Amino Acids (2012) 43:2381–2391
DOI 10.1007/s00726-012-1318-7

ORIGINAL ARTICLE

X-ray crystal structure of CMS1MS2: a high proteolytic activity cysteine proteinase from *Carica candamarcensis*

Marco T. R. Gomes · Raphael D. Teixeira · Míriam T. P. Lopes · Ronaldo A. P. Nagem · Carlos E. Salas

Received: 9 March 2012 / Accepted: 26 April 2012 / Published online: 19 May 2012

Abstract CMS1MS2 (CC-Ib) from *Carica candamarcensis* (*Vasconcellea cundinamarcensis*) is a cysteine proteinase found as a single polypeptide containing 213 residues of 22,991 Da. The enzyme was purified by three chromatographic steps, two of them involving cationic exchange. Crystals of CMS1MS2 complexed with E-64 were obtained by the hanging drop vapor-diffusion method at 291 K using ammonium sulfate and polyethylene glycol 4000/8000 as precipitant. The complex CMS1MS2-E-64 crystallized in the tetragonal space group $P4_12_12$ with unit-cell parameters; $a = b = 73.64$, $c = 118.79$ Å. The structure was determined by Molecular Replacement and refined at 1.87 Å resolution to a final R factor of 16.2 % ($R_{free} = 19.3$ %). Based on the model, the structure of CMS1MS2 (PDB 3IOQ) ranks as one of the least basic cysteine isoforms from *C. candamarcensis,* is structurally closer to papain, caricain, chymopapain and mexicain than to the other cysteine proteinases, while its activity is twice the activity of papain towards BAPNA substrate. Two differences, one in the S2 subsite and another in the S3 subsite of CMS1MS2 may contribute to the enhanced activity relative to papain. In addition, the model provides a structural basis for the sensitivity of CMS1MS2 to inhibition by cystatin, not shown by other enzymes of the group, e.g., glycyl endopeptidase and CMS2MS2.

Keywords *Carica candamarcensis* · Crystal structure · Cysteine protease · Latex · Proteinase · *Vasconcellea cundinamarcensis*

M. T. R. Gomes · R. D. Teixeira · R. A. P. Nagem · C. E. Salas (✉)
Departamento de Bioquímica e Imunologia,
Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais,
Av. Antônio Carlos 6627,
Belo Horizonte, MG 31270-901, Brazil
e-mail: cesbufmg@icb.ufmg.br

M. T. P. Lopes
Departamento de Farmacologia,
Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais,
Av. Antônio Carlos 6627,
Belo Horizonte, MG 31270-901, Brazil

Introduction

Plants store a variety of fluids, including latexes, resins, gums and mucilages within secretory cells and canals. In latex of *Carica papaya L.* a mixture of cysteine endopeptidases including papain (EC 3.4.22.2) (Mitchel et al. 1970), chymopapains A and B and other isoforms (3.4.22.6) (Watson et al. 1990), glycyl endopeptidase (3.4.22.25) (Barrett and Buttle 1985; Ritonja et al. 1989) and endopeptidase omega (caricain) (3.4.22.30) (Dubois et al. 1988) are transiently released after injury until a protein clot forms at the wounded site. The coagulation process is vital as defence ploy against possible pathogen attack. Earlier, we provided evidence that during latex coagulation in Caricaceae a number of peptides are proteolytically processed in a non-random manner (Silva et al. 1997; Moutim et al. 1999). These changes are concomitant with drastic variations in latex proteolytic activity, suggesting the involvement of these enzymes during clot formation. The situation strongly resembles blood coagulation and clot formation during wounding in mammals (Furie and Furie 1988). Increased amounts and/or enhanced activation of these enzymes occurs when the fruit is repeatedly wounded, suggesting an adaptation in stressed plants (Azarkan et al. 2006).

Latex of *Carica candamarcensis* (*Vasconcellea cundinamarcensis*), a cultivar of Caricaceae family common to many areas in South America contains cysteine proteinases involved in clot formation, as well. Interestingly, *C. candamarcensis* proteinases display higher (five to sevenfold) proteolytic activity compared with enzymes from *C. papaya* (Baeza et al. 1990; Bravo et al. 1994). The occurrence of stronger activities is not unique to this species, as elevated activities were reported in other South American species (*V. monoica*, *V. stipulate*, *V. heilbornii*) and might represent a regional adaptation of this genus (Kyndt et al. 2007). Although the latex components of *C. candamarcensis* are less studied, some reports describing their proteolytic constituents are available. CC28 enzyme, the most basic cysteine proteinase from the group with a relative mass of 28.6 kDa (Gravina de Moraes et al. 1994), five cysteine proteinase isoforms (Gomes et al. 2005), the primary structures of two cysteine proteinases (CC-I, CC-III), (Walreavens et al. 1993, 1999), the cloning of a genomic region coding for a cysteine proteinase (Pereira et al. 2001), the cDNA cloning of CMS2MS2-like enzyme (Corrêa et al. 2011), plus the identification and the primary structure of a proliferative proteinase (Silva et al. 2003; Gomes et al. 2007), were reported. The isolation of two cysteine proteinase isoforms displaying proliferative effects on mammalian systems prompted us to study in more detail the proteolytic enzymes of *C. candamarcensis* (Gomes et al. 2005). This remarkable property had been anticipated as ethnopharmacological studies suggested that the plant was indigenously used to treat digestive disorders (Soplin et al. 1996).

The study of plant cysteine proteinases attracts academic interest because of their multiple alleged activities including antihelmintic and digestive action against nematode infection (Stepek et al. 2006), phytobezoar and celiac disease (Gass et al. 2007; Baker et al. 2007), anti-inflammatory activity by attenuating T cell response (Mynott et al. 1999), decreasing levels of prostaglandin (Gaspani et al. 2002), plasma fibrinogen and bradykinin (Brien et al. 2004). In addition, gastric ulcer protective and healing effects (Mello et al. 2006, 2008), burn-healing activity (Gomes et al. 2010a) likely mediated by the proliferative effect of cysteine proteinases CMS2MS2 and CMS2MS3, or as wound debridant (Ayello and Cuddigan 2004; Ford et al. 2002; Melano et al. 2004), were reported. The antitumoral/antimetastatic activity is perhaps one of the most attractive properties of plant cysteine proteinases, as this effect has been demonstrated with enzymes from different plant sources, arguing for a common mechanism (Wald et al. 1998, 2001; Batkin et al. 1988; Grabowska et al. 1997; Chobotova et al. 2010; Cabral et al. 2006; Báez et al. 2007).

Rawlings and Barrett (1993) used structural and evolutionary criteria to group proteinases in families and clans. Most plant cysteine proteinases belong to the papain (C1) and legumain (C13) families. The three-dimensional structures of C1 enzymes show the typical papain-like fold described by Drenth et al. (1968), composed of two domains, an α-helix-rich (L) domain and a β-barrel-like (R) domain, separated by a groove containing the active site formed by residues Cys25 and His159 (papain numbering), each one on each domain.

