCRAKQVGGPIVKTSGVGRVQPNNEG 119
1GEC YGCNRGYQSTSLQYVAQNG-IHLRAKYPYIAKQQTCRANQVGGPKVKTNGVGRVQSNNEG 119
1PPN YGCNGGYPWSALQLVAQYG-IHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNQG 119
VqCP-A VFLQTIANQPVSVVFEAKGRAFQNYRGGIFAGPCGISTDHAVTAVGYG----NGYILIKN 175
VXH-B ALIQAIANQPVSV---------- 132
Cms1ms2 ALIQRIAIQPVSIVVEAKGRAFQNYRGGIFAGPCGTSIDHAVAAVGYG----NDYILIKN 175
VXH-D ALIQRIAIQPVSIVVEAKGRAFQ---------- 142
VXH-A ALIQRIAIQPVSIVVEAKGRAFQNYRGGIFAGPCGTSIDHAVAAVGYG----NGYILIKN 175
VS-A ALIQRIAIQPVSIVVEAKGRAFQNYRGGIFAGPCGTSIDHAVAAVGYG----NGYILIKN 175
MEX1 SLIQAIANQPVSVVTDSRGRGFQFYKGGIYEGPCGTNTDHAVTAVGYG----KTYLLLKN 175
MEX2 SLIQGIGNQPVSVLHESKGRAFQLYKGGIFNGPCGYKNDHAVTAIGYG----KAQLLDKN 176
VXH-C SLIKVIANQPVSVLVDSSERAFHFYRGGI---------- 148
VS-B SLIKTISIQPVSVLVESKGRPFQFYKGGVFGGPCGTKLDHAVTAVGYG----KDYILIKN 175
CC-III SLIKAIAKQPVSVLVESKGKAFQFYKKGIFGGPCGTKVDHAVTAVGYG----KDYILIKN 175
1YAL SFLGALANQPLSVLVEAGGKPFQLYKSGVFDGPCGTKLDHAVTAVGYGTSDGKNYIIIKN 179
1PPO NLLNAIAKQPVSVVVESKGRPFQLYKGGIFEGPCGTKVDHAVTAVGYGKSGGKGYILIKN 179
1GEC SLLNAIAHQPVSVVVESAGRDFQNYKGGIFEGSCGTKVDHAVTAVGYGKSGGKGYILIKN 179
1PPN ALLYSIANQPVSVVLQAAGKDFQLYRGGIFVGPCGNKVDHAVAAVGYG----PNYILIKN 175
Identity
VqCP-A SWGTGWGEQGYLRIRRGFGNPQGTCGVFSDSVFPIKRR- 213
VXH-B ---------- 88%
Cms1ms2 SWGTGWGEGGYIRIKRGSGNPQGACGVLSDSVFPTKNR- 213 85%
VXH-D ---------- 86%
VXH-A SWGTGVTAP---------- 184 85%
VS-A SWGTGVTAP---------- 184 85%
MEX1 SWGPNWGEKGYIRIKRASGRSKGTCGVYTSSFFPIKGYR 214 60%
MEX2 SWGPNWGEKGYIKIKRASGKSEGTCGVYKSSYFPIKGYR 215 59%
VXH-C ---------- 56%
VS-B SWGPKWGDKGYIKIKRASGQSEHAELTGVTAP------- 207 56%
CC-III SWGPXWGEXGYIKIKRASGHCEGICGIYKSSYFPAEGYR 214 56%
1YAL SWGPNWGEKGYMRLKRQSGNSQGTCGVYKSSYYPFKGFA 218 59%
1PPO SWGTAWGEKGYIRIKRAPGNSPGVCGLYKSSYYPTKN-- 216 67%
1GEC SWGPGWGENGYIRIRRASGNSPGVCGVYRSSYYPIKN-- 216 65%
1PPN SWGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN-- 212 64%

