The structural origins of the unusual specificities observed in the isolation of chymopapain M and actinidin by covalent chromatography and the lack of inhibition of chymopapain M by cystatin

Mark P. THOMAS,*† Chandra VERMA,† Susan M. BOYD† and Keith BROCKLEHURST*§

*Laboratory of Structural and Mechanistic Enzymology, Department of Biochemistry, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, U.K., and †Protein Structure Research Group, Department of Chemistry, University of York, Heslington, York, YO1 5DD, U.K.

1. The selectivity observed when the potentially general technique for the isolation of fully active forms of cysteine proteinases, covalent chromatography by thiol–disulphide interchange, is applied to chymopapain M and to actinidin was investigated by a combination of experimentation and computer modelling. Neither of these enzymes is able to react with the original Sepharose–GSH-2-dipyridyl disulphide gel, but fully active forms of both enzymes are obtained by using Sepharose-2-hydroxypropyl-2'-dipyridyl disulphide gel, which is both electrically neutral and sterically less demanding than the GSH gel. Electrostatic potential calculations, minimization and molecular-dynamics simulations provide explanations for the unusual, but different, specificities exhibited by actinidin and chymopapain M in the interactions of their active centres with ligands. 2. The unique behaviour of chymopapain M in exerting an almost

absolute specificity for substrates with glycine at the P₁ position and in resisting inhibition by cystatin was examined by the computer-modelling techniques. A new, modelled, structure of the complete chicken egg-white cystatin molecule based on the crystal structure of a short form of cystatin was deduced as a necessary prerequisite. The results suggest that electrostatic repulsion prevents reaction of actinidin with the GSH gel, whereas a steric 'cap' resulting from a unique arginine-65–glutamic acid-23 interaction in chymopapain M prevents reaction of the gel with this enzyme and accounts for the lack of its inhibition by cystatin and its specificity in catalysis. 3. Use of chymopapain M as a structural variant of papain demonstrates the validity of the predictions of Lowe and Yuthavong [Biochem. J. (1971) 124, 107–115] relating to the structural requirements and binding characteristics of the S₁ subsite of papain.

INTRODUCTION

The natural variation in both structure and chemical behaviour that exists in the cysteine proteinase family of enzymes contributes valuable opportunities for the investigation of structure–function relationships pertinent to active-centre chemistry and catalytic mechanism (see [1,2] for reviews and [3–7] for recent structure-related kinetic studies and for references to recent crystallographic, spectroscopic and molecular biological studies). The latex of *Carica papaya* (papaya, pawpaw) is a particularly rich source of cysteine proteinase variants (see [2,8] for reviews) which include two enzymes for which high-resolution crystal structures are available {papain (EC 3.4.22.2) at 0.16 nm (1.6 Å) resolution [9] and papaya proteinase Ω (PPΩ) (EC 3.4.22.30) at 0.18 nm (1.8 Å) resolution [10], the latter being closely similar to the model built previously [3] by using comparative modelling techniques}. We recently added to the cysteine proteinase structural database by building a model of the three-dimensional structure of chymopapain M, the late-eluted monothiol component of the mixture of chymopapains that is eluted in ion-exchange chromatography between papain, the least basic component of papaya latex, and PPΩ, the most basic component. This modelled structure [11] was built from the amino acid sequence [12] and the three currently known high-resolution crystal structures of the homologous enzymes, papain, PPΩ and actinidin {EC 3.4.22.14; 0.17 nm (1.7 Å) [13]} by using the COMPOSER suite of programs [14–16] and refinement by interactive graphics and energy minimization. The existence of

the model now permits interpretation in structural terms of the properties of chymopapain M, some of which are markedly different from those of other cysteine proteinases (see [11] and references therein). Chymopapain M is probably a component of the cysteine-proteinase fraction of papaya latex discovered by Schack [17], named 'papaya peptidase B' by Lynn [18] and partially characterized by Polgár [19]. It is the enzyme with unusual specificity characteristics (papaya proteinase IV) that Buttle et al. [20] claimed (incorrectly) to be a previously undetected cysteine proteinase eluted from a cation-exchange column near to the early-eluted chymopapains. In fact, when due care is taken to prevent proteolysis, this enzyme is eluted as the last component of the chymopapain band, as is the papaya peptidase B fraction [11]. It was marketed by Novabiochem U.K. initially as Endopeptidase Gly-C and currently as Proteinase Gly-C and has found application in the preparation of protein fragments and in peptide mapping and sequencing (see, e.g., [21]). These applications derive from its almost absolute specificity for substrates with a glycine residue at position P₁ [22].

