Crystal Structure of Papain–Succinyl-Gln-Val-Val-Ala-Ala-*p*-Nitroanilide Complex at 1.7-Å Resolution: Noncovalent Binding Mode of a Common Sequence of Endogenous Thiol Protease Inhibitors†

Atsushi Yamamoto,‡ Kouji Tomoo,‡ Mitsunobu Doi,‡ Hirofumi Ohishi,‡ Masatoshi Inoue,‡ Toshimasa Ishida,*,‡ Daisuke Yamamoto,§ Satoshi Tsuboi,∥ Hiroshi Okamoto,∥ and Yoshio Okada∥

Osaka University of Pharmaceutical Sciences, 2-10-65 Kawai, Matsubara City, Osaka 580, Japan, Biomedical Computation Center, Osaka Medical College, 2-7 Daigaku-machi, Takatsuki 569, Japan, and Faculty of Pharmaceutical Sciences, Kobe-Gakuin University, Nishi-ku, Kobe 673, Japan

Received April 15, 1992; Revised Manuscript Received July 24, 1992

ABSTRACT: Succinyl-Gln-Val-Val-Ala-Ala-*p*-nitroanilide corresponding to a common sequence of endogenous thiol protease inhibitors is a noncompetitive reversible inhibitor of papain. In order to elucidate the binding mode of the inhibitor at the atomic level, its complex with papain was crystallized at ca. pH 7.0 using the hanging drop method, and the crystal structure was analyzed at 1.7-Å resolution. The crystal has space group $P2_12_12_1$, with $a = 43.09$, $b = 102.32$, $c = 49.69$ Å, and $Z = 4$. A total of 47 215 observed reflections were collected on the imaging plates using the same single crystal, and 19 833 unique reflections with $F_o > \sigma(F_o)$ were used for structure determination and refinement. The papain structure was determined by use of the atomic coordinates of papain previously reported, and then refined by the X-PLOR program. The inhibitor molecule was located on a difference Fourier map and fitted into the electron density with the aid of computer graphics. The complex structure was finally refined to $R = 19.6\%$ including 118 solvent molecules. The X-ray analysis of the complex crystal shows that the inhibitor is located at the R-domain side, not in the center of the binding site created by the R- and L-domains of papain. Such a binding mode of the inhibitor explains well the biological behavior that the inhibitor exhibits against papain. Comparison with the structure of papain–stefin B complex indicates that the structure of the Gln-Val-Val-Ala-Gly sequence itself is not necessarily the essential requisite for inhibitory activity. On the basis of the present results, the inhibitory role of the Gln-Val-Val-Ala-Gly sequence common to the thiol protease inhibitors is discussed.

Lysosomal thiol proteases with a highly reactive Cys residue at the active site are abundant in living cells and play important roles in intracellular proteolysis (Kirschke & Barrett, 1987; Katunuma, 1989) such as the catabolism of proteins and peptides (Kominami et al., 1986) and the processing of prohormones (Marks et al., 1986) or in the extracellular degradation of collagen (Delaisse et al., 1984). It is now well recognized that the imbalance of their enzymatic activities causes serious diseases such as muscular dystrophy (Katunuma & Kominami, 1987), osteoporosis (Delaisse et al., 1984), and tumor invasion (Denhardt et al., 1987). Therefore, the development of low-molecular inhibitors that can specifically control the proteolytic activities of thiol proteases is a special requirement for potent therapeutic drugs.

In order to design useful inhibitors, the most effective approach is to analyze the biological functions of endogenous thiol protease inhibitors. Recently many endogenous thiol protease inhibitors have been isolated from various animal tissues (Barrett et al., 1986) and are believed to play a significant role in protecting the organism against uncontrolled action of endogenous or exogenous thiol proteases by virtue of the inhibitors' complementary reversible binding. These inhibitors have been named cystatins and they constitute the cystatin superfamily of proteins. There are two well-conserved regions in the amino acid sequences of the cystatin superfamily: a Gly residue in the N-terminal region and a Gln-Val-Val-Ala-Gly (QVVAG) sequence in the central region (Jerala et al., 1989). These two conserved sites have been presumed to be important in binding to the thiol proteases and/or for inhibitory activity. Recent X-ray crystal structure analysis of chicken egg white cystatin (Bode et al., 1988) and of the complex of recombinant human stefin B and papain (a thiol protease) (Stubbs et al., 1990) suggested this hypothesis. In these structures, the two conserved regions were shown to form a wedge which slots into the papain active site, where the QVVAG region forming a β-hairpin loop is located at the papain S1′ and S2′ subsites and the N-terminal region is directed toward the S1–S3 subsites.

