Structure of Chymopapain at 1.7 Å Resolution†

Dominique Maes,*,‡ Julie Bouckaert,‡ Freddy Poortmans,§ Lode Wyns,‡ and Yvan Looze‖

Ultrastructure Unit, Vlaams Interuniversitair Instituut voor Biotechnologie, Vrije Universiteit Brussel, Paardenstraat 65, B-1640 Sint-Genesius-Rode, Belgium, Vlaams Instituut voor Technologisch Onderzoek, Boeretang 200, B-2400 Mol, Belgium, and Protein Chemistry Unit, Faculté de Médecine, Université Libre de Bruxelles, Campus de la Plaine, Boulevard du Triomphe, B-1050 Bruxelles, Belgium

Received June 21, 1996⊗*; Revised Manuscript Received September 27, 1996*

ABSTRACT: The X-ray structure of chymopapain, a cysteine proteinase isolated from the latex of the fruits of *Carica papaya* L., has been determined by molecular replacement methods and refined to a conventional *R* factor of 0.19 for all observed reflections in the range from 9.5 to 1.7 Å resolution. The crystals used in this study contained a unique molecular species of chymopapain with two moles of thiomethyl attached to the two free cysteines per mole of enzyme. A comparison is made with the other known papaya proteinase X-ray structures: papain, caricain, and glycyl endopeptidase. Their backbone conformations are extremely similar except for two loop regions. Both regions are located at the surface of the protein and far away of the active site cleft. In each X-ray structure the same water network was found at the interface between the two domains of the enzyme. A close examination of the active site groove showed that the specificity restrictions dictated by the S2 subsite did not differ significantly among the four proteinases.

Chymopapain [EC 3.4.22.6] is an extracellular cysteine proteinase secreted in the latex of the tropical tree *Carica papaya* L. Like the three other well-documented papaya proteinases papain [EC 3.4.22.2], caricain [EC 3.4.22.30], and glycyl endopeptidase [EC 3.4.22.25], chymopapain is synthesized with a signal and a propeptide at its N terminus (Cohen et al., 1986; Mc Kee et al., 1986; Revell et al., 1993; Karrer et al., 1993; Taylor et al., 1995). The presence of the pro-regions is required for correct folding of the proteinases (Revell et al., 1993; Vernet et al., 1995). On the other hand, these propeptides are selective high-affinity inhibitors of the mature papaya enzymes (Taylor et al., 1995; Cygler et al., 1996). The biological significance of these observations remains obscure since no clear physiological function for the presence of active proteinases in the papaya latex has yet been firmly established.

About 20 families of proteinases, dependent on a cysteine residue at the active site center, have been recognized. Chymopapain belongs to the papain (C1) family (Jacquet et al., 1989a; Watson et al., 1990) which gathers enzymes found in a wide variety of life forms as well as a wide variety of activities. The papain family indeed contains endopeptidases with either broad (e.g., chymopapain) or very narrow (e.g., glycyl endopeptidase) specificities, aminopeptidases, a peptidylpeptidase, and even peptidases with both endo- and exopeptidase activities (e.g., cathepsins B and H). Finally, there are also family members (such as *Gly m* I from *Glycine max*) that show no proteolytic activity (Rawlings & Barrett, 1994).

In the study of molecular recognition of active center chemistry and catalytic mechanism, the natural structural variation that exists within the papain family is particularly useful. In this context, three-dimensional structures are a necessity. Accurate X-ray structures are now available for several members of the C1 family: papain (Drenth et al., 1962; Kamphuis et al., 1984; Pickersgill et al., 1992), actinidin (Baker, 1980), caricain (Pickersgill et al., 1991, 1993), glycyl endopeptidase (O'Hara et al., 1995), cruzipain (Mc Grath et al., 1995), and rat and human cathepsins B (Jia et al., 1995; Musil et al., 1991).

The need for determining the X-ray structure of chymopapain, which is the object of the present paper, was dictated by several independent considerations.

At the best of our knowledge, chymopapain is unique among the members of the papain family, in existing as a mixture of two forms, named A and B. Designated jointly as EC 3.4.22.6 the two chymopapain forms are immunologically indistinguishable (Buttle & Barrett, 1984) and have identical amino acid sequences (Jacquet et al., 1989a; Watson et al., 1990). However, the two forms differ in their reactivity with two-protonic-state reactivity probes (Khan & Polgar, 1983; Baines et al., 1986) and ester substrates (Korodi et al., 1986). Attempts to determine the structural basis for these observations through modeling or through experimental investigations of the chymopapain physicochemical properties have been unsuccessful (Topham et al., 1990; Jacquet, 1990; Solis-Mendiola et al., 1989; Summer et al., 1993).

In common with papain, chymopapain shares a broad specificity toward substrates (Jacquet et al., 1989b). Papain, nevertheless, shows a preference for aromatic (Phe or Tyr) residues at P2. The side chains of Val 133 and Val 157 in the S2 subsite of papain are substituted by leucyl side chains in chymopapain (Jacquet et al., 1989a; Watson et al., 1990). Such substitutions are predicted to restrict the accessibility of the S2 pocket and thus to affect the reactivity of the enzyme in such a way that only smaller groups at P2 could

† This work was supported by the VLAB project of the Flemish government. D.M. is a research associate of the Belgian National Science Foundation (NFWO). J.B. received a grant from the IWT.