Papain was the first cysteine proteinase to be characterized in latex from *Carica papaya* and became a protein model in structural and kinetic studies for cysteine proteinases. The relatively easy to produce milligram amounts of this enzyme stems from its high concentration in papaya fruit latex before ripening, thus providing another example for the regulated expression of these enzymes. Papain is found in latex along with five isoforms of chymopapain, proteinase omega, and proteinase IV. It is likely that other short-lived proteinases also occur in papaya latex, but their identities were not yet confirmed in latex preparations, obtained by non-rigorous collection protocols. To date, none of the enzymes from *C. candamarcensis* has been structurally characterized, despite their stronger activity and potential therapeutic applications.

The enzyme described here (CMS1MS2) is purified by a combination of chromatographic steps and appears as highly pure on SDS-PAGE or mass spectrometry (Gomes et al. 2008). Using BAPNA (Benzoyl-Arg-p-nitroanilide) as substrate CMS1MS2 displays 18-fold higher catalytic efficiency than CMS2MS2 and its catalytic efficiency is twice higher than that of papain. In addition, the cleavage pattern of the β-chain of insulin demonstrates a preference for hydrophobic residues at P2 and P3 and no preference for P1 subsite, similar to other papain-like enzymes (Gomes et al. 2010b). It is of interest therefore to determine the structure of CMS1MS2 and to analyze the possible determinants responsible for its high activity. In the present work CMS1MS2 was crystallized in a different space group than that we have previously reported (Gomes et al. 2008). The X-ray structure of the enzyme was solved in a tetragonal space group $P4_12_12$ at 1.87 Å resolution and contains a single molecule per asymmetric unit. As far as we know, this is the first crystallographic structure for a proteinase from *C. candamarcensis* which exhibits twice the catalytic efficiency of papain and the closest resemblance with that enzyme.

Materials and methods

Protein purification and crystallization

CMS1MS2 was isolated from *Carica candamarcensis* latex by a combination of chromatographic steps involving:

initial gel filtration followed by two cation-exchange purification steps (CMS-Sephadex and Mono-S-Sepharose) and complex formation with E-64, as described previously (Teixeira et al. 2008; Gomes et al. 2008). Briefly, the purified enzyme at 1 mg ml^{-1} was activated with 0.9 mM dithiothreitol (DTT) for 30 min at 277 K. Then, 210 mM E-64 was added to the solution to attain a 1:5 molar ratio of enzyme to E-64 and further incubated with agitation for 30 min at 277 K. The sample was then dialyzed against Milli-Q water and concentrated to 10 mg ml^{-1}. No residual proteolytic activity was observed with casein following overnight incubation at 310 K. Initially, crystals were obtained using Hampton Research Crystal Screen kit 1 solutions #30 (0.2 M ammonium sulfate, 30 % w/v PEG 8000) and #31 (0.2 M ammonium sulfate, 30 % w/v PEG 4000). In this study, we used the hanging drop vapor-diffusion method at 291 K and home-prepared solutions with refined crystallization conditions to obtain suitable crystals for diffraction. The crystallization drop was prepared by mixing 1 µL of a 10-mg/ml CMS1MS2-E-64 solution with an equal volume of the reservoir solution and depositing it on top of the reservoir cell for diffusion.

X-ray diffraction data acquisition

Hanging drop plates containing single CMS1MS2-E-64 crystals were surface transported to the Brazilian Synchrotron Light Laboratory (Campinas, Brazil). Prior to data acquisition crystals were transferred for few seconds from the crystallization site into a cryogenic solution containing 9 µl of the mother liquor plus 1 µl of ethylene glycol. The crystals were then flash-cooled to 100 K using a cold nitrogen stream followed by data collection. Diffraction data were collected with a rotation range of 1.0° per 30 s per image on a MarMosaic 225 image-plate detector using the W01B-MX2 beamline. The crystal-to-detector distance was set to 110 mm and a total of 360 images were collected to a resolution of 1.87 Å.

Data processing, structure solution and refinement

Data were indexed, integrated and scaled with the HKL2000 package (Otwinowski and Minor 1997). Phasing was performed by Molecular Replacement using AMoRe (Navaza 1994) from CCP4 (1994). The template used for Molecular Replacement was generated by the SWISS-MODEL server at http://swissmodel.expasy.org//SWISS-MODEL.html (Schwede et al. 2003) based on the primary structure of CMS1MS2 and available structures of caricain (PDB 1MEG), chymopapain (PDB 1YAL) and papain (PDB 1PPN). Cycles of restrained refinement were performed using REFMAC 5.2.0019 (Murshudov et al. 1997) concomitantly with ligand identification and manual rebuilding of the model. Water molecules were added to the model based on ($2mF_{obs} - DF_{calc}$; Φ_{calc}) and ($mF_{obs} - DF_{calc}$; Φ_{calc}) electron density maps. The progress in structure refinement was monitored by inspection of the model on graphics and *R*-factor/*R*-free values. The data on collection and refinement statistics are presented in Table 1.

Comparative three-dimensional structure studies and protein presentation

Molecular presentation, superposition and comparison with homologous tertiary structures of papain-like proteases and

Table 1 Data collection and refinement statistics of CMS1MS2 tetragonal crystal isoform

Beamline	MX2–LNLS
Wavelength (Å)	1.458
Temperature (K)	291
Crystal-to-detector distance (mm)	110
Rotation range per image (°)	1.0
Total rotation range (°)	360
Space group	$P4_12_12$
Unit cell parameters (Å)	$a = b = 73.639$, $c = 118.786$
Resolution range (Å)	1.87–50.00
Total # of observations	782,158
Total # of unique reflections	27,792
# of unique reflections (test set)	1,395
Data completeness (%)	100.0 (100.0)
$\langle I/\sigma\ (I)\rangle$	45.8 (17.0)
R^{a}_{merge} (%)	8.1 (25.1)
Redundancy	28.1 (27.8)
Mathews coefficient ($Å^3Da^{-1}$)	3.50
Solvent content (%)	64.82
No. of reflections used in refinement	27,714
R-factor[b] (%)	16.2
R^{c}_{free} (%)	19.3
# protein molecules/a.u.	1
# water molecules/a.u.	216
# E-64 molecules/a.u.	1
Mean *B*-factor ($Å^2$)	22.3
Rmsd bond distances (Å)	0.013
Rmsd bond angles (°)	1.3

Statistical values for the highest resolution shell (1.93–1.87 Å) are given in parentheses

[a] $R_{merge} = \Sigma_{hkl}\,\Sigma_i\,|\,I_i(hkl) - \langle I(hkl)\rangle\,|/\Sigma_{hkl}\,\Sigma_i\,I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of reflection *hkl* and $\langle I(hkl)\rangle$ is its average

[b] *R*-factor $= \sum(|F_{obs}| - |F_{calc}|)/\sum|F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively

[c] R_{free} was calculated using 5 % of reflections randomly selected in a test data set

B-factor calculation were performed using Pymol Program v0.99 (DeLano 2002). To model the subsite interaction (P2–S2) using a putative substrate, the papain structure complexed with the inhibitor N-benzyloxycarbonyl-leucyl-phenylalanyl-glycyl-diazomethane (Z-Leu-Phe-Gly-CHN_2) (PDB 1KHQ) was superposed with various cysteine proteinases aligning the tripeptide Leu-Phe-Gly portion at their active site.

