cause the preferred monoanion has the 6-hydroxyl proton of the quinoline derivative ionized and would not lead to the desired product.

Although pseudobactin A converted spontaneously into pseudobactin in aqueous solution, both compounds were natural products of *Pseudomonas* B10. Pseudobactin A was produced only in iron-limiting, chemically defined minimal media, whereas pseudobactin production occurred even in nutritionally richer iron-deficient media. Pseudobactin A, but not ferric pseudobactin A, exhibited in vitro antibiosis against the bacterium *Erwinia carotovora*, which causes potato soft rot and seedpiece decay (data not shown). Analogous results have been reported for pseudobactin and ferric pseudobactin (Kloepper et al., 1980). The ability of pseudobactin A to promote plant growth or control phytopathogenic microorganisms in soils cannot be distinguished from that of pseudobactin, because pseudobactin A converts into the latter during the time course of the experiments.

Both ferric pseudobactin and ferric pseudobactin A are equally efficient at transporting iron into *Pseudomonas* B10 (J. Leong, unpublished experiments). The reason why *Pseudomonas* B10 should produce two siderophores is not evident nor are the biological implications of the chemical conversion immediately obvious.

The absolute configuration about the iron in ferric pseudobactin A could not be determined from its relatively featureless visible CD spectrum. The configuration of the additional chiral carbon in pseudobactin A is also not known. The determination of the X-ray structure of ferric pseudobactin A, which is now in progress (D. van der Helm, unpublished experiments), could answer these questions.

Acknowledgments

We thank Drs. H. A. Itano and R. F. Doolittle for the use of their equipment and Dr. J. M. Wright and the staff of the Southern California Regional NMR Facility for technical assistance.

References

Avdeef, A., Sofen, S. R., Bregante, T. L., & Raymond, K. N. (1978) *J. Am. Chem. Soc. 100*, 5362-5370.

Kloepper, J. W., Leong, J., Teintze, M., & Schroth, M. N. (1980) *Nature (London) 286*, 885-886.

Ong, S. A., Peterson, T., & Neilands, J. B. (1979) *J. Biol. Chem. 254*, 1860-1865.

Pouchert, C. J., & Campbell, J. R. (1974) *The Aldrich Library of NMR Spectra*, Aldrich Chemical Co., Inc., Milwaukee, WI.

Sadtler Research Laboratories (1980a) *Sadtler Standard Carbon-13 NMR Spectra*, Philadelphia, PA.

Sadtler Research Laboratories (1980b) *Sadtler Standard Proton NMR Collection*, Philadelphia, PA.

Teintze, M., Hossain, M. B., Barnes, C. L., Leong, J., & van der Helm, D. (1981) *Biochemistry* (preceding paper in this issue).

Wüthrich, K. (1976) *NMR in Biological Research: Peptides and Proteins*, pp 42-55, North-Holland/American Elsevier, New York.

Wyler, H., & Dreiding, A. S. (1962) *Helv. Chim. Acta 45*, 638-640.

Direct Determination of the Protonation States of Aspartic Acid-102 and Histidine-57 in the Tetrahedral Intermediate of the Serine Proteases: Neutron Structure of Trypsin†

Anthony A. Kossiakoff* and Steven A. Spencer‡

ABSTRACT: A neutron structure analysis at 2.2-Å resolution has been performed on bovine trypsin covalently inhibited by a transition-state analogue, the monoisopropylphosphoryl (MIP) group. The unique ability of neutron diffraction to locate hydrogen atoms experimentally has allowed the determination of the protonation states of the catalytic site residues (Asp-102 and His-57). Since the bound MIP group mimics the tetrahedral intermediate structure, these correspond to the protonation states at the most crucial step of the hydrolysis. This has resolved a much debated mechanistic issue by showing conclusively that the catalytic base in the transition state of the reaction is His-57, not Asp-102. This finding has important implications for the understanding of the hydrolysis mechanism of the serine proteases. A detailed examination of the stereochemical interaction among the catalytic groups was also conducted to identify their individual roles in the mechanism. Besides functioning as the catalytic group, it was found that His-57 could effectively "steer" the attacking water toward the acyl group during deacylation. Other aspects of protein structure which are observable only by neutron diffraction analysis are also discussed. These include orientation of well-ordered amide side chains, which is made possible by the large scattering difference between nitrogen and oxygen atoms, location and orientation of water molecules, and hydrogen exchange properties of the protein.

A major shortcoming of X-ray determinations of protein structures is that the locations of the hydrogen atoms cannot be observed directly and therefore must be inferred from the stereochemistry of the heavier atoms. This method of extrapolating hydrogen positions makes it difficult to define accurately the chemical interactions taking place among groups within the protein, especially since any unusual or unexpected stereochemistry for the hydrogens will be overlooked. Since a large number of the total atoms in a protein are hydrogen, and since much of the chemistry of proteins involves hydrogen

† From the Biology Department, Brookhaven National Laboratory, Upton, New York 11973. *Received February 12, 1981; revised manuscript received June 30, 1981.* Research carried out at Brookhaven National Laboratory was under the auspices of the U.S. Department of Energy.

‡ Present address: Genentech Inc., South San Francisco, CA 94080.

0006-2960/81/0420-6462$01.25/0

atoms, it is apparent that the ability to determine the hydrogen positions experimentally is important. In this respect, neutron diffraction, by virtue of its ability to locate hydrogen and deuterium atoms directly, is potentially capable of probing certain aspects of the detailed structures of proteins beyond the capability of X-ray methods alone.

The use of neutron diffraction to study biological samples has mainly been limited to solution scattering studies or to systems existing in partially ordered assemblies such as membranes, collagen, and muscle. The advantages of neutron scattering over its X-ray counterpart for structural studies at low resolution have been well documented (Schoenborn, 1975), but until rather recently, high-resolution single-crystal studies of proteins have been confined to X-ray techniques. This situation has been due mainly to certain experimental difficulties which are discussed later in the text. However, because of advances in several important areas, neutron diffraction is now capable of providing resolution comparable to that of X-ray diffraction.

Schoenborn and his co-workers (Schoenborn, 1971; Norvell & Schoenborn, 1975) have shown with their neutron studies of myoglobin that a majority of well-ordered hydrogen atom positions can be located in Fourier maps at a resolution near 2 Å. These results led us to believe that neutron diffraction was the most direct technique to solve a particularly important and elusive question involving the mechanism of a family of enzymes called the serine proteases.

There currently exists a debate in the literature concerning the chemical character of the catalytically active residues of these enzymes and the possibility that environmental influences have a perturbing effect on their normal behavior. Details of the experimental results which make up the basis for this controversy will be discussed in a later section, but briefly, the disagreement centers on the assignment of protonation states of the active groups during the transition state of the reaction (Blow et al., 1969; Robillard & Shulman, 1972; Hunkapillar et al., 1973; Stroud et al., 1975; Bachovchin & Roberts, 1978; Markley, 1978).

X-ray crystallographic studies have shown that trypsin (Stroud et al., 1971, 1974; Chambers & Stroud, 1977a; Bode & Schwager, 1975) and its closely related pancreatic family members, elastase (Watson et al., 1970) and chymotrypsin (Birktoft et al., 1970; Cohen et al., 1969; Tulinsky et al., 1973), as well as quite distant relatives in terms of evolutionary origin (Alden et al., 1970; Delbaere et al., 1975), all have active site regions that are nearly identical and that this region can be subdivided into a catalytic site and a substrate binding pocket. The catalytic site is characterized by three invariant residues, a histidine (57), an aspartic acid (102), and a serine (195). Their direct participation in the hydrolysis reaction has been proven unequivocally, since chemical modification of any of these three residues will greatly diminish or completely abolish catalysis (Jansen et al., 1949; Fahrney & Gold, 1963; Ong et al., 1964; Shaw et al., 1965; Henderson, 1971; Martinek et al., 1971). The hydrolysis mechanism which is consistent with most known experimental data involves nucleophilic addition of the Ser-195 hydroxyl group to the carbonyl carbon of the substrate (Bender & Kezdy, 1964; Inward & Jencks, 1965). The generation of the effective nucleophile is accomplished through a transfer of a proton from the serine hydroxyl to either the imidazole of His-57 or, through the intermediacy of His-57, the carboxylate of Asp-102 (Figure 1). The primary, yet unresolved, mechanistic question concerns the identification of which residue, His-57 or Asp-102, acts as the chemical base in the hydrolysis reaction.

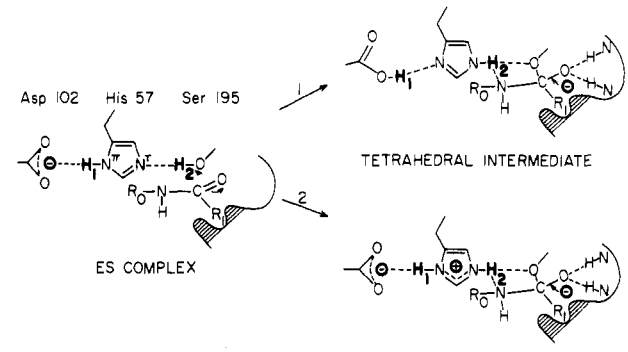

FIGURE 1: Two different locations proposed for proton H(1) in the tetrahedral intermediate (TI) structure. In the formation of TI (1), H(1) shifts from N^{τ} to Asp-102 ($O_{\delta 2}$) as proton H2 transfers from the serine hydroxyl group to N^{τ} of the imidazole. His-57 remains neutral throughout the reaction. In TI (2), proton H(1) remains on the imidazole as H(2) transfers. The imidazole becomes charged and forms a zwitterion with the carboxylate of Asp-102. The intermediate pictured here corresponds to that formed during acylation and differs from the deacylation intermediate (which MIP mimics) only by the replacement of the leaving group (R_O–NH_2) with a hydroxyl group.