In spite of this reasonable interpretation, however, some questions have been raised concerning the biological roles of the conserved N-terminal G residue and the central QVVAG sequence. For example, Nikawa et al. (1989) and Jerala et al. (1990) evidenced by mutation experiments that the QVVAG sequences of cystatin A and stefin B are not essential for the inhibitory activity against thiol proteases, while Abe et al. (1988) and Arai et al. (1991) reported that the papain-inhibitory activity of oryzacystatin, a member of the cystatin superfamily, depends on the central QVVAG region and does not require the N-terminal region. As there is currently no direct proof at the atomic level that the QVVAG sequence itself is necessary for inhibitory activity, we decided to analyze the crystal structure of the papain–succinyl-Gln-Val-Val-Ala-Ala-*p*-nitroanilide (1) complex, in order to determine the reason for the evolutionary conservation of these two regions.

† The atomic coordinates of the complex structure have been deposited in the Brookhaven Protein Data Bank under file name 1ABP.

* To whom correspondence should be addressed.
‡ Osaka University of Pharmaceutical Sciences.
§ Osaka Medical College.
∥ Kobe-Gakuin University.

0006-2960/92/0431-11305$03.00/0

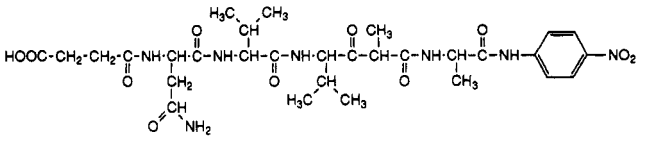

FIGURE 1: Chemical structure of succinyl-Gln-Val-Val-Ala-Ala-*p*-nitroanilide (**1**), together with the label of each residue.

Compound **1**, derived from a series of QVVAG analogues, exhibits potent reversible and noncompetitive inhibition against papain (Okada et al., 1988). The chemical structure of **1** is given in Figure 1.

MATERIALS AND METHODS

Preparation of Complex and Crystallization. Papain was purchased from Sigma. Twice-crystallized papain was prepared by the method of Kimmel and Smith (1954) and further purified according to Sluyterman and Wijdenes (1970). Inhibitor **1** was chemically synthesized using the conventional liquid-phase method and its purity was checked by amino acid and elemental analyses. According to the measurement of Barrett (1977), the inhibition rate (K_i) of **1** against papain was $K_i < 10^{-5}$ M, suggesting that the binding affinity of inhibitor **1** to papain is relatively high compared with other low molecular weight inhibitors such as leupeptin and that complex crystals would be obtained from a sample solution in which an excess of the inhibitor coexists with papain. Confirmation as to whether the crystals obtained were of the complex or not was obtained by measurement of their CD spectra; the CD spectrum of papain alone is noticeably modified by the coexistence of the inhibitor (Okada et al., 1988).

Purified papain (9.1 mg/mL) in an aqueous solution containing 0.1 M KCl, 20 mM EDTA, and 0.1 M cysteine (pH adjusted to 7.0) was incubated for 10 min at 37 °C with a 10-fold molar excess of **1** dissolved in a minimal amount of dimethyl sulfoxide. After incubation, the complex solution was concentrated to 53 mg/mL by ultrafiltration (Amicon YM2 membrane).

Crystallization experiments were performed at 20 °C using the hanging drop vapor diffusion method. The sample solution was prepared by adding 2 M aqueous NaCl (1.2 μL) and a methanol/ethanol (2:1 v/v) mixture (10 μL) to the condensed complex solution (5 μL) and was equilibrated against a 0.1 M aminoethanol/HCl buffer (pH 8.9) containing a 64% methanol/ethanol (2:1 v/v) mixture.

