Swillens, S., Van Cauter, E., and Dumont, J. E. (1974), *Biochim. Biophys. Acta 364,* 250–259.
Tesser, G. I., Fisch, H. U., and Schwyzer, R. (1972), *FEBS Lett. 23,* 56–58.
Tesser, G. I., Fisch, H. U., and Schwyzer, R. (1974), *Helv. Chim. Acta 57,* 1718–1730.
Walsh, D. A., Perkins, J. P., Brostrom, C. O., Ho, E. S., and Krebs, E. G. (1971), *J. Biol. Chem. 246,* 1968–1976.
Walsh, D. A., Perkins, J. P., and Krebs, E. G. (1968), *J. Biol. 243,* 3763–3765.
Wilchek, M., and Miron, T. (1974), *Mol. Cell. Biochem. 4,* 181–187.
Wilchek, M., Salomon, Y., Lowe, M., and Selinger, Z. (1971), *Biochem. Biophys. Res. Commun. 45,* 1177–1184.

Binding of Chloromethyl Ketone Substrate Analogues to Crystalline Papain†

J. Drenth,* K. H. Kalk, and H. M. Swen

ABSTRACT: Papain (EC 3.4.22.2) is a proteolytic enzyme, the three-dimensional structure of which has been determined by x-ray diffraction at 2.8 Å resolution (Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., and Wothers, B. G. (1968), *Nature (London) 218,* 929–932). The active site is a groove on the molecular surface in which the essential sulfhydryl group of cysteine-25 is situated next to the imidazole ring of histidine-159. The main object of this study was to determine by the difference-Fourier technique the binding mode for the substrate in the groove in order to explain the substrate specificity of the enzyme (P_2 should have a hydrophobic side chain (Berger and Schechter, 1970)) and to contribute to an elucidation of the catalytic mechanism. To this end, three chloromethyl ketone substrate analogues were reacted with the enzyme by covalent attachment to the sulfur atom of cysteine-25. The products crystallized isomorphously with the parent structure that is not the native, active enzyme but a mixture of oxidized papain (probably papain-SO_2^-) and papain with an extra cysteine attached to cysteine-25. Although this made the interpretation of the difference electron density maps less easy, it provided us with a clear picture of the way in which the acyl part of the substrate binds in the active site groove. The carbonyl oxygen of the P_1 residue is near two potential hydrogen-bond donating groups, the backbone NH of cysteine-25 and the NH_2 of glutamine-19. Valine residues 133 and 157 are responsible for the preference of papain in its substrate splitting. By removing the methylene group that covalently attaches the inhibitor molecules to the sulfur atom of cysteine-25 we obtained acceptable models for the acyl-enzyme structure and for the tetrahedral intermediate. The carbonyl oxygen of the P_1 residue, carrying a formal negative charge in the tetrahedral intermediate, is stabilized by formation of two hydrogen bonds with the backbone NH of cysteine-25 and the NH_2 group of glutamine-19. This situation resembles that suggested for the proteolytic serine enzymes (Henderson, R., Wright, C. S., Hess, G. P., and Blow, D. M. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 63–70; Robertus, J. D., Kraut, J., Alden, R. A., and Birktoft, J. J. (1972b), *Biochemistry 11,* 4293–4303). The nitrogen atom of the scissile peptide bond was found close to the imidazole ring of histidine-159, suggesting a role for this ring in protonating the N atom of the leaving group (Lowe, 1970). This proton transfer would be facilitated by a 30° rotation of the ring around the C^{β}–C^{γ} bond from an in-plane position with the sulfur atom to an in-plane position with the N atom. The possibility of this rotation is derived from a difference electron-density map for fully oxidized papain vs. the parent protein.

Papain (EC 3.4.22.2), a cysteine protease from the fruits of the papaya tree (Carica papaya), has been studied extensively and in a number of reviews (Glazer and Smith, 1971; Drenth et al., 1971) its biochemical and structural properties have been summarized. Although the position of the catalytic site on the enzyme surface is known, the exact interaction of the enzyme with its substrates has not yet been established. It has been generally agreed upon that the hydrolysis of a substrate by papain, like the serine proteases, proceeds via a covalent acyl-enzyme intermediate, according to the general equation:

SCHEME I

$$\text{acylation: } E + S \rightleftarrows ES \rightleftarrows ES^{\dagger}_{tetr} \rightarrow ES^{*} + P_1$$
$$\text{deacylation: } ES^{*} + H_2O \rightleftarrows ES^{\ddagger}_{tetr} \rightarrow E + P_2$$

In this scheme ES is the Michaelis complex, ES* is the acyl-enzyme formed through the Cys-25 thiol group, P_2 is the acid, and P_1 the alcohol or amine moiety of the hydrolyzed substrate. Both acylation and deacylation most likely proceed via a tetrahedral intermediate ($ES^{\dagger}_{tetr}$, $ES^{\ddagger}_{tetr}$ in Scheme I) in which the carbonyl carbon atom undergoing nucleophilic attack has four tetrahedrally arranged bonds. Whereas the evidence for acyl-enzyme formation in the papain-catalyzed hydrolysis pathway is abundant (Lowe, 1970; Smolarsky, 1974), no direct

† From the Laboratory of Structural Chemistry, University of Groningen, Zernikelaan, Groningen, The Netherlands. *Received February 6, 1976.* This research was supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO).

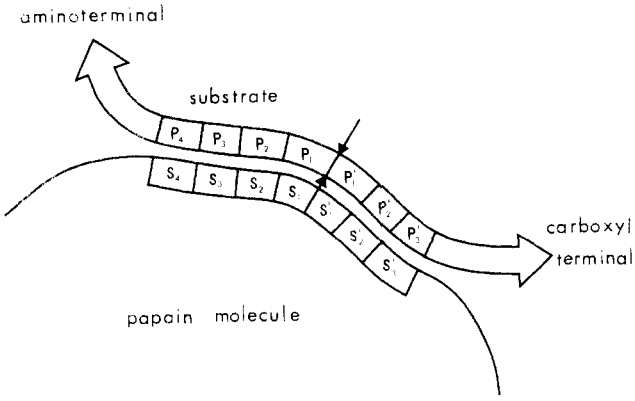

FIGURE 1: Active site of papain according to the concept of Schechter and Berger (1967). The residues of the substrate are denoted by P, the corresponding subsites on the enzyme by S.

evidence is available for the existence of tetrahedral intermediates. The tetrahedral form, however, is generally assumed to exist in nonenzymatic reactions involving a nucleophilic attack on a carbonyl carbon atom (Bruice, 1961; Bender and Kézdy, 1964). Furthermore, structural data in support of the tetrahedral intermediate concept are available for the serine proteases.

