protein structure communications

High-resolution complex of papain with remnants of a cysteine protease inhibitor derived from *Trypanosoma brucei*

Magnus S. Alphey and William N. Hunter*

Division of Biological Chemistry and Molecular Microbiology, School of Life Sciences, University of Dundee, Dundee DD1 5EH, Scotland

Correspondence e-mail:
w.n.hunter@dundee.ac.uk

Attempts to cocrystallize the cysteine protease papain derived from the latex of *Carica papaya* with an inhibitor of cysteine proteases (ICP) from *Trypanosoma brucei* were unsuccessful. However, crystals of papain that diffracted to higher resolution, 1.5 Å, than other crystals of this archetypal cysteine protease were obtained, so the analysis was continued. Surprisingly, the substrate-binding cleft was occupied by two short peptide fragments which have been assigned as remnants of ICP. Comparisons reveal that these peptides bind in the active site in a manner similar to that of the human cysteine protease inhibitor stefin B when it is complexed to papain. The assignment of the fragment sequences is consistent with the specificity of the protease.

1. Introduction

The first cysteine protease structure to be determined was that of papain from *Carica papaya*. Since its discovery, many 'papain-like' proteases, also referred to as thiol or sulfhydryl peptidases, have been characterized and are classified as clan CA proteases. The cysteine proteases are grouped into seven clans defined according to the linear organization of catalytic residues in the sequence, *e.g.* clan CA has the catalytic residues Cys, His and Asn or Asp ordered in sequence, clan CD presents two catalytic residues, His and Cys, in sequence, clan CE has a triad formed by His, Glu or Asp and Cys at the C-terminus, clan CF also presents a catalytic triad, but ordered as Glu, Cys and His, clan CG has a dyad of two cysteine residues and clan CH presents a Cys, Thr and His triad with the catalytic cysteine at the N-terminus (Rawlings *et al.*, 2006). Additionally, clan membership depends upon specificity, with clan CA proteases characterized by sensitivity to the inhibitor E64 [L-*trans*-epoxysuccinyl-leucyl-amido-(4-guanidino)-butane] and by having substrate specificity defined by the S2 pocket (Sajid & McKerrow, 2002). The majority of protozoan parasite cysteine proteases belong to clan CA family C1 papain-like proteases. This family of parasite-derived cysteine peptidases are critical to the life cycle or pathogenicity of many parasites, where they contribute key roles in immunoevasion, enzyme activation, pathogenesis, virulence and tissue and cellular invasion as well as excystment, hatching and moulting, and are considered to be promising chemotherapeutic targets (Sajid & McKerrow, 2002; Mottram *et al.*, 2004).

The actions of mammalian cysteine proteases are controlled in part by endogenous tight-binding inhibitors from the cystatin superfamily (Grzonka *et al.*, 2001; Abrahamson *et al.*, 2003). The *Leishmania* genome lacks genes encoding cystatins. However, in *Trypanosoma cruzi* a potent inhibitor of the parasite's own cysteine protease cruzipain was identified and called chagasin (Besteiro *et al.*, 2004). Subsequently, several homologues of these inhibitors of cysteine proteases (ICPs) were identified in the parasitic protozoa *T. brucei*, *L. major* and *L. mexicana* and the bacterium *Pseudomonas aeruginosa* (Sanderson *et al.*, 2003). ICPs inhibit clan CA family C1 cysteine proteases with varying specificities. The molar ratio of inhibition is 1:1 and inhibition is competitive. The ICP of *T. brucei* (*Tb*ICP) appears to be more potent than the *L. mexicana* ICP and displays low nanomolar K_i values against the clan CA family. Whilst ICPs share low sequence homologies and no significant identity with cystatins or other cysteine protease inhibitors, their functional homology implies
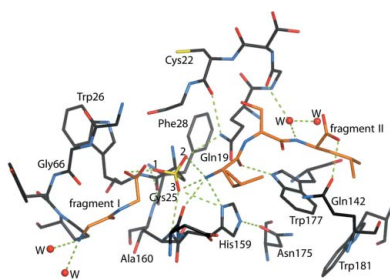
a common evolutionary origin between bacterial and protozoal proteins (Sanderson *et al.*, 2003).