X-ray Data Collection and Processing. Platelet crystals were grown in the same droplet after ca. 5 days. Single crystals sealed in thin-walled glass capillaries with a small amount of mother liquor were used for X-ray work. Precession photographs showed the crystals to be orthorhombic space group $P2_12_12_1$ with the cell constants of $a = 43.09$, $b = 103.32$, and $c = 49.69$ Å, indicating molecular packing similar to that of the papain–E-64-c (an irreversible thiol protease inhibitor) complex (Kim et al., 1992). One complex is contained per symmetric unit of the crystal. X-ray intensity data were collected from a single crystal with dimensions of 0.5 × 0.6 × 0.4 mm^3 using a Rigaku R-Axis IIC diffractometer equipped with two imaging plates. A total of 47 215 reflections [$I > \sigma(I)$] up to 1.7-Å resolution were measured by rotating the crystal from 0° to 90° about the a^* axis with the rotation interval of 2.5°/imaging plate. These data were merged and scaled to yield 19 833 unique reflections, comprising 87.4% of all data expected at 1.7-Å resolution. R_{merge} ($\Sigma|I - \langle I\rangle|/\Sigma|I|$) was 0.045. The scale factors and overall thermal factor were estimated by means of the Wilson plot.

Structure Solution and Refinement. Molecular replacement was attempted as a means to determining the complex structure, where the atomic coordinates of papain in the papain–E-64-c complex (Kim et al., 1992) were used as a search structure. Fortunately, the molecular orientation and translation of this model in the unit cell were almost isomorphous with those in the present crystal structure, giving an R value ($\Sigma(|F_o| - |F_c|)/\Sigma|F_o|$) of 0.438. Thus, the refinement of this initial structure was carried out using a combination of simulated annealing and conventional restrained refinement methods by the X-PLOR program (Brünger, 1990). Conventional restraints were applied for the bond distances, angles, planar groups, chiral volumes, van der Waals contacts, and individual temperature factors. Each stage of the refinement cycle consisted of (1) the least-squares refinements of (a) a weight for the calculation of structure factors, (b) positional parameters, and (c) isotropic thermal parameters of non-H atoms, (2) the conformational check by the presentation of the Ramachandran plot, and (3) the manual fitting of each atom on the electron density map using the FRODO program (Jones, 1978) on an Iris 2400 Turbo interactive computer graphic system.

The progress of refinement is summarized in Table I. After the first stage of refinement, solvent molecules having electron densities greater than 0.4 e/Å^3 in the map and temperature factors less than 50 Å^2 were picked up. After the eighth stage of refinement, a difference Fourier map was calculated, revealing a continuous electron density corresponding to **1**, as is shown in Figure 2. Because of the relatively weak electron density, some doubts remained concerning the validity of the modal assignment of the N-terminal moiety of **1** using only the difference Fourier map. Thus, the $(2|F_o| - |F_c|)$ map using the atomic coordinates of papain and solvent molecules, except for **1**, was calculated by the X-PLOR program. This map clearly identified all the atomic positions of **1**. The R value was reduced to 0.196 in the last refinement of the ninth stage, where a total of 1807 non-H atoms were included using the 18 383 unique reflections [$F_o > 3\sigma(F_o)$] up to 1.7-Å resolution.

All numerical calculations were carried out on a DEC Vax workstation 3100 computer at the Computation Center of Osaka University of Pharmaceutical Sciences. The atomic coordinates of the complex structure have been deposited into the Brookhaven Protein Data Bank (PDB file name 1ABP), from which copies are available.

RESULTS

Overall Structure. The overall structure of papain and the binding mode of **1** are shown in Figure 3. Papain consists of 212 amino acid residues and is folded into two large domains: an L-domain consisting of residues 10–111 and 207–212 and an R-domain consisting of the remaining residues. The Cys-25 active center of papain is located on the left side (L-domain side) of the cleft created by these two domains. Most of the conformational ϕ/φ torsion angles are found within the outer limit boundary of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) and no significant difference from the isolated papain (Kamphuis et al., 1984) was observed in the secondary structures; the root-mean-square deviation between the corresponding Cα-atoms was 0.35 Å.