Indeed, stable tetrahedral intermediates have been found to exist in the crystals of the bovine trypsin–pancreatic trypsin inhibitor complex (Rühlman et al., 1973; Huber et al., 1974) and the porcine trypsin–soybean trypsin inhibitor complex (Blow et al., 1974; Sweet et al., 1974). For an elaboration of Scheme I, detailed structural information about the interactions between papain and its substrates during the catalytic process is essential. With a polypeptide substrate these interactions may involve binding of up to seven consecutive amino acid residues in the substrate chain to an equal number of complementary subsites in the active site of papain (Schechter and Berger, 1967), as shown in Figure 1. The model of the extended active site for papain, resulting from kinetic studies, was supported by the results of the investigation of the crystal structure of papain by x-ray diffraction (Drenth et al., 1968, 1971; Koekoek, 1969). A deep crevice running across the surface of the molecule and containing the essential Cys-25 and His-159 residues was easily recognized as the active site of the enzyme. In order to establish the identity of the subsites and the nature of the enzyme–substrate interactions, further x-ray studies were necessary. Following a line generally adopted in enzyme crystallography, Wolthers studied by difference-Fourier techniques the binding to papain of some inhibitors, corresponding to the acylating portion of good substrates (Wolthers, 1970; Wolthers et al., 1970). The inhibitors used contained a *p*-iodo-L-phenylalanine residue at P_2, in view of the predilection of the S_2 subsite for hydrophobic residues in general and L-phenylalanine in particular (Schechter and Berger, 1968). Because of a low occupancy and lack of sufficient isomorphism and crystalline order he could only locate the I atom present in the inhibitors; it was found close to Val-133, Val-157, and Ser-205 in the active-site groove. From this observation, he suggested a binding mode for the acyl part of a good substrate.

In this paper we wish to report the results of an x-ray investigation of the binding to papain of some peptide chloromethyl ketone inhibitors, containing L-phenylalanine at P_2. A similar use of peptide chloromethyl ketone inhibitors with subtilisin BPN′ (Robertus et al., 1972a) and γ-chymotrypsin (Segal et al., 1971) provided informative structural data about the enzyme–substrate interactions for these serine proteases. Unlike the serine proteases, which are known to react with this kind of inhibitors through their essential histidine residue, papain reacts through its essential cysteine residue, according to the general equation:

$$R\text{-}C(=O)\text{-}CH_2\text{-}Cl + HS\text{-papain} \rightarrow R\text{-}C(=O)\text{-}CH_2\text{-}S\text{-papain} + HCl$$

Electron density difference maps of the papain derivatives gave a detailed picture of the way of binding of the $R\text{-}C(=O)\text{-}CH_2\text{-}$ moiety in the active site of papain. From the molecular model of this structure the methylene group could be removed without a major change in the position of its neighboring atoms, leading to a plausible structure of the acyl-enzyme.

One of the most significant features in both structures is that the carbonyl oxygen of the P_1 residue is near two potential hydrogen-bond donating groups in the active site of the enzyme, the backbone NH of Cys-25 and the $N^{\epsilon}2$ of Gln-19. This situation bears a close resemblance to that found for subtilisin BPN′ (Robertus et al., 1972b). The groups in question possibly play an important role in the papain-catalyzed hydrolytic process by stabilizing the tetrahedral intermediates. A structural model for these intermediates could be built.

Experimental Section

Benzyloxycarbonyl-L-phenylalanyl-L-alanine chloromethyl ketone (ZPACK[1]) and benzyloxycarbonylglycyl-L-phenylalanylglycine chloromethyl ketone (ZGPGCK) were specially prepared for us by Dr. G. I. Tesser. Acetyl-L-alanyl-L-alanyl-L-phenylalanyl-L-alanine chloromethyl ketone (AcAAPACK) was prepared and kindly supplied by Dr. J. C. Powers (Powers and Tuhy, 1972). Anal. Calcd for ZPACK: Cl, 8.8; C, 62.9; H, 5.3; N, 7.0. Found: Cl, 7.6; C, 63.0;H, 5.8; N, 6.8. Anal. Calcd for ZGPGCK: Cl, 8.0; C, 59.5; H, 5.0; N, 9.5. Found: Cl, 7.9; C, 58.7; H, 5.3; N, 9.3.

Purification of Papain. Dried papaya latex from different sources was used for the preparation of twice-crystallized papain by the method of Kimmel and Smith (1954). This papain preparation was further purified by affinity chromatography according to the method described by Sluyterman and Wijdenes (1970). The inactive nonsulfhydryl papain (PapSX) fraction was concentrated by ultrafiltration to 20 mg/ml, then dialyzed exhaustively against water, and finally precipitated with NaCl (0.2 M) at pH 6.5 and centrifuged. This protein preparation was used for the crystallization of PapSX. The sulfhydryl papain fraction obtained as the $HgCl_2$ derivative (PapSHgCl) was concentrated by ultrafiltration to 20 mg/ml, then dialyzed exhaustively against a 0.5 mM $HgCl_2$, 0.5 mM EDTA solution in water, and finally precipitated with NaCl (0.2 M) at pH 6.5 and centrifuged. This mercury derivative was used for the preparation of the chloromethyl ketone derivatives of papain.

The Preparation of the Chloromethyl Ketone Derivatives. A PapSHgCl solution (20 mg/ml) in 10% Me_2SO, 0.1 M KCl, 50 mM cysteine, 10 mM EDTA, pH 7.0, was treated with a 10% molar excess of the chloromethyl ketone inhibitor, dissolved in the minimal amount of pure dimethyl sulfoxide. Inhibition of the enzyme was found to be complete within a few minutes as was checked by standard assay against α-*N*-ben-

[1] Abbreviations used are: ZPACK, benzyloxycarbonyl-L-phenylalanyl-L-alanine chloromethyl ketone; ZGPGCK, benzyloxycarbonylglycyl-L-phenylalanylglycine chloromethyl ketone; AcAAPACK, acetyl-L-alanyl-L-alanyl-L-phenylalanyl-L-alanine chloromethyl ketone; PapSX, nonsulfhydryl papain; EDTA, (ethylenedinitrilo)tetraacetic acid.

zoyl-L-arginine ethyl ester. The inhibited enzyme was dialyzed exhaustively against water and concentrated by ultrafiltration to 100 mg/ml. This protein solution was used for the crystallization of the papain chloromethyl ketone derivatives.

