SERINE PROTEASES: STRUCTURE AND MECHANISM OF CATALYSIS

❖949

Joseph Kraut

Department of Chemistry, University of California, San Diego,
La Jolla, California 92093

CONTENTS

PERSPECTIVES AND SUMMARY

The serine proteases are a class of proteolytic enzymes characterized by the presence of a uniquely reactive serine side chain. They are of extremely widespread occurrence and diverse function. Thus far only two families have been especially well studied, the trypsin family and the subtilisin family, although it is likely that other distinct families of serine proteases also exist. According to current usage, the term "family" connotes a group of these enzymes that are sequentially homologous and closely similar in overall three-dimensional geometry. Members of the trypsin family occur not only in the vertebrates and other phyla of higher organisms, but also among the prokaryotes. On the other hand, subtilisins have thus far been definitely identified only among the bacilli.

As a result of about thirty years of intensive study, most especially concentrated on bovine chymotrypsin because of its ready availability, the mode of operation of these enzymes is now fairly well understood in terms of their three-dimensional geometries, down to the level of individual atomic positions. This has been possible because, to a degree not yet achieved for any other group of enzymes, protein crystallography has yielded extensive and highly detailed structural information on an array of more or less related serine proteases. Consequently, investigators are now able to interpret in stereochemical terms the vast store of chemical and kinetic information so painstakingly and ingeniously accumulated over the years.

An important factor contributing to the assurance with which this interpretation can be made is that the trypsin family and the subtilisin family have their catalytically functional groups arranged in the same geometrical relationship. Nevertheless, the two enzyme families have entirely different overall three-dimensional structures and are therefore very probably descended from unrelated ancestral enzymes. Thus nature appears to have invented the same biochemical mechanism on at least two separate occasions.

Following is a list of the serine protease crystal structures that have been published thus far: bovine α-chymotrypsin (1–4); bovine γ-chymotrypsin (5); bovine chymotrypsinogen (6); subtilisin BPN' (7, 8); subtilisin Novo (9); porcine elastase (10, 11); bovine trypsin (12–15); the complex between bovine trypsin and bovine pancreatic trypsin inhibitor (16–19); the complex between porcine trypsin and soybean trypsin inhibitor (20, 21); *Streptomyces griseus* protease B (22, 23); and bovine trypsinogen (24, 24a). The list includes identical crystal structures that have been independently determined twice (α-chymotrypsin, trypsin, and trypsinogen) and the structures of what subsequently turned out to be two different crystal forms of the same protein, determined independently in different laboratories (subtilisins BPN' and Novo), thus providing a reassuring degree of redundancy. In addition, a number of important binding studies by the difference-Fourier method have been reported.

In this article, I emphasize the view that transition-state theory provides a simple and unifying conceptual framework within which to discuss enzymic structure-function relationships. According to this view an enzyme catalyzes a particular reaction because it is a template for binding the transition-state complex of that reaction. The serine proteases specifically bind the tetrahedral transition-state complex characteristic of acyl transfer reactions, and it is observed that all of the known covalent inhibitors of the serine proteases do indeed form tetrahedral adducts with similar geometry.

This binding template is made up of a number of elements acting together: an antiparallel β-binding site for the acylating polypeptide chain of

the substrate; specific side-chain binding sites that vary with the particular enzyme; a less specific leaving-group site; a site for hydrogen bonding to a tetrahedral oxyanion; and of course the reactive serine side chain for covalent bonding to the substrate's carbonyl carbon atom.

The charge relay system, consisting of a histidine side chain imidazole hydrogen bonded to a buried carboxylate of an aspartic acid, can also be regarded as furnishing the binding site for a proton which, in the transition state, is in the process of being transferred between nucleophiles, e.g. between the reactive serine of the enzyme and the leaving group of the substrate. The charge relay system probably does not directly induce an unusually high degree of intrinsic nucleophilicity at the reactive serine.

Because of limitations of space I must forego the customary background survey of the field. The interested reader is referred to recent reviews by Blow (25), Stroud et al (26), and Bender & Killheffer (27). A good elementary overview is given by Stroud in *Scientific American* (28). Also conspicuous for its omission here is any discussion of zymogen activation (6, 29, 30) or evolutionary relationships (31) as these factors apply to the question of mechanism vs structure.

THE REACTION SEQUENCE AND INTERMEDIATES

A large variety of ingenious kinetic studies by many investigators has firmly established the sequence of reactions that occurs when serine proteases catalyze the hydrolysis of peptide and ester bonds. Most of this work has been done with bovine α-chymotrypsin, and has recently been reviewed in detail by Bender & Killheffer (27). All of the evidence points to the existence of identical reaction sequences not only for the entire family of serine proteases homologous with chymotrypsin (32–34) but also for the subtilisin family (35). In retrospect this is not surprising in view of the close similarity between the three-dimensional structures of the catalytic regions in the two families of enzymes. The general serine protease reaction sequence is shown schematically in Figure 1. This figure is arranged to emphasize the fact that the acylation step (top row) is mechanistically just the reverse of the deacylation step (bottom row), and indeed the symmetry would be exact if X and Y were identical, as they are when X = Y = OH and the net reaction is just the exchange with water of this group on a virtual substrate (36). The group RCO– represents a polypeptide, an amino acid or other acyl group, and HX is normally an amine, alcohol or other conventional leaving group, whereas HY is water. In general terms, however, as this representation of the reaction sequence emphasizes, the reaction catalyzed by serine proteases is fundamentally an acyl transfer or interchange. The enzyme in Figure 1 is symbolized by E–OH, with –OH representing the reactive serine side chain. The first intermediate is the Michaelis complex. This is followed by

$$R-\overset{O}{\overset{\|}{C}}-X + E-OH \xrightarrow{K_s} R-\overset{O}{\overset{\|}{C}}-X : E-OH \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} R-\overset{O^-}{\overset{\|}{\underset{X\cdots H^+}{C}}}-O-E$$

$$\xrightarrow[k_{-2}]{k_2} \quad HX$$

$$R-\overset{O}{\overset{\|}{C}}-O-E$$

$$\xrightarrow[k_3]{k_3} \quad HY$$

$$R-\overset{O}{\overset{\|}{C}}-Y + E-OH \overset{K_p}{\rightleftharpoons} R-\overset{O}{\overset{\|}{C}}-Y : E-OH \underset{k_4}{\overset{k_{-4}}{\rightleftharpoons}} R-\overset{O^-}{\overset{\|}{\underset{Y\cdots H^+}{C}}}-O-E$$

Figure 1   The reaction sequence for serine proteases.

an intermediate tetrahedral adduct in which a covalent bond is formed between the substrate carbonyl carbon atom and the reactive serine $O\gamma$ atom. A hydrogen ion is shown as partially transferred from $O\gamma$ to the leaving group of the substrate in the tetrahedral intermediate. When this transfer is completed, the tetrahedral intermediate breaks down to the acyl-enzyme and liberates the free leaving group. Subsequent deacylation of the acyl-enzyme is, as noted above, simply the reverse of the acylation process with the leaving group replaced by water or some other acyl-accepting nucleophile.

What is the evidence for such a sequence of intermediates in the catalytic pathway of the serine proteases? One line of evidence is that several reasonably stable models of all three intermediates, the Michaelis complex, the tetrahedral adduct, and the acyl-enzyme, have been prepared and isolated, and a few have been characterized structurally by the difference-Fourier method of X-ray crystallography. The term "models" is used here because obviously it is not possible at ordinary temperatures to prepare true intermediates of actual enzyme-catalyzed reactions. However it is possible to find special conditions under which some step of the enzymic process is blocked, either because the substrate-analog is defective or distorted (i.e. inhibitors), or because the enzyme is altered in some way, or both. One early example is the crystallographic study by Steitz, Henderson & Blow (37) of the virtual substrate N-formyl-L-tryptophan bound to α-chymotrypsin, which is presumably a model for the Michaelis complex. Similar models of polypeptide virtual substrates bound to subtilisin were examined crystallographically by Robertus et al (38). A model for the acyl-enzyme was structurally characterized by Henderson (39), who studied indoleacryloyl-α-chymotrypsin. This acyl-enzyme is hydrolyzed very slowly, and the study was performed at pH 4 where the enzyme is relatively inactive. Models of the tetrahedral adduct for subtilisin were described by Matthews et al (40)

and by Poulos et al (41). The former were concerned with subtilisin-boronic acid complexes in which the boronic acid group assumes a tetrahedral geometry. The latter reported on the structure of polypeptide chloromethyl ketone derivatives of subtilisin in which the ketone group forms a stabilized tetrahedral hemiketal with the catalytic serine side chain while the chloromethyl group has simultaneously reacted to alkylate the catalytic histidine side chain.

Despite the existence of these models, however, it may be asked whether the intermediates shown in Figure 1 actually occur in the enzyme-catalyzed reaction sequence. The kinetic evidence is extensive and can only be outlined briefly here.

The Michaelis complex was originally proposed as a ubiquitous feature of all enzymic reactions, including the serine protease–catalyzed reactions, to explain the phenomenon of saturation kinetics. More direct evidence for this intermediate is that binding of substrate, for example N-acetyl-L-tryptophanamide, to α-chymotrypsin in solution has been demonstrated by equilibrium dialysis, and it has been shown that such binding competitively inhibits hydrolysis of other substrates (42). Thus, the reality of the Michaelis complex seems well established. In any event it is not easy to imagine a catalytic mechanism that would not involve prior binding of the substrate to the enzyme.