◀ **Fig. 6** Alignment of the VqCP-A sequence with those of a group of plant proteases belonging to the subfamily C1A. The VqCP-A sequence was aligned with the sequences of *Vasconcellea* × *heilbornii* (VXH A, VXH B, VXH C, VXH D), *Vasconcellea stipulata* (VS-A, VS-B), *Carica candamarcensis* (Cms1ms2, CC-III), *Jacaratia mexicana* (MEX1, MEX2), and *Carica papaya*: chymopapain (1YAL), caricain (1PPO), glycyl endopeptidase (1GEC) and papain (1PPN)

("in silico") and the masses of the peptides obtained were compared with those observed in the PMF of each of the proteases isolated chromatographically. The VQ-III peptidase showed 11 peptides matching with the cloned protease (tolerance 0.085 Da), representing 56 % of the enzyme sequence and covering 72 % of the total intensity of the mass spectrum. These results strongly suggest that the sequence obtained corresponds to the VQ-III peptidase.

Further, this analysis allowed determining that the sequence of a peptide (mass 2,101 Da, GGFPPYALQY-VAQNGIHLR) was highly conserved in most *Vasconcellea quercifolia* peptidases (VQ-II, VQ-III, VQ-IV, VQ-V and VQ-VI), as well as in the detected proenzyme and VXH-B protease. The sequence proposed for the mature enzyme was built replacing the amino acids corresponding to the primer for the first ten residues coming from the N-terminus sequence of VQ-III and was named VqCP-A.

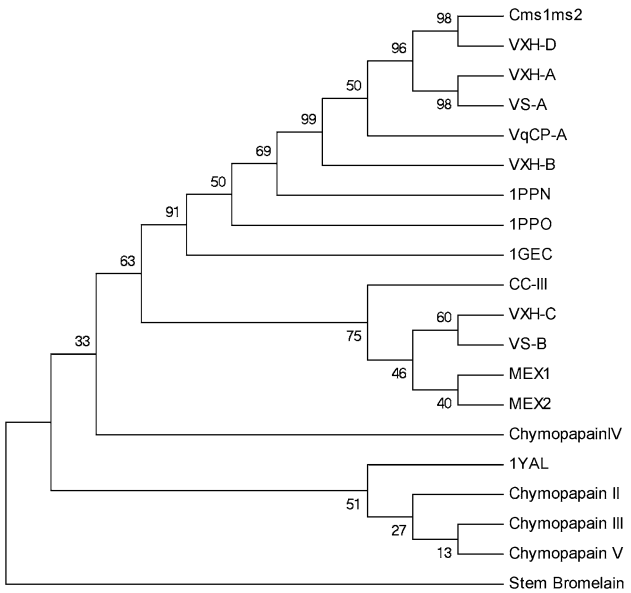


Fig. 7 Phylogenetic tree of cysteine protease sequences present in species of the Caricaceae family. Tree 1 out of four most parsimonious trees (length = 476) is shown. The consistency index is (0.743073), the retention index is (0.798419), and the composite index is 0.627329 (0.593284) for all sites and parsimony-informative sites (in parentheses) The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches

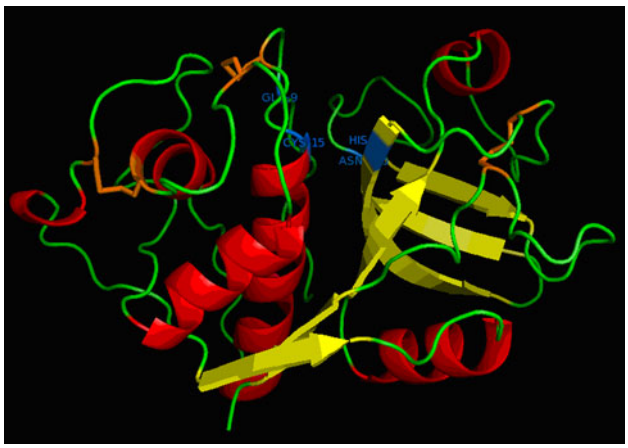


Fig. 8 Predicted molecular model of VqCP-A. Antiparallel β-sheet structures are represented in *yellow*, the α helices in *red*, the catalytic site residues in blue and the cysteines that form disulfide bridges in orange. http://bmm.cancerresearchuk.org/~3djigsaw/job-ref is 573 73092

The VqCP-A sequence is composed of 213 amino acids (from amino acid 11 to 213 derived from the cDNA sequence).