Crystallization

PapSX. To the PapSX preparation containing about 90% water, methanol was added slowly, while stirring, to reach a final methanol concentration of about 67% (v/v). This protein solution (about 30 mg/ml) was divided over small vials, which were then placed in a desiccator already containing a vial of aminoethanol (technical grade) and at the bottom a 67% (v/v) methanol–water mixture. Under these conditions a volatile base present in the aminoethanol raises slowly the pH of the protein solution by vapor diffusion. Crystals of the right modification (type C) and sufficiently large for a detailed x-ray diffraction study grew within a few weeks at 4 °C.

Chloromethyl Ketone Derivatives. To the pure salt-free protein solution (100 mg/ml) saturated-NaCl solution was added to reach a final concentration of 0.2 M NaCl. Subsequently, methanol was added slowly, while stirring, to reach a final concentration of about 67% (v/v) methanol. From this protein solution (about 30 mg/ml) crystals were grown following the same procedure as described for PapSX. Crystals of the right modification (type C) and sufficiently large for a detailed x-ray diffraction study grew within a few weeks at room temperature. Crystals were mounted in the usual way in the presence of a 67% (v/v) methanol buffer (0.1 M aminoethanol, pH 9.3) mixture. Data were collected to 2.8 Å resolution on an Enraf-Nonius diffractometer, Model CAD-3 (PapSX) and Model CAD-4 (other derivatives), in an ω-scan with Ni-filtered Cu Kα radiation from a sealed-off Philips fine-focus x-ray tube. Crystals were discarded when the intensity of any one out of four standard reflections had dropped by 10%.

Cell parameters were calculated from diffractometer settings and found in all cases to be the same as for crystals of the parent protein within 0.5% or less. Electron density-difference maps were calculated using $m(|F_{\text{deriv}}| - |F_{\text{parent}}|) \exp[i \cdot \alpha_{\text{parent}}]$ as coefficients, in which m, F_{deriv}, and α_{parent} are the figure of merit, the amplitude, and the phase angle of the parent protein structure factor, respectively. The figure of merit m and the phase angle α_{parent} had been obtained by the multiple isomorphous replacement technique (Drenth et al., 1968). Maps were calculated on a 2 cm/1 Å scale, contoured manually and copied on transparent sheets. The first contour level was drawn at about three times the root mean-square difference density, 3σ, calculated over the entire unit cell. The next contour levels were drawn at 1.2σ intervals. Interpretation of the difference maps in the active-site region was facilitated by use of a Richards' box (Richards, 1968) with vertical mirror (Matthews et al., 1972).

Results

The Parent Structure. The x-ray investigation of the structure of papain (Drenth et al., 1968; Koekoek, 1969) was performed with crystals grown from twice-crystallized papain prepared by the method of Kimmel and Smith (1954). Such preparations are known to contain three molecular species of papain: active, activatable, and nonactivatable papain, the only difference between them being the state of the essential Cys-25 thiol group. In the active enzyme this group is present as such, in the activatable enzyme it is in the form of a mixed disulfide with cysteine (PapSScys) and in the nonactivatable enzyme it is transformed most likely to either a sulfinic or sulfonic group (PapSX). The conditions under which the crystals for the x-ray study were grown were such that any active enzyme present at first was converted into either activatable or nonactivatable papain eventually. As a result of cocrystallization, both activatable and nonactivatable papain were present in the parent protein crystal used for the x-ray study, presumably in about a 1:1 ratio (Sluyterman and De Graaf, 1969). Therefore, the parent 2.8 Å electron-density map of papain should contain the structural information of both molecular forms of the enzyme, i.e., PapSScys and PapSX. However, only the activatable form was easily recognized in the map: a bifurcate positive density, continuous with the density corresponding to the Cys-25 residue side chain, unmistakably is representing the extra cysteine attached to Cys-25 through a disulfide bridge. The map was less clear about the nonactivatable form of papain. In order to establish the nature of this molecular species (PapSX) it was prepared in a pure form as the byproduct of the separation of active papain and inactive, nonactivatable papain by the method of Sluyterman and Wijdenes (1970). Crystals of nonactivatable papain grown from aqueous 67% (v/v) methanol were exactly isomorphous with the parent protein crystals.

The electron-density-difference map between pure nonactivatable papain and the parent protein was remarkably clean, except for some peaks and holes close to the positions of the Cys-25 sulfur atom and the His-159 imidazole ring. At the position of the extra cysteine, negative density appeared in the difference map but it did not form a continuous region. Apparently, some ordered water molecules are present in the nonactivatable papain structure at the position of the extra cysteine in activatable papain. Particularly the positions of its carboxyl and amino groups are considered to be occupied by water molecules. Apart from this uncertainty the actual features in the difference map permitted a straightforward interpretation.

From the overlapping positive peaks close to the Cys-25 sulfur position, we concluded that its thiol group is oxidized in nonactivatable papain, most likely to the sulfinic acid (SO_2^-). Another positive peak close to one of the oxygen peaks can then be explained as either a water molecule or a cation compensating for the negative charge on the SO_2^- group. Very interesting was the difference density at the position of the imidazole ring of His-159. The two complementary positive and negative peaks can be explained by a different position of the ring in activatable (PapSScys) and nonactivatable (PapSX) papain, respectively.

In the former compound the ring is pushed "downward" by the extra cysteine, while in the latter it is either free to oscillate by rotating around its β–γ bond between the downward position and a position about 30° "upward", or it is fixed in this upward position. For the terminology upward and downward we refer to the orientation of the molecule as in Figure 2. It should be emphasized that the rotation of the ring between the downward and upward position does not disturb in any way the rest of the protein molecular structure. Thus, when the extra cysteine and oxygen atoms are stripped from activatable and nonactivatable papain, respectively, the remaining protein atoms are in the same position in the two molecular species except for the His-159 imidazole ring (Figure 2a,b).