Early kinetic evidence for the acyl-enzyme was the demonstration by Hartley & Kilby (43) that chymotrypsin-catalyzed hydrolysis of p-nitrophenyl esters resulted in the initial release of a burst of p-nitrophenol stoichiometrically equivalent to the number of moles of enzyme present. Zerner, Bond & Bender (44) later showed that the acyl-enzyme hypothesis was very probably valid for less highly reactive ester substrates, as distinct from the highly labile p-nitrophenyl esters, by demonstrating that the same V_{max} is observed for all esters of the same acyl group. This is consistent with a process in which the same acyl-enzyme is formed rapidly and breaks down more slowly. For amides the relative rates of the two steps are reversed; the rate-determining step is acylation, so intermediacy of the acyl-enzyme is harder to establish (45). There is, however, no compelling reason to suppose that the reaction sequence should be different for the two kinds of substrate, and indeed Fastrez & Fersht (46) have shown that a common intermediate, presumably the acyl-enzyme, occurs in δ-chymotrypsin–catalyzed hydrolysis of ester, amide, and peptide derivatives of N-acetyl-L-phenylalanine.

Tetrahedral intermediates, unstable species of high energy, have long been known or inferred in many nonenzymic acyl-transfer reactions, and it is a reasonable expectation, although not absolutely required, that they should also play a role in serine protease catalysis of the same kinds of

reactions (47). In fact, a central thesis in current views of the mechanism of serine proteases is that the activated transition-state complex for the enzyme-catalyzed acyl-transfer reaction must closely resemble the tetrahedral intermediate in structure. The precise meaning of this statement may be clarified by focusing momentarily on the location of the "extra" proton in the tetrahedral intermediate and in the transition state. In the tetrahedral species depicted in Figure 1, the proton in question is shown as partially bonded both to the reactive serine $O\gamma$ atom and to the leaving group X or the entering group Y. However, to qualify as an intermediate, a species must have sufficient stability to exist long enough to be identified experimentally. Depending upon the nature of the groups X and Y, it is conceivable that the most stable species may have the proton on either X or Y, or on the serine $O\gamma$ atom. In any case it is unlikely to be halfway between. On the other hand, the transition state is by definition the most unstable high-energy species in the reaction pathway. Thus it will be a very transient species with bonds in the process of forming and breaking. But regardless of precisely where it occurs in the pathway it is evident that on going from Michaelis complex to acyl-enzyme or from acyl-enzyme to product complex, the transition state will to some degree resemble the tetrahedral intermediate.

Several approaches have yielded kinetic evidence for an additional intermediate, which is taken to be the tetrahedral adduct, lying on the reaction sequence between the Michaelis complex and the acyl-enzyme. However, because the tetrahedral adduct is relatively unstable and must therefore exist in a low steady-state concentration, demonstration of its presence must rely upon rather more esoteric experiments and analyses of kinetic data than are required for the other intermediates. Caplow (48) and Lucas & Caplow (49) have interpreted substituent effects on rates of acylation of chymotrypsin by the p-methoxy- and p-chloroanilides of acetyl-L-tryptophan as evidence for existence of the tetrahedral intermediate. Similar conclusions for subtilisin were drawn by Bender & Philipp (50) from results obtained in studies on hydrolysis of two substituted acetanilides. Frankfater & Kézdy (51) showed that chymotrypsin-catalyzed hydrolysis of p-nitrophenyl acetate and of the corresponding thiol ester proceeds by the same mechanism and at the same rate. Because the –SR group should be displaced at least 250 times faster than the –OR group, they concluded that a rate-determining, bond-forming step precedes the bond-breaking step in acylation of the enzyme and identified the rate-determining step as formation of a tetrahedral adduct. Fersht & Requena (52) observed a change with pH of the rate-determining step in the acylation of some unusually soluble specific hydrazide substrates of chymotrypsin. They interpreted this as implying the existence of a tetrahedral intermediate between the Michaelis complex and the acyl-enzyme, breakdown of which is rate-determining at

high pH and formation of which is rate-determining at low pH. A similar interpretation has been made by O'Leary & Kluetz (53) of the variation with pH in a ^{15}N kinetic isotope effect on the chymotrypsin-catalyzed hydrolysis of N-acetyl-L-tryptophanamide.

More direct spectrophotometric evidence for an additional intermediate between the Michaelis complex and the acyl-enzyme has been obtained with chromophoric substrates of chymotrypsin, i.e. N-furylacryloyl-L-tyrosine methyl ester (54) and N-furylacryloyl-L-tryptophanamide (55) and the corresponding methyl ester (56).

It now seems probable that there are other intermediates in the serine protease reaction sequence as well. Recently Fink (57) was able to detect four reactions prior to acylation in low-temperature (to $-90°C$) spectrophotometric studies on chymotrypsin hydrolysis of N-acetyl-L-phenylalanine p-nitroanilide in aqueous dimethyl sulfoxide. Reaction 1 was ascribed to formation of the Michaelis complex, reactions 2 and 3 to substrate-induced changes in the positions of active-site groups, and reaction 4 to some unidentified process involving His-57 but which is not formation of a tetrahedral intermediate, an oxazolinone [implicated by Coletti-Previero et al (56)] or an acyl-enzyme.

In a related recent development, low-temperature protein crystallography has been introduced by Douzou, Hoa & Petsko (58) and Petsko (59). Using these methods, Alber, Petsko & Tsernoglou (60) in collaboration with Fink & Ahmed (61) have trapped a normally unstable acyl-elastase in X-ray studies at $-55°C$. Clearly the future holds much promise for low-temperature enzymology, both in solution and in the crystal.

The sequence of chemical events depicted in Figure 1 and outlined briefly in the foregoing paragraphs is often referred to as the catalytic mechanism for the serine proteases, and may be regarded as defining that class of enzymes. However, it would seem to be preferable to reserve the term "mechanism" for the molecular machinery responsible for enzymic rate enhancement, i.e. for actual catalysis, and to refer to Figure 1 and similar schemes as the reaction "sequence" or "pathway." Clearly Figure 1 tells little about why amides, esters, and peptides hydrolyze perhaps 10^{10} times faster in the presence of a serine protease than in its absence; there is nothing in the scheme that would distinguish the species E–OH from N-acetyl-L-serineamide, which does not have any particularly notable enzymic properties even though it does have a free serine side chain that could participate in "covalent catalysis."

This review takes the position that the actual "mechanism" of serine protease catalysis in the sense implied by the above, i.e. in the sense of explaining catalysis in structural terms, is based on stabilization of the substrate in its activated transition-state complex by specific binding to the enzyme.

TRANSITION-STATE STABILIZATION

The assertion that the catalytic activity of an enzyme is due to preferential binding of a substrate molecule in a configuration characteristic of its activated transition-state complex might almost seem to be a tautology. This idea was first explicitly enunciated by Pauling thirty years ago (62, 63). It has since then been touched on briefly by Bender, Kézdy & Gunter (64), by Jencks (65) in an article on the possible role of strain and distortion in enzyme catalysis, and by several other authors in a variety of contexts. Wolfenden (66–68) and Lienhard (69, 70) in particular have given the subject its current impetus by invoking arguments based on simple thermodynamic cycles to show exactly how binding constants are related to enzymic rate accelerations, and this theme has been elaborated upon somewhat by Fersht (71) and by Schray & Klinman (72). Tautology or not, transition-state stabilization theory has provided a convincing rationale for the existence of enzyme inhibitors that are orders of magnitude more strongly bound than are substrates and has stimulated efforts aimed at designing new and more potent inhibitors. For the purposes of this review, the theory furnishes a satisfying logical framework for discussing the structural basis of the serine proteases' catalytic mechanism.

Obviously it is beyond the scope of this article to present a detailed review of the theory of enzymic transition-state stabilization. The interested reader is referred especially to the articles by Wolfenden (67) and Lienhard (70). However, the theory is so simple that it may be worthwhile to outline here its basic premises and deductions.

The fundamental assumptions of activated-complex theory or "absolute rate theory" (73) in chemical kinetics are as follows: (a) In every chemical reaction the reactants are in equilibrium with an unstable activated complex, the transition-state complex, which decomposes to give products; the equilibrium constant for formation of the complex is denoted $K^{\ddagger}$. (b) The rate of this decomposition is proportional to the concentration of the complex, with a proportionality constant equal to $k_B T/h$, where k_B is the Boltzmann constant, T is the absolute temperature, and h is Planck's constant. Note that it is assumed the proportionality constant is the same regardless of which reaction is being considered. It follows from a and b that the overall reaction rate constant is given by $K^{\ddagger}(k_B T/h)$.

The activated-complex species are not to be thought of as molecules in the ordinary sense, for they have only transient existence and occur, by definition, at the maxima of the curves relating potential energy to the reaction coordinate (actually at the saddle points of potential energy surfaces). They are characterized by partially formed chemical bonds. A special property of the activated transition-state complex is that it has a unique,

very loose internal mode of vibration, which is unstable with respect to dissociation into products. This "vibration" occurs along a coordinate that is identified as the reaction coordinate.

The additional fundamental assumption relating the theory of enzyme catalysis with the foregoing is, as indicated at the outset of this section, that there is "an active region on the surface of the enzyme which is closely complementary in structure not to the substrate molecule itself, in its normal configuration, but rather to the substrate molecule in a strained configuration, corresponding to the activated complex for the reaction catalyzed by the enzyme" (62). That is to say, the enzyme preferentially binds the activated transition-state complex (referred to for the sake of brevity, albeit somewhat inaccurately, as the transition state), thereby stabilizing it and increasing its concentration, and hence increasing the rate of reaction. It will be seen from this formulation that the venerable "strain" theory of enzyme catalysis (65) is actually equivalent to transition-state stabilization theory, although it has not always been so elegantly and clearly expressed.

