The 2.1 Å Structure of a Cysteine Protease with Proline Specificity from Ginger Rhizome, *Zingiber officinale*[†,‡]

Kyung H. Choi,[§,||] Richard A. Laursen,[§] and Karen N. Allen*[,⊥]

Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, and Structural Biology Group, Department of Physiology, Boston University School of Medicine, 80 East Concord Street, Boston, Massachusetts 02118-2394

Received March 19, 1999; Revised Manuscript Received July 7, 1999

ABSTRACT: A cysteine protease from ginger rhizome (GP-II) cleaves peptides and proteins with proline at the P_2 position. The unusual specificity for proline makes GP-II an attractive tool for protein sequencing and identification of stably folded domains in proteins. The enzyme is a 221 amino acid glycoprotein possessing two N-linked oligosaccharide chains (8% glycosylated by weight) at Asn99 and Asn156. The availability of the sequence of these glycosyl chains afforded the opportunity to observe their structure and impact on protein conformation. The three-dimensional structure of GP-II has been determined by X-ray crystallography to a resolution of 2.1 Å (overall *R*-factor = 0.214, free $R = 0.248$). The overall structure of GP-II is similar to that of the homologous cysteine proteases papain, actinidin, and glycyl endopeptidase, folding into two distinct domains of roughly equal size which are divided by a cleft. The observed N-linked glycosyl chains (half the total carbohydrate sequence) participate in both crystallographic and noncrystallographic contacts, tethering the proteins together via hydrogen bonds to the carbohydrate residues without intervening ordered water molecules. The putative S_2 binding pocket (the proline recognition site) was identified by superposition of the GP-II structure with structures of four previously determined papain−inhibitor complexes. The particular enzymic amino acids forming the S_2 pocket of GP-II (Trp, Met, and Ala) are similar to those found in the proline binding pockets of the unrelated enzymes α-lytic protease and cyclophilin. However, there is no conserved three-dimensional arrangement of these residues between the three enzymes (i.e., no proline binding motif). Thus, the particular amino acids found at S_2 are consistent with a binding pocket for a moiety with the steric characteristics and charge distribution of proline. Size exclusion is also a mechanism for selectivity compared to the S_2 binding pocket of papain. The S_2 binding pocket of GP-II greatly restricts the size of the side chain which could be bound because of the occurrence of a tryptophan in place of the corresponding tyrosine in papain. In light of the nature of the binding pocket, the specificity of GP-II for proline over other small nonpolar amino acids may be attributed to a direct effect of proline on the substrate peptide backbone conformation.

Proteases from ginger rhizome, *Zingiber officinale*, were first reported by Ichikawa et al. in 1973 (*1*) and shown to have a high proteolytic activity against undefined substrates such as casein, bovine serum albumin, and collagen (*2*, *3*). These proteases have been separated by DEAE-cellulose chromatography into two fractions, namely ginger protease I (GP-I)[1] and ginger protease II (GP-II). Recently, one of our laboratories determined the complete amino acid sequences of both the GP-I and GP-II isoforms, which were shown to possess 83% sequence identity, by solid-phase peptide sequencing using Edman chemistry (Protein Information Resource accession codes A59040 and A59041, respectively). The ginger proteases are inactivated by sulfhydryl reagents (*1*) and thus are grouped with other plant cysteine proteases, such as papain, actinidin, and stem bromelain, and with similar enzymes from nonplant sources, such as cathepsin B.

Three-dimensional structures have been elucidated for papain (*4*), actinidin (*5*), proteinase Ω (*6*), and glycyl endopeptidase (*7*). Comparison of these cysteine proteases reveals similar structures in that the polypeptide chain is folded into two distinct domains and the active site is located at the same position in a cleft between the domains. The cysteine proteases of the papain family exhibit a wide variety of activities accommodated by this common structure, including endopeptidases with broad specificity such as papain (*8*), endopeptidases with very narrow specificity such

[†] K.N.A. is funded by a grant from the National Science Foundation (MCB 9630430).

[‡] The coordinates of the refined structure have been deposited with the Brookhaven National Laboratories Protein Data Bank, accession code 1cqd.

* Correspondence should be addressed to K.N.A. E-mail: allen@med-xtal.bu.edu.

[§] Boston University.

[||] Present address: Structural Biology Group, Department of Physiology, Boston University School of Medicine, 80 East Concord St., Boston, MA 02118-2394.

[⊥] Boston University School of Medicine.

[1] Abbreviations: GP-II, ginger protease II; GP-I, ginger protease I; NCS, noncrystallographic symmetry; pNA, *p*-nitroanilide; ES-MS, electrospray mass spectrometry; Man, mannose; Xyl, xylose; Fuc, fucose; GlcNAc, *N*-acetyl-D-glucosamine; N-GlcNAc, *N*-acetyl-D-glucosamine N-linked to Asn; ES-CID, electrospray collision-induced dissociation.

10.1021/bi990651b CCC: $18.00
Published on Web 08/17/1999

as glycyl endopeptidase (*9*), aminopeptidases such as cysteine aminopeptidase (*10*), peptidases with both endopeptidase and exopeptidase activities such as cathepsins B and H (*11*), and the dipeptidyl peptidase, dipeptidyl peptidase I (*12*).

In the case of papain the nature of the residue at the P_2 position has been shown to be the most significant in terms of determining specificity toward peptide substrates, and it is well established that papain exhibits a strong preference for amino acids with aromatic side chains such as Phe and Tyr at this position (*13*). As in papain, the primary determinant of specificity for the ginger proteases is also the residue at the P_2 position. GP-I and GP-II catalyzed hydrolysis of a large number of protein and peptide substrates shows that the predominant determinant for specificity is a proline at the P_2 position, with little preference for residues at P_1 (*14*). This conclusion is confirmed by kinetic analysis of GP-II with some 20 tripeptide *p*-nitroanilide type substrates which show that k_{cat}/K_m values for substrates with P_2 proline are 100−3000 times greater than those with other amino acids at P_2 (*15*). These studies also indicate that there may be some preference for ionic residues (e.g., Lys and Glu) at P_1. The primary specificity of GP-I and GP-II for proline at P_2 can be contrasted with the well-known preference for proline at P_2 exhibited by the serine protease thrombin. In the case of thrombin, the primary determinant of specificity has been demonstrated to be Arg or Lys at P_1 by identifying the cleavage sites of protein (*16*) and peptide substrates (*17*). The preference for proline at P_2 in thrombin is seen as an increase of several-hundredfold in k_{cat}/K_m in peptide substrates with Pro at P_2 (*18*) as well as an increase in k_{on} of 1000-fold for serpin inhibitors with Pro compared to Thr at P_2 (*19*).

Proteases are useful probes of protein conformation and are often used as a tool to identify the most stably folded protein core (*20*). Native proteins are rather resistant to proteases due to the tightly folded structures or domains adapted by most globular proteins. Thus the parts of the proteins which are accessible to proteases are the exposed loops or the linking regions of the domains. Since proline frequently occurs in exposed loops and turns in proteins, ginger protease may prove to be a useful tool for identifying turns and for generating large fragments in protein structure and conformation studies. Since it is of interest to target proline, it would be useful to understand the elements which control proline recognition and binding, thus allowing protein engineering of proline binding sites. The present X-ray crystallographic study was undertaken with the aim of explaining the specificity of GP-II toward proline-containing peptides and to compare its structure with those of other cysteine proteases. The structure of GP-II has been solved by molecular replacement to 2.1 Å resolution and the active site identified. The proline binding subsite of GP-II is made up of amino acids similar to those comprising the proline binding pockets of the unrelated proteins cyclophilin and α-lytic protease.

