THE FACTORS IMPORTANT TO CATALYSIS BY SERINE PROTEASES: STRUCTURE OF DIISOPROPYLFLUOROPHOSPHATE-INHIBITED BOVINE TRYPSINOGEN REFINED AT 2.1 Å RESOLUTION

Thesis by:

Melvin O. Jones

In Partial Fulfillment of the Requirements

for the Degree of Master of Science

1982

California Institute of Technology

Pasadena, California 91125

(Submitted October 1, 1981)

ACKNOWLEDGEMENTS

I would like to thank my advisor Bob Stroud and the other members of the Stroud group who have lent me so much support, both academic and personal over the last several years.

The text of this thesis shall be published and I would like to thank those people who have directly contributed to this study, my co-authors: Dr. John Chambers, Dr. Anthony Kossiakoff, Dr. L.M. Kay and Dr. R.M. Stroud.

I would also like to thank Ms. Rebecca Price for technical assistance and Drs. Monty Krieger and George Kenyon for valuable assistance.

ABSTRACT

Crystals of bovine trypsinogen inhibited with diisopropylfluorophosphate (DFP) were grown at pH 7 and x-ray intensity data to 2.1 Å resolution were collected. The reflections were assigned the phases determined by Kossiakoff (1977) for the isomorphous native trypsinogen structure. Using difference fourier techniques, the resulting protein structure was refined to a residual of 18.2%. As expected, a comparison between (DIP)-trypsinogen and the native trypsinogen structures shows all the structural differences to be in the catalytic site region. However, the structure of the catalytic region was expected to be quite similar to that of DFP-inhibited bovine trypsin at neutral pH, but there exist some marked differences.

In trypsin, one of the isopropyl groups of the inhibitor is analogous to the leaving group of a specific substrate and is hydrolyzed off the inhibitor. In the resulting monoisopropylphosphoryl (MIP)-trypsin, the inhibitor is oriented with a non-esterified phosphoryl oxygen in the oxyanion stabilization site as expected for a tetrahedral intermediate in proteolysis. His 57 Nε2 points toward the phosphoryl oxygen at the leaving group site, and is in a position suitable for proton donation to the leaving group.

By contrast, the inhibitor on DIP-trypsinogen remains intact and the presence of electron density at each of the three possible isopropyl locations indicates that the groups of the inhibitor are less specifically oriented. After independent refinement of each of the three possible orientations, the favored orientation was with isopropyl groups near the oxyanion and leaving group sites and the non-esterified phopsphoryl oxygen hydrogen-bonded through two solvent molecules to Ser 214 O. The imidazole ring of His 57 is displaced from its native position at the catalytic site out into the solvent region where it forms hydrogen-bonds to Tyr 94 O and through a solvent molecule to Ser 214 O. Ser 195 Oγ has moved closer to the position occupied by the His 57 imidazole in native trypsinogen and MIP-trypsin.

Both statistical variation in orientation of the inhibitor and movement of Ser 195 Oγ toward the native imidazole site can be explained by the nonfunctionality of the oxyanion binding site in trypsinogen.

The difference between DIP-Tgen and MIP-trypsin thus suggests both non-productive transition state or substrate binding and stereochemical incompatability between the tetrahedral intermediate and the native imidazole site as possible contributors to the relative inactivity of the proenzyme.

Conversely, and of more general importance, these results emphasize the importance of exact stereochemical alignment in generating a functional catalytic enzyme.

TABLE OF CONTENTS

INTRODUCTION

The pancreatic serine proteases are synthesized as inactive proenzymes, or zymogens. The conversion of trypsin to trypsinogen is an allosteric change mediated by cleavage of the N-terminal six amino acids. The newly formed amino terminus, isoleucine 16, folds into the proenzyme to form a charge interaction with aspartic acid 194. Enzymatic activity is generated by the subsequent conformational changes which alter both the static structures and dynamic properties of the substrate binding and catalytic sites (1, 2). The three-dimensional structures of two of these zymogens have been solved: chymotrypsinogen (3, 4, 5) and trypsinogen (7, 8, 9, 10). Structures of the corresponding enzymes, chymotrypsin (11, 12), and trypsin (13-19) have also been determined. Alterations in conformation of the protein in residues 188-194, which form the specific substrate binding pocket, and around Gly 193 N and Ser 195 N which are involved in stabilization of reaction intermediates (10, 21, 22) are good candidates for explaining the greatly reduced reaction rates in the zymogens (3, 4, 5, 8, 23). The catalytic site residues, Asp 102, His 57, and Ser 195, were in both cases in closely similar conformation in the zymogen and enzyme. Although, in this region, even subtle changes may be very important (6).

Since trypsin and trypsinogen have such similar active site conformations, without prior knowledge it would be difficult to deduce which, if either, of the two structures represented the active species. The initial goal of this study was to better characterize the active enzyme structure by comparing the structures of trypsin and trypsinogen in intermediate stages of hydrolysis.

Diisopropylfluorophosphate (DFP) is a powerful inhibitor of serine proteases (24), which will react with trypsinogen although it does so at a rate about four orders of magnitude slower than with trypsin (25). In the case of DFP inhibited trypsin, the inhibitor, which is here shown to be a negatively charged, tetrahedral, monoisopropyl derivative (MIP), is covalently attached to Ser 195 $O\gamma$ and exhibits characteristics expected of a tetrahedral intermediate in the enzymatic hydrolysis (26, 17, 18): it is a transition state analogue. The structure of DIP-trypsinogen might thus be expected to resemble an intermediate stage in hydrolysis by the zymogen, and so define in more detail some of the features which account for the zymogens greatly reduced activity relative to trypsin. In the following sections we describe the refined 3-dimensional structure of DIP-trypsinogen at 2.1 Å resolution and compare this structure to those of DIP-trypsin and native trypsinogen.

EXPERIMENTAL

DIP-Trypsin:

Conditions for the crystallization of DIP-trypsin have been described in detail by Stroud et al., (14). Bovine trypsin (Worthington, 2x crystallized, salt free lyophilized) was inhibited with DFP according to the procedure of Cunningham (26) and was stored at 4°C after lyophilization. The crystals grew over a period of several months at room temperature from solutions of 5.8% $MgSO_4$ at pH 7. The crystals are orthorhombic, space group $P2_12_12_1$ with unit cell parameters A=54.84 Å, B=58.61 Å, C=67.47 Å. DIP-trypsin used for thin layer chromatography experiments was stored at -20°C after lyophilization.