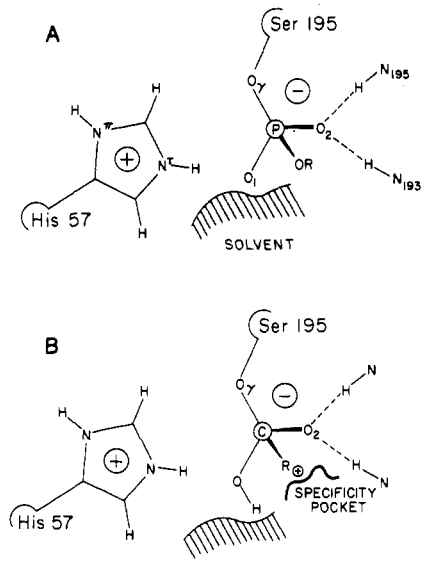

FIGURE 2: Structural comparison of (A) the MIP group with (B) the model of the tetrahedral intermediate (TI). Both substrates possess a net charge of −1 and have an oxygen directed at the NH peptide groups of 193 and 195. In the TI, ligand R is the side-chain moiety and binds in the specificity pocket; in MIP, OR is an isopropyl group.

The neutron diffraction study reported here was initiated in part to settle this controversy by determining conclusively the correct protonation assignments. Toward this goal, a form of bovine trypsin inhibited by a monoisopropylphosphoryl group (MIP)[1] (Figure 2A) was chosen as the most suitable candidate because it had been found that the bound MIP group closely mimicked the expected structure and electrostatic properties of a real substrate–enzyme intermediate (Stroud et al., 1974; Kossiakoff & Spencer, 1980). Analysis of this complex consequently allowed the direct determination of the protonation states of the catalytically important residues during the most crucial stage of the hydrolysis reaction.

In this paper, we place particular emphasis on that portion of the structure concerned with the catalytic function and its implications for the reaction mechanism. Since there are rather substantial differences between X-ray and neutron methods in the application of refinement techniques and

[1] Abbreviations used: MIP, monoisopropylphosphoryl; NMR, nuclear magnetic resonance; TI, tetrahedral intermediate; PTI, pancreatic trypsin inhibitor; STI, soybean trypsin inhibitor.

evaluation of phasing models, the neutron procedures are discussed in some detail. A preliminary report of the structural results has been published (Kossiakoff & Spencer, 1980).

Experimental Procedures

Data Collection. The bovine MIP-trypsin crystal used for data collection was grown from an 8% $MgSO_4$ solution at pH 7 by vapor diffusion (Stroud et al., 1974). The crystal was orthorhombic, space group $P2_12_12_1$, with cell parameters a = 54.85, b = 58.61, and c = 67.47 Å. Its physical dimensions were 0.9 × 0.9 × 2.0 mm, giving an approximate volume of 1.6 mm^3. The crystal was soaked for approximately 1 month in a D_2O–$MgSO_4$ solution, pH 6.2 (uncorrected pH meter reading), to replace exchangeable hydrogen atoms with deuterium. The crystal was sealed in a quartz capillary tube containing a small amount of the mother liquor and then mounted on a four-circle diffractometer with the crystallographic a^* axis parallel to the instrument Φ axis.

The neutron beam from the reactor was monochromatized and focused by using a bent pyrolytic graphite monochromator, which produced a beam flux of 5×10^7 neutrons cm^{-2} s^{-1} at the sample. A wavelength of 1.56 Å was chosen as the best compromise between maximum efficiency and a small $\lambda/2$ contribution to the direct beam. A $\lambda/2$ filter was not used since its attenuation at the desired wavelength was considerable. Instead, the $\lambda/2$ contribution to the direct beam was measured by diffraction from a KBr crystal. The $\lambda/2$ contamination was found to comprise 6.4% of the diffracted beam and was corrected for during the data reduction.

The data were collected in layers perpendicular to the a^* axis by using normal beam rotation geometry. The crystal was rotated around the Φ axis in steps of 0.05° over a range of $(90 + 2\theta_{max})°$ for each lattice level. In all, 17 levels were collected, giving approximately 8700 unique reflections.

The detector system consisted of three linear position-sensitive detectors connected in series (Alberi et al., 1975; Cain et al., 1975) and arranged so that data from 0° to 48° in 2θ could be collected without repositioning the detectors. The long axes of the detectors were coplanar, and angled such that the midpoint for each detector was tangent to the cone of diffraction for the level being collected. This arrangement considerably reduced the layer–line curvature which would occur if upper level cones of data were projected onto a single linear detector of the same total length. The angular range of the detector system was subdivided into 720 data channels, with each channel approximately corresponding to the positional resolution of the detectors.

Data Reduction. Neutron diffraction data from protein crystals tend to have a poor signal-to-noise ratio because of the low beam flux available from a nuclear reactor and the high backgrounds which result from incoherent, inelastic scattering of neutrons by hydrogen atoms in the sample. In the present case, these problems were further amplified by the crystal size, which was over an order of magnitude smaller than any protein crystal previously used to collect high-resolution neutron data. As a result, a data reduction method had to be developed to extract accurate intensity information in the presence of high background levels.

The method which was used takes advantage of the fact that the linear detector subdivides each reflection into steps along the detector axis. This results in a two-dimensional array of intensity points (Φ steps vs. detector channels) for each reflection. These arrays make it possible to improve the intensity determination by allowing the peak boundary to be more accurately defined, thus eliminating considerable extraneous background and substantially improving the statistical quality of the data. Details of this technique have been presented elsewhere (Spencer & Kossiakoff, 1980).

Refinement. Refinement of atomic coordinates was carried out by using the constrained difference Fourier technique (Freer et al., 1975; Chambers & Stroud, 1977a). After each cycle of coordinate shifts, the structure was rebuilt to ideal bond lengths and bond angles and to a lower global energy by using a program developed by Hermans (Hermans & McQueen, 1974; Ferro et al., 1980). The calculated function which was minimized for each atom was

$$F_1 = K_X\sum\tfrac{1}{2}(X - X_0)^2 + K_l\sum\tfrac{1}{2}(l - l_0)^2 + K_\theta\sum\tfrac{1}{2}(\theta - \theta_0)^2 + K_\rho\sum\tfrac{1}{2}(\rho - \rho_0)^2 + K_e(-A/r^6 + B/r^{12} + C/r) + K_d\sum\tfrac{1}{2}E_b[1 - \cos n(\phi - \phi_0)]$$

where the contributions are derived from (a) the approach to initial coordinates, (b) bond lengths, (c) bond angles, (d) fixed dihedral angles, (e) nonbonded energy, and (f) torsional potential.

The first term in the energy function, deviation from the initial coordinates, was weighted in the minimization according to the atom's individual temperature factor. In practice, this meant that well-defined atoms (low temperature factors) were perturbed a smaller distance from their starting positions than those with higher temperature factors. This was a necessary procedure because the more poorly defined atoms have larger and less accurate atomic shifts, and if given equal weight in the rebuilding process, they have a negative influence on the refinement.

Insofar as most close contact regions in a protein involve hydrogen–hydrogen interactions, the last two terms in the energy refinement, the nonbonded and torsional energies, are of special importance in neutron structure analysis. In situations where the methyl rotors appeared to adopt high-energy conformations, the weight of the torsional energy parameter was set to rotate the group only 10–15° toward its energetically preferred orientation. A rotation of this limited magnitude allowed the rotor to readjust back toward its initial position if subsequent difference maps continued to show a strong preference for the higher energy conformer.

The nonbonded energy function used in the refinement consisted of three terms. The first two terms are equivalent to the Lennard-Jones "6–12" potential, where coefficients A and B represent the attractive and repulsive components of the interaction. The third term expresses the electrostatic interaction between the atoms calculated as the monopole potential between the partial charges (Poland & Scheraga, 1967). When a group of atoms was found to be involved in hydrogen bonding, the coefficients A and B of the donor and acceptor atoms were adjusted to redefine the resulting potential energy surface. Rather than the inclusion of a separate set of A and B values for each type of hydrogen bonding couple, an average value for A and B was selected and applied in all cases.

During the initial cycles of refinement, the nonbonded and torsional energy components were downweighted compared to the terms involved in the reidealization of the geometry. This ensured that no large shifts would be performed on the starting model during the first stages of refinement. The weights of these two terms were gradually increased in subsequent refinement cycles, thereby eliminating persistent high-energy interactions.

Results and Dicussion

Starting Phasing Model and Criteria of Confidence. The initial phasing model was calculated by applying the appropriate neutron scattering lengths (Table I) to the refined X-ray

Table 1: Scattering Lengths for Relevant Atom Types[a]

element	scattering length ($\times 10^{-13}$ cm)
H	−3.74
D	6.67
C	6.65
N	9.40
O	5.80
S	2.80
Ca	4.70

[a] From Bacon (1975).

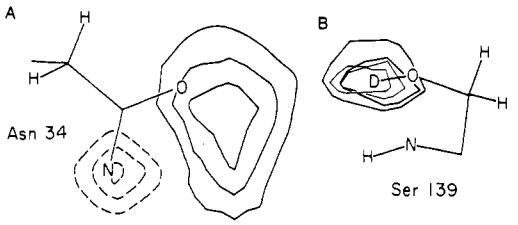

FIGURE 3: (A) Difference map of Asn-34 side chain. The nitrogen and oxygen positions shown are those from the X-ray model. The difference density indicates that the orientation of the nitrogen and oxygen atoms is incorrect. (B) Difference map for Ser-139. On well-ordered hydroxyl side chains, orientation of deuterium atom can sometimes be assigned.

MIP-trypsin coordinates of Chambers & Stroud (1977b). The superb quality of these coordinates is demonstrated by the reported low R factor ($R = 0.15$) and the high resolution of the data used in the refinement (1.5 Å). The initial structure factor calculation for the neutron data, excluding all hydrogen atoms and water molecules, gave an R value of 0.304.

This statistic indicated a strong correlation between the neutron amplitudes and the starting coordinate set, considering that nearly one-half of the total atoms (hydrogen) in the structure were omitted from the structure factor calculation. Perhaps more encouraging, however, was the amount of interpretable information contained in the resulting difference map. Since this map contained no bias toward any features except those of the original coordinates, it was at this stage of the analysis that the quality of the data and the phasing model could best be evaluated.

