Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of a thermostable mutant of pro-papain reveals its activation mechanism

Sumana Roy,‡ Debi Choudhury,‡ Pulakesh Aich, Jiban K. Dattagupta and Sampa Biswas*

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

‡ These authors contributed equally to this work

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

Received 1 August 2012
Accepted 8 September 2012

PDB Reference: thermostable mutant of pro-papain, 3tnx

Papain is the archetype of a broad class of cysteine proteases (clan C1A) that contain a pro-peptide in the zymogen form which is required for correct folding and spatio-temporal regulation of proteolytic activity in the initial stages after expression. This study reports the X-ray structure of the zymogen of a thermostable mutant of papain at 2.6 Å resolution. The overall structure, in particular that of the mature part of the protease, is similar to those of other members of the family. The structure provides an explanation for the molecular basis of the maintenance of latency of the proteolytic activity of the zymogen by its pro-segment at neutral pH. The structural analysis, together with biochemical and biophysical studies, demonstrated that the pro-segment of the zymogen undergoes a rearrangement in the form of a structural loosening at acidic pH which triggers the proteolytic activation cascade. This study further explains the bimolecular stepwise autocatalytic activation mechanism by limited proteolysis of the zymogen of papain at the molecular level. The possible factors responsible for the higher thermal stability of the papain mutant have also been analyzed.

1. Introduction

As in other proteases, spatio-temporal regulation of proteolytic activity is important for papain-like cysteine proteases (clan C1A) to control undesired proteolysis (Khan & James, 1998). This regulation is mediated by different means in different stages of the cell cycle, starting from control at the gene-expression level to translocation to different cellular compartments, pro-peptide-mediated zymogenic control and regulation by different specific protein inhibitors of the mature and active protease and finally degradation of the protease by cell-housekeeping machinery (Turk *et al.*, 2012).

In the inactive (latent) zymogens, proteolysis is regulated by the pro-peptide segment, which blocks the active-site cleft by positioning itself in a nonproductive orientation (the reverse of a natural substrate) and thereby preventing access of substrates to the catalytic site of the enzyme (Cygler *et al.*, 1996; Coulombe *et al.*, 1996). This pro-peptide segment is also known to have foldase activity; it folds itself independently and then catalyses the folding of the mature part into its native state by lowering the free-energy barrier of folding. Subsequently, the pro-peptide is removed by limited proteolysis during maturation (Wiederanders *et al.*, 2003). Pro-peptides sometimes contain intracellular sorting signals which help to traffick the zymogens to appropriate cellular compartments (Wiederanders *et al.*, 2003). Therefore, studies of the mechanism of pro-peptide-regulated latency and activation of

the proteases of this family at the molecular level are necessary in order to understand their basic modes of action. Such studies are also important for the design of potential drugs by mimicking the specific conformation of the active-site-blocking peptide in the latent state as part of a therapeutic approach.

Papain, a cysteine protease of plant origin and an archetypal member of the family, was the first protease in the family for which the crystal structure was determined (Drenth *et al.*, 1968). However, no structural information has been available until now on the precursor form of papain, although zymogen structures of other members of the family have been determined (Coulombe *et al.*, 1996; Groves *et al.*, 1996; LaLonde *et al.*, 1999; Kaulmann *et al.*, 2006). This study reports the X-ray crystallographic structure of a pro-papain mutant and correlates the structure with the pH-regulated stepwise zymogen-activation mechanism of papain at the molecular level. The structural information available from this study will also shed light on the pro-peptide/substrate specificity of the papain active-site cleft since pro-peptide side chains utilize the same binding sites as those of substrates. *In vitro* processing of the papain precursor has been analysed previously by biochemical, kinetic and mutational studies (Vernet *et al.*, 1991, 1995) and demonstrated that the processing occurs in a stepwise manner with a combination of intermolecular and intramolecular cleavages. Probable cleavage sites have also been postulated (Vernet *et al.*, 1995). However, the observations on stepwise activation and the identification of cleavage sites were not conclusive owing to the rapid activation process of pro-papain, which meant that trapping of transition intermediate(s) was not possible, and also owing to the lack of a zymogen structure of papain. In this study, we have attempted to address these problems by generating a slow-activating mutant of the papain precursor and solving its X-ray structure.

Our attempts to crystallize wild-type pro-papain or its active-site cysteine mutant (Cys→Ala) were not successful. However, its thermostable mutant turned out to be amenable to crystallization. It is worth mentioning that the positions of the three mutations (K174mR, V32mS and G36mS) which were generated for higher thermostability of this pro-papain variant are distant from the catalytic cleft and the pro-segment/mature segment interface. Also, it has been established in this study as well as in our previous work (Choudhury *et al.*, 2010) that the catalytic activity and specificity of this mutant are similar to those of wild-type papain. Therefore, structural analysis of this papain mutant should also help us to understand the zymogen structure and the activation mechanism of papain itself. We further noticed that time course of zymogen activation is slower in this thermostable mutant compared with the wild type (Choudhury *et al.*, 2009) and that this slower activation process is advantageous because it enables us to trap the intermediates generated during activation even within a time frame of 10 min. Thus, this papain mutant is a good model system to study the mechanism of zymogen activation of papain at the molecular level, particularly because such a slow activation rate has not been observed for other papain-like proteases (McQueney *et al.*, 1997; Ménard *et al.*, 1998).

2. Experimental procedures

2.1. Cloning, expression and purification of recombinant proteins

The thermostable mutant of the papain precursor, pro-pap-RSS[1] (Choudhury *et al.*, 2010), was used to generate the pro-pap-RSS-C25mA mutant (Roy *et al.*, 2011). The expression, purification and refolding of these two proteins (pro-pap-RSS and pro-pap-RSS-C25mA) were carried out using established protocols (Choudhury *et al.*, 2009). Immediately after the refolding of pro-pap-RSS, 1 m*M* $HgCl_2$ was added to block the catalytic cysteine and to prevent autocatalytic degradation of the refolded protein. Recombinant pro-pap-RSS contains a 44-amino-acid vector sequence including a hexahistidine tag at the N-terminus followed by 107 residues of the pro-segment and 212 residues of the mature segment containing the catalytic domain. The construct has three mutations in the inter-domain region of the mature part: K174mR, G36mS and V32mS. An inactive form, pro-pap-RSS-C25mA, was generated which contains an additional mutation in which the catalytic cysteine is replaced by an alanine (C25mA) to avoid autocatalytic activation. The refolded pro-pap-RSS-C25mA was further purified by size-exclusion chromatography on a Sephacryl S-100 column for structural studies.

2.2. Profiling of activation/processing of pro-pap-RSS to mature pap-RSS

2.2.1. Analysis by SDS–PAGE. To capture the intermediates during the conversion of pro-pap-RSS to mature pap-RSS, the zymogen was incubated at 313 K (at this temperature the activation process is slower, which helps in capturing the intermediates) in 0.1 *M* sodium acetate buffer pH 4.0 with 0.002 *M* EDTA in the presence of 0.005 *M* DTT for 0–2 h. At 10 min intervals, aliquots of the activated samples were inhibited by adding 10 µ*M* E-64, boiled with an equal volume of 2× SDS–PAGE sample buffer and run on a 12% polyacrylamide gel. After electrophoresis, the molecular masses of the bands observed in the gel were determined using the *ImageQuant TL* software (GE Healthcare).

2.2.2. Analysis by nondenaturing gelatine SDS–PAGE. The progress of the conversion of the pro-enzyme to the mature enzyme was also followed by gelatine gel zymography (Michaud, 1998) to detect the proteolytically active intermediates during the activation process. For this, aliquots were removed at 10 min intervals during the activation process (at pH 4.0, as described above), inhibited with the reversible inhibitor antipain (Suzuki *et al.*, 1981) in a molar excess to prevent further activation and autocatalysis of the remaining

[1] Abbreviations: pro-pap-RSS, a thermostable mutant of pro-papain with K174mR, V32mS and G36mS mutations in the mature part; pap-RSS, a thermostable mutant of mature papain with K174mR, V32mS and G36mS mutations in the mature part; pro-pap-RSS-C25mA, active-site cysteine knockout with a C25mA mutation in pro-pap-RSS; p, used as a suffix to indicate amino acid(s) in the pro-segment; m, used as a suffix to indicate amino acid(s) in the mature part; E-64, 1-[L-*N*-(*trans*-epoxysuccinyl)leucyl]amino-4-guanidinobutane; BAPNA, *N*-benzoyl-DL-arginine *p*-nitroanilide; pEFLNA, L-pyroglutamyl-L-phenylalanyl-L-leucyl-*p*-nitroanilide; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; NMA, normal-mode analysis.

pro-enzyme/mature enzyme and immediately frozen to trap the conformations of the proteins. Equal volumes of 2× nonreducing SDS–PAGE sample buffer were added to all samples prior to loading. The samples were run on a 12% polyacrylamide gel copolymerized with 0.1% gelatine at 277 K. After electrophoresis, the gel was washed with 2.5% Triton X-100 for 30 min at room temperature to remove SDS, washed with distilled water and then incubated for 1 h in activity buffer (0.1 *M* sodium acetate pH 5.5 with 0.002 *M* EDTA, 0.005 *M* DTT), which reversed the inhibition by antipain. The gel was subsequently incubated for 15 h in activity buffer without DTT. The processed gel was stained with Coomassie Brilliant Blue R-250 and proteinase activity was detected by appearance of clear bands on a dark blue background.

