Astbury, W. T., Dickinson, S., and Bailey, K. (1935), *Biochem. J. 29*, 2351.
Beaven, G. H., Gratzer, W. B., and Davies, H. G. (1969), *Eur. J. Biochem. 11*, 37.
Bradbury, E. M., Brown, L., Downier, A. R., Elliott, A., Fraser, R., Hanby, W. E., and McDonald, T. R. (1960), *J. Mol. Biol. 2*, 276.
Davidson, B., Tooney, N., and Fasman, J. D. (1966), *Biochem. Biophys. Res. Commun. 23*, 156.
Geddes, A. J., Parker, K. D., Atkins, E. D., and Beighton, E. (1968), *J. Mol. Biol. 32*, 343.
Gerber, B. R., and Noguchi, H. (1967), *J. Mol. Biol. 26*, 197.
Iizuka, E., and Yang, J. T. (1966), *Proc. Nat. Acad. Sci. U. S. 55*, 1175.
Klostermeyer, H., and Humbel, R. E. (1966), *Angew. Chem., Int. Ed. Engl. 5*, 807.
Koltun, W. L., Waugh, D. F., and Bear, R. S. (1954), *J. Amer. Chem. Soc. 76*, 413.
Miyazawa, T., and Blout, E. (1961), *J. Amer. Chem. Soc. 83*, 712.
Parker, K. D., and Rudall, K. M. (1957), *Nature (London) 179*, 905.
Reithel, F. J. (1963), *Advan. Protein Chem. 18*, 159.
Sarkar, P. K., and Doty, P. (1966), *Proc. Nat. Acad. Sci. U. S. 55*, 982.
Senti, F. R., Eddy, C. R., and Nutting, G. C. (1943), *J. Amer. Chem. Soc. 65*, 2472.
Timasheff, S. N., Susi, H., Townend, R., Stevens, L., Gorbunoff, M. J., and Kumosinski, T. F. (1967), *in* Conformation of Biopolymers, Ramachandran, G. N., Ed., New York, N. Y., Academic Press, pp 173–196.
Townend, R., Kumosinski, T. F., Timasheff, S. N., Fasman, G. D., and Davidson, B. (1966), *Biochem. Biophys. Res. Commun. 23*, 163.
Waugh, D. F. (1944), *J. Amer. Chem. Soc. 66*, 663.
Waugh, D. F. (1957), *J. Cell. Comp. Physiol. 49*, 145.

An X-Ray Crystallographic Study of the Binding of Peptide Chloromethyl Ketone Inhibitors to Subtilisin BPN'†

Jon D. Robertus, Richard A. Alden, Jens J. Birktoft, Joseph Kraut,*
James C. Powers,‡ and Philip E. Wilcox§

ABSTRACT: The difference-Fourier method was used to elucidate the mode of binding of polypeptide chloromethyl ketone inhibitors to the proteolytic enzyme subtilisin BPN'. Four inhibitors with an L-phenylalanine chloromethyl ketone group at their C termini were investigated: benzyloxycarbonyl-L-alanyl-L-phenylalanine chloromethyl ketone, benzyloxycarbonyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone, benzyloxycarbonylglycylglycyl-L-phenylalanine chloromethyl ketone, and acetyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone. All were found to alkylate the catalytic site His-64 with their phenylalanine side chain fitting snugly into a hydrophobic crevice. In each case the inhibitor polypeptide chain forms a system of hydrogen bonds of the antiparallel β sheet type with an extended segment of backbone chain in the enzyme consisting of residues 125–127. Furthermore, this extended backbone segment comprises one wall of the phenylalanine side-chain-binding crevice. The assumption that this geometry is a valid model for binding of polypeptide substrates is supported by the convincing way in which it explains the observed specificity of the enzyme toward the first, second, and third residues preceding the cleaved peptide bond in such substrates. When the corresponding catalytic and binding sites, considered as a single entity, are compared between subtilisin and chymotrypsin, it is found that they are sterically indistinguishable at current levels of resolution. Since the overall folding of the two molecules is otherwise entirely different, it is surmized that the two classes of serine proteases represent an example of convergent molecular evolution.

With determination of the crystal structure of the bacterial protease subtilisin BPN' now well advanced (Wright *et al.*, 1969; Alden *et al.*, 1971), it becomes possible to apply the powerful difference-Fourier technique of X-ray crystallography to the problem of how this enzyme interacts with its substrates. There is, however, a fundamental limitation inherent in any attempt to study enzymic mechanisms by X-ray crystallographic methods. Enzyme reactions occur on a time scale in the neighborhood of 10^{-8} sec, whereas several days are required to collect adequately informative X-ray data. It would therefore seem to be beyond present capabilities to "look" directly at any individual step in an enzymic reaction by X-ray techniques. Nevertheless a number of obvious stratagems are available that permit the crystallographer to obtain detailed structural information relevant to substrate binding, information which is unobtainable by any other means and which at least provides a basis for construction of reasonable hypotheses regarding the actual enzymic mechanisms.

For purposes of brief discussion these stratagems can be

† From the Department of Chemistry, University of California at San Diego, La Jolla, California 92037. *Received January 24, 1972.* This work was supported by research grants from the National Institutes of Health (GM 10928, GM 16717) and the National Science Foundation (GB 15684, GB 23054) and by a Public Health Service Research Career Development Award to R. A. A. from the National Institute of General Medical Sciences (GM 15401). A preliminary report has been published in *Cold Spring Harbor Symp. 36* (1971).

‡ Department of Chemistry, Georgia Institute of Technology, Atlanta, Ga. 30332.
§ Department of Biochemistry, University of Washington, Seattle, Wash. 98105. Deceased Nov 2, 1971.

classified under three headings, although clearly there is no strict logical distinction between them. One approach might be to modify the enzyme molecule itself in such a way as to inactivate it while minimizing alteration of its structure, and then examine how substrates are bound to the inactive molecule. The simplest such modification for the particular case of subtilisin is inactivation by lowering the pH, presumably causing protonation of His-64 in the catalytic site. Unfortunately, our crystals are unstable at pH below 5 where there is a significant reduction in activity. Another modification of subtilisin that may be useful in such studies is conversion of the catalytic serine residue Ser-221 to cysteine (Neet *et al.*, 1968; Polgar and Bender, 1969).

A second general line of attack becomes possible (A. R. Fersht, 1971, personal communication) if one can find a reversible enzyme-catalyzed reaction whose equilibrium constant is in the neighborhood of unity. If, furthermore, the crystal is stable in the presence of roughly equimolar and sufficiently high concentrations of both products and reactants, it should be possible to obtain a time-averaged picture of the entire course of the reaction. Such a picture would probably be too complicated to interpret without the aid of additional data, but might be very informative in conjunction with other kinds of experiments. To our knowledge, this latter approach has not yet been used in crystallographic experiments.

A third approach, the one that has in fact been most often applied, might be described under the heading of inhibitor studies. Here the object is to bind to the enzyme, either covalently or noncovalently, a molecule that resembles as closely as possible a true substrate, but which remains bound rather than undergoing further reaction. Clearly one must be cautious in drawing conclusions from such experiments as the failure of the bound molecule to undergo the normal enzyme catalyzed reaction necessarily implies that there is something wrong with the way in which it interacts with the enzyme. A wide range of choices exists for the small molecule, each with its own characteristic advantages and disadvantages, for example, poor substrates, virtual substrates, product inhibitors, and irreversible covalent inhibitors. It is this last approach that has been applied in the series of studies on subtilisin BPN′ reported here. Specifically, we have studied the binding of four chloromethyl ketone analogs of good polypeptide substrates. In each the carboxyl group of a C-terminal L-phenylalanine residue is replaced by a chloromethyl ketone group. These reagents are essentially just extensions of benzyloxycarbonyl-L-phenylalanine bromomethyl ketone, shown by Shaw and Ruscica (1968) to alkylate a single histidine residue in subtilisin BPN′ with concomitant loss of activity. The residue was subsequently identified as His-64 (Markland *et al.*, 1968). Morihara and Oka (1970) later found that extending the polypeptide portion at the N terminus by inserting L-Ala-Gly increased reaction rates by a factor of over 200. All four reagents employed in this study are also found to alkylate the catalytic site His-64, with their polypeptide chains bound in a manner that provides a convincing explanation for the observed specificity[1] of the enzyme toward P_1, P_2, and P_3 (Morihara *et al.*, 1969–1971).