DIP-Trypsinogen:

Bovine trypsinogen inhibited by diisopropylfluorophosphate (25) was generously provided by M. Kerr, (University of Washington). The inhibited protein was further purified by column chromatography at 4°C (Sephadex SP-C-50) according to the procedure of Schroeder and Shaw (27), and then stored at -20°C until used. DIP-trypsinogen was crystallized by vapor diffusion. A 4% solution of DIP-trypsinogen and 10 mM Benzamidine HCl was adjusted to pH 7.3 with dilute NaOH and

allowed to equilibrate with a reservoir of 25% ethanol/H_2O (v/v) at 4°C. Crystallization occurred over a period of two months. The crystals are trigonal, space group $P3_121$, with unit cell parameters: A=B=55.17 Å, C=109.25 Å and are isomorphous with the native trypsinogen crystals described by Kossiakoff (8).

Data Collection and Reduction:

Data were collected to 2.1 Å resolution from a single crystal of dimensions .4 x .4 x .4 mM on a Syntex $P\bar{1}$ automated diffractometer employing the step-scan technique of Wyckoff (28). Data reduction was performed as described by Kossiakoff (8). The data were put on an absolute scale by scaling the amplitudes to those observed for isomorphous native trypsinogen which had been scaled according to the method of Wilson (29). The residual between the DIP and native observed structure factor amplitudes was .1302. The dependence of F(DIP)-F(NAT) /F(NAT) on resolution is shown in Figure 1. The final data set consisted of the 10,149 independent reflections whose structure factor amplitudes were greater than twice their standard deviation, of the 11,972 total independent reflections in the 2.1 Å sphere.

Structure Determination and Refinement:

Initial interpretation was based on a difference Fourier synthesis with terms (F(DIP)-F(NAT)) $\exp(i\phi_c)$ where F(DIP) and F(NAT) are the observed structure factor amplitudes for DIP-trypsinogen and native trypsinogen respectively and ϕ_c is the phase computed from the refined coordinates for the then current 1.8 Å refined native trypsinogen model (8). The resulting DIP-trypsinogen model was refined according to the procedure of Chambers and Stroud (15, 16). Because the reflections at spacings >7 Å are strongly influenced by the solvent continuum (which was not represented in the model), these reflections were omitted from the refinement. Reflections with Fo>>Fc are likely to be poorly phased in the difference syntheses, and reflections with (Fo-Fc)/0.5(Fo+Fc))>1.2 were therefore omitted on each cycle.

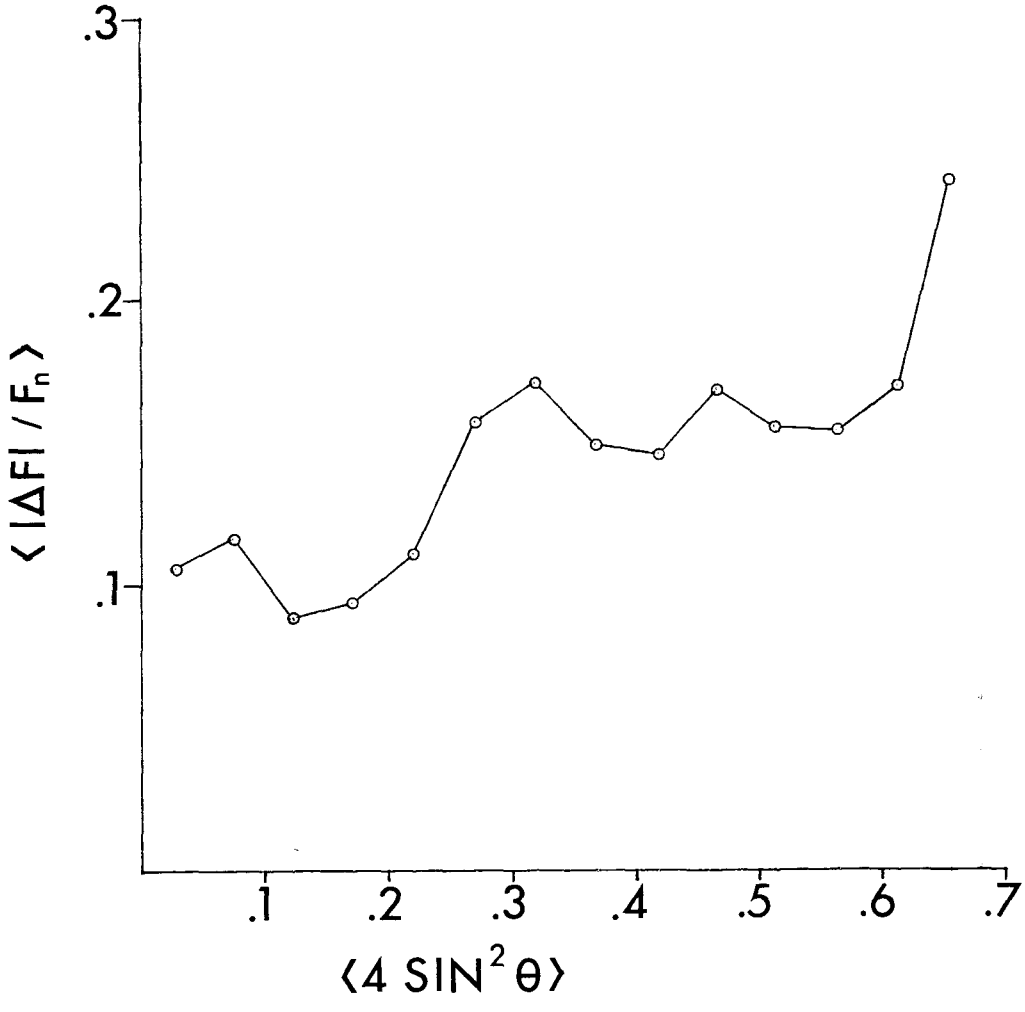

Figure 1. Graph of the dependence of $< F_{o(DIPTg)} - F_{o(DIPTg)} / F_{o(Tg)} >$ on resolution.

Peak Integrations:

Peaks in the electron density maps computed with Fourier coefficients $[2F_o(DIP)-F_c(DIP)]\exp(i\phi_c(DIP))$ were integrated numerically using the expression: $Z=\Sigma\rho_{xyz}\Delta V$ where Z is the integrated density, ρ_{xyz} is the electron density at the point xyz, and ΔV is the volume element defined by the map grid. Integration was carried out over those grid points above a specified electron density level and within specified radii of the chosen atomic positions. The specified electron density levels and radii were determined by examining the map and choosing these values to include all density for a given peak. The integrated values were put on an absolute scale by assuming full occupancy for the imidizole of His 40, the Cys 42 -Cys 58 disulfide bridge, and the indole ring of Trp 215, all of these groups were well defined and gave consistent integrated densities. The rms. error estimated for the integrated densitites of these three groups was 12.1% based on the disagreement between the integrations and the known numbers of electrons.