It is a well-recognized phenomenon that a Fourier synthesis tends to be dominated by the phases rather than by the amplitudes (Ramachandran & Srinivasan, 1970). Therefore, it would be expected that if the quality of the data was poor the resulting Fourier map would approximately mirror, at reduced weight, the parent structure with few additional interpretable features. Since a neutron map contains a set of inherent structural characteristics, particularly those resulting from the existence of observable hydrogen and deuterium atoms, the presence or absence of these features in the map can be used to judge the quality of the data. They also provide a more reliable indication of the quality of the structure than that provided by refinement statistics alone.

One such feature of neutron maps is the ability to orient properly the side-chain amide groups of asparagine and glutamine. Figure 3A shows a graphic example of this capability. In a protein X-ray analysis, the difference in scattering intensity between O and NH_2 is much to small to be detected. In contrast, the neutron scattering magnitudes of oxygen and nitrogen (5.8 vs. 9.3) are quite dissimilar, and there is additional scattering at the nitrogen site from the two bound deuterium atoms. The resulting scattering differential is over 350%, quite large enough to be detected for well-ordered side chains. For example, in the trypsin X-ray model (Chambers & Stroud, 1977b), the positions of the oxygen and nitrogen for Asn-34 are as represented in Figure 3A. It is quite apparent from the neutron difference map that this orientation had been misassigned. For three other amide side chains in the structure, the difference density indicated a similar need to reassign the oxygen and nitrogen positions.

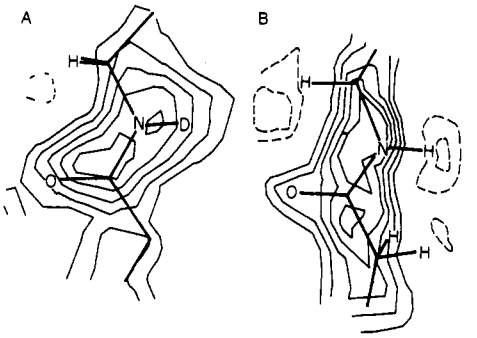

FIGURE 4: Determination of hydrogen–deuterium exchange on peptide groups. (A) Fully exchanged peptide; (B) unexchanged peptide group.

Another important test to establish confidence in the initial structure was the location of water molecules. The water structure was not included in the initial phasing model, so that peaks in the neutron difference map which corresponded to water molecule positions in the X-ray structure owed their existence solely to the neutron data. The 40 best determined water molecules from the X-ray model were readily located in the difference maps. These water molecules were, by and large, those involved in multiple coordination to the protein and therefore the most highly ordered. In addition, there were very few unexplained peaks in the maps which were of comparable height to these molecules.

The correlation between the correctness of the structure and the ability to locate exchanged deuterium atoms on the peptide backbone and the many exchangeable side chains is tenuous at best. Although there were many cases where a full deuterium exchange on the amide nitrogen of the peptide had obviously taken place (Figure 4A), and also cases, in internal regions of the molecule, where the position was fully occupied by a proton (Figure 4B), a common situation was one involving some degree of partial exchange. The results of such a circumstance can be determined from the plot in Figure 5. For example, if the site were 70% exchanged, the scattering contribution from the deuterium would be 4.7 (6.7×0.7), and the component from the hydrogen would be −1.2. The resultant scattering magnitude would be 3.5, a reduction to nearly one-half the full value for deuterium. When this is coupled with the fact that peaks representing atoms not in the phasing model are statistically expected in the map at one-half their true weight (Luzzati, 1953; Henderson & Moffat, 1971), it is quite apparent why this class of atom is a poor indicator of the correctness of the model.

In trypsin, there are about 1600 unexchangeable hydrogen atoms; thus, the ability to locate these atoms in stereochemically reasonable positions is another potential test of the model. However, a hindrance to the straightforward location of hydrogens is that hydrogen atoms are located only 1 Å from their parent atom and possess a scattering magnitude of opposite sign. These overlapping distributions of opposite sign have the effect that, in a Fourier map, the apparent peak position for a hydrogen atom is not located at the atomic center but rather is displaced from it along a vector between the hydrogen and the parent atom (for example, see Figure 6A). This displacement is away from the parent atom, and its magnitude is related to the density of the parent atom and the resolution of the map. Although the use of difference maps in place of

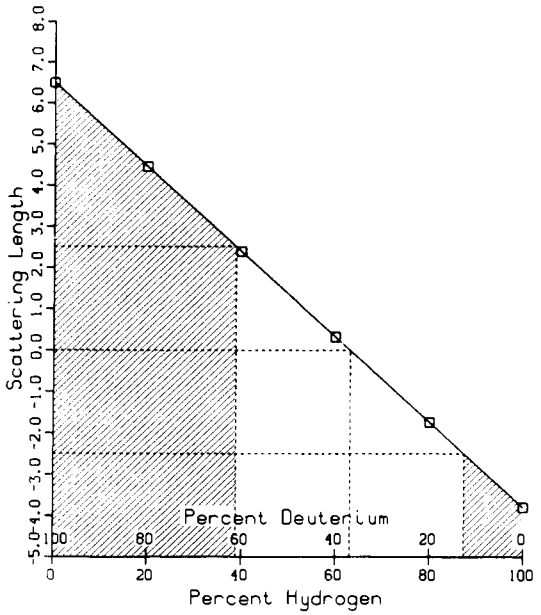

FIGURE 5: Scattering length as a function of the hydrogen–deuterium exchange ratio. An effective scattering length of ±2.5 was chosen as the minimum value that could be used to assign a H or D atom; shaded areas of the plot indicate the H/D ratios where assignments can be made. To assign a D, the atom occupancy must be at least 60% deuterium; in the case of H assignments, the atom has to be nearly 90% hydrogen.

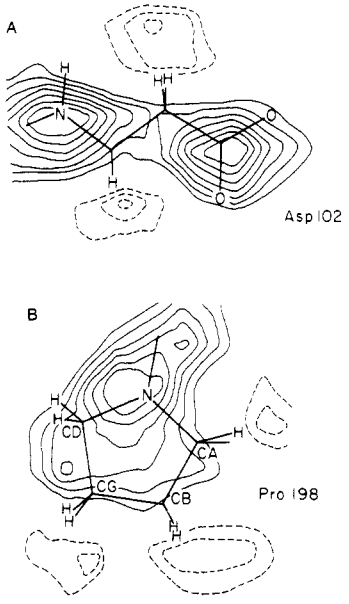

FIGURE 6: Effect of parent atom temperature factors on bound hydrogen atoms. (A) Example of correct parent atom temperature factors. Translation of the negative peaks of the hydrogen atoms away from their actual positions is expected at the resolution of these maps. At 2.2 Å, a portion of the negative density peak of the H overlaps with the positive peak of the parent atom, effectively canceling density between the atoms and giving the illusion that the actual peak has been translated. (B) Example of incorrect temperature factors. Assigned temperature factor of proline CD is too small, causing its phasing contribution to "mask-out" the hydrogen atoms. Temperature factors for the other parent atoms (CB, CG) are somewhat larger than their true values.

Fourier maps should effectively decrease the influence that adjacent atoms have on one another, interpretive errors can be made when the atoms being investigated have incorrect temperature factors (Figure 6B). This of course assumes that the parent atom has been placed near its correct position initially. In situations involving incorrect placement, interpretation becomes considerably more complicated.

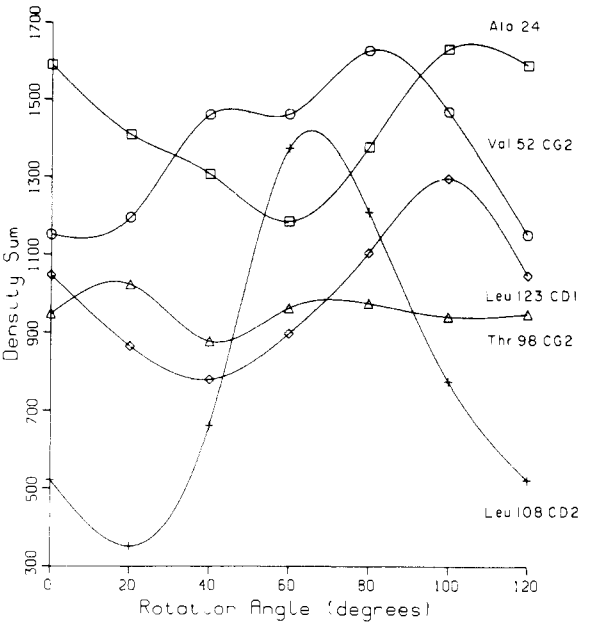

FIGURE 7: Methyl rotor plots. Plots for individual rotors were obtained from difference maps with the parent atom removed. Methyl hydrogens were rotated around their rotor axes in 20° increments. At each interval, densities found at the hydrogen positions were summed. A range of 120° was covered.

Addition of Hydrogen and Deuterium Atoms to the Model. Each residue in the structure was inspected on an individual basis by using Fourier and Fourier-difference maps. Where the residue was deemed "well-ordered", the hydrogen atoms that were in stereochemically constrained positions were added to the model. If a region of the residue showed disorder (or misplacement), the hydrogens pertaining to that portion were omitted. Hydrogen temperature factors were assigned as $B + B^{1/2}$, where B is the temperature factor of the parent atom.

In a similar manner, deuterium atoms were searched for in difference maps by looking at individual residues which contained potential sites of exchange. The criterion for assigning a deuterium position was that there be significant positive density (one-half of the expected density) in the difference map at the predicted atomic center.

The most difficult hydrogens to place were those on methyl groups. It has been well established that in a protein these hydrogens are not fixed in unique orientations but rather spin rapidly about the methyl rotor axis (Stockton et al., 1974). At any given time, however, there exists a greater probability that they reside in their lower energy conformations. If the percentage of time that the methyl rotor spends in any one conformation is large enough, this information should be reflected in the Fourier map.

Because of the possible inaccuracies involved in making realistic assignments of methyl rotor positions by using data extending to only 2.2-Å resolution, the prospects for determining unique rotor conformations were assessed by inspecting a difference map calculated from a synthesized set of data at an equivalent resolution. It was found that for well-ordered methyl groups, the rotor orientations could be assigned to within 20° of their correct positions. These findings corresponded closely to what was observed in the trypsin difference maps, where a well-ordered methyl group was characterized by an area of negative density into which a trio of hydrogen atoms could be placed.