We set out to cocrystallize the *Tb*ICP–papain complex, seeking to generate structural data on an ICP and to understand the mode of inhibition. Here, we report the resulting papain structure with ICP-derived peptide fragments bound within the active-site cleft.

2. Methods and materials

2.1. Sample preparation

The gene encoding *Tb*ICP was previously cloned into plasmid pBP117 (Sanderson *et al.*, 2003), which produces recombinant protein carrying an N-terminal histidine tag. This plasmid was heat-shock transformed into *Escherichia coli* strain BL21(DE3). Cells were grown in Luria–Bertani medium supplemented with ampicillin ($100 \ \mu g \ l^{-1}$) to an optical density of 0.7. The culture was cooled to 288 K, gene expression was induced with $0.2 \ mM$ isopropyl β-D-thiogalactopyranoside and cell growth was continued overnight. Cells were harvested by centrifugation ($2500g$) at 277 K, resuspended in binding buffer ($25 \ mM$ Tris–HCl pH 7.5, $500 \ mM$ NaCl, $5 \ mM$ imidazole) and lysed using a OneShot cell disrupter (Constant Systems). Insoluble debris was separated by centrifugation ($40 \ 000g$) at 277 K for 20 min and the supernatant was filtered through a $0.45 \ \mu m$ syringe filter and then applied onto an Ni^{2+}-resin column (GE Healthcare) pre-equilibrated with binding buffer using a BioCAD 700e (Perseptive Biosystems). The resin was washed with $25 \ mM$ Tris–HCl, $10 \ mM$ imidazole pH 7.5 and the product was eluted with an increasing imidazole gradient. Fractions were analyzed by SDS–PAGE and those containing *Tb*ICP were pooled and dialysed overnight against $25 \ mM$ Tris–HCl pH 7.5 in the presence of 80 units of thrombin (Amersham). The resulting mixture was filtered ($0.45 \ \mu m$) and applied onto a ResourceQ anion-exchange column (Amersham). *Tb*ICP does not bind to this column, whilst thrombin and the cleaved histidine-tag fragment do. Fractions containing *Tb*ICP were pooled, dialyzed overnight against $25 \ mM$ Tris–HCl pH 7.5 at 277 K and then concentrated to $3.4 \ mg \ ml^{-1}$.

2.2. Crystallization and data collection

Purified *Tb*ICP was mixed with papain (Sigma–Aldrich) to final concentrations of $1.4 \ mg \ ml^{-1}$ (*Tb*ICP) and $2 \ mg \ ml^{-1}$ (papain) in $25 \ mM$ Tris–HCl pH 7.5. This mixture was used in hanging-drop crystallization trials with commercially available screens. No crystals or promising conditions were identified over a period of several months and the conditions were set aside at room temperature. Following storage for 2 y, a crystal was observed in conditions that were originally established by combining $1 \ \mu l$ protein mixture with $1 \ \mu l$ of a reservoir consisting of 50% ethanol, $0.01 \ M$ sodium acetate. The crystal was cooled in a stream of nitrogen to 103 K and used for data collection on beamline ID29 of the European Synchrotron Radiation Facility, Grenoble. The orthorhombic crystal diffracted to $1.5 \ \text{\AA}$. A data set comprising 360 images, each of $1°$ oscillation, were collected, processed with *MOSFLM* (Leslie, 1992) and scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994) with details presented in Table 1. At this stage the composition of the crystal was unknown, but since the crystallization conditions resembled those previously reported for papain (Kamphuis *et al.*, 1984) and the unit-cell parameters are similar to those reported for an orthorhombic crystal form of the enzyme, albeit with a 5% difference in unit-cell lengths, we thought it likely that papain itself had been crystallized. The Matthews coefficient calculated for one molecule per asymmetric unit of papain was $2.2 \ \text{\AA}^3 \ Da^{-1}$, with 44% solvent

Table 1
Crystallographic statistics.

Values in parentheses represent the highest resolution bin of approximate width 0.08 Å.