EXPERIMENTAL PROCEDURES

Crystallization. GP-II was purified according to the protocol previously described (*1*) with slight modification as follows. A linear gradient of 0−0.25 M NaCl in 20 mM sodium phosphate buffer, pH 7.0, was used for DEAE-cellulose ion-exchange chromatography. Before crystallization, the GP-II sample was exchanged into low ionic strength buffer, 1 mM HEPES−HCl, pH 7.4, containing 5 mM sodium tetrathionate as a protecting reagent for free cysteines, and concentrated to ∼15 mg·mL^{-1} as determined by a Bradford protein assay. The initial crystallization conditions were found by the sparse factorial matrix (*21*) method using Crystal Screen I and II (Hampton Research). Crystals were grown for data collection at 18 °C using the vapor diffusion method with hanging-drop geometry in which the 0.5 mL well solution was 26% poly(ethylene glycol) monomethyl ether 2000, 0.2 M ammonium sulfate, and 0.1 M sodium acetate buffered to pH 5.2. The protein solution (5 μL) was mixed with an equal volume of the well solution. After the hanging drops had equilibrated (24 h), each drop was seeded with 1 μL of seeding mixture. The seeding mixture was made by diluting (by 10-fold) the mother liquor from a drop containing crystals into the initial well solution plus 5% poly(ethylene glycol) monomethyl ether 2000. The crystals (which are plates) grew to a size of 0.7 × 0.7 × 0.05 mm in dimension in 1−2 weeks.

Enzyme Activity in Dissolved Crystals. To determine if the active site cysteine of GP-II was covalently modified by a thiosulfate protecting group, the activity of the crystals was measured in the presence and absence of DTT. A few GP-II crystals were removed from a small drop into an Eppendorf tube and washed twice with the well solution (which contains no DTT). The well solution was then decanted, and the crystals were dissolved in an assay buffer (100 μL) composed of 50 mM sodium phosphate buffer, pH 7.0. From this stock enzyme solution, aliquots (10 μL) were removed and mixed with substrate (1 mM Gly-Pro-Ala-*p*-nitroanilide) dissolved in assay buffer (total 1 mL). The reaction was monitored by UV/vis spectrophotometry on a Hewlett-Packard 8452A diode array spectrophotometer at a wavelength of 405 nm at 25 °C. No activity was observed in the absence of DTT, even using long (30 min) incubation times. The addition of 2 mM DTT to the reaction buffer, which regenerates the active site sulfhydryl group, caused the immediate appearance of an increasing absorption signal with respect to time at 405 nm. From this experiment it is clear that the active site cysteine in the crystalline enzyme is covalently bound to a thiosulfate group.

Data Collection. Crystals were grown as described above and soaked for >1 h in a cryosolvent consisting of the well solution plus 15% 2-methyl-2,4-pentanediol. The crystals were flash-frozen in a stream of nitrogen gas cooled to −180 °C by liquid nitrogen and maintained at that temperature throughout data collection. Data to 2.1 Å resolution were collected on a Rigaku rotating anode X-ray source (50 mA, 100 kV) equipped with an RaxisII image plate. Data were collected using 1° oscillations and a crystal-to-detector distance of 110 mm. These data were processed and scaled with DENZO (*22*) and SCALEPACK (*22*), respectively. Data collection and processing statistics are listed in Table 1. The GP-II crystals are monoclinic, crystallizing in the space group $P2_1$ with unit cell dimensions $a = 94.0$ Å, $b = 45.5$ Å, $c = 110.0$ Å, and $\beta = 105.0°$. The number of GP-II monomers per asymmetric unit was calculated to be 4, using a solvent content of 50% for the crystals.

Molecular Replacement. The structure was solved by molecular replacement with AMoRE (*23*). The search model

Table 1: Data Collection and Refinement Statistics

data collection	
resolution	25.0−2.1 Å
reflections	58449
completeness (overall/outer shell) (%)	99.7/94.9
R_{sym}[a]	0.191
I/σ	3.8
refinement statistics	
protein atoms (non-hydrogen)	864
no. of water molecules	375
no. of thiosulfate molecules	4
no. of carbohydrate molecules	12 NAG and 4 fucose
resolution	2.1
F/σ cutoff	0
R-factor	0.214
free *R*-factor[b]	0.248
mean *B*-factors (main chain and side chain, Å^2)	13.5
RMS deviation	
bonds	0.005
angles	1.060
dihedrals	24.80
impropers	0.527

[a] $R = \sum((I - \langle I \rangle)^2)/\sum I^2$. [b] The free *R*-factor (*26*) was calculated from a random selection constituting 10% of the data.

was derived from three currently known high-resolution crystal structures of the homologous enzymes actinidin, papain, and protease Ω using the Swiss-Model Protein Modeling server (*24*). The enzyme monomer thus obtained was used as the search model with data in the resolution range 10−4 Å and 2.0 F/σ intensity cutoff. Once the rotation and translation searches were completed for the first monomer, a second translation search with two molecules was performed, fixing the solution with the lowest *R*-factor from the first translation search. This procedure was continued until all four monomers were found. Rigid body fitting within AMoRE of the final solution yielded an *R*-factor of 0.400.

Refinement. Crystallographic refinement and model building were performed using alternating cycles of X-PLOR refinement and manual refitting of the model to the electron density using the program O (*25*). Ten percent of all reflections were excluded from the refinement for the calculation of the free *R*-factor during the refinement. Refinement of the tetramer as a rigid body was carried out, followed by refinement of each monomer as a separate rigid body in X-PLOR (*26*) over the resolution range 8.0−3.0 Å. Since the free *R*-factor did not decrease in value after subsequent positional refinement, it was necessary to remove the amino acid side chains in the refinement model. The regions of the model possessing low sequence identity, or containing insertions or deletions in the sequence alignment with actinidin, papain, and protease Ω, were replaced by alanine residues (a total of 61 residues out of 216). This new model was then refined using strict noncrystallographic symmetry (NCS) in X-PLOR. Each round of refinement involved positional, preparational, simulated annealing (temperature = 2500 K) and *B*-group refinement. With the incorporation of a bulk solvent correction (*27*) in X-PLOR, data were included in the resolution range 25.0−2.1 Å with no F/σ cutoff. In the early stages of refinement, amino acid side chains were built into the electron density in the NCS averaged map calculated with the coefficients $2F_o - F_c$ contoured at 1.2σ. Building was also guided by $F_o - F_c$ maps with electron density contoured at 1.0σ and -1.0σ. The NCS averaged $2F_o - F_c$ map showed substantial errors in the regions of Ser100−Pro105 and Asn198−Pro202. These regions were rebuilt using a simulated annealing omit map in the next cycle of refinement. When building and refinement of the main chain were completed, the main chain was checked with a series of simulated annealing omit maps spanning the entire length of the protein (20 residues at a time).