The physicochemical properties calculated for VqCP-A were molecular mass (23,449 Da), pI (9.43) and molar extinction coefficient at 280 nm (49,390 M^{-1} cm^{-1}). The calculated pI value agrees with the results from isoelectric focusing of the VQ preparation, where the alkaline bands (pI > 9.3) show proteolytic activity (Torres et al. 2010).

The analysis of the VqCP-A sequence by the BLAST algorithm revealed high degree of identity (54–88 %) with 13 cysteine proteases of the C1A family, including papain. The alignments of VqCP-A with the complete sequences of other cysteine proteases belonging to species of the Caricaceae family are shown in Fig. 6. VXH-B showed the highest identity (88 %), followed by VXH-D (86 %), and VXH-A, VS-A and CMS1MS2 with 85 % of identity (Kyndt et al. 2007; Gomes et al. 2008).

The amino acid sequence of VqCP-A contained the characteristic features of all "papain-like" cysteine proteinases (Fig. 6). A high conservation degree was observed particularly for those amino acid residues 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 Asn 175, characteristic of Clan CA. Phe141, Trp177 and Trp 181, involved in the hydrophobic pocket of the catalytic site (Barrett et al. 2004a), are also preserved. On the basis of the aforementioned information, VqCP-A should be included in the cysteine peptidase subfamily C1A. The partial sequence of this enzyme was submitted to the EMBL nucleotide sequence database (Accession Number HE613670).

Molecular phylogeny

The evolutionary relationships between VqCP-A and other cysteine proteases present in the species of the family Caricaceae were derived using the maximum parsimony method (Fig. 7). The amino acid sequence of VqCP-A along with the proteases VXK-B, VXH-A, VXH-D (*V.* × *heilbornii*), VS-A (*V. stipulata*), and Cms1ms2 (*V. cundinamarcensis*), form a group with a high bootstrap value, revealing a close phylogenetic relationship. This group is closer to papain, suggesting them to be papain homologues, while in other cluster are the chymopapain (1 YAL) and its isoforms, in accordance to the report of Kyndt et al. (2007). Stem bromelain was chosen as the outgroup because of its low degree of similarity with the other sequences.

Tertiary structure of VqCP-A predicted by molecular modeling

The predicted molecular model of VqCP-A is shown in Fig. 8. The enzyme structure shows the typical papain-like fold composed of two domains separated by a groove containing the active site. The L domain (α-helix-rich) contains the catalytic residues Gln19 and Cys25 and it is stabilized by two disulfide bridges (Cys22–Cys63 and Cys56–Cys95), and the R domain (β-barrel-like) contains the catalytic residues His159 and Asn175 and it is stabilized by another disulfide bridge (Cys153–Cys200).

The specificity subsite that is dominant in most peptidases of subfamily C1A is the S2 subsite, which commonly displays a preference for occupation by a bulky hydrophobic side chain and not a charged one (Barrett et al. 2004b). In papain, the S2 subsite is essentially composed of residues Tyr67, Pro68, Trp69, Val133, Val157, Ser205, and Phe207 (Maes et al. 1996), while in VqCP-A the S2 pocket (based on a model in which VqCP-A was superimposed on that of papain) is composed of Phe67, Pro68, Pro69, Val133, Thr157, Asp205, and Val207 (papain numbering). The comparison of S2 subsites revealed that the residues involved in the S2 pocket of VqCP-A are mainly hydrophobic, as papain, but, unusually, in VqCP-A there is an acidic residue (Asp) instead of the hydrophilic residue (Ser) of papain. This replacement also occurs in *Entamoeba* sp., stem bromelain, and lobster digestive endopeptidase, while in cathepsin B and other enzymes this Ser is replaced by Glu. These changes can be explained by a distinctive specificity (Barrett et al. 2004b).