Space group	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 42.32$, $b = 46.12$, $c = 95.70$
Resolution range (Å)	96–1.5
No. of observed/unique reflections	418881/60928
Wilson B ($\text{\AA}^2$)	16.24
Completeness (%)	100 (100)
Multiplicity	13.6 (13.8)
R_{merge} (%)	9.5 (62.9)
Mean $\langle I/\sigma(I) \rangle$	21.4 (3.7)
R factor/R_{free} (%)	17.67 (22.0)/22.54 (29.4)
R.m.s.d. from ideal values, bond lengths (Å)	0.013
R.m.s.d. from ideal values, bond angles (°)	1.416
Average B values ($\text{\AA}^2$)	
Overall	17.9
Main chain	16.9
Side chain	18.9
Waters	34.4
Fragment I	38.1
Fragment II	35.1
Ramachandran analysis	
Residues in most favourable regions (%)	88.4
Residues in additional allowed regions (%)	11.6
DPI† (Å)	0.25

† Diffraction precision indicator (Cruickshank, 1999).

content. However, since the diffraction data extended to slightly higher resolution than the best resolved data available for this protease (structures in the PDB fall in the range 2.8–1.6 Å resolution), we continued with the analysis.

2.3. Structure determination and model refinement

Molecular replacement (*MOLREP*; Vagin & Teplyakov, 2000) using the papain model with PDB code 9pap (Kamphuis *et al.*, 1984) produced a solution with an R factor of 38% and a correlation coefficient of 0.64. Rigid-body refinement (*REFMAC*5; Murshudov *et al.*, 1997) and further restrained refinement interspersed with model building, adjustment and water placement using *COOT* (Emsley & Cowtan, 2004) resulted in a complete model with an R factor of 17.6% and an R_{free} of 22.5%. The R_{merge} for data in the highest resolution range exceeded 60%. In general, we would not normally use such data but, given that the $\langle I/\sigma(I) \rangle$ value was nearly 4 for this resolution bin and with high redundancy approaching 14, we were content to include these diffraction terms and trust the benefits of maximum-likelihood weighting (Murshudov *et al.*, 1997). The approach appears to have been successful given that the statistics (R factor = 22.0%, R_{free} = 29.4%) for the highest resolution data are acceptable.

The completed model comprises residues 1–212 and 161 waters. Eight residues (76–79 and 193–196) are relatively poorly defined in the electron-density maps and 17 (3, 9, 13, 21, 34, 70, 73, 74, 84, 91, 98, 99, 133, 145, 155, 173 and 197) are modelled in dual conformations. Residues 35, 118 and 135 are all assigned as glutamine in the starting model (PDB code 9pap), but on the basis of hydrogen-bonding considerations our model contains glutamic acid at these positions, a point discussed below. In addition to acetate (included in the crystallization conditions), glycerol (likely to have been acquired from the dialysis tubing) and the three O atoms bound to the active-site cysteine, which is in the form of sulfonic acid, two short peptide fragments have been modelled into the active-site cleft. It is likely that these are remnants of the *Tb*ICP that was mixed with papain prior to crystallization. The geometry of this high-resolution model was acceptable, with all residues in the most favourable or allowed regions of the Ramachandran plot (Table 1).

protein structure communications

3. Results

3.1. Overall structure

The structure of papain has been well characterized (Drenth *et al.*, 1976; Kamphuis *et al.*, 1984; Pickersgill *et al.*, 1992; Tsuge *et al.*, 1999). The protein is assembled from two domains, each comprising residues from both the N- and C-terminal sections of the polypeptide. One domain consists of a six-stranded antiparallel β-sheet and the other domain consists mainly of three α-helices. The elongated active-site cleft is formed between them and is lined by residues from both domains. The active-site Cys25 is positioned at the N-terminus of α1 and is likely to be influenced by the helix dipole. As noted from previous structural studies on papain (Kamphuis *et al.*, 1984), this cysteine has been oxidized to sulfonic acid, probably owing to the highly reactive nature of the thiol group in the active enzyme.