Subsequently, the NCS strict was substituted by NCS restraints with a high (300 kcal mol^{-1} Å^{-2}) weight. Water molecules (375 in total) were added to the model using a $F_o - F_c$ map with electron density contoured at 3σ. The $2F_o - F_c$ maps also clearly defined parts of carbohydrate chains linked through the side chain amide groups of Asn99 and Asn156. The *N*-glycosyl chain from the structure of glycoamylase was used for modeling *N*-acetyl-D-glucosamines (GlcNAc), and the nanoate complex of L-fucose was used for fucose residues (Brookhaven Protein Data Bank entries 1agm and 1liv, respectively). Topology and parameter files of GlcNAc and fucose were generated using the xplor2d program in the X-UTIL package (*28*). In addition, a thiosulfate molecule in each of the four enzyme monomers was assigned to the electron density in the vicinity of Cys27 on the basis of the electron density in a $2F_o - F_c$ map contoured at 3σ. In the final stages of structure refinement, NCS restraints on some side chains (residue 66, 87, 92, 94, 97, 101, 106, 113, 125, 141, 145, 147, 156, 171, 172, 174, and 218), but not the main chain, were removed. Rebuilding was aided by the program OOPS (*29*). The final structure had an *R*-factor of 0.214 and a free *R*-factor of 0.248 using data in the resolution range 25.0−2.1 Å and no F/σ intensity cutoff. The final structure contains 375 water molecules, 4 thiosulfates, 12 GlcNAc rings, and 4 fucose rings. The geometric parameters of the model have been analyzed with PROCHECK and are all inside the range of expected deviations from ideality of other structures determined to 2.0 Å resolution.

RESULTS AND DISCUSSION

Structure Determination. GP-II crystallized in the space group $P2_1$ ($a = 94.0$ Å, $b = 45.5$ Å, $c = 110.0$ Å, and $\beta = 105.0°$) with four monomers per asymmetric unit. The crystal structure was determined by molecular replacement (Table 1) and refined to 2.1 Å resolution with an *R*-factor of 0.214 and a free *R*-factor of 0.248. A total of 375 water molecules were also refined with the protein structure, as well as a single covalently bound thiosulfate group at each active site cysteine (sodium tetrathionate was added as a protecting group for the nucleophilic cysteine before crystallization, as described in Experimental Procedures). The final model comprises residues 3−218, that is, the complete GP-II molecule except the two N-terminal and the three C-terminal amino acids, for which no electron density was observed. A Ramachandran plot of the final molecule shows that 88% of residues are in the "most favored" regions defined by the program PROCHECK (*30*), with the remaining 12% in the allowed regions and none in generally allowed or disallowed regions.

The structure is well-defined with an average *B*-factor of 13.5 Å^2 for both main chain and side chain atoms, and good side chain electron density is visible for all residues except

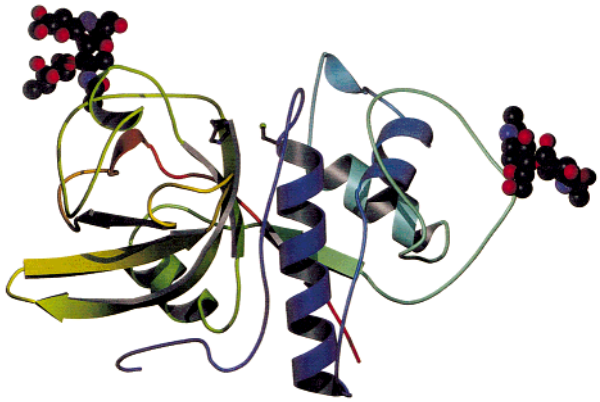

FIGURE 1: Ribbon diagram of the GP-II monomer. The molecule is colored from the N (blue) to the C terminus (red), and the carbohydrate moieties are depicted with space-filling models. The catalytic diad is depicted as a ball-and-stick model with gray bonds. This figure was rendered using the programs MOLSCRIPT (*48*) and POVRAY (*49*).

Table 2: Cα Comparison of GP-II with Other Cysteine Proteases

protein	rms deviation (Å)[a]	no. of common Cα atoms	sequence identity (%)
papain	1.4	212	47
protease Ω	1.6	216	48
actinidin	0.69	214	58
glycyl endopeptidase	1.4	216	50

[a] The comparison was performed using LSQ_EXPLICIT in the program O.

the terminal side chain atoms of Arg66, Gln94, and Lys218. Although the average temperature factors for the protein are very low, a plot of average temperature factor of each residue as a function of residue number reveals an increase in average *B*-factor to 27.1 for the main chain from residues 97−101. These residues lie in a surface loop containing one of the enzyme's two N-glycosylation sites. The *B*-factors for the glycosyl residues vary from 15 to 49 Å^2 and are lower for those moieties that participate in protein−protein contacts (see following section) and for those proximal to the attachment site of the carbohydrate chain.

The Cysteine Protease Family and Identification of the Active Site. The polypeptide chain of GP-II folds into two distinct domains of roughly equal size divided by a cleft, as depicted in Figure 1. Domain I (residues 13−112 and 215−218) consists primarily of residues from the first half of the chain and domain II (residues 3−12 and 113−214) primarily of residues from the second half. The N- and C-terminal residues cross over into each domain and help to hold the two domains together. Domain I is mainly α-helical and domain II is an antiparallel β-sheet structure. The GP-II protein fold is also stabilized by three disulfide bonds between Cys24 and Cys65, Cys58 and Cys98, and Cys155 and Cys206. These residues are strictly conserved throughout the papain family (*10*).

As expected from the sequence identity between them, the overall structure of GP-II is similar to that of the other members of the papain family. This fact is demonstrated by the low rms deviation for the common Cα atoms (Table 2). In comparison, the average rms deviation between four different structures of papain is 0.51 Å. When compared to papain, the major differences in the Cα trace between the two structures are the conformations of the surface loops comprising residues 98−109, 156−159 (both of which contain a glycosylation site in GP-II), and 197−201. There is also an insertion of four extra residues (171−174) in a turn in domain II of GP-II. GP-II is similar to the other members of the cysteine protease family, including actinidin, glycyl endopeptidase, and protease omega, in that it contains this insertion. In fact, papain is the only family member which does not contain the loop.

The surface electrostatic potential of GP-II reveals polar residues concentrated on one side of the protein. This charged face extends over both domains. The opposite side of the GP-II molecule which surrounds the substrate binding pocket is fairly neutral. This neutral face extends over a 10 Å radius around the active site cysteine. Such a neutral surface surrounding the active site pocket may be advantageous for making protein−protein interactions with substrates of varied amino acid makeup. The active site lies in a 5.5 Å deep and 9.5 Å long cleft at the interface of domains I and II. Cys27 is the only free sulfhydryl group as determined by titration of the protein with 5,5′-dithiobis(2-nitrobenzoic acid) (*14*, *31*). On the opposite side of the active site cleft from Cys27 is His161, which we propose from the primary sequence alignment as well as the structural superposition with papain to be the group in the protease mechanism which polarizes the cysteine thiol enabling deprotonation. Figure 2 depicts the amino acid residues and hydrogen-bonding network in the active site of GP-II. Cys27 is contributed by domain I while His161 resides in domain II. His161 is oriented such that ND1 is 3.8 Å from the sulfur atom of Cys27, while the other nitrogen atom, NE2, is within hydrogen-bonding distance (2.8 Å) of the carbonyl group of the side chain of Asn181. The function of Asn181 is most probably to correctly orient His161 with respect to the nucleophilic Cys27 in order to stabilize the thiolate/imidazolium pair as in papain (*32*). In papain, it has been proposed that, upon acylation of the enzyme, rotation of this histidine imidazole allows protonation of the leaving amine (*33*). The electron density corresponding to a thiosulfate protecting group is observed connected to Cys27 (represented as a ball-and-stick model in Figure 2). It persists at an electron density cutoff of 3σ in a $F_o - F_c$ map, as does the active site Cys27 side chain sulfur. The thiosulfate has an average *B*-factor of 44.5 Å^2. The presence of a covalently bound protecting group was confirmed by demonstrating that the dissolved GP-II crystal regains enzymatic activity against the substrate Gly-Pro-Ala-*p*-nitroanilide (Gly-Pro-Ala-pNA) only in the presence of DTT (see Experimental Procedures). As depicted in Figure 2 the sulfate portion of the thiosulfate group binds in the oxyanion hole, making contacts with the backbone NH of Cys27 and to the amide NH of the Gln21 side chain. The relative positions of the catalytic diad Cys27 and His161 as well as Asn181 and the residues comprising the oxyanion hole superimpose in the structure of GP-II and both native papain (cysteine sulfur oxidized) and inhibited papain (PDB accession codes 1ppn, 9pap, 1pad, 1pe6, and 1pop). Thus there is no change in position of any of these residues relative to one another in liganded and unliganded active sites of GP-II and papain.