One of the unusual characteristics of chymopapain M is the lack of reactivity of its catalytic-site thiol group towards the electrophilic sites of the Sepharose–GSH-2-pyridyl disulphide (-S-S-2-Py) gel, which prevents its purification in fully active form by covalent chromatography [11] using the original protocol for this technique [23,24]. Covalent chromatography by thiol–disulphide interchange using the GSH gel was introduced in 1973 for the preparation of fully active papain [23] and has been used subsequently for the isolation of numerous thiol-containing

Abbreviations used: 2-Py-S-S (or -S-S-2-Py), 2-pyridyl disulphide; 2-Py-S-S-2-Py, 2,2'-dipyridyl disulphide; Bz-L-Arg-4NA, N^ε-benzoyl-L-arginine 4-nitroanilide; Boc-Gly-4NPE, t-butoxycarbonylglycine 4-nitrophenyl ester; PPΩ, papaya proteinase Ω; a.b.n.r., adopted basis newton raphson; m.d., molecular dynamics; e.u., electron unit.

† Present address: Department of Biochemistry, University of Bristol, Bristol BS8 1TD, U.K.

§ To whom correspondence and requests for reprints should be addressed.

proteins including cysteine proteinases (see [25,26] for reviews). The method has been considered to be essentially general for thiol-containing molecules when used at pH ~ 8 , with selectivity for molecules with low pK_a thiol groups (notably cysteine proteinases) when the mercaptan is mixed with the gel at pH ~ 4 . The atypical failure of the technique to permit preparation of fully active chymopapain M was suggested to be due to steric restrictions in its S_1 subsite [11]. The present paper provides evidence in support of this suggestion by reporting the successful isolation of fully active chymopapain M by covalent chromatography using the less sterically-demanding Sepharose-2-hydroxypropyl-S-S-2'-Py gel and computer modelling to demonstrate a steric rather than an electrostatic interpretation of the failure of chymopapain M to react with the GSH gel.

Until the observation that chymopapain M fails to react with the GSH gel [11] the only other protein that, as far as we are aware, had exhibited this behaviour was actinidin. In view of the unusually low pI value (3.1) of actinidin [27], we had supposed its failure to react with the GSH gel to be due to electrostatic repulsion and had successfully isolated fully active actinidin [28] by using the same (electrically neutral as well as less sterically demanding) hydroxypropyl gel that we here report permits the isolation of chymopapain M. We now provide evidence from computer modelling that whereas the failure of chymopapain M to bond to the GSH gel is due to a steric clash in the S_1 -subsite region, the failure of actinidin to bond to this gel is due to the large negative potential that envelops actinidin almost completely and extends out to ~ 0.8 nm (8 Å) from the mouth of the active-centre cleft. In addition, we have used the modelled structure of chymopapain M and a new computer-modelled structure of the complete chicken egg-white cystatin molecule based on the crystal structure of a short form of cystatin [29] to demonstrate the structural basis for the lack of inhibition of chymopapain M by cystatin and to show how use of chymopapain M as a structural variant of papain permits evaluation of the predictions of Lowe and Yuthavong [30] relating to the structural requirements and binding characteristics of the S_1 subsite of papain.

MATERIALS AND METHODS

Isolation of the chymopapain M-rich fraction by cation-exchange chromatography

Spray-dried papaya latex (Powell and Scholefield Ltd., Liverpool, U.K.) was used as the source of chymopapain M [11]. A solution of the latex (15 g) in 100 ml of 0.1 M sodium acetate buffer, pH 5.0, containing 1 mM EDTA and saturated with 2,2'-dipyridyl disulphide (2-Py-S-S-2-Py) was dialysed overnight against 3×5 litres of the acetate buffer devoid of 2-Py-S-S-2-Py. The protein solution was clarified by centrifugation at 10000 g and 4 °C for 15 min and was then applied to a column (30 cm $\times$ 2.5 cm) of S-Sepharose Fast Flow. Ion-exchange chromatography was carried out using a linear gradient with a start buffer of 0.1 M sodium acetate buffer, pH 5.0, containing 1 mM EDTA and finish buffer comprising the start buffer containing, in addition, 1 M NaCl. Up to 300 fractions, each of 12 ml, were collected and analysed by measurement of A_{280} and catalytic activity towards both *N* α -benzoyl-L-arginine 4-nitroanilide (Bz-L-Arg-4NA) and t-butoxycarbonylglycine 4-nitrophenyl ester (Boc-Gly-4NPE) in the presence of 20 mM L-cysteine. The fractions rich in chymopapain M-S-S-2-Py, which are eluted after chymopapain B and papaya lysozyme [11], were identified by their characteristic discrimination between the P_1 -arginine and P_1 -glycine substrates in favour of the latter, after activation by cysteine in the assay mixtures.