Our model is essentially identical to published structures of papain with r.m.s.d. values determined by overlaying C^{α} positions of 0.41 Å (PDB code 1stf; Stubbs *et al.*, 1990), 0.32 Å (9pap; Kamphuis *et al.*, 1984), 0.38 Å (1bp4; LaLonde *et al.*, 1998), 0.46 Å (1bqi; LaLonde *et al.*, 1998), 0.30 Å (1cvz; Tsuge *et al.*, 1999), 0.32 Å (1khq; Janowski *et al.*, 2004), 0.35 Å (1pip; Yamamoto *et al.*, 1992) and 0.32 Å (1pe6; Yamamoto *et al.*, 1991). There are minor differences owing to the flexibility of surface residues Arg41, Gln73, Arg98, Glu99, Arg111, Gln114, Arg145 and Lys156. The N-terminus (Glu3) and C-terminus (Asn212) also exhibit some flexibility and the electron density in these regions is not as well defined as for the rest of the molecule.

In all but three of the deposited papain structures (1khp, 1khq and 1ppn), residues 35, 118 and 135 are assigned as glutamine. Using the hydrogen-bonding networks as a guide, we assign these residues as glutamic acid and as an example show Glu118 in Fig. 1. Glu118 OE1 accepts hydrogen bonds donated from the backbone amide of Gly192 and the hydroxyl of Tyr203, whilst Glu118 OE2 accepts a hydrogen bond from Arg191 NH1. The carboxylate side chain of Glu135 participates in a three-centre hydrogen bond with the amide of Gly54. The distances between the OE1 and OE2 atoms and Gly54 N are 3.07 and 3.09 Å, respectively. Glu35 OE2 accepts a hydrogen bond donated from the amide of Tyr48 and a water molecule; OE1 interacts with two water molecules and the side-chain hydroxyl of Thr14. This hydroxyl group accepts hydrogen bonds from NZ of Lys17 and Lys174, thus defining that it must donate a hydrogen bond to Glu35 OE1.

In early amino-acid sequences of papain, residues 118 and 135 were initially assigned as glutamic acids, but on the basis of a re-evaluation of the sequence were changed to glutamine (Mitchel *et al.*, 1970). It is possible that there is variation in papain sequences depending upon the exact source of the enzyme. We note that only small structural perturbations would occur if the hydrogen-bonding patterns were to be altered by incorporation of glutamines at these positions in the sequence.
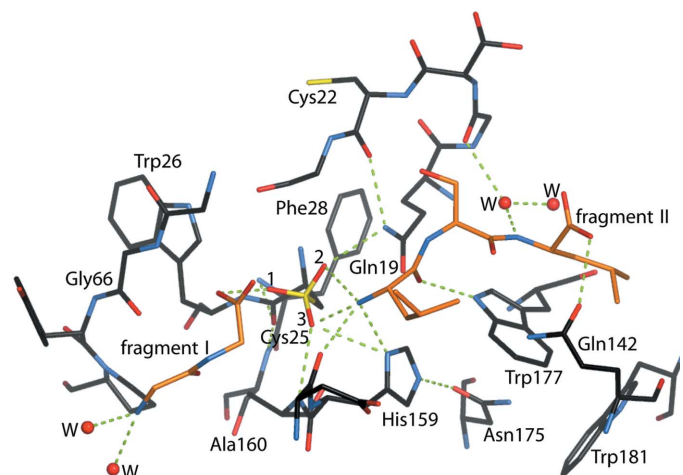
3.2. The active site and peptide fragments