2.2.3. Amino-acid sequencing. The separated intermediate products from SDS–PAGE were transferred onto a PVDF membrane using a Hoefer TE 22 Mighty Small Transphor Tank Transfer Unit with CAPS buffer pH 11. The transferred protein fragments on PVDF membrane were stained with Coomassie Brilliant Blue (0.1% Coomassie Brilliant Blue R-250 in 40% methanol, 1% acetic acid). The desired bands were cut and N-terminal sequencing was performed on an Applied Biosystems Procise Sequencer at the National Facility for Protein Sequencing, IIT, Bombay, India. Five cycles were performed to identify the N-termini of the cleaved fragments.

2.3. Measurement of kinetic parameters

Enzyme activity (K_m and k_{cat}) was determined at 298 K by continuously monitoring the release of *p*-nitroaniline (*p*-NA) from BAPNA and pEFLNA (Sigma, Missouri, USA) at 410 nm using a $\Delta\varepsilon$ of 8800 M^{-1} cm^{-1} for *p*-NA (Mole & Horton, 1973). Complete conversion of pro-pap-RSS to mature pap-RSS was optimized (Choudhury *et al.*, 2010). Accordingly, the pro-enzyme was incubated at 333 K in 0.1 *M* sodium acetate buffer pH 4.0 containing 0.002 *M* EDTA and 0.02 *M* cysteine for 45 min in order to obtain mature pap-RSS. Immediately after conversion, a pre-determined amount of BAPNA or pEFLNA in 0.05 *M* sodium acetate buffer pH 6.5 containing 0.002 *M* EDTA, 0.002 *M* cysteine and 0.1% Brij 35 was added and the release of *p*-NA was monitored for 20 min at 410 nm. *GraphPad Prism* (http://www.graphpad.com/prism) was used to obtain K_m and V_{max} values by fitting the reaction velocities to the Michaelis–Menten equation. The k_{cat} values were determined using the equation $k_{cat} = V_{max}/[E]_T$. $[E]_T$ is the total concentration of the active enzyme, the values of which were measured by active-site titration with the irreversible inhibitor E-64, which binds in a 1:1 molecular ratio to the active sites of most cysteine proteases (Barrett *et al.*, 1982). The kinetic parameters for commercial papain were also determined with pEFLNA for comparison.

2.4. Fluorescence measurements

The intrinsic tryptophan fluorescence of recombinant pro-pap-RSS-C25mA was determined at a protein concentration of 1.28 µ*M* with appropriate blanks for baseline correction. To measure the tryptophan fluorescence intensity quenching at acidic pH, the protein was equilibrated in 0.1 *M* sodium acetate buffer at different pH values (4.0 and 5.5 separately) at 298 K for 10 min and the fluorescence spectra were recorded as described below. At the same time, fluorescence spectra were also measured after incubating the protein at 333 K for 45 min at pH 4.0 to mimic the conditions of zymogen activation. All fluorescence experiments were performed with the help of a Varian Cary Eclipse fluorescence spectrophotometer at room temperature (298 K) using a fixed cuvette of dimensions 2 × 10 mm and a scan rate of 300 nm min^{-1}. Tryptophan emission spectra were recorded from 305 to 550 nm by exciting the protein solution at 295 nm. The band pass for both excitation and emission was set at 5 nm.

Table 1
Crystallographic data and refinement statistics for pro-pap-RSS-C25mA.

Values in parentheses are for the highest resolution shell.

X-ray source	BM14, ESRF
Wavelength (Å)	0.9785
Temperature (K)	100
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 42.9, b = 74.8, c = 116.5, β = 93.0
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}† (%)	8.6 (87.9)
$\langle I/\sigma(I)\rangle$	28.4 (2.0)
Refinement	
Resolution (Å)	50–2.6
R_{work}†/R_{free}‡ (%)	18.29/23.59
No. of refined atoms	
Proteins	4945
Waters	133
Cl^- ion	1
Average B factor (Å^2)	65.18
Ramachandran statistics	
Allowed (%)	99.6
Outliers (%)	0.4
R.m.s.d. bonds (Å)	0.017
R.m.s.d. angles (°)	1.64

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the average intensity of reflection hkl. ‡ $R_{work} = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$. R_{free} is the cross-validation R factor for the test set (5%) of reflections omitted in model refinement.

2.5. Crystal structure determination of pro-pap-RSS-C25mA

Crystals of pro-pap-RSS-C25mA were grown by the hanging-drop vapour-diffusion method at 293 K as described previously (Roy *et al.*, 2011). X-ray diffraction experiments were performed on beamline BM14 at the ESRF, Grenoble, France; a complete diffraction data set was collected to 2.6 Å resolution. The data were processed and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

The structure of pro-pap-RSS-C25mA was determined by the molecular-replacement method using *MOLREP* within the *CCP*4 suite (Winn *et al.*, 2011). The coordinates for a model of a chimeric pro-protease structure were generated by combining the structures of native mature papain (Kamphuis *et al.*, 1984; PDB entry 9pap) and the pro-domain

of pro-caricain (Groves *et al.*, 1996; PDB entry 1pci) as described previously (Roy *et al.*, 2011). Since Matthews analysis (Matthews, 1968) suggested the presence of two molecules in the crystal asymmetric unit, 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%. After a few cycles of rigid-body refinement followed by restrained refinement using *REFMAC*5 (Murshudov *et al.*, 2011), we started mutating amino acids in the electron density and model building with *Coot* (Emsley & Cowtan, 2004). TLS refinement in *REFMAC*5 and *CNS* (Brünger *et al.*, 1998) was used in the final stage of refinement, which yielded an *R* factor of 18.29% and an $R_{\rm free}$ of 23.59%. The first nine residues of the pro-segment in both molecules (*A* and *C*) and the side chain of three residues, Tyr97p in molecules *A* and *C* and Asn102p in molecule *C*, could not be identified in the electron density. In the difference electron-density map ($F_{\rm o} - F_{\rm c}$) a strong peak was observed which was assigned as a chloride ion since 150 m*M* NaCl was present in the protein buffer used for crystallization (Roy *et al.*, 2011). The location of the chloride ion was also consistent with a positively charged environment. A total of 133 water molecules were located in the structure. Data-collection and refinement statistics are given in Table 1.

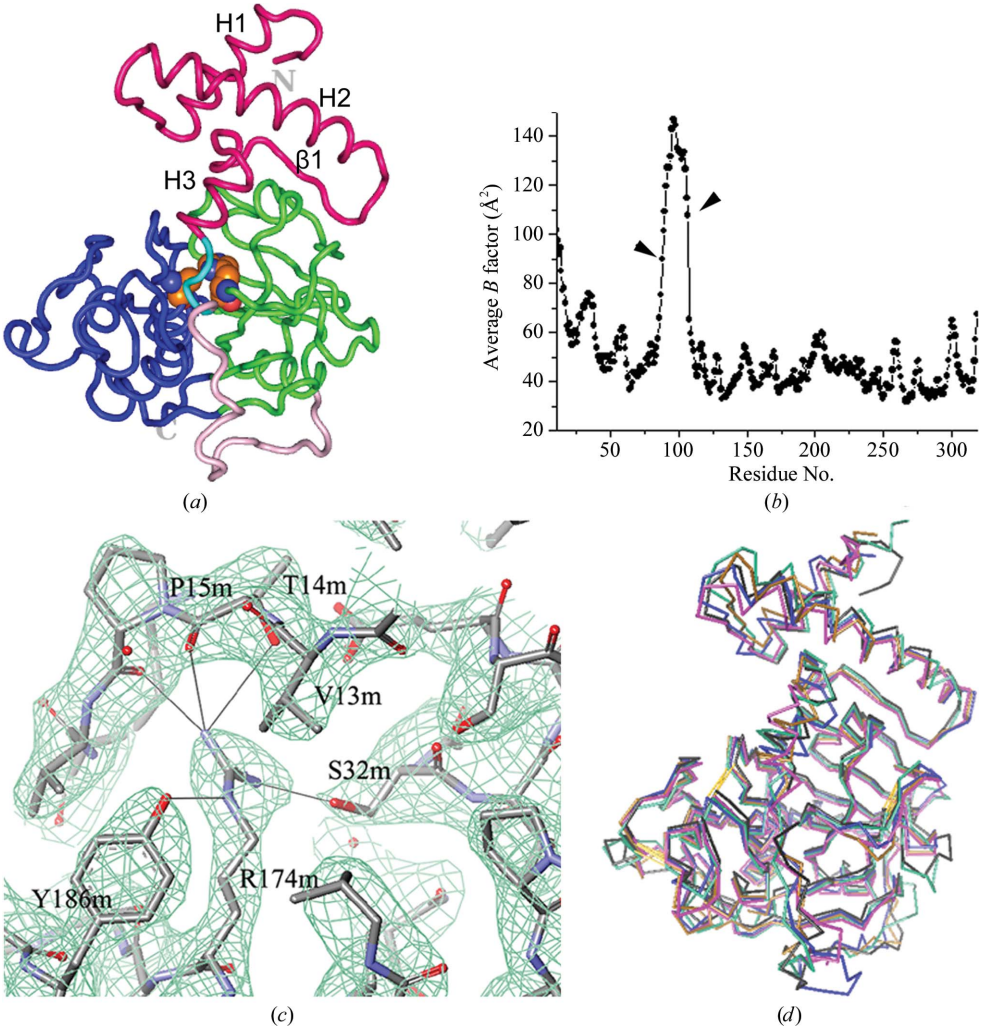

Figure 1
Overall structure of pro-pap-RSS. (*a*) Structural overview of pro-pap-RSS-C25mA represented as a cartoon diagram. The pro-domain part of the pro-segment is represented in deep magenta, while the extended pro-peptide part blocking the catalytic cleft is represented in cyan and the rest of the extended loop connecting the mature part is represented in light pink. The L-domain and R-domain of the mature segment are represented in blue and green, respectively. The catalytic residues Cys25mA and His159m are represented as spheres. (*b*) A plot of the average *B* factor against residue number of the protein. The region within the two arrows represents the high-*B*-factor residues (87p–107p) of the extended loop of the pro-segment connecting the mature part (light pink region of the cartoon diagram). (*c*) A representative section of electron-density map ($2F_{\rm o} - F_{\rm c}$) contoured at 1.2σ. (*d*) Superposition of the C^{α} traces of pro-pap-RSS-C25mA (black), pro-caricain (light green) and pro-cathepsins L (yellow), K (blue) and S (magenta).