Enzymes of the subtilisin family have long been known to bear a marked resemblance to chymotrypsin in both mechanism and specificity (Markland and Smith, 1971). The beginnings of a stereochemical rationale for the similarity of mechanism was provided by the discovery that a similar geometrical arrangement of key groups at the catalytic site (the charge relay system) exists in the two enzymes (Blow *et al.*, 1969; Alden *et al.*, 1970). However, no resemblance at all was found in the overall folding of the two molecules, suggesting that they evolved independently to converge upon the same catalytic mechanism. Moreover, simple inspection of the subtilisin model initially failed to reveal a structural feature analogous to chymotrypsin's "tosyl hole," now identified as the hydrophobic specificity cavity (Steitz *et al.*, 1969; Henderson, 1970). Nevertheless, as will be shown in this communication, an analogous hydrophobic specificity cavity does indeed exist in subtilisin, and what is more, the stereochemistry of polypeptide substrate binding by subtilisin is exactly the same as that deduced from parallel studies on γ-chymotrypsin (Segal *et al.*, 1971). Thus a structural basis is now established for the similarity in specificity, as well as mechanism, between the two enzymes and the concept of convergent molecular evolution is further reinforced.

Methods and Materials

Benzyloxycarbonylglycylglycyl-L-phenylalanine chloromethyl ketone and acetyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone were prepared as previously described by Segal *et al.* (1971). Benzyloxycarbonyl-L-alanyl-L-phenylalanine chloromethyl ketone and benzyloxycarbonyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone were a gift from Dr. K. Morihara. Acetyl-L-tyrosine ethyl ester was obtained from Aldrich Chemical Co.

Subtilisin BPN′ was purchased from the Enzyme Development Corp., New York, N. Y. Subtilisin Novo was purchased from Novo Industries, Copenhagen. Subtilisin BPN′ and subtilisin Novo are crystallographically identical and can be used interchangeably (Robertus *et al.*, 1971). The commercial product is rather impure, containing degradation products and colored contaminants which cannot be removed by gel filtration on Sephadex, and which inhibit crystal growth. The crude powder was dissolved (10 mg/ml) in cold 0.05 M sodium acetate buffer (pH 5.9), immediately brought to 65% saturation with solid ammonium sulfate, and allowed to stand overnight at 5°. Subsequent steps were also carried out in the cold. The suspension was then centrifuged and the precipitate resuspended in 0.01 M potassium phosphate buffer (pH 6.5). This solution was passed over Sephadex G-25 equilibrated with 0.01 M potassium phosphate (pH 6.5). The eluted protein was immediately placed onto a Whatman CM52 carboxymethylcellulose column previously equilibrated with 0.01 M potassium phosphate. A gradient of 0.01–0.10 M potassium phosphate (pH 6.5) was applied to separate the components (Polgar and Bender, 1969).

The elution profile measured by absorption at 278 nm showed four peaks, the first containing the colored material and the fourth and major peak containing the active subtilisin. The contaminants were not further investigated. The purified enzyme fraction, at a concentration of 2 mg/ml was made approximately 0.1 M in potassium phosphate (pH 7) by addition of an appropriate amount of solid salt, assuming the eluent from the CM-cellulose column to be 0.05 M in potassium phosphate. A tenfold molar excess of chloromethyl ketone dissolved in sufficient dioxane to yield a final dioxane concentration of 2.5% was added and allowed to react for 24–36 hr at room temperature. Assay against acetyl-L-tyrosine ethyl ester (Schwert and Takenaka, 1955) performed on the reaction mixture showed no detectable activity (<1%).

[1] For definition of the notation P_1, P_2, P_3, . . . and S_1, S_2, S_3, . . . , see Schechter and Berger (1967).

Preparation of the ZAP[2] derivative differed from the others in that the chloromethyl ketone was added to a 2.5-mg/ml solution of subtilisin BPN′ obtained by redissolving already crystallized active enzyme. These active crystals, prepared as previously described (Wright *et al.*, 1969), were dissolved by addition of water to the crystals suspended in mother liquor, followed by dropwise addition of 0.1 N NH_4OH to pH 10 at 5°. The enzyme was concentrated by precipitating with ammonium sulfate, then resuspended in pH 5.9 sodium acetate buffer, and passed over Sephadex G-25. The eluent, at 5 mg/ml, was diluted 1:2 with 0.2 M sodium phosphate (pH 7.2). The pH was adjusted to 7 and a 20-fold molar excess of ZAPCK added in sufficient dioxane to make the final solution 1% dioxane.

When inhibition was complete each protein was precipitated with solid ammonium sulfate, the precipitate redissolved in 0.05 M sodium acetate buffer (pH 5.9) and desalted on a Sephadex G-25 column. The eluent was concentrated by vacuum dialysis to 15 mg/ml and crystallized as previously described (Wright *et al.*, 1969). Usable crystals were obtained after 1–2 months.

Crystals were mounted in the usual way and data collected on a Hilger and Watts four-circle automatic diffractometer (Wright *et al.*, 1969). A single crystal was used to collect all reflections to 2.5-Å resolution in the positive k hemisphere of reciprocal space for each derivative.

Cell parameters were calculated from diffractometer settings and found in all cases to be the same as for active crystals within 1% or less. Electron density difference maps were calculated using $F_{deriv} - F_{BPN'}$ as coefficients and centroid phases from the original structure determination of phenylmethanesulfonyl subtilisin BPN′ (Wright *et al.*, 1969). The notation F_{deriv} refers to reflection amplitudes for crystals of a given derivative of subtilisin, and $F_{BPN'}$ to reflection amplitudes for the native active subtilisin BPN′ crystals. Maps were calculated on a 1 Å × 0.7 Å × 1 Å grid and plotted on a scale of 2 cm/Å, with the lowest contour level at three times the root-mean-square difference density, 3σ, taken over all points in the unit cell. This 3σ value ranged from 0.12 e/Å^3 to 0.15 e/Å^3. Subsequent contour levels were drawn at 1.5σ intervals. Interpretation of the difference maps was facilitated by use of an optical comparator (Richards, 1968).

Fourier maps using $2F_{deriv} - F_{PMS\cdot BPN'}$ coefficients and centroid phases for PMS·BPN′ were also calculated for the ZAP and ZAGP derivatives. This map should tend to show the correct electron density distribution of the inhibited enzyme itself, minimizing spurious effects due to the use of PMS·BPN′ phases. These maps were plotted on a 1-cm/Å scale with the lowest contour level at 0.1 e/Å^3 and subsequent levels contoured at 0.1 e/Å^3 intervals. The maps were then transferred onto transparent sheets with a Xerox copier, and the sheets affixed to Plexiglas plates for vertical spacing.

TABLE I: Cartesian Coordinates of AAGP Subtilisin.[a]

Atom	Residue	x	y	z
CH_2	P_1 Phe	18.4	25.7	21.0
C	P_1 Phe	18.8	27.2	20.8
O	P_1 Phe	19.3	27.6	19.7
Cα	P_1 Phe	18.5	28.5	21.7
Cβ	P_1 Phe	19.6	29.3	22.2
Cγ	P_1 Phe	18.7	30.5	22.5
Cδ1	P_1 Phe	17.4	30.7	22.0
Cε1	P_1 Phe	16.7	31.9	22.2
Cζ	P_1 Phe	17.2	32.9	23.0
Cε2	P_1 Phe	18.5	32.7	23.6
Cδ2	P_1 Phe	19.1	31.6	23.3
N	P_1 Phe	17.8	28.1	22.9
C	P_2 Gly	16.5	28.0	23.2
O	P_2 Gly	15.6	28.1	22.4
Cα	P_2 Gly	16.1	27.7	24.6
N	P_2 Gly	14.8	28.3	25.0
C	P_3 Ala	14.7	29.6	25.6
O	P_3 Ala	15.6	30.3	25.7
Cα	P_3 Ala	13.3	29.0	26.1
Cβ	P_3 Ala	12.4	30.3	24.8
N	P_3 Ala	13.3	31.1	26.8
C	P_4 Ac	12.4	31.3	27.9
O	P_4 Ac	11.6	30.4	28.3
Cα	P_4 Ac	12.6	32.7	28.6
N	His-64	19.3	19.4	23.8
Cα	His-64	19.9	20.8	23.6
Cβ	His-64	18.9	21.8	24.2
Cγ	His-64	19.1	23.2	23.4
Nδ1	His-64	19.8	24.3	24.0
Cε1	His-64	19.7	25.3	23.1
Nε2	His-64	18.9	24.9	22.1
Cδ2	His-64	18.5	23.7	22.3
C	His-64	21.2	20.7	24.6
O	His-64	22.2	21.2	24.2
N	Ser-221	22.9	29.1	19.4
Cα	Ser-221	23.0	28.2	20.5
Cβ	Ser-221	21.6	27.5	20.7
Oγ	Ser-221	21.6	26.6	21.6
C	Ser-221	24.1	27.0	20.6
O	Ser-221	24.4	26.3	21.5

[a] x, y, and z are measured in Å units in the Cartesian coordinate system defined by Alden *et al.* (1971).

