Biochemistry

 Volume 32, Number 43 November 2, 1993

Accelerated Publications

A Glutamic Acid Specific Serine Protease Utilizes a Novel Histidine Triad in Substrate Binding†,‡

Vicki L. Nienaber,§,‖ Klaus Breddam,⊥ and Jens J. Birktoft*,§,#

Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, Missouri 63110, and Department of Chemistry, Carlsberg Laboratory, Copenhagen, Denmark

Received April 14, 1993; Revised Manuscript Received August 23, 1993®

ABSTRACT: Proteases specific for cleavage after acidic residues have been implicated in several disease states, including epidermolysis, inflammation, and viral processing. A serine protease with specificity toward glutamic acid substrates (Glu-SGP) has been crystallized in the presence of a tetrapeptide ligand and its structure determined and refined to an *R*-factor of 17% at 2.0-Å resolution. This structure provides an initial description of the design of proteolytic specificity for negatively charged residues. While the overall fold of Glu-SGP closely resembles that observed in the pancreatic-type serine proteases, stabilization of the negatively charged substrate when bound to this protein appears to involve a more extensive part of the protease than previously observed. The substrate carboxylate is bound to a histidine side chain, His213, which provides the primary electrostatic compensation of the negative charge on the substrate, and to two serine hydroxyls, Ser192 and Ser216. Glu-SGP displays maximum activity at pH 8.3, and assuming normal pK_a's, the glutamate side chain and His213 will be negatively charged and neutral, respectively, at this pH. In order for His213 to carry a positive charge at the optimal pH, its pK_a will have to be raised by at least two units. An alternative mechanism for substrate charge compensation is suggested that involves a novel histidine triad, His213, His199, and His228, not observed in any other serine protease. The C-terminal α-helix, ubiquitous to all pancreatic-type proteases, is directly linked to this histidine triad and may also play a role in substrate stabilization. The amino acid sequence of this protease near the primary substrate binding site has been compared with those of distantly related viral and bacterial proteases. These considerations and comparison with a recently reported structure of a chymotrypsin-like protease from Sindbis virus core protein (Tong et al., 1993) tentatively suggest that a putative general substrate binding scheme for proteases with specificity toward glutamic acid may involve a histidine residue, His213, and a hydroxyl function at residue 192.

Traditionally, the pancreatic-type serine proteases have been classified into three groups: those which cleave after positively charged residues (trypsin-like), those which cleave after large hydrophobic residues (chymotrypsin-like), and those which cleave after small hydrophobic residues (elastase-like) (Stroud, 1974). Serine proteases specific for other types of amino acids, including negatively charged residues, have been characterized, and many of these have found widespread use in protein

† V.L.N. was a fellow of the American Heart Association, Missouri Affiliate.

‡ The refined atomic coordinates have been deposited with the Protein Data Bank at Brookhaven National Laboratory (PDB code 1HPG).

* To whom correspondence should be addressed at Hoffmann La-Roche Inc.

§ Washington University.

‖ Current address: DuPont Merck Pharmaceuticals, Experimental Station, Wilmington, DE 19880

⊥ Carlsberg Laboratory.

Current address: Roche Research Center, Hoffmann-La Roche Inc., 340 Kingsland St., Nutley, NJ 07110.

® Abstract published in *Advance ACS Abstracts,* October 1, 1993.

0006-2960/93/0432-11469$04.00/0

degradation and sequence studies. While many of the properties of these proteases, including their amino acid sequences, have been described, structural studies of such serine proteases have until now not been reported.

A serine protease with high specificity for cleavage after glutamic acid residues, glutamic acid specific *Streptomyces griseus* protease (Glu-SGP),[1] has been isolated and characterized (Yoshida *et al.*, 1988; Svendsen *et al.*, 1991; Breddam & Meldal, 1992). The catalytic efficiency of Glu-SGP for cleavage after aspartic acid residues is much lower than for glutamic acid and is in fact no better that in the case of cleavage after phenylalanine and alanine side chains (Breddam & Meldal, 1992). In addition, this protease has a high sequence identity to *S. griseus* protease A (SGP-A, 59%) and B (SGP-B, 56%), both having chymotrypsin-like specificity, as well as to the elastase-like *Myxobacter* 495 α-lytic protease (α-LP, 36%) (Svendsen *et al.*, 1991). Because these proteases contain about 30 fewer amino acids than the pancreatic enzymes, they have been classified into a subgroup of pancreatic serine proteases, the small bacterial serine proteases (James *et al.*, 1978).

Several other acid-specific proteases which have important biological roles have been identified. Epidermolytic toxins A and B from *Staphylococcus aureus* which induce staphylococcal scalded skin syndrome in newborns (Dancer *et al.*, 1990) are serine proteases homologous to the glutamic acid specific V8 protease also from *S. aureus*. The aspartic acid specific interleukin-1β converting enzyme, present in monocytes, activates interleukin-1β, which is important in the pathogenesis of inflammatory diseases (Thornberry *et al.*, 1992). Granzyme B found in cytotoxic T lymphocytes where it functions in the defense against tumor cell proliferation and viral infections is likewise an aspartic acid specific protease (Odake *et al.*, 1991). Additionally, a number of viral proteases have been demonstrated to cleave after glutamic acid and glutamine residues. These viral proteases include those found in hepatitis-A, human polio virus, and human rhino virus (common cold), where they serve a principal role in the processing of viral polyproteins into functional gene products (Wellink & van Kammen, 1988).

In this paper we report an initial description of the unique specificity design in the acid-specific proteases, together with a preliminary examination of the crystal structure of Glu-SGP. A more detailed analysis of other aspects of the refined structure and comparison with other proteases will be presented elsewhere. Examination of substrate binding to Glu-SGP indicates that this enzyme is unique among the known serine protease structures, as its primary specificity is defined in part by residues outside the primary binding site. This is partly necessitated by the lack of an obvious direct counter charge for the glutamic acid side chain, an unanticipated finding (Graf *et al.*, 1987; Komiyama *et al.*, 1991). Thus, rather than using a lysine or arginine side chain as a counter charge, this protease appears to utilize three histidines which have been termed the histidine triad.