Since the conformations of the two molecules (*A* and *C*) in the asymmetric unit are similar (C^{α} r.m.s.d. of 0.35 Å), all subsequent structural analyses are based on molecule *A*.

The stereochemistry of the final model of the structure was checked using *PROCHECK* (Laskowski *et al.*, 1993). Molecular images were generated using the programs *Discovery Studio* (Accelrys Inc.) and *PyMOL* (http://www.pymol.org). Atomic coordinates and structure factors have been deposited in the Protein Data Bank as entry 3tnx.

2.6. Normal-mode analysis and comparative modelling

Normal-mode analysis (NMA) is a powerful tool for predicting collective molecular motions of biological macromolecules (Tama, 2003). The web-based program *ElNémo* (http://igs-server.cnrs-mrs.fr/elnemo/index.html), an interface to the elastic network model, was used on the pro-pap-RSS-C25mA structure for this purpose to generate the low-frequency mode(s). We calculated five normal modes (NMODE = 5) and the perturbation was from −100 (DQMIN = −100) to +100 (DQMAX = +100) in 20 steps (DQSTEP = 20). Default values were used for the other parameters. Of the five normal modes given by *ElNémo* (modes 7, 8, 9, 10 and 11), mode 10 was discarded owing to low

collectivity. The other modes were chosen for domain analysis by the web-based program *DynDom* (http://fizz.cmp.uea.ac.uk/dyndom/). The PDB files containing the conformations of each mode were given as inputs for domain analysis in *DynDom*, in which the default parameters were used.

The structures of papain complexed with two proteinaceous inhibitors, stefin (Stubbs *et al.*, 1990; PDB entry 1stf) and chagasin (Redzynia *et al.*, 2009; PDB entry 3e1z), were chosen to compare the mode of binding of the inhibitors and the pro-peptide segment blocking the catalytic cleft of papain. The papain parts of the structures of the two papain–inhibitor complexes and pro-pap-RSS-C25mA were superimposed for this analysis. The contact surface areas of the inhibitors and the pro-peptide part with the mature papain part were calculated using the program *Discovery Studio* (Accelrys Inc.). The mature parts of four zymogen structures of the papain-like proteases pro-caricain and pro-cathepsins L, K and S (Groves *et al.*, 1996; Coulombe *et al.*, 1996; LaLonde *et al.*, 1999; Li *et al.*, 1993; PDB entries 1pci, 1cjl, 1by8 and 1coy, respectively) were superimposed to analyze the similarity/dissimilarity and the relative orientation of the pro-domains with respect to the superimposed mature part.

3. Results

3.1. Overall structure of pro-pap-RSS-C25mA

The properly folded pro-pap-RSS-C25mA was crystallized; the crystal provided X-ray diffraction data to 2.6 Å resolution and yielded a detailed picture of the zymogenic structure of the pro-papain mutant (Fig. 1*a*). The molecular structure of the zymogen has an ellipsoidal shape with approximate dimensions of 55 × 45 × 35 Å. The structure of the molecule can be divided into two major regions: the N-terminal pro-segment (Asn10p–Asn107p) and the mature segment (Ile1m–Asn212m) that contains the catalytic centre. A section of the representative electron-density map is shown in Fig. 1(*c*).

3.2. The pro-segment structure

The N-terminal part of the pro-segment (10p–83p) forms a domain (the pro-domain, which acts as a scaffold) that is structurally similar to the pro-domains of other cathepsin L-type papain-like proteases (Coulombe *et al.*, 1996; Groves *et al.*, 1996; LaLonde *et al.*, 1999; Kaulmann *et al.*, 2006) and consists of three α-helices, H1 (14p–29p), H2 (35p–57p) and H3 (74p–83p), and a short β-strand, β1 (63p–65p) (Fig. 1*a*). The rest of the pro-segment beyond helix H3 (84p–107p) is in an extended conformation ('extended pro-peptide'). The structure of the pro-domain of pro-pap-RSS-C25mA is similar to that observed in the structures of pro-caricain and pro-cathepsins L, K and S (C^{α} r.m.s.d. of $\leq$ 0.8 Å) with certain differences in helix H1 and in the adjoining loop region and in helix H3 and the extended peptide of the pro-segment that blocks the catalytic cleft (Fig. 1*d*).

Three residues from the C-terminus of helix H3 (Thr83p, Lys81p and Phe78p) nestle in the primed subsites of the substrate-binding cleft. The pro-peptide beyond Ile86p has poor electron density and high thermal parameters (Fig. 1*b*) compared with the other parts of the molecule. Helices H1 and H2 are almost orthogonal to each other and their crossover forms a hydrophobic core mediated by Phe22p and Trp26p from helix H1 and Phe43p and Phe46p from helix H2 (Fig. 2*a*). A salt bridge between Glu23p OE2 and Lys39p NZ stabilizes the two long helices H1 and H2 of the pro-domain. The side chain of Asp72p

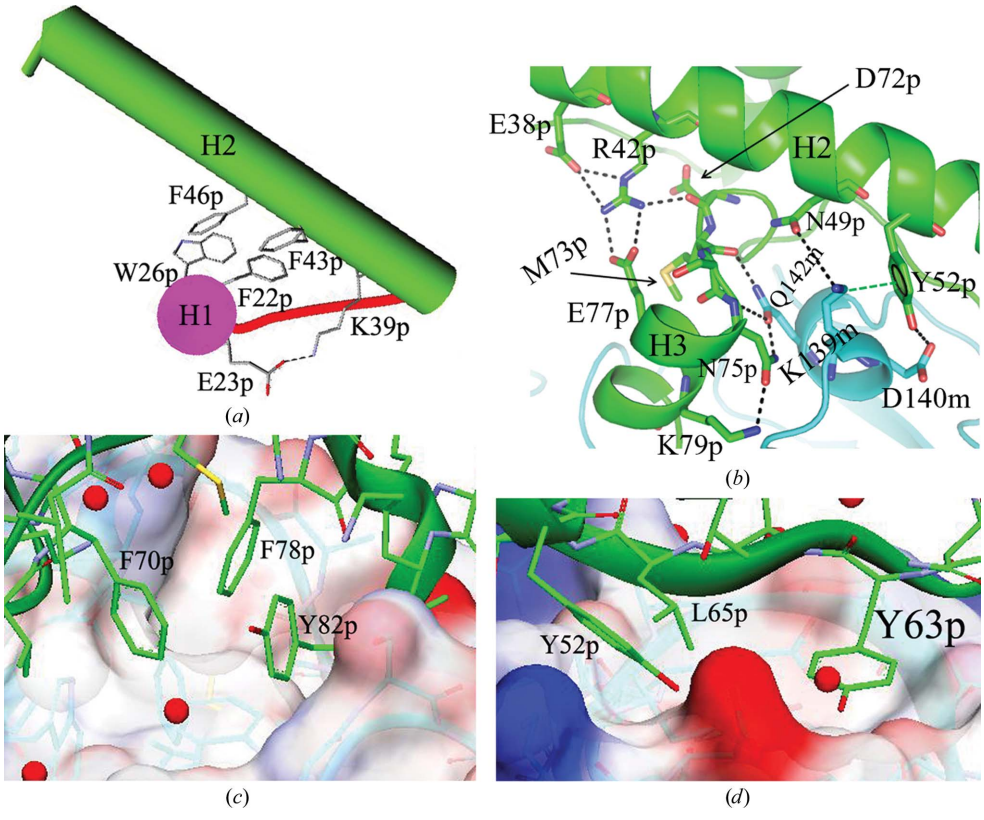

Figure 2
Structural stability of the pro-domain. (*a*) Crossover of helices H1 and H2 in cartoon representation. (*b*) Electrostatic interactions within the pro-domain and at the interface between the pro-segment and mature segment. Hydrogen bonds and salt bridges within 3.5 Å and the π–cation interaction are indicated by black and green lines, respectively. The mature segment is represented by a cyan ribbon and the pro-domain is represented as a green ribbon. (*c*, *d*) Two hydrophobic pockets of mature enzymes formed by Phe149m, Trp181m, Phe141m, Trp177m, Cys200m and Cys153m for the first pocket and Phe149m, Pro151m and Tyr144m for the second pocket in which hydrophobic residues of the pro-domain are stabilized. The residues of the mature segment are not labelled in the figure.