Inhibitor **1** was positioned at the R-domain side of the cleft which corresponds to S subsites according to Schechter and Berger (1967); its electron density was clearly distinguished

Table I: Summary of the X-PLOR Refinement

	stage								
	1	2	3	4	5	6	7	8	9
resolution (Å)	40–1.7	40–1.7	40–1.7	40–1.7	40–1.7	40–1.7	40–1.7	40–1.7	10–1.7
$F_o/\sigma(F_o)$	1	1	1	1	1	1	1	3	3
no. of reflections	19833	19833	19833	19833	19833	19833	19833	18515	18383
no. of atoms (solvent)	1655 (0)	1693 (38)	1718 (63)	1713 (58)	1719 (64)	1710 (55)	1732 (77)	1773 (118)	1807 (118)
R(initial)	0.438	0.272	0.266	0.257	0.246	0.247	0.251	0.299	0.232
R(final)	0.282	0.254	0.249	0.249	0.243	0.246	0.247	0.237	0.196

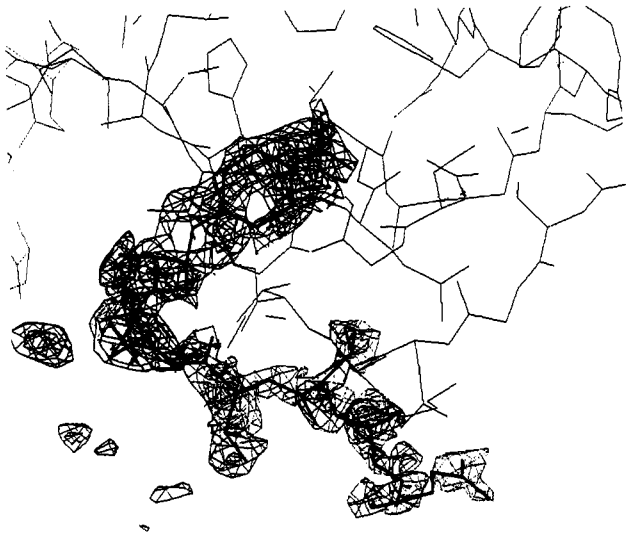

FIGURE 2: Electron density map of **1** binding to the R-domain interface of the papain S active site in the final stage of refinement. Possible positioning of **1** on the map is also shown with thick lines. Thin lines represent the amino acid residues of papain.

from those of the symmetry-related neighboring papains and no atomic short contacts less than 3.5 Å were observed. **1** did indeed project its pNA moiety into the inner part of the R-domain interface. The N-terminal moiety of **1** was relatively less defined than the C-terminal one on the electron density map, probably because the former region is located at the entry of the cleft and is bordered with solvent molecules.

Binding Mode of Inhibitor **1**. A stereoscopic view of the interface between the inhibitor and papain is shown in Figure 4. Figure 5 (a) illustrates schematically its interaction mode for the sake of easy interpretation, and the significant atomic interactions are summarized in Table II. Two oxygen atoms of the pNA-7 I nitro group of **1** are hydrogen-bonded to the amide $N^{\epsilon}H_2$ group of Gln-142. One of them further participates in an electrostatic interaction with the Ala-137 NH group. The NH group of Ala-6 I of **1** is linked with the Asp-158 C=O group by a hydrogen-bond formation. The marked electron density observed for the Ala-Ala-pNA sequence of **1** is probably due to the fixation of this moiety through these interactions with papain. In the Suc-Gln-Val-Val moiety of **1**, on the other hand, a relatively weak hydrogen-bond formation (Gln-2 I C=O···Lys-156 $N^{\zeta}H$) is possible. Furthermore, this moiety is held by hydrophobic interactions with papain Lys-156 and Val-157 residues.

In order to estimate the binding fitness of inhibitors to papain, the difference accessibility was calculated for each residue and the results are shown in Figure 6. This value is based on the difference between the ASA values of papain structures with and without inhibitor **1**, where ASA represents the accessible surface area given by Lee and Richards (1971). The residues interacting with the inhibitor, especially Lys-156, Val-157, and Asp-158 main and side chains, showed significant decreases in their accessibility, suggesting that these residues comprising the R-domain interface are largely shielded from the solvent phase by binding with **1**. This result illustrates the tight binding of inhibitor **1** to papain and explains its relatively high inhibitory activity ($K_i < 10^{-5}$ M). In conclusion, the present X-ray analysis clarified that **1** binds to the R-domain side of the cleft composed of papain S subsites.