Representative examples of the rotor conformations are pictured in Figure 7. These curves describe the summed intensities found at the atom sites of the three hydrogens

Table II: Development and Refinement of Trypsin Model

difference map no.	no. of atoms in model	no. of D_2O	R^a
	Addition of Atoms to Model		
1	1654 (no hydrogen)[b]	0	0.304
2	2394	41	0.272
3	2569	41	0.264
4	2716	55	0.258
5	3034[c]	55	0.255
	Refinement of Model		
9 cycles	3034	80	0.187
rms deviation from ideal bond lengths and angles: bond lengths, 0.013 Å; bond angles, 2.3°; torsional angles, 2.7°			

[a] $R = \Sigma|F_o - F_c|/\Sigma F_o$. [b] Chambers-Stroud model for MIP-trypsin. [c] 92% of total possible atoms in structure.

plotted as a function of their rotor angle. The predicted lowest energy (staggered) conformer is designated to be at (60°, 180°, 300°). From the figure, the orientation of the CD(2) methyl group of Leu-108 can clearly be assigned to the staggered conformation. As a rule, however, most conformer plots were not as definitive; the four other rotor curves in the figure are more representative of the shape for the majority of the methyl rotors.

Although many of the methyl rotors could be assigned to positions approximating the staggered conformation, there were some instances, such as the curves for Ala-24 and Leu-123 CD(1), where the eclipsed conformer was clearly favored. In subsequent refinements of these rotor groups, only a small reorientation was observed. However, it must be emphasized that the refinement of methyl groups is inherently difficult; small coordinate or temperature factor errors on the parent carbon atom can greatly influence the observed densities at hydrogen sites. Thus, the accuracy of the eclipsed conformers will have to be reevaluated when higher resolution data become available before any significance can be assigned to them.

In all, the iterative process of locating the present number of hydrogen and deuterium atoms took five difference map cycles (Table II). As would be expected, the number of newly observed hydrogen (deuterium) atoms decreased with each successive cycle, until it was decided that it would be unproductive to continue the procedure. At this stage in the analysis, before commencing refinement, the R factor was 25.5%.

Difficulties Presented in Refinement. In any protein structure refinement, there exists a number of inherent difficulties: the large number of atoms in the structure, the limited resolution, errors in the model, and so forth. However, neutron structures present several unique problems not encountered in X-ray analysis. These special problems predominantly arise from the close proximity of hydrogen atoms to their parent atoms, coupled with the effects of the negative scattering length of hydrogen atoms. Clearly, potential problems exist when the difference density generated from positional errors of one atom overlaps an adjacent atom site. This situation is further complicated by the fact that, because of this negative scattering length, an error in a hydrogen atom position is minimized by moving the atom down the difference density gradient, that is, opposite to the direction required for correcting parent atom positions.

The success of the curvature–gradient refinement technique is predicated on the assumption that since the measurement of the gradients is restricted to a small range around the position of the atom, the influence of the neighboring atoms is largely minimized (Leung et al., 1957). However, at the resolution to which most protein data are collected, there exists a real danger that gradient values can be adversely affected by the influences of proximate atoms. In viewing this problem, we were concerned that at the resolution of our maps, productive refinement would be difficult, or perhaps even impossible, because of the short hydrogen–parent atom bond distances.

To evaluate the refinement capabilities of the technique, we devised a test in which all the coordinates in the model were perturbed by a varying but known amount from their ideal positions, and a difference synthesis was computed by using a set of 2.2 Å data equivalent to the collected trypsin data. It was determined that, in general, dependable shifts could be obtained when the coordinate errors were less than 0.3 Å. (A dependable shift was defined as one which was in the right direction and would likely converge to the proper atomic position in three cycles or less.) Also, it was found that if a parent atom (an atom with one or more hydrogens attached to it) was displaced by over 0.6 Å from its correct position, the effect of the neighboring hydrogens rendered the gradient at the parent atom center inaccurate. This problem was not investigated exhaustively. However, this test did give helpful general indications as to what kind of response could be expected from the technique in certain defined situations, and it clearly emphasized the necessity that a neutron analysis begin with a well-determined X-ray model.

As was expected, owing to the high quality of the trypsin starting structure, the calculated shifts from the difference maps showed that a substantial majority of the neutron structure was closely compatible with the X-ray model. The largest shifts in the well-ordered regions occurred on parent atoms with two or more attached hydrogens. The overall shift of all nonhydrogen atoms from the initial X-ray model was 0.21 Å.

The following subsections discuss examples of specific refinement problems with brief explanations of how they were handled.

(*a*) *Constraining Hydrogen Atom Temperature Factors.* When two or more hydrogen atoms were bound to the same parent atom, their individual temperature factors (B factors) were adjusted toward parity, since it is hard to rationalize a situation in a refined structure where two stereochemically constrained hydrogens, bound to the same parent atom, would have substantially differing temperature factors. The adjustment was performed by calculating the average B of the atoms involved and then readjusting the value of each B back to one-third the difference between the average and its original value. This scheme effectively reduces unwarranted discrepancies between structurally similar hydrogen atoms but does allow some influence by the independently arrived at temperature factor value. It was found that moderate adjustments toward equalization, extended over several cycles, constituted a more effective procedure in arriving at the stabilized set of B values than one which simply sets the B values to the average value in a single step.

(*b*) *Refinement of Partial Occupancy Deuterium Atoms.* Assigning a proper temperature factor to a deuterium atom proved to be quite difficult because, unlike an unexchanged hydrogen atom which can be treated as an extension of the parent atom, the true density for the deuterium is affected by both an occupancy and a temperature factor. We felt that there was no straightforward way to deal easily with this problem; accordingly, a reasonable guess as to the temperature factor was made, and it was allowed to converge to its desired value during refinement.

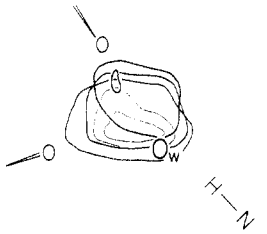

FIGURE 8: Orientation of D_2O molecules. Difference map density with the X-ray-determined position for the oxygen atom (O_w) subtracted. Resulting density is due to scattering of the two deuteriums alone.

At present, our refinement program adjusts only temperature factors. A more satisfactory approach would be to assign a temperature factor derived from that of the parent atom and refine the occupancy factor. The actual effect on the phasing model of the present mode of refinement is probably quite small, since the only atoms which suffer from this treatment are those of low occupancy. However, since quantification of the exchange ratios of these atoms is of considerable biological interest, their treatment in the refinement will be corrected when higher resolution data become available.

(*c*) *Refinement of Water Molecules.* Each deuterium atom of a water molecule (D_2O) has the potential to scatter at a magnitude comparable to that of the parent oxygen (see Table I). It is therefore possible, in principle, to establish the coordination geometry of the well-ordered waters from the density features in a Fourier map. Very helpful in this respect is the knowledge of the oxygen atom position from the X-ray analysis. The oxygen can be subtracted from the map by a difference synthesis, in effect isolating the scattering density for the deuterium atoms alone.

Before water molecule refinement was initiated, a significant percentage of the largest peaks in the "oxygen-removed" difference map could be attributed to these isolated deuterium atoms associated with D_2O molecules. The difference density was usually not characterized by two distinct lobes, but rather by a single broad asymmetric peak with the oxygen position located on the periphery, as shown in Figure 8. After studying several approaches to orienting the D_2O molecules by using both empirical and energetic criteria, it became apparent that it was not realistic to pursue detailed conclusions of this nature at the current resolution. However, these attempts did indicate that at an extended resolution, the orientation of D_2O molecules should be readily obtainable from the maps.

The method used for D_2O refinement in this treatment was to refine the molecule as a single atom having a scattering magnitude approximately 1.5 times that of oxygen. This essentially established the centroid position of the molecule. Although this procedure added little new information concerning water structure, it was necessary because it eliminated errors in the map which could be propagated from the non-negligible scattering effect of the deuterium atoms.

Refinement Results. The progress of the refinement and its current status after nine cycles are summarized in Table II. The structure is highly constrained toward ideal bond lengths and angles, as it was felt that the limited resolution of the data did not warrent the introduction of unusual geometric features into the model. The average deviation in bond lengths is 0.013 Å, bond angles 2.3°, and torsional angles 2.7°. The structure contains no high-energy interactions over 5.2 kcal (as calculated by the energy refinement program) and only 27 over 2 kcal.

Attempts To Resolve the Issue of the Protonation State of the Catalytic Residues by Other Techniques. The number of studies in the literature which have attempted to assign definitively the role of the Asp-His-Ser triad in the mechanism of the serine proteases gives a good indication of the wide interest held in the answer to this question, as well as the experimental difficulties which have been encountered in attempting its solution. A basic problem which arises is the fact that all techniques (excepting diffraction methods) which have the theoretical capacity to discriminate phenomena at the atomic level in proteins are indirect in the sense that their results must be interpreted on the basis of the observed character of a related model system. While this approach is often successful, in the case of elucidating the serine protease mechanism it has led to contradictory results. We therefore feel that it is both important and instructive to review these findings.

Nuclear magnetic resonance (NMR) has been used by a number of workers to study this reaction. In their pioneering work on the serine proteases, Robillard & Shulman (1972) used ^1H NMR to determine a titration curve for the His-57 imidazole. They interpreted their results as indicating a pK_a of 7.5 for His-57. Later ^1H NMR studies by these authors and by Markley and co-workers (Robillard & Shulman, 1974a,b; Markley & Porubcan, 1976; Porubcan et al., 1978; Markley & Ibanez, 1978) on a variety of serine proteases, protease–inhibitor complexes, and zymogens further supported the conclusion that His-57 had a pK_a around 7. Finally, an ^{15}N experiment by Bachovchin & Roberts (1978), wherein the His-57 imidazole of the bacterial serine protease, α-lytic protease, was enriched with ^{15}N, showed that the imidazole titrated with a pK_a of approximately 7.