Assays

Chymopapain M protein was quantified by using A_{280} ($E_{280}^{1\% 1\text{cm}} = 38\,500\text{ M}^{-1}\cdot\text{cm}^{-1}$ [20]) and total thiol content and intact catalytic site content by spectroscopic titration at 343 nm ($\Delta E = 8.08 \times 10^3\text{ M}^{-1}\cdot\text{cm}^{-1}$) following reaction with 2-Py-S-S-2-Py at pH 4 and 8 respectively (see [1,2]). Catalytic activity towards Bz-L-Arg-4NA (potassium phosphate buffer, pH 6.0, 10.1 M, containing 1 mM EDTA, 250 μM substrate and 13 mM-L-cysteine at 25 °C) was evaluated by spectral analysis at 410 nm [31]. Catalytic activity towards Boc-Gly-4NPE (sodium acetate buffer, pH 4.5, 10.1 M containing 1 mM EDTA and 13 mM-L-cysteine) was evaluated by spectral analysis at 400 nm. It is convenient to assay fractions for activity towards Boc-Gly-4NPE at 400 nm even at pH 4.5 to obviate problems of alkaline hydrolysis and to avoid excessive dilution prior to assay [11]. Reaction was started by addition of 0.1 ml of 20 mM substrate in acetonitrile. Assessment of potential inhibition by chicken cystatin was performed by using active chymopapain M (0.1 μM) preincubated with chicken cystatin (1.0 μM) for 20 min at 25 °C and Boc-Gly-4NPE (1.0 mM) as substrate.

Isolation of fully active chymopapain M by covalent chromatography

Fractions rich in chymopapain M-S-S-2-Py were pooled and adjusted to 85% saturation with respect to $(\text{NH}_4)_2\text{SO}_4$. The mixture was stirred for 30 min at room temperature and then centrifuged for 15 min (10000 g, 4 °C). The pellet was dissolved in 100 ml of 0.1 M Tris/HCl buffer, pH 8.0, containing 40 mM L-cysteine and 1 mM EDTA to convert the enzyme-2-pyridyl mixed disulphide to active enzyme. The solution was acidified to pH ~ 5 by the addition of 1 M HCl, and the enzyme was then precipitated with 85%-satd. $(\text{NH}_4)_2\text{SO}_4$ and isolated by centrifugation as described above. The pellet was dissolved in 300 ml of 0.1 M sodium acetate buffer, pH 4.0, containing 0.3 M NaCl and 1 mM EDTA and the solution was mixed with Sepharose-2-hydroxypropyl-S-S-2'-Py gel (Pharmacia GB; 200 g wet weight) which had been equilibrated previously in the acetate buffer. The mixture was stirred at room temperature for 30 min, isolated by suction on a sintered-glass funnel and washed with 200 ml of 0.1 M Tris/HCl buffer, pH 8.0, containing 0.3 M NaCl and 1 mM EDTA to re-equilibrate the gel at the higher pH in preparation for release of chymopapain M. This was effected by washing the gel on the sinter with 300 ml of a 40 mM solution of cysteine in the Tris/HCl buffer. The filtrate was adjusted to pH ~ 5 by addition of 1 M HCl, and chymopapain M was precipitated with 85% $(\text{NH}_4)_2\text{SO}_4$ and isolated by centrifugation as described above. A sample was dissolved in 0.1 M sodium acetate buffer, pH 5.0, and evaluated by analytical cation-exchange f.p.l.c. using a Mono S HR 5/5 column [11]. When the analysis showed the presence of traces of other chymopapains, preparative cation-exchange chromatography was carried out as described above to produce pure, fully active enzyme. A preparation on the scale described provides sufficient chymopapain M for about 500 stopped-flow kinetic runs with 2-Py-S-S reactivity probes or up to about 5000 initial-rate determinations using Boc-Gly-4NPE as substrate.

Investigation of chymopapain M, papain and actinidin and their interactions with ligands by computer modelling

The GSH-S-S-2-Py gel was represented by the pyridyl derivative of GSH [extracted from the 3D Fine Chemicals Directory (3D) (MDL, San Leandro, CA, U.S.A.); the structures in this directory