Table 2
Interactions in the structure of pro-pap-RSS-C25mA involving side chains of aspartic acid and glutamic acid residues of the pro-segment and those of the mature segment at the interface with the pro-domain which are likely to be affected owing to protonation at the lower pH of 4.0.

Residue	Position	Interacting residue	Position	Distance (Å)
Glu23p OE2	Helix H1 of pro-domain	Lys39p NZ	Helix H2 of pro-domain	2.76
Glu38p OE2	Helix H2 of pro-domain (in ERFNIN motif)	Arg42p NH2	Helix H2 of pro-domain (in ERFNIN motif)	3.46
Glu38p OE1		Lys34p N		2.98
Glu44p OE2	Helix H2 of pro-domain	Lys47p NZ	Helix H2 of pro-domain	3.44
Asp72p OD2	Loop region at the N-terminus of helix H3	Lys31p NZ	Loop region connecting helices H1 and H2	2.85
Asp72p OD1	(in GNFD motif)	Tyr33p OH		2.54
Glu77p OE1	Helix H3 of pro-domain	Arg42p NH1	Helix H2 of pro-domain (in ERFNIN motif)	3.30
Glu77p OE2		Arg42p NH2		3.00
Asp140m OD1	Surface loop of R-domain of mature segment	Tyr52p OH	Helix H2 of pro-domain	3.14
Asp158m OD1	Surface loop of R-domain of mature segment	Thr83p OG1	P2′ position of the pro-peptide region blocking the catalytic cleft	3.97

stabilizes the loop joining helix H1 and H2 by making two hydrogen bonds to Lys31p NZ and Tyr33p OH (Table 2). The H2 helix contains the conserved ERFNIN motif (Coulombe *et al.*, 1996; Groves *et al.*, 1998; Wiederanders *et al.*, 2003; Kumar *et al.*, 2004) found in cathepsin L-type proteases. The residues of this motif are involved in a series of interactions between conserved residues of the pro-domain (Table 2), thus generating conserved interactions which not only stabilize the pro-domain but also the zymogen structure as a whole. The residues Glu77p from the H3 helix and Glu38p from the H2 helix form salt bridges with Arg42p of helix H2 (Fig. 2*b*) of the pro-domain. Glu38p and Arg42p are two conserved residues of the ERFNIN motif, while Glu77p is also conserved in cathepsin L-type zymogens. Thus, the scaffold of the pro-domain is found to be stabilized by concerted hydrophobic and electrostatic interactions between amino-acid residues (Figs. 2*a* and 2*b*).

3.3. The mature segment

In papain-like cysteine proteases, the mature part is subdivided into two domains referred to as the left domain (L-domain) and the right domain (R-domain) in accordance with the standard view (Fig. 1*a*). The catalytic centre is at the interface of these two domains. The L-domain contains three α-helices and the longest helix contains the active-site cysteine residue, which has been mutated to alanine in this structure. The R-domain is a β-barrel-like structure and the barrel is capped by an α-helix (118m–127m) at one end. The β-strand (144m–149m) at the top of the barrel forms an antiparallel β-sheet with the short β-strand (β1) of the pro-domain (63p–65p). This mature part is structurally similar to mature papain (Pickersgill *et al.*, 1992; PDB entry 1ppn; C^{α} r.m.s.d. of 0.4 Å) in the zymogen pro-pap-RSS-C25mA and the catalytic dyad residues in both the structures are almost superimposable, revealing that the catalytic cleft is fully formed in the zymogen.

The mutations created in the mature segment of papain, K174mR, V32mS and G36mS (pro-pap-RSS), are involved in the formation of interdomain salt bridges and hydrogen bonds (Fig. 1*c*). These mutations were designed by taking structural signatures from the interdomain region of another thermostable papain-like protease of plant origin, ervatamin C (Thakurta *et al.*, 2004; Ghosh *et al.*, 2007; Choudhury *et al.*, 2010). Owing to these mutations, the number of interdomain (L-domain and R-domain) hydrogen bonds has increased compared with the wild-type protease, resulting in higher thermal stability of the protease.

3.4. Interactions between the pro-segment and the mature segment

A general characteristic of C1A proteases is that the mature segment interacts with the pro-segment *via* three regions: (i) the pro-segment binding loop (PBL) of the R-domain, (ii) a short β-strand of the R-domain and (iii) the substrate-binding region (primed and unprimed). In the pro-pap-RSS-C25mA structure a similar binding pattern is observed, in which regions 138m–142m and 144m–149m of the R-domain constitute the PBL and the short β-strand, respectively. These two regions and the primed subsites of the R-domain form two hydrophobic patches which stabilize the hydrophobic side chains of the pro-domain together with the formation of a β-sheet between a β-strand of the pro-segment (63p–65p) and a β-strand from the mature segment (144m–149m). The first hydrophobic patch is formed by Phe149m, Trp181m, Phe141m, Trp177m, Cys200m and Cys153m, which interact with Phe78p and Tyr82p from helix H3 and Phe70p from the preceding region of helix H3 in the pro-domain (Fig. 2*c*). The second hydrophobic patch is formed by Phe149m, Pro151m and Tyr144m, which are involved in hydrophobic packing with Trp63p and Leu65p from the β-strand and Trp52p from the H2 helix of the pro-domain (Fig. 2*d*). Apart from these hydrophobic interactions, the scaffold of the pro-domain is also stabilized by various electrostatic interactions with the mature segment (Fig. 2*b*). Hydrogen bonds are formed between residues from helices H2 and H3 and residues from the connecting peptide region between helices H2 and H3 to residues Ala137m, Gln142m, Lys139m and Asp140m of the mature segment (Fig. 2*b*). These interactions mainly help in positioning the extended part of the pro-peptide, which finally blocks the catalytic cleft. Ser85p OG fills the oxyanion hole by making hydrogen bonds to Ala25m N and Gln19m NE2. The hydrophobic side chain of Ile86p occupies the hydrophobic S2 subsite pocket, which is known to be a specificity-determining subsite (Fig. 3).

3.5. Zymogen activation and kinetic analysis

The progress of activation of pro-pap-RSS to mature pap-RSS was monitored by SDS–PAGE and gelatine gel zymography (Figs. 4*a* and 4*b*). The maturation occurs in a stepwise manner, in which pro-pap-RSS (calculated molecular mass 41 kDa) is converted to the mature form, pap-RSS (expected molecular mass 23.4 kDa) through intermediate truncated products. The major intermediate band observed has a molecular mass of ∼35 kDa (Fig. 4*a*) and complete conversion to the mature form of the enzyme (which appears as a diffuse band near 24 kDa) only occurs after 90 min of incubation.

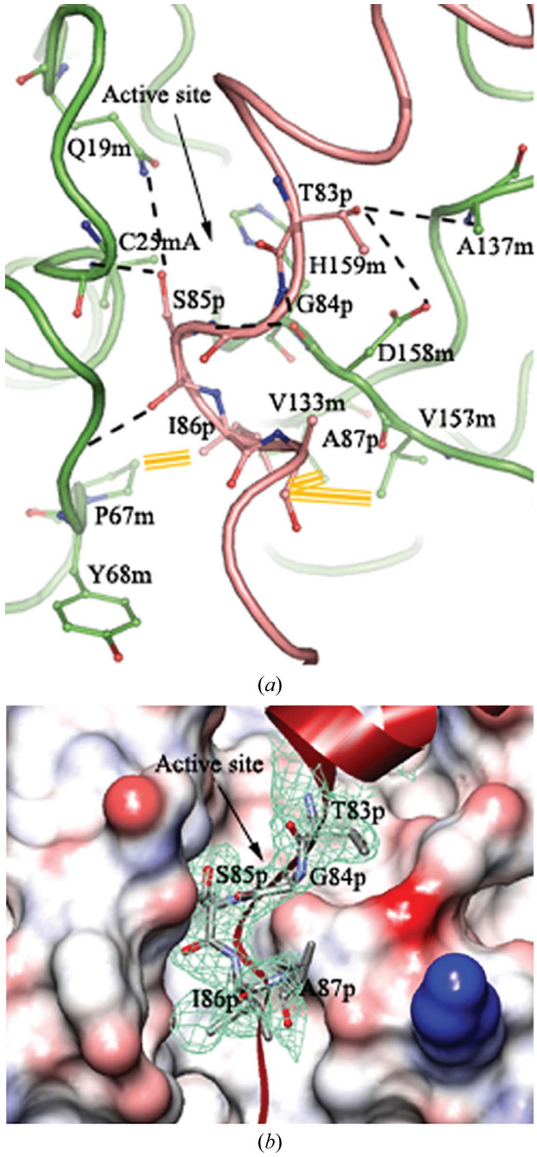

Figure 3
Interdomain interactions. (*a*) Interactions of residues of the pro-segment (Thr83p, Gly84p, Ser85p, Ile86p and Ala87p) occupying subsites S2′, S1′, S1, S2 and S3 of the mature segment of pro-pap-RSS-C25mA. (*b*) $2F_o - F_c$ electron-density map contoured at 1.2σ associated with the same residues of the pro-segment as in (*a*). The mature part is represented by an electrostatic surface and the pro-segment by a red-coloured ribbon. The side chains of the P2′, P1′, P1, P2 and P3 residues of the pro-domain are represented as stick models.