Contradicting these results are the findings of Hunkapillar et al. (1973), who reported that they were unable to reproduce the original results of Robillard and Shulman. These authors then studied α-lytic protease selectively enriched with ^{13}C at the C(2) carbon of His-57. They found that the C(2) carbon exhibited significant chemical shifts at both pH 6.7 and 3.3, but the carbon–hydrogen spin–spin coupling constant characteristic of a protonated imidazole was observed only below pH 3.8. Based on this result, they proposed that environmental effects in the protein had so raised the apparent pK_a of Asp-102 that the titration at pH 6.7 reflected the loss of a proton from the Asp-102 carboxyl group in the presence of a neutral imidazole rather than the loss of a proton from an imidazolium cation in the presence of a carboxylate anion.

Further support for this hypothesis came from the solvent isotope experiments by Hunkapiller et al. (1976). A second-order rate dependence was observed in these experiments and was interpreted to indicate the simultaneous movement of two protons in the transition state, in agreement with their NMR interpretation. It was noted, however, that this result conflicted with the first-order rate dependence reported in several prior studies (Pollock et al., 1973; Elrod et al., 1975). These seemingly contradictory results were explained as being the result of differences in the degree of specificity of the substrates used in the different experiments (Hunkapiller et al., 1976).

Further evidence supporting the assignment of Asp-102 as the protonated species in the transition state came from the difference infrared absorption work of Koeppe & Stroud (1976). By chemically modifying all the carboxyl groups except Asp-102 and Asp-194, they were able to observe that one carboxyl group titrated at pH 6.8. This was assigned as Asp-102, based on a competitive binding study using Cu ion. In addition, the theoretical studies reported by Scheiner et al. (1975) and Scheiner & Lipscomb (1976) indicated that Asp-102 was the catalytic base in the reaction.

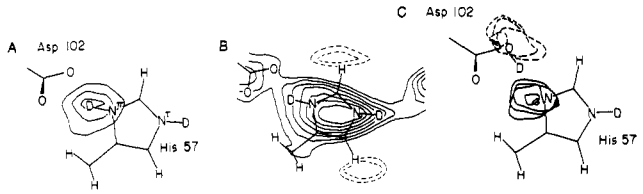

FIGURE 9: (A) A difference map $[(F_o - F_c) \exp(i\Phi_c)]$ calculated with only the deuterium [H(1)] between the His-57 and Asp-102 side chains left out of the phases. The difference peak is approximately 5.5σ above the background level and shows the deuterium to be bound to the imidazole nitrogen. (B) A Fourier map computed with terms $(2F_o - F_c) \exp(i\Phi_c)$. In this map, both of the catalytically important deuteriums [corresponding to H(1) and H(2) in Figure 1] were omitted from the phases. It is clear from this map that both are located on the imidazole. (C) A difference map in which the deuterium was placed by stereochemistry on atom $O_{\delta 2}$ of Asp-102 (if there were one, it would be on $O_{\delta 2}$). The difference density peak clearly indicates that the preferred location of the deuterium is on the imidazole of His-57.

These contradictory conclusions make it evident that the role of the Asp–His couple had not yet been determined conclusively. These disagreements also demonstrated the inherent uncertainties in making definitive spectral assignments in systems as complex as proteins.

Determination of the Protonation States of the Catalytic Residues. There was some concern initially that the mechanistically important proton between His-57 and Asp-102 [H(1) in Figure 1] would not be detectable in our Fourier maps due to the possibility of it being only partially exchanged with deuterium. Such a situation was possible because the small cavity in which this proton resides is well insulated from solvent by the imidazole ring and the carboxylate side chain, and these side chains are themselves restricted in their motion by a set of strong hydrogen bonds. It was, therefore, very reassuring to find a large positive peak between the N^{π} of His-57 and $O_{\delta 2}$ of Asp-102 in the first difference map. The size of the peak represented virtually a fully occupied deuterium atom. In order to eliminate any question of bias, no attempt was made to assign this deuterium to either the imidazole or the carboxylate side chain until after a majority of the other atoms was included in the phasing model.

Since a major objective of the study was to identify conclusively the group acting as the base, it was important that the interpretation of the proton positions in the catalytic site be unambiguous. Toward this end, three methods were devised to test the preferred location of proton H(1) (Figure 1). In method 1, the deuterium [proton H(1) was exchanged for deuterium during the soaking procedure] was omitted from the model and the full structure refined through one cycle of coordinate shifts and model reidealization. This refinement cycle was run to eliminate any bias in the phases that may have been introduced by a prior placement of this atom in the model. Figure 9A showed the result of the difference synthesis calculated from these structure factors and is interpreted as showing a distinct preference for the deuterium to reside on the imidazole.

In method 2, a deuterium atom was placed by stereochemistry on the nitrogen of the imidazole and its position refined for two cycles. Throughout the refinement, the deuterium atom exhibited stable refinement characteristics and a relatively low temperature factor, indicating that it was tightly bound to the parent N^{π} atom. The same process was repeated but with the deuterium atom placed in its alternative position on the $O_{\delta 2}$ of the Asp-102 side chain. In this case, the outcome was quite different; the refinement was unstable with respect to both positional and thermal parameters, and the deuterium atom resided along a distinct gradient, a clear indication of misplacement.

Table III: Individual Atomic Charge Densities[a] for the Model of the Tetrahedral Intermediate and the MIP Group

TI charges[b]		MIP charges[c]	
O_γ	−0.34	O_γ	−0.27
O_H	−0.23	O_1	−0.49
R	−0.04	OR	−0.27
O_2	−0.51	O_2	−0.49
C	+0.12	P	+0.52

[a] Refer to Figure 2 for geometry of the atoms. [b] Kleier et al. (1976). [c] Perahia et al. (1975).

Method 3 involved the qualitative examination of two independent difference maps. For the first map, structure factors were calculated with the deuterium atom on the N^{π} of His-57. In agreement with the refinement results, a nearly featureless map was produced, signifying that this placement was in close accord with the data. In contrast, the placement of the deuterium on the $O_{\delta 2}$ of Asp-102 resulted in the interpretable difference map shown in Figure 9C. The directionality indicated by the difference peak is distinctly toward the imidazole ring.

The results from each of these three techniques independently support the conclusion that the mechanistically important proton is coordinated to the imidazole of His-57. Taken together, we believe they offer compelling proof that, at physiological pH, the protonated species in the tetrahedral intermediate is the imidazole of His-57 rather than the carboxylate of Asp-102.

MIP-Trypsin Structure as a Model for a Tetrahedral Intermediate Complex. It is accepted by most investigators that both the acylation and deacylation steps of the enzymatic peptide hydrolysis of serine proteases proceed through a tetrahedral intermediate (TI) (Caplow, 1969; Fersht & Requena, 1971a; Hunkapiller et al., 1976). These intermediates have virtually identical structures, with the exception that the –NHR group of the peptide in the acylation intermediate is replaced by an OH group in the deacylation intermediate. The close correspondence between the structures of the MIP group and the TI formed in deacylation (Stroud et al., 1974) makes the MIP-trypsin structure an ideal starting point for a detailed examination of the interactions among the catalytic residues at this step of the reaction.

In order to ensure that the location of the critical proton in MIP-trypsin, as determined by the neutron diffraction experiment, correctly represents that in the tetrahedral intermediate (TI), it is important to ascertain whether the differences in charge distribution between the MIP and TI are substantial enough to alter the protonation states of the catalytic groups (Asp vs. His) under the experimental conditions of the neutron diffraction experiment.

Two methods were used to quantify this effect. In the first, the relative amounts of electrostatic potential energies derived from the ion–ion interaction between the imidazolium ion and the MIP group on one hand and the TI on the other were compared. Both the phosphate ester, MIP, and the TI possess a net charge of −1, and because of their close proximity to the imidazole (Figure 2), they should raise the pK_a of His-57 to some extent.

Electrostatic potentials[2] were calculated for the His^+–MIP^- and His^+–TI^- systems from the individual atomic charge

[2] Electrostatic potentials were calculated by using the expression $E = \sum_{i \neq j} q_i q_j / (\epsilon_0 r)$. ϵ_0, the dielectric constant, was 3.0; r is the distance between charges q_i and q_j.

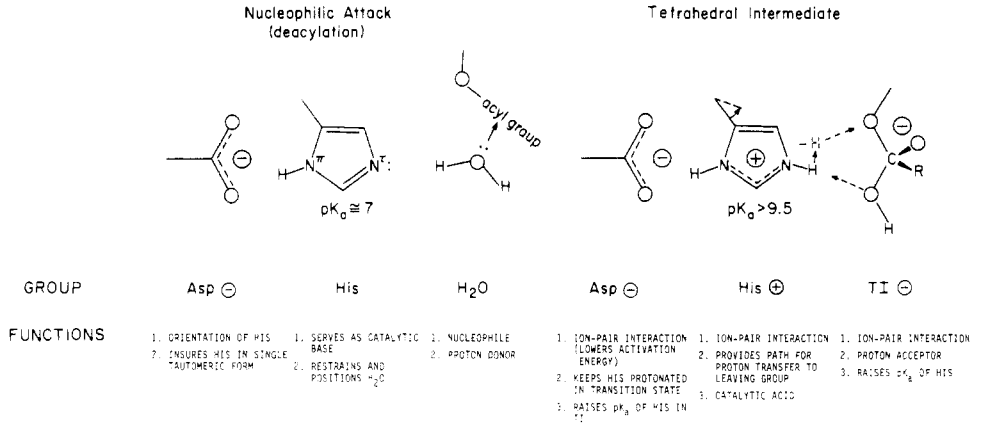

FIGURE 10: Functions of the catalytic groups during the deacylation step of the hydrolysis reaction. As the TI is formed, the electrostatic and dielectric environment of the His-57 imidazole is altered by the influences of Asp-102 and the TI. As a consequence, the pK_a of His-57 is raised, and because of the relatively low proton affinities of the Asp and TI, the imidazole remains protonated in the transition state to form a set of ion-pair interactions ($Asp^-His^+TI^-$).

densities for the MIP group and the model of the TI given in Table III. Ligand R is the side-chain moiety in the TI structure, while in MIP this is an isopropyl group. The phosphoryl oxygen, O(1), in MIP corresponds to an OH group in the TI. From the table, it can be seen that the charges on the Ser-195 O_γ and oxygen, O_2, are similar in both the MIP and TI species. While the R and O(1) groups are significantly more negative in MIP, the phosphorus is substantially more positive than a transition-state carbon, with the net result that the two effects approximately cancel in the electrostatic energy calculation.