Results

All of the difference maps were similar in general appearance. That is, the major feature in all cases was a well-defined continuous chain of positive density running from the catalytic site near His-64, along the backbone chain from Ser-125 to Gly-127 and terminating in the vicinity of Tyr-104. We interpreted this chain as the polypeptide inhibitor covalently attached to His-64. The P_1 phenylalanine side chain was bound in a cleft in the same fashion for all inhibitors and the remainder of the polypeptide chain was hydrogen bonded to form an antiparallel β-sheet arrangement with the Ser-125 to Gly-127 backbone chain.

Figure 1a–c show models of the ZAP, ZAGP, and AAGP inhibitors superimposed on the appropriate electron density difference maps. In each case the continuous positive density is seen to be the most prominent feature of the map and the models can be seen to conform quite well to that density.

[2] Abbreviations used are: Z, benzyloxycarbonyl; ZAP, benzyloxycarbonyl-L-alanyl-L-phenylalanine; ZAPCK, benzyloxycarbonyl-L-alanyl-L-phenylalanine chloromethyl ketone; ZAGP, benzyloxycarbonyl-L-alanylglycyl-L-phenylalanine; ZGGP, benzyloxycarbonylglycyl glycyl-L-phenylalanine; AAGP, acetyl-L-alanylglycyl-L-phenylalanine; PMS, phenylmethanesulfonyl; PMS·BPN′, phenylmethanesulfonyl-subtilisin BPN′; TPCK, tosyl-L-phenylalanine chloromethyl ketone; ZPBK, benzyloxycarbonyl-L-phenylalanine bromomethyl ketone.

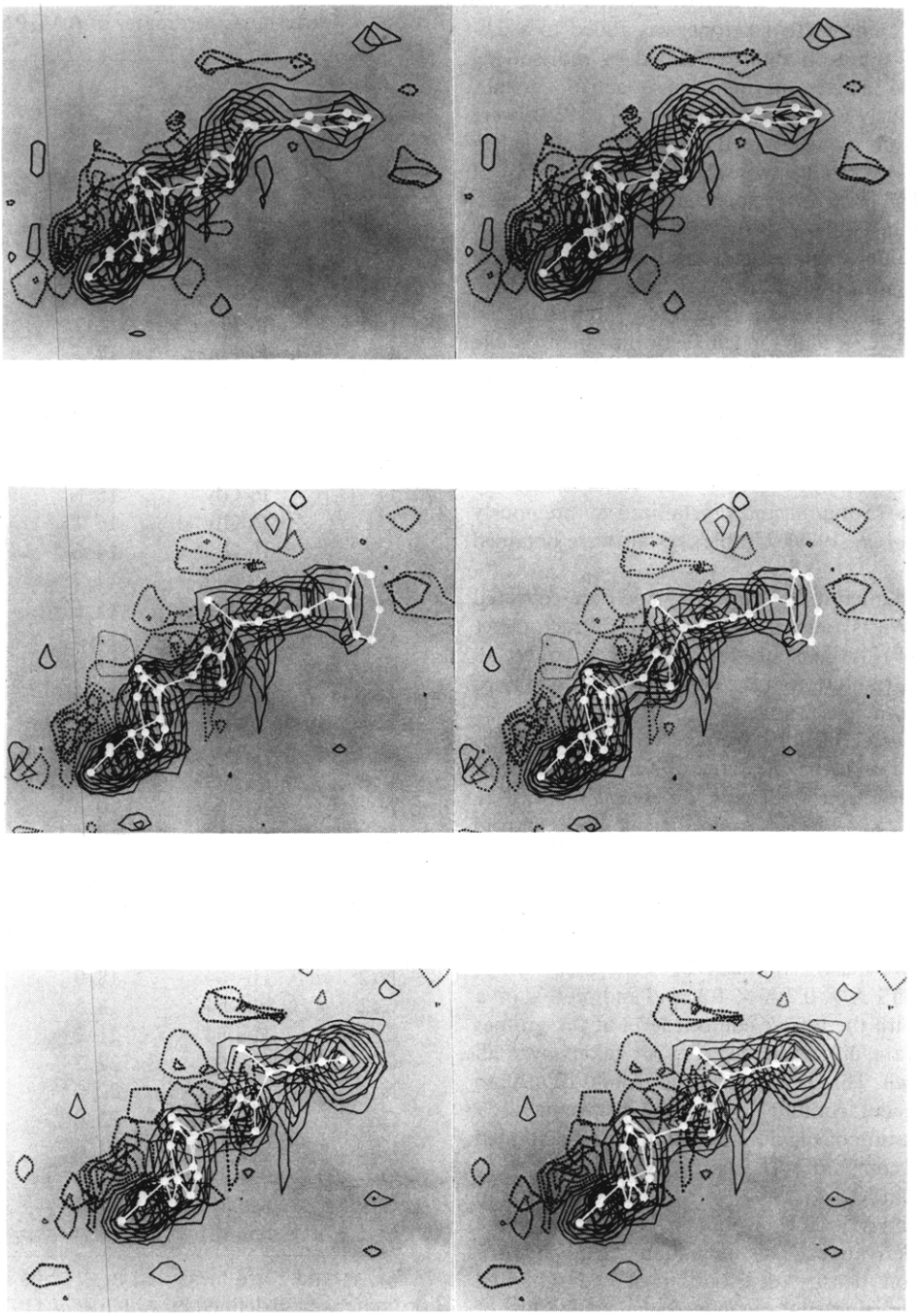

FIGURE 1: 2.5-Å resolution electron density maps illustrating the differences between subtilisin–inhibitor complexes and active subtilisin. Positive and negative electron density is represented by solid and dotted contouring, respectively. The outermost contours are shown at a density of ± 0.12 e/Å^3 with subsequent levels at intervals of ± 0.06 e/Å^3. Superimposed on the appropriate density is an outline of (a, top) benzyloxycarbonyl-L-alanyl-L-phenylalanylmethylene, (b, middle) benzyloxycarbonyl-L-alanylglycyl-L-phenylalanylmethylene, and (c, bottom) acetyl-L-alanylglycyl-L-phenylalanylmethylene moieties. The inhibitor amino-terminal blocking group appears at the upper right. The P_1 phenylalanine carbonyl group is at the lower left, with the methylene carbon covalently attached to the Nϵ2 of His-64 below the carbonyl carbon. The negative peaks to the left and above the inhibitor density indicate the sites for displaced solvent molecules.

Peculiar aspects of each inhibitor which affect this main density will be discussed below.

In addition, almost all peaks above 3σ in electron density difference elsewhere in the maps were the same for all maps, showing that the same minor adjustments in the protein and solvent structure had occurred for all four derivatives.

Stereoscopic drawings of the ZAGP and AAGP inhibitor groups bound to subtilisin BPN′, prepared from ORTEP plots, are shown in Figures 2 and 3, respectively. Cartesian coordinates for the AAGP derivative are given in Table I.

Attachment of the inhibitors to the side chain of His-64 proved to be the most vexing problem in building the models. To make this linkage required a slight rotation of about 10° around the histidine Cα–Cβ bond and a rotation of about 30° around the Cβ–Cγ bond in our present model of the active enzyme. This arrangement, although consistent with the difference density, placed the carbonyl oxygen of the P_1 Phe only 2.8 Å from the Cβ of Ser-221, or about 0.6 Å closer than standard van der Waals contact. It should be emphasized, however, that the electron density corresponding to this

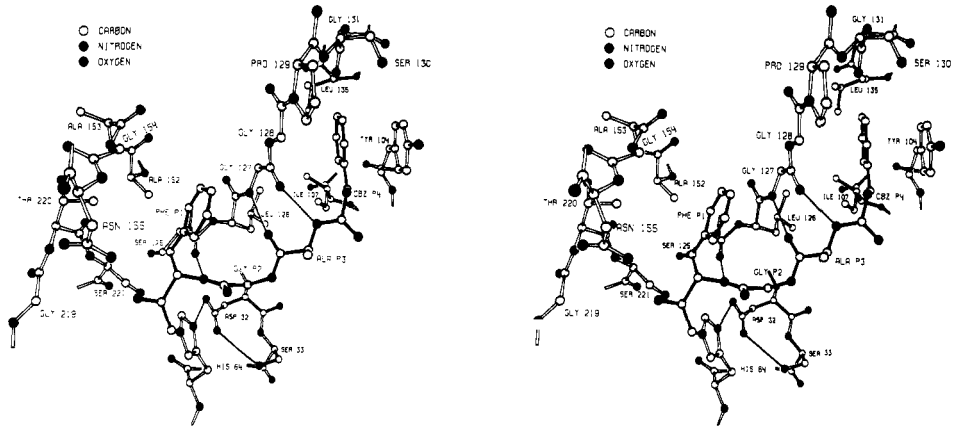

FIGURE 2: The active-site region of subtilisin BPN′ with the inhibitor benzyloxycarbonyl-L-alanylglycyl-L-phenylalanylmethylene covalently bound to Nε2 of His-64. The illustration is a stereoscopic pair. CBZ denotes the benzyloxycarbonyl group.