Circular dichroism spectra

The data of the secondary structure for papain, chymopapain and caricain were reported early (Solís-Mendiola et al. 1992; Arroyo-Reyna et al. 1994). Purified CMS1MS2 or CMS2MS2 at 0.1 mg/mL in 10 mM phosphate buffer was scanned using far-UV CD spectra between 190 and 260 nm at 18 °C. A Jasco J-810 spectropolarimeter was employed for circular dichroism (CD) measurements, using a 2-mm path-length cell. The CD intensities were expressed as $\Delta\varepsilon$ (mdeg M^{-1} cm^{-1}). Percent distributions of different secondary structural elements (α-helix, β-sheet, β-turn and random coil) were estimated using the CDNN program (Sreerama and Woody 1993).

Results and discussion

CMS1MS2 crystal structure

The purification, characterization and initial crystallization attempts of CMS1MS2 of *C. candamarcensis* were reported earlier (Gomes et al. 2008). CMS1MS2 crystals suitable for diffraction have been obtained in two isoforms. Plate-like crystals belonging to the monoclinic space group $P2_1$ with two CMS1MS2 molecules per asymmetric unit were initially reported (Gomes et al. 2008). A new tetragonal crystal form of CMS1MS2 was unexpectedly obtained after recrystallization of dissolved monoclinic crystals during transport to the synchrotron. The new crystal isoform belonged to the tetragonal system (space group $P4_12_12$) with unit-cell parameters: $a = b = 73.64$, $c = 118.79$ Å and diffracted beyond 1.87 Å resolution, as reported in Table 1.

Both crystals isoforms were subjected to phase determination by Molecular Replacement with success. Since the latter tetragonal crystal exhibited better isotropic diffraction, data completeness and refinement statistics, its structure determination was conducted to a final R factor of 16.2 % ($R_{free} = 19.3$ %). The model contains a single CMS1MS2 molecule in the asymmetric unit with a solvent content of 64.8 %.

The asymmetric unit of the tetragonal isoform includes one CMS1MS2 molecule with 213 amino acid residues, 216 water molecules, 10 ethylene glycol molecules, 8 sulfate ions, plus one E-64 molecule covalently bound to the nucleophilic Cys25 residue. The net charge for CMS1MS2 is +9 $(R + K) - (D + E)$ compared to +16 for CMS2MS2 (another isoform from *C. candamarcensis*), similar to mexicain (+8) (*Jacaratia mexicana*), halfway between papain (+6) and chymopapain (+12) from *C. papaya.*

The stereo chemical quality of the model was analyzed with PROCHECK V 3.4.4 Program (Morris et al. 1992). The protein model places 207 residues corresponding to 97.3 % of the entire protein within the allowed regions of the Ramachandran plot, while six glycine residues are out of the boundary regions (data not shown). In addition, 11 residues (Ser30, Ser47, Arg64, Asn77, Glu89, Arg96, Gln99, Gln123, Ser156, Gln197 and Arg213) exhibit two possible conformations. The overall three-dimensional structure of CMS1MS2 is similar to other proteinases from the papain family (family C1A) (Drenth et al. 1968; Grzonka et al. 2001); it contains two domains (L and R) separated by a groove encompassing the active site formed by residues Cys25 and His159, each one located on each domain (Fig. 1a, b). Structurally, the model predicts a hydrogen bond interaction between the side-chain O atom of Asn175 and the imidazole ring from His159 (2.67 Å), providing proper positioning of His159 relative to the nucleophilic Cys25, as observed in similar reported cases (Fig. 1b).

The L domain is composed of three α-helixes: the R domain contains a twisted β-sheet and two α-helixes. The C- and N-termini of the R and L domains, respectively, establish interactions to stabilize the substrate-binding region. CMS1MS2 contains seven cysteine residues like other members of papain family, six forming disulfide bonds (Cys22–Cys63, Cys56–Cys95 and Cys153–Cys200), plus the catalytically active Cys25.

Comparative analysis with related structures

The crystal unit falls into a single assymmetric unit, confirming data that the protein exhibits monomeric state in solution (Teixeira et al. 2008). Analogous situation was found in other plant cysteine proteinases; however, the mexicain crystal contained four slightly dissimilar units (Gavira et al. 2007). The modeled molecule created by the electron density map of the main chain and lateral residues allows a precise representation of the structure. The average *B*-factor for the model is 19.7 $Å^2$ (18.4 $Å^2$ main chain, 21.1 $Å^2$ side chain); the lower *B*-factors localize on residues Gly11–Val16, the α-helix containing residues Trp26–Ile40 within the active site and residues Gln128–Ile132, opposite to the α-helix Trp26–Ile40. The higher *B*-factors locate on Ala97–Lys101 corresponding to a single