* Corresponding author.
‡ Vrije Universiteit Brussel.
§ Vlaams Instituut voor Technologisch Onderzoek.
‖ Université Libre de Bruxelles.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1996.

S0006-2960(96)01491-2 CCC: $12.00

provide effective substrates (Jacquet, 1990; Watson et al., 1990; Topham et al., 1990). This prediction has not been confirmed experimentally with protein substrates (Jacquet et al., 1989b) or with synthetic substrates (Walraevens et al., 1993).

EXPERIMENTAL PROCEDURES

Crystallization and Data Collection

A chymopapain solution was prepared in which the two free cysteines (Cys 25 and Cys 117) were protected with a thiomethyl group. Plate shaped crystals were grown at 20 °C from a 6 mg/mL (0.25 mM) solution of this fully protected chymopapain containing 2 M sodium acetate adjusted to a pH of 7.6 by the addition of acetic acid. Full details concerning the purification and the crystallization procedure have been published previously (Azarkan et al., 1996).

The protein crystallized in the monoclinic space group *C*2 with $a = 145.18(10)$, $b = 32.35(7)$, $c = 47.42(6)$ Å, $\beta = 98.37(7)°$, and with one molecule in the asymmetric unit.

Diffraction intensities were collected on an Enraf-Nonius Fast Area detector at 293 K, using Cu Kα radiation generated by a rotating anode source operated at 45 kV and 98 mA. Crystals were mounted in a glass capillary with an arbitrary orientation with respect to the camera axes. The detector distance was 120 mm. Diffraction extended up to 1.4 Å. Data to 2.2 Å resolution were collected on a first crystal of size 0.40 × 0.40 × 0.08 mm at 60 and 150 s/frame with each consisting of a 0.12° rotation in ω with 2θ equal to 11.5° and 30°, respectively. On a second crystal of size 0.60 × 0.25 × 0.03 mm, high-resolution data from 2.8 to 1.7 Å were measured with the same parameters except for an exposure time of 150 s/frame and a 2θ of 43°. The data were processed using the MADNES software (Messerschmidt & Pflugrath, 1987) and merged, scaled, and truncated using the CCP4 programs ROTAVATA/AGROVATA and TRUNCATE. Details of the data collection statistics are summarized in Table 1.

Molecular Replacement

Following alignment there is about 60% sequence homology between the four papaya proteinases. Molecular replacement was performed using the crystal structure of caricain (papaya protease ω) (entry 1ppo in the Brookhaven Protein Data Banks; Bernstein et al., 1977; Pickersgill et al., 1991) as a search model as it required no deletions and insertions, except for the addition of two amino acids at the C-terminal end of chymopapain. The model was altered in such a way that in positions where the sequence of caricain and chymopapain differed, the residue was replaced by alanine (or glycine, when the corresponding residue in chymopapain was of that type). The orientation of the molecule in the unit cell was determined using the Crowther rotation function as implemented in the ALMN program (CCP4). Structure factors were calculated from the search model placed in a triclinic cell with $a = b = c = 80$ Å and $\alpha = \beta = \gamma = 90°$. The largest peak, 8σ above background, in the cross rotational function was found for the Eulerian angles $\alpha = 211.8°$, $\beta = 114.7°$, and $\gamma = 34.4°$. This orientation was found consistently using different resolution ranges between 12 and 4 Å. The rotated molecule was positioned within the *C*2 cell by using the translation function of the program TRANSZ (CCP4). A clear peak was observed 10σ above background for $f_a = -15.3$ Å and $f_c = -16.1$ Å.

Table 1

Data Collection Statistics

resol. shell	no. of unique refl.	% complete	I/σ	R_{merge}	cumulative R_{merge}
∞−5.21	888	99.2	13.2	0.043	0.043
5.21−3.75	1437	98.9	11.9	0.051	0.049
3.75−3.08	1744	96.0	10.0	0.058	0.051
3.08−2.67	1949	91.5	8.1	0.079	0.054
2.67−2.40	2226	92.7	6.7	0.103	0.059
2.40−2.19	2435	91.9	5.4	0.124	0.064
2.19−2.03	2326	80.8	5.9	0.104	0.065
2.03−1.90	2372	77.1	5.0	0.130	0.065
1.90−1.79	2469	75.3	3.7	0.186	0.066
1.79−1.70	2374	68.8	3.3	0.222	0.066

total no. of observed reflections: 39 180
total no. of unique reflections: 20 199

Refinement Statistics	
resolution range	9.5−1.7 Å
R factor (all data)	0.192
no. of non-hydrogen protein atoms	1667
no. of water molecules	222
no. of others	2 thiomethyl (SCH_3)
RMS deviation in bond lengths	0.011 Å
RMS deviation in bond angles	1.57°
RMS deviation in improper torsions	1.46°

Refinement

To improve the molecular replacement solution, several rigid body refinement cycles with data ranging from 10 to 2.5 Å resolution were performed using the X-PLOR program package (Brunger et al., 1992) running on a Silicon Graphics Indy workstation. The latter yielded an *R* factor of about 0.45 with a correlation of 73.2% between the observed and calculated structure factors.