We assume that this protein structure is also the structure of the native, active enzyme with the Cys-25 thiol group fully reduced. The rotational angle of the His-159 imidazole ring is unknown in the native structure. It has been suggested that in the active enzyme the Cys-25 thiol group and the His-159 imidazole are present as the thiolate–imidazolium ion pair

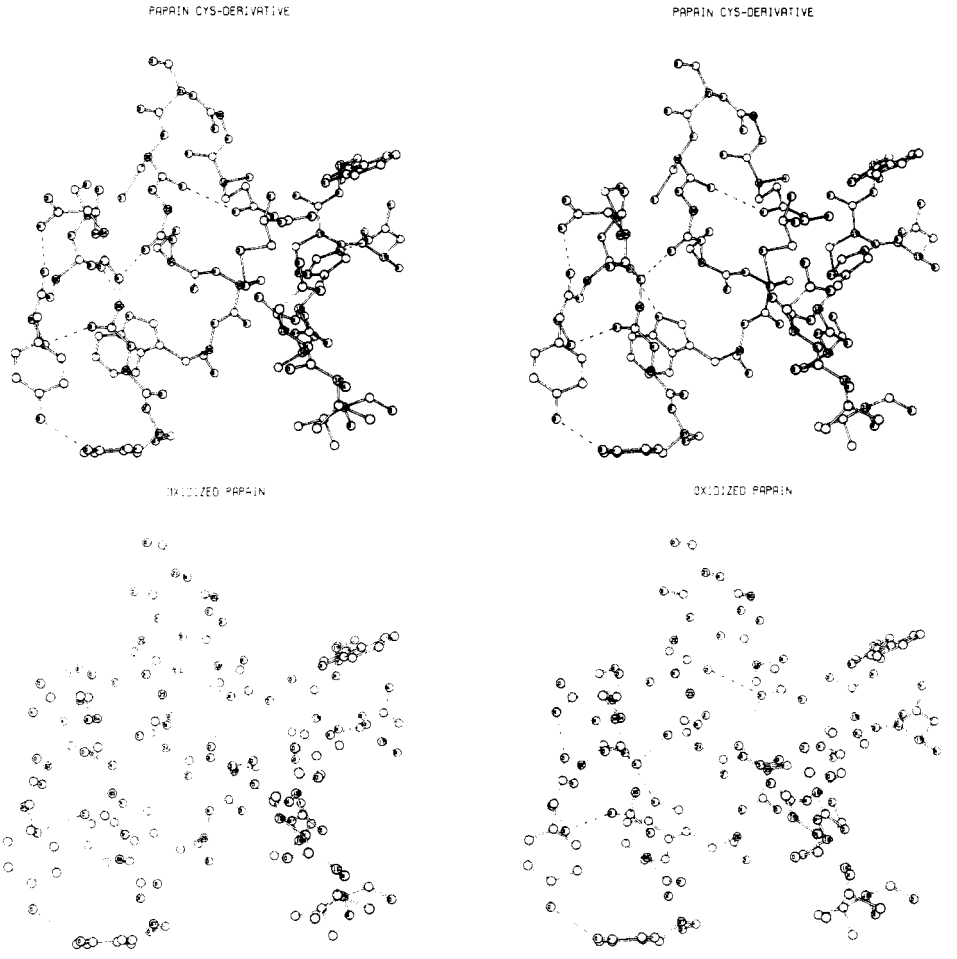

FIGURE 2: The active-site region of papain with the Cys-25 SH. (a) (Top) blocked by an extra Cys residue; (b) (bottom) transformed into a sulfinic acid. Note the orientation of the imidazole ring of His-159. The illustrations are stereoscopic pairs.

(Polgár, 1974; Drenth et al., 1975). This might lead to a preference for the ring being in one plane with the sulfur atom.

Previously, no allowance had been made for the actual state of the Cys-25 residue in the parent crystals. Since this might affect unfavorably the interpretation of the difference maps for the chloromethyl ketone derivatives in the vicinity of the Cys-25 residue, we carefully checked the structure of the active-site region of papain itself as it was derived previously from the 2.8 Å resolution map. A new model for this region was constructed in the Richards' box, using the old parent-electron-density map. The present model was essentially equal to the previous one. However, we now arrived at a somewhat different interpretation of the map for Cys-25: its C^{β} has been moved slightly toward the C^{β} of Ser-176, these atoms now being in van der Waals contact; by a rotation around the C^{α}–C^{β} bond S^{γ} has been brought from its original position to a position where it is virtually in a direct line with the C^{β} of Ser-176 and the C^{β} of Cys-25 (Figure 2a,b); necessary adjustments of neighboring protein atoms were of minor importance. The present conformation of Cys-25 permits without further changes the attachment of an extra cysteine residue fitting nicely in the above mentioned bifurcate density. Likewise, it permits without further adjustment the attachment of two extra oxygen atoms in the two positive peaks close to the Cys-25 S^{γ} position present in the difference map of pure nonactivatable papain vs. parent protein.

Quite unexpectedly we now arrived at a different interpretation of the map at two more places. One concerns a rotation over 180° of the peptide plane between residues Cys-22 and Gly-23 justified by the electron density present for the carbonyl oxygen atom. The other concerns the side chain of Gln-19: $O^{\epsilon}1$ and $N^{\epsilon}2$ have virtually changed places and the position of $N^{\epsilon}2$ is now fixed by a hydrogen bond to the carbonyl oxygen of Cys-22. The rest of the protein atoms in the active site region have remained in essentially the same positions as in the older model. Comparing the old and new model we found a root mean square difference of 0.7 Å for the positions of the atoms in the active-site region not affected by the above model changes. This value is an indication for the precision that can be obtained in building a protein molecular model from an electron density map at 2.8 Å resolution. This new model was used as the reference for the inhibitor derivatives.