We were unable to crystallize a papain–*Tb*ICP complex and conclude that during storage digestion of *Tb*ICP has occurred and the protease has crystallized with two peptide fragments bound in the active site (Fig. 2). Papain is a relatively promiscuous protease releasing an array of peptide fragments and it is possible that a
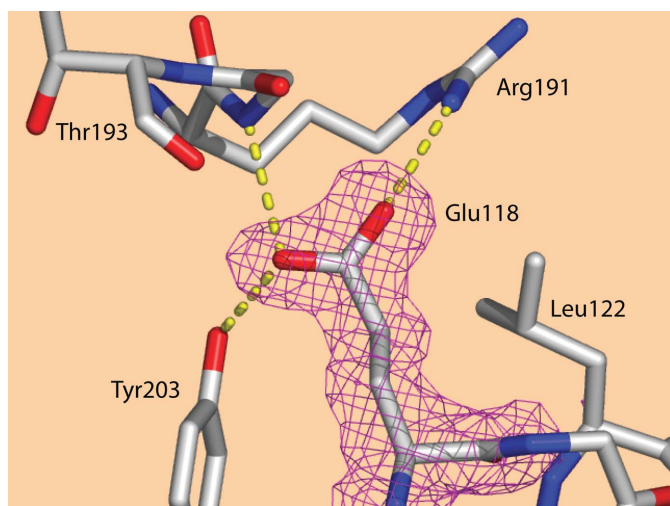
Figure 1
An omit difference electron-density map for Glu118 calculated with coefficients $(F_{o} - F_{c})$ and contoured at 3σ (magenta) revealing the hydrogen-bonding (yellow dashed lines) pattern that defines the side-chain properties. F_{o} and F_{c} represent the observed and calculated structure factors, respectively. The refined model is shown in sticks with O atoms in red, N atoms in blue and C atoms in white. All figures were prepared using *PyMOL* (DeLano, 2002).
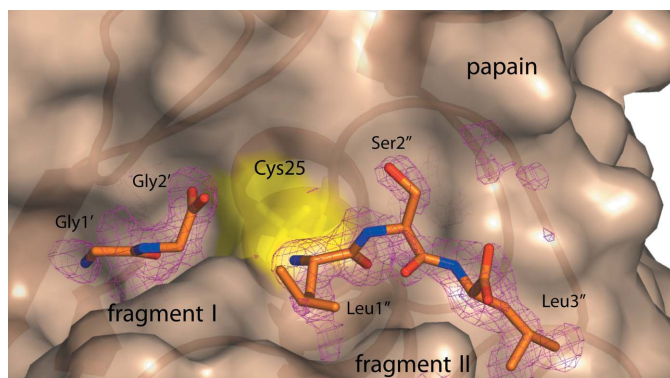
Figure 2
Molecular-surface representation of the papain active site showing the position of the catalytic Cys25 (yellow) and the two peptide fragments (sticks coloured C orange, O red, N blue) with associated omit difference electron-density $(F_{o} - F_{c})$ map contoured at 1.5σ (magenta).

Figure 3
Selected active-site details. Putative hydrogen-bonding interactions (green dashed lines) between papain (sticks coloured C black, O red, N blue, S yellow) and the peptide fragments derived from *Tb*ICP (sticks coloured C orange, O red, N blue) are depicted. The OD1, OD2 and OD3 atoms associated with the sulfonic acid group of Cys25 are labelled 1, 2 and 3, respectively; water molecules are shown as red spheres and labelled W.

mixture of such fragments occupy the active site. However, careful inspection of electron-density and difference-density maps, taking into consideration the amino-acid sequence of *Tb*ICP, has allowed us to model fragment I as the dipeptide Gly-Gly (corresponding to residues Gly78-Gly79 of *Tb*ICP). Fragment II has been modelled as a tripeptide Leu-Ser-Leu which corresponds to Leu95-Ser96-Leu97 of *Tb*ICP. The dipeptide occupies the S subsite and the tripeptide is placed in the S′ subsite of papain. The active-site Cys25 is modified by covalent attachment of three O atoms, as mentioned previously, and the position of each allows a number of activating and stabilizing interactions with surrounding residues and also the two short peptide fragments bound in the active-site cleft. Selected interactions are depicted in Fig. 3.