Several rounds of positional and temperature factor least-squares refinement, difference map calculation ($F_o - F_c$, $2F_o - F_c$), and model refitting were undertaken using the X-PLOR package (Brunger et al., 1992) for the calculations. Modification of the model was carried out on the same computer using the program O (Jones et al., 1991).

At completion of the first round of refinement the side chains were built into the difference density. The C-terminal end from residue Thr 203 on was discarded from the model. In the consecutive cycles these residues were rebuilt and the model was updated to include the two thiomethyl (SCH_3) moieties, which formed disulphide bridges with Cys 25 and Cys 117, respectively. Individual *B* factors were introduced in the refinement.

The solvent molecules were selected from peaks which appeared in both electron density maps and made reasonable bonds with the protein atoms and other solvent molecules. A total of 222 solvent molecules, all having a fixed occupancy of 1, were successfully included in the refinement.

Finally, alternative side-chain conformations were modeled into the structure and the position as well as the occupancy of these alternative sites were refined.

The final *R* value is 19.2% for all reflections between 9.5 and 1.7 Å resolution. The results of the refinement procedure are summarized in Table 1. The quality of the final electron

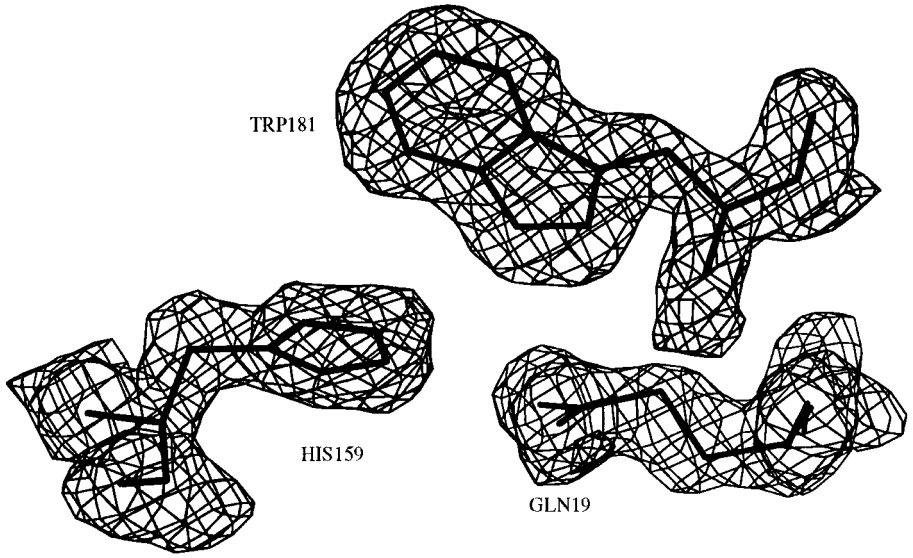

FIGURE 1: Gln 19, His 159, and Trp 181 shown in the final electron density map.

Table 2[a]

Intermolecular Contacts	
no. of intermolecular contacts (d < 4 Å)	135
hydrogen bonds	14
salt bridges	0
van der Waals contacts	121
no. of residues involved in crystal contacts	44
no. of intermolecular hydrogen bonds mediated by one water molecule	15

Intermolecular Hydrogen Bond Contacts		distance (Å)
1 Tyr O	93 Tyr OH	3.41
67 Tyr OH	184 Asn OD1	2.76
67 Tyr OH	184 Asn ND2	3.21
188 Lys NZ	209 Ser OG	3.00
104 Lys O	218 Ala N	2.88
106 Lys N	218 Ala O	2.98
215 Lys NZ	218 Ala O	3.15
215 Lys NZ	218 Ala OT	2.75
66 Gly O	139 Lys NZ	3.07
170 Asp OD2	197 Ser N	3.03
170 Asp OD2	197 Ser OG	3.19
172 Lys NZ	197 Ser O	2.94
172 Lys NZ	197 Ser OG	3.45
194 Lys NZ	199 Asn OD1	2.83

[a] Packing contacts were determined with the program X-PLOR (Brunger, 1992) using the following criteria: 2.4 < d < 3.5 Å for hydrogen bonds; 1.75 < d < 3.5 Å for salt bridges; and d < 4.0 Å for van der Waals contacts.

density is illustrated in Figure 1. The refined atomic parameters and structure factors have been deposited at the Protein Data Bank at Brookhaven (Bernstein et al., 1977). Intermolecular contacts were calculated with the X-PLOR package (Brunger et al., 1992) and are tabulated in Table 2. Figures were drawn using MOLSCRIPT (Kraulis, 1991).

Comparative Study

An overview of the details of the X-ray structures used for the comparative analysis of the four papaya proteases is given in Table 3 (Pickersgill et al., 1991, 1992; O'Hara et al., 1995; Kamphuis et al., 1984). Structural data on papain are abundant; for the purpose of generality a monoclinic as well as an orthorhombic structure was chosen. The choice of these structures was based on resolution and active site occupancy; when possible a structure with a resolution similar to 1.7 Å was chosen and furthermore we opted for a structure with the smallest distortion of the active site compared to the native enzyme.