Thin Layer Chromatography:

Synthesis of reference esters: DIP and MIP reference esters were synthesized from triisopropyl phosphite (Pfaltz & Bauer). The starting

material was oxidized to triisopropyl phosphate by the method of Cox and Westheimer (31) with N_2O_4 gas (Matheson), then refluxed with NaI in methyl-ethyl ketone to produce a mixture of the di- and mono-phosphate esters. Protein preparation: The phosphate inhibitors were removed from 0.5 ml solutions of lyophilized DIP-trypsin, DIP-trypsinogen, and crystalline DIP-trypsin by the method of Oosterbaan, et al., (31). The protein was then removed by addition of 0.5 ml, 20% trichloroacetic acid and centrifugation. The pH of the supernatant was adjusted to 7.0 with NaOH before application to the TLC sheet.

TLC System:

The solvent employed for inhibitor identification was butanol-acetic acid-water (BAW)(2:1:5 by volume). The chromatograms were run on 5 cm x 10 cm Eastman #13255 cellulose sheets. Sheets were pre-run using the same solvent system.

Development:

Chromatograms were developed as described in Hettler (32) using the ammonium molybdate reagent of Hanes and Isherwood (33). Photo-reduction was used to bring about the blue color development of the phosphorous compounds.

Sensitivity and Controls:

In order to identify the reference esters and to determine whether they were altered during the protein preparation procedures, the reference esters were subjected to the conditions described under protein preparation. Chromatograms of thus treated reference solutions were not visibly different from untreated solutions. Chromatograms run in BAW (4:1:5) on Whatman #1 filter paper gave results (R_f(DIP)=.83, R_f(MIP)= .50) in agreement with those reported by Cohen et al (34).

A dilute solution of Na_2HPO_4 was employed to determine the sensitivity of the chromatograms. With this solution, .1-.2 μgm phosphate/cm^2 could be detected.

RESULTS

Studies of the three-dimensional crystal structure of DFP-inhibited trypsin (17, 18) indicated that conversion of the diisopropyl- to the monoisopropyl-phosphate ester had occurred, reflecting the 'ageing' process observed for a number of phosphate-ester inhibited serine proteases (35, 36, 37, 38). The conversion was verified by TLC as shown in Figure 2. It was therefore desireable to determine by TLC whether this conversion was also occurring in DIP-trypsinogen, the

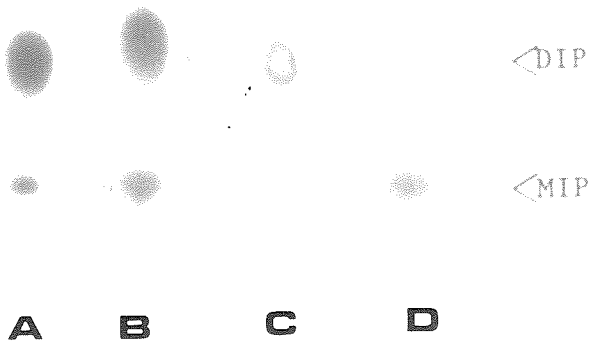

Figure 2. Thin-layer chromatograms prepared as described in the text. The spots at R_f arise from MIPOH and DIPOH, respectively. The samples on each chromatogram were: A) lyophilized DFP-inhibited trypsin; B) reference mixture of DIPOH and MIPOH; C) lyophilized DFP-inhibited trypsinogen; D) dissolved crystals of DFP-inhibited trypsin. The major trypsin species is monoisopropylphosphoryl-, and the major trypsinogen is diisopropylphosphoryl-.

resulting chromatogram is also shown in Figure 2. The reference compound was clearly resolved into two components corresponding to DIPOH (R_f = 0.42) and MIPOH (R_f= 0.20). The DIP-trypsinogen sample, once hydrolyzed with hydroxylamine, also gave two spots, but contained only a minor portion of MIPOH. In the case of the DIP-trypsin sample, a much greater percentage of the diester had been hydrolyzed to MIP, and in that of the DFP-inhibited trypsin crystal sample where the ageing had progressed nearly to completion, only a trace of the diester remained intact. These results clearly show that the predominant species in these trypsin crystals is MIP-trypsin, confirming the crystallographic results previously reported by Chambers & Stroud (17, 18).

In the case of DIP-trypsinogen, DIP constitutes the major component of the inhibitor and the crystallographic results suggest that the di-isopropyl ester also predominates in the DIP-trypsinogen crystals. A portion of the initial electron density map, computed with terms $[F_{o(DIPg)}-F_{o(Tg)}]e^{i\phi_c(Tg)}$ is presented in Figure 3. This region contained all of the prominent features in the map. The largest peak consisted of positive electron density connected to Ser 195 Oγ, arising from the inhibitor. There was also a large negative peak at the site occupied by the His 57 imidizole in the native zymogen, and a corresponding positive peak connected to His 57 Cβ in the immediately

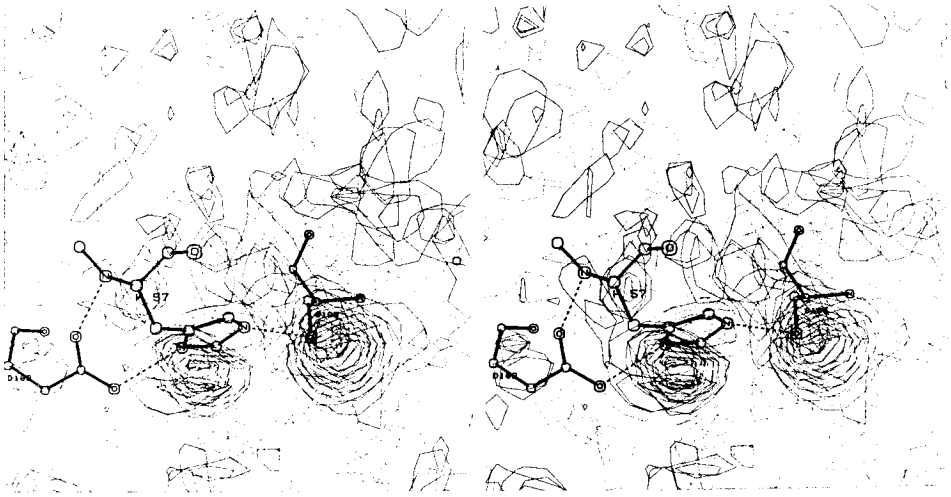

Figure 3. A portion around the catalytic site of the initial 2.1 Å difference map computed with terms $[F_{o(DIPTg)} - F_{o(Tg)}]e^{i\phi_c(Tg)}$, viewed down the crystallographic Z-axis (i.e., top view). The native trypsinogen model used in the phasing is superimposed. Solid contours represent positive electron density; dotted contours, negative density. The most prominent features are a) presence of the DIP-inhibitor covalently attached to Ser 195 O_γ; b) displacement of the His 57 imidazole into the solvent, and c) movement of Ser 195.

adjacent solvent region (Figure 3). Thus, movement of the imidazole from its native position between the side chains of Asp 102 and Ser 195 out into the solvent was indicated.