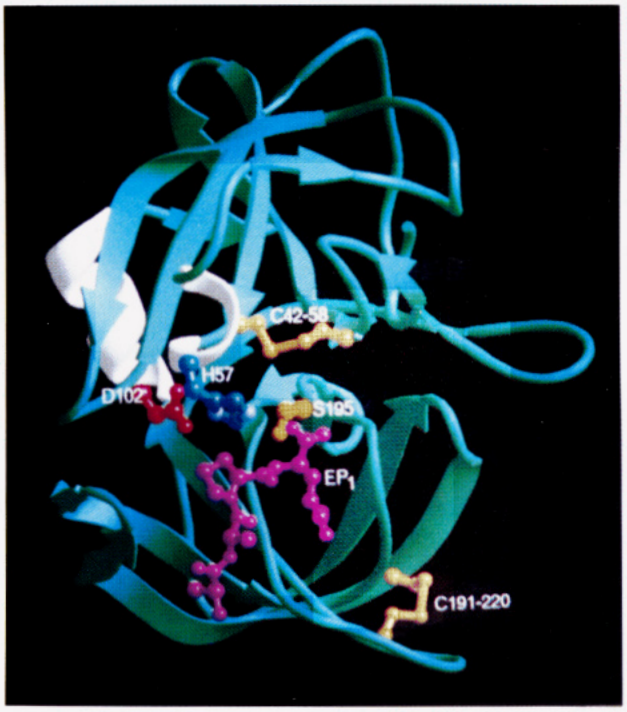

FIGURE 1: Ribbon drawing of the main chain fold of Glu-SGP. Helices are shown in white, while β-strands are drawn as green arrows. The N-terminal β-barrel 1 forms the top portion of the structure, while β-barrel 2 containing the substrate binding site is located at the bottom. The catalytic triad (D102, H57, and S195), color coded red, blue, and yellow, respectively, and the tetrapeptide ligand (Ala-Ala-Pro-Glu), in pink, are represented as ball and stick models. The two disulfide bridges, C42–C58 in β-barrel 1 (top) and C191–C220 in β-barrel 2 (bottom), are also shown.

MATERIALS AND METHODS

Glu-SGP was isolated from Pronase (Actinase, Kaken Seiyaku, Tokyo, Japan) as described (Svendsen *et al.*, 1991). Boc-Ala-Ala-Pro-Glu-*p*-nitroanilide was prepared as described (Breddam & Meldal, 1992). Using the factorial random screening method (Carter, 1990) a number of conditions were identified which yielded crystals. The best crystals and those used in the structure determination reported here were grown as follows: The protease was incubated with a 10-fold molar excess of Boc-Ala-Ala-Pro-Glu-*p*-nitroanilide overnight at 4 °C in 20 mM Hepes, pH 7.5, and then brought to a final protein concentration of 33 mg/mL in 0.2 M $MgCl_2$ and 0.1 M Hepes. Crystals were grown by vapor diffusion against 50% saturated sodium citrate, pH 7.5 at 20 °C, using the sitting drop method, reaching a size of ~0.5 mm over a period of 36–48 h. They belong to space group $C2$, with a = 77.26 Å, b = 36.29 Å, c = 51.22 Å, and β = 101.8° and diffract to better than 1.3-Å resolution. Each asymmetric unit contains one protease molecule, and the Matthews volume, V_m, is 1.9 Å^3/Da.

Data were collected on a Xoung–Hamlin-type multiwire area detector and processed using the program package of Chris Nielsen. A total of 90 655 observations of the 13 006 possible reflections to 1.8 Å were collected and merged together with an R_{merge} of 8.3%. Initial phases were determined by molecular replacement using the program package MERLOT

[1] Abbreviations: Glu-SGP, glutamic acid specific *Streptomyces griseus* protease; SGP-A, *Streptomyces griseus* protease A; SGP-B, *Streptomyces griseus* protease B; α-LP, α-lytic protease; Tryp-SG, *Streptomyces griseus* trypsin; Tryp-Bov, bovine trypsin; CHT-Bcv, bovine α-chymotrypsin; Elast-Porc, porcine pancreatic elastase; Elast-HN, human neutrophil elastase; HepBP-Hum, human heparin binding protein; Colla-Crab, Uca pugilator collagenolytic serine protease; RMCPII-rat, rat mast cell protease II; Cathep-G, cathepsin G; SA-V8, *Staphylococcus aureus* V8 protease; SA-ETOX-A,-B, *Staphylococcus aureus* epidermolytic toxin A and B; Glu-BL, *Bacillus licheniformis* glutamic acid specific protease; TBRV-24K, tomato black ring virus; CPMV-24K, cowpea mosaic virus; PPV-NIa, plum pox virus; FMDV-3C, foot and mouth disease virus; EMCV-3C, encephalomyocarditis virus; HAV-3C, hepatitis-A virus; CVA21-3C, Coxsackie virus A21; HRV14-3C, human rhino virus; YFN-NS3, yellow fever virus; WNV-NS3, west Nile virus; SCP, Sindbis virus core protein.

(Fitzgerald, 1988) with a stripped model (backbone and Cβ) of SGP-A (PDB code 2SGA; Sielecki *et al.*, 1979) as the search probe. Using data between 4- and 8-Å resolution and a cutoff radius of 20 Å origin removed Patterson maps were calculated with the subroutine CROSUM. The rotation function was searched over an asymmetric unit of Euler angle space on a grid of 2.5° for α and 5° for β and γ, with the highest peak being the right one. A single peak over 65% was obtained by the subroutine PAKFUN for the translation function and proved to be the correct solution. The *x*–*z* parameters were further refined by calculating an *R*-value map (RVAMAP) and using the minimum of this calculation; the starting *R*-factor as calculated by XPLOR (Brunger, 1991) was 0.444 using 4–8-Å data and a global temperature factor of 20 Å^2.