The charge centroids of the imidazolium ion, the MIP group, and the TI were computed from their individual atomic charge densities (Desmeules, 1980; Perahia et al., 1975; Kleier et al., 1976). The use of charge centroids to calculate electrostatic potential energies is not rigorously correct, since each atom in the group has a somewhat different dielectric environment. However, since relative rather than absolute energies are being compared, this should not be a significant problem because corresponding atoms in the MIP and TI groups have virtually identical environments.

The electrostatic potential energy of a system is a function of the interacting charges, the intercharge distance, and the effective dielectric environment separating them. The independent parameter in the energy expression is the relative distance between the charge centroids. The distance between the charge centroids of His^+ and MIP^- was calculated to be 4.35 Å and that between His^+ and TI^- to be 4.65 Å. If the effect of charge transfer in attenuating the charges on the imidazole and TI (Kollman, 1981) is considered, the differential of 0.25 Å corresponds to an energy change of approximately 1 kcal/mol. An even smaller energy difference would be obtained if the effect of solvation of the ligands [O(1) and R] were taken into consideration. In any case, the comparable magnitudes of the energies of the two interactions strongly suggest that their influence on the protonation of His-57 should be essentially the same.

A second method was used to assess the significance of the MIP–TI electrostatic differences on the pK_a of His-57 to corroborate these results. This was a procedure developed by Gurd and his colleagues (Shire et al., 1974; Friend & Gurd, 1979; Matthew et al., 1979), which can be used to derive a mathematic model to estimate the effects of neighboring charge interactions on the pK of a titrating group. ΔpK values for His-57 were calculated by using the charge distributions in Table III. No attenuation due to individual solvent accessibilitites was incorporated in the calculation, so that the resulting changes represented a maximum limit on ΔpK. The calculated effects of the MIP and TI interaction were found to be the same within 0.2 of a pK unit (+1.13 pK and +0.97 pK units, respectively).

Based on the results from the two methods, it is evident that the overall electrostatic character of the MIP group closely resembles that of a real tetrahedral intermediate. Hence, it can be stated with confidence that this neutron diffraction result showing that His-57 is the catalytic base in the hydrolysis mechanism also correctly describes the protonation states of the catalytic residues in the tetrahedral intermediate.

It should be noted that these calculations further show that the interaction with the TI species affects the chemical characteristics of His-57 to a significant degree. In the native enzyme, this important charge interaction is absent. In addition, the orientation and effective dielectric around the imidazole are different in the native configuration compared to those of the enzyme–substrate complex (Robertus et al., 1972; Matthews et al., 1975; Ruhlmann et al., 1973; Sweet et al., 1974). Because of these factors, the results of those studies which based their assignments of the Asp–His protonation states on the titration of the active site group in the native enzyme (Hunkapiller et al., 1973; Robillard & Shulman, 1972; Koeppe & Stroud, 1976; Bachovchin & Roberts, 1978; Markley & Ibanez, 1978) have to be viewed with the appropriate reservations.

Implication of the Structural Findings to the Roles of the Catalytic Residues. The structural findings reported here offer compelling proof that the protonated species in the tetrahedral intermediate formed during deacylation is the imidazole of His-57 rather than the carboxylate of Asp-102, and the structural similarities between the deacylation and acylation intermediates suggest that this is also true for the acylation step of the reaction. In light of this finding, it is appropriate to review other aspects of the reaction mechanism and to discuss the functions performed by the catalytic groups. The results of this review are summarized in Figure 10 and discussed in subsequent sections. In addition to the relevant experimental data, the discussion includes references to model-building experiments (Steitz et al., 1969; A. A. Kossiakoff, unpublished experiments) which identify important stereochemical and entropic aspects of the initial step of the deacylation reaction.

(*a*) *Role of Asp-102.* The neutron diffraction experiment clearly shows that His-57 is the catalytic base in the hydrolysis reaction. Hence, the double proton transfer mechanism (Hunkapiller et al., 1973), in which the negative charge on

Asp-102 is in effect transferred to Ser-195 by the simultaneous movement of protons H(1) and H(2) (Figure 1), is clearly incorrect. In the often cited "charge–relay" mechanism (Blow et al., 1969), it is proposed that the aspartate channels some of its negative charge through the His-57 imidazole to the Ser-195 O_γ, effectively enhancing the nucleophilic properties of the hydroxyl group. This charge transfer is said to be accomplished through a series of "electronic" rearrangements of hydrogen bonds and is distinct from the double proton transfer mechanism in that it does not necessarily involve the actual transfer of the H(1) and H(2) protons. However, this "charge–relay" concept has been challenged on both experimental and theoretical grounds (Matthews et al., 1977; Kraut, 1977; Nakagawa et al., 1980; Desmeules, 1980).

The importance of the aspartate to the reaction mechanism, therefore, does not apparently include the special properties invoked by either of these two mechanisms. In fact, as in the case of His-57, the function of Asp-102 can be explained satisfactorily by its normal chemical properties. In brief, the importance of the aspartate to the catalytic effectiveness of the serine proteases can be attributed to both its precise positioning in the active site and its negatively charged state under the conditions of the reaction.

In the tetrahedral intermediate, there exists a charge interaction among the triad ($Asp^-His^+TI^-$). The thermodynamic properties of a similar arrangement of ionized groups (i.e., − + −) in the catalytic site of lysozyme have been studied (Warshel & Levitt, 1976; Warshel, 1978), and for this enzyme, it was proposed that the ion-pair interactions play an important role in lowering the activation energy of the reaction. In this case, compared to the reaction in water, a rate enhancement of 10^6 was estimated. Although a similar rate enhancement cannot be directly extrapolated to the serine protease system, the similarities between the ion-pair interactions in these two enzymes suggest that the electrostatic potential energy derived from these interactions is also important in the serine protease system.

Two other functions which Asp-102 performs are listed in Figure 10. (Although deacylation is shown, they apply to the acylation reaction as well.) Besides its role in the charge triad, it is involved through hydrogen bonding in orienting His-57 and ensuring that during nucleophilic attack by water (deacylation) or the Ser-195 O_γ (acylation) the imidazole remains in the correct tautomeric form with N^τ unprotonated. These two functions increase the effectiveness of His-57 as a catalytic base by ensuring the correct stereochemistry for proton transfer from the nucleophile. When the composite of these functions is analyzed, it is apparent that the group best qualified to perform them is one having an acid side chain. These functions are sufficient to explain the importance of Asp-102 without postulating unusual chemical properties for this residue.

(*b*) *Entropic Effects of His-57 during Deacylation.* It is known from crystallographic evidence that in acylation the precise positioning of the nucleophile (Ser-195 O_γ) relative to a highly developed binding template is an integral part of the protein structure. In contrast, during deacylation, the attacking water molecule approaches the bound acyl group from the surrounding solvent region, and thus, this process would not seem to have the same degree of stereospecificity toward the substrate as its acylation counterpart. However, the model-building studies discussed below suggest that His-57 can serve as an orienting group for the attacking water molecule.

It has been suggested by Steitz et al. (1969) that a water molecule could be positioned in such a manner that it could hydrogen bond to the N^τ of His-57 and direct its oxygen lone-pair electrons at the acyl group. A similar arrangement of the His-57 side chain and a water molecule has been observed experimentally in the indoleacryloyl-chymotrypsin structure by Henderson (1970); however, it was pointed out by this author that this derivative of chymotrypsin is not a particularly good model for an acyl enzyme.

In order to quantify in detail the extent to which nucleophilic attack by the water molecule on a bound acyl group is compatable with the stereochemical parameters accepted for most simple nucleophilic additions to carbonyl groups, a computerized model-building study was performed (A. A. Kossiakoff, unpublished experiments). The groups incorporated in the model (Asp-102, His-57, the acyl group, and water) were adjusted to the predicted stereochemistry at three successive stages of the "reaction" (corresponding to water molecule to acyl carbon distances of 3.5, 2.6, and 1.53 Å) by a computerized model-building procedure based on a steepest descent gradient minimization. Details of these model-building experiments will be presented elsewhere; however, since the findings bear directly on many of the stereochemical factors of the reaction mechanism, they will be briefly discussed.

At each stage of the model reaction, it was found that the water molecule could be adjusted to fulfull the following requirements: (1) the existence of a good hydrogen bond to His-57, with one hydrogen directed toward N^τ and an oxygen lone pair pointed at the acyl carbon; (2) an approach direction for the water closely perpendicular (<12°) to the acyl plane. These stereochemical relationships suggest that the orientation effect of the His–H_2O hydrogen bond could be an important factor in increasing the efficiency of nucleophilic attack in the deacylation step.

This hydrogen-bonding alignment also significantly restricts the modes of movement of the water, in effect reducing the possible degrees of freedom of the water from 6 to 1 (only the rotation around the hydrogen-bonding axis is unconstrained). Since an enzyme's ability to limit the translational and rotational degrees of freedom of the reacting groups provides the important entropic driving force for catalysis (Bruice, 1971; Page & Jencks, 1971), this alignment of the water molecule by the H_2O–His hydrogen bond should significantly enhance the reaction rate.

Bruice & Benkovic (1964) have suggested that the entropic effects on reaction rates can be approximated by the simple empirical relationship that each increase in reaction order adds a factor of 4–5 kcal/mol to the system. In the present case, the H_2O–His interaction effectively decreases the order from 2 to 1; so a decrease of 4–5 kcal in the overall free energy of activation would be expected, producing a 1000-fold rate enhancement. Although this is an approximation, clearly the H_2O–His interaction is important in minimizing the unfavorable energy of activation resulting from restricting the degrees of freedom of the reactants in the transition state.

(*c*) *Proton Affinity of His-57 in the Tetrahedral Intermediate.* In both the acylation and deacylation processes, His-57 performs a bifunctional catalytic role, acting first as the base by accepting a proton from the attacking group and then as an acid by transferring this proton to the leaving group. One factor governing the timing of the rate-determining proton transfer event is the proton affinity of the catalytic group; that is, the proper relationship must exist between the proton affinities of the nucleophile, His-57, and the leaving group in order to ensure an energetically favorable pathway for proton

transfer both to and from the histidine.