Gly1′ (′ and ″ denote fragments I and II, respectively) is positioned in a hydrophobic region of the active-site cleft surrounded by Trp69, Val133 and Phe207, with Pro68 at the base of the cleft (not shown). The Gly1′ amide forms two hydrogen bonds with water molecules (Fig. 3). Gly2′ is placed near Ala160 and its carbonyl O is within hydrogen-bonding distance of the main-chain amide of Gly66 and Cys25 OD1. The latter association suggests that Cys25 OD1 represents the hydroxyl group of the sulfonic acid, an assignment consistent with the other interactions observed with the modified Cys25. The Cys25 OD2 group accepts hydrogen bonds from Gln19 NE2 and the amino-terminus of fragment II, the Leu1″ amide, while Cys25 OD3 interacts with His159 ND1 and the Ala160 amide. His159, part of the protease catalytic triad, is held in position by Asn175 and is 5.4 Å distant from the side chain of Asp158, traditionally considered to be the third member of the triad (not shown). Both Asn175 and His159 have low *B* factors, 14 and 11 Å^2, respectively, whilst the *B* factor for Asp158 is around 20 Å^2. There has been discussion in the literature on whether the catalytic triad for papain is Cys25–His159–Asp158 or alternatively Cys25–His159–Asn175 (Wang *et al.*, 1994). However, it has been shown that Asn175 is not essential for enzyme activity and is more likely to be involved in enzyme

stability and orientation of the catalytic His159 (Vernet *et al.*, 1995). In addition to its interactions with His159, Cys25 is held in position through interactions of its carbonyl group with the backbone amides of Phe28 and Ser29.

The amino end of fragment II is held in place by hydrogen bonds donated to Cys25 OD3 and the carbonyl group of Asp158. The Leu1″ carbonyl group accepts a hydrogen bond donated from Trp177 NE1, whilst the side chain nestles comfortably in a pocket created mainly by the side chains Ala137, Gln142, Asp158 and Trp177. Ser2″ is solvent accessible and does not make any direct hydrogen bonds to the protein. The Leu3″ side chain binds in a hydrophobic patch created by Trp177 and Trp181, whilst the amide group interacts with a water molecule. The fragment II carboxylate group interacts with Gln142 OE1, suggesting that it is protonated. Gln142 NE2 donates a hydrogen bond to the carbonyl group of Ala136.

Comparison of the structure reported here with the complex formed between papain and the protease inhibitor human cystatin stefin B (Stubbs *et al.*, 1990) was carried out by overlaying papain. This indicates that the positions of the bound peptide fragments closely resemble the positions of the cleft-binding N-terminus and first loop of stefin B (Fig. 4). The direction of the stefin B polypeptide is consistent with that observed for fragments I and II.
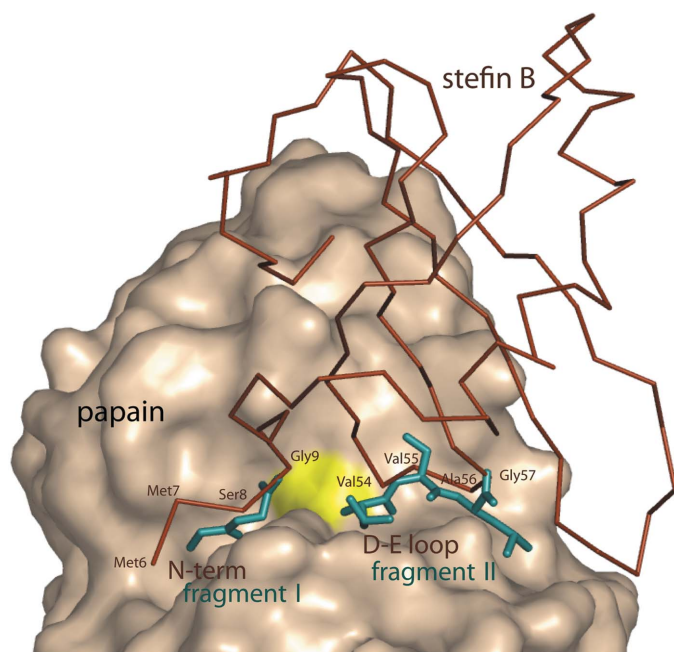
We have assigned the fragments described in this study to products of *Tb*ICP digestion with sequences defined solely on the basis of interpreting the electron density and on successful refinement. For fragment I, two glycine residues corresponding to Gly78-Gly79 were assigned. This agrees with a theoretical model of *L. mexicana* ICP (*Lm*ICP) bound to papain which places the inhibitor *BC* loop in the S subsite (Smith *et al.*, 2006) and suggests that this part of the papain active-site cleft can accept small side chains. Alignment of the two ICP sequences places a Gly76-Ala77-Gly78-Gly79 motif of *Tb*ICP alongside the *BC* loop of *Lm*ICP (data not shown).