The structure was refined using the program package XPLOR, employing a combination of rigid body, conjugate gradient, and simulated annealing energy minimization. The refinement used only data which was 2σ above background and began with the structure of SGP-A (including side chains) positioned into the correct orientation within the Glu-SGP unit cell. After five cycles of refinement, a $2F_o - F_c$ map was calculated, and the Glu-SGP sequence was incorporated into the structure using the program TURBO (Roussel & Cambillau, 1989). Most of the side-chain changes could be fitted very easily at this stage. Unique insertion loops were less obvious and were modeled during the latter stages of the refinement process. The resolution was extended gradually from 4.0 to 2.0 Å, at which stage restrained individual *B*-factor refinement was carried out. The current model includes 66 ordered solvent molecules and corresponds to an *R*-factor of 0.17 with rms bond-length deviations of 0.014 Å and rms bond angle deviations of 3.18°. Secondary structure assignment was calculated by the method of Kabsch and Sander (1983), and solvent accessibility was determined with the Lee and Richards (1971) algorithm as implemented in XPLOR.

RESULTS

In Figure 1, the tertiary structure of Glu-SGP is shown as a ribbon diagram. This structure is very similar[2] to that of SGP-A (rms 1.19 Å; Sielecki *et al.*, 1979), SGP-B (rms 1.24 Å; Read *et al.*, 1983), and α-LP (rms 1.18 Å; Fujinaga *et al.*, 1985) and somewhat similar to that of the pancreatic serine proteases (Birktoft & Blow, 1972). Like other serine proteases, Glu-SGP is composed of two β-barrel cylindrical structures and a C-terminal α-helix. The first β-barrel contains His57[3] and Asp102 of the catalytic triad, while β-barrel 2 contains Ser195 of the catalytic triad and the S_1[4] specificity pocket.

Hydrogen-bonding of the Ala-Ala-Pro-Glu tetrapeptide ligand at the S_1 site of Glu-SGP is detailed in Figure 2A. In this orientation, the pocket will be described as having a top, bottom, left, and right side. The right side of the pocket is formed by the backbone extending up from Ala192A-N (see Figure 4 for partial sequence alignment) to Ser195-N, while the left side of the pocket arises from the backbone at Ser214-N down through Ser216-Oγ. Side chains contributing to binding specificity are His213, located at the back of the pocket, as well as Ser216 and Ser192, both located at the front. The ligand has two carboxyl groups: the C-terminal formed after substrate cleavage and the functional group of the glutamic acid (Glu-P_1) side chain. The C-terminal OT1 atom hydrogen bonds with His57-Nε2 of the catalytic triad and with a solvent molecule (not shown), while OT2 is located in the oxyanion hole forming hydrogen bonds with Ser195-N and Gly193-N. The oxyanion hole is important in stabilizing a negative charge which develops on the tetrahedral intermediate during catalysis and should therefore serve no role as a substrate specificity determinant (Kraut, 1977). Cα and Cβ primarily interact with the right side of the pocket, while Cγ and Cδ are located in the groove formed by the left and right side of the P_1 binding site. Oε1 points toward the back side of the pocket forming hydrogen bonds with His213-Nε2 and Ser192-Oγ, while Oε2 points toward solvent and is also hydrogen bonded to Ser216-Oγ and with solvent molecules. The overall geometry of the Glu-SGP S_1 site is very similar to that observed in SGP proteases A and B, as well as in α-LP. SGP-A and SGP-B both have an alanine at 192, a threonine at 213, and a glycine at 216 (see Figure 4). In α-LP, methonines are at positions 192 and 213, while a glycine is at position 216. Thus, none of the residues implicated in the binding specificity of Glu-SGP are present in these other highly homologous enzymes. An $F_o - F_c$ electron density omit map, superimposed on selected residues of the Glu-SGP model, is shown in Figure 2B. Well-defined density is indicated for all residues of Glu-SGP as well as for the Glu-peptide bound at the active site of this protease.

Inspection of the Glu-SGP S_1 site initially suggested that the enzyme–substrate complex might have reduced stability at pH 8.3 where the enzyme exhibits the highest activity.[5] Assuming normal pK_a values, the Glu-P_1 side chain would be negatively charged at this pH while His213 should be neutral;[6] thus, there is no positive counter charge in the vicinity of the substrate carboxyl group. However, a closer examination of the structure indicates that binding of the P_1 moiety at the S_1 site might be further stabilized by an extended hydrogen-bonding network which involves three histidines: His213 of the S_1 specificity pocket, His199, and His228 (Figure 3A). This His triad stretches through the hydrophobic center of β-barrel 2 and forms a direct link between the S_1 site and a solvent channel located at opposite ends of this barrel. Specifically, hydrogen bonds are formed between His213-Nε2 and Glu-P_1-Oε2, between His213-Nδ1 and His199-Nε2, and between His199-Nδ1 and His228-Nδ1. His228-Nε2 is in turn hydrogen bonded to an ordered buried solvent molecule, labeled W1, located at the opposite end of β-barrel 2 (Figure 3A).

The bottom of β-barrel 2 is capped by a loop extending from Asn129 to Val132 (Figure 3A), which together with the side chains of Pro230, Val181, and Val165 form a deep pocket. This pocket contains four ordered solvent molecules (W1–W4), of which only one, W2, is accessible to bulk solvent. Solvent molecule W1 is hydrogen bonded to His228, to the solvent-accessible W2, and to W4. The key interactions within this solvent channel center around solvent molecule W4, which

[2] rms deviations between Glu-SGP and the other small bacterial proteases were calculated using the program HOMOLOG (Rao & Rossmann, 1973) employing only Cα atoms. The small bacterial proteases have previously been compared in detail to the larger pancreatic enzymes by James et al. (1978).

[3] The numbering system is based on that of α-chymotrypsin (James et al., 1978).

[4] Substrate amino acids are represented by P*n*, ..., P2, P1, P1′, P2′, ..., P*n*′, where P1–P1′ is the bond hydrolyzed, while S*n*, ..., S2, S1, S1′, S2′, ..., S*n*′ represents the corresponding binding sites on the protein following the convention of Schechter and Berger (1968).

[5] Breddam, unpublished results.

[6] If it is assumed that the pK_a for His213 is 6.4 in the absence of substrate, then less than 1% of the imidazole ring should be protonated at pH 8.5. Additional considerations of the protonation state of His213 are presented in the Discussion.