have been run through CONCORD and provide three- rather than two-dimensional coordinate versions, modified using QUANTA (MSI, Burlington, MA, U.S.A.) so that both of its carboxy groups were ionized and a terminal amino group was added (where it is attached to the agarose oxygen atoms). The structures of papain and actinidin were taken from the Brookhaven protein database (entries 9pap [9] and 2act [13] respectively), while chymopapain M was the homology-modelled structure described previously [11]. The sulphur atom of the catalytic-site cysteine residue was found to be bonded to oxygen atoms in both of the Brookhaven co-ordinate sets, and these oxygen atoms were removed before carrying out any modelling studies. Electrostatic-potential calculations were performed by using the UHBD program [32], which solves the linearized Poisson–Boltzmann equation using a finite-difference method. Each system was represented with a constant interior dielectric constant of 2.0, a solvent dielectric constant of 78.0 and an ionic strength of zero at a temperature of 300 K. The presence or absence of crystal or added water molecules does not vastly change the overall distribution of the potentials, as has been found for the wild-type and a mutant yeast triosephosphate isomerase [33]. Maps were evaluated across a $0.65 \text{ nm} \times 0.65 \text{ nm} \times 0.65 \text{ nm}$ ($65 \text{ \AA} \times 65 \text{ \AA} \times 65 \text{ \AA}$) grid with a spacing of 0.1 nm ($1.0 \text{ \AA}$). The potential at the grid boundary was set to zero. QUANTA (MSI, Burlington, MA, U.S.A.) was used to examine the maps. Minimization and molecular-dynamics (m.d.) simulations were carried out using the extended-atom CHARMM force field [34], in which only the polar hydrogen atoms are represented explicitly. Polar hydrogen atoms were added using the HBUILD functionality [35] of CHARMM. Charges were assigned using the standard CHARMM parameter sets. The non-bonded interactions were calculated using a Coulomb potential, with the solvent screening effected by the use of a distance-dependent dielectric and were truncated by using a shifted potential [34] with a 1.1 nm ($11 \text{ \AA}$) cut-off [36], as were the Lennard-Jones–van der Waals interactions. In the presence of explicit solvent a constant dielectric of 1.0 was used. Electrostatic interactions between atoms separated by three covalent bonds were scaled to 0.5 of their value. For the m.d. simulations, the protein was solvated to 1.5 nm ($15 \text{ \AA}$) from the S^γ atom of cysteine-25. This resulted in the active centre being ‘capped’ by a shell of water molecules. All non-hydrogen atoms were constrained by a harmonic force constant of $418.4 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ ($1.0 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$) and the whole system was subject to 100 steps of steepest descent and 200 steps of adopted basis newton raphson (a.b.n.r.) minimizations [34] to relax the protein–water system. All polar hydrogen atoms were rebuilt in the mutual fields of the proteins and waters using the HBUILD option of CHARMM. Any residue which had an atom within 1.2 nm ($12 \text{ \AA}$) of the catalytic cysteine residue was kept free while the rest of the system, including the outer 0.3 nm ($3 \text{ \AA}$) layer of the ‘water-cap’, was fixed; m.d. simulations were started by assigning random velocities (corresponding to an average temperature of 300 K) to the atoms from a Maxwell–Boltzmann distribution. The SHAKE algorithm [37] was used to constrain all bond lengths to their equilibrium values, enabling an integration time step of 0.002 ps . The system was coupled to an external bath with constant temperature and a coupling constant of 0.05 ps [38]. Non-bonded lists were updated every 0.05 ps using a 1.3 nm ($13 \text{ \AA}$) cut-off, and any net translation and rotation in the system was removed every 4 ps . This procedure was carried out for all three proteins. The systems were equilibrated until the total potential energy was stable (this took about 20 ps for each system) and the co-ordinates were then saved every 0.2 ps for 80 ps . In order to model the binding of papain to the gel, the

disulphide bond of the pyridyl-GSH moiety was cleaved, the GSH sulphur atom was covalently bonded to the S^γ atom of cysteine-25 of the protein, and the atomic potential was redefined to be analogous to the cysteine disulphide potential used in CHARMM. The conformational surface of the bound ligand (in the absence of experimental data) was evaluated by restrained minimizations in a manner similar to the adiabatic minimization method used to model reaction paths [39]. The ligand was rotated across the disulphide linkage to the cysteine residue from -180° to $+180^\circ$ in increments of 20° . At each point, residues that had any atom within 1.0 nm ($10 \text{ \AA}$) of any atom of the ligand were constrained by a harmonic force constant of $41.84 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ ($10^{-1} \cdot \text{kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$), while the rest of the system was fixed. This force constant was relaxed to 4.184 and $0.4184 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ (10^{-2} and $10^{-3} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$) for further examinations of the conformational flexibility of the bound ligand. The ligand itself was left free to relax into its local minimum. This resulted in several local minima being found. The structure for further study was chosen by visual inspection of the models. The whole system was subjected to steps of conjugate gradient and a.b.n.r. minimizations until the change in the average force on the atoms between subsequent steps was less than, or equal to, $0.04184 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ ($10^{-3} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$).