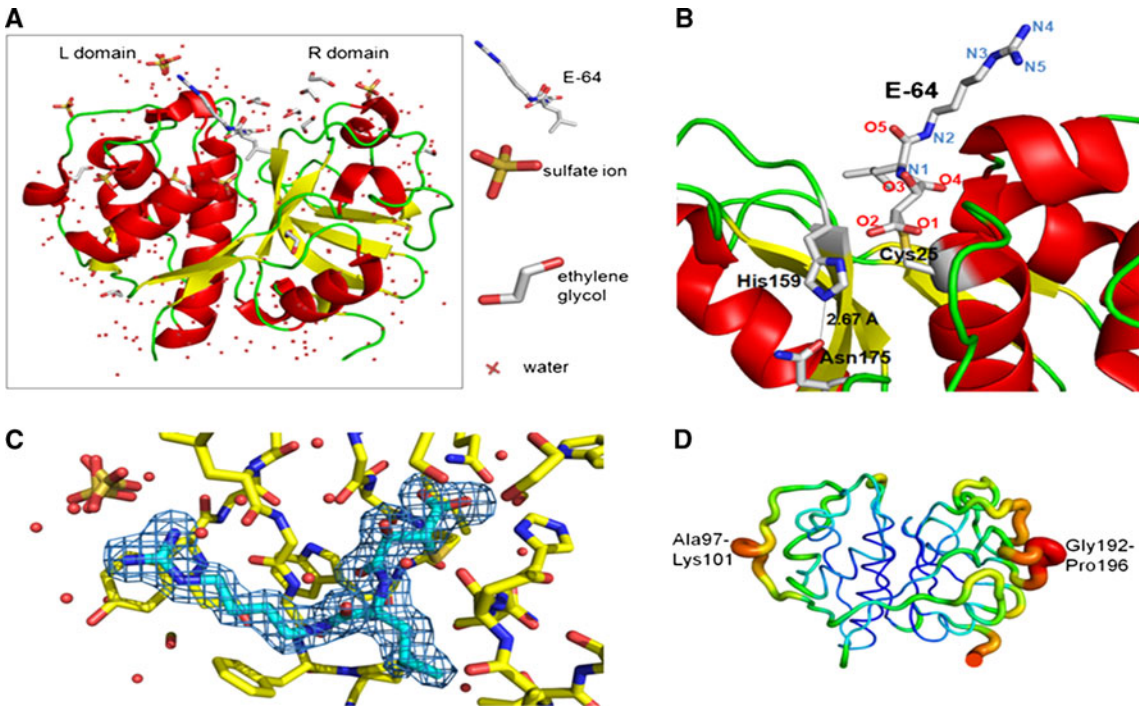

Fig. 1 The tridimensional structure of CMS1MS2. **a** Cartoon-like structure of CMS1MS2-E-64 complex: The model includes sulfate ions, ethylene glycol and water molecules. **b** Active site cleft of CMS1MS2. Residues forming the catalytic triad and the E-64 molecule are labeled. **c** The electron density around E-64 inhibitor. An OMIT electron density map ($mF_{obs} - DF_{calc}$; Φ_{calc}) around the E-64 inhibitor contoured at 3 σ is depicted in blue. **d** The *B*-factor distribution is shown as color gradient (*red*, larger stick diameter) indicating high *B* values, (*yellow*, *green*) intermediate values, (*blue*, *violet*) low values

turn-helix and in a coiled domain covering residues Gly192-Pro196, both located on the surface (Fig. 1d). A similar distribution of *B*-factors occurs in another model for cysteine proteinase (Gavira et al. 2007).

Circular Dichroism spectra from CMS1MS2 suggests a secondary structure composed by 33 % α-helix, 22.6 % β-sheet, 18.6 % β-turn and 25.8 % random coil (Table 2), similar to CMS2MS2 from the same species. The major difference observed when compared with *C. papaya* enzymes is the lower content of random coil in *C. candamarcensis* proteinase (25.8 vs. 45–47 %). The secondary structure composition of the modeled molecule obtained applying the Stride package (Heinig and Frishman 2004) is similar to the distribution determined in CD experiments (data not shown).

The primary structure alignment of CMS1MS2 with plant cysteine proteinases demonstrates, as expected, a high similarity to other enzymes from the Caricaceae (Gomes et al. 2010b). The proteinase CMS1MS2 contains the conserved regions 10-KGAVTPVK-17, 22-CGSCWAFS-29 based on papain notation. The residues 168-TSDG-171 found in chymopapain, glycyl endopeptidase and proteinase omega are absent in CMS1MS2, CMS2MS2 and papain. Other primary structure differences, namely in loops 191–196 and 99–105 were observed (Gomes et al. 2010b). These regions display large *B*-factors, suggesting that their structural differences arise from intrinsic flexibility of residues in these loops (Fig. 1d).

Table 3 shows the computed data following 3D structural comparison of various cysteine proteinase models. We used the Protein Model Comparator server (http://pm-cmp.appspot.com/) to determine Q-score, rmsd (root mean square deviation) GDT_TS and TM-score. The highest Q-score, GDT_TS and TM-S were found with papain, mexicain, proteinase omega and chymopapain, while glycyl endopeptidase exhibited intermediate scores followed by cathepsin K and actinidin displaying lower scores. CMS1MS2 shares 69 % amino acid identity with proteinase omega, 67 % with papain and 64 % with mexicain (Table 3). The rmsd confirms that CMS1MS2 resembles more Caricaceae enzymes than others cysteine proteinases.

Detailed inspection of structural deviations between Cα carbons from CMS1MS2 and other plant cysteine enzymes is shown in Fig. 2. Major deviations occur at the N-terminal residues (1–4), in residues 17–22, before the active site, residues 57–61, residues 97–106, in residues 165–171, and near the C-terminal region encompassing residues 191–204, the latter heterogeneity probably explained by structural flexibility and sequence differences within this region. A comparison of papain and CMS1MS2 which are structurally closer shows Cα clear divergences at residues

Table 2 Secondary structures of plant cysteine proteinases

Secondary structure	Papain[a]	Caricain[a]	Chymopapain[a]	CMS2MS2	CMS1MS2
α-Helix	24.1	24.5	28.7	28.5	33.0
β-Sheet antiparallel	10.7	12.7	6.5	14.7	15.4
β-Sheet parallel	4.5	1.8	8.3	8.8	7.2
β-Turn	13.4	15.5	12.0	17.6	18.6
Random coil	47.3	45.5	44.5	30.4	25.8

The numbers represent the percent distribution of secondary structure for each protein. Purified CMS1MS2 or CMS2MS2 at 0.1 mg/mL in 10 mM phosphate buffer was scanned using far-UV CD spectra between 190 and 260 nm at 18 °C. A Jasco J-810 spectropolarimeter was employed for CD measurements, using a 2-mm path-length cell

[a] Solís-Mendiola et al. (1992)

Table 3 Structural comparison between CMS1MS2 and plant proteases

PDB code	rmsd (Å)	Q-score	GDT_TS	TM-score	Length (aligned)	Match	Identity	Ligand
1KHP[a]	0.535	0.978	0.989	0.985	212 (208)	139	0.668	Diazomethylketone
2BDZ[b]	0.586	0.966	0.978	0.976	212 (209)	133	0.636	E-64
1PPO[c]	0.558	0.948	0.973	0.969	216 (209)	144	0.689	Hg
1YAL[d]	0.585	0.943	0.969	0.963	218 (209)	126	0.603	Thiomethyl group
1GEC[e]	0.672	0.917	0.953	0.943	216 (210)	138	0.657	Diazomethyl group
1ATK[f]	0.985	0.746	0.851	0.806	215 (205)	94	0.459	E-64
1AEC[g]	0.697	0.655	0.781	0.701	218 (206)	104	0.505	E-64

rmsd root mean square deviation (Å), *Q-score* fraction of target contacts, *GDT_TS* global distance test (total score), *TM-score* template modelling score: the structural similarity score is normalized within the range [0–1], *Match* identical residues between proteins compared, *Identity* ratio between matching and aligned residues

[a] Papain, Janowski et al. (2004)

[b] Mexicain, Gavira et al. (2007)

[c] Proteinase Omega, Pickersgill et al. (1990)

[d] Chymopapain, Maes et al. (1996)

[e] Glycyl endopeptidase, O'Hara et al. (1995)

[f] Cathepsin K, Zhao et al. (1997)

[g] Actinidin, Varughese et al. (1992)

2, 78, 99, 100, 104, 144, 198. The deviation at position 104 is caused by a Tyr residue in papain which is absent in CMS1MS2, CMS2MS2 and remaining enzymes from *C. papaya*. The missing residue places instead of Tyr104 a basic Lys residue at the equivalent position of CMS1MS2.