By considering a simple thermodynamic cycle, Wolfenden (66, 67) showed for a single-substrate reaction how the dissociation constants for binding of the substrate, K_S, and for binding of the transition state, K_T, are related to formation constants for the transition state when the substrate is bound to the enzyme, $K_E^{\ddagger}$, and when the substrate is free, $K_N^{\ddagger}$, namely $K_S/K_T = K_E^{\ddagger}/K_N^{\ddagger}$. Applying assumption b above, it follows that $K_S/K_T = k_E/k_N$, where k_E is the first-order rate constant for conversion of the enzyme-substrate complex (the Michaelis complex) into enzyme-product complex and k_N is the corresponding rate constant for nonenzymic conversion of the substrate into product. Evidently k_E is the same as k_{cat} in the customary Michaelis-Menten treatment when dissociation of the enzyme-product complex is rapid.

This simple relationship deserves some comment. For one thing, it says that the dissociation constant for binding the substrate must be many orders of magnitude greater than that for binding the transition state, typically by factors in the range 10^8 to 10^{14} (69). Molecules that bear a structural resemblance to the transition state are therefore expected to be bound to the enzyme very strongly indeed. In fact, according to Wolfenden (67), in view of certain possible exceptions the binding ratio K_S/K_T must always be greater than the experimentally observed rate ratio k_E/k_N; however, there appears to be some uncertainty on this point (72). In any event, realization that a rational approach to the design of enormously potent enzyme inhibitors might be possible has stimulated the search for such transition-state analogs with potential clinical applications (74).

From the viewpoint of the theory of enzyme catalysis the Wolfenden relation can be regarded as saying that a necessary condition for catalysis

to occur is that the enzyme should bind the substrate in its transition state more strongly than in its ground state. The task of the structural enzymologist, then, is to guess what the transition-state structure might be and to identify those features of its interaction with the enzyme that contribute to enhancement of its binding over that of the ground state.

A slight complication arises when we come to consider the specific case of the serine proteases. The reactions they catalyze involve not single substrates but two substrates, for example hydrolysis of a peptide bond by water. Additionally the serine protease reaction sequence includes formation of a covalent intermediate, the acyl-enzyme. These factors do not alter the fundamental deductions of the theory outlined above, as may be seen by considering the thermodynamic cycle shown in Figure 2, which is based on Lienhard's treatment of such reactions (70). In this scheme we compare the uncatalyzed hydrolysis of an O-acyl derivative of N-acetyl-L-serine-amide (top row) with the self-catalyzed hydrolysis of the corresponding acyl-enzyme (bottom row). As usual, $K_N^{\ddagger}$ and $K_E^{\ddagger}$ refer to the formation constants of the corresponding activated transition-state complexes. A new feature is that K_A and K_{TX} are no longer dissociation constants but are now equilibrium constants, respectively, for interchange of the acyl group from enzyme to N-acetyl-L-serineamide and for the corresponding interchange of the activated transition-state complex. By the same arguments as before, we find that $K_A/K_{TX} = K_E^{\ddagger}/K_N^{\ddagger} = k_E/k_N$, and we can conclude that the transition state is bound many orders of magnitude more strongly to the enzyme than is the acyl group itself.
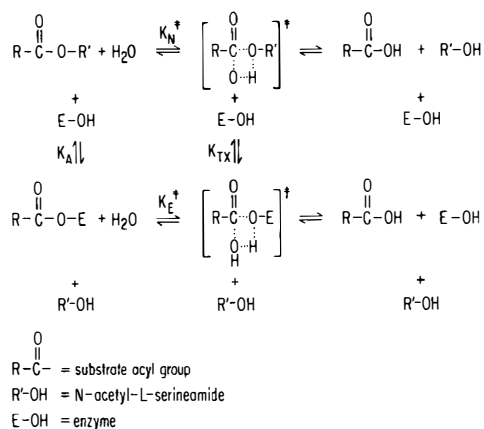

Figure 2 Thermodynamic cycle for transition-state interchange in the deacylation reaction. Based on Lienhard (70).

The reason for focusing on the deacylation step instead of the acylation step is that in doing so we eliminate a complication in the latter case due to the necessity for considering differences in binding entropy between the substrate and the transition state (70). As Lienhard notes, binding of the substrate in the acylation step is accompanied by net loss of translational and overall rotational entropy of one species, and consequently binding of the transition state would be greatly favored by this effect alone. In contrast, transition-state interchange does not result in the net immobilization of one species. Page & Jencks have estimated that such an entropy effect could be worth as much as a factor of 10^8 M in the ratio of the two binding constants (75). No such effect is involved in the deacylation reaction of Figure 2, so transition-state stabilization for this step of the reaction sequence must involve enthalpic interactions between the enzyme and the substrate in its transition state that are not present for the acyl-enzyme. An important point is that by invoking the principle of microscopic reversibility we can argue that any such interactions which contribute to the deacylation reaction must also contribute to the (reverse) acylation reaction, and consequently that more than just entropy effects of the Page & Jencks type must be involved even for acylation.

STRUCTURAL FEATURES PARTICIPATING IN CATALYSIS

Now that I have presented the background for a unified view of the serine protease mechanism, I examine the structural basis of that mechanism in the various members of this class of enzymes.

All of the serine proteases for which X-ray structural studies have been carried out have five features in common. Although the subdivision is somewhat arbitrary, it may be considered that the function of each of these features is to bind a different part of the natural polypeptide substrate in its transition-state configuration. The features referred to are (a) the extended polypeptide binding site on the acyl-group side of the susceptible peptide bond; (b) a number of sites for binding, with greater or lesser specificity, the side chains of a polypeptide substrate; (c) a site for binding the substrate on its leaving-group side; (d) a site [somewhat loosely referred to as the "oxyanion hole" by Robertus et al (33)] for binding the carbonyl oxygen atom of the susceptible peptide bond when the carbonyl group is in a tetrahedral configuration; and (e) the reactive serine side chain, which forms a covalent bond with the carbonyl carbon atom of the susceptible peptide bond, and the well-known "charge relay system" (76), which may be regarded as a site for binding a proton that, in the transition state, is being

shuttled between the reactive serine and the leaving or entering group (this is not the heretofore generally accepted view of the charge relay system).

Each of these five features is found throughout the trypsin family of serine proteases, except that, of course, the side-chain specificity sites are suitably modified from one member of the family to another. But more remarkably, each of the features is also found in the structure of the totally unrelated bacterial serine protease subtilisin, and with almost precisely the same geometrical relationships among the parts (7, 8, 38, 77–79). It is reasonable to suppose that the trypsin family and the subtilisin family evolved independently from different ancestors because the two three-dimensional structures are totally dissimilar and the structural features thought to function in catalysis are contributed by amino acid residues with different orders of occurrence in the sequence. This evolutionary convergence provides powerful evidence that each of the features in question must actually play an important part in enzymic catalysis.

The location, stereochemistry, and probable roles of these various functional sites within the structures of the trypsin family and subtilisin were first established by crystallographic study of small substrate-like inhibitors bound to the enzymes. Earlier ideas have now been definitively confirmed by two new structural determinations on the complex of bovine trypsin with bovine pancreatic trypsin inhibitor (16–19) and of porcine trypsin with soybean trypsin inhibitor (20, 21). These inhibitors are small proteins, respectively 58 and 181 residues long, which bind very strongly and specifically to trypsin. They are representative of a widely occurring group of protease inhibitors that participate in controlling protease activity in diverse biological systems (80, 81).

Before I consider details, it is appropriate to emphasize a point that has been made by Wolfenden (67). In the context of the theory of transition-state stabilization it is "meaningless to distinguish between binding sites and catalytic sites." Clearly this point applies to the five structural features described below and should be kept in mind even though it is sometimes convenient to refer to certain of them as binding sites and others as catalytic sites.

Polypeptide Binding on the Acyl-Group Side

This site was first recognized as such in γ-chymotrypsin (82, 83) and in subtilisin (78, 84). Polypeptides with an L-phenylalanine chloromethyl ketone group at their carboxy-terminal ends were bound to the enzymes and located by difference-Fourier methods. These studies showed that in both enzymes there is a stretch of extended polypeptide backbone chain, Ser-214–Trp-215–Gly-216 in chymotrypsin and Ser-125–Leu-126–Gly-127 in subtilisin, that forms a typical antiparallel β-pair hydrogen-bounded struc-

ture with the polypeptide chain of the inhibitor. The COOH-terminus of the latter is close to the catalytic serine residue on the enzyme, and a later study of halomethyl ketone binding to subtilisin (85) led to the conclusion that members of this group of inhibitors in fact form tetrahedral hemiketal adducts with the serine in addition to alkylating the catalytic histidine. Thus the inhibitor must mimic the way in which the acyl-group of a substrate is bound to the enzyme as a tetrahedral intermediate during catalysis.

It is customary for convenience of discussion (86) to divide a polypeptide substrate into amino acid residues extending from the cleaved bond toward the NH_2-terminus, denoting them as P_1, P_2, P_3 etc (the acyl-group side) and those extending from the cleaved bond toward the COOH-terminus, denoting the latter as P'_1, P'_2, P'_3 etc (the leaving-group side). The corresponding binding sites on the enzyme are denoted S_1, S_2, S_3 and S'_1, S'_2, S'_3.

