Acta Crystallographica Section F
**Structural Biology and Crystallization Communications**
ISSN 1744-3091

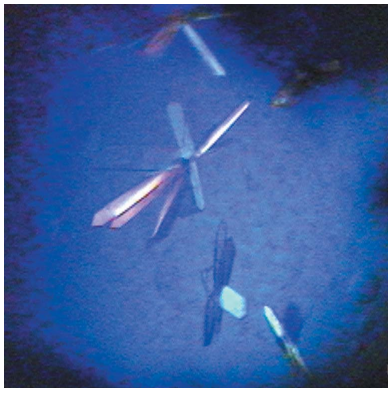
# Purification, crystallization and preliminary X-ray analysis of CMS1MS2: a cysteine proteinase from *Carica candamarcensis* latex

**Marco Túlio Ribeiro Gomes,[a] Raphael Dias Teixeira,[a] Henrique de Assis Lopes Ribeiro,[a] Andréia Pereira Turchetti,[a] Caroline Furtado Junqueira,[a] Míriam Tereza Paz Lopes,[b] Carlos Edmundo Salas[a]* and Ronaldo Alves Pinto Nagem[a]***

[a]Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG, CEP 31270-901, Brazil, and [b]Departamento de Farmacologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Avenida Antônio Carlos 6627, Belo Horizonte, MG, CEP 31270-901, Brazil

Correspondence e-mail: cesbufmg@icb.ufmg.br, nagem@icb.ufmg.br

Received 4 March 2008
Accepted 24 April 2008





Cysteine proteinases from the latex of plants of the family Caricaceae are widely used industrially as well as in pharmaceutical preparations. In the present work, a 23 kDa cysteine proteinase from *Carica candamarcensis* latex (designated CMS1MS2) was purified for crystallization using three chromatography steps. The enzyme shows about fourfold higher activity than papain with BAPNA as substrate. Crystals suitable for X-ray diffraction experiments were obtained by the hanging-drop method in the presence of PEG and ammonium sulfate as precipitants. The crystals are monoclinic (space group $P2_1$), with unit-cell parameters $a$ = 53.26, $b$ = 75.71, $c$ = 53.23 Å, $\beta$ = 96.81°, and diffract X-rays to 1.8 Å resolution.

## 1. Introduction

The latex from plants of the family Caricaceae contains several compounds that confer protection against environmental damage. For instance, papain, a cysteine proteinase from *Carica papaya* latex, is crucial for the defence of the papaya tree against lepidopteran larvae (Konno *et al.*, 2004). Furthermore, upon fruit injury, latex exudates transiently until a clot forms around the wounded area. The coagulation process is vital as it creates a physical barrier against predator attack. Previously, we provided evidence that a number of peptides are proteolytically processed in a nonrandom manner during latex coagulation. Clot formation takes place when various isoforms of cysteine proteases are transiently activated in a fashion similar to mammalian blood coagulation (Moutim *et al.*, 1999).

The highly active cysteine proteases from *C. papaya* generically designated papain are widely used industrially and in pharmaceutical preparations. These applications are facilitated by the relatively simple procedure required to obtain these proteases in a form suitable for use and because of their *in vitro* nonspecific mode of cleavage, which allows the digestion of a range of substrates (Bravo *et al.*, 1994). The European (EU) market is estimated to be several hundred tons per year, while the US market is projected to be double that of the EU (Agribusiness Development Centre, 2000).

To date, several crystallographic structures of cysteine proteases from the Caricaceae family have been determined, such as those of papain (Kamphuis *et al.*, 1984), chymopapain (Maes *et al.*, 1996) and caricain (Katerelos *et al.*, 1996), all of which are from *C. papaya*. More recently, Gavira *et al.* (2007) described the structure of mexicain, a cysteine protease isolated from *Jacaratia mexicana*, another member of the Caricaceae family. These enzymes share between 58% and 69% residue similarity and their reaction mechanisms are well understood. The participation of Cys25 and His159 in the catalytic site of papain was established when the three-dimensional structure of the enzyme was solved (Drenth *et al.*, 1968).