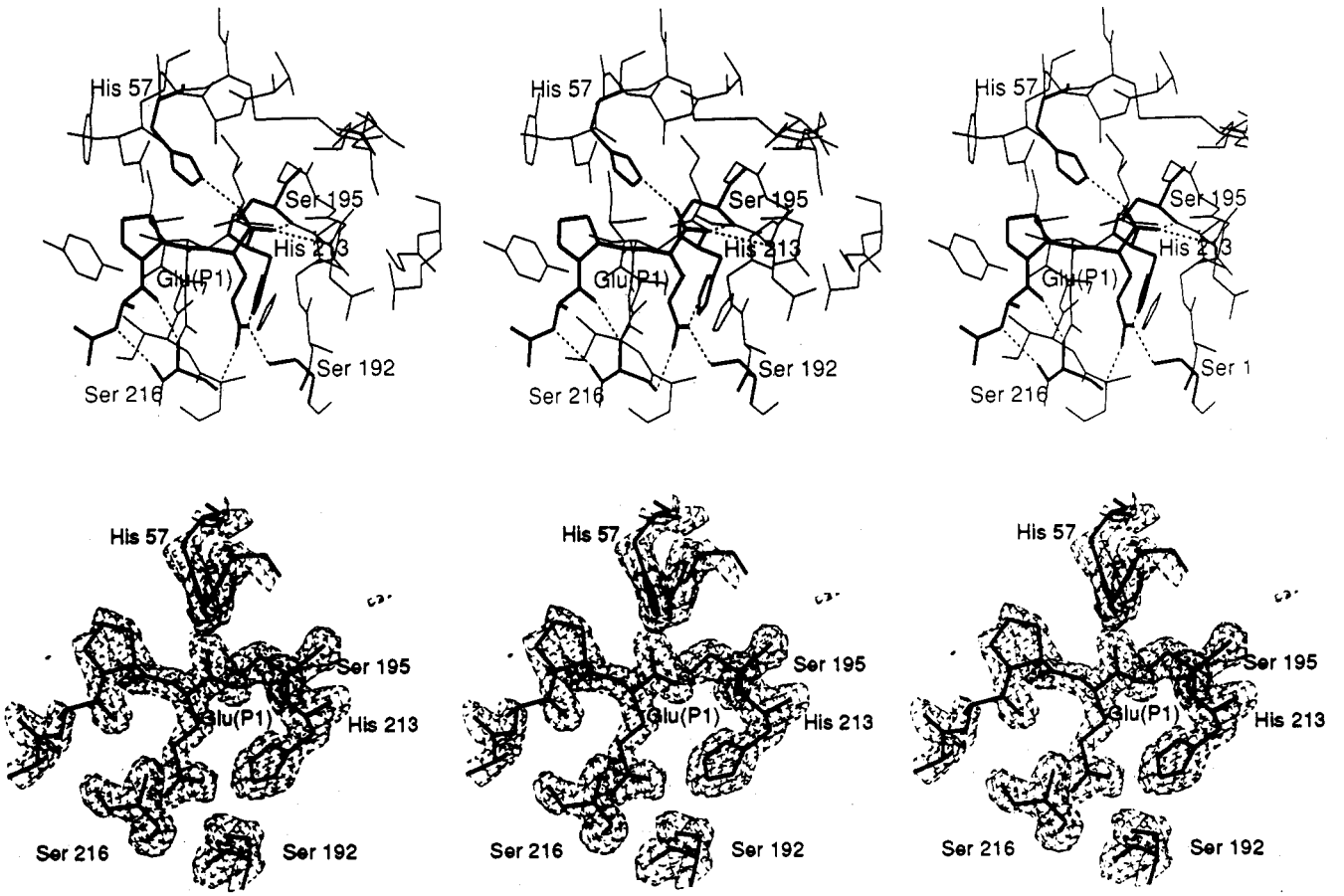

FIGURE 2: (A, top) Stereo representation of the active site of Glu-SGP. The Ala-Ala-Pro-Glu part of the ligand as well as residues His57, Asp102, Ser195, His213, Ser192, and Ser216 are shown as the thicker dark lines. The view is the same as in Figure 1. Hydrogen bonds between the peptide and Glu-SGP are represented by dashed lines. The left pair of figures gives a parallel stereoview while the right pair provides a cross-eyed stereo. All stereo diagrams in this paper are presented similarly. (B, bottom) $F_o - F_c$ omit map contoured at 2.7σ. Residues His57, Ser192, Ser195, His213, and Ser216 and the Boc-Ala-Ala-Pro-Glu peptide were omitted in the calculations. A break in density between the proline and alanine residues of the bound peptide can be seen in this figure. However, continuous density for the peptide ligand is present in the $2F_o - F_c$ map contoured at 1.0σ and in the $F_o - F_c$ omit map contoured at 1.8σ.