Participation of N-Linked Glycosyl Chains in the Protein Interface. Like the other proteins in the cysteine protease family, GP-II is a monomer in solution as demonstrated by

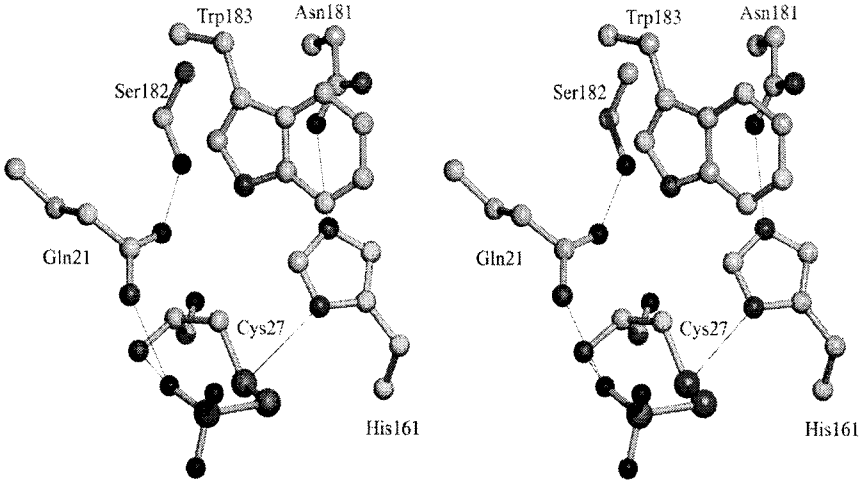

FIGURE 2: Stereo diagram of the amino acids and hydrogen-bonding network in the active site of GP-II. All molecules are depicted as ball-and-stick models. All amino acids are labeled. Hydrogen bonds described in the text are denoted by thin black lines. The thiosulfate protecting group forms a covalent bond to the Cys27 sulfhydryl group. This figure was rendered using the programs MOLSCRIPT and POVRAY.

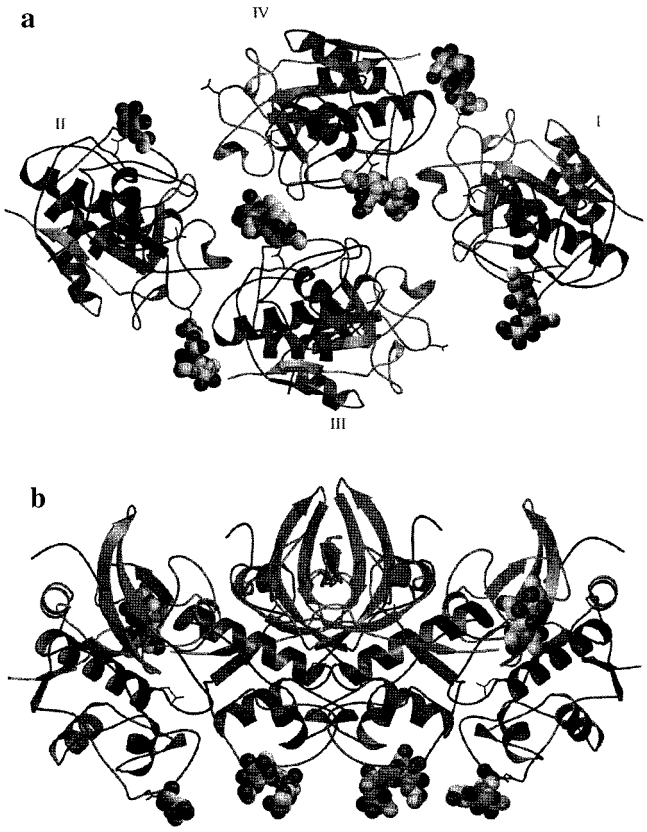

FIGURE 3: Ribbon diagram of the GP-II tetramer. The carbohydrate moieties are depicted with space-filling models. The thiosulfate protecting group covalently bonded to the catalytic cysteine and Asn residues demonstrated to be glycosylated are depicted as bonds. The top view (a) shows that the protein−protein interface is not extensive. Hydrogen bonds formed between carbohydrate residues and protein monomers account for a number of the intra tetramer interactions. The bottom view (b) is rotated 90° from the view depicted in (a). This figure was rendered using the programs MOLSCRIPT and POVRAY.

an estimated molecular weight of 22 500 by gel filtration corresponding to the monomer (*1*). In the crystal structure, GP-II forms a loose tetramer (a dimer of dimers, labeled monomers I−IV; see Figure 3a) in the crystallographic asymmetric unit with four distinct oligomerization interfaces. As depicted, monomer I interacts with monomers III and IV, and monomer II interacts with monomers III and IV. Monomers III and IV do not interact directly despite their close proximity (∼4 Å). It is unusual to obtain crystals of proteins which are highly glycosylated (GP-II is 8% glycosylated by weight). Often the carbohydrate is removed from the protein prior to crystallization. Even when such crystals are obtained, the carbohydrate portion of the structure is often disordered and thus cannot be observed in the electron density map. Thus, although 234 X-ray crystallographic structures are found in the Protein Data Bank using the query string "glycoprotein", the majority have one or no glycosyl residues in the model and less than 10% of these structures have three or more consecutive glycosyl residues per site. In the case of GP-II, half the total carbohydrate sequence was observed. GP-II has two glycosyl chains which are N-linked to Asn99 and Asn156 as determined previously by electrospray mass spectrometry (ES-MS) (unpublished data). Thus, the tetramer contains a total of eight glycosylation sites. Although the entire glycoside structure, $(Man)_3(Xyl)_1(Fuc)_1$-$(GlcNAc)_1(N\text{-}GlcNAc)_1$ was not observed for any of the sites (see Figure 4a for abbreviations), electron density for three glycosyl residues was observed linked to Asn156 of monomers I and II and linked to Asn99 of monomers III and IV. Either two glycosyl residues (monomer II Asn99), one glycosyl residue (monomer I Asn99), or no glycosyl residues (monomers III and IV Asn156) were observed in the other four sites of the tetramer. Electron density for the observed glycosyl residues was consistent with the linkage deduced by ES-CID mass spectrometric analysis. The N-GlcNAc residue has one α1→3 linked fucosyl residue and one β1→4 GlcNAc. The fit of the model into the $2F_o - F_c$ map for the *N*-glycosyl chain linked to Asn156 of monomer II is presented in Figure 4b.