Kinetic studies of cysteine proteinases from Caricaceae suggest that CMS1MS2 displays the highest catalytic efficiency (*kcat*/Km) of these enzymes (288.7 M^{-1} s^{-1}) with BAPNA substrate, followed by papain that attains 145.0 M^{-1} s^{-1}. The catalytic efficiency of chymopapain (4.1 M^{-1} s^{-1}), CMS2MS2 (16.2 M^{-1} s^{-1}) and other proteolytic enzymes is well below these figures (Gomes et al. 2010b). The specific activity of CMS1MS2 using BAPNA was 7.13 nM μg s^{-1} compared with 0.49 nM μg s^{-1} for CMS2MS2 (Teixeira et al. 2008).

The higher *kcat*/Km of papain compared with chymopapain has been explained by papain preference at P2 subsite for an aromatic (Phe or Tyr) residue. The substitutions of Val133 and Val157 at S2 in papain for the larger residues Leu133 and Leu157 in chymopapain explain this catalytic difference (Watson et al. 1990). These replacements restrict the accessibility of the Phe residue from peptide substrates to the S2 pocket (Fig. 3), decreasing the reactivity of chymopapain. Interestingly, CMS1MS2 with its *kcat*/Km superior to papain contains Val133 and Ile157 at these positions, the second one (Ile157) more voluminous than Val157 in papain, arguing for the relevance of Val133 and against the importance of Val157.

A comparison at the active sites from CMS1MS2 and papain, the enzymes with best enzymatic performance within the group shows the following differences: Gly21, Phe67, Leu69, Ile132, Ile157, Val207 are found in CMS1MS2 while the corresponding residues in papain are Ser21, Tyr67, Trp69, Val132, Val157 and Phe207 (Table 4). While most of these changes are conservative, the substitution of Trp69 in papain by Leu69 in CMS1MS2 results in a less voluminous residue in S2, which may facilitate substrate entry. Besides, the presence in

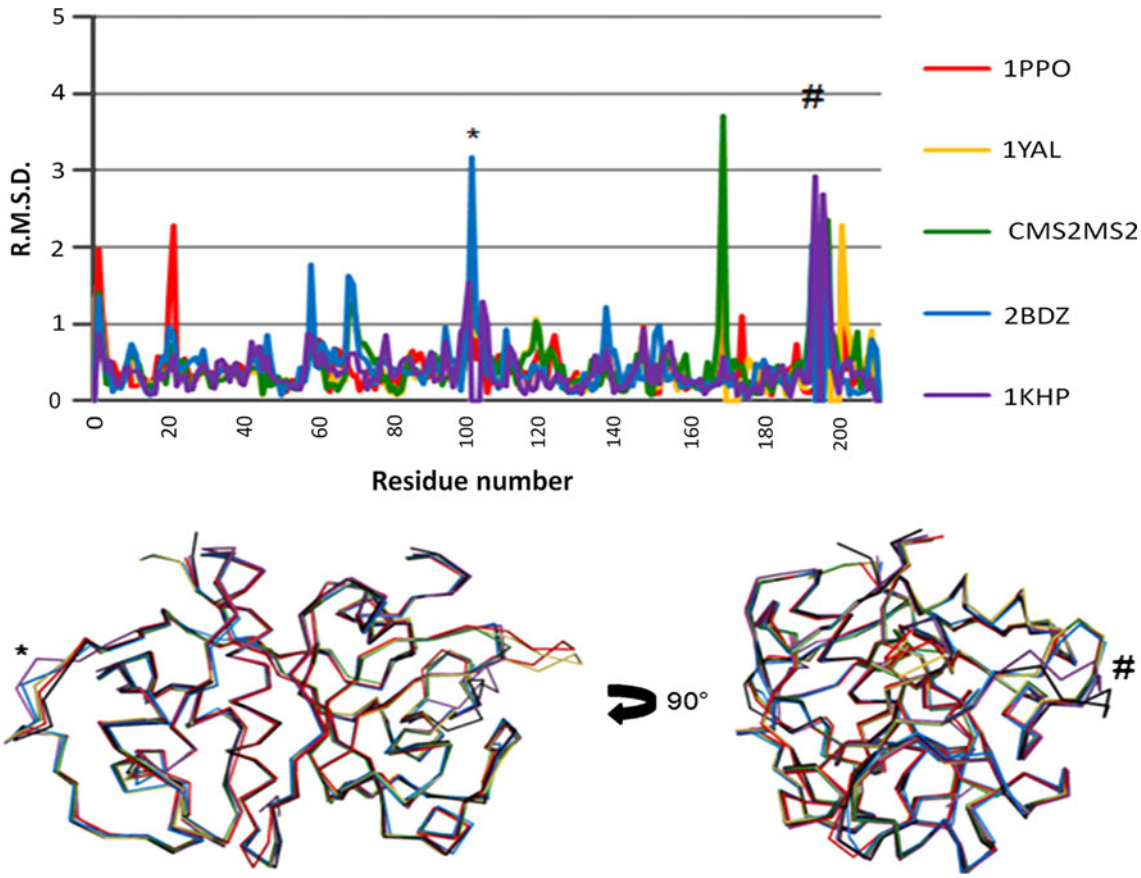

Fig. 2 Structural comparison of cysteine proteases from Caricaceae. The rmsd (Å) between various cysteine proteinases and CMS1MS2 is represented by the graph. The distances were calculated for each residue based on the superimposed model of CMS1MS2 and each of the molecules. Below is shown the structure alignment using the Cα backbones of proteinase omega (PPO), chymopapain (1YAL), mexicain (2BDZ), papain (1KHP) and CMS2MS2. * and # indicate regions showing large variations among proteins

CMS1MS2 of Phe instead of the more polar Tyr67 residue in the equivalent position of papain at S3 could favor the interaction with the substrate. An additional difference is the presence of Gly21in CMS1MS2 S1′ site instead of Ser21 in S1′ from papain; however, it is unclear how this change could favor the higher activity of CMS1MS2.