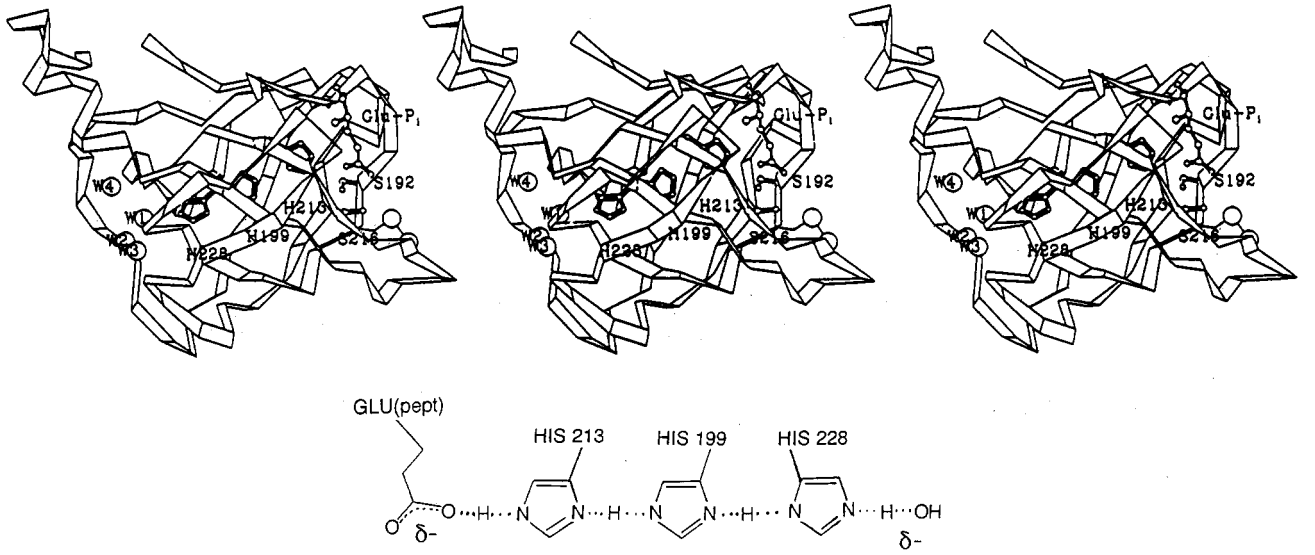

FIGURE 3: (A, top) Ribbon drawing of the second β-barrel of Glu-SGP. The view is rotated 90° to the right of that in Figures 1 and 2A. Side chains of the His triad, consisting of His213, His199, and His228, are represented by thick lined ball and stick models. Solvent molecules W1, W2, W3, and W4 are represented by circles, and Glu-P_1, Ser192, and Ser216 are shown as thinly lined ball and stick models. The 191–220 disulfide bridge is shown as two larger balls at the lower right side. (B, bottom) Schematic representation of the histidine triad charge relay system. The location of the protons of His213, His199, His228, and W1 are shown for a transient state. In this state, a partial negative charge resides on both Glu-P_1 and W1.

in addition to its interaction with W1 also accepts hydrogen bonds from Val231-N and Thr232-N which form the N-terminal rim of α-helix 2. The significance of this interaction is the observation that α-helices do possess microdipoles resulting from the placement and, to some extent, the alignment of the intrinsic dipole moments of the peptide bond

Enzyme group	Enzyme name	CHYM#--> P1	78-	9	19 012--34	5	6	20 78901	21 234---567---8901	23456	22 ----78-901-2	--345	6	789
Large Pancreatic Proteases	Tryp-SG	R/K	GV-	D	TCQ--GD	S	G	GPMFR	RKDNA-DEW---IQVG	IVSWG	----Y--GCARP	--GYP	G	VYT
	Tryp-Bov	R/K	GK-	D	SCQ--GD	S	G	GPVVC	--S---G-----KLQG	IVSWG	----S--GCAQK	--NKP	G	VYT
	CHT-Bov	F/Y/W	GV-	S	SCM--GD	S	G	GPLVC	KKN---GAW---TLVG	IVSWG	----SS-TCS-T	--STP	G	VYA
	Elast-Porc	A	GVR	S	GCQ--GD	S	G	GPLHC	LVN---GQY---AVHG	VTSFV	----SRLGCNVT	--RKP	T	VFT
	Elast-HN[1]	V	QA-	G	VCF--GD	S	G	SPLVC	--N---G-----LIHG	IASFV	----RG-GCASG	--LYP	D	AFA
	HepBP-Hum[2]	K	RG-	G	ICN--GD	G	G	TPLVC	--E---G-----LAHG	VASFS	----LG-PC--G	--RGP	D	FFT
	Colla-Crab	Y/L/K	GK-	G	TCD--GD	S	G	GPLNY	--D---G-----LTYG	ITSFG	----AAAGCE-A	--GYP	D	AFT
	RMCP-Rat[3]	F	LR-	A	AFM--GD	S	G	GPLLC	--A---G-----VAHG	IVSYG	-----HPDAK--	---PP	A	IFT
	Cathep-G[4]	F/L	RK-	A	AFK--GD	S	G	GPLLC	--N---------VAHG	IVSYG	-----KSSGV--	---PP	E	VFT
	Granzyme B	D	KR-	A	SFR--GD	S	G	GPLVC	KK----------VAAG	IVSYG	----YK-DGS--	---PP	R	AFT
Small Bacterial Proteases	Glu-SGP	E	---	-	ACSAGGD	S	G	GAHFA	------G-S---VALG	IHSGS	----S--GCSGT	--AGS	A	IHQ
	SGP-A	F/Y/W	---	-	VCAQPGD	S	G	GSLFA	------G-S---TALG	LTSGG	----SG-NCR-T	--GGT	T	FYQ
	SGP-B	F/Y/W	---	-	VCAEPGD	S	G	GPLYS	------G-T---RAIG	LTSGG	----SG-NCS-S	--GGT	T	FFQ
	α-LP	A/G	---	-	ACMGRGD	S	G	GSWIT	TSA---G-----QAQG	VMSGG	NVQSNGNNCGIP	ASQRS	S	LFE
α-virus	SCP	W	RG-	V	GGR--GD	S	G	RPIMD	NS----G-----RVVA	IVLGG	----A--DEG-T	--RTA	L	SVV
SA-V8-like Bacterial Proteases	SA-V8[5]	E	---	-	LSITGGN	S	G	SPVFN	EK----NE-----VIG	IHWGG	----VPNEFNGA			
	SA-ETOX-A[6]	unk	LRY	Y	GFIVPGN	S	G	SGIFN	SN----GE-----LVG	IHSSK	----VS-HLDRE	HQINY	G	VGI
	SA-ETOX-B[6]	unk	SQY	F	GYIEVGN	S	G	SGIFN	LK----GE-----LIG	IHSGK	----GGQHNLPI	-GVFF	N	RKI
	Glu-BL[7]	E	LQY	A	MDIYGGQ	S	G	SPVFE	EQSSSRTNCSGPCSLA	VHTNG	----VYGGSSYN	RGTRI	T	KEV
Flavi-virus	YFN-NS3	R/K	AL-	D	YPS--GT	S	G	SPIVN	RN----GE-----VIG	LYGNG	----ILVGDNSF	VSAIS	Q	TEV
	WNV-NS3	R/K	TL-	D	YPI--GT	S	G	SPIVD	KN----GD-----VIG	LYGNG	----VI.....	.....	.	...
Plant-nepo	TBRV-24K	R/K	HY-	E	SRN--DD	C	G	MIILC	QIK---GKM---RVVG	MLVAG	----KDKTSWAD	IMPPN	T	LAE
Plant-como	CPMV-24K	Q/E	---	-	APIIPED	C	G	SLVIA	HIG---GKH---KIVG	VHVAG	----IQGKIGCA			
Plant-poty	PPV-NIa	Q/E	---	-	ISIKDGH	C	G	LPIVS	TRD---GS-----ILG	LHSLA	----NSTNTQNF			
Pic-aptho	FMDV-3C	Q/E	---	-	AAIKAGY	C	G	GAVLA	KD----GADT--FIVG	THSAG	----GN.....			
Pic-entero	HAV-3C	Q/E	---	-	GEGLPGM	C	G	GALVS	SNQ---SIQN--AILG	IHVAG	----GN.....			
	CVA21-3C	Q/E	---	-	FPIRAGQ	C	G	GIITC	T-----G-----KVIG	MHVGG	----NGSH-GFA			
Pic-rhino	HRV14-3C	Q/E	---	-	YPIKSGY	C	G	GIITC	T-----G-----KIFG	IHVGG	----NG.....			
Pic-entero	HPV1-2A	Y/T/A	---	-	GFESPGD	C	G	GILRC	HH----G------VIG	IITAG	----GE.....			