The Structure of the Chloromethyl Ketone Derivatives. The difference maps calculated for the papain derivatives of ZPACK, ZGPGCK, and AcAAPACK were very similar in general appearance, the major feature being a well-defined continuous chain of positive density running from the catalytic site near S^{γ} of Cys-25 right through the active-site groove and terminating in the vicinity of the rings of Tyr-61 and 67. This main density was characterized further by a conspicuous side-chain density continuous with it and appearing in the three maps at exactly the same position near Val-133. We interpreted this chain of positive density as the peptide inhibitor covalently attached to S^{γ} of Cys-25. The P_1 and P_2 residues were bound in the active site groove in the same way for the three inhibitors, with the P_2 phenylalanine hydrogen bonded to the Gly-65–Gly-66 backbone chain and its side chain bound

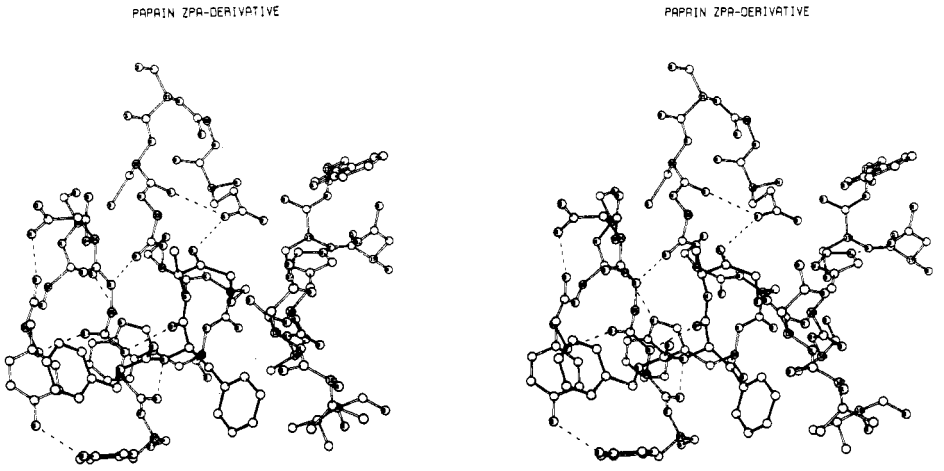

FIGURE 3: The active-site region of papain with the inhibitor benzyloxycarbonyl-L-phenylalanyl-L-alanylmethylene (ZPA). The illustration is a stereoscopic pair.

to a hydrophobic area in the active-site groove. Besides the above major feature, an almost identical pattern of positive and negative density was present in the three maps in the protein region surrounding the active-site groove. Partly, this pattern was consistent with a widening of the groove caused by the interaction of papain with the inhibitors. The remaining features could be explained by either minor adjustments in the protein molecular structure or a rearrangement of the solvent structure. Figure 3 shows the active-site region of papain with the ZPA inhibitor bound covalently to S^{γ} of Cys-25.

We shall now describe in more detail the structure of the papain derivatives as deduced from the respective electron-density difference maps. In this description "higher" or "lower" refers to the structure as viewed in Figure 3. The distribution of the difference electron density near the catalytic site, which is identical for the three papain derivatives, left no room for doubt about the attachment of the inhibitors to the S^{γ} of Cys-25. The positive density in the difference maps corresponding to the methylene group was continuous with the electron density in the parent map corresponding to this sulfur atom. However, to make this linkage the sulfur atom had to be moved outwards about 0.9 Å into the active-site groove along the direct line through C^{β} of Ser-176 and of Cys-25. Positive difference density consistent with this shift of the sulfur atom was present in the three maps, extending the positive density corresponding to the methylene group. Furthermore, negative density close to the original sulfur position confirmed our interpretation. The shift of the sulfur atom together with a shift of C^{β} of Cys-25 required a slight repositioning of neighboring protein atoms. The difference maps gave no decisive answer as to the precise adjustment of the protein structure. Negative density, chiefly present at the main-chain peptide planes of the first turn of the five turn α helix comprising the residues Ser-24 to Gly-43, merely indicated that there had been some structural change to allow for the movement of the Cys-25 side chain. A more specific indication for this change was produced by two complementary positive and negative peaks at either side of the indole ring of Trp-26 from which we concluded that a small translation parallel to the sulfur shift formed part of the structural change to which the first turn of the α helix is subjected. At one more site the maps revealed a small but significant change in the protein structure as a result of the movement of the Cys-25 side chain away from the C^{β} of Ser-176. A negative peak at the position of O^{γ} of Ser-176 together with a positive peak nearby were found to be consistent with a rotation of the oxygen by 120° around the serine C^{α}–C^{β} bond. In its new position the oxygen was at hydrogen-bond distance from the $O^{\epsilon}1$ of Gln-19.

The difference density corresponding to the P_1 carbonyl group was weak, due to the displacement of a nearby solvent molecule present in the parent structure. However, sterical considerations permitted this group to be placed unambiguously near the Gly-23–Ser-24 backbone chain, with the carbonyl oxygen pointing roughly towards the $N^{\epsilon}2$ of Gln-19. The protein region adjacent to the position assigned to the P_1 carbonyl oxygen is pointedly dominated by the backbone NH of Cys-25 and the above-mentioned $N^{\epsilon}2$ of Gln-19. One of the two oxygen atoms of the sulfinic group in nonactivatable papain was also near this protein region. However, neither this oxygen atom nor the P_1 carbonyl oxygen was found to be located optimally with respect to the three potential hydrogen-bond donating amide groups. The P_1 C^{α} atom was in the middle of the active-site groove between the Gly-23 and the Asp-158 carbonyl groups, with the P_1 alanine side chain in both ZPA and AcAAPA right on top of it. Actually the P_1 alanine C^{β} did not manifest itself as a marked positive density bulge in the ZPA and AcAAPA difference maps, but rather as an extension of the positive density corresponding to the P_1 C^{α} atom. This extension was absent in the ZGPG difference map.

Indeed, the latter map showed a small negative peak approximately at the position assigned to the P_1 C^{β} atom. The presence, in the ZPA and AcAAPA difference maps only, of a negative feature higher up in the active-site groove, consistent with the displacement of a solvent molecule by the P_1 alanine side chain, further confirmed our interpretation. The peptide group between the P_1 and P_2 residues was easily positioned into the corresponding difference density, with its NH pointing roughly towards the Asp-158 main-chain carbonyl oxygen and its CO group towards the protein region in between the Trp-26 indole ring and the Gly-65–Gly-66 backbone chain. The position of this peptide group coincided with a marked positive density feature present in the parent protein resembling a methanol molecule. In all three difference maps a water molecule appeared near the P_1NH and the Asp-158 CO groups.