The structural comparison between CMS1MS2 derived by X-ray crystallography and the structural model proposed for CMS2MS2 (Gomes et al. 2007) supports the rationale for sensitivity of CMS1MS2 to cystatin, unlike the resistance of CMS2MS2 (Gomes et al. 2010b). At least two possible changes justify this difference. While CMS2MS2 has a Ser136 like glycyl endopeptidase which is also insensitive to cystatin, CMS1MS2 has the usual Ala136 found in most papain-like enzymes sensitive to cystatin. The presence of Ala136 is suggested as required for interaction with hydrophobic residues Leu54 and Val55 located within the first flexible loop of cystatin (Bode et al. 1988). Moreover, it is proposed that Trp177 and Trp181 are essential for protein binding to cystatin. Then, the non-bulky Gly180 in CMS1MS2 would not interfere during this hydrophobic interaction. Interestingly, the presence of the long and basic amino acid Arg180 between Trp177 and Trp181 in CMS2MS2 imposes a restriction on the inhibitor-enzyme interaction that can be crucial for cystatin resistance (Gomes et al. 2010b).

The active site interaction with the inhibitor

The electron density observed around E-64 in the CMS1MS2 complex is well defined; the inhibitor is positioned at the enzyme cleft entrance, between the L and R domains, within binding distance to Cys25 (Fig. 1b, c) as in similar structures determined earlier (Varughese et al. 1992; Zhao et al. 1997; Gavira et al. 2007).

The distance between C2 from E-64 and the SG atom from Cys25 is 2.08 Å, somewhat higher than the values for caricain, actinidin, cathepsin K and papain complexes (≈1.80 Å), the latter crystallized with the E-64c analogue, but shorter than the distance for mexicain (2.33 Å) (data not shown). However, while the reactive side of E-64 is positioned relatively constantly among complexes, the guanidine tail shows different positioning depending on the complex. The conformation of the CMS1MS2-E-64

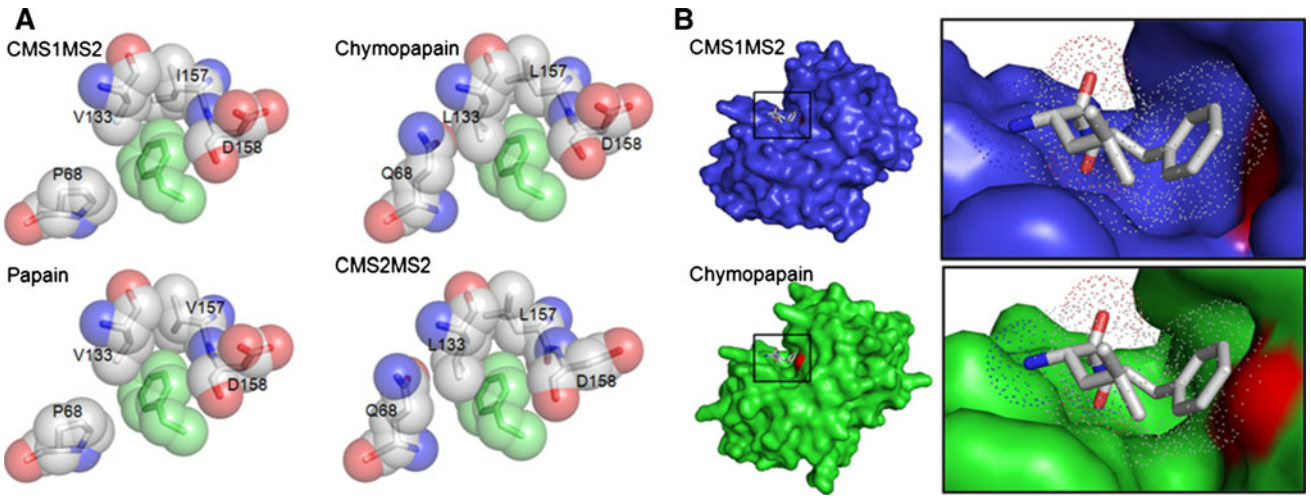

Fig. 3 Comparison of subsite structure at the active site cleft in proteases from Caricaceae. **a** View of the P2 side chain residue (Phe, shown in *green*) of a putative substrate located in S2 pocket in the indicated proteases. The residues forming the S2 pocket are indicated and colored according to atom's identity (carbon in gray, nitrogen in *blue* and oxygen in *red*). **b** Overall view of a putative substrate composed by Leu-Phe-Gly colored as described above inserted in the active site cleft (*left*). A detailed view (*right*) shows the protuberance due the presence of Leu133 (shown in *red*) in chymopapain instead a Val133 in CMS1MS2

Table 4 Subsite residues at the active site cleft in papain-like proteases

Subsite	Residue[a]	Papain	CMS1MS2	Chymopapain	CMS2MS2	Caricain	Glycyl endopep.
S3	61	Y	Y	Y	H	H	Y
	66	G	G	G	G	G	G
	67	Y	F	Y	Y	Y	Y
S2	68	P	P	Q	Q	P	Q
	69	W	L	T	T	P	S
	132	V	I	V	V	V	V
	133	V	V	L	L	V	V
	157	V	I	L	L	V	V
	158	D	D	D	D	D	D
	207	F	V	Y	Y	Y	Y
S1	23	G	G	G	G	G	E
	25	C	C	C	C	C	C
	65	G	G	G	G	G	R
S1′	20	G	G	G	N	G	G
	21	S	G	A	P	S	Y
	22	C	C	C	C	C	C
S2′	142	Q	Q	Q	Q	Q	Q
S3′	177	W	W	W	W	W	W
	181	W	W	W	W	W	W

[a] Papain numbering

complex is similar to the positioning observed in the actinidin-E-64 complex. The E-64 tail lies straight over the L domain, while in the cathepsin complex the inhibitor is bent towards the loop 181–186 and in the proteinase omega D158E complex the E-64 tail points away from the protein structure. On the other hand, the positioning of E-64 in papain is somewhat different, with the guanidine residue leaning towards the R domain, although this difference could be due to the modified character of E-64 used in the papain complex (data not shown). The distances between O1–O5 atoms from E-64 and Gln19, Ser24, Cys25, Gly66 and His159 from CMS1MS2, respectively, are similar to those observed in the protein-E-64 models compared in Table 5. Particularly, the hydrogen bond between E-64-O1 and His159-ND1 results stronger (2.73 Å) if compared with similar structures (2.80–3.29 Å), as it is the interaction between E-64-O2 and Gln19-NE2 (2.77 Å) matched in the complex caricain-E-64.