Satterthwait & Jencks (1974b) have estimated that the pK_a of the leaving-group nitrogen in the acylation intermediate approximates that of the parent amine (ranging from 8 to 11). If one assumes that the activated complex has a TI structure, the question is raised as to how a histidine group with a pK around 7 can function effectively in its role as a catalytic acid unless its pK is altered during the formation of the TI. Based on the estimate of the leaving-group pK_a (and assuming a normal imidazole pK), Komiyama & Bender (1979) have proposed that the enzymatic cleavage of amides cannot proceed through a normal TI in the rate-determining step since the proton affinity of the leaving group (p$K_a \simeq$ 8–11) would presumably exceed that of imidazole (p$K_a \simeq$ 7) before the TI could be completely formed. Such a situation would require that transfer of the proton to the leaving group in the acid-catalyzed step be completed prior to the formation of the activated complex. The resulting structure is similar to that of the dipolar intermediates produced during the uncatalyzed aminolysis of acetate esters (Satterthwait & Jencks, 1974a), and Satterthwait & Jencks (1974a) have questioned whether the intermediate products formed in this mechanistic pathway would be stable enough to be viable intermediates in the course of the reaction.

The arguments put forth by Komiyama & Bender (1979) are based mainly on estimates of the relative pKs of the constituent groups and are thermodynamically reasonable if the pK of His-57 is about 7 in the transition state, as it is in the parent enzyme. However, the experimental evidence described below suggest that the effective pK of His-57 in the TI is probably in excess of 9.5.

In addressing this issue, it must be recognized that the conventional definition of pK applies only to acids (or bases) in which the ionizing site is accessible to the solvent. This is the case for His-57 in the native enzyme for which NMR evidence indicates that the pK_a is only slightly higher than that for a normal histidine (Robillard & Shulman, 1972; Hunkapiller et al., 1973; Porubcan et al., 1978; Bachovchin & Roberts, 1978). However, during the reaction, His-57 becomes hydrogen bonded to the TI, thereby insulating it from the solvent and generating a neighboring charged group, altering both the electrostatic and dielectric medium of the imidazole. As was evidenced from the calculations performed for the MIP–TI comparison described earlier, these changes would be expected to affect significantly the proton affinity of His-57 relative to its value in the native enzyme.

Robillard & Shulman (1974a), using ^{1}H NMR to study the effects of a series of TI analogues (derivatives of boronic acid) on the electronic properties of the catalytic site groups in chymotrypsin and subtilisin, reported that the N$^\pi$ proton resonance of His-57 was very sensitive to the binding of these negatively charged TI analogues. Furthermore, the protonation of the histidine was invariant over the pH range 6–9.5, where it would be expected that an unperturbed imidazole would change protonation state. These findings led them to conclude that these negatively charged analogues raise the pK of the catalytic site region so that the N$^\tau$ of His-57 remains protonated at pHs up to 9.5.

The pH-dependent activity profiles of the pancreatic serine proteases also can be interpreted to indicate that the proton affinity of His-57 is apparently raised in the TI. The observed pH-activity curves for hydrolysis of peptides and amides by the pancreatic serine proteases have a bell-shaped profile and reflect a maximum activity close to pH 8. The low-pH limb of the curve corresponds to ionization of the catalytic base, which has a pK_a of about 7 (Bender & Kezdy, 1964). The high-pH limb of the curve exhibits an apparent pK_a of about 9.5 (Spomer & Wootton, 1971; Fersht & Requena, 1971b; Kaplan & Dugas, 1969). Fersht & Requena (1971b), however, have demonstrated that this limb does not correspond to the deprotonation of the imidazole in the acid-catalyzed step [a process which should be kinetically distinguishable since it has been shown that in acylation this is the rate-determining step (O'Leary & Kluetz, 1970; Fersht, 1971; Caplow, 1969)] but is associated with the deprotonation of the amino terminus, an event which controls enzyme conformation and substrate binding. In the case of the chymotrypsin-catalyzed hydrolysis of a neutral amide substrate, K_{cat} (the rate-limiting, bond-breaking step) is pH independent at alkaline pH, and it is the enzyme–substrate dissociation constant, K_m, which is the pH-dependent factor (Himoe et al., 1967). The clear implication of these kinetic observations is that even at alkaline pH the imidazole still has the capability to function as a catalytic acid and that the decrease in activity results from the enzyme's inability to bind the substrate properly rather than any decrease in the bond-breaking capacity of the catalytic site groups.

These results are suggestive that the proton affinity of His-57 is raised during the formation of the TI, an effect presumably due to changes in the local environment of the imidazole. The value of the pK of His-57 is indicated to be close to that of the leaving group; therefore, from a thermodynamic standpoint, a reaction pathway through a tetrahedral intermediate is indeed feasible and is in agreement with kinetic investigations (Caplow, 1969; Fersht & Requena, 1971a; Hunkapiller et al., 1976).

Conclusion

The primary question concerning the mechanism of the serine proteases has been resolved by using neutron diffraction. An unambiguous assignment of the proton position in the Asp-102 and His-57 hydrogen bond has been made by using three independent tests. In each case, the result indicated that the proton was located on the imidazole of His-57. Since the bound inhibitor mimics a transition-state analogue, in both structural and electrostatic properties (Kossiakoff & Spencer, 1980), the imidazole of His-57 clearly functions as the catalytic group during the transition state of the reaction. This role for His-57 is one for which it is chemically well-suited. It can be concluded, therefore, that the catalytic effectiveness of the Asp-His-Ser triad is not due to some chemical effect perpetuated through the hydrogen bonding system as proposed in the charge–relay (Blow et al., 1969) and the double proton transfer (Hunkapiller et al., 1973) mechanisms but is rather a function of the standard chemical characteristics of these groups made maximally effective and efficient by their precise stereochemical positioning in the enzyme–substrate complex.

The results of the neutron analysis of trypsin are noteworthy in other respects as well. The analysis was performed with data collected from a crystal over an order of magnitude smaller than any crystal previously used for neutron protein crystallography; yet, the refinement statistics and the wealth of information obtained from the resulting maps clearly indicate that the data are of high quality. Also, it has been demonstrated that even at 2.2 Å the locations of hydrogen and deuterium atoms can be determined. The usability of crystals previously considered too small for neutron studies, coupled with the number of detailed structural features which can be evaluated at a resolution above 2.0 Å, should dramatically increase the number of protein systems which can be studied profitably by neutron diffraction.

Acknowledgments

We thank Drs. B. Schoenborn and J. Cain, who designed and developed the crystallographic station at the Brookhaven High Flux Beam Reactor, for their help and suggestions during the course of this work. We also thank Dr. R. Stroud for donating the trypsin crystal used in this work and Dr. J. Hanson for help with the computer graphics.

References

Alberi, J., Fischer, J., Radeka, V., Rogers, L. C., & Schoenborn, B. P. (1975) *IEEE Trans. Nucl. Sci. NS-22*, 255.

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

Bachovchin, W. W., & Roberts, J. D. (1978) *J. Am. Chem. Soc. 100*, 8041.

Bacon, G. E. (1975) in *Neutron Diffraction*, p 39, Clarendon Press, Oxford.

Bender, M. L., & Kezdy, F. J. (1964) *J. Am. Chem. Soc. 86*, 3704.

Birktoft, J. J., Blow, D. M., Henderson, R., & Steitz, T. A. (1970) *Philos. Trans. R. Soc. London, Ser. B 257*, 67.

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

Bode, W., & Schwager, P. (1975) *J. Mol. Biol. 98*, 693.

Bruice, T. C. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 569.

Bruice, T. C., & Benkovic, S. J. (1964) *J. Am. Chem. Soc. 86*, 418.

Cain, J. E., Norvell, J. C., & Schoenborn, B. P. (1975) *Brookhaven Symp. Biol. No. 27*, VIII-43.

Caplow, M. J. (1969) *J. Am. Chem. Soc. 91*, 3639.

Chambers, J. L., & Stroud, R. M. (1977a) *Acta Crystallogr., Sect. B B33*, 1824.

Chambers, J. L., & Stroud, R. M. (1977b) *Protein Data Bank*, Brookhaven National Laboratory.

Cohen, G. H., Silverton, E. W., Matthews, B. W., Braxton, H., & Davies, D. R. (1969) *J. Mol. Biol. 44*, 129.

Delbaere, L. T., Hutcheon, W. L., James, M. N., & Theissen, W. E. (1975) *Nature (London) 257*, 758.

Desmeules, P. (1980) PhD. Thesis, Princeton University.

Elrod, J. P., Gandour, R. D., Hogg, J. L., Kise, M., Maggiora, G. M., Schowen, R. L., & Venkatasubban, K. S. (1975) *Symp. Faraday Soc. No. 10*, 145.

Fahrney, D. E., & Gold, A. M. (1963) *J. Am. Chem. Soc. 85*, 997.

Ferro, D. R., McQueen, J. E., McCown, J. T., & Hermans, J. R. (1980) *J. Mol. Biol. 136*, 1.

Fersht, A. R. (1971) *J. Am. Chem. Soc. 93*, 3504.

Fersht, A. R., & Requena, Y. (1971a) *J. Am. Chem. Soc. 93*, 7079.

Fersht, A. R., & Requena, Y. (1971b) *J. Mol. Biol. 60*, 279.

Freer, S. T., Alden, R. A., Carter, C. W., & Kraut, J. (1975) *J. Biol. Chem. 250*, 46.

Friend, S. H., & Gurd, F. R. (1979) *Biochemistry 18*, 4612.

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

Henderson, R. (1971) *Biochem. J. 124*, 13.

Henderson, R., & Moffat, J. K. (1971) *Acta Crystallogr., Sect. B B27*, 1414.

Hermans, J., & McQueen, J. E. (1974) *Acta Crystallogr., Sect. A A30*, 730.

Himoe, A., Parks, P. C., & Hess, G. P. (1967) *J. Biol. Chem. 242*, 919.

Hunkapiller, M. W., Smallcombe, S. H., Whitaker, D. R., & Richards, J. H. (1973) *Biochemistry 12*, 4732.

Hunkapiller, M. W., Forgac, M. D., & Richards, J. H. (1976) *Biochemistry 15*, 5581.