Although the tetrahedral arrangement of the oxygen atoms of the inhibitor could be discerned at this resolution, placement of the two isopropyl groups was ambiguous since there was significant positive electron density at all three possible sites. The isopropyl groups were therefore, omitted from the first phase calculations, and the starting model for refinement consisted of the then current 1.8 Å refined native trypsinogen coordinates (8) with the inhibitor (excluding isopropyl groups) attached to Ser 195 and with the imidazole of His 57 positioned out into the solvent. The calculated R-factor for this starting model was R = 0.274. The refinement of this model by the difference Fourier method is summarized in Table 1, and the resulting structure around the catalytic site is shown in Figure 4.

By stage 3 of the refinement (see Table 1), inspection of the F_o-F_c and $2F_o$-F_c maps showed that there was still significant electron density at each of the three possible isopropyl group locations. Such a situation could arise from either statistical variation in the orientation of the inhibitor or from a solvent molecule being hydrogen-bonded to the nonesterified phosphoryl oxygen. In Stage 3,

TABLE 1

Summary of DIP-Trypsinogen Refinement

Stage	R* at beginning of stage	R at end of stage	Number ΔF cycles	Comments
1	0.274	0.256	1	Refined only His 57 imidazole and phosphate group of inhibitor. Isopropyl groups omitted.
2	0.274	0.206	6	Refined all positions and thermal parameters. Isopropyl groups tentatively included on 02 and 03 (see Figure 4), but placement not unambiguous. 11 solvent molecules included.
3a	0.206	0.199	3	Isopropyl groups on 02 and 03. Solvent hydrogen bonded to 01. Unfavorable contacts between isopropyl group on 02 and His 57 imidazole.
3b	0.207	0.196	3	Isopropyl groups on 01 and 03. Solvent hydrogen bonded to 02.
3c	0.207	0.198	3	Isopropyl groups on 01 and 02. Solvent hydrogen bonded to 03. Unfavorable close contacts between isopropyl group on 02 and His 57 imidazole.
4	0.204	0.182	4	Began with stage 3b coordinates. Located 62 additional solvent molecules bringing total to 74. Addition of the solvent increased R from 0.196 to 0.204. Refinement of all positions and thermal parameter led to final model.

* R is the standard crystallographic R-factor for ∞ to 2.1 Å resolution.

$$R = \frac{\Sigma \left| F_{o(DIPTg)} - F_{c(DIPTg)} \right|}{\Sigma F_{o(DIPTg)}}$$

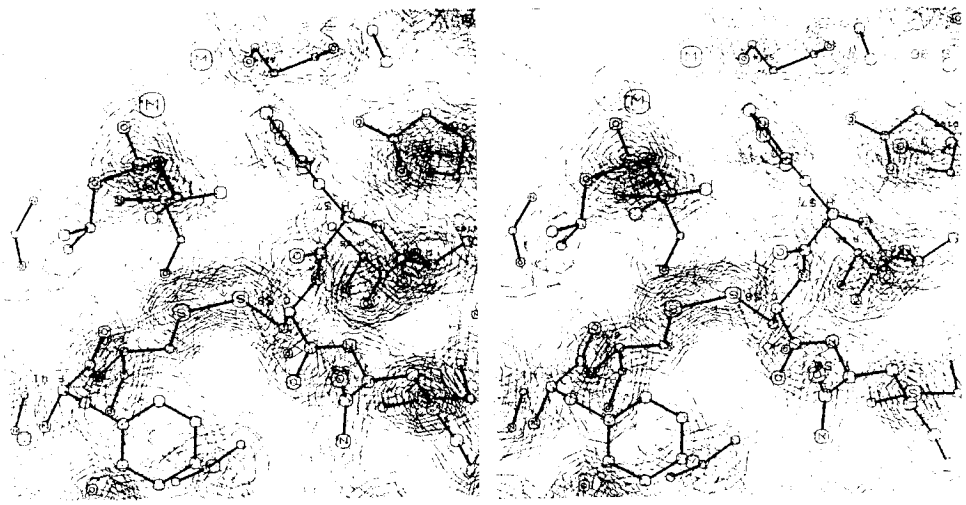

Figure 4. The refined DIP-trypsinogen structure in the catalytic site region, viewed down the crystallographic x-axis, (i.e., front view). The superimposed electron density was computed with terms $[F_{o(DIPTg)}-F_{c(DIPTg)}]e^{i\phi_c(DIPTg)}$. The imidazole of His 57 has been displaced into the solvent. In the orientation of the inhibitor which appears to be most favored, isopropyl groups are connected to phosphate oxygens at the oxyanion site (O1) and the leaving group site (O3). The phosphoryl oxygen at O2 is in a position analogous to that for the alpha-carbon of a peptide substrate.

models of the three possible orientations of the DIP group, with a solvent molecule hydrogen bonded to the phosphoryl oxygen, were separately refined and the resulting 2.1 Å, F_o-F_c and $2F_o$-F_c maps were carefully examined to see if one orientation was favored. It appeared that the best orientation of the inhibitor was that with isopropyl groups attached at phosphate oxygens 02 and 03 (see Figure 4).

