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

# Crystallization and preliminary X-ray diffraction studies of the precursor protein of a thermostable variant of papain

**Sumana Roy, Debi Choudhury, Chandana Chakrabarti, Sampa Biswas* and J. K. Dattagupta**

Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700 064, India

Correspondence e-mail: sampa.biswas@saha.ac.in

Received 2 March 2011
Accepted 23 March 2011

The crystallization of a recombinant thermostable variant of pro-papain has been carried out. The mutant pro-enzyme was expressed in *Escherichia coli* as inclusion bodies, refolded, purified and crystallized. The crystals belonged to space group $P2_1$, with unit-cell parameters $a$ = 42.9, $b$ = 74.8, $c$ = 116.5 Å, $\beta$ = 93.0°, and diffracted to 2.6 Å resolution using synchrotron radiation. Assuming the presence of two molecules in the asymmetric unit, the calculated Matthews coefficient is 2.28 Å$^3$ Da$^{-1}$, corresponding to a solvent content of 46%. Initial attempts to solve the structure using molecular-replacement techniques were successful.

## 1. Introduction

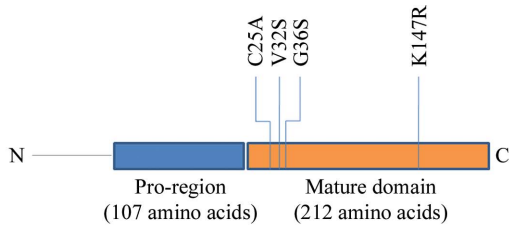
Papain-like cysteine proteases represent a major component of the lysosomal proteolytic enzymes and are involved in various biological functions (Dickinson, 2002); the enzymes from plant sources are often used in industry (Grzonka *et al.*, 2007). The cysteine proteases are expressed as an inactive precursor in a pre-pro-enzyme form which contains a signal peptide (pre-region), an inhibitory pro-region and a mature catalytic domain. The pro-region is generally cleaved at acidic pH to produce the active enzyme (Brocklehurst *et al.*, 1985). The pro-peptide parts of the cysteine proteases are essential for the correct folding, transport, maturation and regulation of activity of the enzyme (Winther & Sørensen, 1991; Baker *et al.*, 1992; Wetmore *et al.*, 1992; Fukuda *et al.*, 1994). It has been observed that the pro-parts of cysteine proteases act as effective inhibitors of their cognate enzymes and also to some extent of other cysteine proteases of the same family (Mach *et al.*, 1994; Taylor *et al.*, 1995; Fox *et al.*, 1992).

Knowledge of the three-dimensional structure of a pro-protease will help us to better understand the molecular basis of inhibition by the pro-region and its role in the activation and folding processes of the enzyme. Although the crystal structure of mature papain is known (Drenth *et al.*, 1968), no report exists on the structure of pro-papain. Attempts to crystallize pro-papain in our laboratory were not successful, mainly because it precipitated during the concentration of the protein for crystallization trials. However, we could generate a thermostable variant (pro-pap-RSS) of pro-papain with three mutations in the interdomain region, K174R, G36S and V32S; the other functional properties of pro-pap-RSS (apart from the enhancement of thermostability) remained the same as those of the wild-type pro-papain (Choudhury *et al.*, 2010). We have performed a preliminary X-ray analysis of this variant of pro-papain with the aim of deciphering the three-dimensional structure of the pro-region and its interactions with the catalytic domain. A recombinant mutant of the variant was constructed by replacing the active-site cysteine residue Cys25 with an alanine (pro-pap-RSS-C25A; Fig. 1) for this study in order to avoid the autocatalytic activation process.

## 2. Materials and methods

### 2.1. Expression and purification

The thermostable variant of pro-papain, pro-pap-RSS, cloned in pET-30 Ek/LIC (Novagen, USA; Choudhury *et al.*, 2010) was used




(*a*)

DFSIVGYSQNDLTSTERLIQLFESWMLKHNKIYKNIDEKIYRFEIFKDNLKYIDETNK
KNNSYWLGLNVFADMSNDEFKEKYTGSIAGNYTTTELSYEEVLNDGDVNIPEYVDW
RQKGAVTPVKNQGSCGSAWAFSAVSTIESSIIKIRTGNLNEYSEQELLDCDRRSYGCN
GGYPWSALQLVAQYGIHYRNTYPYEGVQRYCRSREKGPYAAKTDGVRQVQPYNEG
ALLYSIANQPVSVVLEAAGKDFQLYRGGIFVGPCGNKVDHAVAAVGYGPNYILIRNS
WGTGWGENGYIRIKRGTGNSYGVCGLYTSSFYPVKN

(*b*)

**Figure 1**
A schematic representation of recombinant pro-pap-RSS-C25A expressed in *E. coli* using pET-30 Ek/LIC vector. (*a*) Recombinant pro-pap-RSS-C25A with a 44-amino-acid vector sequence with a hexahistidine tag at the N-terminus (black line), a 107-residue pro-region (blue bar) and a 212-residue mature catalytic domain (orange bar). The positions of the mutated residues are shown. (*b*) The amino-acid sequence of the pro and mature domains of pro-pap-RSS-C25A are shown in the corresponding colours, with the four mutations indicated in bold.