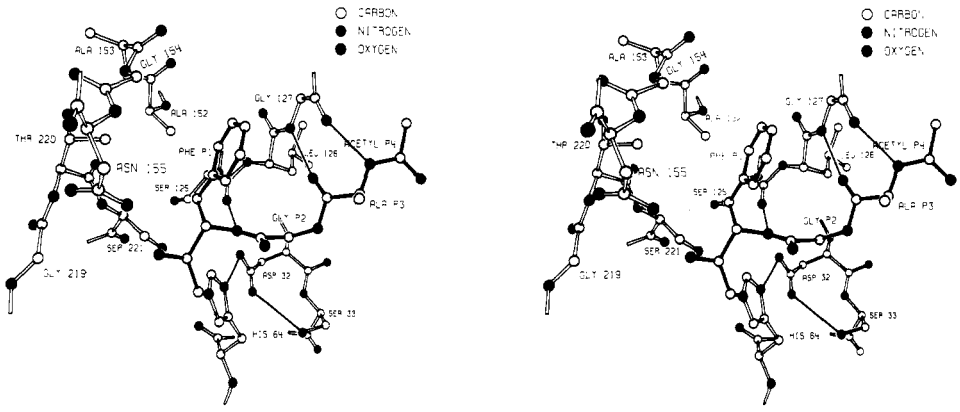

FIGURE 3: The active-site region of subtilisin BPN′ with the inhibitor acetyl-L-alanylglycyl-L-phenylalanylmethylene covalently bound to Nε2 of His-64. The illustration is a stereoscopic pair.

carbonyl group adjoins negative density corresponding to an ejected water molecule, Wat-202 (Alden *et al.*, 1971), which is hydrogen bonded to Ser-221 Oγ in the active enzyme. This negative density may well distort the P_1 Phe carbonyl density enough to cause the appearance of the short contact with Ser-221 Cβ referred to above.

A small positive peak in the difference maps near the Ser-221 hydroxyl groups was interpreted as a rotation of 60° around the α–β bond of that side chain. The corresponding negative peak for this movement is obscured by the added methylene carbonyl density, making precise interpretation of the rotation difficult. However, such a rotation is stereochemically reasonable. Similarly, several other minor peaks and holes of about 3σ amplitude located along the His-64 backbone are consistent with a slight adjustment of the region induced by the alkylation of His-64.

To help confirm that our interpretation was correct the $2F_{ZAGP} - F_{PMS\cdot BPN'}$ and $2F_{ZAP} - F_{PMS\cdot BPN'}$ Fouriers were calculated. In both maps the added methylene carbonyl density appeared to be continuous with density corresponding to His-64. However, one side of this added density was also continuous with density for the Cβ of Ser-221, but the second continuum was not as strong as the linkage to His-64. Furthermore, it was puckered, giving the impression of contact between two separate groups. The packing in this region appears to be very close and the best position for the P_1 Phe carbonyl groups seen in the $2F_{deriv} - F_{PMS\cdot BPN'}$ maps, like that for the conventional difference maps, is also about 0.6 Å inside normal van der Waals contact distance with the Ser-221 side chain.

On balance, the best interpretation is that the inhibitors are covalently attached to His-64 Nε2. This finding is in accord with results of Markland *et al.* (1968) showing that His-64 was specifically alkylated by ZPBK and of Morihara and Oka (1970) that subtilisin BPN′ loses one histidine upon reaction with ZAGPCK. It is interesting to note that in constructing similar models for peptide chloromethyl ketone inhibited γ-chymotrypsin, Segal *et al.* (1971) also found it necessary to place the P_1 carbonyl group atoms closer than a normal van der Waals distance to the Cβ of Ser-195.

The most readily identified feature of P_1 is the Phe side-chain ring, which lies in a surface crevice of the enzyme. One wall of this S_1 crevice is formed by the planar peptide groups of the extended backbone chain segment Ser-125–Leu-126–Gly-127. The Phe side-chain ring lies parallel to this wall and appears to make van der Waals contact with the plane of peptide bonds 125–127. The opposite wall is formed by side chains and backbone of residues Ala-152–Ala-153–Gly-154, while the top of the crevice is formed by backbone atoms of Val-165–Gly-166–Tyr-167–Pro-168. It is noteworthy that the extended backbone chain segment Leu-126–Gly-127 appears to have moved slightly in all the chloromethyl ketone derivatives. A vector moment calculation of the unresolved groups involved (Henderson, 1970) yields an average movement of

0.5 Å away from the bound Phe side-chain ring. The measured distance between the centers of atoms comprising the walls of the S_1 crevice in the active enzyme model is approximately 7 Å (from Cα of Gly-127 to Cα of Ala-154). Assuming a standard van der Waals radius of 2 Å for these atoms, the width of the unoccupied crevice is therefore about 3 Å. Thus the above movement of 0.5 Å is consistent with a slight opening of the crevice to accommodate an aromatic ring of roughly 3.5-Å van der Waals thickness. Water molecule Wat-308, which hydrogen bonds to the backbone NH of Asn-155 and which occupies the entrance to this cleft in the active enzyme, is displaced by the bound P_1 Phe ring.

The last feature of the P_1–S_1 interaction in these derivatives we shall describe is the formation of a hydrogen bond of 3 Å length between the amido NH of the P_1 Phe and the backbone carbonyl oxygen of Ser-125. A water molecule, Wat-201, hydrogen bonded to the carbonyl oxygen of Ser-125 in the active enzyme is displaced by the inhibitor.

The interaction between P_2 and S_2 is much less specific, in all four cases, than that seen at S_1. Both the CO and NH of P_2 point out toward the solvent, away from the enzyme. The α-carbon of P_2 makes van der Waals contact with Cα of Leu-126. In the ZAP derivative P_2 is alanine, and in this case the side-chain Cβ is seen to make contact with Cδ1 of Leu-96 and with the imidazole ring of His-64. In the three tripeptide inhibitors P_2 is Gly, but in each case steric constraints at P_1 and P_3 require a conformation at P_2 such that any L-amino acid there would also have its β carbon placed identically with that seen at P_2 in the ZAP derivative.

Interactions between subsite S_3 and inhibitor segment P_3 are more restrictive than those at S_2. In the tripeptide inhibitors, two hydrogen bonds are formed, one between CO of P_3 and the backbone NH of Gly-127 and another between the NH of P_3 and the backbone carbonyl of Gly-127. Both bonds are about 3 Å in length. These bonds, together with the hydrogen bond at S_1–P_1 already described, create a slightly twisted antiparallel pleated-sheet arrangement between the polypeptide chain of the inhibitor and the backbone segment 125–127 of the enzyme. The twist may be seen clearly in Figures 2 and 3 by noting the relative angle to one another of the hydrogen bonds, shown as thin solid lines. Solvent molecule Wat-88 is displaced by the P_3 alanine backbone. In the active enzyme this solvent molecule is hydrogen bonded to the carbonyl function of Gly-127, to Wat-201, and to Wat-308.

The side chain of P_3 alanine in both AAGP and ZAGP derivatives is seen to point out into the surrounding medium, displacing solvent molecule Wat-82. As expected, negative density corresponding to this displaced solvent molecule and a positive density bulge corresponding to P_3 alanine's Cβ are not present in the ZGGP difference map where P_3 is glycine, further confirming our interpretation.

In the dipeptide derivative ZAP, the S_3 subsite is occupied by the benzyloxycarbonyl group. Although a hydrogen bond appears to exist between the CO function of the Z group and the backbone NH of Gly-127, its length of 3.5 Å is longer than that in the tripeptides. The benzene ring of the benzyloxycarbonyl group lies in the lower part of a large depression in the surface of the enzyme. One face of this benzene ring is in contact with the side chain of Leu-126; the other face of the benzene ring is in contact with the backbone of residues Gly-102–Gln-103. As the model of the ZAP derivative is presently constructed, the plane of the benzene ring lies parallel to the plane of the peptide bond of those residues. However, the electron density in this region of the original subtilisin Fourier map is very weak and may be subject to reinterpretation.

