Structure of Bovine Trypsinogen at 1.9 Å Resolution†

Anthony A. Kossiakoff,‡ John L. Chambers,§ Lois M. Kay, and Robert M. Stroud*,¶

ABSTRACT: The three-dimensional crystal structure of bovine trypsinogen at ~pH 7.5 was initially solved at 2.6 Å resolution using the multiple isomorphous replacement method. Preliminary refinement cycles of the atomic coordinates for trypsinogen have been carried out first to a resolution of 2.1 Å, and later to 1.9 Å, using constrained difference Fourier refinement. During the process, structure factors F_c and φ_c were calculated from the trypsinogen structure and final interpretation was based on an electron-density map computed with terms $(2|F_o| - |F_c|)$ and phases φ_c at a resolution of 1.9 Å. Crystals of trypsinogen grown from ethanol–water mixtures are trigonal, with space group $P3_121$, and cell dimensions a = 55.17 Å and c = 109.25 Å. The structure is compared with the bovine diisopropylphosphoryltrypsin structure at ~pH 7.2, originally determined from orthohombic crystals by Stroud et al. (Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol. 36,* 125–140; Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1974), *J. Mol. Biol. 83,* 185–208), and later refined at 1.5 Å resolution by Chambers and Stroud (Chambers, J. L., and Stroud, R. M. (1976), *Acta Crystallogr.* (in press)). At lower pH, 4.0–5.5, diisopropylphosphoryltrypsin forms crystals which are isomorphous with trypsinogen, with cell dimensions a = 55.05 Å and c = 109.45 Å. This finding was used in the solution of the six trypsinogen heavy-atom derivatives prior to isomorphous phase analysis, and as a further basis of comparison between trypsinogen and the low pH trypsin structure. There are small differences between the two diisopropylphosphoryltrypsin structures. Bovine trypsinogen has a large and accessible cavity at the site where the native enzyme binds specific side chains of a substrate. The conformation and stability of the binding site differ from that found in trypsin at ~pH 7.5, and from that in the low pH form of diisopropylphosphoryltrypsin. The catalytic site containing Asp-102, His-57, and Ser-195 is similar to that found in trypsin and contains a similar hydrogen-bonded network. The carboxyl group of Asp-194, which is salt bridged to the amino terminal of Ile-16 in native trypsin or other serine proteases, is apparently hydrogen bonded to internal solvent molecules in a loosely organized part of the zymogen structure. The unusually charged N-terminal hexapeptide of trypsinogen, whose removal leads to activation of the zymogen, lies on the outside surface of the molecule. There are significant structural changes which accompany activation in neighboring regions, which include residues 142–152, 215–220, 188A–195. The NH group of Gly-193, normally involved in stabilization of reaction intermediates (Steitz, T. A., Henderson, R., and Blow, D. M. (1969), *J. Mol. Biol. 46,* 337–348; Henderson, R. (1970), *J. Mol. Biol. 54,* 341–354; Robertus, J. D., Kraut, J., Alden, R. A., and Birkoft, J. J. (1972), *Biochemistry 11,* 4293–4303) in the enzyme, is moved 1.9 Å away from its position in trypsin. Alterations in the binding site for substrate side chains are prime candidates for explaining the relative inactivity of trypsinogen.

Tryptic activation of the family of pancreatic proenzymes or zymogens, which include trypsinogen, chymotrypsinogen, procarboxypeptidase, and proelastase, takes place in the duodenum by limited proteolysis (Davie and Neurath, 1955; Neurath, 1975). The primary activating cleavage for the serine proteases is hydrolysis of the bond between a basic residue 15,[1] and isoleucine (valine in elastase and plasmin). This cleavage is accompanied by a conformation change which generates, or at least greatly enhances, enzymatic activity (Neurath and Dixon, 1957; Sigler et al., 1968).

The conversion of trypsinogen to trypsin is the first step of the digestive cascaded-activation sequence, and it relies on the known high specificity of the activator (enterokinase) for the unusual N-terminal sequence Val-Asp-Asp-Asp-Asp-Lys unique to trypsinogens (Maroux et al., 1971). The activation of further trypsinogen by trypsin at the second-cascade step requires that the most favored tryptic cleavage of that zymogen be the activating split between residues 15 and 16. Yet, at the same time, it has been shown that the presence of four aspartic acids in the activation peptide (residues 10–15) has a markedly negative effect on this tryptic cleavage (Abita et al., 1969). In all species, the activation peptide must generate a good substrate quality for the activator, and a poor, but nonetheless most favorable, tryptic cleavage site on the entire trypsinogen molecule. In this light, it is not too surprising to find that the activation peptide sequence of trypsinogen, with its dual role, is so highly conserved throughout all species (De Haen et al., 1975).