Three hydrogen bonds can be formed between substrate and enzyme on the acyl-group side: between the carbonyl oxygen of S_1 and the amido nitrogen of P_1; between the backbone amido nitrogen of S_3 and the P_3 carbonyl oxygen; and between the carbonyl oxygen of S_3 and the amido nitrogen of P_3. I discuss the S_1–P_1 hydrogen bond again when the Michaelis complex is considered.

The geometry of this functional component is slightly different in trypsin. Because Ser-218 is deleted in the trypsin sequence, a type I β-bend occurs at residues 216–217–219–220, and the carbonyl group of residue 216 [CO(216)] is turned inward and hydrogen-bonded to NH(220). Thus CO(216) is unavailable for making an S_3–P_3 hydrogen bond (21). Otherwise the antiparallel β-pair interaction between enzyme and substrate at this site is identical with that in chymotrypsin.

Although the antiparallel β-binding segment in chymotrypsin and trypsin is probably not longer than three residues, 214–216, in subtilisin (38) it is likely to include as well Gly-128, Pro-129, and the backbone NH of Ser-130, and thus to comprise five or six residues, 125–130.

There is some question about whether this interaction is quite the same at S_1–P_1 in the Michaelis complex as it is in the covalently bound intermediates. Robertus et al (38) studied the binding to subtilisin of polypeptide virtual substrates corresponding to the acylating portion of good substrates. They were found to be involved in the same antiparallel β-pair just described, but at least in the case of the polypeptide Z-Gly-Gly-L-Tyr, for which an especially clear difference-Fourier map was obtained, the P_1 portion was seen to be bound about 1 Å further out from the enzyme surface than were the covalent polypeptide inhibitors (78, 79). More specifically, the S_1–P_1 hydrogen bond distance involving CO(125) was too long by about 1 Å to make a normal hydrogen-bonded interaction and the P_1 tyrosine side chain did not fit all the way into the S_1 specificity crevice (described below).

Robertus et al (38) hypothesized that this looser binding at P_1 is characteristic of the Michaelis complex in contrast to the tetrahedral intermediate or the acyl-enzyme. On the other hand, Blow (25) states that there is no difference in this regard between the binding geometries seen in the trypsin-inhibitor complexes (presumed to represent a tetrahedral covalent intermediate) and in the complex of the virtual substrate N-formyl-L-tryptophan with chymotrypsin (a model for the Michaelis complex). Although the $NH \cdots CO(214)$ distance in question in the trypsin complexes is 3.2–3.3 Å and greater than 3.5 Å in the chymotrypsin complex, as Blow (25) points out this difference is probably not experimentally significant. Also, it should be emphasized that the proposal of Robertus et al (38) was based on comparison of two similar difference-Fourier maps, a very sensitive technique for detecting small movements.

Several lines of evidence indicate that the interpretation of Robertus et al (38) regarding this point is also valid for the trypsin-chymotrypsin family. Henderson (39) notes that the indolyl ring of indoleacryloyl-chymotrypsin lies 0.5–1.0 Å deeper into the S_1 specificity crevice than does the corresponding ring in the complex with N-formyl-L-tryptophan. Moreover, Ingles & Knowles (87) have presented kinetic evidence that the presence of an acylamido group at P_1 is important for catalysis (transition-state binding) but not for initial formation of the Michaelis complex in chymotrypsin. A similar conclusion has been reached by Kobayashi & Ishii (88) for trypsin. Thus closer binding of covalent intermediates compared with Michaelis complexes is almost certainly real for subtilisin and cannot be dismissed for the trypsin family of enzymes. This aspect of substrate binding is important because it provides a mechanism contributing to tighter binding in the transition state than in the Michaelis complex, as required to explain enzymic catalysis.

Side-Chain Specificity Sites

Trypsin, the most specific of the serine proteases, attacks preferentially at peptide bonds following an arginine or a lysine residue at P_1. Chymotrypsin rapidly hydrolyzes peptide bonds following an aromatic side chain at P_1. Elastase does not display quite such a marked specificity, but generally prefers an uncharged nonaromatic side chain, especially alanine, at the P_1 position. Subtilisin most nearly resembles chymotrypsin with respect to specificity at P_1, but will accept nonaromatic apolar side chains and even charged side chains at P_1 when the substrate residue at P_3 or P_4 is hydrophobic.

This behavior is readily understood in structural terms. When a polypeptide substrate is bound at the site described in the previous section, the P_1 side chain fits into a crevice on the enzyme surface (37). One side of this

crevice is composed of planar peptide links in the stretch of the enzyme's backbone chain that forms the S_1–S_2–S_3 antiparallel β-binding segment, residues 214 to 216 in the trypsin family, and 125 to 127 in subtilisin (79). The other wall of the crevice is made up of the backbone peptide links 190–191 and 191–192 in the trypsin family, and of a somewhat more irregular surface composed of the side chains and backbone of segment Ala-152–Ala-153–Gly-154 in subtilisin. Thus the S_1 specificity crevice in chymotrypsin will best accommodate a planar aromatic side chain, but is not well suited to accept other shapes, whereas subtilisin will accept aromatic side chains quite well and other side chains if necessary. The S_1 specificity crevice in trypsin is very similar to that in chymotrypsin, except that Ser-189 is replaced by Asp-189. Consequently lysine and especially arginine side chains at P_1 are stabilized when bound in the S_1 crevice by interaction with Asp-189 (12, 13, 16, 21). In elastase (11) the S_1 crevice is partially occluded by the side chain of Val-216, and the bottom is partly filled by the side chain of Thr-226, leaving room for binding of small P_1 side chains only. Both of these residues are glycine in trypsin and chymotrypsin. Because of the small binding energy available from the interaction between the P_1 side chain and the S_1 specificity crevice in elastase, catalysis is more dependent upon enzyme-substrate contacts remote from the susceptible bond. Thus elastase is about 4000 times more efficient at hydrolyzing polypeptide substrates like Ac-Ala-Ala-Ala-OMe than the single amino acid esters Ac-Ala-OMe (89), and displays considerable subsite specificity out to subsite S_5 (90–93). However it is not yet clear how the extended polypeptide substrate interacts with elastase (93, 94). Such greatly enhanced substrate reactivity when P_2 or P_3 residues are added on is not seen for chymotrypsin and trypsin, although there is a small effect (95).

Although subtilisin resembles chymotrypsin with respect to side-chain specificity at P_1, it also behaves somewhat like elastase in that it is much more effective against extended polypeptide substrates (96–98). Subtilisin differs from elastase, however, in that it possesses an elaborate S_4 binding crevice that can accommodate an aromatic side chain at P_4 (38, 79). This crevice is evidently rather adjustable; when it is occupied by an aromatic group, the side chain of Tyr-104 swings out of the way and lies parallel to the former. It is also noteworthy that the S_4 crevice apparently has an even higher affinity for a benzene ring than does the S_1 crevice (38).

In an unpublished study, J. J. Birktoft has examined the literature on subtilisin-mediated peptide cleavages in proteins and has found that there are numerous examples where P_1 is an "abnormal" residue for subtilisin (that is, not hydrophobic). In most cases these abnormalities can be explained by overriding specificity for an aromatic residue at P_4 instead. Such a tandem arrangement of side-chain specificity sites would be appropriate

for promoting hydrolysis of longer polypeptide chains to tripeptides, which might then be absorbed more efficiently than are single amino acids by the bacilli secreting a subtilisin type of enzyme.

Polypeptide Binding on the Leaving-Group Side

The geometry of this interaction is less well-established than that on the acylating-group side. One reason is that it is probably weaker and much less specific, and therefore it is not easy to design inhibitors that will bind strongly in the productive mode and with a well-defined geometry on the leaving-group side of the bond-splitting site. In fact, the only clear depiction of such binding that has been obtained comes from X-ray studies of two trypsin-inhibitor complexes (16–21).

Fersht, Blow & Fastrez (99) presented the first kinetic data clearly showing favorable interaction between a P'_1 residue and an S'_1 site in chymotrypsin. They investigated the reaction that is the reverse of acylation of chymotrypsin by amides, namely, deacylation of the acyl-enzyme by amines. It was found for example that with specific acyl-enzymes, such as acetyl-phenylalanyl-chymotrypsin, the latter reaction went 21 times faster with alaninamide acting as the attacking amine than with the normally very reactive hydrazine, and 800 times faster than with ammonia. Alanine itself was relatively ineffective. This is consistent with the presence in chymotrypsin of a S'_1 site capable of binding a P'_1 residue if the latter does not have a free carboxyl group. A similar conclusion for subtilisin is suggested by relative hydrolysis rates of synthetic polypeptides of the type Z-Gly-Leu- and Z-Tyr- (96, 98, 100). Moreover, there appears to be an S'_2 and possibly an S'_3 site in subtilisin. Kinetic parameters for a series of $(Ala)_4$- and $(Ala)_3$-peptides, amides, and esters (93) clearly show that elastase has binding sites for P'_1, P'_2, and P'_3 but not for P'_4. Moreover the data reveal that S'_1 and S'_2 are hydrophobic. For none of the serine proteases, however, are there sufficient data to reveal any possible preference for specific side chains on the leaving-group side.