The depression which contains the ZAP benzyloxycarbonyl group is fairly extensive, and in fact continues on up into subsite S_4, now to be described. As just mentioned the lower part of this depression is lined by the side chain of Leu-126 and the backbone of Gly-102–Gln-103. It also contains the side chain of Leu-96. This rather broad lower mouth then funnels up and back into a very deep cavity. One wall is lined by the side chain of Leu-126 and the planar faces of the backbone peptide bonds from Leu-126 to Gly-128. The other side of the first two of these peptide bonds forms one wall of the S_1 pocket already described. Thus the extended backbone chain 125–128 may be thought of as a partition between the S_1 and S_4 cavities. The backbone of Ser-130 and Gly-131 covers the front portion of the ceiling of the S_4 cavity, while the side chain of Leu-135 contributes the innermost part of that ceiling and the back wall. The side chain of Ile-107 contributes to the wall opposite from Leu-126. The side chain of Tyr-104 closes off the deepest part of the cavity.

Subsite S_4 is occupied by the benzyloxycarbonyl group in the ZAGP and ZGGP derivatives, and by an acetyl group in the AAGP derivative. Interpretation of the exact location of the benzyloxycarbonyl group is hindered by the fact that its benzene ring displaces the Tyr-104 ring and the resulting electron density differences tend to cancel one another, leaving only truncated positive density corresponding to the carbonyl end of the Z group. This is seen in Figure 1b where only the bottom of the Z ring model is shown within contoured electron density. Some negative density corresponding to the displaced Tyr-104 ring is seen on its periphery, but no positive density for the new location of the Tyr side chain. The missing positive density may in turn be obscured by some displaced solvent molecules, since the trend of density indicates the side chain is rotated outward around the Cα–Cβ bond. The most likely interpretation is shown in Figure 2. The Z ring has been placed deep in the second hydrophobic crevice previously described. It makes contact with Cδ1 of Ile-107, with the Leu-126–Gly-127–Gly-128 peptide bonds, with Cδ2 of Leu-135, and with the Tyr-104 side chain. The latter has been rotated to lie parallel with the Z ring at a distance of 3.5 Å, that is, in van der Waals contact.

The acetyl group at P_4 in AAGP abuts the Tyr-104 side chain, but no motion of the latter is apparent in this case. The methyl carbon of the acetyl group appears to occupy essentially the same position as the corresponding benzyloxy oxygen of the Z group in the ZAGP derivative. It is interesting to note in Figure 1c the rather large size and density of the acetyl group; its integrated electron density in the difference map is roughly 80% of the integrated density for the P_1 Phe ring. This observation is in agreement with the finding by Segal *et al.* (1971) that their acetyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone reagent, which was also used in this study, was indeed partially contaminated with trifluoroacetyl-L-alanylglycyl-L-phenylalanine chloromethyl ketone.

Several minor but significant peaks and holes are also scattered throughout the four difference maps. Most of these are constant from map to map and are probably due to small adjustments in the enzyme and surrounding solvent structure resulting from inhibitor binding, and to slight alteration in parameters from the native unit cell.

Discussion

A question arises immediately concerning the validity of

TABLE II: Kinetic Constants for Some Subtilisin BPN′ Substrates.[a]

Peptide	K_m (mM)	k_{cat} (sec^{-1})
P_5 P_4 P_3 P_2 P_1 ↓		
Z -Gly-Leu-NH_2	43.4	0.51
Z -Ala-Leu-NH_2	46.0	7.35
Z -Gly-Gly-Leu-NH_2	33.3	14.9
Z -Gly-Ala-Leu-NH_2	11.6	73.1
Z -Ala-Gly-Leu-NH_2	66.7	76.5
Z-D-Ala-Gly-Leu-NH_2	80.0	1.1
Z -Phe-Gly-Leu-NH_2	19.5	57.5
Z -Ala-Gly-Gly-Leu-NH_2	66.7	92.0
Z-D-Ala-Gly-Gly-Leu-NH_2	71.5	6.58

[a] Condensed from Morihara *et al.* (1970). Residues are in the L configuration unless otherwise noted.

TABLE III: Hydrolysis Rates for Some Subtilisin BPN′ Substrates.[a]

Peptide	Rel Rate
P_5 P_4 P_3 P_2 P_1 ↓	
H -Leu-NH_2	0.00
Z -Leu-NH_2	0.23
Z -Gly-Leu-NH_2	1.0
Z -Ala-Leu-NH_2	14.0
Z -Leu-Leu-NH_2	2.4
Z -Pro-Leu-NH_2	4.8
Z -His-Leu-NH_2	0.6
Z -Tyr-Leu-NH_2	0.5
Z-D-Ala-Leu-NH_2	0.056
Z -Gly-Gly-Leu-NH_2	36.0
Z -Ala-Gly-Leu-NH_2	93.0
Z-D-Ala-Gly-Leu-NH_2	2.3
Z -Ala-Gly-Gly-Leu-NH_2	92.0
Z-D-Ala-Gly-Gly-Leu-NH_2	7.0

[a] Condensed from Morihara *et al.* (1970). Residues are in the L configuration unless otherwise noted.

assuming that the mode of binding observed here for polypeptide chloromethyl ketones is a good model for substrate binding. There are several reasons to believe that it is. (1) In a recent review, Shaw (1970) points out that for serine proteases the haloketone analogs of good substrates appear to share the corresponding substrates' affinity for the active-site region. For example, benzylamide retards inactivation of chymotrypsin by TPCK with the same inhibition constant that it shows in retarding hydrolysis of true substrates. Further, TPCK will not alkylate the catalytic site His-57 in chymotrypsin if the active-site region has been previously blocked, for example, by phosphorylation of Ser-195 with DFP. Finally, Morihara and Oka (1970) have shown that the relative rates of subtilisin BPN′ inhibition by three different polypeptide chloromethyl ketones parallel the rates of hydrolysis of the corresponding amide substrates. Therefore, it seems very likely that the chloromethyl ketone substrate analogs used here do indeed undergo substrate-like binding. (2) According to the widely accepted picture of reaction between the serine proteases and a typical substrate, RC(O)X, one expects to observe at least two distinct stages of binding. First the substrate is bound in a noncovalent but highly stereospecific Michaelis–Menten complex. This gives way, possibly *via* a tetrahedral intermediate, to the acyl-enzyme RC(O)Enz in which an ester linkage is formed to the reactive serine. Now it is reasonable to suppose that the difference in geometry between the Michaelis–Menten complex and the acyl-enzyme is small. Our inhibitors, of course, are not bound to the reactive Ser-221 *via* an ester linkage, but to the nearby His-64 side chain *via* a methylene carbonyl linkage. However, in the model it is possible to remove the methylene carbon atom from the methylene carbonyl group, and to construct instead an ester linkage to Ser-221 without appreciably perturbing the structure of the complex. Therefore in this respect as well the enzyme–inhibitor complex seems to provide a reasonable model for a hypothetical enzyme–substrate complex. (3) Very strong support for the model we propose for substrate binding also comes from recent experiments with polypeptide substrates (Morihara *et al.*, 1969, 1970). Kinetic data are given in Tables II and III for hydrolysis by subtilisin BPN′ of a variety of polypeptide substrates with leucinamide at P_1. These data were selected from Morihara *et al.* (1970). Table II shows values obtained by these investigators for kinetic constants k_{cat} and K_m, while Table III shows only relative rates of hydrolysis for several such substrates. In making any correlation between kinetic data and the X-ray results reported here it is important to bear in mind the inherent limitations of these data. For example, the low solubility of many substrates forced Morihara and coworkers to use solutions containing 15% formamide. In the case of Z-Gly-Leu↓-NH_2 (↓ indicates site of hydrolysis), a change from 0–15% formamide increased K_m by a factor of 8 and decreased k_{cat} by one-half. How other substrates are affected is not known, but it seems prudent to consider the kinetic data as significant only to within an order of magnitude. One should also bear in mind the implications of the kinetic equations for a two-step mechanism such as applies to the serine proteases: $K_m = K_s[k_3/(k_2 + k_3)]$ and $k_{cat} = k_2k_3/(k_2 + k_3)$, where K_s is the enzyme–substrate dissociation constant, k_2 is the acylation rate and k_3 the deacylation rate (Gutfreund and Sturtevant, 1956). For amide substrates such as those in Tables II and III, it is generally accepted that $k_2 \ll k_3$ (Markland and Smith, 1971). Under these circumstances $K_m \approx K_s$, and $k_{cat} \approx k_2$ (Zerner and Bender, 1964). Thus the constants in Table II can be interpreted as physically simple kinetic parameters.