The data obtained provide an important contribution to the biochemical and structural knowledge of the proteolytic system present in the latex of a species belonging to the family Caricaceae. Additionally, the presence of a propeptidase in plant latex was detected for the first time. The isolated enzymes are similar to those obtained from related species, but at the same time showed specific characteristics that allowed their identification as individual chemical entities. Finally, the use of proteomics tools allowed to correlate the sequence of the cloned enzyme (VqCP-A) with one of the proteases (VQ-III peptidase) purified from latex.

Acknowledgments The present work was supported by grants from ANPCyT, CONICET (PIP 0477), UNLP and CIC (Argentina), CYTED (Project IV.22) and BIO2007-68046-C02 (MEC, Spain). M.J. Torres is a CONICET fellow, L.M.I. López and W.D. Obregón belong to the CONICET Researcher Career, and C.L. Natalucci belongs to the CICPBA Researcher Career. The MALDI-TOF MS experiments were carried out in the proteomics facility (SePBioEs) from IBB-UAB, a member of ProteoRed-ISCIII. We are grateful to Prof. Pedro Balatti (Instituto de Fisiología Vegetal, UNLP-CONICET) for assisting in the phylogenetic analysis.

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(17):3389–3402

Azarkan M, El Moussaoui A, van Wuytswinkel D, Dehon G, Looze Y (2003) Fractionation and purification of the enzymes stored in the latex of *Carica papaya*. J Chromatogr B 790:229–238

Badillo VM (2000) *Carica L.* vs. *Vasconcella* St. Hil. (*Caricaceae*) con la rehabilitación de este último. Ernstia 10:74–79

Badillo VM (2001) Nota correctiva *Vasconcellea* St. Hil. y no *Vasconcella* (*Caricaceae*). Ernstia 11:75–76

Baeza G, Correa D, Salas CE (1990) Proteolytic enzymes in *Carica candamarcensis*. J Sci Food Agric 51:1–9

Barrett AJ, Rawlings ND, Woessner JF (2004a) Handbook of proteolytic enzymes, 2nd edn. Elsevier Academic Press, Amsterdam, pp 1125–1134

Barrett AJ, Rawlings ND, Woessner JF (2004b) Handbook of proteolytic enzymes, 2nd edn. Elsevier Academic Press, Amsterdam, pp 1051–1071

Bates PA, Sternberg MJE (1999) Model building by comparison at CASP3: using expert knowledge and computer automation. Proteins Suppl 3:47–54

Bates PA, Kelley LA, MacCallum RM, Sternberg MJE (2001) Enhancement of protein modelling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins Suppl 5:39–46

Bravo LM, Hermosilla L, Salas CE (1994) A biochemical comparison between latex from *C. candamarcensis* and *C. papaya*. Braz J Med Biol Res 27:2831–2842

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

Contreras-Moreira B, Bates PA (2002) Domain fishing: a first step in protein comparative modelling. Bioinformatics 18:1141–1142

Corrêa NCR, Mendes IC, Gomes MTR, Kalapothakis E, Chagas BCA, Lopes MTP, Salas CE (2011) Molecular cloning of a mitogenic proteinase from *Carica candamarcensis*: its potential use in wound healing. Phytochemistry 72:1947–1954

Dimitri MJ (1979) Enciclopedia Argentina de Agricultura y Jardinería, 3ra edn, Tomo I, 2do vol. ACME, Buenos Aires, p 765

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

Dubey VK, Jagannadham MV (2003) Procerain, a stable cysteine protease from the latex of *Calotropis procera*. Phytochemistry 62:1057–1071

Dubois T, Kleinschmidt T, Schnek AG, Looze Y, Braunitzer G (1988) The thiol proteinases from the latex of Carica papaya L. II. The primary structure of proteinase omega. Biol Chem Hoppe-Seyler 369:741–754

Gavira JA, González-Ramírez LA, Oliver-Salvador MC, Soriano-García M, García-Ruiz JM (2007) Structure of the mexicain-E-64 complex and comparison with other cysteine proteases of the papain family. Acta Crystallogr D 63:555–563