N-terminal sequence analysis of the major truncated intermediate (Fig. 4*a*) gave the sequence GY*X*QN, which reveals that cleavage occurs between the fifth and sixth positions of the pro-peptide at its N-terminus during autocatalytic zymogen activation. Therefore, this proteolytic cleavage removes the entire N-terminal tag of 44 amino acids (including the hexahistidine) together with five amino acids of the pro-peptide. The calculated molecular weight of this truncated intermediate product is 35.5 kDa, which matches that of the product determined by SDS–PAGE analysis (Fig. 4*a*). A cleavage near the N-terminal part of the pro-segment was also observed for pro-cathepsin L (Ménard *et al.*, 1998). N-terminal sequencing of the final product corresponding to the mature protease was not conclusive, possibly owing to heterogeneity of the product resulting from incomplete truncation.

The time course of activation as followed by nonreducing nondenaturing substrate gel zymography (Fig. 4*b*) also demonstrated that the process of conversion to the mature form occurs in a sequential manner. As described in §2, protease samples at different time intervals were quickly chilled at 253 K after activation in the presence of the reversible inhibitor antipain before electrophoresis on a gelatine gel where they were finally immobilized. Thus, it can be presumed that the conformations of the pro-protease, intermediates and mature enzyme at pH 4.0 are retained on gelatine SDS–PAGE and the corresponding activities for all the forms are for the conformations attained at the lower pH of 4.0, even though the activity was monitored at pH 5.5 (Fig. 4*b*). An important observation was that the pro-form of the protein exhibits substantial activity, although a much lower activity was observed for the mature form under the conditions described for gelatine gel zymography.

We were able to capture the intermediates formed during maturation of pro-pap-RSS to mature pap-RSS by incubating the pro-enzyme at the lower temperature of 313 K in the presence of DTT as an activator during gelatine gel zymography. However, for kinetic analysis the incubation was performed at 333 K with cysteine as an activator, where complete conversion of the pro form to the mature form occurred after 40–45 min. The k_{cat}, K_m and k_{cat}/K_m values of the protease against BAPNA and pEFLNA are summarized in Table 3. The results indicate that although the K_m values of pap-RSS (this study) and commercial native papain for BAPNA (Mole & Horton, 1973) are similar, the k_{cat}/K_m value is lower for pap-RSS. This lower value of k_{cat}/K_m probably arises from the fact that the value obtained for pap-RSS is based on the protease concentration measured by titration with E-64, whereas that for commercial papain is calculated based on the protein concentration of papain itself (Mole & Horton, 1973). This lower value may also arise from the presence of pro-peptide parts in the reaction mixture generated during the maturation process of the protease, which are known to have an inhibitory effect on the enzyme (Taylor *et al.*, 1995; Wiederanders *et al.*, 2003). The results also indicate that pEFLNA is a better substrate for pap-RSS than BAPNA. This increase in specificity suggests that longer peptides may provide additional anchors at the catalytic site. The result

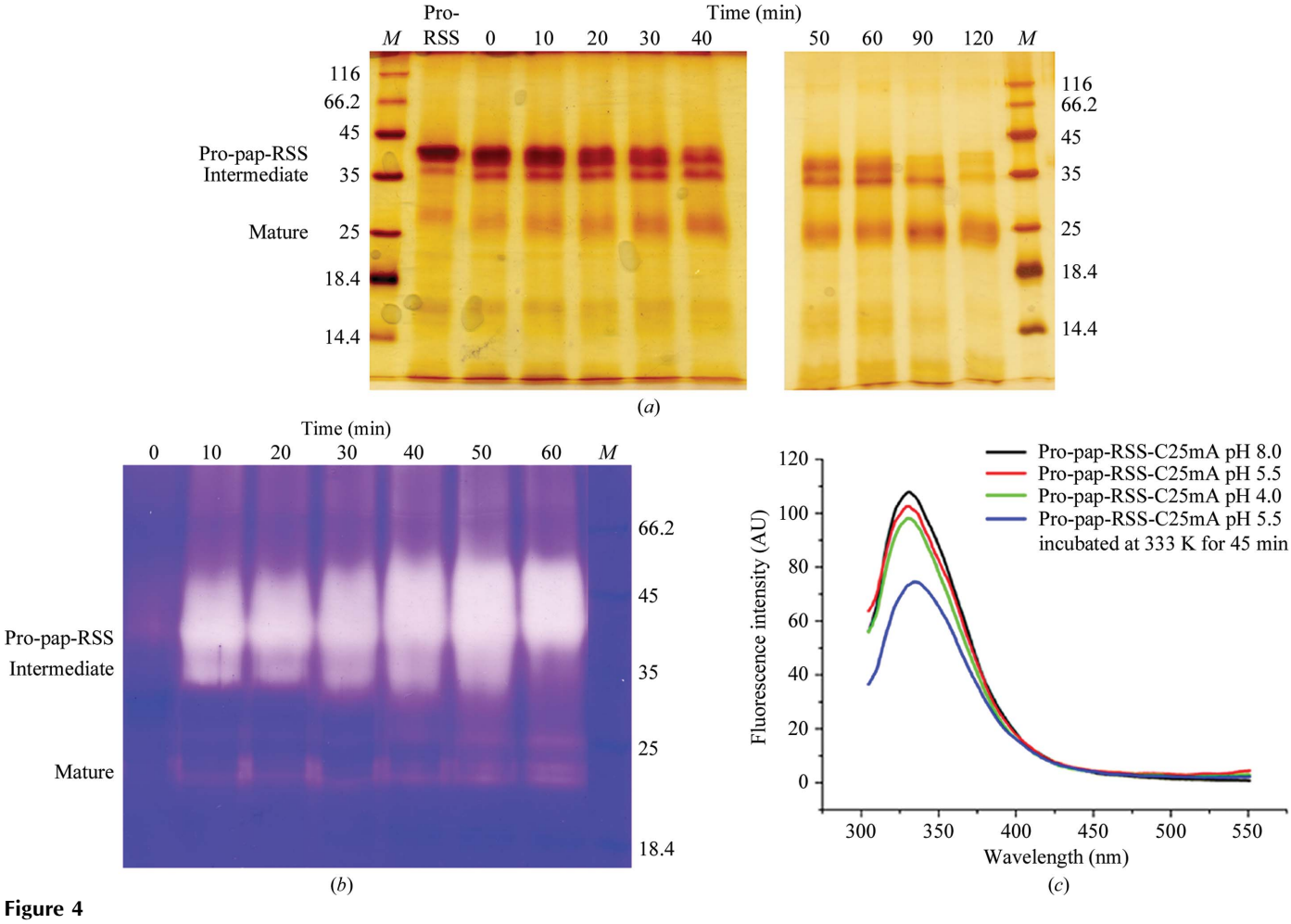

Figure 4
Activation/processing of pro-pap-RSS to mature pap-RSS. (*a*) SDS–PAGE analysis: pro-pap-RSS was activated as described in §2 for 0–2 h and inhibited with 10 µ*M* E-64 after activation to stop the activation process. The samples were run on a 12% polyacrylamide gel and silver-stained after electrophoresis. Lane *M* contains molecular-mass markers (labelled in kDa). (*b*) Substrate gel assay: aliquots of activated pro-pap-RSS removed at 10 min intervals were inhibited with the reversible inhibitor antipain (∼50 µ*M*) and immediately frozen. The samples were processed and run on a nonreducing 12% polyacrylamide gel copolymerized with 0.1% gelatine, stained with Coomassie Brilliant Blue R-250 and proteinase activity was detected by the appearance of clear bands on a dark blue background. (*c*) Fluorescence spectra of pro-pap-RSS-C25mA at different pH values and temperatures.

obtained here is consistent with a previous study showing that papain has higher activity towards carbobenzoxy-phenylalanine-L-arginine-7-amino-4-methylcoumarin hydrochloride (Cbz-FR-MCA) than towards FR-MCA (which lacks an N-terminal Cbz group; Nägler *et al.*, 1999).

3.6. Fluorescence study

At neutral pH, pro-pap-RSS-C25mA exhibits maximum emission at around 330.9 nm (Fig. 4*c* and Supplementary Table S1[2]), which is characteristic of a protein containing buried tryptophan residues. The crystal structure of pro-pap-RSS-C25mA shows that all six of the tryptophan residues are buried, which corroborates the results from fluorescence studies. The intensities in the fluorescence spectrum of pro-pap-RSS-C25mA remained almost unaltered at pH 4.0 and pH 5.5 after 10 min of incubation (Fig. 4*c*). A slight variation in intensity was only observed after 45 min of incubation at 333 K at pH 4.0 together with a 3 nm red shift of the maximum wavelength (333.8 nm), which indicates that no significant change in the microenvironment of the tryptophan residues occurs even after a long exposure at this pH and temperature.

3.7. Normal-mode analyses

We have investigated the collective motions of the crystal structure by NMA (Tama, 2003), in which domain motions were analyzed for the conformations generated in this study (Supplementary Table S2[2]). Two lowest frequency modes (9 and 11) were identified in this study, in which three interacting domains (the first two in the pro-segment and the third in the mature segment) undergo twisting and opening/closure motions. The fixed domain (domain 1) is comprised of the C-terminal part of the pro-segment and the pro-peptide binding loop (PBL) of the mature segment. Domain 2 is the

[2] Supplementary material has been deposited in the IUCr electronic archive (Reference: MV5074). Services for accessing this material are described at the back of the journal.