It has been known for some time that subtilisin exhibits a distinct preference for L-aromatic amino acids and L-leucine at P_1 (Morihara and Tsuzuki, 1969). This observation is readily explained by the presence of a "hydrophobic" binding cleft at site S_1, as already described. Although our proposed substrate binding model is based exclusively upon experiments with inhibitors incorporating a phenylalanine at P_1, in fact it is not difficult to construct an otherwise identical model with a leucine residue at P_1 instead. The binding cleft is lined principally with hydrophobic groups but in the active enzyme it does contain one or two solvent molecules. As we noted, Wat-308 occupies the entrance to this cleft in the active enzyme crystals, and is displaced by the bound inhibitor. Also, a partially occupied water site exists deep inside the cleft

where Wat-310 can apparently make hydrogen bonds with the carbonyl functions of Ala-152 and Leu-126, and with the backbone NH of Gly-169. Thus, it is evident that this cleft is not lined with hydrophobic groups exclusively. Nevertheless, we refer to it as a "hydrophobic" cleft simply because subtilisin shows a preference for aromatic and apolar residues at P_1.

That subtilisin shows a broader range of specificities at P_1 than chymotrypsin (Morihara and Tsuzuki, 1969) is also consistent with the detailed nature of this cleft. For subtilisin the hydrophobic cleft is planar on only one side, while the other side is a more irregular conglomeration of backbone groups and methyl carbons. In chymotrypsin, on the other hand, both sides of the hydrophobic cleft are planar, presumably accounting for its more specific affinity for aromatic side chains.

The stereospecificity of site S_1 is very high, with Z-Gly-D-Leu$\downarrow$NH$_2$ showing no measurable hydrolysis after 16-hr exposure to subtilisin under optimum catalytic conditions, while Z-Gly-L-Leu$\downarrow$NH$_2$ was more than 95% hydrolyzed (Morihara and Tsuzuki, 1969). Such a degree of stereospecificity is readily explained by our model which clearly shows that binding a D-aromatic amino acid with its aromatic ring and acyl amido group in the same position as for the corresponding L-amino acid would effectively remove the scissle bond from the vicinity of the catalytic site.

The importance of the P_1–S_1 hydrogen bond between the P_1 amido group and the carbonyl oxygen of Ser-125 is also evident from the observation (Table III) that leucinamide is not measurably hydrolyzed by subtilisin BPN′, although the rate for Z-Leu$\downarrow$NH$_2$ is appreciable. This is, of course, consistent with the well-known fact that subtilisin is an endopeptidase. It seems likely that a free amino group at P_1, in contrast to an amido group at that position, would be too highly solvated and have the wrong geometry to form an acceptable P_1–S_1 hydrogen bond.

The kinetic parameters of Table II and the hydrolysis rate data of Table III indicate that an alanine at P_2 is optimum in the series Z-X-Leu$\downarrow$NH$_2$ and in the series Z-Gly-X-Leu$\downarrow$NH$_2$. Although not conclusive, the kinetic parameters do suggest that orientation of the substrate, as reflected by k_{cat}, is affected to a greater degree than is substrate binding affinity, as measured by K_m. Further, the data in Table III for the series Z-X-Leu$\downarrow$NH$_2$ show that a bulkier side chain than alanine at P_2 also interferes with the efficiency of the hydrolytic mechanism. This observation is consistent with our model. As described earlier the Cβ of alanine at P_2 makes van der Waals contact with the Cδ1 of Leu-96, and with the side chain of His-64. Presumably these contacts allow optimum orientation of the substrate in the active site region. Accommodation of a larger side chain, such as a Leu or Tyr at P_2 would disrupt the alignment, while glycine at this position, on the other hand, would allow greater freedom of motion and hence retard proper alignment. The geometry of our binding model also predicts that the Cβ of a D-Ala at P_2 would make unfavorable contact with the Cα of Leu-126, necessitating a substantial realignment of the substrate. This is borne out by the data of Table III which shows a 250-fold increase in hydrolysis rate upon comparing Z-L-Ala-L-Leu$\downarrow$NH$_2$ to Z-D-Ala-L-Leu$\downarrow$NH$_2$ as substrate.

In concluding our discussion of subsite S_2 we should comment upon the observation that subtilisin is not inhibited by TPCK (a specific inhibitor for chymotrypsin) while ZPBK is a good subtilisin inhibitor (Markland *et al.*, 1968). It is also of interest that, on the other hand, subtilisin does hydrolyze tosyl-L-phenylalanine ethyl ester, albeit an order of magnitude more slowly than benzyloxycarbonyl-L-phenylalanine ethyl ester (Morihara *et al.*, 1970). Model building experiments were undertaken to investigate the underlying structural basis for this apparent anomaly. It was found that if the methylene carbonyl linkage to His-64 was maintained, a stereochemically permissible tosyl group at P_2 could not be constructed. More specifically, any rotation about the N–S bond which left the sulfonyl oxygens in an acceptable position with respect to the methylene carbon forced the tosyl benzene ring into excessively close contact with the enzyme in the vicinity of Leu-96. This was not the case for the more flexible benzyloxycarbonyl blocking group. When, however, the methylene linkage is not present, as for hydrolysis of tosyl-L-phenylalanine ethyl ester, a sterically acceptable acyl intermediate can be constructed. Again, our model satisfactorily explains the observed properties of the enzyme.

The P_3–S_3 interaction involves two additional hydrogen bonds and as expected the kinetic data in Tables II and III indicate that L-tripeptides are better substrates than L-dipeptides. For example, extention of the best dipeptide substrate Z-Ala-Leu$\downarrow$NH$_2$ to Z-Gly-Ala-Leu$\downarrow$NH$_2$ increases k_{cat} by a factor of 10 and decreases K_m by a factor of 4 (Table II).

As already described, the side chain of P_3 points outward into the solvent, and in agreement with this picture, replacement of Gly at X in the series Z-X-Gly-Leu$\downarrow$NH$_2$ even by a residue with a side chain as bulky as that of Phe has a relatively minor effect on binding or catalysis (Table II). However, the side chain of a D-amino acid at P_3 would make unacceptably close contact with Cδ2 of Leu-96, necessitating some rearrangement of the bound polypeptide. Again, support for the present substrate binding model is provided by the kinetic data, where the K_m is little affected, but k_{cat} is drastically reduced in changing to D-Ala in Z-Ala-Gly-Leu$\downarrow$NH$_2$ (Table II).

It is clear from the kinetic data that a subsite S_4, exerting some stereospecific influence in subtilisin, must exist, as may be seen by comparing the last two entries in Tables II and III. Very little can be said about S_4 from the results reported here, however, except that it may be occupied by the benzyloxycarbonyl blocking group in the ZAGP and ZGGP derivatives. It is noteworthy that the Z group would bear some stereochemical resemblance to a phenylalanine residue at P_4, suggesting that indeed S_4 may be specifically designed to accommodate such a residue.

Subsites S_1' and S_2' on the leaving-group side of the scissile bond are also known from kinetic experiments (Morihara *et al.*, 1970) to affect the rate of hydrolysis, and model building shows that there are a very limited number of stereochemically convincing possibilities for their location on the surface of the subtilisin molecule. However, X-ray crystallographic evidence on this point is lacking and the subject will not be discussed in this paper.

(4) Finally, a very important line of evidence supporting the model for substrate binding reported here is the remarkable similarity at the binding site between subtilisin and chymotrypsin. Both enzymes are well studied endopeptidases characterized by a reactive serine in the active site. Subtilisin is representative of a family of proteases, thus far found only in

TABLE IV: Comparison of Active-Site Atoms from Subtilisin and Chymotrypsin.[a]

Subtilisin		Chymotrypsin		Deviation
Atom	Residue	Atom	Residue	(Å)
Cβ	Asp-32	Cβ	Asp-102	0.9
Cγ	Asp-32	Cγ	Asp-102	1.1
Oδ2	Asp-32	Oδ2	Asp-102	2.1
Oδ1	Asp-32	Oδ1	Asp-102	0.6
Cβ	His-64	Cβ	His-57	0.8
Cγ	His-64	Cγ	His-57	0.5
Nδ1	His-64	Nδ1	His-57	0.5
Cε1	His-64	Cε1	His-57	0.5
Nε2	His-64	Nε2	His-57	0.4
Cδ2	His-64	Cδ2	His-57	0.8
Cα	Ser-125	Cα	Ser-214	1.7
C	Ser-125	C	Ser-124	1.0
O	Ser-125	O	Ser-214	0.9
N	Leu-126	N	Trp-215	0.8
Cβ	Leu-126	Cβ	Trp-215	1.2
Cα	Leu-126	Cα	Trp-215	0.5
C	Leu-126	C	Trp-215	0.3
O	Leu-126	O	Trp-215	1.2
N	Gly-127	N	Gly-216	0.6
Cα	Gly-127	Cα	Gly-216	0.3
C	Gly-127	C	Gly-216	0.6
O	Gly-127	O	Gly-216	0.7
N	Gly-128	N	Ser-217	0.8
Cα	Gly-128	Cα	Ser-217	1.2
Cβ	Ser-221	Cβ	Ser-195	0.4
Oγ	Ser-221	Oγ	Ser-195	1.1
Cα	Ser-221	Cα	Ser-195	0.3
		Mean deviation = 0.81 Å		

[a] Coordinates for subtilisin are from Alden *et al.* (1971). Chymotrypsin coordinates are from Birktoft and Blow (1972).