*Carica candamarcensis*, also named *Vasconcellea cundinamarcensis*, another member of the Caricaceae family, is common to many areas of South America. Its latex contains cysteine proteinases with a proteolytic activity that is five to seven times higher than the corresponding enzymes from *C. papaya* (Baeza *et al.*, 1990; Bravo *et al.*, 1994). Although the proteolytic components of *C. candamarcensis*

have been less studied than their counterparts in *C. papaya*, some reports describing the proteolytic and biological activities of this species are available. Walreavens *et al.* (1993) described the isolation of four cysteine proteases (CCI, CCII, CCIII and CCIV). CCI displayed the highest amidase activity against the substrate BAPNA ($N^{\alpha}$-benzoyl-DL-arginine-*p*-nitroanilide hydrochloride), being about three times more active than papain (Walreavens *et al.*, 1993). Subsequently, CCI was resolved into two components (CCIa and CCIb) by reverse-phase HPLC under denaturing conditions (Walraevens *et al.*, 1999). Their primary structures (213 amino-acid residues per isoform) were elucidated and shown to share 89% identity (Walraevens *et al.*, 1999). Recently, our group described a group of cysteine proteinases designated P1G10 that displayed protective and healing effects on gastric ulcers when given to animals prior to indomethacin or following ulcer induction with acetic acid (Mello *et al.*, 2008). The proteases CMS2MS2 and CMS2MS3 were purified from the P1G10 source and shown to enhance proliferation of cultured mammalian fibroblasts (Gomes *et al.*, 2005). In addition, another enzyme, CMS1MS2, was isolated from P1G10 and identified in this study as CCIb (as designated earlier) by electrospray ionization mass spectrometry (ESI–MS) and N-terminal Edman sequencing. The latter enzyme shares 66% amino-acid residue identity with papain.

In the present work, we describe the purification, crystallization and preliminary X-ray diffraction analysis of CMS1MS2 covalently bound to the protease inhibitor E-64 [*n*-(3-carboxyoxirane-2-carbonyl)-leucyl-amino(4-guanido)butane]. This is the first crystallographic report of a cysteine proteinase from *C. candamarcensis*.

## 2. Purification

Latex from *C. candamarcensis* was freed from low-molecular-weight contaminants by gel filtration on G-10 as described by Mello *et al.* (2008). The first protein fraction (P1G10) recovered was applied onto a CM-Sephadex C-50 column as described by Gomes *et al.* (2005). During each chromatographic step the active fractions were monitored at 405 nm for amidase activity against BAPNA as described by Gravina *et al.* (1994). The first peak (CMS1) eluted from CM-Sephadex C-50 displayed the highest amidase activity (data not shown). Further purification of CMS1 was attained by Mono-S FPLC (0.5 × 100 mm) chromatography. The Mono-S equilibration buffer was 1 m*M* $Na_2EDTA$, 16 m*M* NaCl and 0.025 *M* sodium borate pH 9.5. An isocratic solution set at the same concentration of the equilibration buffer allowed the separation of three fractions (CMS1MS1, CMS1MS2 and CMS1MS3, according to elution order). Each of the pooled fractions was concentrated by Amicon ultrafiltration through YM10 membranes (10 kDa molecular-weight cutoff). CMS1MS2 displayed the highest amidase activity of these fractions, followed by CMS1MS1 and then CMS1MS3 (Fig. 1).

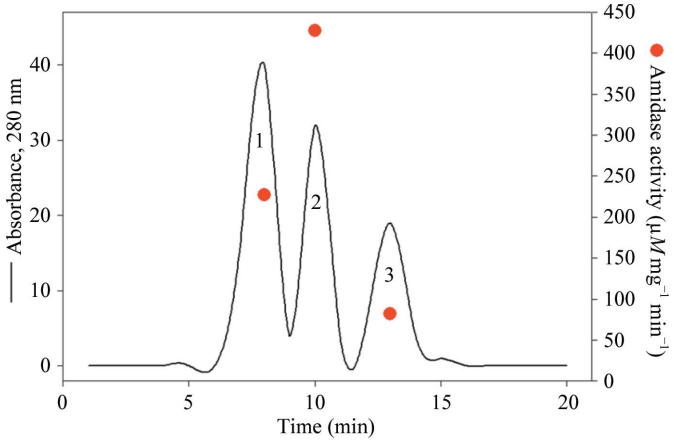



**Figure 1**
Mono-S chromatography of the CMS1 fraction from *C. candamarcensis*. The first protein peak from CM-Sephadex was applied (1 mg) onto a Mono-S column and eluted at a flow rate of 1 ml $min^{-1}$ with a buffer solution containing 1 m*M* $Na_2EDTA$, 16 m*M* NaCl and 0.025 *M* sodium borate pH 9.5. The red dots show the mean amidase activity of each pooled fraction. 1, 2 and 3 correspond to CMS1MS1, CMS1MS2 and CMS1MS3, respectively.

Denaturing polyacrylamide-gel electrophoresis of CMS1MS2 (SDS–PAGE) showed a single band with an apparent molecular weight of 23 kDa, which was in accordance with the ESI–MS data (22 991 Da). This weight resembles the weight of CCIb (22 990 Da) previously reported by Walraevens *et al.* (1999). Sequencing of the first 21 amino-acid residues of the isolated CMS1MS2 (IPTSIDWRQKGAVTPVRNQGG) by Edman degradation confirms that CMS1MS2 and CCIb are the same protein.