The structures were superimposed, excluding insertions/deletions and regions with large shifts (vide infra), and their basic folds were compared using the program LSQMAN (Kleywegt & Jones, 1994). The solvent structures were studied with the FIXWAT program (Lisgarten et al., 1993). Waters from different structures were considered to be equivalent if they were within 4 Å of the corresponding protein and if the centers of both waters were within a distance of 1 Å after optimal superposition of both structures.

RESULTS AND DISCUSSION

Chymopapain Structure

The basic fold of chymopapain is extremely similar to that of the other papaya protease structures. The polypeptide chain is folded into two domains of roughly the same size but different conformation (Figure 2). One domain (residues 10−111 and 211−218; the L-domain) is mainly α-helical; the other domain (residues 1−9 and 112−210; the R-domain) essentially consists of extensive antiparallel β-sheet interactions. The active site is located at the interface between these domains, with the active center Cys 25 positioned at the start of the α-helix, composed of residues 24−42 in the L-domain.

The stereochemical quality of the refined model was analyzed with the program PROCHECK (Laskowski et al., 1993). The Ramachandran plot shows all residues in the allowed regions of conformational space, with 87.6% of the residues in most favored regions and 12.4% in additional allowed regions. All stereochemical parameters for the main-chain and side-chain atoms are better than those obtained for protein structures refined to a similar resolution. The rms-coordinate error, estimated from a Luzzati plot, is 0.15 Å (Luzzati, 1952).

The side chains of Lys 58, 64, 84, 91, 94, 145 and of Gln 3, located on the surface of the protein, are not completely visible in the electron density map. There is no clear density for Asp 99. For the OG atoms of Ser 29 and Ser 115 as well as for the NZ atoms of Lys 39 and Lys 188, two alternate conformations were found. After refinement the occupancies of these alternate sites range from 0.4 to 0.6.

Table 3: Five Papaya Proteinase X-ray Structures under Investigation

compound	papain	papain	caricain	glycyl endopeptidase	chymopapain
pdb code	9pap	1ppn	1ppo	1gec	1yal
space group	*P*212121	*P*21	*P*3112	*P*31	*C*2
cell (Å)	45.2; 104.6; 50.9	65.7; 50.7; 31.5	74.1; 74.1; 77.8	55.8; 55.8; 64.4	145.2; 32.3; 47.4
(deg)	90; 90; 90	90; 98.4; 90	90; 90; 120	90; 90; 120	90; 98.4; 90
resolution (Å)	1.65	1.6	1.8	2.1	1.7
R factor	16.1	16	15.5	19.6	19.2
no. of protein atoms	1652	1659	1641	1642	1667
no. of waters	195	226	133	117	222
Cys 25	O_3H	SCH_2CH_2OH[a]	Hg	CBZ-Leu-Val-GCM	SCH_3
others	29 CH_3OH	3 CH_3OH			SCH_3 at Cys 117

[a] In the monoclinic papain structure a mercaptoethanol group was attached to Cys 25. This group was refined as an oxygen atom (coordinates of this O atom are in the pdb file).

FIGURE 2: Chymopapain structure; α-helices are shown as helical ribbons and β-strands as arrows. The side chains of the residues belonging to the catalytic triad are shown as ball and stick.

The average temperature factors of the main-chain atoms and of the side-chain atoms along the polypeptide chain display essentially the same pattern of variation. The highest values are found in the loop extending from Ala 97 to Pro 113 and at both ends of the protein chain, with the C-terminal end having lower average temperature factors (of about 20 Å^2) than the N-terminal end (avrage temperature factors of about 30 Å^2). This is due to the multiple intermolecular contacts of residues Lys 215, Gly 216, Phe 217, and Ala 218, tabulated in Table 2.

The final number of non-protein atoms consists of 222 water molecules and the two thiomethyl (SCH_3) groups bound to the two free cysteines Cys 25 and Cys 117.

A number of solvent molecules are seen in the vicinity of the active site. Two water clusters can be distinguished. The first group comprises water molecules 21, 23, 24, 25, 26, 27, 46, and 47 (indicated as black circles in Figure 3). These eight molecules are buried in the interface between the two domains of the protein; each one is hydrogen bonded to either one of the other water molecules of the group and/or to a protein atom (belonging to residue Ala 12, Thr 14, Lys 17, Phe 28, Ala 32, Glu 35, Glu 50, Ala 160, Thr 162, Lys 178, Asn 179, or Ser 180). They all have low temperature factors ranging from 9.6 to 14.7 Å^2 with an average value of 11.6 Å^2. They may be considered as an integral part of the protein.

Water molecules 37, 42, 43, 44, 109, 110, and 146 form the second network, located in the active site cleft. They have higher temperature factors than the members of the first group (values ranging from 16.4 to 52.7 Å^2 with an average value of 32.2 Å^2). Hydrogen bond are made with residues Gly 66, Gln 68, Thr 69, and Asp 158.