The electron density for isopropyl groups on 01 and 03 is nicely forked, as shown in Figure 5, and the isopropyl groups fit quite well. This orientation also gave the lowest residual (Stage 3B in Table 1), although the improvement over those calculated for the other orientations may not be significant. All subsequent refinement was carried out on the model with isopropyl groups attached to 01 and 03. At this stage, 62 additional ordered solvent molecules were located and added to the twelve (including the tightly bound Ca^{++} ion; (8)) already incorporated. Addition of the solvent molecules increased the R-factor to 0.204. Four cycles of refinement of all positional and thermal parameters gave the final residual for the constrained model of 0.182 at 2.1 Å. Ninety-six reflections with very poor agreement (see Experimental) and those corresponding to spacings less than 7 Å were omitted from the Fourier map calculations (but not the R-factor) on the

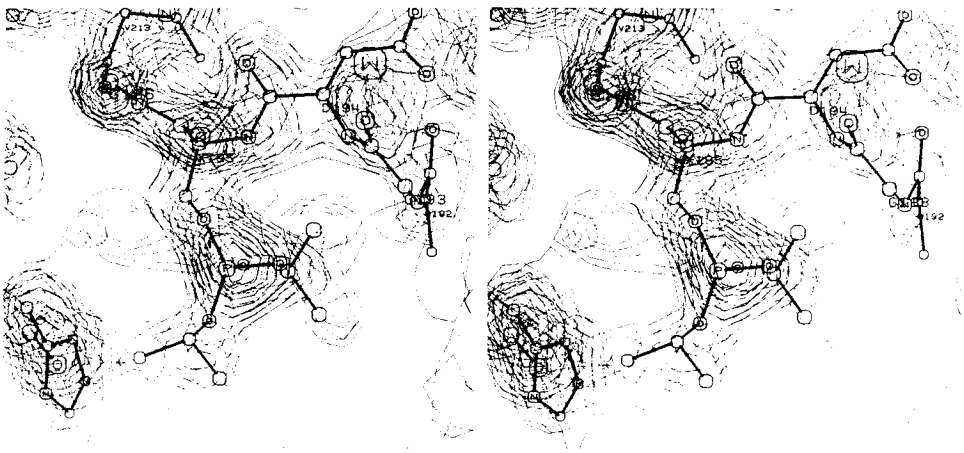

Figure 5. Top view, down the crystallographic Z-axis, of the catalytic site in DIP-trypsinogen, showing placement of the isopropyl groups in the favored orientation. The 2.1 Å electron density was computed with coefficients $2F_{o(DIPTg)}-F_{c(DIPTg)}$.

final cycle. Average deviations of bond lengths, bond angles, and dihedral angles from "ideal values" (obtained from Marsh and Donahue (37) were 0.025 Å, 3.05 Å and 5.49 Å, respectively for the final model. Standard deviations of the atomic positions were estimated for each atom in the refined model according to the method of Cruickshank (40) (Table 2). These estimates agreed well with the overall estimate of 0.25 Å made by Luzzati's (42) method, and with the magnitude of the changes still taking place in the structure over the last few cycles of refinement. The mean positional standard deviation estimated for each residue in the DIP-trypsinogen sequence is presented in Figure 6.

In DIP-trypsinogen the His 57 side chain has undergone rotations of approximately 90° about both the α-β and the β-γ bonds (see Figures 4 and 7). The electron density for the imidazole indicates that it is slightly mobile in this position with respect to rotation about the β-γ bond. With rotations about this bond the imidazole can form several hydrogen bonds with groups on the surface of the zymogen:

$$d_{H57N_{81}\text{-}Y940_n} = 3.0\ \text{Å};\quad \overline{\chi_2} = \chi_2 - 40.5°$$

$$d_{H57N_{81}\text{-}H570} = 2.6\ \text{Å};\quad \overline{\chi_2} = \chi_2 + 29°$$

$$d_{H57N\varepsilon 2\text{-}W12} = 2.8\ \text{Å};\quad \overline{\chi_2} = \chi_2 - 2°$$

$$d_{H57N\varepsilon 2\text{-}W11} = 3.4\ \text{Å};\quad \overline{\chi_2} = \chi_2 - 39.5°$$

TABLE 2

Estimated Positional Standard Deviation for Atom Types in the Refined DIPTg Structure*

B(Å^2)†	C	N	O	P	S	Ca^{2+}
4	0.236	0.186	0.153	-	0.077	-
5	0.245	0.193	0.159	-	-	-
6	0.254	0.200	0.164	-	-	-
7	0.263	0.207	0.170	-	0.086	-
8	0.272	0.215	0.177	-	0.089	-
9	0.282	0.222	0.183	-	-	-
10	0.292	0.231	0.190	-	-	0.075
11	0.303	0.239	0.197	-	0.099	-
12	0.314	0.247	0.204	-	-	-
14	0.336	0.265	0.219	-	-	-
16	0.360	0.284	0.234	-	-	-
18	0.385	0.304	0.251	-	-	-
20	0.412	0.326	0.269	-	-	-
24	0.470	0.372	0.307	-	-	-
28	0.534	0.423	0.350	-	-	-
32	0.606	0.481	0.398	-	-	-
36	0.685	0.545	0.451	-	-	-
40	0.773	0.615	0.510	-	0.257	-
50	1.031	0.823	0.684	-	-	-

* Computed according to the expression of Cruickshank (1949).

† Value of the refined individual isotropic thermal parameter. The mean value of B for all atoms in the refined structure was 19 Å^2.

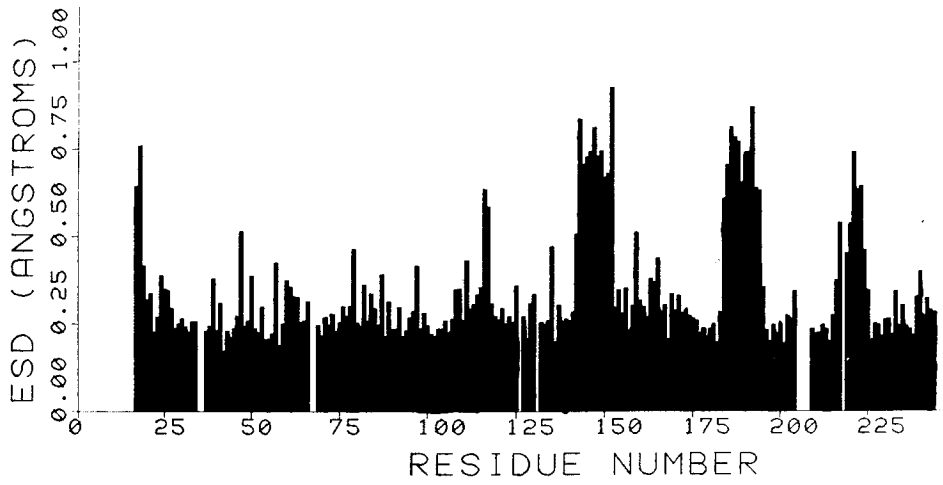

Figure 6. Histogram of the mean estimated positional standard deviation for each residue in the DIP-trypsinogen sequence.