## 3. Crystallization

For crystallization trials, CMS1MS2 was inhibited with E-64 (Sigma–Aldrich) in order to avoid self-digestion. This inhibitor was first isolated from *Aspergillus japonicus* by Hanada *et al.* (1978) and shown to be specific for cysteine proteases. The purified enzyme at 1 mg $ml^{-1}$ was activated with 0.9 m*M* dithiothreitol (DTT) for 30 min at 277 K. 210 μ*M* E-64 was then 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 $ml^{-1}$. The residual proteolytic activity was verified with casein as substrate and no measurable activity was detected following overnight incubation at 310 K (Baeza *et al.*, 1990). The mass of the CMS1MS2–E-46 complex was 23 351 Da as determined by ESI–MS. These data show that one E-64 molecule binds to one CMS1MS2 molecule (1:1 complex), as expected.

Crystallization was performed by the hanging-drop vapour-diffusion method at 291 K using Hampton Research Crystal Screens I and II. The crystallization assay was performed by mixing 1 µl 10 mg $ml^{-1}$ CMS1MS2–E-64 complex solution with an equal volume of reservoir solution. The drop was suspended over 1.0 ml reservoir solution. Plate-like crystals suitable for diffraction experiments (Fig. 2) with approximate dimensions of 0.25 × 0.05 × 0.01 mm appeared within a week in two Crystal Screen I conditions: Nos. 30

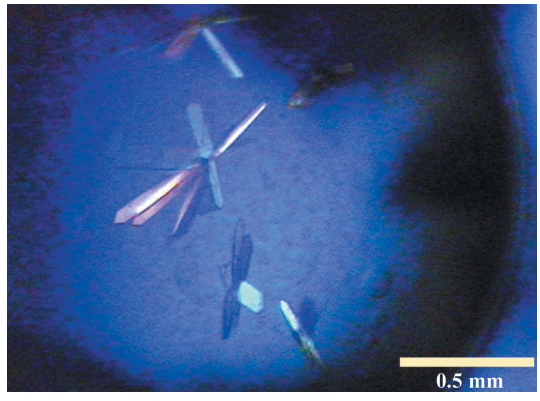



**Figure 2**
Typical crystals of CMS1MS2. These crystals appeared in a crystallization drop suspended over a reservoir containing 0.2 *M* ammonium sulfate, 30%(*w*/*v*) PEG 8000.

**Table 1**
Summary of diffraction data statistics for the CMS1MS2 crystal.

Values in parentheses are for the highest resolution shell (1.84–1.80 Å).

| | |
|---|---|
| Wavelength (Å) | 1.425 |
| Space group | $P2_1$ |
| Unit-cell parameters (Å, °) | $a$ = 53.26, $b$ = 75.71, $c$ = 53.23, $\beta$ = 96.81 |
| Resolution range (Å) | 50.00–1.80 (1.84–1.80) |
| No. of observations | 139844 (8961) |
| No. of unique reflections | 36999 (2404) |
| Data completeness (%) | 95.0 (92.0) |
| $I/\sigma(I)$ | 20.3 (12.5) |
| $R_{\rm merge}$† (%) | 6.4 (10.5) |
| Redundancy | 3.8 (3.7) |
| Mosaicity (°) | 0.65 |
| $B$ factor (Å$^2$) | 16.2 |

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$.

[0.2 *M* ammonium sulfate, 30%(*w*/*v*) PEG 8000] and 31 [0.2 *M* ammonium sulfate, 30%(*w*/*v*) PEG 4000]. The crystals obtained in the screening experiments were used for data acquisition as attempts to fine-tune the crystallization conditions and/or improve crystal quality using home-prepared solutions were unsuccessful.

## 4. Data collection and processing

Single crystals were harvested using nylon loops (Hampton Research) and transferred from the crystallization drop to 10 µl of a cryogenic solution containing the mother liquor and 10%(*v*/*v*) ethylene glycol for a few seconds. Crystals were then flash-cooled to 100 K in a cold nitrogen stream and used for data collection. Diffraction data were collected with a rotation range of 0.5° per 90 s image on a MAR CCD image-plate detector using the D03B-MX1 beamline at the Brazilian Synchrotron Light Laboratory (Polikarpov *et al.*, 1998). The crystal-to-detector distance was set to 75 mm and a total of 360 images were collected to a resolution of 1.80 Å.