Table 5 Hydrogen bonding interactions of E-64 with cysteine proteinases

Ligand atom	Protein atom#	1AEC	2BDZ	1MEG	1ATK	3IOQ
(A) Distances in Å between the ligand and atoms in proteins						
E-64-O1	159-His-ND1	2.80	3.29	2.88	2.95	2.73
E-64-O2	25-Cys-N	2.94	3.13	3.01	3.06	2.98
E-64-O2	25-Cys-SG	3.24	3.18	3.36	3.24	3.41
E-64-O2	19-Gln-NE2	2.91	2.92	2.74	2.95	2.77
E-64-O2	24-Ser-N	3.38	3.44	3.32	3.31	3.21
E-64-O4	66-Gly-N	2.92	2.85	2.78	3.07	2.97
E-64-O4	25-Cys-SG	3.51	3.30	3.41	3.38	3.42
E-64-N1	158-Asp-O	3.15	3.17	3.27	3.29	3.48
E-64-N2	66-Gly-O	2.91	2.65	2.94	2.83	3.00
E-64-N5	64-Arg-O	3.41	2.71	–	–	3.04
(B) It shows the distance between the ligand and H_2O						
E-64-O1	HOH	2.81	2.61	2.98	–	2.63
E-64-O1	HOH	3.13	–	–	2.66	–
E-64-O3	HOH[b]	2.69	2.43	–	2.91	2.58
E-64-O3	HOH[a]	–	2.82	–	–	3.22
E-64-N1	HOH[a]	–	3.28	–	–	3.18
E-64-O5	HOH[a]	–	2.77	–	–	3.15
E-64-O5	HOH	–	–	–	–	2.76
E-64-N3	HOH	3.02	–	–	–	–
E-64-N4	HOH	3.23	–	–	–	3.76
E-64-N4	HOH	–	–	–	–	2.93
E-64-C10	HOH	3.08	–	–	–	–
E-64-C3	HOH[b]	3.52	3.16	–	3.77	3.42

Cysteine proteinases whose models were built with E-64 as ligand are included. A dash means an interaction ≥3.8 Å (cut-off)

1AEC actinidin, *2BDZ* mexicain, *1MEG* chymopapain, *1ATK* cathepsin K, *3IOQ* CMS1MS2

[a] HOH: same water molecule

[b] HOH: the same water molecule

Besides the protein-inhibitor contacts, solvent-inhibitor interactions account for stabilization of the ligand into the protein cleft. It is possible to see in Table 5 that five well-defined water molecules locate within hydrogen bond distances (<3.3 Å) from oxygen or nitrogen atoms from E-64 molecule. Interestingly, one of these water molecules (HOH[a]) is responsible for mediating three hydrogen bonds to O3, N1 and O5 E-64 atoms, a pattern observed only in mexicain.

Conclusions

The cysteine proteinase described in this study belongs to the papain family and represents the first solved structure from *C. candamarcensis*, a species native and maintained in South America. The net charge of CMS1MS2 (+9) explains its early elution with the cationic exchanger used for purification. The enzyme displays the highest proteolytic activity if compared with the homologues from the same species and is twice the activity of papain towards BAPNA substrate, the most active enzyme from *C. papaya*. The increased activity of CMS1MS2 is shared by other proteolytic enzymes identified in species autochthonous to the continent suggesting an adaptive process.

The structural analysis demonstrates similarities between CMS1MS2 and papain; however, minor differences are observed, some of them within the Cα backbone, others at the active site, e.g., Leu69 in CMS1MS2 instead of Trp69 in papain. The availability of the recombinant form of CMS2MS2, an isoform of CMS1MS2 allows the design of site-directed mutants to probe the effect of such changes.

Acknowledgments The research was funded by CNPq, CAPES and Fapemig. We thank Dr. Abraham Schnaiderman for his financial aid to this research and the Brazilian Synchrotron Light Laboratory (LNLS) for providing access to their facilities for X-ray crystallography experiments. Each author confirms that there is no conflict of interest in connection with this manuscript.

References

Arroyo-Reyna A, Hernandez-Arana A, Arreguin-Espinosa R (1994) Circular dichroism of stem bromelain: a third spectral class within the family of cysteine proteinases. Biochem J 300:107–110

Ayello E, Cuddigan JE (2004) Debridement: controlling the necrotic/cellular burden. Adv Skin Wound Care 17:66–75

Azarkan M, Dibiani R, Baulard C, Baeyens-Volant D (2006) Effects of mechanical wounding on *Carica papaya* cysteine endopeptidases accumulation and activity. Int J Biol Macromol 38:216–224

Báez R, Lopes MT, Salas CE, Hernández M (2007) In vivo antitumoral activity of stem pineapple (*Ananas comosus*) bromelain. Planta Med 73:1377–1383

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

Baker EL, Baker WL, Cloney DJ (2007) Resolution of a phytobezoar with Aldoph's Meat Tenderizer. Pharmacother 27:299–302

Barrett AJ, Buttle DJ (1985) Names and numbers of papaya proteinases. Biochem J 228:527

Batkin S, Taussig S, Szekerczes J (1988) Modulation of pulmonary metastasis (Lewis lung carcinoma) by bromelain, an extract of the pineapple stem (*Ananas comosus*). Cancer Invest 6:241–242

Bode W, Engh R, Musil D, Thiele U, Huber R, Karshikov A, Brzin J, Kos J, Turk V (1988) The 2.0 A X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J 7:2593–2599

Bravo LM, Hermosilla L, Salas CE (1994) A biochemical comparison between latex from *C. candamarcensis* and *C. papaya*. Braz J Med Biol Res 26:2831–2842

Brien S, Lewith G, Walker A, Hicks S, Middleton D (2004) Bromelain as a Treatment for Osteoarthritis: a Review of Clinical Studies. Evid Based Complement Alternat Med 1:251–257

Cabral H, Leopoldino AM, Tajara EH, Greene LJ, Faça VM, Mateus RP, Ceron CR, de Souza Judice WA, Juliano L, Bonilla-Rodriguez GO (2006) Preliminary functional characterization, cloning and primary sequence of Fastuosain, a cysteine peptidase isolated from fruits of *Bromelia fastuosa*. Protein Pept Lett 13:83–89

CCP4 Program Suite 6.0.2 (1994) Collaborative computational project number 4. Acta Cryst D50:760–763

Chobotova K, Vernallis AB, Majid FA (2010) Bromelain's activity and potential as an anti-cancer agent: Current evidence and perspectives. Cancer Lett 290:148–156

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

DeLano WL (2002) The PyMOL molecular graphics system. DeLano Scientific, Palo Alto, CA, USA. http://www.pymol.org

Drenth J, Jansonius JN, Koekoek R, Swen HM, Wolthers BG (1968) Structure of papain. Nature (London) 218:929–932

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

Ford CN, Reinhard ER, Yeh D, Syrek D, De Las Morenas A, Bergman SB, Williams S, Hamori CA (2002) Interim analysis of a prospective, randomized trial of vacuum-assisted closure versus the healthpoint system in the management of pressure ulcers. Ann Plast Surg 49:55–61