The P_2 carbonyl oxygen appeared to be hydrogen bonded to the backbone NH of Gly-66. A second hydrogen bond was formed between the NH of P_2 and the backbone CO of Gly-66. Difference density consistent with a small adjustment of both the peptide plane between the Gly-65 and -66 residues and the

peptide plane between the Gly-66 and Tyr-67 residues facilitating the formation of these P_2–S_2 hydrogen bonds was observed in the maps. The position of the remaining atoms of the P_2 residue was easily deduced from the difference maps. The C^α atom was placed off the Gly-65–Gly-66 backbone chain, with its free hydrogen atom pointing toward the peptide plane between the residues Asp-158 and His-159. The C^β atom was found to be in van der Waals contact with C^δ of Pro-68 and C^β of Ala-160; its position virtually coincided with a solvent molecule present in the parent protein structure.

The P_2 phenylalanine side chain fitting nicely into its corresponding density was in hydrophobic contact with $C^{\gamma}1$ and $C^{\gamma}2$ of Val-133 and $C^{\gamma}1$ of Val-157. The $C^{\delta}2$ and $C^{\epsilon}1$ atoms of the ring virtually coincided with peaks present in the parent protein map identified as solvent molecules. In addition, the C^{ζ} atom of the ring displaced a nearby solvent molecule, in the parent protein hydrogen bonded to O^{γ} of Ser-205. The binding site for the P_3 residue being the benzyloxycarbonyl group in the ZPA inhibitor, a glycine residue in the ZGPG and alanine residue in the AcAAPA inhibitor were in the vicinity of the side chain of Tyr-67. The P_3 carbonyl group pointed away from the protein towards the solvent. In the ZGPG inhibitor this group approached the protein body a few tenths of an Ångstrom more closely than in the AcAAPA inhibitor. This effect increased for the next three atoms along the backbone of the inhibitors. We attributed this positional difference to the fact that the P_3 alanine side chain was pointing towards the protein making hydrophobic contact with the ring of Tyr-61 and C^α of Gly-65. In the ZGPG inhibitor where this side chain is absent, the C^α of P_3 was found to be in van der Waals contact with the protein. The benzyloxy oxygen atom in the ZPA inhibitor occupied about the same position as the C^α atom of P_3 in the ZGPG inhibitor. The difference density for the remaining atoms of the benzyloxycarbonyl group of the ZPA inhibitor bore a marked resemblance both in setting and appearance to the density for the NH of P_3, and the CO, O, and CH_2 of the benzyloxycarbonyl group of the ZGPG inhibitor. The interpretation was complicated somewhat by the fact that in the ZPA map positive density was missing for the top of the Z ring, while in the ZGPG map the entire Z ring did not show up.

Figure 3 shows the most likely interpretation for the ZPA inhibitor in this region. As for the AcAAPA inhibited enzyme, its difference map gave only weak density beyond its P_3 C^α atom. Besides, positive density was missing for the P_4 NH and the acyl group. Therefore, it seems likely that P_4 is not bound tightly to the protein surface in the present medium.

All three maps showed quite clearly the impact on the protein structure caused by the inhibitors in the active-site groove. Two complementary strings of positive and negative density along the Arg-59 to Pro-68 backbone chain including the Cys-63–Cys-22 disulfide bridge, as well as some features along the Ser-21 to Gly-23 backbone chain, proved to be consistent with a shift of these chains away from the inhibitors. A similar pattern of difference density, although less pronounced, existed along the Lys-156 to His-159 backbone chain indicating that this chain also was pushed aside by the inhibitors. The overall effect was that the active-site groove had widened to accommodate the inhibitors. According to our estimation this widening amounted to about 1 Å.

Discussion

The structure determination of papain and its peptide chloromethyl ketone derivative has been performed on type C crystals equilibrated to a 67% (v/v) methanol buffer (0.1 M aminoethanol, pH 9.3) mixture.

Papain does not exhibit any catalytic activity under these conditions. One might argue, therefore, that the binding mode of the chloromethyl ketone inhibitors to papain observed in the crystalline state is not representative for that in water within the enzyme's pH activity range (pH 4.0–8.0).

Another question that arises concerns the validity of the assumption that the binding mode of these inhibitors to papain bears relevance to that of real substrates during the catalytic process. We have good reason to believe that neither the medium nor the extra methylene group in the ketone derivatives appreciably influence the binding of the peptide chain in the active-site groove. Previously, Wolthers (1970) used x-ray diffraction techniques to study the binding to papain of product inhibitors corresponding to the acylating portion of good substrates, containing a *p*-iodophenylalanine residue at P_2, viz. Ala-Ala-*p*-I-Phe-Arg and Boc-*p*-I-Phe-Arg. In his experiments, Kimmel and Smith papain-type C crystals grown from a 67% (v/v) methanol–water mixture were cross-linked by glutaraldehyde and then transferred to 15% Na_2SO_4. Due to this modification of the conditions the cell dimensions changed and the crystalline order diminished somewhat. However, Wolthers was able to show that the orientation of the molecules relative to the unit cell axes had remained unaltered. Under these new conditions the crystalline enzyme displays normal catalytic activity (Sluyterman and De Graaf, 1969). The enzyme–inhibitor complexes were prepared by soaking activated crystals in 15% Na_2SO_4 containing the inhibitor, at pH 4.3. The iodine atom in both enzyme–inhibitor complexes was found at exactly the same position in the unit cell, with fractional coordinates: $x = 0.45$, $y = 0.63$, $z = 0.32$. This iodine position appears to be very close to the position we assigned to the C^{ζ} atom of the P_2 phenylalanine residue in the present papain–inhibitor derivatives, with fractional coordinates: $x = 0.48$, $y = 0.64$, $z = 0.34$. This observation, combined with the fact that at low pH values the product inhibitors are known (Smolarsky, 1974) to bind covalently to S^{γ} of Cys-25 as the acyl-enzyme intermediate, and considering that definite stereochemical constraints are imposed on both types of inhibitors by the S_1 and S_2 subsites, justifies our conclusion that the binding of the peptide chloromethyl ketones observed by us is a good model for substrate binding.