Two other water molecules 45 and 206, relatively isolated (not hydrogen bonded to any other water molecule), are found in the interface between the two domains. Water molecule 45 is hydrogen bonded to Trp 26 O, Ser 29 OG (alternate conformation), Thr 30 OG1 of the L-domain, and to Ser 131 OG of the R-domain, while water molecule 206 connects Gln 19 NE2 in the L-domain with Trp 181 NE1 in the R-domain.

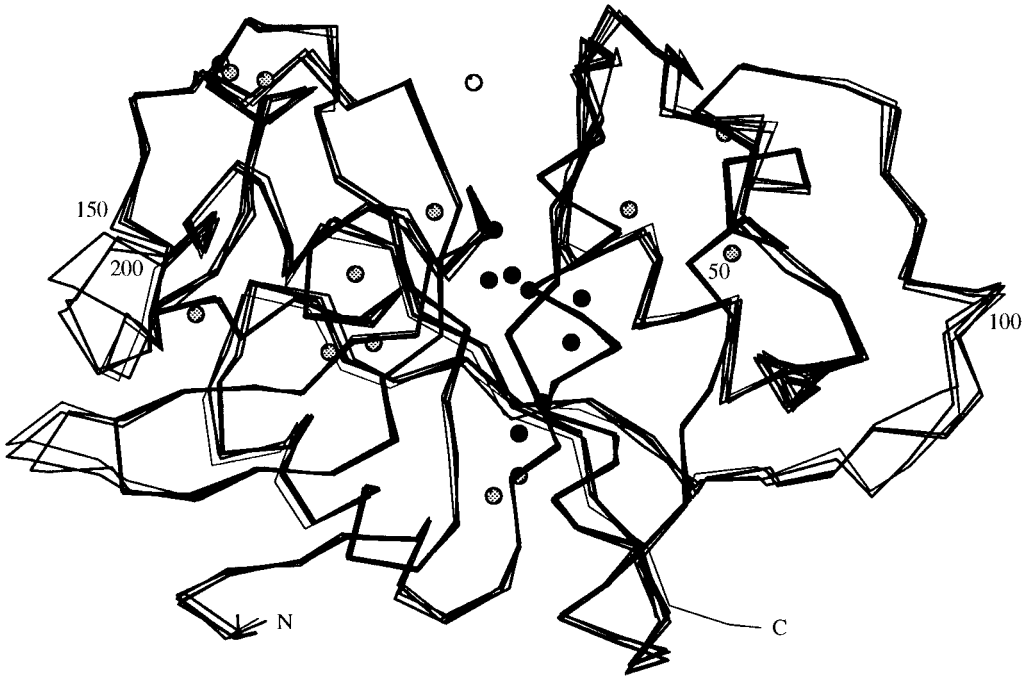

FIGURE 3: C_α-trace superposition of chymopapain, monoclinic and orthorhombic papain, caricain, and glycyl endopeptidase. The 22 conserved water sites are shown as circles (black, water sites belonging to the conserved water cluster, comprising water sites 21, 23−27, 46, and 47; white, water site 206; grey, others).

Comparative Study

(A) Alignment of the sequence of the four papaya proteinases reveals that papain has one deletion of four residues (169−172, chymopapain numbering) in a loop connecting two antiparallel β-strands. Chymopapain is the sole proteinase having two additional residues Phe 217 and Ala 218 at the C-terminal end. In Table 4 the number of identical residues for each pair of papaya proteinases is given.

All cysteine residues are homologous, except for an additional cysteine located at position 117 on the protein surface of chymopapain.

(B) Structural Comparison. Table 4 tabulates the rms deviation between the C_α atoms for each pair of papaya protease structures. Dihedrals φ and ψ are also compared. It turns out that although chymopapain has the highest sequence homology with glycyl endopeptidase, its structure is most similar to that of caricain. The major backbone differences between chymopapain and the other four X-ray structures are located in two loop regions, at the protein surface, containing residues 102 and 103 on the one hand and 197, 198, and 199 on the other hand. Omitting the residues of these regions, with a C_α deviation larger than 3.8 Å, results in an improved fit with a rms difference of about 0.6 Å for all pairs of structures. Figure 3 shows this optimal superposition of the C_α trace for the five structures under investigation.

The above mentioned loops are flexible with very high average temperature factors of the main chain atoms. An exception is the loop extending from residue 196 to 200 in chymopapain and caricain. In chymopapain the average temperature factors of the main-chain atoms in this loop are below the average values, as a result of the numerous packing contacts. Furthermore, some atoms belonging to this loop are hydrogen bonded to residues Asn 116 and Glu 118 (Gly 198 O to Asn 118 N; Gly 198 N to Asn 118 OE1; Ser 200 O to Asn 116 O; Ser 200 N to Asn 116 O). In caricain the main chain of the loop follows the same trajectory as in chymopapain. The average temperature factors of the main-chain atoms in this loop in caricain are somewhat higher than for the rest of the protein but not as high as in papain or glycyl endopeptidase. The loop is not involved in packing contacts but the same hydrogen bonding network as mentioned above does exist. An additional hydrogen bond is formed between Asn 117 ND2 and Asn 199 OD1. This hydrogen bond does not exist in chymopapain due to the presence of the protecting group on Cys 117. This thiomethyl group pushes the side chain of Asn 199 outward, allowing the OD1 atom to form a packing contact with 194 Lys NZ. One may speculate that this could well be one important reason for the formation of the high-quality crystals, obtained when fully protected chymopapain solutions are used. Previously by using a chymopapain preparation, in which the Cys 117 did exist as a mixture of sulfonic acid and thiomethyl derivatives, crystals diffracting up to only a resolution of 2.4 Å could be obtained.