The crystal structures of bovine trypsin with pancreatic trypsin inhibitor (16–19) and of porcine trypsin with soybean trypsin inhibitor (20, 21) have yielded the only reliable picture of interaction between the substrate leaving group and the enzyme. Perhaps the most significant fact about these interactions in the two structures is that they are somewhat different, as might be expected for interactions that are relatively weak and nonspecific. The P'_1 side chain, which is alanine in the pancreatic inhibitor and isoleucine in the soybean inhibitor, is in contact with a hydrophobic region including the disulfide bridge Cys-42-Cys-58. The P'_2 residue is arginine in both inhibitors (although not in all trypsin inhibitors), and in both it is seen to

participate in what is probably a charge-transfer interaction with Tyr-151. However, in spite of the similar side-chain interactions, the two inhibitor backbone chains are not in the same ϕ, ψ conformation. Possibly the difference is only an artifact of interpretation due to the relatively lower resolution (2.6 Å vs 1.9 Å) and level of refinement of the soybean trypsin inhibitor complex structure. In neither case, however, does the S' backbone chain form an antiparallel β-structure with the P' chain of the inhibitor.

Robertus et al (38) had to rely on model building alone to derive a proposed leaving-group binding geometry in subtilisin-substrate complexes, although its probable location was based on weak density observed in difference-Fourier maps. In contrast to the trypsin case, they depict the leaving-group polypeptide chain as forming a second short stretch of antiparallel β-pair structure with the backbone chain segment at Asn-218. The leaving-group side chains are shown as interacting with the enzyme surface at Tyr-217 and Phe-189.

A question remains concerning the binding site for an entering water molecule when the acyl-enzyme undergoes hydrolysis. The obvious choice is simply to assume that it goes in the same place as does the proton-accepting terminal NH of the leaving group in the tetrahedral intermediate. This assumption would satisfy the principle of microscopic reversibility, but it is not a thermodynamic requirement when the leaving and entering groups are not identical. Henderson (39) has proposed that the water molecule in question is hydrogen-bonded to both the carbonyl oxygen atom of the acylating group and to $N\epsilon2$[1] of His-57, where a water molecule is seen in indoleacryloyl-chymotrypsin. This position appears to be close to the terminal NH site. As Henderson emphasizes, however, indoleacryloyl-chymotrypsin is an abortive acyl-enzyme.

There are serious difficulties associated with identifying the site of water binding. First, because they are crystallized from media high in sulfate, a bound sulfate ion is observed (when the pH is below 8) very close to this location in the native crystal form of α-chymotrypsin (101), γ-chymotrypsin (102), elastase (94), benzamidine-trypsin (15), and subtilisin (103). A phosphate ion is similarly bound in *Streptomyces griseus* protease B (22, 23). Displacement of this anion on binding covalent inhibitors and virtual substrates tends to confuse the resulting difference maps in its vicinity. The significance of this anion binding is obscure; possibly it may be a phenomenon associated with the presence at this location of a specific binding site for the proton-accepting terminal NH of the leaving group or for the attacking water molecule. A second difficulty in identifying this site is that

[1] *Biochemistry,* 1970, 9:3471-79.

it may be incomplete or deformed unless an acyl group is properly in place, as suggested by kinetic data on hydrolysis of N-methyl-N-acetyltyrosyl-chymotrypsin (104).

One important reason for precisely defining the leaving-group and water-binding site is that it would be interesting to know whether their geometrical relationship to the other binding sites could promote a tetrahedral distortion at the carbonyl carbon of the substrate bond being cleaved. Closely related is the question of whether this site could also favor pyramidalization of the terminal proton-acceptor atom of the leaving group, as might be anticipated from considerations of transition-state stabilization. Very little comment on these questions can be found in the literature.

Tetrahedral Binding and the Oxyanion Binding Site

The fundamental thesis of transition-state theory applied to enzymes is that an enzyme functions by furnishing a template complementary to the substrate in its transition-state configuration. An immediate corollary is that any molecule bearing a structural resemblance to the enzymic transition state (a transition-state analog) will be bound very tightly. Accordingly, serine proteases should exhibit high affinity for substances that can form stable tetrahedral adducts with the reactive serine. This phenomenon is now recognized to be the explanation for the behavior of several classical serine protease inhibitors.

In the simplest structural terms, the four ligands attached to the substrate's carbonyl carbon atom are constrained by their interactions with the enzyme to be tetrahedrally disposed about that atom. I have already discussed the acylating or P_1 portion and the P'_1 or leaving-group portion. In the next section I deal with covalent bond formation at the reactive serine and later the substrate's carbonyl oxygen atom, which must be an oxyanion in the addition intermediate.

The first suggestion that there is a specific binding site for a tetrahedrally skewed carbonyl oxygen anion was made by Henderson (39, 105). He proposed that there might exist a pair of hydrogen bonds from the backbone NH of the reactive Ser-195 and of Gly-193 to the carbonyl oxygen in the (nontetrahedral, planar trigonal) acyl-enzyme as represented by indoleacryloyl-chymotrypsin. He further noted that such hydrogen bonds "might be stronger in a tetrahedral intermediate thereby effectively lowering its free energy and favoring its formation" (39). Robertus et al (38) decided, on the basis of difference maps and model building, that a corresponding pair of hydrogen bonds was present in the subtilisin-substrate complex, involving the backbone NH of reactive Ser-221 and the side chain NH_2 of Asn-155. However, they found that these two hydrogen bonds cannot form in the Michaelis complex or in the acyl-enzyme, but only when the substrate

carbonyl is covalently linked to the enzyme and in the tetrahedral configuration. This site of oxyanion binding was called an "oxyanion hole," and it was suggested that this feature supplies much of the required transition-state stabilization with respect to both the Michaelis complex and the acyl-enzyme.

Subsequently it has been found that in all of the covalent inhibitor-serine protease complexes the inhibitor also assumes a tetrahedral geometry with an oxygen or hydroxyl group similarly placed in the oxyanion binding site. This situation is summarized in Figure 3.
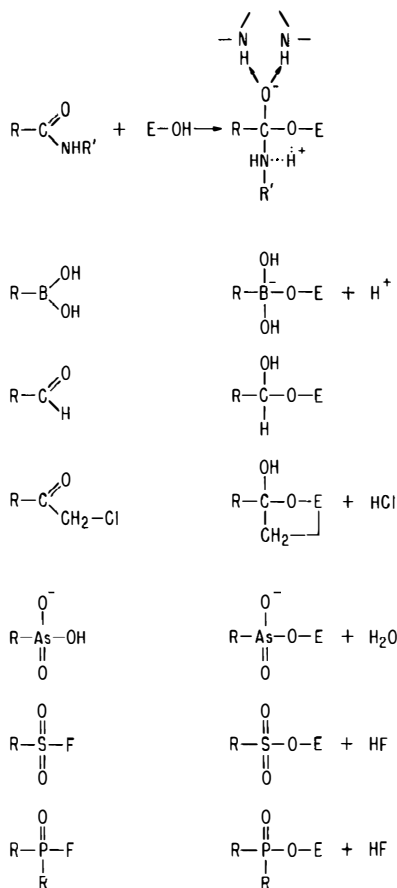
Figure 3   Comparison between substrate in the tetrahedral transition state complex (*top*) and covalent inhibitors. Note that all of the inhibitors also form tetrahedral adducts with an oxygen or hydroxyl group in position to accept hydrogen bonds from the same pair of donors.

The first entry in Figure 3 represents interaction with the enzyme of either a normal amide or a peptide substrate in its transition state or a specific protein inhibitor like bovine pancreatic trypsin inhibitor (16–19) or soybean trypsin inhibitor (20, 21). In both enzyme-inhibitor complexes the reactive carbonyl group in the inhibitor is decidedly tetrahedral, even though in the bovine pancreatic inhibitor-trypsin complex it appears that the distance between the serine Oγ and the carbonyl carbon is about 2.5 Å, too long for a covalent bond but too short for a comfortable van der Waals approach. Moreover, Huber et al (18) have also determined the structure of the same complex when the enzyme's reactive Ser-195 has been chemically converted to dehydroalanine, and find that the same high degree of pyramidalization at the reactive carbonyl carbon in the inhibitor is maintained. Evidently tetrahedral geometry of the complex is dependent upon the existence of a suitably complementary surface on the enzyme, predisposition by the inhibitor itself to become tetrahedrally distorted, and interaction between the two molecules. Markley & Porubcan (106) have presented proton magnetic resonance data indicating that the extra proton is bound to His-57 in the complex between porcine β-trypsin and bovine pancreatic trypsin inhibitor.

The second entry in Figure 3 represents the boronic acid adducts with serine proteases. Matthews et al (40) have described the crystal structures of two aromatic boronic acid adducts with subtilisin and concluded that they are good models for the transition-state geometry proposed by Robertus et al (38).

Aldehyde-enzyme complexes are represented on the third line of Figure 3. Although no crystal structures have been reported for such adducts, it seems very probable that the enzymic transition state is in fact mimicked by hemiacetal complexes of the kind depicted here. This inference is supported by the observation that specific polypeptide aldehydes are known to be powerful serine protease inhibitors (107, 108), by ^{13}C magnetic resonance data on a tripeptide aldehyde adduct with α-lytic protease (109), and by analogy with the crystallographic structures of polypeptide halomethyl ketone derivatives of subtilisin, which form tetrahedral hemiketal complexes (41) with the enzyme. The latter are depicted immediately below the aldehydes in Figure 3, with an additional covalent bond to the enzyme representing alkylation of the catalytic hisidine residue as well.