Inward, P. W., & Jencks, W. P. (1965) *J. Biol. Chem. 240*, 1986.

Jansen, E. F., Nutting, M. D., & Balls, A. K. (1949) *J. Biol. Chem. 179*, 201.

Kaplan, H., & Dugas, H. (1969) *Biochem. Biophys. Res. Commun. 34*, 681.

Kleier, D. A., Scheiner, S., & Lipscomb, W. N. (1976) *Int. J. Quantum Chem., Quantum Biol. Symp. No. 3*, 161.

Koeppe, R. E., & Stroud, R. M. (1976) *Biochemistry 15*, 3450.

Kollman, P. (1981) *J. Am. Chem. Soc.* (in press).

Komiyama, M., & Bender, M. L. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 557.

Kossiakoff, A. A., & Spencer, S. A. (1980) *Nature (London) 288*, 414.

Kraut, J. (1977) *Annu. Rev. Biochem. 46*, 331.

Leung, V. C., Marsh, R. E., & Schomaker, V. (1957) *Acta Crystallogr. 10*, 650.

Luzzati, V. (1953) *Acta Crystallogr. 6*, 142.

Markley, J. L. (1978) *Biochemistry 17*, 4648.

Markley, J. L., & Porubcan, M. A. (1976) *J. Mol. Biol. 102*, 487.

Markley, J. L., & Ibanez, I. B. (1978) *Biochemistry 17*, 4627.

Martinek, K., Savin, Y. V., & Berezin, I. V. (1971) *Biokhimiya (Moscow) 36*, 806.

Matthew, J. B., Hanania, G. I., & Gurd, F. R. (1979) *Biochemistry 18*, 1919.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1975) *J. Biol. Chem. 250*, 7120.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T., & Kraut, J. (1977) *J. Biol. Chem. 252*, 8875.

Nakagawa, S., Umeyama, H., & Kudo, T. (1980) *Chem. Pharm. Bull. 28*, 1342.

Norvell, J. C., & Schoenborn, B. P. (1975) *Brookhaven Symp. Biol. No. 27*, II-12.

O'Leary, M. H., & Kluetz, M. D. (1970) *J. Am. Chem. Soc. 92*, 6089.

Ong, E. B., Shaw, E., & Schoellmann, G. (1964) *J. Am. Chem. Soc. 86*, 1271.

Page, M. I., & Jencks, W. P. (1971) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 30*, 1240.

Perahia, D., Pullman, A., & Berthod, H. (1975) *Theor. Chim. Acta 40*, 47.

Poland, D., & Scheraga, H. A. (1967) *Biochemistry 6*, 3791.

Pollock, E., Hogg, J. L., & Schowen, R. L. (1973) *J. Am. Chem. Soc. 95*, 968.

Porubcan, M. A., Neves, D. E., Rausch, S. K., & Markley, J. L. (1978) *Biochemistry 17*, 4640.

Ramachandran, G. N., & Srinivasan, R. (1970) in *Fourier Methods in Crystallography*, Chapter 4, Wiley-Interscience, New York.

Robertus, J. D., Kraut, J., Alden, R. A., & Birktoft, J. J. (1972) *Biochemistry 11*, 4293.

Robillard, G., & Shulman, R. G. (1972) *J. Mol. Biol. 71*, 507.

Robillard, G., & Shulman, R. G. (1974a) *J. Mol. Biol. 86*, 519.

Robillard, G., & Shulman, R. G. (1974b) *J. Mol. Biol. 86*, 541.

Ruhlman, A., Kukla, D., Schwager, P., Bartels, K., & Huber, R. (1973) *J. Mol. Biol. 77*, 417.

Satterthwait, A. C., & Jencks, W. P. (1974a) *J. Am. Chem. Soc. 96*, 7018.

Satterthwait, A. C., & Jencks, W. P. (1974b) *J. Am. Chem. Soc. 96*, 7031.

Scheiner, S., & Lipscomb, W. N. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 432.
Scheiner, S., Kleier, D. A., & Lipscomb, W. N. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2606.
Schoenborn, B. P. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 569.
Schoenborn, B. P. (1975) *Brookhaven Symp. Biol. No. 27*, I-10.
Shaw, E., Mares-Guia, M., & Cohen, W. (1965) *Biochemistry 4*, 2219.
Shire, S. J., Hanania, G. I., & Gurd, F. R. (1974) *Biochemistry 13*, 2967.
Spencer, S. A., & Kossiakoff, A. A. (1980) *J. Appl. Crystallogr. 13*, 563.
Spomer, W. E., & Wootton, J. F. (1971) *Biochim. Biophys. Acta 235*, 164.
Steitz, T. A., Henderson, R., & Blow, D. M. (1969) *J. Mol. Biol. 46*, 337.
Stockton, G. W., Polnazzek, C. F., Leitch, L. C., Tullock, A. P., & Smith, I. C. (1974) *Biochem. Biophys. Res. Commun. 60*, 844.
Stroud, R. M., Kay, L. M., & Dickerson, R. E. (1971) *Cold Spring Harbor Symp. Quant. Biol. 36*, 125.
Stroud, R. M., Kay, L. M., & Dickerson, R. E. (1974) *J. Mol. Biol. 83*, 185.
Stroud, R. M., Kreiger, M., Koeppe, R. E., Kossiakoff, A. A., & Chambers, J. L. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E.) pp 13-32, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Sweet, R. M., Wright, H. T., Janin, J., Chothea, C. H., & Blow, D. M. (1974) *Biochemistry 13*, 4212.
Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V., & Wright, L. H. (1973) *Biochemistry 12*, 4185.
Warshel, A. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 5250.
Warshel, A., & Levitt, M. (1976) *J. Mol. Biol. 103*, 227.
Watson, H. C., Shotton, D. M., Cox, J. M., & Murihead, H. (1970) *Nature (London) 225*, 806.

Effects of an Atherogenic Diet on Apolipoprotein E Biosynthesis in the Rat†

Yen-Chiu Lin-Lee, Yoshio Tanaka, Chin-Tarng Lin, and Lawrence Chan*

ABSTRACT: The effects of an atherogenic diet on apolipoprotein E biosynthesis in the rat liver were studied by immunocytochemical and biochemical techniques. Two groups of rats were fed either a normal laboratory chow or a special atherogenic diet containing 5% lard, 1% cholesterol, 0.35% taurocholic acid, and 0.1% propylthiouracil for 25 days. The atherogenic diet fed animals developed increased plasma cholesterol concentrations (134 ± 22 vs. 70 ± 15 mg/dL) and apolipoprotein E (apoE) concentrations (42.5 ± 10.2 vs. 15.6 ± 4.8 mg/dL). Plasma albumin levels were unchanged (7.3 ± 3.0 vs. 7.4 ± 2.0 g/dL). When liver sections from these animals were studied by indirect immunoperoxidase staining, by using a rabbit anti-rat apoE serum, the number of apoE-positive grains increased from 5.45 ± 0.49 to 6.79 ± 0.54 per hepatocyte, and the intensity of staining of individual grains also increased. Rat liver slices were incubated in vitro in culture medium containing [^{35}S]methionine for 90 min at 37 °C. Radioactivity incorporated into immunoprecipitable apoE were 2.89 ± 0.2% of Cl_3AcOH-precipitable radioactivity in special diet fed animals compared to 1.32 ± 0.1% in normal controls. Again, radioactivity incorporated into immunoprecipitable albumin was unchanged (11.9% in special diet group vs. 11.5% in controls). Total RNA was isolated from the liver of both groups of animals. Poly(A) RNA was purified by oligo(dT)-cellulose chromatography. Translation of the poly(A) RNA in a wheat germ system in vitro indicates that the atherogenic diet fed rat liver contained significantly higher concentrations of apoE mRNA activity (3.17% of total Cl_3AcOH-precipitable cpm) when compared to that of the controls (1.5% of total activity). In contrast, albumin mRNA activity in these RNA samples stayed constant at 9.1% (atherogenic diet fed) and 8.9% (controls), respectively. Our observations have provided unequivocal evidence that feeding the rats the special atherogenic diet for 25 days leads to an increase in apoE synthesis in the liver. This increase was mediated at least in part by an accumulation of translatable apoE mRNA activity in this organ.

Hypercholesterolemia is an important factor in the development of atherosclerosis (Goldstein et al., 1973). Plasma cholesterol levels in man can be altered by supplementation of the diet with polyunsaturated or unsaturated fats, as well as with cholesterol (Shore et al., 1974). The rat has been used by a number of laboratories as an experimental model for diet-induced hypercholesterolemia (Mahley & Holcombe, 1977; Swaney et al., 1977). In this animal, the various lipoproteins are well characterized and the metabolic interconversions of these lipoproteins have also been studied extensively (Swaney et al., 1977; Wong & Rubinstein, 1977).

The rat does not develop atherosclerosis naturally. It is also relatively resistant to hypercholesterolemia. However, high serum cholesterol and lipoprotein levels and even atherosclerosis can be induced in this animal by feeding it a diet containing 5% lard, 1% cholesterol, 0.35% taurocholic acid, and 0.1% propylthiouracil (Mahley & Holcombe, 1977; Mahley, 1978). Rats as well as other animals fed such a diet (referred to below as atherogenic diet) develop characteristic changes in their lipoprotein patterns. These include the following: the occurrence of β-VLDL[1] (i.e., cholesterol-enriched

† From the Departments of Medicine and Cell Biology, Baylor College of Medicine, and the Methodist Hospital, Houston, Texas 77030. *Received May 15, 1981.* This work was supported by grants from the National Institutes of Health (HL 23470), the American Heart Association (78-1102), and the Texas Affiliate of the American Heart Association.

[1] Abbreviations used: VLDL, very low density lipoproteins; IDL, intermediate density lipoproteins; LDL, low density lipoproteins; HDL, high density lipoproteins; apo, apolipoproteins; oligo(dT), oligothymidylate; EDTA, ethylenediaminetetraacetic acid; Gdn, guanidine; Tris, tris(hydroxymethyl)aminomethane; $NaDodSO_4$, sodium dodecyl sulfate; Cl_3AcOH, trichloroacetic acid.

0006-2960/81/0420-6474$01.25/0