(C) Solvent Structure. For each pair of X-ray structures under investigation the number of conserved water sites is tabulated in Table 4. Twenty-two water sites, making similar hydrogen bonds, were conserved in all five structures. They all have low temperature factors (average value of 17.8 $Å^2$ in the chymopapain structure). They are drawn on Figure 3.

The majority of them are found in the interface between the two domains. The first water network of eight waters, mentioned above and displayed as black circles on Figure 3, is fully conserved. This is the well-documented water cluster found in all cysteine protease structures solved up to date. The second network in the active site cleft of chymopapain is not found in any of the other four X-ray structures. It was pointed out by Rullmann et al. (1989) that structural fluctuations in this region resulted in only minor energy fluctuations, accounting for the variation of water sites

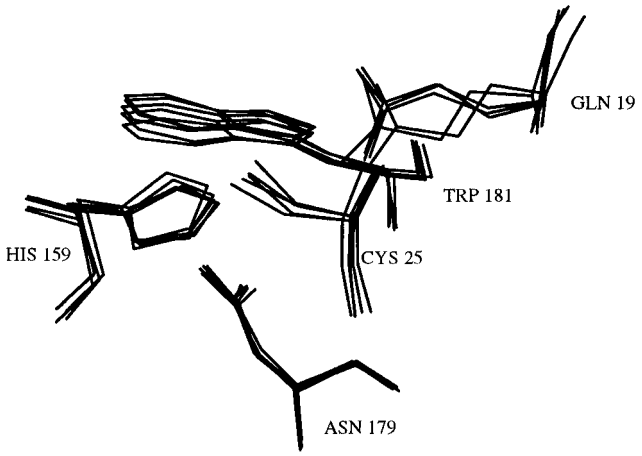

FIGURE 4: Superposition of the active site residues for the five papaya protease X-ray structures under investigation.

among the structures and for the high temperature factors of the solvent molecules in this region.

Water molecule 206 (white circle on Figure 3) is conserved, despite its high temperature factor of 59.0 Å^2. On the other hand, water molecule 45 of chymopapain is only found in the glycyl endopeptidase structure. The substitution of Gln 68 to Pro 68 in papain and caricain accounts for the fact that the main chain is displaced by about 1 Å in this region. As a result no space is left for a solvent molecule at this position in papain and caricain.

In chymopapain position 107 is occupied by Ile while Thr is found in the other three papaya proteases. Due to this substitution the CD1 atom of Ile 107 partly occupies the water site at an hydrogen bonding distance from Leu 72 O, Thr 107 OG1, and Gly 109 O, found in papain, caricain, and glycyl endopeptidase.

One might argue that the lower resolution of the glycyl endopeptidase X-ray structure as well as the covalent bonding of an inhibitor to the active site center Cys 25 accounts for the fact that fewer water molecules were included in the structure of this proteinase and makes any comparison of water sites questionable. If one only considers the structures of chymopapain, papain, and caricain, 29 water sites instead of the 22 mentioned before are conserved. The additional conserved water sites (8, 19, 39, 51, 61, 124, and 137) do not occupy the position of the inhibitor in glycyl endopeptidase. They are all located at the protein surface, except for water molecule 61, which is hydrogen bonded to residues 35, 44, and 46 in the structures under investigation.

(D) Active Site and Specificity Subsites. Residues near the active site and those actually involved in the catalytic process (Cys 25, His 159, and Asn 179) superimpose to an extent that approaches the atomic coordinate accuracy (Figure 4). Changes do occur in residue Gln 19. It has two conformations: one is found in papain and chymopapain and the other in caricain and glycyl endopeptidase. The S1 subsite, formed by the catalytic Cys 25 and the conserved Gly 23 and Gly 65, is a wide pocket, which exerts relatively little influence on the substrate specificity for all cystein proteases. There is one exception. In glycyl endopeptidase, where Gly 23 is replaced by Glu, this site is glycine specific.

Table 4: Comparison of the Five Papaya Protease X-ray Structures[a]

	9pap	1ppn	1ppo	1gec	chym
9pap		*212 [87]*	*146 [51]*	*144 [41]*	*126 [66]*
1ppn	**0.29**		*146 [55]*	*144 [41]*	*126 [67]*
1ppo	**1.02 (0.61)**	**1.07 (0.63)**		*175 [30]*	*143 [51]*
1gec	**0.64**	**0.66**	**1.09 (0.54)**		*151 [47]*
chym	**1.01 (0.63)**	**1.07 (0.69)**	**0.56**	**1.08 (0.56)**	

[a] Italic type: number of residues in common after alignment. The number of conserved water sites is between square brackets. Bold type: root mean square difference for all the C_α atoms. Root mean square difference for the C_α atoms differing by less than 3.8 Å is between brackets.