Arsonic acids were discovered by Glazer (110, 111) to inhibit serine proteases with a typical enzymic pH dependence. He proposed that they bind covalently at the reactive serine, and an unpublished difference-Fourier map for the subtilisin complex with 4-(4'-aminophenylazo)phenylarsonic acid obtained by J. D. Robertus was consistent with this hypothesis.

The aromatic sulfonyl fluorides are well-known covalent inhibitors of serine proteases. Several of the original serine protease crystal structures were in fact solved using the enzymes inhibited by sulfonyl fluorides to avoid autocatalytic degradation: tosyl-α-chymotrypsin (1), phenylmethanesulfonyl-subtilisin BPN' (7), tosyl-γ-chymotrypsin (5), and tosyl-elastase (10, 11). An especially clear stereoscopic depiction of the toluenesulfonyl group bound to α-chymotrypsin, with one sulfonyl oxygen atom obviously occupying the oxyanion binding site, can be seen in Figure 5 of the article by Henderson (39).

Inhibition of any protease by diisopropyl fluorophosphonate, represented by the last line in Figure 3, is generally considered diagnostic for its identification as a serine protease. We believe that this reagent is also a transition state analog, which forms a stable covalent tetrahedral adduct with the enzyme. In the article by Stroud et al (12), Figure 12 shows a stereoscopic view of the diisopropylphosphoryl group (DIP) bound to trypsin in which O_3 of the DIP group accepts hydrogen bonds from the backbone NH of Ser-195 and of Gly-193, in conformity with all of the foregoing inhibitors.

The Reactive Serine and the Charge Relay System

The class of serine proteases is defined by the presence of a uniquely reactive serine side chain, which makes a covalent ester bond to the carbonyl carbon atom of the susceptible bond in substrates to form an acyl-enzyme, and which reacts similarly with a number of covalent inhibitors (see Figure 3), most notably diisopropyl fluorophosphonate. It has long been known from chemical and kinetic pH studies that a histidine side chain is also involved in the enzymic process, and the first interpretable electron-density map of chymotrypsin (1) showed that indeed the side chain of His-57 is close to that of the reactive Ser-195. On the basis of an improved electron-density map and of the invariance of Asp-102 in several homologous sequences, Blow, Birktoft & Hartley (76) proposed that the Ser-195 side chain is hydrogen-bonded to the His-57 imidazole, which is in turn hydrogen-bonded to the buried carboxylate of Asp-102. They further proposed that polarization of the system, which they dubbed the "charge relay system," due to negative charge on the carboxylate would make the serine oxygen nucleophilic and thus reactive toward substrates.

Almost immediately subtilisin was discovered to contain a reactive Ser-221 close to and presumably hydrogen-bonded to His-64 (7), which in turn is hydrogen-bonded to a buried Asp-32 (8). The geometrical relationship between the functional groups in this hydrogen-bonded network is the same in subtilisin as in chymotrypsin, despite the fact that the two are otherwise structurally unrelated; this suggests strongly that indeed the charge relay system must perform the same fundamental mechanistic function through-

out the entire class of serine proteases. The same arrangement is also found, as expected, in the other members of the chymotrypsin family, elastase (11, 94), trypsin (12, 15, 112), and *S. griseus* protease B (22, 23), which are all structurally and sequentially homologous with chymotrypsin.

The geometry of the charge relay system, at pH above about 7, can be schematically represented for purposes of discussion as $-COO^- \cdots HN-Im-N \cdots HO-CH_2-$, where the nitrogen atom to the left is designated $N\delta1$ (or N1), that to the right is $N\epsilon2$ (or N3), and the serine oxygen is $O\gamma$. This functional grouping has been the subject of considerable study by the methods of magnetic resonance spectroscopy (106, 109, 113–116), physical organic chemistry on model compounds and reactions (117, 118), and molecular orbital theory (119–122). Most of these studies [but not all—see (117)] have yielded results that are interpreted as supporting the generally accepted view put forward on the basis of X-ray crystallography.

To anticipate somewhat, a rather modified conception of the structure and function of the charge relay system (103) now appears to be more acceptable. It holds that the hydrogen bond between $O\gamma$ of the reactive serine and $N\epsilon2$ of the histidine is distorted to the point where it is nonexistent or at best very weak; that the serine is not intrinsically nucleophilic, but instead reacts simply because it is in the optimum position to attack a tetrahedrally distorted carbonyl carbon atom in the substrate; that this distortion is induced by binding to the enzyme; and that the His-Asp couple functions to facilitate transfer of a proton either from the attacking serine $O\gamma$ to the leaving group in the acylation step, or from an attacking nucleophile to the serine $O\gamma$ in deacylation. Thus if this modified view is correct, a better name for the charge relay system might be "proton relay system." In other language, the His-Asp couple can be said to stabilize the enzymic transition state by binding the proton being transferred.

From the outset it was evident that there were difficulties inherent in the original formulation of the operation of the charge relay system. Its upper pK_a is 6.5 to 7, which is normal for histidine; therefore, one would expect very little proton transfer to the histidine from a serine with a pK of about 14, and very little tendency for the serine to exist as a highly nucleophilic alkoxide ion. Polgár & Bender (123) have argued from pH and kinetic deuterium isotope effects on acyl-enzyme formation with subtilisin and thiol-subtilisin (in which Ser-221 is chemically converted to cysteine) that thiol-subtilisin, but not subtilisin itself, contains a hydrogen bond to His-64. Polgár (124) has also pointed out that on the basis of the previously accepted picture of the charge relay system, only promotion of the first proton transfer, i.e. the general base-catalyzed removal of the proton from serine, would be explained; the second proton transfer, general acid-catalyzed donation of a proton to the leaving group, would actually be inhibited.

Earlier X-ray crystallographic data are also puzzling. Chymotrypsino-
gen, the zymogen precursor of chymotrypsin, was shown to contain a
charge relay system that was very similar to the one in chymotrypsin (6,
29). If nucleophilic reactivity of the serine is caused by this particular
structural feature alone, why is the zymogen inactive or, as was later
discovered (125), only about 10^{-7} as active as the enzyme? The converse
problem exists for methyl-chymotrypsin, in which His-57 is methylated at
Nϵ2 (126–128). It was expected that the modified enzyme would show no
activity at all. Henderson (127) reported, however, that the modified en-
zyme still possesses a low level of inherent activity, with catalytic rates
reduced by 2×10^{-4} to 5×10^{-6} for specific substrates, and crystal struc-
ture analysis (128) revealed only very slight shifts in the positions of the
His-57 and Ser-195 side chains. Methylation of His-57 should have entirely
abolished reactivity of the serine if the original theory of the charge relay
system were correct.

The foregoing is suggestive; however the most compelling evidence that
the reactive serine probably does not receive an induced negative charge by
hydrogen bonding to histidine comes from improved treatment of the X-ray
data. Recently Matthews et al (103) have reported that although crystallo-
graphic refinement of the subtilisin structure is incomplete, it justifies the
conclusion that there is either no hydrogen bond between the reactive
Ser-221 and His-64 or at most a very weak one; examination of other
published serine protease structures supports the same conclusion. Con-
versely, chymotrypsinogen is now reported to contain an apparently normal
hydrogen bond between Ser-195 and His-57 (129).

Also, although recent molecular orbital calculations by Demoulin &
Allen (122) were based on atom coordinates that were adjusted to minimize
the energy of their model system, these calculations showed that there is
only slight activation of Ser-195 in the native enzyme, but that binding of
a substrate in its tetrahedral configuration promotes transfer of the serine's
proton. Similar conclusions were drawn by Umeyama et al (119) and by
Scheiner, Kleier & Lipscomb (121).

Whether or not there is actually a hydrogen bond between Nϵ2 of the
histidine in the charge relay system and Oγ of the reactive serine, all
available evidence clearly supports the existence of a good hydrogen bond
between the histidine Nδ1 and the carboxylate side chain of a buried
aspartic acid. How this couple might operate in the enzymic mechanism has
been suggested by ^{13}C NMR data reported by Hunkapiller et al (113) for
α-lytic proteases. This is a bacterial serine protease from *Myxobacter 495*
that is sequentially homologous with and therefore presumed to have the
same structure as the trypsin family. Hunkapiller et al identified the reso-
nance corresponding to Cδ2 after artificial ^{13}C enrichment at this position

in the enzyme's single histidine. Both the chemical shift of $C_\delta 2$ and coupling with its directly bonded hydrogen atom were monitored as a function of pH. Hunkapiller et al concluded that the imidazole ring of the histidine residue becomes positively charged only below pH 4, and that the carboxylate of the aspartic acid residue is ionized only above pH 6.7. This result strongly supports the idea that transfer of an additional proton onto $N\epsilon 2$ of the imidazole ring causes a simultaneous shift of the first proton from $N\delta 1$ onto the carboxylate group, leaving the imidazole neutral instead of positively charged, and the carboxyl group neutral instead of negatively charged. It should not matter whether the additional proton comes from the surrounding medium at pH below 6.7 or from the reactive serine $O\gamma$ as it attacks the substrate. Evidently this special property of the His-Asp couple is designed to minimize the activation energy of proton transfer in the latter case as it can then occur without unfavorable charge separation.

Thus, a schematic representation of the charge relay system when the proton is being transferred would be $-COOH \cdots N-Im-NH \cdots {}^-O-CH_2-$. This tautomeric form would not be expected to occur, however, except when a substrate is present in its tetrahedral transition state.