The conformations of the glycosyl chains at Asn156 are similar in monomers I and II. The root-mean-square difference in relative atomic positions after superposition of these glycosyl chains from these monomers is 0.47 Å. This is also true of the conformation of the glycosyl chains at Asn99 in monomers III and IV, with the greatest difference in the conformation of the fucose moiety. Residues in the glycosyl

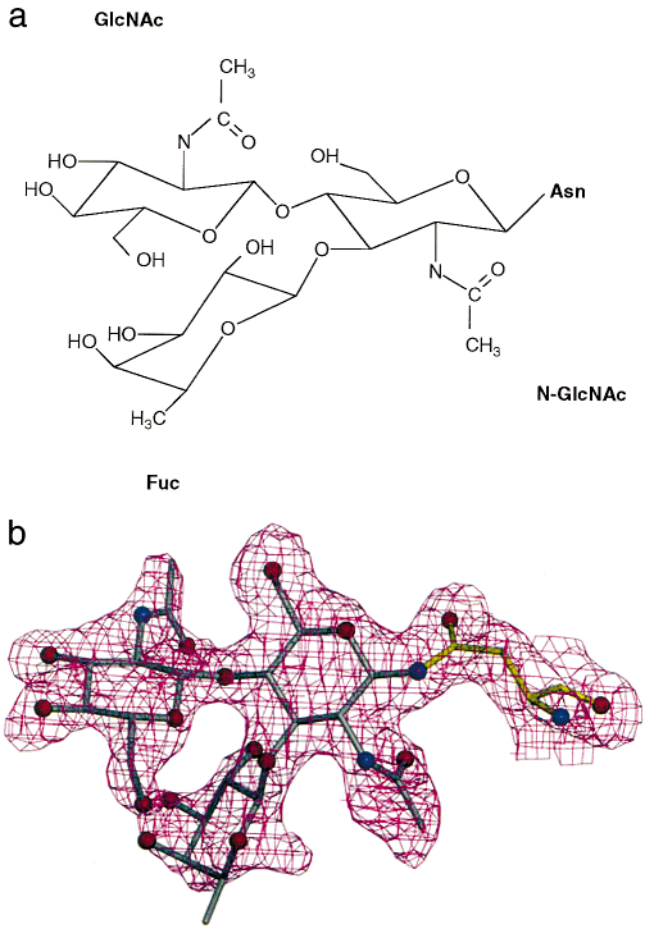

FIGURE 4: Structure of the N-linked carbohydrate chains of GP-II. (a) The chemical structure of three carbohydrate moieties linked to Asn156 or Asn99. The carbohydrate moieties are depicted in the same relative orientation as they appear in the molecular model shown in panel b. (b) A typical region of the electron density corresponding to the carbohydrate chains. The electron density from a $2F_o - F_c$ map is depicted as magenta cages while the atomic coordinates are depicted as ball-and-stick models. The three carbohydrate moieties linked to Asn156 of monomer II are shown. This figure was rendered using the programs MOLSCRIPT and POVRAY.

chains are involved in a hydrogen-bonding network that fixes the relative orientations of these three sugars with respect to one another. The hydrogen-bonding interactions between sugar moieties within a glycosyl chain and between sugar and protein are summarized in Table 3. One must consider the possibility that conformations of the sugars within the carbohydrate chains observed represent "artifacts" of crystallization since they are stabilized by interactions of the carbohydrate moieties with adjacent symmetry-related protein molecules. The dihedral angles of the majority of the glycosyl bonds present in the GP-II carbohydrate structure fall within the range of dihedral angles determined for solution structures of similar oligosaccharides by NMR and modeling (*34*, *35*). At Asn99 (in monomers III and IV) the dihedral angles for the GlcNAc−GlcNAc linkage fall just outside the allowed region. A "solution-like" carbohydrate structure was also found in the X-ray crystallographic structures of legume lectins (*36*−*38*). From evidence collected thus far via macromolecular X-ray crystallography, carbohydrate structures are not greatly distorted from those found in solution. The conformations of the glycosyl chains at Asn156 and Asn99 differ from one another. As depicted in Figure 3b, the glycosyl chains modifying Asn99 are all disposed on a single face of the tetramer, allowing their participation in crystal contacts. In contrast, the carbohydrate moieties at Asn156 participate in contacts related to noncrystallographic symmetry.

Table 3: Direct and Indirect Polar Contacts between Atoms of Carbohydrates and GP-II Molecules[a]

first group		bridging	second group	
monomer	residue	water[b]	monomer	residue
I	Fuc501 O2		I	GlcNAc503 O7
	Fuc501 O2			GlcNAc500 O4
	N-GlcNAc500 O7			Asn201 ND2
	GlcNAc503 O6		IV	Val43 O
	GlcNAc503 O6		IV	Ile42 O
	N-GlcNAc900 O7		I	Asn99 ND2
	N-GlcNAc900 O7	W299	I	Ser100 OG
II	GlcNAc503 O6			Fuc501 O4
	GlcNAc503 O6			Fuc501 O5
	GlcNAc503 O7			N-GlcNAc500 O4
	GlcNAc503 O7			Fuc501 O3
	N-GlcNAc500 O7			Asn156 ND2
	GlcNAc500 N2		III	Thr44 O
	GlcNAc500 N2		III	Val43 O
III	N-GlcNAc900 O6		III	GlcNAc901 N2
	N-GlcNAc900 O6			GlcNAc901 O4
	N-GlcNAc900 O6			Ser101 OG1

[a] All contacts listed are between 2.5 and 4.0 Å. [b] If no bridging water is listed, then the atom of the left-hand column is in direct contact with the corresponding atom of the right-hand column.

Contacts related to noncrystallographic symmetry consist of both protein−protein and protein−carbohydrate interactions. The protein−protein interface between monomers consists mostly of hydrogen-bonding interactions between side chains of Arg, Asn, and Ser and main chain atoms. The glycosyl chains involved in the noncrystallographic interface provide a "tether" between monomers of the tetramer. Shaanan et al. (*36*) reported in the structure of a legume lectin tethering of a N-linked heptasaccharide from one protein monomer to the adjacent protein monomer by an intricate network of intra- and intermolecular hydrogen bonds. For the six N-linked glycosyl chains of GP-II we observed two such networks. The participation of the *N*-glycosyl chain in a major contact is observed in the dimerization interface between monomers I and IV and monomers II and III (see Table 3). The N-linked glycosyl chain at Asn156 of monomer I makes intramolecular contacts with protein in monomer I and intermolecular contacts with protein in monomer IV. Also, the *N*-glycosyl chain at Asn156 of monomer II makes intramolecular contacts with protein in monomer II and intermolecular contacts with protein in monomer III. This tethering effect may be responsible for both the low mobility of the glycosyl chains and the formation of the tetramer. Borne et al. (*37*, *38*) observed hydrogen-bonding networks via water that act as bridging links between the glycosyl chains and the corresponding protein monomer. Extensive, ordered water structure of this type was not observed in GP-II. Although crystal packing contacts in a monomeric protein such as GP-II can be considered artifacts of crystallization, they make use of the same forces that govern contacts in protein complexes and oligomeric proteins. Thus the type of protein−carbohydrate contacts observed in the crystal structure of GP-II are likely to occur in other glycosyl−protein systems.