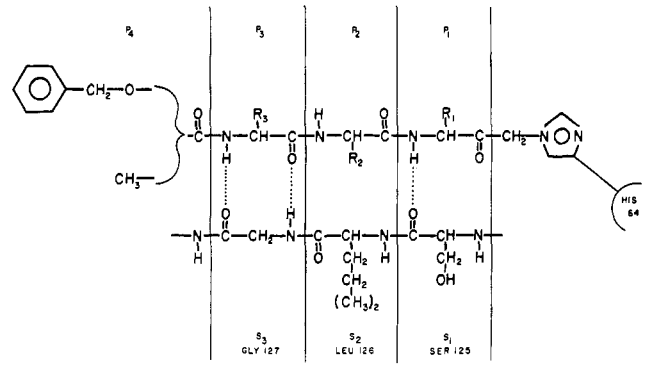

FIGURE 4: A schematic representation of the observed binding of tripeptide chloromethyl ketone inhibitors to active subtilisin. The blocking group on the inhibitor peptide is either an acetyl or a benzyloxycarbonyl group.

bacteria and molds (Mikeš *et al.*, 1969) characterized by the amino acid sequence Gly-Thr-Ser*-Met-Ala around the reactive serine, while chymotrypsin is representative of a family found in both bacteria and higher organisms possessing Gly-Asp-Ser*-Gly within the corresponding region of the active-site sequence (Dayhoff, 1969). Both enzymes readily hydrolyze substrates in which the carbonyl group of the susceptible bond is donated by an aromatic amino acid, and both enzymes possess a histidine residue essential for catalysis. Finally, both enzymes have been the subject of rather extensive studies by X-ray crystallography which have yielded considerable insight into the stereochemistry of catalysis by serine proteases in general.

The structure of α-chymotrypsin has been studied extensively by Blow and his coworkers (Blow, 1971). Two main features of the active-site area are first, the "tosyl hole" binding pocket which accommodates the aromatic side chain of specific substrates, and second, the "charge-relay" hydrogen-bond system linking the side chains of Asp-102, His-57, and Ser-195. This latter feature may account for the extraordinary reactivity of the Ser-195 hydroxyl group.

The X-ray structure of subtilisin BPN′ (Wright *et al.*, 1969) revealed that, despite the general similarity of enzymic properties between subtilisin and chymotrypsin, the two molecules bore no overall structural resemblance to one another. Subtilisin, for example, contains 31% α helix and an extensive region of parallel β-pleated sheet, while chymotrypsin contains only 8% helix but many segments of antiparallel β sheet. However, it was discovered (Alden *et al.*, 1970) that a charge-relay system of identical geometry involving Asp-32, His-64, and Ser-221 existed in the subtilisin active site. The finding suggested that the two enzymes had arrived at the same catalytic mechanism despite their overall dissimilarities—a case, perhaps, of convergent evolution at the molecular level. The work reported in this paper extends and strengthens this notion.

The model for polypeptide substrate binding to subtilisin proposed here is strikingly similar to that arrived at by Segal *et al.* (1971) in their parallel study of the binding of polypeptide chloromethyl ketones to γ-chymotrypsin. Both enzymes bind the substrate polypeptide in exactly the same way, as can be seen by comparing Figure 3 of this paper to Figure 4 in Segal *et al.* (1971). An extended segment of backbone chain in the enzyme, Ser-214–Trp-215–Gly-216 in chymotrypsin, and Ser-125–Leu-126–Gly-127 in subtilisin, forms an antiparallel β-pleated sheet with the substrate. The arrangement is illustrated schematically in Figure 4, and is displayed in detail in stereo views of the subtilisin model in Figures 2 and 3. Further, the plane of this pleated sheet comprises one wall of the crevice that binds the aromatic side chain at P_1 in both enzymes. It is also remarkable that the same residue types are present in both enzymes at S_1 and S_3 where hydrogen bonds between substrate and enzyme backbone chains occur, namely, serine at S_1 and glycine at S_3. Why this should be so is not completely obvious, but it is probably relevant that if S_3 were other than glycine, its side chain would block the S_1 crevice in both enzymes, as is actually the case for elastase (Shotton and Watson, 1970).

In order to compare the geometries of the catalytic and binding sites of the two enzymes in a more quantitative manner, the coordinates of 27 atoms comprising this region in chymotrypsin were rotated into a least-squares fit to the corresponding atoms in subtilisin. As seen in Table IV the resulting root-mean-square difference in relative position is about 0.8 Å with a maximum deviation of 2.1 Å between subtilisin Asp-32 Oδ2 and chymotrypsin Asp-102 Oδ2. To place this result in perspective, it should be noted that a recent comparison between the crystallographically determined structures of subtilisin BPN′ and Novo (Hol, 1971), believed to be identical molecules (Robertus *et al.*, 1971) resulted in a

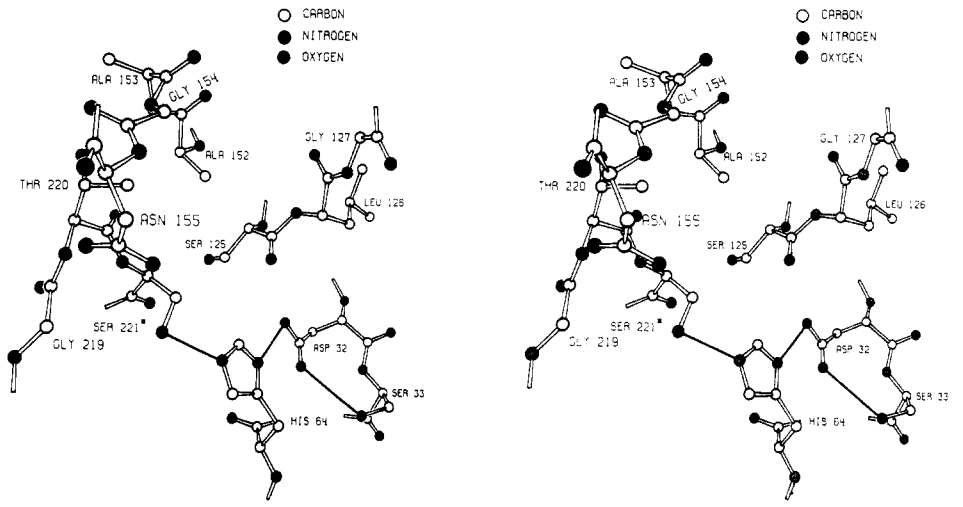

FIGURE 5: The active site region of subtilisin BPN'. Hydrogen bonds are depicted as thin solid lines. The orientation is identical to that used for Figures 2 and 3.