FIGURE 4: Sequence alignment, adopted in part from Table II of Bazan and Fletterick (1990), aligning residues near the catalytic Ser195 and including those forming the primary specificity site in proteases of known structure. The two top groupings, the pancreatic and bacterial proteases, are aligned on the basis of their 3-D structures. Insertions are placed in loop regions between β-sheet regions marked with a shaded background in the figure.The active site Ser/Cys is enclosed between double lines as are residues 189 and 226 which serve roles in substrate binding. His, Ser, and Thr residues assumed to be involved in substrate binding are double underlined and in bold. The lower half of the figure includes structures for which no 3-D structures have been determined with the exception of the SCP-protease (Tong et al., 1993). In this part of the structure residues have been omitted at the beginning and the end since there is little basis for alignments. Additional sequence references: 1, Powers and Harper (1986); 2, Petersen et al. (1993); 3, Powers et al. (1985); 4, Tanake et al. (1985); 5, Breddam and Meldal (1992); 6, Bailey and Smith (1990); 7, Svendsen and Breddam (1992).

(Åqvist *et al.*, 1991). These microdipoles are oriented such that a net positive charge resides at the N-terminus of the α-helix, and in Glu-SGP this net positive charge might stabilize the negative charge on W1 and ultimately on the substrate carboxyl group.

A proposed transient state of the His triad, upon substrate binding, is shown in Figure 3B, where the protons within this hydrogen-bonding network are depicted as being shared between adjacent residues. In the extreme case, the protons would be shuttled to the left, resulting in a neutral glutamic acid side chain with a full negative charge being present on W1. A direct counter charge for this negative potential cannot be identified. However, a possible role of the His triad hydrogen-bonding network upon substrate binding is to facilitate delocalization of the net negative charge of the substrate through the center of β-barrel 2, such that at least a partial negative charge may reside on the solvent molecule at the end of the β-barrel opposite to the S_1 site (Figure 3B). When substrate is bound, all three histidines are completely shielded from solvent, and no other residues are in position to form hydrogen bonds with this network. Furthermore, the His triad is not found in any other known serine protease structures (Greer, 1990).

DISCUSSION

The S1 substrate binding site of the small bacterial proteases (Glu-SGP, SGP-A, SGP-B, α-LP) is shallower than that of the pancreatic enzymes (chymotrypsin, trypsin, elastase) and may best be described as a groove lying on the protein surface rather than a deep pocket. One edge of this groove in the small bacterial proteases is formed by a two-residue insertion at position 192 (see Figure 4), which also causes the 191–220 disulfide bridge to be located away from the S1 site. In the pancreatic enzymes, this disulfide bridge forms part of the rim of the binding specificity pocket while residue 189 forms its base (Steitz *et al.*, 1970; Birktoft & Blow, 1972). In the pancreatic enzyme trypsin Asp189 provides a counter charge for the substrate (Marquart *et al.*, 1983), a role which in Glu-SGP could be assumed by His213 located at the back of the binding groove. Ser192 and Ser216 form the front and appear to be important in properly orienting the substrate for cleavage. Thus, specificity for basic and acidic substrates is generated by two totally different mechanisms and has been engineered into two unique versions of the S1 site. It is interesting to note that, although the bacterial *S. griseus* trypsin (Tryp-SG) has a pocket identical to that of the pancreatic enzyme (Read & James, 1988), the acidic amino acid specificity pocket of Glu-SGP was built into the simpler binding pocket of SGP-A and SGP-B. Hence, changing the specificity from basic to acidic amino acids involved more than simply changing the counter charge of Tryp-SG from negative to positive.

The lack of a direct counter charge at the S1 site of Glu-SGP was somewhat unexpected (Graf *et al.*, 1987). Komiyama *et al.* (1991) suggested that a negatively charged P1 residue would be stabilized by a positively charged lysine or arginine residue at the S1 site of Glu-SGP. In addition, Graf *et al.* (1987), in order to provide the counter charge for a negatively charged substrate, replaced the specificity-determining amino acid, Asp-189, of the Arg/Lys-specific protease, trypsin, with a lysine in the hope of creating a protease with Glu-SGP-like activity. However, instead of generating a

protein with specificity toward acidic side chains, an enzyme with chymotrypsin-like activity was obtained. In this case, the inability to generate a tight and well-defined binding site for negatively charged substrates was rationalized to be caused by flexibility of the lysine side chain.