DISCUSSION

The complex crystal structure of papain with inhibitor **1** can explain well the biological phenomena that **1** and its analogues exhibit an inhibitory activity against papain. Some examples are given in Table III. Significant decrease in the inhibitory activity due to OMe substitution (**2**) for the pNA of **1** is probably due to the lack of hydrogen-bonding ability with the Gln-142 residue. The absence of a specific interaction of the Gln-2 I side chain with the papain binding site indicates that the inhibitory activity of **3** is almost the same as that of **1**. The results given in Table III also indicate the importance of hydrophobic interactions between the Suc-Gln-Val-Val region of **1** and the papain Lys-156 and Val-157 residues for prominent inhibitory activity, because inhibition against papain decreases drastically as the distance between the Suc and pNA groups decreases. On the other hand, Okada et al. (1988) reported that **1** acts as a noncompetitive inhibitor against papain, based on the kinetic analysis according to the Lineweaver–Burk plot. The binding mode shown in Figures 3 and 4 can explain why **1** inhibits papain noncompetitively. The cleft created by the R- and L-domains has a large space, and the binding of **1** to the R-domain interface would allow access of other inhibitors to the Cys-25 active center located at the L-domain interface, depending on the concentrations of inhibitors; in other words, the binding of **1** to papain would not be large enough to completely block the access of other inhibitors to the Cys-25 residue.

As previously mentioned in the introduction, the Gly residue in the N-terminal region and the QVVAG sequence in the middle portion of thiol protease inhibitors are well conserved and have been presumed to be important for inhibitory activity. The X-ray crystal analysis of papain–stefin B complex (Stubbs et al., 1990) showed that both of the conserved regions play an important role in inhibiting enzymatic activity by forming a wedge-shaped tertiary structure which projects into the papain active site. A schematic representation of the interaction mode of the QVVAG region in stefin B with papain is shown in Figure 5 (b). It is obvious from Figure 5 that the binding mode of inhibitor **1** is quite different from that of the QVVAG region in stefin B, although both bind to the cleft comprising the papain active site. This fact clearly indicates that papain does not have a binding pocket selective and specific for the QVVAG sequence; in other words, there is no specific interface complementary only for the QVVAG amino acid arrangement. It appears reasonable to consider that the QVVAG sequence itself, which is conserved in thiol protease inhibitors, is not necessarily the structural requisite for inhibitory activity but that the β-hairpin loop structure that

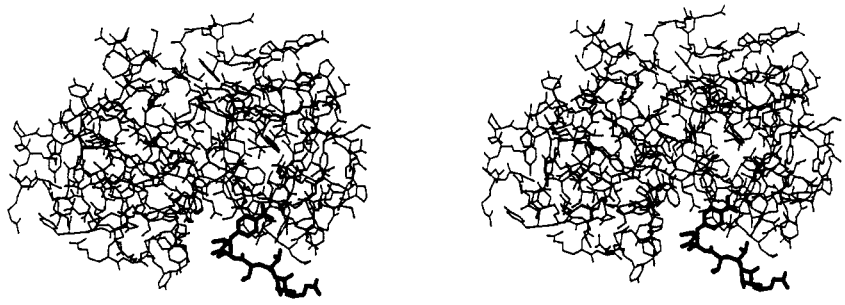

FIGURE 3: Stereoscopic view of the overall structure of papain complexed with **1**. The bold solid lines represent **1**.

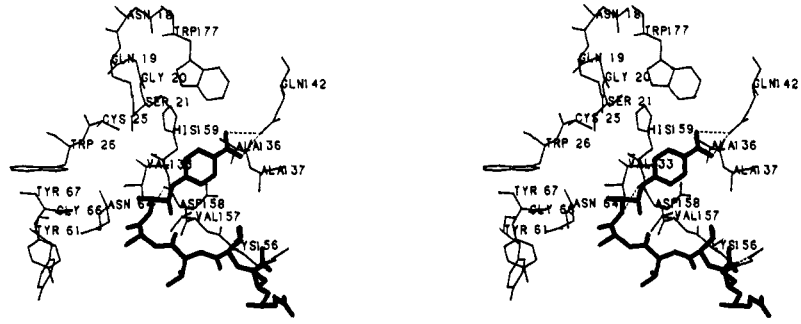

FIGURE 4: Stereoscopic drawing of binding mode of **1** to papain R-domain. Possible hydrogen bonds or electrostatic interactions are shown with dotted lines.