to generate the pro-pap-RSS-C25A mutant following the standard protocol of the QuikChange site-directed mutagenesis kit (Stratagene, USA) with forward primer 5′-CAGGGTTCTTGTGGTAGT-**GCG**TGGGCATTCTCAGCTGTTG-3′ and reverse primer 5′-CA-ACAGCTGAGAATGCCCA**CGC**ACTACCACAAGAACCCTG-3′. The PCR reaction involved 16 cycles of amplification using *Pfu* DNA polymerase (according to the manufacturer's instructions) and the amplified product was transformed into *Escherichia coli* strain DH10B. The identity of the mutant clone was verified by automated DNA sequencing using the MegaBACE 1000 sequencing system (Amersham Biosciences, USA).

A recombinant clone with the desired mutation was transformed into *E. coli* strain BL21 (DE3) and a single colony was cultured in LB medium with 50 µg ml$^{-1}$ kanamycin at 310 K to an $OD_{600}$ of 0.6–0.8 and then induced with 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside

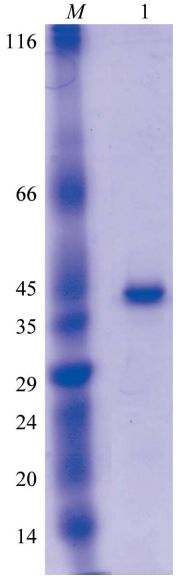



**Figure 2**
15% SDS–PAGE analysis of purified pro-pap-RSS-C25A. Lane *M*, protein molecular-weight markers (kDa); lane 1, purified mutant protein after gel-filtration chromatography. The gel was stained with Coomassie Brilliant Blue R-250.

**Table 1**
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

| | |
|---|---|
| Data-collection parameters | |
| X-ray source | BM14, ESRF, Grenoble, France |
| Wavelength (Å) | 0.9785 |
| Temperature (K) | 100 |
| Oscillation range (°) | 1 |
| Crystal-to-detector distance (mm) | 279.95 |
| Data-integration statistics | |
| Space group | $P2_1$ |
| Unit-cell parameters (Å, °) | $a$ = 42.9, $b$ = 74.8, $c$ = 116.5, $\beta$ = 93.0 |
| Mosaicity (°) | 1.04 |
| Resolution range (Å) | 50–2.6 (2.64–2.60) |
| Observed reflections | 157456 |
| Unique reflections | 21685 |
| Multiplicity | 7.3 (6.9) |
| Completeness (%) | 97.0 (94.6) |
| $R_{merge}$† (%) | 5.7 (68.2) |
| $\langle I/\sigma(I)\rangle$ | 28.4 (2.0) |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the $i$th observation of reflection $hkl$ and $\langle I(hkl)\rangle$ is the average intensity of reflection $hkl$.

(IPTG) for 5 h with vigorous shaking. The cells were harvested by centrifugation, resuspended in a buffer consisting of 20 m*M* Tris–HCl pH 8.0, 1% Triton X-100, 5% sucrose and 2.5 m*M* DTT and lysed by sonication. The recombinant protein was expressed as inclusion bodies and the expressed protein was solubilized in urea, purified on an Ni–NTA (Novagen, USA) column and refolded using the dilution method as described previously (Choudhury *et al.*, 2009).

After refolding, the protein was concentrated and purified further by gel-filtration chromatography on a Sephacryl S-100 (Amarsham, USA) column that had been pre-equilibrated with 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl and 1% glycerol. The protein was eluted in the same buffer and the peak fractions were collected. The purity of the protein was checked by SDS–PAGE analysis. Around 300 mg refolded and purified protein was obtained from 1 l bacterial culture.

### 2.2. Crystallization and data collection

The data were processed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997) and the data-collection and processing statistics are summarized in Table 1.