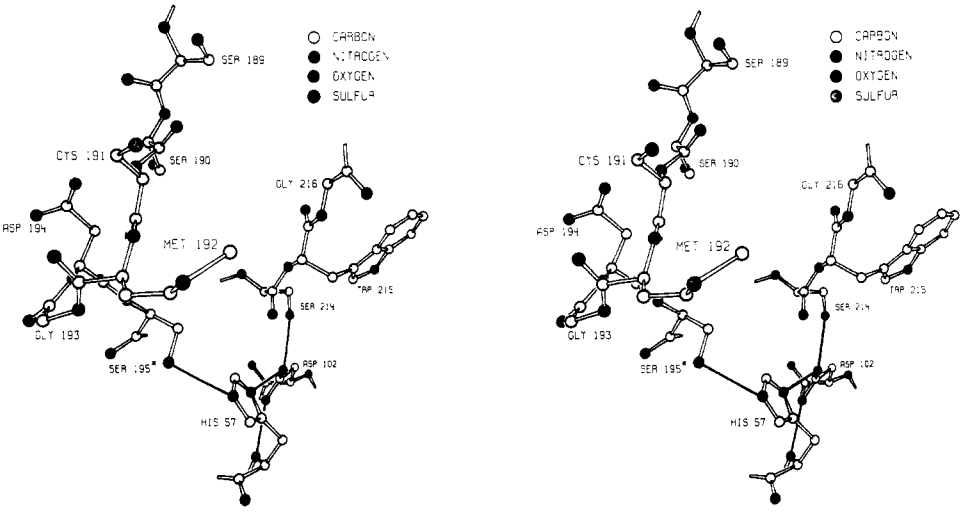

FIGURE 6: The active-site region of α-chymotrypsin. The orientation was arrived at by rotating the atoms listed in Table IV into a least-squares fit to the equivalent atoms of subtilisin. Hydrogen bonds are shown as thin solid lines.

root-mean-square difference of 1.6 Å for all atoms and 1.2 Å for backbone atoms. Thus our confidence is reinforced that the active-site geometries seen here for subtilisin and chymotrypsin are not just qualitatively similar, but are indistinguishable at the present resolution limits. The active site of subtilisin is shown in Figure 5 and the active site of chymotrypsin, rotated into the subtilisin orientation is shown in Figure 6 for comparison.

In Figures 5 and 6 the charge relay system in subtilisin is shown as two hydrogen bonds, one between the Oγ of Ser-221 and the Nϵ2 of His-64, and a second between the Nδ1 of His-64 and the Oδ1 of Asp-32. In chymotrypsin the corresponding hydrogen bonds extend between Oγ of Ser-195 and Nϵ2 of His-57 and between Nδ1 of His-57 and Oδ1 of Asp-102. However, it should be mentioned that the electron density maps of both proteins indicate that the second hydrogen bond, from the Nδ1 of the active-site His, may in fact be directed between the carboxyl oxygens of the aspartate instead of toward one or the other of them.

In Figure 5 we have included, in addition to the charge relay system, a hydrogen bond between the Oδ2 of Asp-32 and Oγ of Ser-33 and similarly in Figure 6 a hydrogen bond between Oδ1 of Asp-102 and Oγ of Ser-214. A hydrogen bond between the backbone NH of His-57 and the Oδ2 of Asp-102 has also been included. These bonds are assumed to help maintain rigidity in the charge-relay system. It should also be pointed out that the apparent minor differences between the two active sites as depicted in Figures 5 and 6 are within the present probable error in the locations of the various groups.

In considering the case for convergent evolution of the catalytic machinery of these two enzymes it is important to note not only that the overall tertiary structures of the molecules are very different, but that the sequential order of the residues involved is also quite different in the two, as shown schematically in Figure 7. This fact makes it highly unlikely that subtilisin and chymotrypsin descended from a common ancestral serine protease with all but the original key residues completely altered in such a way as to destroy any vestige of structural similarity.

That the two enzymes have independently evolved both the same catalytic machinery (the charge-relay system), and

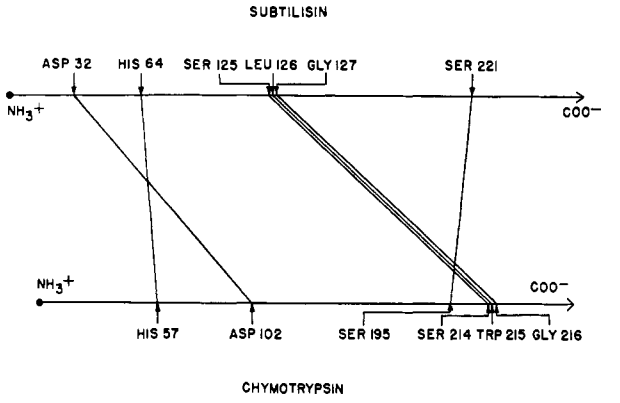

FIGURE 7: Positions of the residues involved in the charge-relay system and in substrate binding as they occur along the primary sequence of subtilisin (upper) and chymotrypsin (lower).

essentially the same substrate binding apparatus invites speculation that these two entities must somehow function together for optimum enzyme efficiency. It is conceivable that the charge-relay system may act most effectively when the natural peptide substrate is oriented in precisely the geometry imposed by the binding site, and that the required stereochemical situation can be achieved only when the binding site has the specific structure we find in the two enzymes. Wang and Parker (1967) proposed a mechanism of action for chymotrypsin in which a system of preformed hydrogen bonds in the catalytic site facilitated proton transfer to the leaving group of the substrate during the acylation step. For this mechanism to function optimally, it is not unreasonable to suppose that the substrate must be precisely oriented with respect to the catalytic groups. Further, Koshland and his coworkers (Storm and Koshland, 1970; Dafforn and Koshland, 1971) have recently proposed an "orbital steering" theory, in which very accurate orientation of substrates contributes significantly to the catalytic power of enzymes. The specifics of this particular theory have raised some controversy (Bruice *et al.*, 1971), but nevertheless the concept that accurate substrate alignment is important has received considerable attention lately (Page and Jencks, 1971).

References

Alden, R. A., Birktoft, J. J., Kraut, J., Robertus, J. D., and Wright, C. S. (1971), *Biochem. Biophys. Res. Commun. 45*, 337.

Alden, R. A., Wright, C. S., and Kraut, J. (1970), *Phil. Trans. Roy. Soc. London, Ser. B 257*, 119.

Birktoft, J. J., and Blow, D. M. (1972), *J. Mol. Biol.* (in press).

Blow, D. M. (1971), *Enzymes 3*, 185.

Blow, D. M., Birktoft, J. J., and Hartley, B. S. (1969), *Nature (London) 221*, 337.

Bruice, T. C., Brown, A., and Harris, D. O. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 658.

Dafforn, A., and Koshland, D. E., Jr. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 2463.

Dayhoff, M. O. (1969), Atlas of Protein Sequence and Structure, Vol. 4, Silver Spring, Md., National Biomedical Research Foundation, p 49.

Gutfreund, H., and Sturtevant, J. M. (1956), *Biochem. J. 63*, 656.

Henderson, R. (1970), *J. Mol. Biol. 54*, 341.

Hol, W. G. J. (1971), Ph.D. Thesis, University of Groningen.

Markland, F. S., Shaw, E., and Smith, E. L. (1968), *Proc. Nat. Acad. Sci. U. S. 61*, 1440.

Markland, F. S., Jr., and Smith, E. L. (1971), *Enzymes 3*, 561.

Mikeš, O., Turková, J., Toan, Ng. B., and Šorm, F. (1969), *Biochim. Biophys. Acta 178*, 112.

Morihara, K., Oka, T., and Tsuzuki, H. (1969), *Biochem. Biophys. Res. Commun. 35*, 210.

Morihara, K., and Oka, T. (1970), *Arch. Biochem. Biophys. 138*, 526.

Morihara, K., Oka, T., and Tsuzuki, H. (1970), *Arch. Biochem. Biophys. 138*, 515.

Morihara, K., and Tsuzuki, H. (1969), *Arch. Biochem. Biophys. 129*, 620.

Morihara, K., Tsuzuki, H., and Oka, T. (1971), *Biochem. Biophys. Res. Commun. 42*, 1000.

Neet, K. E., Nanci, A., and Koshland, D. E., Jr. (1968), *J. Biol. Chem. 243*, 6392.

Page, M. I., and Jencks, W. P. (1971), *Proc. Nat. Acad. Sci. U. S. 68*, 1678.

Polgar, L., and Bender, M. L. (1969), *Biochemistry 8*, 136.

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

Robertus, J. D., Alden, R. A., and Kraut, J. (1971), *Biochem. Biophys. Res. Commun. 42*, 334.

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

Schwert, G. W., and Takenaka, Y. (1955), *Biochim. Biophys. Acta 16*, 570.

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

Shaw, E. (1970), *Enzymes 1*, 91.

Shaw, E., and Ruscica, J. (1968), *J. Biol. Chem. 243*, 6312.

Shotton, D. M., and Watson, H. C. (1970), *Nature (London) 225*, 811.

Steitz, T. A., Henderson, R., and Blow, D. M. (1969), *J. Mol. Biol. 46*, 337.

Storm, D. R., and Koshland, D. E., Jr. (1970), *Proc. Nat. Acad. Sci. U. S. 66*, 445.

Wang, J. H., and Parker, L. (1967), *Proc. Nat. Acad. Sci. U. S. 58*, 2451.

Wright, C. S., Alden, R. A., and Kraut, J. (1969), *Nature (London) 221*, 235.

Zerner, B., and Bender, M. L. (1964), *J. Amer. Chem. Soc. 86*, 3669.