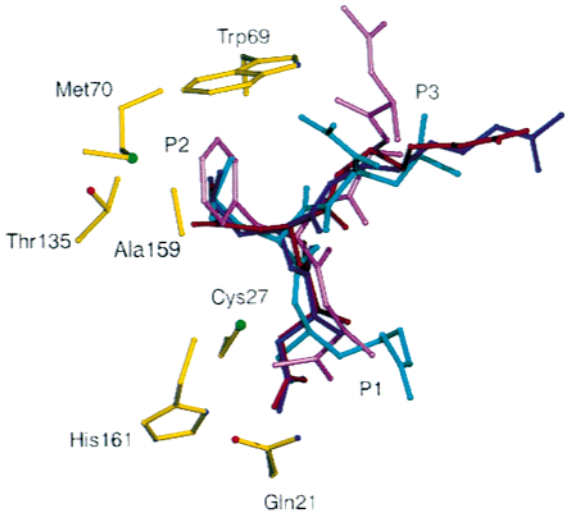

FIGURE 5: Superposition of the GP-II active site with inhibited complexes of proteases in the papain family. Only the GP-II active site residues are displayed for clarity. All amino acid side chains from GP-II are drawn using ball-and-stick models with yellow bonds. The inhibitors are depicted as licorice whips from superposition with the inhibited structures of papain (cyan and pink), protease Ω (red), and actinidin (dark blue) corresponding to the PDB accession codes 1pop, 1pad, 1meg, and 1aec, respectively. The inhibitor moieties P_1, P_2, and P_3 are labeled. This figure was rendered using the programs MOLSCRIPT and POVRAY.

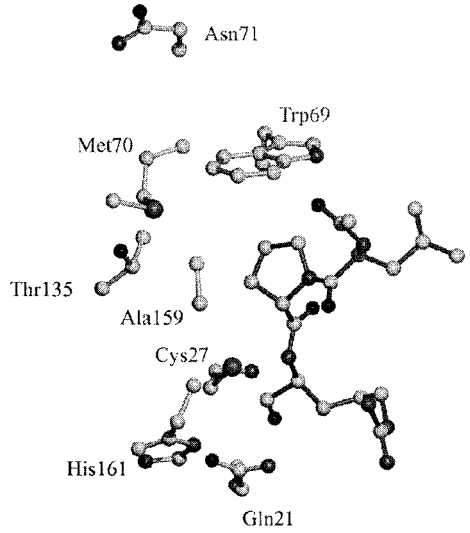

FIGURE 6: Interactions between GP-II and the model inhibitor Ac-Leu-Pro-arginal. All molecules are depicted as ball-and-stick models; residues from GP-II are drawn using light gray bonds while the inhibitor is drawn using black bonds. Those enzymic amino acid residues involved in binding proline at the P_2 site are depicted. The catalytic diad (Cys27 and His161) and oxyanion hole forming Gln21 are shown for reference. The inhibitor coordinates were obtained from the superposition of GP-II with the papain−leupeptin complex and substitution of the P_2 leucine to proline using the mutate-replace option in the program O.

Superposition of Cysteine Protease Inhibitors onto the Active Site of GP-II. The unique cyclic structure of proline places conformational restrictions on peptides and proteins and confers particular biological properties upon these biomolecules (*39*). Manipulating biomolecules containing proline requires either engineering proline binding sites into enzymes with a desired activity or changing the catalytic activity of enzymes with existing proline binding sites. Such protein engineering requires knowledge of the determinants of proline specificity. The S_2 site of GP-II is of special interest because of the enzyme's specificity for substrate with proline in the P_2 position. Complexes between proteases and peptide inhibitors can provide structural information about enzyme−substrate interactions in the ground state or transition state. Since a structure of GP-II with inhibitor bound is not yet available, we identified the substrate binding sites of GP-II by superposition of the protein on the homologous cysteine proteases. Comparison of the superimposed structures of inhibited complexes of papain, protease Ω, and actinidin (PDB entries 1pad and 1pop for papain, 1meg, and 1aec, respectively, for the other proteases) to the native structure of GP-II was made using the option Lsq in the graphics program O. Because of the structural similarity between these homologues (see previous section and Table 2), the active sites are extremely well aligned, and thus the inhibitors are reasonably positioned within the site (see Figure 5 for superposition). Overall, the peptide backbones of the inhibitors are in an extended conformation similar to one another, with the P_1 and P_2 backbone atoms bound in the active site cleft in the same relative orientation for all four inhibitors. The carbonyl oxygen of the P_1 residue, which would carry a formal negative charge in the corresponding tetrahedral intermediate, is in the appropriate position to form two stabilizing hydrogen bonds with the backbone NH of Cys25 (papain numbering) and the NH_2 group of Gln19. Other stabilizing interactions can be observed for the docked inhibitors. The P_2 carbonyl oxygen is positioned within hydrogen-bonding distance of the NH of the Gly66 backbone. Also, the P_2 side chain (Phe and Leu for those inhibitors) is bound to a hydrophobic area in the active site. The P_3 carbonyl group points away from the protein toward the solvent and shows different orientations for each inhibitor. Since Leu and Pro are both nonpolar, hydrophobic amino acids, the papain inhibitor Ac-Leu-Leu-arginal was chosen and changed to Ac-Leu-Pro-arginal in the program O within the papain−inhibitor complex (PDB entry 1pop). This new molecule was then analyzed in the active site of GP-II after superposition of the protein on the papain structure. Because no energy minimization is performed, only the positions of the enzyme residues comprising the proline binding pocket are considered in the analysis and not the position of the proline-containing peptide itself. The fit of the Ac-Leu-Pro-arginal to the GP-II active site is shown in Figure 6.

In papain, the structural elements that interact with the P_2 substrate side chains are comprised of the side chains of Tyr67, Pro68, Trp69, Val133, and Val157, which together form a hydrophobic pocket. In GP-II these residues are replaced by the corresponding residues Trp69, Met70, Tyr71, Thr135, and Ala159 (see Figure 6). The most distinct difference in the two structures is the location of Tyr67 (papain) and Trp69 (GP-II). The replacement of Ser60 in papain by Asn62 in GP-II pushes the indole ring of Trp69 in GP-II away from the position assumed by the side chain of Tyr67 seen in both the native and inhibited papain structures. Thus, the Cα of Trp69 (GP-II) relative to Tyr67 (papain) shifts by 2.2 Å. This causes the large inflexible indole ring of Trp69 to protrude into the S_2 site and reduces the volume of the site to almost half of that of papain. Thus, large residues cannot fit into the S_2 subsite due to steric hindrance with this Trp69. This steric clash is illustrated for the papain inhibitor Cbz-Gly-Phe-Gly in Figure 7. We

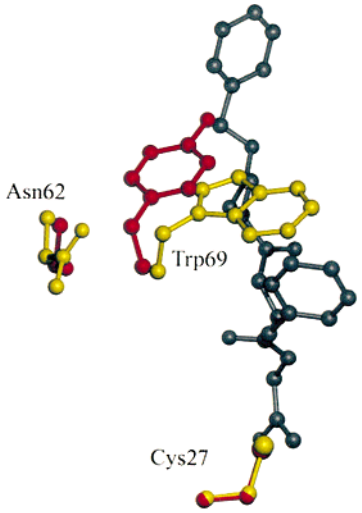

FIGURE 7: Superposition of the active sites of GP-II (yellow) and papain (red) depicted with ball-and-stick models. The peptidic inhibitor Cbz-Gly-Phe-Gly is shown in gray. The catalytic Cys27 is shown for reference. With respect to the Trp69 of GP-II, Tyr67 of papain is rotated out of the pocket and positioned with the phenolic hydroxyl group pointing toward the solvent. This position is allowed by the occurrence of the relatively small serine side chain in papain as opposed to the larger Asn62 in GP-II (also shown). This model predicts that the size exclusion effect of Trp69 on the P_2 binding pocket prevents inhibitors with bulky P_2 side chains, such as Cbz-Gly-Phe-Gly, from being accommodated in the GP-II active site.

interpret this change in the S_2 site as an important structural determinant of the specificity of GP-II toward Pro at the P_2 position via size exclusion. The contribution of the other hydrophobic residues forming the S_2 binding site toward proline specificity cannot be underestimated.