It is noteworthy that in the description of papain activity provided by the commercial supplier of the enzyme, Sigma–Aldrich, the enzyme is defined as having activity towards the peptide bonds of basic residues, leucine or glycine. Our observation and assignment of the peptide fragments bound in the active site is consistent with such a definition.

Figure 4
Surface representation of papain with peptide fragments (cyan) *in situ*. Overlaid in brown is a C$^\alpha$ trace of the cysteine protease inhibitor stefin B, showing that the positions of the peptide fragments correlate well with the binding loops of the inhibitor.

References

Abrahamson, M., Alvarez-Fernandez, M. & Nathanson, C. M. (2003). *Biochem. Soc. Symp.* **70**, 179–199.

Besteiro, S., Coombs, G. H. & Mottram, J. C. (2004). *Mol. Microbiol.* **54**, 1224–1236.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cruickshank, D. W. J. (1999). *Acta Cryst.* D**55**, 583–601.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.

Drenth, J., Kalk, K. H. & Swen, H. M. (1976). *Biochemistry*, **15**, 3731–3738.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Grzonka, Z., Jankowska, E., Kasprzykowski, F., Kasprzykowska, R., Lankiewicz, L., Wiczk, W., Wieczerzak, E., Ciarkowski, J., Drabik, P., Janowski, R., Kozak, M., Jaskolski, M. & Grubb, A. (2001). *Acta Biochim. Pol.* **48**, 1–20.

Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z. & Jaskolski, M. (2004). *J. Pept. Res.* **64**, 141–150.

protein structure communications

Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.

LaLonde, J. M., Zhao, B., Smith, W. W., Janson, C. A., DesJarlais, R. L., Tomaszek, T. A., Carr, T. J., Thompson, S. K., Oh, H. J., Yamashita, D. S., Veber, D. F. & Abdel-Meguid, S. S. (1998). *J. Med. Chem.* **41**, 4567–4576.

Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.

Mitchel, R. E. J., Chaiken, I. M. & Smith, E. L. (1970). *J. Biol. Chem.* **245**, 3485–3492.

Mottram, J. C., Coombs, G. H. & Alexander, J. (2004). *Curr. Opin. Microbiol.* **7**, 375–381.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Pickersgill, R. W., Harris, G. W. & Garman, E. (1992). *Acta Cryst.* B**48**, 59–67.

Rawlings, N. D., Morton, F. R. & Barrett, A. J. (2006). *Nucleic Acids Res.* **34**, D270–D272.

Sajid, M. & McKerrow, J. H. (2002). *Mol. Biochem. Parasitol.* **120**, 1–21.

Sanderson, S. J., Westrop, G. D., Scharfstein, J., Mottram, J. C. & Coombs, G. H. (2003). *FEBS Lett.* **542**, 12–16.

Smith, B. O., Picken, N. C., Westrop, G. D., Bromek, K., Mottram, J. C. & Coombs, G. H. (2006). *J. Biol. Chem.* **281**, 5821–5828.

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). *EMBO J.* **9**, 1939–1947.

Tsuge, H., Nishimura, T., Tada, Y., Asao, T., Turk, D., Turk, V. & Katunuma, N. (1999). *Biochem. Biophys. Res. Commun.* **266**, 411–416.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Vernet, T., Tessier, D. C., Chatellier, J., Plouffe, C., Lee, T. S., Thomas, D. Y., Storer, A. C. & Menard, R. (1995). *J. Biol. Chem.* **270**, 16645–16652.

Wang, J., Xiang, Y. F. & Lim, C. (1994). *Protein Eng.* **7**, 75–82.

Yamamoto, A., Tomoo, K., Doi, M., Ohishi, H., Inoue, M., Ishida, T., Yamamoto, D., Tsuboi, S., Okamoto, H. & Okada, Y. (1992). *Biochemistry*, **31**, 11305–11309.

Yamamoto, D., Matsumoto, K., Ohishi, H., Ishida, T., Inoue, M., Kitamura, K. & Mizuno, H. (1991). *J. Biol. Chem.* **266**, 14771–14777.