The substrate preferences shown by the members of the papain family are generally believed to be dictated by the S2 subsite (Drenth, 1976). In papain, the S2 subsite is essentially composed of residues 67, 68, 69, 133, 157, and 207. Chymopapain is the sole member among the papaya proteases in which the highly conserved Val 133 and Val 157 are substituted. Leucine residues occupy these positions in chymopapain, and it was generally expected that these changes would profoundly modify the geometry of subsite S2 and accordingly the specificity of this proteinase (Jacquet, 1990; Topham et al., 1990). These predictions were based on a model (see Figure 5) in which the C_α trace of chymopapain was superimposed on that of papain and the conformations of Leu 133 and Leu 157 side chains were calculated in order to minimize interactions with adjacent atoms (Jacquet, 1990). Comparison of the model of chymopapain with the X-ray structure reveals that the side chains of Leu 133 and Leu 157 do not have the predicted conformation. For Leu 133 the largest discrepancy is found for the χ_2 angle having values of 79° and −172° for the predicted and the X-ray structure, respectively. For Leu 157 on the other hand the χ_1 angle has values of 50° and −46°, respectively. This means that in the X-ray structure of chymopapain the side chain of Leu 157 is "turned away" from the S2 pocket and does not modify the accessibility of this subsite as compared to the valines in the three other proteases. The position of the CG atom of Leu 133 is close

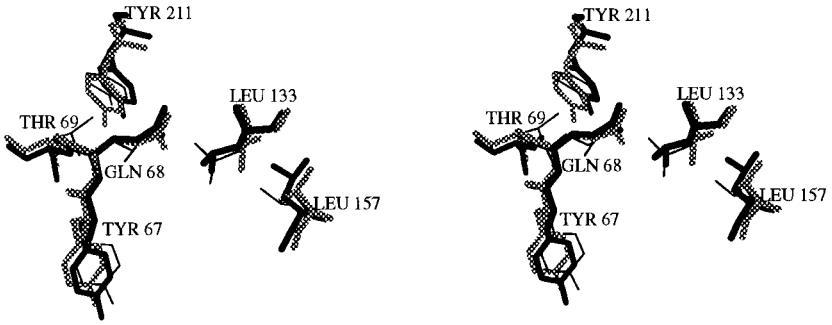

FIGURE 5: Stereopair of the S2 subsite for chymopapain (thick black line) and the predicted model (thin black line). For reason of comparison, the S2 subsite of glycyl endopeptidase (thick grey line) is also included.

to the CG1 atom of Val 133 in papain. The CD1 and CD2 atoms are protruding in the S2 pocket. To have an idea of the consequences on the specificity of this site a single phenyl ring and the chloromethyl ketone inhibitor AcAAPACK containing a phenylalanine residue (Drenth et al., 1976) was modeled into the binding site of chymopapain. For both molecules an energy minimum was found when the phenyl ring was positioned in the S2 subsite.

CONCLUSION

The X-ray structure of chymopapain was determined and compared with the X-ray structures of papain, caricain and glycyl endopeptidase. Chymopapain is the last among the four papaya cysteine proteinases for which one obtained high-quality crystals suitable for the accurate determination of its three-dimensional structure. The structure of chymopapain is extremely similar to those of the three other papaya proteinases including the same water network at the interface between the two domains. Differences in backbone conformation were found only for two loops at the surface of the protein, far away from the active site cleft. As a consequence of this similarity, one may conclude as suggested by Topham, that the structure, determined here, corresponds to the expected structure for chymopapain B (Topham et al., 1990).

While the existence of another form of chymopapain called A has never been in doubt, few reliable experimental data do actually document its physico-chemical and its catalytic properties (Buttle et al., 1989). Taking into account that both forms have identical amino acid sequences (Jacquet et al., 1989a; Watson et al., 1990) and are immunologically indistinguishable (Buttle & Barrett, 1984), one may speculate that only subtle chemical and/or conformational differences do differentiate these forms. Unfortunately, the three-dimensional structure reported here does not afford any further insight into this question. Altogether it seems thus reasonable presently to regard chymopapain as a unique enzyme, the structure of which being reported here.

ACKNOWLEDGMENT

We are indebted to Henri Moereels from the Center for Molecular Design of Janssen Research Foundation who kindly performed the energy minimization calculations.

REFERENCES

Azarkan, M., Maes, D., Bouckaert, J., Dao Thi, M. H., Wyns, L., & Looze, Y. (1996) *J. Chromatogr*. (in press).

Baines, B. S., Brocklehurst, K., Carey, P. R., Jarvis, M., Salih, E., & Storer, A. C. (1986) *Biochem. J. 233*, 119.

Baker, E. N. (1980) *J. Mol. Biol. 141*, 441.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. J., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535.

Brunger, A. T. (1992) in *X-PLOR Version 3.1: A System for X-Ray Crystallography and NMR*, Yale University Press, New Haven, CT.

Buttle, D. J., & Barrett, A. J. (1984) *Biochem. J. 223*, 81.