N-terminal part of the pro-segment and domain 3 is the rest of the mature segment (Fig. 5). Domains 2 and 3 undergo rotational motion with respect to the fixed domain 1 through the hinge-bending residues at the middle of helix H2 of the pro-segment, the residues at the pro-segment blocking the primed subsites (82p–84p) and the residues at the interface of domains 1 and 2 (Fig. 5*a*). In mode 9 of the normal-mode analysis (Fig. 5), domain 2 undergoes a twisting movement of up to 41.6° compared with the fixed domain 1. In contrast, domain 3 moves by up to 17.6° compared with the same domain 1 (Supplementary Table S2), which reveals that the movement of domain 2 is greater than that of domain 3. The NMA studies indicate that the scaffold part of the pro-segment (the pro-domain) is not a tightly packed single domain; rather, it can be divided into two small domains. In contrast, the R-domain and L-domain of the mature part are structurally coupled by nonbonded interactions and can be treated as a single domain.

Table 3
Kinetic parameters of substrate hydrolysis for Pap-RSS and commercial papain.

	Specific activity†	Substrate	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (M^{-1} s^{-1})
Pap-RSS	130.2 ± 1.5‡	pEFLNA	3.252 ± 1.24	0.281 ± 0.029	11344.3 ± 3360.14
		BAPNA	0.413 ± 0.032	3.731 ± 0.279	114.24 ± 6.31
Commercial papain	221.73 ± 0.67‡	pEFLNA	1.021 ± 0.26	0.293 ± 0.028	3500.73 ± 574.69
		BAPNA	0.73 ± 0.03§	2.95 ± 0.16§	247.8 ± 22.3§

† One unit is the amount of soluble protease required to release 1 µg of soluble azopeptides per minute. ‡ Values taken from Choudhury *et al.* (2010). § Values taken from Mole & Horton (1973).

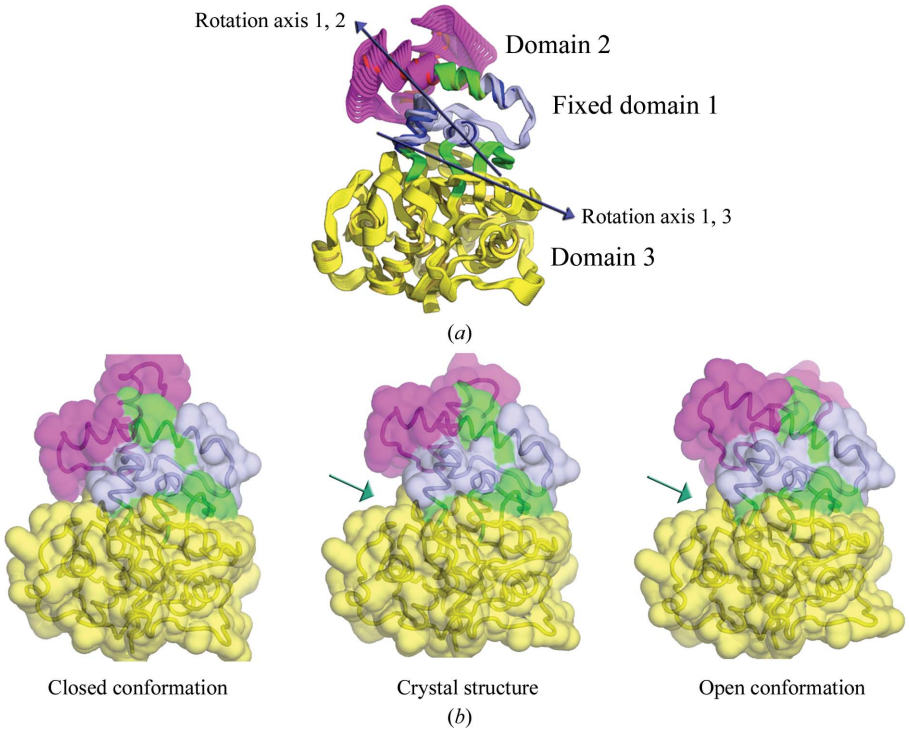

Figure 5
Domain-motion analysis of mode 9 of NMA. (*a*) The two dynamic domains (domains 2 and 3; see Supplementary Table S2) are coloured purple and yellow, while the fixed domain (domain 1) is coloured blue. The hinge-bending regions are coloured green. The rotation axes between domain 1 and domain 2 and between domain 1 and domain 3 are indicated by long arrows. The X-ray structure which is used as input for NMA is merged (shown in darker colours for each domain). (*b*) Surface representation of the three domains in the crystal structure (middle) and in the closed and open conformations. The green arrow indicates the open cleft for an incoming substrate in the crystal structure and in the open conformation.

4. Discussion

Caricain, another papain-like protease from papaya latex, has a high (69%) amino-acid sequence identity to papain and its zymogen structure is known (Groves *et al.*, 1996). Superpositions of the overall structures of the zymogens of caricain and papain as a whole and of their mature parts and pro-domains separately show C^{α} r.m.s.d.s of 0.48, 0.47 and 0.62 Å, respectively. However, relative orientation of the pro-domains as rigid bodies shows a rotation of ∼7° when only the mature parts are superposed (Figs. 6*a* and 6*b*). The extended part of the pro-segment blocking the catalytic cleft is also differently oriented in the two proteases (Figs. 6*c* and 6*d*). These changes in the orientation of the pro-segment as a whole are mainly attributed to a few substitutions, two in the pro-segment itself (Ly79p→Asn and Thr83p→Val) and two in the PBL region (Lys139m→Arg and Asp140m→Pro) of the mature part. The presence of a lysine residue (Lys79p) in helix H3 in the pro-pap-RSS-C25mA structure helps in forming a network of hydrogen bonds in this region which pulls helix H3 towards the R-domain (Fig. 6*c*) and the presence of an aspartic acid Asp140m and a lysine residue Lys139m in the PBL make overall rotation of the pro-domain possible (Fig. 6*c*). In the pro-caricain structure (Groves *et al.*, 1996; PDB entry 1pci), the substitution Lys79p→Asn makes the H3–R-domain interactions weaker compared with those in papain. The presence of an arginine residue in the corresponding position to Lys139m of pro-pap-RSS-C25mA also pushes helix H2 to accommodate the bulkier side chain of arginine; this pushing is additionally facilitated by the absence of the aspartic acid (Asp140m→Pro) and its hydrogen bond to Tyr52 OH from helix H2 is lost (Fig. 6*d*). These three substitutions as a whole may be responsible for the rotation of the scaffold part (pro-domain) of the pro-segment. Another residue in papain, Thr83p, is involved in forming hydrogen-bond interactions with the R-domain. The substitution of this polar threonine residue (Thr83p) by a hydrophobic valine in caricain may be

responsible for the movement of the valine side chain towards a hydrophobic environment. As a result, the backbone of the extended part of this region is also oriented differently. Although the next two residues, Gly84p and Ser85p, that block the catalytic dyad are the same in papain and caricain, their interactions and orientations are totally different in the two proteases. Ser85p OG occupies the oxyanion hole in papain, but in caricain there are no such atoms from the pro-domain in the close vicinity of the oxyanion hole. These comparisons show that pro-segment–mature segment interactions are stronger in papain compared with caricain. Papain, which has a stronger catalytic efficiency (Zucker *et al.*, 1985), probably requires stronger binding interactions between the pro-segment and the mature segment to maintain proteolytic latency in the zymogen condition.

Cysteine proteases are involved in many physiological processes and their spatio-temporal activity is strategically regulated by different means in the cell. Nature has developed various proteinaceous inhibitors to control/regulate the proteolytic activities of proteases. The pro-peptide also acts as an inhibitor in the zymogen form of the enzyme. A comparison of the inhibitory mechanisms of the pro-peptide and proteinaceous inhibitors is important in order to understand the regulation of the enzyme. The inhibitory function of the pro-peptide has been studied previously *in vitro* using isolated pro-peptide of papain and mature papain (Groves *et al.*, 1998),