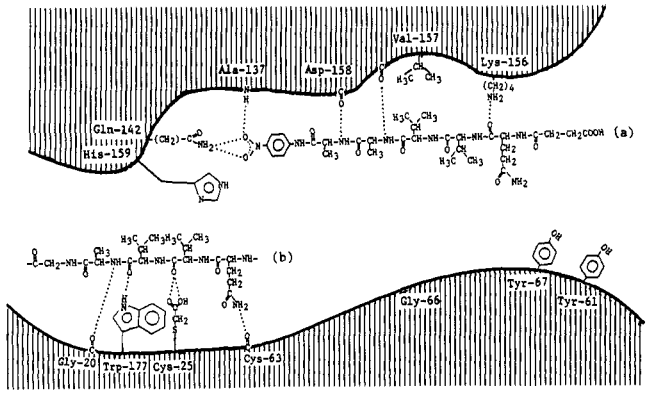

FIGURE 5: Schematic representation of the interaction modes of **1** (a) and the QVVAG region of stefin B (b) with the papain active site. Possible hydrogen bonds and short contacts are represented with dotted lines.

Table II: Possible Hydrogen Bond or Electrostatic Interactions between **1** and Papain Binding Site

1 residue	atom	papain residue	atom	distance (Å)
Gln-2 I	O	Lys-156	$N^{\zeta}H$	3.27
Ala-5 I	O	H_2O^a	OH	3.20
	NH	Val-157	O	3.81
Ala-6 I	O	H_2O	OH	3.60
	NH	Asp-158	O	2.95
pNA-7 I	O1[b]	Gln-142	$N^{\epsilon}H$	2.72
	O1	Ala-137	NH	2.80
	O2[b]	Gln-142	$N^{\epsilon}H$	3.07

[a] Water molecule of solvent. [b] O1 and O2 represent two nitro oxygen atoms of pNA.

this sequence forms in the tertiary structure of the inhibitor plays an important role. The contradicting results concerning the inhibitory roles of the conserved N-terminal Gly residue and the central QVVAG sequence would also be explained by considering their spatial positions in the tertiary structure. The conserved region spatially located nearest to the active Cys residue of thiol protease, which depends on the interaction mode of each tertiary structure of thiol protease and its inhibitor, would be indispensable for the inhibitory activity.

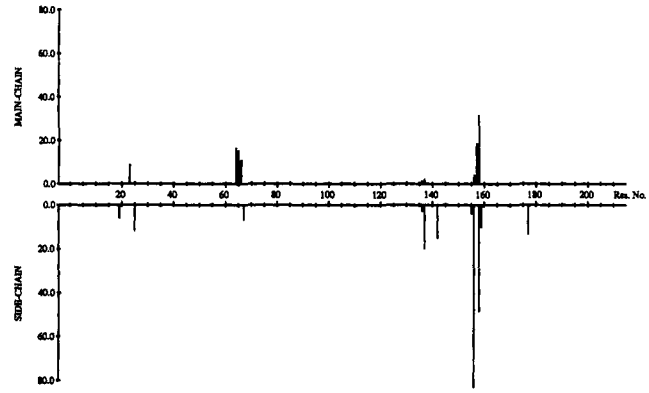

FIGURE 6: Difference of accessible surface area (Å^2) of main chain (top) and side chain (bottom) of each amino acid residue due to binding with **1**.