By removing the extra methylene group, which is between the P_1 carbonyl group and the S^{γ} atom of Cys-25, and keeping the latter atom in its position, it was possible to construct the thio ester linkage to Cys-25, providing the structural model for the acyl-enzyme (Figure 4a). It is emphasized that in order to make the thio ester linkage the structure of the rest of the inhibitor was only perturbed slightly comprising mainly a small shift of the atoms of the P_1 residue. The P_1 carbonyl group in the acyl-enzyme structure is still in the protein region containing the Cys-25 NH and Gln-19 NH_2 potential hydrogen-bond donating groups. The main difference between the acyl-enzyme structure proposed here and those by Wolthers et al. (1970) and by Lowe and Yuthavong (1971) concerns the orientation of the thio ester group. In their models the P_1 carbonyl oxygen is pointing roughly toward the His-159 imidazole ring. Another difference concerns the P_1NH group that in the models proposed by Wolthers et al. (1970) and Lowe and Yuthavong (1971) is hydrogen bonded to the Asp-158 main chain CO group. In the peptide chloromethyl ketone derivatives of papain, a water molecule appears near the P_1 NH and Asp-158 CO and is probably hydrogen bonded to either group, suggesting that the distance between them is too wide for a direct P_1–S_1 hydrogen bond. It is, of course, possible that in the real acyl-enzyme structure the widening of the groove at

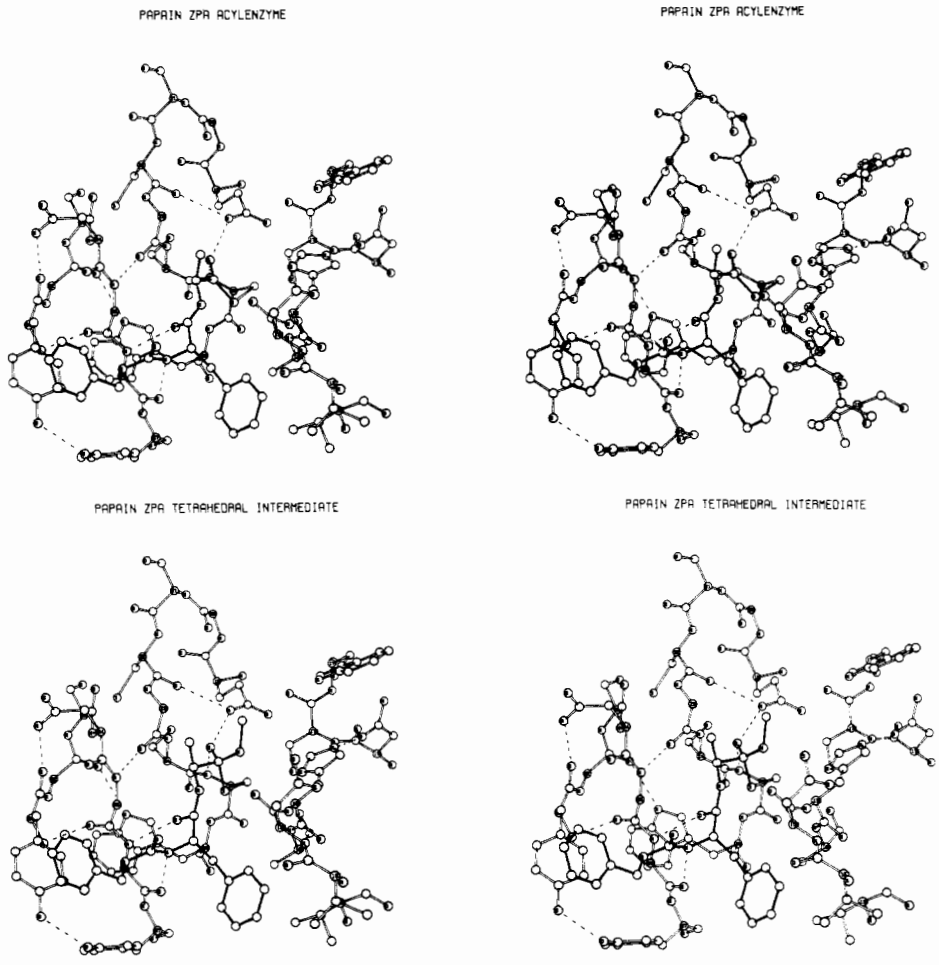

FIGURE 4: (a) (Top) the proposed structure for the ZPA acyl-enzyme; (b) (bottom) the proposed structure for the ZPA tetrahedral intermediate. The illustrations are stereoscopic pairs.

S_1 is less than found in the present investigation and that there this hydrogen bond is indeed present. Finally, the widening of the groove, already predicted by Lowe and Yuthavong (1971) at S_1, is shown by the present work to occur along the entire binding area, most conspicuous in the region near Tyr-61 and -67.

The stereospecificity requirements for the substrate as given by Berger and Schechter (1970) are for P_1 and P_2 in agreement with our results. For P_3 the side chain of an L-residue is found pointing towards the protein body and it is difficult to assign an absolute stereospecificity to the enzyme for this residue, also in agreement with Berger and Schechter's results. Our results do not indicate a strong binding for P_4, although this might be partly due to the medium. The position of the phenylalanine side chain near Val-133 and -157 nicely explains the specificity of the enzyme for a hydrophobic residue at P_2 (Berger and Schechter, 1970).

Starting from the acyl-enzyme structure we obtained the tetrahedral intermediate structure by modifying the carbonyl carbon into a tetrahedral carbon atom, keeping the sulfur atom in its position and the carbonyl oxygen in its binding site (Figure 4b). Also the rest of the substrate chain did not move. The additional fourth bond of the tetrahedral carbon atom has to be perpendicular to the plane through the sulfur atom, the carbonyl oxygen and the C^{α} of P_1. In this way the position of the nitrogen atom of the scissile peptide bond could easily be located. It was interesting to find it close to the imidazole ring of His-159. This suggests a role for the imidazole ring in protonating the N atom of the leaving group as has been proposed by Lowe (1970).