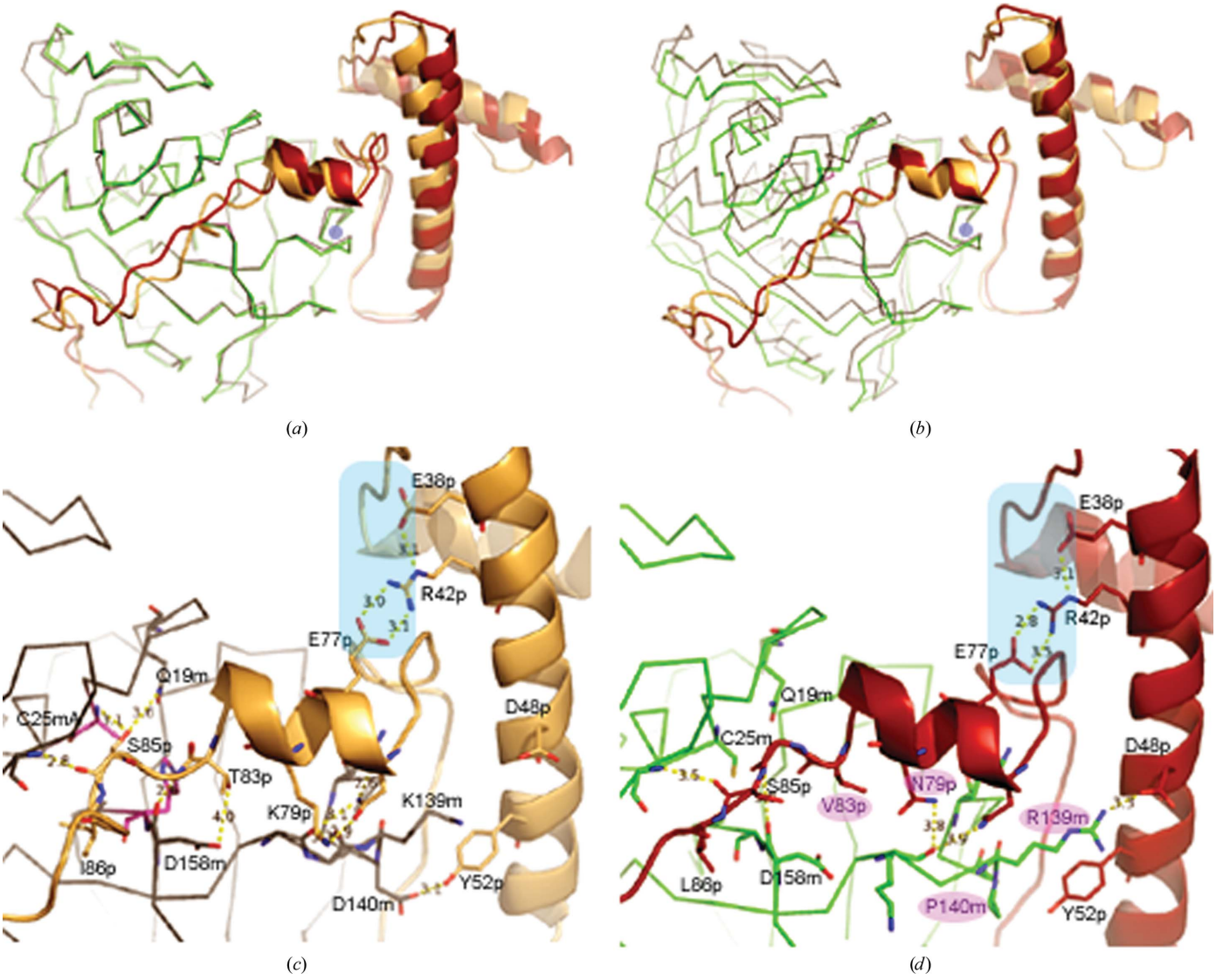

Figure 6
Comparison of pro-pap-RSS with pro-caricain (PDB entry 1pci). (*a*) Superposition of mature parts showing the relative orientation of the pro-domains. The pro-segments of the two proteases are represented as ribbon diagrams, while the mature parts are shown as C^{α} traces. The pro-segments and mature parts of pro-pap-RSS-C25mA are coloured golden yellow and grey and those of pro-caricain are coloured deep orange and green, respectively. (*b*) Superposition of the pro-domains. The pro-domain of pro-caricain rotates anticlockwise by $\sim 7^{\circ}$ to superpose with that of pro-pap-RSS around an axis passing through the blue circle and perpendicular to the plane of the paper. (*c*, *d*) The interactions of the pro-domain and the extended pro-peptide of pro-pap-RSS and pro-caricain with their cognate mature parts. The blue-shaded interaction is conserved in the family and is important for pH-induced zymogen activation. The pink-shaded amino acids represent the replacements which lead to the change in the relative orientation of the pro-domain and the mode of blocking of the active site by the extended pro-peptide in pro-caricain compared with pro-pap-RSS.

which revealed that the pro-peptide is a tight-binding inhibitor with a K_i value in the nanomolar range at neutral pH. The structure of pro-pap-RSS-C25mA shows that the pro-peptide poses steric hindrance to substrate binding at the catalytic cleft. Similar steric hindrance is also observed for the protein inhibitors chagasin and stefin when complexed with papain (Redzynia *et al.*, 2009; Stubbs *et al.*, 1990). The overall folds of chagasin and stefin differ from each other and from the pro-peptide of papain, and their modes of association with the enzyme are also different. Three inhibitory wedges (three β-hairpin loops for chagasin and two β-hairpin loops and one extended trunk for stefin) of the inhibitors hinder the accessibility of the substrate to the catalytic cleft, in contrast to the extended region of the pro-peptide which covers the catalytic cleft throughout, thus inhibiting substrate access (Fig. 7). The three wedges of the inhibitors mainly block the subsites at either side of the catalytic dyad without any direct contact with the catalytic cysteine in order to avoid proteolytic cleavage. However, the pro-peptide directly interacts with the catalytic residues and oxyanion-hole elements and also with the residues of the primed and unprimed subsites in a non-productive orientation opposite to that of a substrate. The scaffold of chagasin and stefin is centrally oriented at the top of the papain molecule with a slight tilt towards the R-domain and the L-domain, respectively (Fig. 7). In contrast, the scaffold of the pro-segment (pro-domain) of papain is inclined more towards the R-domain and covers a larger surface area (1253 Å^2) of papain compared with chagasin and stefin, which have contact surface areas of 419 and 363 Å^2, respectively.

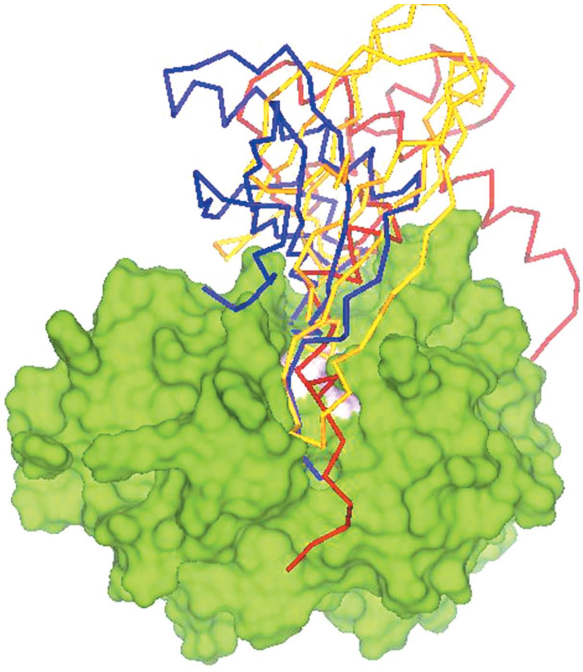

Figure 7
Comparison of inhibition mechanisms: superposition of stefin (PDB entry 1stf), chagasin (PDB entry 3e1z) and propeptide (residues 9p–96p). The corresponding ribbons are coloured blue, yellow and red, respectively. The mature part of papain is represented by a green surface, with the catalytic centre coloured pink.

We have used the X-ray structure of pro-pap-RSS-C25mA to understand the activation mechanism of pro-papain at the molecular level. The crystal used in this study was grown at near-neutral pH (Roy *et al.*, 2011). It has been well established that zymogen activation of papain-like cysteine proteases is triggered at acidic pH (Cygler & Mort, 1997; Wiederanders *et al.*, 2003) with a few exceptions (Hellberg *et al.*, 2002; Quintas-Granados *et al.*, 2009; Dutta *et al.*, 2011). Accordingly, in order to understand the conformational changes in the zymogen structure at lower pH and its influence on the activation mechanism of pro-papain, we also attempted to obtain the structure of the papain mutant from a crystal grown at pH 4.0 or by soaking an existing crystal (grown at higher pH) at pH 4.0. We did not succeed in obtaining an acceptable diffraction-quality crystal for structural analysis. Therefore, intrinsic tryptophan fluorescence studies were carried out on pro-pap-RSS-C25mA in solution. The fluorescence studies indicated that total dissociation of the pro-domain from the mature segment does not occur at lower pH. This differs from the previous postulation that the zymogen is activated only when the pro-domain dissociates from the mature part at the activation pH (Groves *et al.*, 1996). The observations from the fluorescence studies were further supported by NMA, and subsequent domain-motion calculations showed that the relative movement of domains 2 and 3 generates a type of breathing motion, which in its open form creates a larger solvent-accessible surface area (Fig. 5*b*) for a substrate to approach the catalytic cleft. In the structure of pro-pap-RSS-C25mA, we observed some important electrostatic interactions in which four residues, Glu23p, Glu38p, Asp72p and Glu77p, of the pro-domain and two residues, Asp140m and Glu158m, of the R-domain (in the mature segment) are involved (Table 2, Figs. 2*a* and 2*b*) in forming intradomain and interdomain salt bridges and hydrogen bonds, respectively. These interactions help in maintaining the structural stability/integrity of the pro-domain and in anchoring the pro-domain to the mature segment, which in turn helps to maintain the latency of the enzyme at high pH. Since glutamic acid and aspartic acid side chains are protonated at the activation pH (∼4.0), lowering the pH may lead to loss of these salt bridges and the hydrogen bonds involving residues Glu23p, Glu38p, Asp72p, Glu77p, Asp140m and Glu158m. This disruption of salt bridges/hydrogen bonds at low pH will reduce the scaffold integrity (structural loosening) in general and thus may destabilize to a large extent the interface between the pro-domain and the mature part of the zymogen. As a result, the conformational flexibility of the extended part of the pro-peptide blocking the catalytic cleft may increase, which is additionally facilitated by the presence of a flexible glycine residue Gly84p at the P1′ site (Fig. 3*a*). This increased mobility of the extended part possibly permits the access of an N-terminal flanking peptide or an extended loop region of another zymogen molecule with a papain-specific sequence to the catalytic cleft for cleavage, thus initiating the zymogen-activation cascade.