Table III: Effects of **1** and Its Derivatives on the Amidolytic Activity of Papain[a]

	compound	inhibition of papain[b] (%)
1	Suc-Gln-Val-Val-Ala-Ala-pNA	90.0
2	Suc-Gln-Val-Val-Ala-Ala-OMe	58.0
3	Suc-Ala-Val-Val-Ala-Ala-pNA	87.0
4	Suc-Gln-Val-Val-Ala-pNA	18.2
5	Suc-Gln-Val-Val-pNA	8.6
6	Suc-Gln-Val-pNA	0

[a] Odaka et al. (1988). [b] The inhibitory effects of the peptides were determined by adding the peptides (0.18 mM) to the assay medium. The values represent the means of four experiments.

REFERENCES

Abe, K., Emori, Y., Kondo, H., Arai, S., & Suzuki, K. (1988) *J. Biol. Chem. 263*, 7655–7659.

Arai, S., Watanabe, H., Kondo, H., Emori, Y., & Abe, K. (1991) *J. Biochem. (Tokyo) 109*, 294–298.

Barrett, A. J. (1977) in *Proteinases in Mammalian Cells and Tissues* (Barrett, A. J., Ed.) pp 181–208, North-Holland Publishing, Amsterdam.

Barrett, A. J., Rawlings, N. D., Davies, M. E., Machleidt, W., Salvesen, G., & Turk, V. (1986) in *Proteinase Inhibitors*

(Barrett, A. J., & Salvesen, G., Eds.) pp. 515–569, Elsevier, Amsterdam.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., & Turk, V. (1988) *EMBO J. 7*, 2593–2599.

Brünger, A. T. (1990) X-PLOR Manual, Version 2.1, Yale University, New Haven, CT.

Delaisse, J.-M., Eeckhout, Y., & Vaes, G. (1984) *Biochem. Biophys. Res. Commun. 125*, 441–447.

Denhardt, D., Greenberg, A. H., Egan, S. E., Hamilton, R. T., & Wright, J. A. (1987) *Oncogene 2*, 55–59.

Jerala, R., Trstenjak, M., Kroon-Zitko, L., Lenarcic, B., & Turk, V. (1989) in *Proteinases and Their Inhibitors: Recent Developments* (Auerswald, E., Fritz, H., & Turk, V., Eds.) pp 55–58, Kernforschungsanlage Juelich, Juelich, Germany.

Jerala, R., Trstenjak-Prebanda, M., Kroon-Zitke, L., Lenarcic, B., & Turk, V. (1990) *Biol. Chem. Hoppe-Seyler 371*, 157–160.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

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

Katunuma, N. (1989) in *RBC: Cell Biology Reviews* (Knecht, E., & Grisolia, S., Eds.) Vol. 20, pp 35–61, Springer-Verlag, Berlin.

Katunuma, N., & Kominami, E. (1987) *Rev. Physiol. Biochem. Pharmacol. 108*, 1–20.

Kim, M.-J., Yamamoto, D., Matsumoto, K., Inoue, M., Ishida, T., Mizuno, H., Sumiya, S., & Kitamura, K. (1992) *Biochem. J.* (in press).

Kimmel, J. R., & Smith, E. L. (1954) *J. Biol. Chem. 207*, 515–531.

Kirschke, H., & Barrett, A. J. (1987) in *Lysosomes: Their Role in Protein Breakdown* (Glaumann, H., & Ballard, F. J., Eds.) pp 193–283, Academic Press, London.

Kominami, E., Ohshita, T., & Katunuma, N. (1986) in *Cysteine Proteinases and Their Inhibitors* (Turk, V., Ed.) p 229, Walter de Gruyter, Berlin.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

Marks, N., Berg, M. J., & Benuck, M. (1986) *Arch. Biochem. Biophys. 249*, 489–499.

Nikawa, T., Towatari, T., Ike, Y., & Katunuma, N. (1989) *FEBS Lett. 255*, 309–314.

Okada, Y., Teno, N., Tsuboi, S., Nakabayashi, K., Itoh, N., Okamoto, H., & Nishi, N. (1988) *Chem. Pharm. Bull. 36*, 1982–1989.

Ramakrishnan, C., & Ramachandran, G. N. (1965) *Biophys. J. 5*, 909–933.

Schechter, I., & Berger, A. (1967) *Biochem. Biophys. Res. Commun. 27*, 157–162.

Sluyterman, L. A., & Wijdenes, L. (1970) *Biochim. Biophys. Acta 200*, 593–595.

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