In a subsequent ^{13}C-NMR experiment, Hunkapiller, Smallcombe & Richards (109) also examined a polypeptide aldehyde complex with α-lytic protease. This enzyme-inhibitor complex should resemble the transition-state complex, and indeed the data further confirmed the foregoing ideas.

Evidence that the His-Asp couple in trypsin behaves in the same way as in α-lytic protease is also accumulating. Markley & Porubcan (106) have identified the proton magnetic resonance peak corresponding to the hydrogen atom attached to $C_\delta 2$ of the charge relay histidine in porcine β-trypsin. The peak showed abnormal titration behavior, with a pK of 5.0. Furthermore, in the complex between porcine β-trypsin and bovine pancreatic trypsin inhibitor, its chemical shift resembled that of the same atom in the protonated charge relay system of the free enzyme. Additionally Koeppe & Stroud (130) have demonstrated by difference-infrared titration that the pK_a of Asp-102 in bovine trypsin is 6.8. Both sets of experiments show that the expected pK reversal within the His-Asp couple proposed by Hunkapiller et al (113) also occurs in trypsin.

Some degree of confusion remains, however, about the nature of the charge relay system. Robillard & Shulman (114–116) have ascribed a low-field resonance in α-chymotrypsin, chymotrypsinogen, trypsin, trypsinogen, subtilisin, and α-lytic protease to a N-H proton in a hydrogen bond between the His and Asp of the charge relay system. In both δ-chymotrypsin and chymotrypsinogen the chemical shift of this resonance titrates with a pK of 7.5. Robillard & Shulman dispute the interpretation by Hunkapiller et al (113) and assert that the simplest explanation for their own results

would be net protonation of His-57 with a pK of 7.5. Resolution of these apparent differences is awaited with interest.

An unusual suggestion has been made by Brinigar & Chao (131), who note that a sulfate or phosphate anion is always bound in the neighborhood of the charge relay system at pH's below the pK of the system. They propose that the aspartic acid side chain is in fact still protonated at the higher, active pH and that at the lower pH a positive charge would be present on the histidine, promoting binding of a counterionic anion. In this scheme the charge relay system contains one more proton than it does in the conventional scheme. However, Demoulin & Allen (122) discount this possibility, arguing that their molecular orbital calculations demonstrate that a negatively charged aspartate is essential to the mechanism of proton transfer, and pointing out that charged species (the anion) can also form strong hydrogen bonds with neutral molecules (the histidine side chain).

There remain several additional open questions that can only be touched upon. Is proton transfer between the reactive serine and the leaving group or the attacking nucleophile a separate step, or is it concerted with the other bond-breaking or bond-making steps? Satterthwait & Jencks (132) considered the evidence that there is little charge development on a leaving or attacking amine group in the transition state and that a kinetic deuterium isotope effect of 2 to 3 is observed, and concluded that the data are insufficient to distinguish among six closely related but distinct detailed mechanisms. An interesting subsidiary question is whether the charge relay histidine operates as a stationary relay as proposed by Wang (133), or by acting as a mobile flip-flop between the serine $O\gamma$ and the leaving or attacking group. Is the His-Asp hydrogen bond ever bifurcated between $O\delta1$ and $O\delta2$ of the aspartate, as it has sometimes been shown in published illustrations, or does it always involve only $O\delta2$? What is the role, if any, of the second serine, Ser-214 in the chymotrypsin family or Ser-33 in subtilisin, hydrogen-bonded to the charge relay aspartate side chain? Are small movements within the enzyme structure that are observed on binding of inhibitors in the S_1 crevice accidental, or are they related to reported substrate activation effects in trypsin (134, 135)?

Literature Cited

1. Matthews, B. W., Sigler, P. B., Henderson, R., Blow, D. M. 1967. *Nature* 214:652-56
2. Birktoft, J. J., Blow, D. M. 1972. *J. Mol. Biol.* 68:187-240
3. Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V., Wright, L. H. 1973. *Biochemistry* 12:4185-92
4. Vandlen, R. L., Tulinsky, A. 1973. *Biochemistry* 12:4193-4200
5. Cohen, G. H., Silverton, E. W., Matthews, B. W., Braxton, H., Davies, D. R. 1969. *J. Mol. Biol.* 44:129-41
6. Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T., Nguyen-Huu-Xuong 1970. *Biochemistry* 9:1997-2009
7. Wright, C. S., Alden, R. A., Kraut, J. 1969. *Nature* 221:235-42
8. Alden, R. A., Wright, C. S., Kraut, J. 1970. *Philos. Trans. Roy. Soc. London* B257:119-24
9. Drenth, J., Hol, W. G. J., Jansonius, J. N., Koekoek, R. 1972. *Eur. J. Biochem.* 26:177-81
10. Watson, H. C., Shotton, D. M., Cox, J. M., Muirhead, H. 1970. *Nature* 225:806-11
11. Shotton, D. M., Watson, H. C. 1970. *Nature* 225:811-16
12. Stroud, R. M., Kay, L. M., Dickerson, R. E. 1974. *J. Mol. Biol.* 83:185-208
13. Krieger, M., Kay, L. M., Stroud, R. M. 1974. *J. Mol. Biol.* 83:209-30
14. Fehlhammer, H., Bode, W. 1975. *J. Mol. Biol.* 98:683-92
15. Bode, W., Schwager, P. 1975. *J. Mol. Biol.* 98:693-717
16. Rühlmann, A., Kukla, D., Schwager, P., Bartels, K., Huber, R. 1973. *J. Mol. Biol.* 77:417-36
17. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., Steigemann, W. 1974. *J. Mol. Biol.* 89:73-101
18. Huber, R., Bode, W., Kukla, D., Kohl, U., Ryan, C. A. 1975. *Biophys. Struct. Mechanism* 1:189-201
19. Bode, W., Schwager, P., Huber, R. 1975. *Proc. 10th FEBS Meet.* pp. 3-20
20. Blow, D. M., Janin, J., Sweet, R. M. 1974. *Nature* 249:54-57
21. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., Blow, D. M. 1974. *Biochemistry* 13:4212-28
22. Codding, P. W., Delbaere, L. T. J., Hayakawa, K., Hutcheon, W. L. B., James, M. N. G., Jurášek, L. 1974. *Can. J. Biochem.* 52:208-20
23. Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G., Thiessen, W. E. 1975. *Nature* 257:758-63
24. Kossiakoff, A. A., Chambers, J. L., Kay, L. M., Stroud, R. M. 1977. *Biochemistry.* In press.
24a. Bode, W., Fehlhammer, H., Huber, R. 1976. *J. Mol. Biol.* 106:325-35
25. Blow, D. M. 1976. *Acc. Chem. Res.* 9:145-52
26. Stroud, R. M., Krieger, M., Koeppe, R. E., Kossiakoff, A. A., Chambers, J. L. 1975. In *Proteases and Biological Control,* ed. E. Reich, D. B. Rifkin, E. Shaw, pp. 13-32. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 1021 pp.
27. Bender, M. L., Killheffer, J. V. 1973. *CRC Crit. Rev. Biochem.* 1:149-99
28. Stroud, R. M. 1974. *Sci. Am.* 231:74-88
29. Kraut, J. 1971. *The Enzymes* 3:165-83
30. Neurath, H. 1975. See Ref. 26, pp. 51-64
31. De Haen, C., Neurath, H., Teller, D. C. 1975. *J. Mol. Biol.* 92:225-59
32. Keil, B. 1971. *The Enzymes* 3:249-75
33. Magnusson, S. 1971. *The Enzymes* 3:277-321
34. Hartley, B. S., Shotton, D. M. 1971. *The Enzymes* 3:323-73
35. Glazer, A. N., Smith, E. L. 1971. *The Enzymes* 3:561-608
36. Bender, M. L., Kemp, K. C. 1957. *J. Am. Chem. Soc.* 79:116-20
37. Steitz, T. A., Henderson, R., Blow, D. M. 1969. *J. Mol. Biol.* 46:337-48
38. Robertus, J. D., Kraut, J., Alden, R. A., Birktoft, J. 1972. *Biochemistry* 11:4293-4303
39. Henderson, R. 1970. *J. Mol. Biol.* 54:341-54
40. Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., Kraut, J. 1975. *J. Biol. Chem.* 250:7120-26
41. Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J., Kraut, J. 1976. *J. Biol. Chem.* 251:1097-1103
42. Johnson, C. H., Knowles, J. R. 1966. *Biochem. J.* 101:56-62
43. Hartley, B. S., Kilby, B. A. 1954. *Biochem. J.* 56:288-97
44. Zerner, B., Bond, R. P. M., Bender, M. L. 1964. *J. Am. Chem. Soc.* 86:3674-79
45. Zerner, B., Bender, M. L. 1964. *J. Am. Chem. Soc.* 86:3669-74
46. Fastrez, J., Fersht, A. R. 1973. *Biochemistry* 12:2025-34
47. Jencks, W. P. 1969. *Catalysis in Chemistry and Enzymology,* pp. 463-554. New York: McGraw-Hill. 644 pp.
48. Caplow, M. 1969. *J. Am. Chem. Soc.* 91:3639-45
49. Lucas, E. C., Caplow, M. 1972. *J. Am. Chem. Soc.* 94:960-63