The interaction between cystatin and papain was carried out by first building the missing eight N-terminal residues of cystatin. These were built from the C-terminal side to the N-terminal side by successively building the residues, finding the minimum energy conformation of the backbone and minimizing the residues built. After all of the residues had been built, they were subjected to m.d. simulations at 1000 K for 100 ps , and each structure (of the last 80 ps) was then minimized until the average force was less than, or equal to, $0.04184 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ ($10^{-3} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$). The calculations were carried out for cystatin alone as well as while it was complexed to papain (after superposing the C^α atoms of cystatin on to the stefin part of the stefin–papain complex [40]). Finally, the complex was subjected to extensive minimizations with only a central region around the interface of radius 1.5 nm ($15 \text{ \AA}$) free to move while the rest of the system was constrained by a force constant of $0.4184 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-2}$ ($10^{-5} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$), until the change in the average force on the atoms between subsequent steps was less than, or equal to, $0.04184 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ ($10^{-3} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$). All calculations were carried out on a network of Silicon Graphics workstations. Analyses were carried out by using QUANTA and SQUID [41].

Figure 1 Structures of (a) the Sepharose–GSH–S–S–2–Py gel and (b) the Sepharose–2–hydroxypropyl–S–S–2′–Py gel

Figure 2 Electrostatic maps at ± 2 kT from the UHBD calculations (see the text) showing papain, chymopapain M and actinidin with the L domain at the bottom

The structure labelled 'glutathione' is the GSH-S-S-2-Py ligand of the Sepharose–GSH-S-S-2-Py gel; its orientation is such that the pyridyl ring is to the left and the whole molecule lies parallel to the active-centre clefts of the enzymes shown; the ligand and the enzymes are not on the same scale.

RESULTS AND DISCUSSION

Isolation of fully active chymopapain M by covalent chromatography

The failure of chymopapain M to react with the GSH gel (Figure 1a) permitted its separation from a chymopapain B contaminant, which did bond to the gel, but did not allow its separation from oxidized chymopapain M (presumably the sulphinic acid) because both oxidized and non-oxidized forms pass through the gel unchanged [11]. In the present work, use of the hydroxypropyl gel (Figure 1b) instead of the GSH gel has been shown to permit isolation of fully active chymopapain M containing 1 mol of catalytic-site cysteine-25/histidine-159 ion-pair/mol of protein. Cysteine proteinase catalytic-site ion-pairs are characterized by rapid reaction (instantaneous on the time scale of a conventional spectrophotometer) with 2-Py-S-S-2-Py at pH 4.0 (see [1,2]). The enzyme that reacted with the hydroxypropyl gel at pH 4.0 to release pyridine-2-thione ($\lambda_{\max}$ 343 nm) and was subsequently released by elutive reaction with L-cysteine at pH 8.0 was identified as chymopapain M by its failure to catalyse the hydrolysis of the P_1 -arginine substrate Bz-L-Arg-4NA, its successful catalysis of the hydrolysis of the P_1 -glycine substrate Boc-Gly-4NPE and the lack of inhibition of the latter catalysed reaction by chicken cystatin. This combination of properties is

unique to chymopapain M, at least among the cysteine proteinase components of papaya latex [11,20].

Assessment of steric and electrostatic interpretations of the inability of chymopapain M and of actinidin to bond to the GSH gel

Computer modelling was used to assess possible reasons for the inability of both chymopapain M and actinidin to bond to the GSH gel. One potential reason for the lack of reaction is electrostatic repulsion. The net charges on papain and chymopapain M are -8.75 and -8.80 electron units (e.u.) respectively, while on actinidin the net charge is -16.20 e.u. The polarities of the residues in the vicinity of the active centre are almost the same in papain and chymopapain M, while the analogous residues in actinidin are more polar, with several ionic residues replacing neutral residues in papain. Such residues (one-letter code; 15D = aspartic acid-15 etc.) in actinidin, with the corresponding papain residues in parentheses, include: 15D (15P); 21E (21S); 66D (64N); 72D (70S); 145K (142Q); 185D (178G); 191E (184N); and 213S (207F). We examined several isopotential contours from calculations using the UHBD program [32] and saw trends similar to the 2 kT potential shown in Figure 2. In papain the 2 kT level extends to about 0.2 nm (2 Å) from the

Figure 3 Stereo-view snapshots from m.d. simulations at 300 K of the active centres of (a) papain, (b) actinidin and (c) chymopapain M solvated to 1.5 nm (15 Å) around the catalytic cysteine residue (labelled C25)

The active centres (including the labelled residues C25, G23, G65 and D158 in papain and C25, G23, G67 and D161 in actinidin) remain exposed in papain and actinidin, whereas the cap formed by the labelled residues E23-R65-D158 in chymopapain M shows how the catalytic cysteine residue is inaccessible to large ligands such as GSH. For clarity, hydrogen bonds between E23, R65, D158 and bridging water molecules are not shown.

Figure 3 For caption see opposite.