In our studies, we identified a cleavage site (Val5p–Gly6p) during the autocatalytic activation cascade (Fig. 4*a*). The

observed intermediate and the final products (~35 and 24 kDa; Fig. 4*a*) show that the cleavage occurs at the two ends of the pro-segment, removing the pro-segment almost entirely from the mature enzyme. It is to be noted that the cleavage site (Ala71p–Asp72p) previously postulated by Vernet *et al.* (1995) differs from our observation (Val5p–Gly6p), although the important role of Asp72p in regulation of the zymogen-activation process mentioned by Vernet and coworkers tallies with our structural interpretation.

The zymography of the products during the time course of zymogen activation showed that pro-pap-RSS and the subsequent intermediates generated during activation are proteolytically active (Fig. 4*b*). When the relative band intensities of the pro-protease, the intermediates and the mature enzyme on SDS–PAGE (Fig. 4*a*) are compared with the corresponding activity bands on gelatine SDS–PAGE (Fig. 4*b*), we observe that the pro-protease is much more active than the mature enzyme for the trapped conformations during the activation process at low pH. We also note from the first lane (0 min) of the zymogram that pro-pap-RSS does not have activity: this is because, unlike in the other lanes, the pro-enzyme for this lane was not previously incubated at a higher temperature at pH 4.0. It is already known that mature papain has a pH optimum for activity at pH 6.5 (Mole & Horton, 1973; Ménard *et al.*, 1990). Therefore, the lower activity of the mature papain at pH 4.0 may be of help in avoiding autocatalytic degradation of the mature protease during zymogen activation. This pH regulation of the zymogen-activation cascade probably protects the mature enzyme products from autocatalysis.

Our observations and a previous report (McQueney *et al.*, 1997) indicate that proteolytic activity is also associated with zymogens for papain-like cysteine proteases at lower pH. It is now established in the literature that the pro-segment catalyses the folding of the native enzyme in almost all types of proteases and that after folding it is proteolytically removed under certain conditions (Wiederanders *et al.*, 2003). The question of why nature has designed a process to remove the pro-segment if the zymogen itself is sufficiently enzymatically active (at lower pH) in the presence of the pro-segment can be addressed by reference to a recent work on pepsin, which demonstrates that removal of the pro-segment results in a mature enzyme that is more kinetically stable but is not stable in the reversible thermodynamic sense (Dee & Yada, 2010). The kinetic stabilization allows an increased resistance to proteolytic degradation (Jaswal *et al.*, 2002; Truhlar *et al.*, 2004), which is more relevant to an enzyme system with a longer lifetime under different environmental conditions. The generation of the mature enzyme is probably not only required for enhanced kinetic stability but is also important in creating a larger surface area (owing to the removal of the pro-part) near the active site to facilitate the access of protein substrates.

In summary, our successful determination of the three-dimensional structure of the zymogen of a thermostable mutant of pro-papain, together with biophysical and biochemical characterization of the zymogen, has established the structural basis of the initiation of the activity of papain in its zymogen form. This study has also helped in the identification of probable cleavage site(s) in its stepwise activation process.

This project was supported by an institutional (SINP) grant and extramural funding from CSIR, DBT and DST, Government of India. We thank the staff of beamline BM14, ESRF, Grenoble for help during diffraction data collection and Mr B. Pahari of SINP for fluorescence experiments.

References

Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. & Hanada, K. (1982). *Biochem. J.* **201**, 189–198.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
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.
Coulombe, R., Grochulski, P., Sivaraman, J., Ménard, R., Mort, J. S. & Cygler, M. (1996). *EMBO J.* **15**, 5492–5503.
Cygler, M. & Mort, J. S. (1997). *Biochimie*, **79**, 645–652.
Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C. & Mort, J. S. (1996). *Structure*, **4**, 405–416.
Dee, D. R. & Yada, R. Y. (2010). *Biochemistry*, **49**, 365–371.
Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968). *Nature (London)*, **218**, 929–932.
Dutta, S., Choudhury, D., Dattagupta, J. K. & Biswas, S. (2011). *FEBS J.* **278**, 3012–3024.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Ghosh, R., Dattagupta, J. K. & Biswas, S. (2007). *Biochem. Biophys. Res. Commun.* **362**, 965–970.
Groves, M. R., Coulombe, R., Jenkins, J. & Cygler, M. (1998). *Proteins*, **32**, 504–514.
Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W. & Jenkins, J. A. (1996). *Structure*, **4**, 1193–1203.
Hellberg, A., Nowak, N., Leippe, M., Tannich, E. & Bruchhaus, I. (2002). *Protein Expr. Purif.* **24**, 131–137.
Jaswal, S. S., Sohl, J. L., Davis, J. H. & Agard, D. A. (2002). *Nature (London)*, **415**, 343–346.
Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.
Kaulmann, G., Palm, G. J., Schilling, K., Hilgenfeld, R. & Wiederanders, B. (2006). *Protein Sci.* **15**, 2619–2629.
Khan, A. R. & James, M. N. (1998). *Protein Sci.* **7**, 815–836.
Kumar, A., Dasaradhi, P. V., Chauhan, V. S. & Malhotra, P. (2004). *Biochem. Biophys. Res. Commun.* **317**, 38–45.
LaLonde, J. M., Zhao, B., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M. & Smith, W. W. (1999). *Biochemistry*, **38**, 862–869.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Li, J., Vrielink, A., Brick, P. & Blow, D. M. (1993). *Biochemistry*, **32**, 11507–11515.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
McQueney, M. S., Amegadzie, B. Y., D'Alessio, K., Hanning, C. R., McLaughlin, M. M., McNulty, D., Carr, S. A., Ijames, C., Kurdyla, J. & Jones, C. S. (1997). *J. Biol. Chem.* **272**, 13955–13960.
Ménard, R., Carmona, E., Takebe, S., Dufour, E., Plouffe, C., Mason, P. & Mort, J. S. (1998). *J. Biol. Chem.* **273**, 4478–4484.
Ménard, R., Khouri, H. E., Plouffe, C., Dupras, R., Ripoll, D., Vernet, T., Tessier, D. C., Lalberté, F., Thomas, D. Y. & Storer, A. C. (1990). *Biochemistry*, **29**, 6706–6713.

Michaud, D. (1998). *Anal. Chim. Acta*, **372**, 173–185.
Mole, J. E. & Horton, H. R. (1973). *Biochemistry*, **12**, 816–822.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Nägler, D. K., Tam, W., Storer, A. C., Krupa, J. C., Mort, J. S. & Ménard, R. (1999). *Biochemistry*, **38**, 4868–4874.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pickersgill, R. W., Harris, G. W. & Garman, E. (1992). *Acta Cryst.* B**48**, 59–67.
Quintas-Granados, L. I., Orozco, E., Brieba, L. G., Arroyo, R. & Ortega-López, J. (2009). *Protein Expr. Purif.* **63**, 26–32.
Redzynia, I., Ljunggren, A., Bujacz, A., Abrahamson, M., Jaskolski, M. & Bujacz, G. (2009). *FEBS J.* **276**, 793–806.
Roy, S., Choudhury, D., Chakrabarti, C., Biswas, S. & Dattagupta, J. K. (2011). *Acta Cryst.* F**67**, 634–636.
Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). *EMBO J.* **9**, 1939–1947.
Suzuki, K., Tsuji, S. & Ishiura, S. (1981). *FEBS Lett.* **136**, 119–122.
Tama, F. (2003). *Protein Pept. Lett.* **10**, 119–132.
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.
Thakurta, P. G., Biswas, S., Chakrabarti, C., Sundd, M., Jagannadham, M. V. & Dattagupta, J. K. (2004). *Biochemistry*, **43**, 1532–1540.
Truhlar, S. M., Cunningham, E. L. & Agard, D. A. (2004). *Protein Sci.* **13**, 381–390.
Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B. & Turk, D. (2012). *Biochim. Biophys. Acta*, **1824**, 68–88.
Vernet, T., Berti, P. J., de Montigny, C., Musil, R., Tessier, D. C., Ménard, R., Magny, M.-C., Storer, A. C. & Thomas, D. Y. (1995). *J. Biol. Chem.* **270**, 10838–10846.
Vernet, T., Khouri, H. E., Laflamme, P., Tessier, D. C., Musil, R., Gour-Salin, B. J., Storer, A. C. & Thomas, D. Y. (1991). *J. Biol. Chem.* **266**, 21451–21457.
Wiederanders, B., Kaulmann, G. & Schilling, K. (2003). *Curr. Protein Pept. Sci.* **4**, 309–326.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zucker, S., Buttle, D. J., Nicklin, M. J. & Barrett, A. J. (1985). *Biochim. Biophys. Acta*, **828**, 196–204.