50. Bender, M. L., Philipp, M. 1973. *J. Am. Chem. Soc.* 95:1665–66
51. Frankfater, A., Kézdy, F. J. 1971. *J. Am. Chem. Soc.* 93:4039–43
52. Fersht, A. R., Requena, Y. 1971. *J. Am. Chem. Soc.* 93:7079–87
53. O'Leary, M. H., Kluetz, M. D. 1972. *J. Am. Chem. Soc.* 94:3585–89
54. Yu, S. H., Viswanatha, T. 1969. *Eur. J. Biochem.* 11:347–52
55. Hess, G. P., McConn, J., Ku, E., McConkey, G. 1970. *Philos. Trans. Roy. Soc. London* B257:89–104
56. Coletti-Previero, M.-A., Axelrud-Cavadore, C., Previero, A. 1970. *FEBS Lett.* 11:213–17
57. Fink, A. 1976. *Biochemistry* 15:1580–86
58. Douzou, P., Hui Bon Hoa, G., Petsko, G. A. 1975. *J. Mol. Biol.* 96:367–80
59. Petsko, G. A. 1975. *J. Mol. Biol.* 96:381–92
60. Alber, T., Petsko, G. A., Tsernoglou, D. 1976. *Nature* 263:297–300
61. Fink, A., Ahmed, A. I. 1976. *Nature* 263:294–97
62. Pauling, L. 1946. *Chem. Eng. News* 24:1375–77
63. Pauling, L. 1948. *Am. Sci.* 36:51–58
64. Bender, M. L., Kézdy, F. J., Gunter, C. R. 1964. *J. Am. Chem. Soc.* 86:3714–21
65. Jencks, W. P. 1966. In *Current Aspects of Biochemical Energetics,* ed. N. O. Kaplan, E. P. Kennedy, pp. 273–98. New York: Academic. 486 pp.
66. Wolfenden, R. 1969. *Nature* 223:704–5
67. Wolfenden, R. 1972. *Acc. Chem. Res.* 5:10–18
68. Wolfenden, R. 1974. *Mol. Cell. Biochem.* 3:207–11
69. Lienhard, G. E. 1972. *Ann. Rep. Med. Chem.* 7:249–58
70. Lienhard, G. E. 1973. *Science* 180:149–54
71. Fersht, A. R. 1974. *Proc. R. Soc. London Ser. B* 187:397–407
72. Schray, K., Klinman, J. P. 1974. *Biochem. Biophys. Res. Commun.* 57:641–48
73. Laidler, K. J. 1969. *Theories of Chemical Reaction Rates,* p. 41. New York: McGraw-Hill. 234 pp.
74. Lindquist, R. N. 1975. In *Drug Design,* ed. E. J. Ariëns, 5:23–80. New York: Academic. 356 pp.
75. Page, M. I., Jencks, W. P. 1971. *Proc. Natl. Acad. Sci. USA* 68:1678–83
76. Blow, D. M., Birktoft, J. J., Hartley, B. S. 1969. *Nature* 221:337–40
77. Kraut, J. 1971. *The Enzymes* 3:547–60
78. Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E., Powers,

J. C. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:117–23
79. Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., Wilcox, P. E. 1972. *Biochemistry* 11:2439–49
80. Fritz, H., Tschesche, H., Greene, L. J., Truscheit, E., eds. 1974. *Proteinase Inhibitors: Proceedings of the 2nd International Research Conference.* New York: Springer-Verlag. 751 pp.
81. Reich, E., Rifkin, D. B., Shaw, E., eds. 1975. *Proteases and Biological Control: Cold Spring Harbor Conferences on Cell Proliferation,* Vol. 2. Cold Spring Harbor, NY: Cold Spring Harbor Lab. 1021 pp.
82. Segal, D. M., Cohen, G. H., Dawes, D. R., Powers, J. C., Wilcox, P. E. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:85–90
83. Segal, D. M., Powers, J. C., Cohen, G. H., Dawes, D. R., Wilcox, P. E. 1971. *Biochemistry* 10:3728–38
84. Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., Wilcox, P. E. 1972. *Biochemistry* 11:2439–49
85. Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J., Kraut, J. 1976. *J. Biol. Chem.* 251:1097–1103
86. Schechter, I., Berger, A. 1967. *Biochem. Biophys. Res. Commun.* 27:157–62
87. Ingles, D. W., Knowles, J. R. 1968. *Biochem. J.* 108:561–69
88. Kobayashi, R., Ishii, S. 1974. *J. Biochem. Tokyo* 75:825–35
89. Gertler, A., Hofmann, T. 1970. *Can. J. Biochem.* 48:384–86
90. Thompson, R. C., Blout, E. R. 1970. *Proc. Natl. Acad. Sci. USA* 67:1734–40
91. Atlas, D., Levit, S., Schechter, I., Berger, A. 1970. *FEBS Lett.* 11:281–83
92. Atlas, D., Berger, A. 1972. *Biochemistry* 11:4719–23
93. Atlas, D. 1975. *J. Mol. Biol.* 93:39–53
94. Shotton, D. M., White, N. J., Watson, H. C. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:91–105
95. Yoshida, N., Yamamoto, T., Izumiya, N. 1968. *Arch. Biochem. Biophys.* 123:165–71
96. Morihara, K., Oka, T., Tsuzuki, H. 1969. *Biochem. Biophys. Res. Commun.* 35:210–14
97. Morihara, K., Tsuzuki, H. 1969. *Arch. Biochem. Biophys.* 129:620–34
98. Morihara, K., Oka, T., Tsuzuki, H. 1970. *Arch. Biochem. Biophys.* 138:515–25
99. Fersht, A. R., Blow, D. M., Fastrez, J. 1973. *Biochemistry* 12:2035–41
100. Morihara, K., Tsuzuki, H., Oka, T.

1971. *Biochem. Biophys. Res. Commun.* 42:1000-6

101. Tulinsky, A., Wright, L. H. 1973. *J. Mol. Biol.* 81:47-56

102. Matthews, B. W., Cohen, G. H., Silverton, E. W., Braxton, H., Davies, D. R. 1968. *J. Mol. Biol.* 36:179-83

103. Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., Kraut, J. 1977. *J. Biol. Chem.* In press

104. Caplow, M., Harper, C. 1972. *J. Am. Chem. Soc.* 94:6508-12

105. Henderson, R., Wright, C. S., Hess, G. P., Blow, D. M. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:63-69

106. Markley, J. L., Porubcan, M. A. 1976. *J. Mol. Biol.* 102:487-509

107. Ito, A., Tokawa, K., Shimizu, B. 1972. *Biochem. Biophys. Res. Commun.* 49: 343-49

108. Thompson, R. C. 1973. *Biochemistry* 12:47-51

109. Hunkapiller, M. W., Smallcombe, S. H., Richards, J. H. 1975. *Org. Magn. Reson.* 7:262-65

110. Glazer, A. N. 1968. *Proc. Natl. Acad. Sci. USA* 59:996-1002

111. Glazer, A. N. 1968. *J. Biol. Chem.* 243:3693-3701

112. Stroud, R. M., Kay, L. M., Dickerson, R. E. 1971. *Cold Spring Harbor Symp. Quant. Biol.* 36:125-40

113. Hunkapiller, M. W., Smallcombe, S. H., Whitaker, D. R., Richards, J. H. 1973. *Biochemistry* 12:4732-43

114. Robillard, G., Shulman, R. G. 1972. *J. Mol. Biol.* 71:507-11

115. Robillard, G., Shulman, R. G. 1974. *J. Mol. Biol.* 86:519-40

116. Robillard, G., Shulman, R. G. 1974. *J. Mol. Biol.* 86:541-58

117. Rogers, G. A., Bruice, T. C. 1974. *J. Am. Chem. Soc.* 96:2473-81

118. Wallerberg, G., Boger, J., Haake, P. 1971. *J. Am. Chem. Soc.* 93:4938-39

119. Umeyama, H., Imamura, A., Nagata, C., Hanano, M. 1973. *J. Theor. Biol.* 41:485-502

120. Amidon, G. L. 1974. *J. Theor. Biol.* 46:101-9

121. Scheiner, S., Kleier, D. A., Lipscomb, W. N. 1975. *Proc. Natl. Acad. Sci. USA* 72:2606-10

122. Demoulin, D., Allen, L. C. 1977. *J. Mol. Biol.* In press

123. Polgár, L., Bender, M. L. 1969. *Proc. Natl. Acad. Sci. USA* 64:1335-42

124. Polgár, L. 1972. *Acta Biochim. Biophys. Acad. Sci. Hung.* 7:29-34

125. Gertler, A., Walsh, K. A., Neurath, H. 1974. *Biochemistry* 13:1302-10

126. Nakagawa, Y., Bender, M. L. 1970. *Biochemistry* 9:259-67

127. Henderson, R. 1971. *Biochem. J.* 124:13-18

128. Wright, C. S., Hess, G. P., Blow, D. M. 1972. *J. Mol. Biol.* 63:295-303

129. Birktoft, J. J., Kraut, J., Freer, S. T. 1976. *Biochemistry* 15:4481-85

130. Koeppe, R. E., Stroud, R. M. 1976. *Biochemistry* 15:3450-58

131. Brinigar, W. S., Chao, T. L. 1975. *Biochem. Biophys. Res. Commun.* 63: 78-83

132. Satterthwait, A. C., Jencks, W. P. 1974. *J. Am. Chem. Soc.* 96:7018-31

133. Wang, J. H. 1970. *Proc. Natl. Acad. Sci. USA* 66:874-81

134. Inagami, T., York, S. S. 1968. *Biochemistry* 7:4045-52

135. Nakata, H., Ishii, S. 1972. *J. Biochem. Tokyo* 72:281-90