Data were indexed, integrated and scaled with the *HKL*-2000 package (Otwinowski & Minor, 1997). The crystals belong to the monoclinic system (space group $P2_1$), with unit-cell parameters $a$ = 53.26, $b$ = 75.71, $c$ = 53.23 Å, $\beta$ = 96.81°. A $V_{\rm M}$ value of 2.28 Å$^3$ Da$^{-1}$ was calculated according to Matthews (1968), assuming the presence of two protein molecules in the asymmetric unit, and the solvent content was estimated as 46.1%. Data statistics are given in Table 1.

Phasing was performed by molecular replacement using *AMoRe* (Navaza, 1994) from the *CCP*4 program suite v.6.0.2 (Collaborative Computational Project, Number 4, 1994). The template used for molecular replacement was generated by the *SWISS-MODEL* server available at http://swissmodel.expasy.org/SWISS-MODEL.html (Schwede *et al.*, 2003) based on the primary structure of CMS1MS2 and the available structures of caricain (PDB code 1meg), chymopapain (PDB code 1yal) and papain (PDB code 1ppn).

Initial refinement considering each monomer as an independent rigid body was performed with *REFMAC* v.5.2.0019 (Murshudov *et al.*, 1997), yielding an $R$ factor of 43.8% and an $R_{\rm free}$ of 42.5%. Further manual rebuilding of the model and cycles of restrained refinement with water insertion and ligand identification are in progress.

We acknowledge Dr Adriano M. C. Pimenta, Dr Marcelo P. Bemquerer and MSc Alexandre M. C. Santos for their support in mass-spectrometry and protein-sequencing experiments. We also thank Andréia N. Meza and Celisa C. Tonoli for technical assistance in crystallization trials performed at the Brazilian Synchrotron Light Laboratory. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Process 140341/2005-6), Fundação de Desenvolvimento à Pesquisa do Estado de Minas Gerais (FAPEMIG, Process CBB-719/06) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Process PRODOC-0060/05-4).

## References

Agribusiness Development Centre (2000). *ADC Commercialization Bulletin #13: Papain.* Agribusiness Development Centre, Kampala, Uganda.
Baeza, G., Correa, D. & Salas, C. E. (1990). *J. Sci. Food Agric.* **51**, 1–9.
Bravo, L. M., Hermosilla, L. & Salas, C. E. (1994). *Braz. J. Med. Biol. Res.* **27**, 2831–2842.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968). *Nature (London)*, **218**, 929–932.
Gavira, J. A., González-Ramírez, L. A., Oliver-Salvador, M. C., Soriano-García, M. & García-Ruiz, J. M. (2007). *Acta Cryst.* D**63**, 555–563.
Gomes, M. T. R., Mello, V. J., Rodrigues, K. C., Bemquerer, M. P., Lopes, M. T. P., Faça, V. M. & Salas, C. E. (2005). *Planta Med.* **71**, 244–248.
Gravina, M. M., Termignoni, C. & Salas, C. E. (1994). *Plant Sci.* **102**, 11–18.
Hanada, K., Tamai, M., Yamagishi, M., Ohmura, S., Sawada, J. & Tanaka, I. (1978). *Agric. Biol. Chem.* **42**, 523–528.
Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.
Katerelos, N. A., Taylor, M. A., Scott, M., Goodenough, P. W. & Pickersgill, R. W. (1996). *FEBS Lett.* **392**, 35–39.
Konno, K., Hirayama, C., Nakamura, M., Tateishi, K., Tamura, Y., Hattori, M. & Kohno, K. (2004). *Plant J.* **37**, 370–378.
Maes, D., Bouckaert, J., Poortmans, F., Wyns, L. & Looze, Y. (1996). *Biochemistry*, **35**, 16292–16298.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mello, V. J., Gomes, M. T. R., Lemos, F. O., Delfino, J. L., Andrade, S. P., Lopes, M. T. P. & Salas, C. E. (2008). *Phytomedicine*, **15**, 237–244.
Moutim, V., Silva, L. G., Lopes, M. T. P., Fernandes, G. W. & Salas, C. E. (1999). *Plant Sci.* **142**, 115–121.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Polikarpov, I., Perles, L. A., de Oliveira, R. T., Oliva, G., Castellano, E. E., Garratt, R. C. & Craievich, A. (1998). *J. Synchrotron Rad.* **5**, 72–76.
Schwede, T., Kopp, J., Juex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Walreavens, V., Jaziri, M., Van Beeumen, J., Schneck, A. G., Kleinschmidt, T. & Looze, Y. (1993). *Biol. Chem. Hoppe-Seyler*, **374**, 501–506.
Walraevens, V., Vandermeers-Piret, M. C., Vandermeers, A., Gourlet, P. & Robberecht, P. (1999). *Biol. Chem.* **380**, 485–488.