The most interesting and unique aspect of the Glu-SGP structure is the presence of the His triad extending through the center of β-barrel 2. As mentioned in the Results section, this hydrogen-bonding network is completely shielded from external solvent, and furthermore, His199, situated in the middle of the triad, is hydrogen bonded exclusively to His213 and His228. The coagulation proteases factors IX and X contain a His at position 199, but otherwise the residues of the His triad are not observed in any other member of the pancreatic family of serine proteases. The uniqueness of this hydrogen-bonding network suggests that it serves a major role in stabilization of the negatively charged Glu-P_1 side chain. It should also be mentioned that Glu-SGP does not utilize glutamine as a substrate (unpublished data), further indicating the importance of electrostatic effects in substrate binding specificity by Glu-SGP. While it is likely that binding of the negatively charged P_1 moiety could induce an increase in the pK_a of His213, it is important to note that all three histidine side chains present in the triad are shielded from bulk solvent in the presence of substrate. Hence, the only proton source is that provided by their respective hydrogen-bonding partners. These considerations give rise to the mechanism proposed in Figure 3B, where the protons of the His triad are being shared to compensate for the negatively charged P_1 moiety. The importance of the electrostatic interaction can be appreciated from studies that show that Glu-SGP is more than 3 orders of magnitude more specific for glutamic acid than glutamine substrates.[5] Considering the similar structures of the glutamine and glutamic acid side chains, it is reasonable to assume that the interactions between these amino acids and Glu-SGP are similar. In addition to the lack of a negative charge, only the interactions between Glu-SGP and the substrate ε2 atom differ. However, the interactions made with the side-chain Glu-P_1-Oε2 may also be made by the Gln-P_1-Nε2 atom, i.e., interacting with solvent atoms and perhaps also hydrogen bonding to Ser216-Oγ. These considerations suggest that the negative charge on the substrate is recognized by Glu-SGP, and the crystal structure reported here suggests that His213 serves a primary role in this process. The pK_a and thus the extent of the assumed positive charge on the imidazole ring of His213 are unknown and in fact cannot be determined by X-ray diffraction analysis. In analogy with other protein systems discussed below, it cannot be excluded that the pK_a of His213 is perturbed by its environment (the His triad) to the extent that His213 carries a full positive charge in the absence of substrate. It should be noted that histidine residues with pK's higher than 9.0 have been observed (Anderson *et al.*, (1990). The unanswered question concerning the role of the His triad and its level of protonation requires additional physical studies such as NMR that currently are in their preliminary stages.

Unlike the His triad, which is unique to Glu-SGP, the C-terminal α-helix is present in all serine proteases of known structures. However, this acid-specific protease is the first instance of an enzyme which appears to utilize the positive microdipole induced at the N-terminus of the α-helix for substrate stabilization. Several examples where hydrogen bonds to the positive end of an α-helix are important in charge stabilization have been described. In the *Salmonella typhimurium* sulfate binding protein (Pflugrath & Quiocho, 1988) a buried sulfate anion is stabilized solely by the dipoles of three α-helices. In addition, phosphate residues such as those found in adenine dinucleotides have been found to be stabilized primarily by interactions with α-helix peptide amides (Hol, 1985). In triose-phosphate isomerase, the pK_a of His95 has been lowered to 4.5 via interactions with an α-helix microdipole (Knowles, 1991). Theoretical calculations have indicated that charge stabilization by an α-helix results primarily from hydrogen-bonding interactions between the charged atom and microscopic dipoles (peptide amido groups) located at the end of the helix (Åqvist *et al.*, 1991). Thus, in Glu-SGP, it may be possible that, by accepting two hydrogen bonds from the N-terminus of α-helix 2, W4 could share a proton or partial positive charge with W1 which could, in turn, share a proton with His228. Taken to the extreme, this mechanism might facilitate shuttling of a partial positive potential through the center of β-barrel 2 to the S_1 binding pocket via the His triad (Figure 3B). In the case of the sulfate binding protein, it was proposed that Ser130 could share its hydroxyl proton with one of the sulfate oxygens yielding HSO_4^- (Pflugrath & Quiocho, 1988) by virtue of the fact that this residue was donating a hydrogen bond to the sulfate dianion while accepting a hydrogen bond from the N-terminus of an α-helix. This would be analogous with the interaction between solvent molecules W4 and W1 found at the base of the His triad in Glu-SGP.

Recently Bazan and Fletterick (1990) suggested that the trypsin-like (or pancreatic) eucaryotic and the bacterial serine proteases may be structurally and functionally related to viral serine or cysteine (where the active site serine has been replaced by a cysteine) proteases. This proposal was based on sequence alignments that focused primarily on residues around the catalytic triad and indicated that this wide range of proteases may all have the same basic catalytic site geometry. However, it is important to note that away from the catalytic site the small bacterial and pancreatic serine proteases display little, if any, sequence homology to the viral proteases. This initial proposal has been substantially reinforced by the recently described structure of the chymotrypsin-like protease from the Sindbis virus core protein (Tong *et al.*, 1993). Although this protease is about 50 and 80 amino acids shorter than the small bacterial and pancreatic proteases, respectively, its structure is very similar to that of chymotrypsin (rms for Cα atoms 1.37 Å) despite little evidence of sequence homology. The structural similarity to the pancreatic-type proteases includes the catalytic residues, the substrate binding site, and the core components of the two β-barrels. Figure 4 has been partly adapted from Bazan and Fletterick (1990) but modified to include proteases with glutamic acid substrate specificity (Glu-SGP, V8 protease, and epidermolytic toxins A and B) as well as other proteases of interest. The primary cleavage specificities of these enzymes have also been included in Figure 4. Further examination of Figure 4 indicates that all viral or bacterial proteases with specificity for cleavage after glutamic acid or glutamine residues have a histidine at position 213, while a basic residue is not observed at position 189. Granzyme B, which prefers to cleave after aspartic acid residues (Odake *et al.*, 1991), has an arginine at position 226. Furthermore, with the exception of HAV-3C, the hydroxyl function at 192 is consistently maintained, most often as a threonine. In addition, the two residues preceding His213 are highly conserved; a glycine is nearly always found at position 211, and a branched hydrophobic chain is usually at position 212. The exceptions are Glu-BL and SCP-protease, which both have alanines at position 212 (see Figure 4). Thus, because

of the high homology around the active site Ser/Cys195, at residues 211 through 213 and at position 192, it is possible that the active site of Glu-SGP may serve as a crude model for the substrate binding site of these Glu/Gln specific enzymes. However, since the sequences of these proteases display little homology away from the active site (Bazan & Fletterick, 1990), it is difficult to propose a more extensive mechanism, such as that provided by the His triad in Glu-SGP. The requirements for charge stabilization would, of course, be of lesser importance for glutamine binding.