Figure 4 A model, in stereo view, of the conformation of GSH (thick broken lines) linked in disulphide-bond formation to the catalytic-site cysteine residue (labelled C25) of papain produced from the minimization studies described in the text

The terminal amine (labelled N-A) represents the nitrogen exposed to solvent and bonded in the gel to the agarose oxygen atoms. The residues shown as thick continuous lines and labelled are the residues from chymopapain M superposed on papain, glutamic acid-23 (glycine-23 in papain), arginine-65 (glycine-65) and aspartic acid-158 (aspartic acid-158), and show how the conformations of glutamic acid-23 and arginine-65 in chymopapain M would clash with any stable conformation of the GSH moiety and hence could inhibit its binding.

surface of the protein and is largely positive, with traces of amphipathicity along the active centre, while chymopapain M is similar with some larger areas of negative potential. However, while the distribution close to the protein surface in actinidin is similar to those of papain and chymopapain M, a large negative potential envelops actinidin almost completely and extends out to about 0.8 nm (8 Å) from the mouth of the active-centre cleft. This is not surprising in view of the presence of the ionic side chains in the vicinity of the active-centre region of actinidin. The potential around the GSH-S-S-2-Py ligand is predominantly negative, owing especially to the two carboxylate anions and the nitrogen atom of the pyridyl ring. This leads to a non-complementarity between the potentials enveloping the GSH ligand and the actinidin active centre and explains why actinidin does not bond to this gel. The changes to the sequence of actinidin discussed by Varughese et al. [42] should not alter this conclusion in any way.

In the case of papain and chymopapain M, the potentials are complementary to that of the GSH ligand and would allow for association and subsequent reaction. Bonding is observed, however, only in the case of papain, and so we examined the m.d. simulations to identify any steric factors that would inhibit the bonding of chymopapain M to the gel. Throughout the 100 ps simulations, the active centres of papain and actinidin remain exposed and solvated (Figures 3a and 3b). Although there is no crystallographic evidence of side-chain orientations in and around the active centre of chymopapain M, in the modelled structure a network of hydrogen bonds exists between E23, R65, D158 and several water molecules, forming a 'cap' over the catalytic-site cysteine residue. This network, which cannot exist in papain because residues 23 and 65 are both glycine in this enzyme, could prevent a large branched ligand such as GSH (Figure 1a) from reaching and bonding to the catalytic site of chymopapain M (Figure 3c). Hydrogen-bond statistics from the simulation show that hydrogen bonds between R65 and E23 (between the N atom and the NH₂ group of R65 and the carboxy oxygen atoms of E23) exist for about 98 % of the simulation, while a hydrogen bond between R65 (NH₂ group) and D158 (carboxy oxygen atom) exists for about 70 % of the time and the

latter is further bridged by two water molecules through hydrogen bonds for about 20–50 % of the time.

Among the restrained minimization studies, only structures with the initial papain–GSH disulphide torsion angle between -180° and -120° and between 100° and 180° resulted in acceptable conformers (all resulted in minima which clustered around the torsion angle between -160° to $+160^\circ$ in the *gauche* region), while the rest of the structures involved steric clashes between the GSH ligand and the active-centre residues. Among the accepted structures, the active-centre residues did not change positions between the studies of decreasing constraints. This suggests that these do not move much in response to the binding of the GSH ligand, a conclusion consistent with the fact that the active-centre cleft is wide enough to accommodate such ligands. These models suggest that the interaction between the GSH ligand and papain is stabilized by contact with the backbone atoms of residues G20, C22, G23, A136, D158, H159, A160 and A180, and with the side-chain atoms of Q19, A137, Q142, D158, H159 and W177; these contacts are further complemented by hydrogen bonds between the carboxy groups of the ligand and protons on N19, N64, H159 and W177. Figure 4 shows a model of GSH bonded to papain as a disulphide with C25. The three cap-forming residues of chymopapain M are also shown. The conformations of E23 (G23 in papain) and R65 (G65 in papain) would clearly result in steric clashes between chymopapain M and the ligand (in all energetically feasible conformations of the ligand). The presence of the side chain of E23 accounts also for the almost absolute specificity exerted by chymopapain M for substrates with glycine in the P₁ position. The kinetically influential ionization with pK_a 3.4 that is unique to this enzyme has been assigned to the carboxy group of E23 [11], a residue that is glycine in papain, actinidin and PPΩ.