In contrast to these differences at the S_2 sites the residues forming the S_1 sites are identical in GP-II and papain; thus the specificity of GP-II at P_1 is predicted to be similar to that of papain. Papain will cleave substrates with Arg, Lys, Glu, His, Gly, or Tyr at P_1, although Lys and Arg are preferred (*40*). The superposition of the four cysteine protease inhibitors with the GP-II structure described above and depicted in Figure 5 shows that the major interaction at P_1 is between the ligand backbone carbonyl and the protein oxyanion hole. The S_1 binding pocket is fairly shallow, allowing medium-length and long side chains to point out into solvent when ligand is bound. There are no interactions between the superimposed inhibitor side chains at P_1 and enzymic residues. This structural analysis is consistent with the finding that GP-II has some preference for ionic residues (e.g., Lys and Glu) at P_1 (*15*). That is, the solvent accessibility of the S_1 pocket makes the absolute charge of the P_1 side chain irrelevant but requires that the ligand side chain be hydrophilic.

Proline Binding in Enzymes. Is the proline binding site in GP-II similar to proline binding sites in other proteins? With the structure of GP-II now in hand, it is possible to compare the proline binding site of GP-II with those of other enzymes in which the primary specificity is for proline binding. α-Lytic protease is a serine protease that can bind peptides through six subsites, of which S_2 is of primary importance since favorable binding interactions in that pocket can increase k_{cat}/K_m or k_{cat} by several-hundredfold and the P_1 side chain has little effect on binding of peptide substrates (*41*). A preference for Pro at the P_2 position has also been noted in the binding of peptide boronic acid inhibitors (*42*). Structural features of the S_2 subsite of α-lytic protease are drawn from crystallographic (*43*) and modeling studies (*44*). One edge of the pocket is composed of the imidazole ring of His157, which is part of the catalytic triad. Residues Phe94 and Tyr171 form sides of the P_2 pocket with the aromatic ring of Tyr171 making a face-on interaction with the proline ring.

Cyclophilin has been shown to have peptidyl−prolyl *cis/trans* isomerase activity in vitro. The crystal structures of *Escherichia coli* and human cyclophilin A complexed with inhibitors have been solved (*45*, *46*); the human numbering system will be used here. The proline binding pocket is composed of Phe113, Phe60, His126 (Tyr120 in *E. coli*), Met61, and Gln63. There is no face-on interaction of the proline ring as in α-lytic protease. Instead, the aromatic rings of Phe113 and Phe61 make an edge-on interaction with the proline ring of the inhibitor. The proline binding pockets of GP-II, cyclophilin A, and α-lytic protease utilize the aromatic side chains of Tyr, Phe, His, and Trp and the hydrophobic side chains of Met and Ala. There is no consensus three-dimensional arrangement of these side chains with respect to one another (or the face of the proline ring, with both edge-on and face-on interactions being observed for α-lytic protease and cyclophilin). Instead, in all cases the selectivity for proline is achieved, in part, by the formation of a small, hydrophobic, pocket.

These selection mechanisms do not explain the exclusion of small, nonpolar amino acids such as Leu and Val from the S_2 pocket. Evidence from substrate kinetics for proline-containing peptides for elastase and some other serine proteases suggests that the degree to which long peptides display multiple binding modes is reduced with proline-containing substrates because of the inability of the S_3 pocket to bind proline (*47*). This is believed to be due to the fact that these substrate peptides are more restricted in their backbone conformation. The restriction of the number of available substrate conformations may be contributing to the specificity for proline exhibited by GP-II. Since the peptide ϕ angle enforced by proline is −60°, which is the approximate angle of residues found in type I and type II turn conformations, the presence of proline at P_2 may hold the substrate in a more "turn-like" conformation which provides a close enzyme−peptide fit with less entropic cost.

In future studies it will be of interest to determine the contribution of the proline residue to the overall backbone conformation of the peptide ligand. Are substrates and inhibitors in which the ligand is held in a "turn" conformation superior? Such a conformational effect of proline may be distinguished from the interactions made by the proline side chain with enzyme specificity pockets. Undoubtedly, future kinetic studies of the specificity of each of the subsites, coupled with X-ray crystallographic studies of enzyme−inhibitor complexes, should provide a deeper knowledge of the structural basis for the specificity of GP-II.

CONCLUSIONS

The structure of GP-II, a cysteine protease with specificity for proline in the P_2 position, has been solved by X-ray crystallography to 2.1 Å resolution and refined to an *R*-factor

and *R*-free of 0.214 and 0.248, respectively. The protein is 8% glycosylated by weight. Half of the total carbohydrate sequence was observed in the structure, and these N-linked glycosyl chains participate in both crystallographic and noncrystallographic contacts via hydrogen bonds (without intervening ordered water molecules). The GP-II carbohydrate structure as well as carbohydrate structures from the limited number of available glycoprotein crystal structures is not distorted significantly from those determined in solution studies. The docking of peptides to GP-II based on superposition with inhibited papain structures identifies the residues involved in proline binding. Surprisingly, comparison of the proline binding site of GP-II to the proline binding sites of the enzymes cyclophilin and α-lytic protease does not reveal a common proline binding motif, but all three proteins utilize similar amino acids to produce the pocket (e.g., Phe, Trp, Met, and Ala). Since such a binding subsite would not exclude other small, nonpolar amino acids, added selectivity may be achieved because proline holds the substrate peptide in a turn-like conformation.

ACKNOWLEDGMENT

We thank Marc C. Morais for his help in data processing and Carla Mattos for her thoughtful comments on the manuscript.