The resulting model for the tetrahedral intermediate confirms, as far as papain is concerned, the generally accepted idea that an enzyme is designed to bind most tightly to a substrate when it is in the transition state of the reaction catalyzed by the enzyme (Pauling, 1946). First, the carbonyl oxygen of the scissile peptide bond is found to be optimally located with respect to the NH of Cys-25 and the NH_2 of Gln-19. In this connection it is interesting to note that a number of other cysteine proteinases are known to have in their primary structure a Gln residue at the same position relative to their essential Cys residue as papain. These proteinases are ficin (Wong and Liener, 1964), stem bromelain (Husain and Lowe, 1968), actinidin (Carne and Moore, 1975), and streptococcal proteinase (Liu et al., 1965). Second, the NH of the scissile peptide bond is in a position to form a hydrogen bond with the Asp-158 main chain CO, while its lone pair is pointing towards the $N^{\delta}1$ of the imidazole of His-159. In the active enzyme this ring is protonated and the sulfhydryl group of Cys-25 is deprotonated (Drenth et al., 1975) and it seems reasonable to assume that the ring and the sulfur atom are fixed in one plane. However, when the S^- attacks the carbonyl carbon atom of the substrate the ring could rotate "upward" by 30° around the C^{β}–C^{γ} bond coming in one plane with the N atom of the leaving group and in exactly the right position for donating its proton to this N atom. The possibility of a rotation by 30° follows from our x-ray crystallographic studies of papain de-

rivatives with the essential sulfur atom blocked by a cysteine residue or oxidized to a sulfinic group. The rotation of the ring does not require a disconnection of the hydrogen bond from the $N^{\epsilon}2$ atom to Asn-175, because this bond is roughly colinear with the C^{β}–C^{γ} bond. In this respect papain differs from the serine enzymes because there the imidazole ring is hydrogen bonded through its $N^{\delta}1$ atom and this does not allow an easy rotation of the ring.

In the proteolytic serine enzymes the substrate chain is in a stretched conformation and connected in an antiparallel way to an extended segment of the backbone chain in the enzyme by three hydrogen bonds. In papain the inhibitor chain forms two hydrogen bonds with an extended segment of backbone chain, consisting of residues 65–67. Here, the two chains are perpendicular to each other and the hydrogen bonds are present at the point of crossing, connecting P_2 with Gly-66. A striking similarity with the proteolytic serine enzymes is the following. The P_1 carbonyl oxygen of the substrate chain is found in papain near an oxygen binding site consisting of the amide group of Gln-19 and the main chain NH group of Cys-25. In subtilisin and in chymotrypsin a very similar binding site for this carbonyl oxygen has been found.

Acknowledgments

We are indebted to Mr. J. Spoelstra and Mr. H. Nijland for skillful technical assistance. To Professor J. N. Jansonius we are grateful for many stimulating discussions and to Drs. J. H. Ploegman for the processing of the oxidized papain data.

References

Bender, M. L., and Kézdy, F. J. (1964), *J. Am. Chem. Soc. 86,* 3704–3714.

Berger, A., and Schechter, I. (1970), *Philos. Trans. R. Soc. London, Ser. B: 257,* 249–264.

Blow, D. M., Janin, J., and Sweet, R. M. (1974), *Nature (London) 249,* 54–57.

Bruice, T. C. (1961), *Proc. Natl. Acad. Sci. U.S.A. 47,* 1924–1928.

Carne, A., and Moore, C. H. (1975), (personal communication).

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M., and Wolthers, B. G. (1968), *Nature (London) 218,* 929–932.

Drenth, J., Jansonius, J. N., Koekoek, R., and Wolthers, B. G. (1971), *Adv. Protein Chem. 25,* 79–115.

Drenth, J., Swen, H. M., Hoogenstraaten, W., and Sluyterman, L. A. AE. (1975), *Proc. Kon. Akad. Wetensch. Amsterdam B78,* 104–110.

Glazer, A. N., and Smith, E. L. (1971), *Enzymes, 3rd Ed. 3,* 501–546.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., and Steigemann, W. (1974), *J. Mol. Biol. 89,* 73–101.

Henderson, R., Wright, C. S., Hess, G. P., and Blow, D. M. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 63–70.

Husain, S. S., and Lowe, G. (1968), *Chem. Commun.,* 1387–1389.

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

Koekoek, R. (1969), Ph.D. Thesis, Rijksuniversiteit Groningen, The Netherlands.

Liu, T. Y., Stein, W. H., Moore, S., and Elliott, S. D. (1965), *J. Biol. Chem. 240,* 1143–1149.

Lowe, G. (1970), *Philos. Trans. R. Soc. London, Ser. B: 257,* 237–248.

Lowe, G., and Yuthavong, Y. (1971), *Biochem. J. 124,* 107–115.

Matthews, B. W., Jansonius, J. N., Colman, P. M., Schoenborn, B. P., and Dupourque, D. (1972), *Nature (London), New Biol. 80,* 37–41.

Pauling, L. (1946), *Chem. Eng. News 24,* 1375–1377.

Polgár, L. (1974), *FEBS Lett. 47,* 15–18.

Powers, J. C., and Tuhy, P. M. (1972), *J. Am. Chem. Soc. 94,* 6544–6545.

Richards, F. M. (1968), *J. Mol. Biol. 37,* 225–230.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., and Wilcox, P. E. (1972a), *Biochemistry 11,* 2439–2449.

Robertus, J. D., Kraut, J., Alden, R. A., and Birktoft, J. J. (1972b), *Biochemistry 11,* 4293–4303.

Rühlmann, A., Kukla, D., Schwager, P., Bartels, K., and Huber, R. (1973), *J. Mol. Biol. 77,* 417–436.

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

Schechter, I., and Berger, A. (1968), *Biochem. Biophys. Res. Commun. 32,* 898–902.

Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R., and Wilcox, P. E. (1971), *Biochemistry 10,* 3728–3738.

Sluyterman, L. A. AE, and De Graaf, M. J. M. (1969), *Biochim. Biophys. Acta 171,* 277–287.

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

Smolarsky, M. (1974), *Isr. J. Chem. 12,* 615–629.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., and Blow, D. M. (1974), *Biochemistry 13,* 4212–4228.

Wolthers, B. G. (1970), Ph.D. Thesis, Rijksuniversiteit Groningen, The Netherlands.

Wolthers, B. G., Drenth, J., Jansonius, J. N., Koekoek, R., and Swen, H. M. (1970), in: Proc. Int. Symp. Structure-Function Relationships of Proteolytic Enzymes Desnuelle, P., Neurath, H., and Ottesen, M., Ed., Munksgaard, Copenhagen, Denmark pp 272–288.

Wong, R. C., and Liener, I. E. (1964), *Biochem. Biophys. Res. Commun. 17,* 470–474.