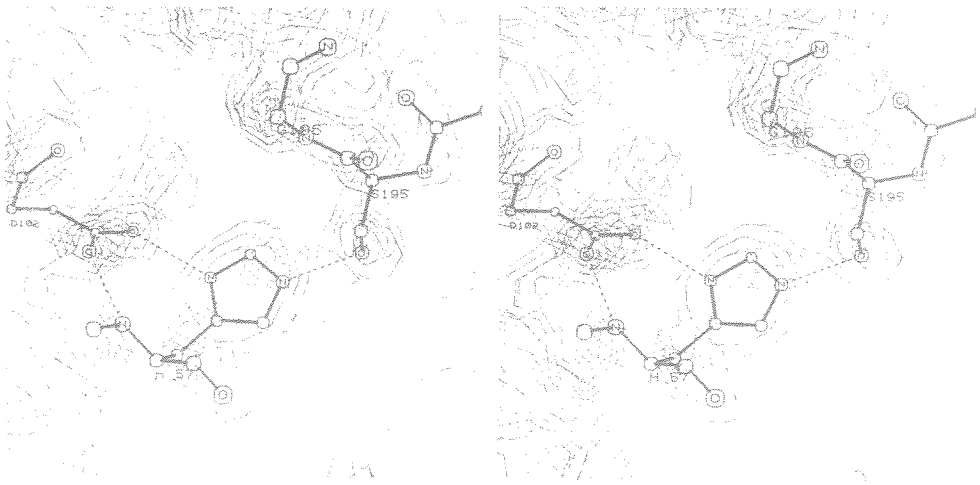

Figure 7. Top view, down the Z-axis, of the catalytic side in native trypsinogen. The imidazole of His 57 appears to form good hydrogen bonds to both Asp 102 and Ser 195. The electron density was computed with coefficients $2F_{o(Tg)}-F_{c(Tg)}$ to 2.1 Å resolution.

where $\overline{\chi}_2$ represents the value of the H57C_α, C_β, C_γ, $N_{\delta 1}$ torsion angle for the refined "average" imidazole position. Sovlent molecule W12, which is also present in native trypsinogen and in DIP-trypsin is hydrogen bonded to the carbonyl oxygen of Ser 214.

$$d_{S2140\text{-}W12} = 3.12 \pm 0.3 \text{ Å}$$

The phosphoryl oxygen atom of the inhibitor, O2 in Figure 4, can form a hydrogen bond through solvent W11 to W12 ($d_{O2\text{-}W11}$=2.5 ± 0.4 Å; $d_{W11\text{-}W12}$= 3.2 ± 0.4 Å). The isopropyl groups are positioned such that the one connected to O1 (Figure 4) fills the site which becomes the oxyanion hole (22) in the active eyzme, and the other which is connected to O3 occupies the position expected for the substrate leaving group. This is in marked constrast to MIP-trypsin where this second isopropyl group is absent, leaving a single isopropyl group pointing into the specific binding pocket (18).

The (F_o-F_c) difference maps over the course of the refinement contained a weak positive disc-shaped peak in the area occupied by the His 57 imidazole in native trypsinogen (8). Since this density

appeared connected to H57 C_β, and resembled an imidazole ring, some variation in position of the imidazole was suspected. To examine this idea the electron density peaks corresponding to the inhibitor and imidazole in the refined $2F_{o(DIPTg)}-F_{c(DIPTg)}$ map were integrated to determine their total electron content. The results are summarized in Table 3. They indicate that the inhibitor is present in approximately two-thirds of the protein molecules in the crystals and that the imidazole is also positioned out in the solvent in approximately two-thirds of the molecules. This result suggests that the peak at the native imidazole site arises from native trypsinogen molecules in the crystal, although statistical variation in the imidazole position in the DIP-trypsinogen molecules, or presence of a pocket of partially ordered solvent molecules cannot be ruled out absolutely.

DISCUSSION

It was not unexpected that all the major differences between trypsinogen and DIP-trypsinogen were in the active site region; however, the nature of some of these differences was quite surprising. The most unexpected feature was the repositioning of the His 57 imidazole outward into the solvent reigon.

TABLE 3

DIP-Trypsinogen Peak Integrations

Group	Formula	True # electrons	Integrated electron density	% Occupany
His 40 Imidazole	C_3H_3N	35	34.5	*
Trp 215 Indole	C_8H_6N	61	50.5	*
Cys 42-Cys 58 disulfide	S_2	32	42.9	*
His 57 imidazole in native position	$C_3H_3N_2$	35	21.7	62.0±12.8
DIP inhibitor	$C_6H_{14}O_3P$	89	53.9	60.6±12.5

* The side chains of residues 40 and 215, and the 42-58 disulfide were assumed to be fully occupied, and were used to normalize the integrated electron densities. The rms. error in the integrated values based on these three groups was 20.6%.

In the native enzyme and zymogen, the imidazole ring of His 57 is located in the catalytic site, where it forms a strong (i.e., linear and rather short --2.65 ± 0.2 Å) hydrogen bond from $N_{\delta 1}$ to Asp 102 $O_{\delta 2}$ and presumably a hydrogen bond from $N_{\varepsilon 2}$ to Ser 195 Oγ. Birkoft, et al. (6) observed that this latter hydrogen bond was strained in the structure of α-chymotrypsin with Ser 195 Oγ 2.5 Å from the optimal position for such an interaction, wherease less strain was present in the case of chymotrypsinogen. Matthews, et al., (42) extended this observation to include other native serine protease enzymes. In the case of the refined trypsin-trypsinogen pair at neutral pH, the difference although still present was not as striking. In benzamidine-inhibited trypsin Ser 195 Oγ is less than 1.5 Å from the optimal position; in MIP-trypsin the distance is 1.15 ± 0.2 Å, and in native trypsinogen, is 0.67 ± 0.3 Å (8). Distortion of this hydrogen bond could be important in initiation of proton transfer in the catalytic site upon substrate binding. The extent to which this observation affects current theories of proton transfer at the catalytic site of serine proteases depends on whether such a distortion also exists after substrate is bound. A high-resolution structure of an enzyme-substrate complex between a serine protease and a "normal" substrate has not yet been determined; however, in the trypsin-pancreatic trypsin inhibitor complex of Huber et al., (43), this hydrogen bond is linear and 2.7 Å long.

A number of studies have indicated that the active site histidine is movable from its native position under various conditions. In the structure of bovine trypsin inhibited by Ag^+ ion the silver ion was bound between the side chains of Asp 102 and His 57, causing the imidazole to swing about the α-β bond into the solvent (44). The structure of DFP-inhibited bovine trypsin differs from that of benzamidine-inhibited trypsin in that the His 57 side chain has rotated about the α-β bond, producing a 0.2 ± 0.07 Å movement of the imidazole (45). Thus, His 57 $N_{\varepsilon 2}$ points toward the leaving group oxygen in MIP-trypsin, whereas it forms a somewhat distorted hydrogen bond to Ser 195 O_γ in benzamidine-trypsin. The ^{13}C nmr experiments of Hunkapiller et al., (46) suggested that at pH ∿3, the catalytic site histidine of α-lytic protease is in equilibrium between its native positon and one where it had swung into the solvent. Results of tritium exchange at $H57C_{\varepsilon 1}$ also suggested that the imidazole in bovine trypsin is in this "out" conformation a small fraction of the time at neutral pH (47).