In summary, the S_1 site of Glu-SGP has three residues which are important in binding and stabilizing the glutamic acid side chain: His213, Ser192, and Ser216. His213 could serve as the counter charge in stabilization of the negatively charged P_1 side chain while Ser192 and Ser216 should serve to properly position the substrate for catalysis. His213, located at the top of β-barrel 2, is at the beginning of the His triad (His213, His199, His228), which extends through the center of this cylinder, connecting the substrate glutamic acid to a solvent channel at the other end of this β-barrel. This interaction is predicted to contribute to a net negative potential at the base of the His triad which could be counteracted by a net positive potential provided by the C-terminal α-helix microdipole. Finally, because of the striking homology among the mammalian, bacterial, and viral serine or cysteine proteases at the catalytic site, it is possible that the mechanism for acid substrate binding observed in Glu-SGP may also serve as a general crude model for the active site geometry for all of these glutamic acid specific or glutamine-specific enzymes.

REFERENCES

Anderson, D. E., Becktel, W. J., & Dahlquist, F. W. (1990) *Biochemistry 29*, 2403–2408.

Åqvist, J., Luecke, H., Quiocho, F. A., & Warshel, A. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2026–2030.

Bailey, C. J., & Smith, T. P. (1990) *Biochem. J. 269*, 535–537.

Bazan, J. F., & Fletterick, R. J. (1990) *Virology 1*, 311–322.

Birktoft, J. J., & Blow, D. M. (1972) *J. Mol. Biol. 68*, 187–240.

Breddam, K., & Meldal, M. (1992) *Eur. J. Biochem. 206*, 103–107.

Brunger, A. T. (1990) *X-PLOR (version 2.1) Manual*, Yale University, New Haven, CT.

Carter, C. W. (1990) *Methods (San Diego) 1*, 12–24.

Dancer, S. J., Garrett, R., Saldanha, J., Jhoti, H., & Evans, R. (1990) *FEBS Lett. 268*, 129–132.

Fitzgerald, P. M. D. (1988) *J. Appl. Crystallogr. 21*, 273–278.

Fujinaga, M., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1985) *J. Mol. Biol. 184*, 479–502.

Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J., & Rutter, W. J. (1987) *Biochemistry 26*, 2616–2623.

Grant, G. A., & Eisen, A. Z. (1980) *Biochemistry 19*, 6089–6095.

Greer, J. (1990) *Proteins 7*, 317–334.

Hol, W. G. J. (1985) *Prog. Biophys. Mol. Biol. 45*, 149–195.

James, M. N. G., Delbaere, L. T. J., & Brayer, G. D. (1978) *Can. J. Biochem. 56*, 396–402.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.

Knowles, J. R. (1991) *Nature 350*, 121–124.

Komiyama, T., Bigler, T. L., Yoshida, N., Noda, K., & Laskowski, M. (1991) *J. Biol. Chem. 266*, 10727–10730.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W., & Huber, R. (1983) *Acta Crystallogr., Sect. B 39*, 480–490.

Odake, S., Kam, C. M., Narasimhan, L., Poe, M., Blake, J. T., Krahenbuhl, O., Tschopp, J., & Powers, J. C. (1991) *Biochemistry 30*, 2217–2227.

Petersen, L. C., Birktoft, J. J., & Flodgaard, H. (1993) *Eur. J. Biochem.* (in press).

Pflugrath, J. W., & Quiocho, F. A. (1988) *J. Mol. Biol. 200*, 163–180.

Powers, J. C., & Harper J. W. (1986) *Proteinase Inhibitors* (Barrett, A. J., & Salvesen, G. S., Eds.) pp 55–152, Elsevier Science Publishers, Amsterdam and New York.

Powers, J. C., Tanaka, T., Harper, J. W., Minematsu, Y., Barker, L., Lincoln, D., Crumley, K. V., Fraki, J. E., Schechter, N. M., Lazarus, G. G., Nakajima, K., Nakashino, K., Neurath, H., & Woodbury, R. G. (1985) *Biochemistry 24*, 2048–2058.

Read, R. J., & James, M. N. G. (1988) *J. Mol. Biol. 200*, 523–551.

Read, R. J., Fujinaga, M., Sielecki, A. R., & James, M. N. G. (1983) *Biochemistry 22*, 4420–4433.

Roussel, A., & Cambillau, C. (1989) in *Silicon Graphics Geometry Partner Directory (Fall, 1989)* (Silicon Graphics, Ed.) pp 77–78, Silicon Graphics, Mountain View, CA.

Schechter, I., & Berger, A. (1968) *Biochem. Biophys. Res. Commun. 27*, 157–162.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1979) *J. Mol. Biol. 134*, 781–803.

Steitz, T. A., Henderson, R., & Blow, D. M. (1970) *J. Mol. Biol. 46*, 337–348.

Stroud, R. M. (1974) *Sci. Am. 231*, 74–88.

Svendsen, I., & Breddam, K. (1992) *Eur. J. Biochem. 204*, 165–171.

Svendsen, I., Jensen, M. J., & Breddam, K. (1991) *FEBS Lett. 292*, 165–167.

Tanaka, T., Minematsu, Y., Reilly, C. F., Travis, J., & Powers, J. C. (1985) *Biochemistry 24*, 2040–2047.

Thornberry, N. A., Bull, H. G., Calaycay, J. R., Chapman, K. T., Howard, A. D., Kostura, M. J., Miller, D. K., Molineaux, S. M., Weidner, J. R., Aunins, J., Elliston, K. O., Ayala, J. M., Casano, F. J., Chin, J., Ding, G. J. F., Egger, L. A., Gaffney, E. P., Limjuco, G., Palyha, O. C., Raju, S. M., Rolando, A. M., Schmidt, J. A., & Tocci, M. J. (1992) *Nature 356*, 768.

Tong, L., Wengler, G., & Rossmann, M. G. (1993) *J. Mol. Biol. 230*, 228–247.

Walter, J., Steigemann, W., Singh, T. P., Bartunik, H., Bode, W., & Huber, R. (1982) *Acta Crystallogr. 38*, 228–247.

Wellink, J., & van Kammen, A. (1988) *Arch. Virol. 98*, 1.

Yoshida, N., Tsuruyama, S., Nagata, K., Hirayama, K., Noda, K., & Makisumi, S. (1988) *J. Biochem. 104*, 451–456.