Buttle, D. J., Kembhavi, A. A., Sharp, S. L., Shute, R. E., Rich, D. H., & Barrett, A. J. (1989) *Biochem. J. 261*, 469.

Cohen, L. W., Coghlan, V. C., & Dihel, L. C. (1986) *Gene 48*, 219.

Collaborative Computational Project No. 4 (CCP4) (1994) *Acta Crystallogr. A D50*, 760.

Cygler, M., Sivaraman, J., Grouchulski, P., Coulombe, R., Storer, A. C., & Mort, J. S. (1996) *Structure 4*, 405.

Drenth, J., Jansonius, J. N., Koekoek, R., Marrink, J., Munnik, J., & Wolthers, B. G. (1962) *J. Mol. Biol. 5*, 398.

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

Jacquet, A. (1990) Ph.D. Thesis. Université Libre de Bruxelles.

Jacquet, A., Kleinschmidt, T., Schnek, A. G., Looze, Y., & Braunitzer, G. (1989a) *Biol. Chem. Hoppe-Seyler 370*, 425.

Jacquet, A., Dubois, T., Kleinschmidt, T., Schnek, A. G., Looze, Y., & Braunitzer, G. (1989B) *Biol. Chem. Hoppe-Seyler 370*, 819.

Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R., & Huber, C. P. (1995) *J. Biol. Chem. 270*, 5527.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjelgaard, M. (1991) *Acta Crystallogr. A47*, 110.

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

Karrer, K. M., Peiffer, S. L., & Di Tomas, M. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3063.

Khan, I., & Polgar, L. (1983) *Biochem. Biophys. Acta 760*, 350.

Kleywegt, G. J., & Jones, T. A. (1994) *ESF/CCP4 Newsl. 31*, 9.

Korodi, I., Asboth, B., & Polgar, L. (1986) *Biochemistry 25*, 6895.

Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946.

Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283.

Lisgarten, J. N., Gupta, V., Maes, D., Wyns, L., Zegers, I., Palmer, R. A., Dealwis, C. G., & Hemmings, A. M. (1993) *Acta Crystallogr. D49*, 541.

Luzatti, P. V. (1952) *Acta Crystallogr. 5*, 802.

Mc Grath, M. E., Eakin, A. E., Engel, J. C., Mc Kerrow, J. H., Craik, C. S., & Fletterick, R. J. (1995) *J. Mol. Biol. 247*, 251.

Mc Kee, R. A., Adams, S., Matthews, J. A., Smith, C. J., & Smith, H. (1986) *Biochem. J. 237*, 105.

Messerschmidt, A., & Pflugrath, J. W. (1987) *J. Appl. Crystallogr. 20*, 306.

Musil, D., Zucic, D., Turk, D., Engh, R. A., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katanuma, N., & Bode, W. (1991) *EMBO J. 10*, 2321.

O'Hara, B. P., Hemmings, A. M., Buttle, D. J., & Pearl, L. H. (1995) *Biochemistry 34*, 13190.

Pickersgill, R. W., Rizkallah, P., Harris, G. W., & Goodenough, P. W. (1991) *Acta Crystallogr. B47*, 766.

Pickersgill, R. W., Harris, G. W., & Garman, E. (1992) *Acta Crystallogr. B48*, 59.

Rawlings, N. D., & Barrett, A. J. (1994) *Methods Enzymol. 244*, 461.

Revell, D. J., Cummings, N. J., Baker, K. C., Collins, M. E., Taylor, M. A. J., Summer, I. G., Pickersgill, R. W., Connerton, I. F., & Goodenough, P. W. (1993) *Gene 127*, 221.

Rullmann, J. A. C., Bellido, M. N., & Van Duijnen, P. Th. (1989) *J. Mol. Biol. 206*, 101.

Solis-Mendiola, S., Zubillaga-Luna, R., Rojo-Dominguez, A., & Hernandez-Arana, A. (1989) *Biochem. J. 257*, 183.

Summer, I. G., Harris, G. W., Taylor, M. A. J., Pickersgill, R. W., Owen, A. J., & Goodenough, P. W. (1993) *Eur. J. Biochem. 214*, 129.

Taylor, M. A. J., Baker, K. C., Briggs, C. S., Connerton, I. F., Cummings, N. J., Pratt, K. A., Revell, D. F., Freedman, R. B., & Goodenough, P. W. (1995) *Protein Eng. 8*, 59.

Topham, C. M., Overington, J., O'Driscoll, M., Salik, E., Thomas, M., Thomas, E. W., & Brocklehurst, K. (1990) *Biochem. Soc. Trans. 18*, 933.

Vernet, T., Berti, P. J., Demontigny, C., Musil, R., Tessler, D. C., Menard, R., Magny, M. C., Storer, A. C., & Thomas, D. Y. (1995) *J. Biol. Chem. 270*, 10838.

Walraevens, V., Jaziri, M., Van Beeumen, J., Schnek, A., Kleinschmidt, T., & Looze, Y. (1993) *Biol. Chem. Hoppe-Seyler 374*, 501.

Watson, D. C., Yaguchi, M., & Lynn, K. R. (1990) *Biochem. J. 266*, 75.

BI961491W