In the crystal structure of the related S. griseus protease A inhibited with tripeptide aldehyde at low pH (4.1)(48) the catalytic site histidine is forced into solution in a manner similar to that for DIPTg. The nmr experiments of Hunkapiller et al., (49) indicated that the active site histidine was forced into solution at low pH in an

α-lytic protease peptide-aldehyde complex, but was in its native position at neutral pH. The bacterial proteases thus appear to behave at low pH more like the pancreatic zymogens at physiological pH. Although these bacterial proteases probably do not have corresponding zymogens, they still contain a salt bridge analogous to the one formed by Asp 194 and the N-terminus upon activation of the pancreatic zymogens--between the internal side chains of Arg 138 and Asp 194 (50). The state of protonation of this salt bridge in the bacterial protease - peptide aldehyde complexes could presumably contribute to their more zymogen-like behavior at low pH.

The stereochemical incompatibility which forces the imidazole in DIPTg from its native position probably arises from the difference in structure between the oxyanion sites of the enzymes and zymogen. In DFP-inhibited bovine trypsin the inhibitor is ordered in a well-defined fashion with a phosphoryl oxygen atom hydrogen-bonded to the main chain amides of Gly 193 and Ser 195 (see Figure 8). His 57 $N_{\varepsilon 2}$ is properly positioned to donate a proton to the leaving group oxygen atom, and the isopropyl group formerly connected at that position has been lost. This configuration is as expected for an intermediate in the hydrolysis of specific amide or ester substrates. In trypsinogen, however, the part of the chain containing the Gly 193 and Ser 195 amides

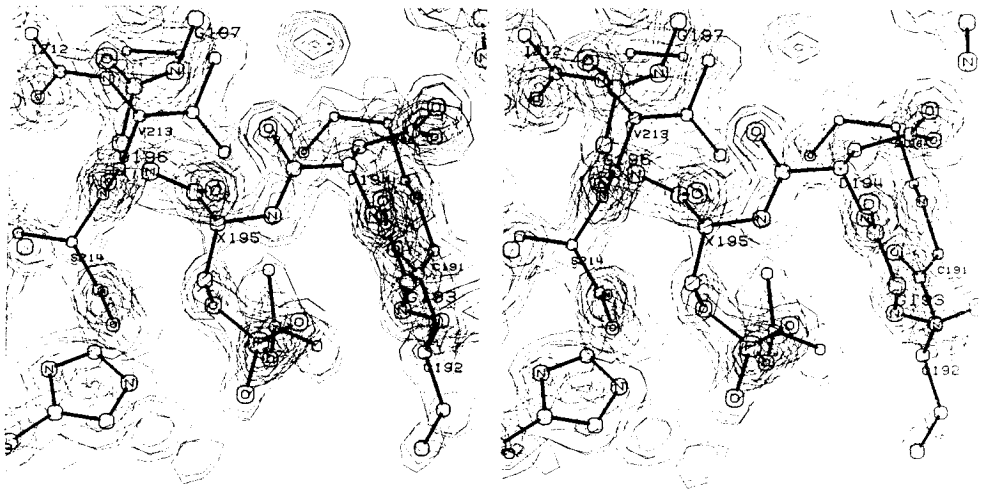

Figure 8. Front view of the catalytic site in MIP-trypsin. The imidazole of His 57 points toward the phosphoryl oxygen at the leaving group site. The isopropyl group formerly attached at this site has been lost. The superimposed electron density was computed with terms $[2F_{o(MIPT)} - F_{c(MIPT)}]e^{i\phi_{c}(MIPT)}$ to 1.5 Å resolution.

is moved approximately 2 Å back toward the cavity occupied by the N-terminus after activation of the proenzyme (8), and these amides are not available for hydrogen bonding (see Figure 7). As a result, the inhibitor in DIPTg is not held in this site in a well-defined manner. The inhibitor appears to occur in more than one orientation, indicated by presence of significant electron density at all three of the possible isopropyl group sites. In the favored orientation, the nonfunctional oxyanion site contains one of the isopropyl groups of the inhibitor. Such a situation is not possible in MIP-trypsin, where steric constraints at the oxyanion site do not permit accomodation of such a large group, and a phosphoryl oxygen is thus positioned there. Without the oxyanion site to hold the inhibitor in position, Ser 195 O_γ in DIPTg is considerably closer to the imidazole location of native trypsinogen than it is in MIP-trypsin or in the native zymogen. The distance between the Ser 195 O_γ postion in DIPTg and the His 57 $N_{\varepsilon 2}$ postion in native trypsinogen is only 2.45 Å; between Asp 102 $O_{\delta 2}$ in DIPTg and His 57 $N_{\delta 1}$ in native trypsinogen the distance is 2.88 Å. Although it is impossible to say from the crystallographic result whether the movement of Ser 195 O_γ is a cause or an effect of the imidazole displacement in DIPTg, it is likely that this steric crowding makes the native imidazole position energetically unfavorable and forces the His 57 side chain into the solvent.

There is some statistical variation in the orientation of the inhibitor in DIPTg, which is permitted by the extra room at the unformed oxyanion site. The favored orientation would correspond to a nonproductive binding mode in the case of an amide or ester substrate, with either the leaving group of the specific side chain positioned at that site. The structural results thus suggest that the probability of substrate binding in a nonproductive mode is higher in trypsinogen than in trypsin, although it appears that in a reaction between trypsinogen and a protein substrate, the normal mode of binding is forced on the complex by the rigid PTI structure, and at considerable energetic cost. The fact that the phosphate group is well localized in DIPTg in spite of the statistical variation in position of the isopropyl groups implies that the closeness of Ser 195 O_γ to the native imidazole site does not depend on this favored orientation of the inhibitor. Thus, displacement of the imidazole may occur in reaction of trypsinogen with a specific substrate ("productively" bound) as well, and could be a significant factor contributing to the relative inactivity of the zymogen.