REFERENCES

1. Ichikawa, Y., Sasa, H., and Michi, K. (1973) Purification of ginger protease, *J. Jpn. Soc. Food Nutr. 26*, 377−383.
2. Hashimoto, A., Takeuti, Y., Kawahara, Y., and Yasumoto, K. (1991) Proteinase and collagenase activities in ginger rhizome, *J. Jpn. Soc. Nutr. Food Sci. 44*, 127−132.
3. Thompson, E. H., and Allen, C. E. (1973) Ginger rhizome: a new source of proteolytic enzyme, *J. Food Sci. 38*, 652−655.
4. Kamphuis, I. G., Kalk, K. H., Swarte, M. B. A., and Drenth, J. (1984) Structure of papain refined at 1.65 Å resolution, *J. Mol. Biol. 179*, 233−256.
5. Baker, E. N. (1980) Structure of actinidin, after refinement at 1.7 Å resolution, *J. Mol. Biol. 141*, 441−485.
6. Pickersgill, R. W., Rizkallah, P., Harris, G. W., and Goodenough, P. W. (1991) Determination of the structure of papaya protease Ω, *Acta Crystallogr. B47*, 766−771.
7. O'Hara, B. P., Hemmings, A. M., Buttle, D. J., and Pearl, L. H. (1995) Crystal structure of glycyl endopeptidase from *carcica papaya*: a cysteine endopeptidase of unusual substrate specificity, *Biochemistry 34*, 13190−13195.
8. Glazer, N., and Smith, E. L. (1971) Papain, in *The Enzymes, 3rd Ed.*, Vol. III, pp 519−523, Academic Press, New York and London.
9. Buttle, D. J., Ritonja, A., Pearl, L. H., Turk, V., and Barrett, A. J. (1990) Selective cleavage of glycyl bonds by papaya proteinase IV, *FEBS Lett. 260*, 195−197.
10. Rawlings, N. D., and Barrett, A. J. (1994) Families of cysteine peptidases, *Methods Enzymol. 244*, 461−485.
11. Musil, D., Zucic, D., Turk, D., Engh, R., Mayr, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma, N., and Bode, W. (1991) The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity, *EMBO J. 10*, 2321−2330.
12. McGuire, M. J., Lipsky, P. E., and Thiele, D. L. (1992) Purification and characterization of dipeptidyl peptidase I from human spleen, *Arch. Biochem. Biophys. 295*, 280−295.
13. Schechter, I., and Berger, A. (1968) On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain, *Biochem. Biophys. Res. Commun. 32*, 898−902.
14. Choi, K. H. (1998) Ph.D. Dissertation, Boston University.
15. Zhu, C. (1999) Ph.D. Dissertation, Boston University.
16. Fenton, J. W., II (1981) Thrombin specificity, *Ann. N.Y. Acad. Sci. 370*, 468−495.
17. Chang, J.-Y. (1985) Thrombin specificity: requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate, *Eur. J. Biochem. 151*, 217−224.
18. Lottenberg, R., Hall, J. A., Blinder, M., Binder, E. P., and Jackson, C. M. (1983) The action of thrombin on peptide *p*-nitroanilide substrates: substrate selectivity and examination of hydrolysis under different reaction conditions, *Biochim. Biophys. Acta 742*, 539−557.
19. Djie, M. Z., Le Bonniec, B. F., Hopkins, P. C. R., Hipler, K., and Stone, S. R. (1996) Role of the P_2 residue in determining the specificity of serpins, *Biochemistry 35*, 11461−11469.
20. Fontana A. (1989) Limited proteolysis of globular proteins occurs at exposed and flexible loops, *Highlights Mod. Biochem. 2*, 1711−1726.
21. Jancarik, J., and Kim, S. H. (1991) Sparse matrix sampling: a screening method for crystallization of proteins, *J. Appl. Crystallogr. 24*, 409−411.
22. Otwinowski, Z. (1990) *DENZO Data Processing Package*, Yale University, New Haven, CT.
23. Navaza, J. (1994) AMORE: an automated package for molecular replacement, *Acta Crystallogr. A50*, 157−163.
24. Peitsch, M. C. (1996) ProMod and Swiss-Model: Internet-based tools for automated comparative protein modeling, *Biochem. Soc. Trans. 24*, 274−279.
25. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these methods, *Acta Crystallogr. A47*, 110−119.
26. Brünger, A. T. (1992) *X-PLOR, Version 3.1. A system for X-ray crystallography and NMR*, Yale University Press, New Haven, CT.
27. Jiang, J. S., Brünger, A. T. (1994) Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures, *J. Mol. Biol. 243*, 100−115.
28. Kleywegt, G. J., and Jones, T. A. (1997) Model-building and refinement practice, *Methods Enzymol. 277*, 208−230.
29. Kleywegt, G. J., and Jones, T. A. (1996) Efficient rebuilding of protein structures, *Acta Crystallogr. D52*, 829−832.
30. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
31. Cavallini, D., Graziani, M. T., and Dupré, S. (1966) Determination of disulfide groups in proteins, *Nature 212*, 294−295.
32. Otto, H., and Schirmeister, T. (1997) Cysteine proteases and their inhibitors, *Chem. Rev. 97*, 133−171.
33. Drenth, K. H., and Swen, H. M. (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain, *Biochemistry 15*, 3731−3738.
34. Imberty, A., Gerber, S., Tran, V., and Pérez, S. (1990) Data bank of three-dimensional structures of disaccharides, a tool to build 3-D structures of oligosaccharides: part 1. oligomannose type N-glycans, *Glycoconjugate J. 7*, 27−54.
35. Bouwstra, J. B., Spoelstra, E. C., Ge Waard, P., Leeflang, B. R., Kamerling, J. P., and Vliegenthart, J. F. (1990) Conformational studies on the N-linked carbohydrate chain of Bromelain, *Eur. J. Biochem. 190*, 113−122.
36. Shaanan, B., Lis, H., and Sharon, N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose, *Science 254*, 862−866.
37. Bourne, Y., Rougé, P., and Cambillau, C. (1990) X-ray structure of a (α-Man(1−3)*β*-Man(1−4)GlcNAc)-Lectin complex at 2.1 Å resolution, *J. Biol. Chem. 265*, 18161−18165.
38. Bourne, Y., Rougé, P., and Cambillau, C. (1992) X-ray structure of a biantennary octasaccharide-lectin complex refined at 2.3-Å resolution, *J. Biol. Chem. 267*, 197−203.
39. Cunningham, D. F., and O'Connor, B. (1997) Proline specific peptidases, *Biochim. Biophys. Acta 1343*, 160−186.

40. Hill, R. L. (1965) Hydrolysis of proteins, *Adv. Protein Chem. 20*, 37−107.
41. Bauer, C. A., Brayer, G. D., Sielecki, A. R., and James, M. N. G. (1981) Active site of alpha-lytic protease: enzyme−substrate interactions, *Eur. J. Biochem. 120*, 289−294.
42. Bone, R., Shenvi, A. B., Kettner, C. A., and Agard, D. A. (1987) Serine protease mechanism: substrate of an inhibitory complex of alpha-lytic protease and a tightly bound peptide boronic acid, *Biochemistry 26*, 7609−7614.
43. Fujinaga, M., Delbaere, L. T., Brayer, G. D., and James, M. N. G. (1985) Refined structure of alpha-lytic protease at 1.7 Å resolution: analysis of hydrogen bonding and solvent structure, *J. Mol. Biol. 184*, 479−502.
44. Epstein, D. M., and Abeles, R. H. (1992) Role of Ser 214 and Tyr 171, components of the S_2 subsite of α-lytic protease, in catalysis, *Biochemistry 31*, 11216−11223.
45. Zhao, Y., and Ke, H. (1996) Mechanistic implication of crystal structures of the cyclophilin-dipeptide complexes, *Biochemistry 35*, 7362−7368.
46. Konno, M., Ito, M., Hayano, T., and Takahashi, N. (1996) The substrate-binding site in *Escherichia coli* cyclophilin A preferably recognizes a *cis*-proline isomer or a highly distorted form of the *trans* isomer, *J. Mol. Biol. 256*, 897−908.
47. Thompson, R. C., and Blout, E. R. (1973) Restrictions on the binding of proline-containing peptides to elastase, *Biochemistry 12*, 51−56.
48. Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
49. Persistence of Vision Ray Tracer (POV-Ray) Version 3.01.

BI990651B