Gomes MT, Mello VJ, Rodrigues KC, Bemquerer MP, Lopes MT, Faça VM, Salas CE (2005) Isolation of two plant proteinases in latex from *Carica candamarcensis* acting as mitogens for mammalian cells. Planta Med 71:244–248

Gomes MTR, Teixeira RD, Ribeiro HA, Turchetti AP, Junqueira CF, Lopes MT, Salas CE, Nagem RA (2008) Purification, crystallization and preliminary X-ray analysis of CMS1MS2: a cysteine proteinase from *Carica candamarcensis* latex. Acta Crystallogr F 64:492–494

Gomes MTR, Turchetti AP, Lopes MTP, Salas CE (2009) Stimulation of fibroblast proliferation by the plant cysteine protease CMS2MS2 is independent of its proteolytic activity and requires ERK activation. Biol Chem 390:1285–1291

Gomes MTR, Ribeiro HA, Lopes MTP, Guzman F, Salas CE (2010) Biochemical comparison of two proteolytic enzymes from *Carica candamarcensis*: structural motifs underlying resistance to cystatin inhibition. Phytochemistry 71:524–530

Hegnauer R (1964) Chemotaxonomie der Pflanzen, Band 3, Dicotyledonae: Acanthaceae-Cyrillaceae. Birkhauser Report, Basel, Stuttgart, Germany, p 743

Jaziri M, Kleinschmidt T, Walraevens V, Schnek AG, Looze Y (1994) Primary structure of CC-III, the glycosylated cysteine proteinase from the latex of Carica candamarcensis Hook. Biol Chem H-S 375:379–385

Kyndt T, Van Damme EJ, Van Beeumen J, Gheysen G (2007) Purification and characterization of the cysteine proteinases in the latex of *Vasconcellea* spp. FEBS J 274:451–462

Kyndt T, Van Droogenbroeck B, Romeijn-Peeters E, Romero-Motochi JP, Scheldeman X, Goetghebeur P, Van Damme P, Gheysen G (2005) Molecular phylogeny and evolution of Caricaceae based on rDNA Internal Transcribed Spacer (ITS) and chloroplast sequence data. Mol Phylogenet Evol 37:442–459

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

Lian Z (1999) Modification of calcite crystal morphology by designed phosphopeptides and primary structures and substrate specificities of the cysteine proteases mexicain and chymomexicain. PhD Thesis, University of Boston, MA, USA

Maes D, Bouckaert J, Poortmans F, Wyns L, Looze Y (1996) Structure of chymopapain at 1.7 Å resolution. Biochemistry 35:16292–16298

Mitchel RE, Chaiken IM, Smith EL (1970) The complete amino acid sequence of papain. Additions and corrections. J Biol Chem 245:3485–3492

Moutim V, Silva LG, Lopes MTP, Wilson FG, Salas CE (1999) Spontaneous processing of peptides during coagulation of latex from *Carica papaya*. Plant Sci 142:115–121

Napper AD, Bennett SP, Borowski M, Holdridge MB, Leonard MJ, Rogers EE, Duan Y, Laursen RA, Reinhold B, Shames SL (1994) Purification and characterization of multiple forms of the pineapple-stem-derived cysteine proteinases ananain and comosain. Biochem J 301:727–735

Natalucci CL, Brullo A, López LMI, Hilal RM, Caffini NO (1996) Macrodontin, a new protease isolated from fruits of *Pseudananas macrodontes* (Morr.) Harms (Bromeliaceae). J Food Biochem 19:443–454

Obregón WD, Liggieri CS, Morcelle SR, Trejo SA, Avilés FX, Priolo NS (2009) Biochemical and PMF MALDI-TOF analyses of two novel papain-like plant proteinases. Protein Pept Lett 16:1323–1333

Oliver-Salvador MC, Gonzalez-Ramirez LA, Gavira JA, Soriano-Garcia M, Garcia-Ruiz JM (2004) Purification, crystallization and preliminary X-ray analysis of mexicain. Acta Crystallogr D 60(PT 11):2058–2060