The structural basis of the lack of inhibition of chymopapain M by cystatin

The crystal structure of a short form of chicken egg-white cystatin (starting from residue G9) has been reported previously

Figure 5 Stereo models of the C α atoms of (a) cystatin complexed with papain and (b) cystatin complexed with papain, but showing also the three 'cap'-forming residues of chymopapain M

[29]. In the present work this structure has been complemented by the addition, using computer modelling, of the missing eight N-terminal residues. A model of papain complexed with this full-length cystatin is shown in Figure 5(a). The N-terminal 'trunk' extends along the side of the papain cleft, and the interaction of papain and cystatin is characterized by 31 close [≤ 0.3 nm (3 Å)] contacts (all but two of them being hydrogen bonds), most of which are localized to the N-terminal trunk region. The finding that the greater the number of N-terminal residues present (before G9) the tighter the binding of the inhibitor to the enzyme [43] is explained by this modelled structure, in particular the contacts that it predicts between the enzyme and the N-terminal trunk of the inhibitor. The net van der Waals interaction is negative in energy, and the electrostatic interaction is about 10 % of the overall electrostatic energy of the whole enzyme-inhibitor complex. The conformations of E23 and R65 of chymopapain M would sterically prevent cystatin from binding along the cleft, at least in a manner similar to the binding modelled for papain (Figure 5b).

Validation of an aspect of papain specificity by using chymopapain M as a structural variant

The principal specificity determinants of papain are the hydrophobic S_2 subsite, the backbone-NH of glycine-66 and the backbone $>C=O$ of aspartic acid-158 (see [44] for a critical appraisal of binding modes). Some additional specificity is exerted by the structure of the S_1 subsite in that although substrates with lysine, arginine or citrulline at P_1 are accepted by papain, the β -branched side chain of valine cannot be accommodated between the α -CH₂ groups of glycine-23 and glycine-65 [30,45]. Lowe and Yuthavong [30] suggested that replacement of glycine-23 by any other residue would result in papain being unable to accommodate residues larger than alanine in the P_1 position of substrates. The close similarity in most respects between the structure of papain and the modelled structure of chymopapain M and the particular difference at residue 23 (glycine in papain and glutamic acid in chymopapain M) [11] provides a means of testing the suggestion made by Lowe and Yuthavong [30]. The fact that chymopapain M will catalyse the hydrolysis of substrates containing only glycine, or much less effectively alanine [22], at P_1 provides support for the hypothesis that mutation of glycine-23 of papain to any other amino acid residue would severely restrict its S_1/P_1 molecular-recognition characteristics.

We thank R. Huber and W. Bode for supplying the co-ordinates of chicken egg-white cystatin and recombinant human stefin B prior to their public release, the Science and Engineering Research Council for an Earmarked Research Studentship and a Postdoctoral Research Assistantship for M.P.T., the Wellcome Trust for a Research Fellowship for C.V., and Zeneca Pharmaceuticals (Alderly Edge, Macclesfield, Cheshire, U.K.) and Zeneca Agrochemicals (Jealott's Hill, Berkshire, U.K.) for funding S.M.B.