**Figure 3**
A crystal of pro-pap-RSS-C25A. The approximate crystal dimensions are 300 × 200 × 100 µm.

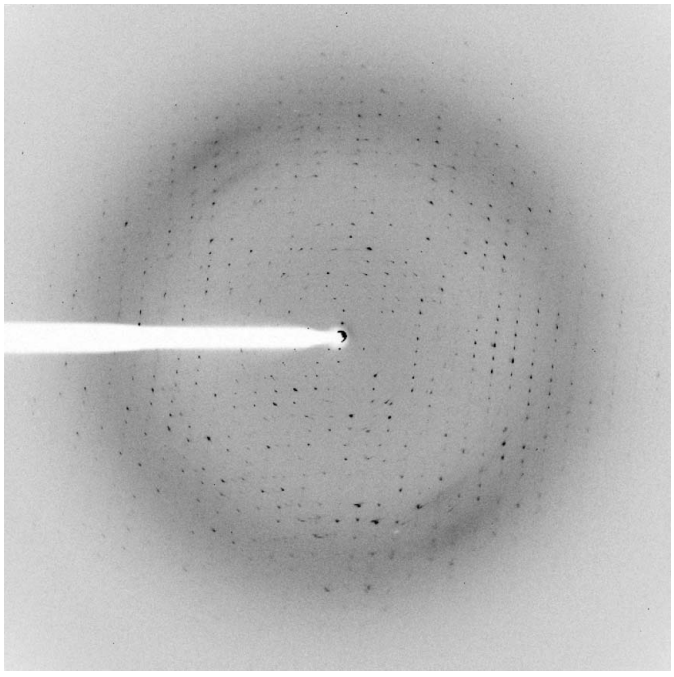


**Figure 4**
Diffraction of a pro-pap-RSS-C25A crystal. The resolution at the edge of the plate is 2.6 Å.

## 3. Results and discussion

Recombinant pro-pap-RSS-C25A was expressed in inclusion bodies in *E. coli*. The purity and molecular weight (41 kDa) of the refolded purified protein were checked by SDS–PAGE (Fig. 2) and it was concentrated to 5 mg ml$^{-1}$ for crystallization trials. Diffraction-quality crystals (of dimensions ∼300 × 200 × 100 µm) were obtained at 293 K after about 70 d (Fig. 3) using the hanging-drop vapour-diffusion method in condition No. 37 of Crystal Screen (Hampton Research, USA). The pro-pap-RSS-C25A crystals diffracted to 2.6 Å resolution (Fig. 4) and belonged to space group $P2_1$, with unit-cell parameters $a$ = 42.9, $b$ = 74.8, $c$ = 116.5 Å, $\beta$ = 93.0°. Assuming the presence of two molecules in the asymmetric unit, the calculated Matthews coefficient is 2.28 Å$^3$ Da$^{-1}$ with a solvent content of 46% (Matthews, 1968). A model of a chimeric pro-protease structure was generated from the structures of mature native papain (PDB code 9pap; Kamphuis *et al.*, 1984) and the pro-domain of pro-caricain (PDB code 1pci; Groves *et al.*, 1996), which has 89% identity and 98% similarity in amino-acid sequence to the pro-region of pro-pap-RSS-C25A. This model, with mismatched residues mutated to alanine, was subsequently used as a search model in molecular-replacement calculations using *MOLREP* (Vagin & Teplyakov, 2010) from the *CCP*4 program suite (Winn *et al.*, 2011). The cross-rotation function for this model was calculated for the entire resolution range and the first two peaks in the output appeared as distinct solutions. These two peaks became more prominent in the translation calculations, with an overall correlation coefficient of 56.0% and an $R$ factor of 45.3%. This solution revealed good crystal packing and no clashes between symmetry-related molecules. Further analysis is in progress.

The work was partially supported by CSIR (Grant No. 21/0653/06/EMR-II), DST (SR/WOS-A/LS-161/2004) and DBT (BM14 beamline project for synchrotron data collection at ESRF, Grenoble), Government of India. We are grateful to the beamline scientists for help in data collection.

## References

Baker, D., Silen, J. L. & Agard, D. A. (1992). *Proteins*, **12**, 339–344.
Brocklehurst, K., Salih, E., McKee, R. & Smith, H. (1985). *Biochem. J.* **228**, 525–527.
Choudhury, D., Biswas, S., Roy, S. & Dattagupta, J. K. (2010). *Protein Eng. Des. Sel.* **23**, 457–467.
Choudhury, D., Roy, S., Chakrabarti, C., Biswas, S. & Dattagupta, J. K. (2009). *Phytochemistry*, **70**, 465–472.
Dickinson, D. P. (2002). *Crit. Rev. Oral Biol. Med.* **13**, 238–275.
Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968). *Nature (London)*, **218**, 929–932.
Fox, T., de Miguel, E., Mort, J. S. & Storer, A. C. (1992). *Biochemistry*, **31**, 12571–12576.
Fukuda, R., Horiuchi, H., Ohta, A. & Tagagi, M. (1994). *J. Biol. Chem.* **269**, 9551–9561.
Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W. & Jenkins, J. A. (1996). *Structure*, **4**, 1193–1203.
Grzonka, Z., Kasprzykowski, F. & Wiczk, W. (2007). *Industrial Enzymes: Structure, Function and Applications*, edited by J. Polaina & A. P. MacCabe, pp. 181–195. Dordrecht: Springer.
Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.
Mach, L., Mort, J. S. & Glössl, J. (1994). *J. Biol. Chem.* **269**, 13030–13035.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Taylor, M. A., Baker, K. C., Briggs, G. S., Connerton, I. F., Cummings, N. J., Pratt, K. A., Revell, D. F., Freedman, R. B. & Goodenough, P. W. (1995). *Protein Eng.* **8**, 59–62.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wetmore, D. R., Wong, S.-L. & Roche, R. S. (1992). *Mol. Microbiol.* **6**, 1593–1604.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Winther, J. R. & Sørensen, P. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 9330–9334.