Otto HH, Schirmeister T (1997) Cysteine proteases and their inhibitors. Chem Rev 97:133–171

Pereira MT, Lopes MT, Meira WO, Salas CE (2001) Purification of a cysteine proteinase from *Carica candamarcensis* L. and cloning of a genomic putative fragment coding for this enzyme. Protein Expr Purif 22:249–257

Ritonja A, Buttle DJ, Rawlings ND, Turk V, Barrett AJ (1989) Papaya proteinase IV amino acid sequence. FEBS Lett 258:109–111

Salas CE, Gomes MT, Hernandez M, Lopes MT (2008) Plant cysteine proteinases: evaluation of the pharmacological activity. Phytochemistry 69:2263–2269

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

Scheldeman X, Willemen L, Coppens d'Eeckenbrugge G, Romeijn-Peeters E, Restrepo MT, Romero Motoche J, Jimenez D, Lobo M, Medina CI, Reyes C, Rodriguez D, Ocampo JA, Van Damme P, Goetghebeur P (2007) Distribution, diversity and environmental adaptation of highland papaya (*Vasconcellea* spp.) in tropical and subtropical America. Biodivers Conserv 16:1867–1884

Sequeiros C, Torres MJ, Trejo SA, Esteves JL, Natalucci CL, López LMI (2005) Philibertain g I, the most basic cysteine endopeptidase purified from the latex of *Philibertia gilliesii* Hook. et Arn. (Apocynaceae). Protein J 24:445–453

Silva LG, Garcia O, Lopes MTP, Salas CE (1997) Changes in protein profile during coagulation of latex from *Carica papaya*. Braz J Med Biol Res 30:615–619

Souza DP, Freitas CDT, Pereira DA, Nogueira FC, Silva FDA, Salas CE, Ramos MV (2011) Laticifer proteins play a defensive role against hemibiotrophic and necrotrophic phytopathogens. Planta 234:183–193

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Teixeira RD, Ribeiro HA, Gomes MT, Lopes MT, Salas CE (2008) The proteolytic activities in latex from *Carica candamarcensis*. Plant Physiol Biochem 46:956–961

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Trejo SA, López LMI, Caffini NO, Natalucci CL, Canals F, Avilés FX (2009) Sequencing and characterization of asclepain f: the first cysteine peptidase cDNA cloned and expressed from *Asclepias fruticosa* latex. Planta 230:319–328

Torres MJ, Trejo SA, Martin MI, Natalucci CL, Avilés FX, López LMI (2010) Purification and characterization of a cysteine endopeptidase from *Vasconcellea quercifolia* A. St.-Hil. latex displaying high substrate specificity. J Agric Food Chem 58: 11027–11035

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

Van Droogenbroeck B, Breyne P, Goetghebeur P, Romeijn-Peeters E, Kyndt T, Gheysen G (2002) AFLP analysis of genetic

relationships among papaya and its wild relatives (Caricaceae) from Ecuador. Theoret Appl Genet 105:289–297

Van Droogenbroeck B, Kyndt T, Maertens I, Romeijn-Peeters E, Scheldeman X, Romero-Motochi JP, Van Damme P, Goetghebeur P, Gheysen G (2004) Phylogenetic analysis of the highland papayas (*Vasconcellea*) and allied genera (Caricaceae) using PCR-RFLP. Theor Appl Genet 108:1473–1486

Walraevens V, Vandermeers-Piret M, Vandermeers A, Gourlet P, Robberecht P (1999) Isolation and primary structure of the CCI papain-like cysteine proteinases from the latex of Carica candamarcensis Hook. Biol Chem 380:485–488

Walreavens V, Jaziri M, van Beeumen J, Schnek AG, Kleinschmidt T, Looze Y (1993) Isolation and preliminary characterization of the cysteine-proteinases from the latex of *Carica candamarcensis* Hook. Biol Chem Hoppe-Seyler 374:501–506

Watson DC, Yaguchi M, Lynn KR (1990) The amino acid sequence of chymopapain from *Carica papaya*. Biochem J 266:75–81