Furie B, Furie BC (1988) The molecular base of blood coagulation. Cell 53:505–518

Gaspani L, Limiroli E, Ferrario P, Bianchi M (2002) In vivo and in vitro effects of bromelain on PGE(2) and SP concentrations in the inflammatory exudate in rats. Pharmacol 65:83–86

Gass J, Bethune MT, Siegel M, Spencer A, Khosla C (2007) Combination enzyme therapy for gastric digestion of dietary gluten in patients with celiac sprue. Gastroenterol 133:472–480

Gavira JA, Gonzalez-Ramirez LA, Oliver-Salvador MC, Soriano-Garcia M, Garcia-Ruiz JM (2007) Structure of the mexicain-E-64 complex and comparison with other cysteine proteases of the papain family. Acta Cryst D63:555–563

Gomes MTR, Mello VJ, Rodrigues KC, Bemquerer MP, Lopes MTP, 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, Bemquerer MP, Lopes MTP, Richardson M, Oyama S, Salas CE (2007) The structure of CMS2MS2, a mitogenic protein isolated from latex of *Carica candamarcensis*. Biol Chem 388:819–822

Gomes MT, Teixeira RD, Ribeiro H, 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 Cryst F64:492–494

Gomes FS, Spínola CV, Ribeiro HA, Lopes MT, Cassali GD, Salas CE (2010a) Wound-healing activity of a proteolytic fraction from *Carica candamarcensis* on experimentally induced burn. Burns 36:277–283

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

Grabowska E, Eckert K, Fichtner I, Schulze-Forster K, Maurer H (1997) Bromelain proteases suppress growth invasion and lung metastasis of B16F10 mouse melanoma cells. Int J Oncol 11:243–248

Gravina de Moraes M, Termignoni C, Salas CE (1994) Biochemical characterization of a new cysteine endopeptidase from *Carica candamarcensis* L. Plant Sci 102:11–18

Grzonka Z, Jankowska E, Kasprzykowski F, Kasprzykowska R, Lankiewicz L, Wiczk W, Wieczerzak E, Ciarkowski J, Drabik P, Janowski R, Kozak M, Jaskólski M, Grubb (2001) Structural studies of cysteine proteases and their inhibitors. Acta Biochim Pol 48:1–20

Heinig M, Frishman D (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucl Acids Res 32:W500–W502

Janowski R, Kozak M, Jankowska E, Grzonka Z, Jaskólski M (2004) Two polymorphs of a covalent complex between papain and a diazomethylketone inhibitor. J Pept Res 64:141–150

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

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

Melano E, Rodriguez HL, Carrillo R Jr, Dillon L (2004) The effects of Panafil when using topical negative pressure to heal an infected sternal wound. J Wound Care 13:425–426

Mello VJ, Gomes MTR, Rodrigues KCL, Sanchez EF, Lopes MTP, Salas CE (2006) In: Govil JN, Singh VK, Arunachalam C (eds) Recent progress in medicinal plants Drug development from molecules, vol 11. Studium Press, LLC, Houston, USA, pp 211–224

Mello VJ, Gomes MTR, Lemos FO, Delfino JL, Andrade SP, Lopes MTP, Salas CE (2008) The gastric ulcer protective and healing

role of cysteine proteinases from *Carica candamarcensis*. Phytomed 15:237–244

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

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins 12:345–364

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

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Cryst D53:240–255

Mynott TL, Ladhams A, Scarmato P, Engwerda CR (1999) Bromelain, from pineaple stems, proteolytically blocks activation of extracellular regulated kinase-2 in T cells. J Immunol 163:2568–2575

Navaza J (1994) AMoRe: an automated package for molecular replacement. Acta Cryst A50:157–163

O'Hara BP, Hemmings AM, Buttle DJ, Pearl LH (1995) Crystal structure of glycyl endopeptidase from *Carica papaya*: a cysteine endopeptidase of unusual substrate specificity. Biochemistry 34:13190–13195

Otwinowski Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol 276:307–326

Pereira MT, Lopes MT, Meira O, 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

Pickersgill RW, Sumner IG, Goodenough PW (1990) Preliminary crystallographic data for protease omega. Eur J Biochem 190:443–444

Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. Biochem J 290:205–218

Ritonja A, Buttle DJ, Rawlings ND, Turk V, Barrett AJ (1989) Papaya proteinase IV aminoacid sequence. FEBS Lett 258:109–112

Schwede T, Kopp J, Juex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res 31:3381–3385

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

Silva CA, Gomes MTR, Ferreira RS, Rodrigues KCL, do Val CG, Lopes MTP, Mello VJ, Salas CE (2003) A mitogenic protein fraction in latex from *Carica candamarcensis.* Planta Med 69:926–932

Solís-Mendiola S, Arroyo-Reyna A, Hernández-Arana A (1992) Circular dichroism of cysteine proteinases from papaya latex. Evidence of differences in the folding of their polypeptide chains. Biochim Biophys Acta 1118:288–292

Soplin SP, Millones EA, Campos JLA, Deza LG, Rios EJ, Sanchez IB, Benítez MR, Pando LG, Panizo RS, Merino C del C, Rivera SJC (1996) Informe Nacional para la Conferencia Técnica Internacional de FAO sobre los Recursos Fitogenéticos

Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 209:32–44

Stepek G, Lowe AE, Buttle DJ, Duce IR, Behnke JM (2006) In vitro and in vivo anthelmintic efficacy of plant cysteine proteinases against the rodent gastrointestinal nematode. Parasitol 132:681–689

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

Varughese KI, Su Y, Cromwell D, Hasnain S, Xuong NH (1992) Crystal structure of an actinidin-E-64 complex. Biochemistry 31:5172–5176

Wald M, Závadová E, Poucková P, Zadinová M, Boubelik M (1998) Polyenzyme preparation Wobe-Mugos inhibits growth of solid tumors and development of experimental metastases in mice. Life Sci 62:43–48

Wald M, Olejár T, Sebková V, Zadinová M, Boubelík M, Poucková P (2001) Mixture of trypsin, chymotrypsin and papain reduces formation of metastases and extends survival time of C57Bl6 mice with syngeneic melanoma B16. Cancer Chemother Pharmacol 47:S16–S22

Walraevens V, Vandermeers-Piret MC, 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, Schneck 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 C, Yaguchi M, Lynn KR (1990) The aminoacid sequence of chymopapain from *Carica papaya*. Biochem J 266:75–81

Zhao B, Janson CA, Amegadzie BY, D'Alessio K, Griffin C, Hanning CR, Jones C, Kurdyla J, McQueney M, Qiu X, Smith WW, Abdel-Meguid SS (1997) Crystal structure of human osteoclast cathepsin K complex with E-64. Nat Struct Biol 4:109–111