REFERENCES

- 1 Brocklehurst, K. (1987) in *Enzyme Mechanisms* (Page, M. I. and Williams, A., eds.), pp. 140–158, Royal Society of Chemistry, London
- 2 Brocklehurst, K., Willenbrock, F. and Salih, E. (1987) In *Hydrolytic Enzymes* (Neuberger, A. and Brocklehurst, K., eds.), pp. 39–158, Elsevier, Amsterdam, New York and Oxford
- 3 Topham, C. M., Salih, E., Frazao, C., Kowlessur, D., Overington, J. P., Thomas, M., Brocklehurst, S. M., Patel, M., Thomas, E. W. and Brocklehurst, K. (1991) *Biochem. J.* **280**, 79–92
- 4 Patel, M., Saleem, I., Mellor, G. W., Sreedharan, S., Templeton, W., Thomas, E. W., Thomas M. and Brocklehurst, K. (1992) *Biochem. J.* **281**, 553–559
- 5 Mellor, G. W., Sreedharan, S. K., Kowlessur, D., Thomas, E. W. and Brocklehurst, K. (1993) *Biochem. J.* **290**, 75–83
- 6 Mellor, G. W., Thomas, E. W., Topham, C. M. and Brocklehurst, K. (1993) *Biochem. J.* **290**, 289–296
- 7 Mellor, G. W., Patel, M., Thomas, E. W. and Brocklehurst, K. (1993) *Biochem. J.* **294**, 201–210
- 8 Brocklehurst, K., Baines, B. S. and Kierstan, M. P. J. (1981) *Top. Enzyme Ferment. Biotechnol.* **5**, 262–335
- 9 Kamphuis, I. G., Drenth, J. and Baker, E. N. (1985) *J. Mol. Biol.* **182**, 317–329
- 10 Pickersgill, R. W., Rizhalla, P., Harris, G. W. and Goodenough, P. W. (1991) *Acta Crystallogr.* **B47**, 766–771
- 11 Thomas, M. P., Topham, C. M., Kowlessur, D., Mellor, G. W., Thomas, E. W., Whitford, D. and Brocklehurst, K. (1994) *Biochem. J.* **300**, 805–820
- 12 Ritonja, A., Buttle, D. J., Rawlings, N. D., Turk, V. and Barrett, A. J. (1989) *FEBS Lett.* **258**, 109–112
- 13 Baker, E. N. (1980) *J. Mol. Biol.* **141**, 441–484
- 14 Blundell, T., Carney, D., Gardiner, S., Hayes, F., Howlin, B., Hubbard, T., Overington, J., Singh, D. A., Sibanda, L. and Sutcliffe, M. (1988) *Eur. J. Biochem.* **172**, 513–520
- 15 Sutcliffe, M. J., Haneef, I., Carney, D. and Blundell, T. L. (1987) *Protein. Eng.* **1**, 377–384
- 16 Sutcliffe, M. J., Hayes, F. R. F. and Blundell, T. L. (1987) *Protein Eng.* **1**, 385–392
- 17 Schack, P. (1967) *Compt. Rend. Trav. Lab. Carlsberg.* **36**, 67–83
- 18 Lynn, K. R. (1979) *Biochim. Biophys. Acta* **569**, 193–201.
- 19 Polgár, L. (1981) *Biochim. Biophys. Acta* **658**, 262–269
- 20 Buttle, D. J., Kembhavi, A. A., Sharp, S. L., Shute, R. E., Rich, D. H. and Barrett, A. J. (1989) *Biochem. J.* **261**, 469–476
- 21 Ritonja, A., Čolić, A., Dolenc, I., Ogrinc, T., Podobnik, M. and Turk, V. (1991) *FEBS Lett.* **283**, 329–331
- 22 Buttle, D. J., Ritonja, A., Pearl, L. H., Turk, V. and Barrett, A. J. (1990) *FEBS Lett.* **260**, 195–197
- 23 Brocklehurst, K., Carlsson, J., Kierstan, M. P. J. and Crook, E. M. (1973) *Biochem. J.* **133**, 573–584
- 24 Brocklehurst, K., Carlsson, J., Kierstan, M. P. J. and Crook, E. M. (1974) *Methods Enzymol.* **34B**, 531–544
- 25 Brocklehurst, K., Carlsson, J. and Kierstan, M. P. J. (1985) *Top. Enzyme Ferment. Biotechnol.* **10**, 146–188
- 26 Brocklehurst, K. (1995) in *Proteins Labfax* (Price, N. C. ed.), Bios Scientific Publishers, Oxford, in the press
- 27 McDowell, M. A. (1970) *Eur. J. Biochem.* **14**, 214–221
- 28 Brocklehurst, K., Baines, B. S. and Malthouse, J. P. G. (1981) *Biochem. J.* **197**, 739–746
- 29 Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988) *EMBO J.* **7**, 2593–2599
- 30 Lowe, G. and Yuthavong, Y. (1971) *Biochem. J.* **124**, 107–115
- 31 Malthouse, J. P. G. and Brocklehurst, K. (1976) *Biochem. J.* **159**, 221–234
- 32 Davis, M. E., Madura, J. D., Sines, J., Luty, B. A., Allison, S. A. and McCammon, J. A. (1991) *Methods Enzymol.* **202A**, 473–497
- 33 Joseph-McCarthy, D., Lolis, E., Komives, E. A. and Petsko, G. A. (1994) *Biochemistry*, **33**, 2815–2823
- 34 Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) *J. Comp. Chem.* **4**, 187–217
- 35 Brunger, A. and Karplus, M. (1988) *Protein Struct. Funct. Genet.* **6**, 32–45
- 36 Loncharich, R. J. and Brooks, B. R. (1989) *Protein Struct. Funct. Genet.* **6**, 32–45
- 37 van Gunsteren, W. F. and Karplus, M. (1982) *Macromolecules* **15**, 1528–1544
- 38 Berendsen, H. J. C., Postma, P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. (1984) *J. Chem. Phys.*, **81**, 3684–3690
- 39 McCammon, J. A. and Harvey, S. C. (1987) *Dynamics Of Proteins And Nucleic Acids*, Cambridge University Press, Cambridge
- 40 Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, R. and Turk, V. (1990) *EMBO J.* **9**, 1939–1947
- 41 Oldfield, T. J. (1992) *J. Mol. Graphics* **10**, 247–252
- 42 Varughese, K. I., Su, Y., Cromwell, D., Hasnain, S. and Xuong, N. (1992) *Biochemistry* **31**, 5172–5176
- 43 Björk, L., Pol, E., Raub-Segall, E., Abrahamson, M., Rowan, A. D. and Mort, J. S. (1994) *Biochem. J.* **299**, 219–225
- 44 Templeton, W., Kowlessur, D., Thomas, E. W., Topham, C. M. and Brocklehurst, K. (1990) *Biochem. J.* **266**, 645–651
- 45 Berger, A. and Schechter, L. (1970) *Philos. Trans. R. Soc. B*, **257**, 249–264