CONCLUSION

The refined crystal structures of native bovine trypsin and trypsinogen are quite similar in the catalytically important regions; the most significant alternatives in these areas involved small (~2 Å) rearrangements of the structure at the oxyanion stabilization site and in the specific binding pocket (23, 8). Both of these segments form hydrogen bonds to the new N-terminal residues Ile 16 -Val 17 upon activation of the zymogen (Asp 194 carboxyl-Ile 16 N; Asp 189-Val 170, Asp 189 O- Val 17 N). Comparison of the inhibited enzyme and zymogen as transition state or intermediate-analogues yields structural differences which are less subtle than those between the native structures, and suggests in a more direct way how those subtle differences in structure can lead to large differences in reactivity of the enzyme and zymogen.

REFERENCES

1. Davie, E.W. and Neurath, H. (1955) J. Biol. Chem. 212, 515-529.
2. Neurath, H. and Dixon, G.H. (1957) Fed. Proc. Am. Soc. Exp. Biol. 16, 791-801.
3. Freer, S.T., Kraut, J., Robertus, J.D., Wright, H.T. and Xuong, Ng. H. (1970) Biochemistry 9, 1997-2009.
4. Wright, H.T. (1973a) J. Mol. Biol. 79, 1-11.
5. Wright, H.T. (1973b) J. Mol. Biol. 79, 13-23.
6. Birktoft, J.J., Kraut, J. and Freer, S.T. (1976) Biochemistry 15, 4481-4485.
7. Stroud, R.M., Krieger, M., Koeppe, R.E. II, Kossiakoff, A.A. and Chambers, J.L. (1975) "Proteases and Biological Control" in Cold Spring Harbor Symp. pp. 13-32.
8. Kossiakoff, A.A., Chambers, J.L., Kay, L.M. and Stroud, R.M. (1977) Biochemistry 16, 654.
9. Bode, W., Fehlhammer, H. and Huber, R. (1976) J. Mol. Biol. 106, 325-335.
10. Fehlhammer, H. Bode, W. and Huber, R. (1977) J. Mol. Biol. 111, 415-438.

11. Matthews, B.W., Sigler, P.B., Henderson, R. and Blow, D.M. (1967) Nature 214, 652.

12. Birktoft, J.J. and Blow, D.M. (1972) J. Mol. Biol. 68, 187-240.

13. Stroud, R.M., Kay, L. and Dickerson, R.E. (1971) Cold Spring Harbor Symp. Quant. Biol. 36, 125-140.

14. Stroud, R.M., Kay, L. and Dickerson, R.E. (1971) J. Mol. Biol. 83, 185-208.

15. Chambers, J.L. and Stroud, R.M. (1977a) Acta Cryst. B33, 1824-1837.

16. Chambers, J.L. and Stroud, R.M. (1977b) Amer. Cryst. Assn. Abstr. Vol. 5, No. 2, p. 59.

17. Chambers, J.L. and Stroud, R.M. (1979a) Submitted to Acta Cryst.

18. Chambers, J.L. and Stroud, R.M. (1979b) -in preparation.

19. Bode, W. and Schwager, P. (1975) J. Mol. Biol. 98, 693-717.

20. Steitz, T.A., Henderson, R. and Blow, D.M. (1969) J. Mol. Biol 46, 337-348.

21. Henderson, R (1970) J. Mol. Biol. 54, 341-354.

22. Robertus, J.D., Kraut, J., Alden, R.A., and Birktoft, J.J. (1972) Biochemistry 11, 4293-4303.

23. Stroud, R.M., Kossiakoff, A.A. and Chambers, J.L. (1977) Ann. Rev. Biophys. Bioeng. 6, 177-193.

24. Jansen, E.F., Nutting, M.D.F., Jang, R. and Balls, A.K. (1949) J.B.C. 179, 189-199.

25. Morgan, P.H., Robinson, N.C., Walsh, K.A. and Neurath, H. (1972) Proc. Natl. Acad. Sci. USA 69, 3312-3316.

26. Cunningham, L.W. (1954) J. Biol. Chem. 207, 443-458.

27. Schroeder, W.A. and Shaw, E. (1968) J. Biol. Chem. 243, 2943-2949.

28. Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B. and Richards, F.M. (1970) J. Biol. Chem. 245, 305-328.

29. Wilson, A.J.C. (1942) Nature 150, 151-152.

30. Cox, J.R. Jr., and Westheimer, F.H. (1958) JACS 80, 5441-5443.

31. Oosterbaan, R.A., Kunst, P, Van Rotterdam, J. and Cohen, J.A. (1958) Biochim. Biophys. Acta 27, 549-555.

32. Hettler, H. (1959) Chromatographic Reviews 1, 225-245.

33. Hanes, C.S. and Isherwood, F.A. (1949) Nature 164, 1107-1112.

34. Cohen, J.A., Oosterbaan, R.A., and Berends, F. (1967) Methods in Enzymology, Vol. XI, pp. 686-702.

35. Jandorf, B.J., Michel, H.O., Schaffer, N.K., Egan, R. and Summerson, W.H. (1955) Discus. Faraday Soc. 20, 134-142.

36. Davies, D.R. and Greun, A.L. (1956) Biochem. J. 63, 529-535.

37. Wagner-Jauregg, T and Hackley, B.E. Jr., (1953) JACS 75, 2125-2130.

38. Bender, M.L. and Wedler, F.C. (1972) JACS 94, 2101-2109.

39. Marsh, R.E. and Donohue, J. (1967) Advanc. Protein Chem. 22, 235-256.

40. Cruickshank, D.W.J. (1949) Acta Cryst. 2, 65-82.

41. Luzzatti, V. (1952) Acta Cryst. 5, 802-810.

42. Matthews, D.A., Alden, R.A., Birktoft, J.J., Freer, S.T. and Kraut, J. (1977) J. Biol. Chem. 252, 9975-8883.

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

44. Chambers, J.L., Christoph, G.G., Kreiger, M., Kay, L. and Stroud, R.M. (1974) Biochem. Biophys. Res. Comm. 59, 70-74.

45. Krieger, M., Kay, L.M., and Stroud, R.M. (1974) J. Mol. Biol. 83, 209-230.

46. Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.E. and Richards, J.H. (1973) Biochemistry 12, 4732-4743.

47. Krieger, M., Koeppe, R.E. II, and Stroud, R.M. (1976) Biochemistry 15, 3458-3464.

48. Brayer, G.D., Delbaere, L.T.J., and James, M.N.G. (1979) J. Mol. Biol. 124, 261-283.

49. Hunkapiller, M.W., Smallcombe, S.H. and Richards, J.H. (1975) J. Org. Mag. Reson. 7, 262-265.

50. Brayer, G.D., Delbaere, L.T.J., James, M.N.G., Bauer, C-A. and Thompson, R.C. (1978) Proc. Natl. Acad. Sci. USA 76, 96-100.