It seems likely that the fundamental basis for inactivity of the proenzymes will be shared by all of the trypsin-like enzymes. The comparison between the structures of chymotrypsinogen (determined by Freer et al., 1970) and trypsinogen should serve to isolate characteristics common to both proenzymes, which will limit the primary reasons for inactivity, to the lowest common factor. In this respect, there are real and significant differences between the two proenzymes. Taking for granted that the enzymes themselves are catalytically active

† From the Norman W. Church Laboratory of Chemical Biology, California Institute of Technology, Pasadena, California 91125. *Received June 16, 1976.* Contribution No. 5342. Supported by National Institutes of Health Grant GM-19984 and National Science Foundation Grant No. BMS75-01405.

‡ Recipient of a NIH Postdoctorial Fellowship. Current address: Department of Biology, Brookhaven National Laboratories, Upton, Long Island, N.Y. 11973.

§ Recipient of a NIH Predoctorial Traineeship.

¶ Recipient of a NIH Career Development Award and a Sloan Foundation Fellow.

[1] The numbering referred to is that of chymotrypsinogen sequence (Hartley, 1970). The sequence for trypsinogen is that of Walsh and Neurath (1964), and Mikes et al. (1967).

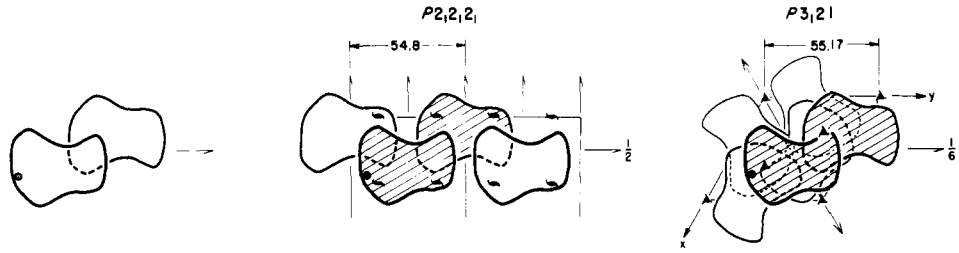

FIGURE 1: The packing diagram for trypsinogen (right) and trypsin (center) reveals a common intermolecular arrangement between molecules around a particular 2_1 axis.

(something which would not be obvious if the structures alone were viewed in the absence of such knowledge), we will compare the two proenzymes emphasizing the question as to why they should be relatively inactive. The answer to this question implies acceptance of the many well-established details of substrate binding and activity summarized, for example, by Blow (1974a).

Experimental Procedures

Crystallization and Derivative Preparation. Bovine trypsinogen (Worthington) was chromatographically purified using the procedure of Schroeder and Shaw (1968). Crystals were obtained by vapor diffusion of a 3% solution of protein against 30% ethanol-water. The crystals grew at 5 °C, as modified rhombohedra which generally averaged 0.25 mm across. A trypsin inhibitor, benzamidine, was added to the solution in a concentration of 2.5 mg/ml to prevent possible trypsin-mediated autolysis of the zymogen during the several weeks required for crystallization. It was essential to eliminate divalent cations from the crystallizing solutions, since they mimic the effects of Ca^{2+} and lead to autoactivation and autolysis of the trypsinogen in solution (Northrop et al., 1948). Crystals used for structure determination were grown from solutions adjusted to about pH 7.5, although isomorphous crystals were obtained throughout the pH range 5–8. Crystals grown this way are trigonal, space group $P3_121$ with cell parameters a = 55.17 Å and c = 109.25 Å. The unit cell volume is approximately 288 000 Å^3 and contains one molecule per asymmetric unit.

In view of the possible autolysis problem, the protein content of the crystals was checked by chromatography and by end-group amino acid analysis. Crystals were dissolved and found to contain pure trypsinogen by chromatography (Schroeder and Shaw, 1968). End-group analysis, using a modified Edman procedure, showed that the protein contained an N-terminal valine with no detectable alternative amino acid. This analysis rules out the possibility that any trypsinogen had been converted to trypsin during crystallization, since trypsin has an N-terminal isoleucine.

At low pH, DIP-trypsin[2] crystals were obtained which were isomorphous with the trypsinogen crystals. These crystals could be grown at pH 4.0–5.5 from either ethanol-water mixtures or from $MgSO_4$ solutions. At higher pH (5.7–8.5), trypsin crystallizes exclusively in the orthorhombic space group $P2_12_12_1$ (Stroud et al., 1971, 1974). At pH's between 4.0 and 5.5, the trigonal form of DIP-trypsin crystallized first. After several weeks, orthorhombic crystals grew from the same solution. The trigonal crystals were found to be unstable above pH 5.5, which suggests that there is some type of structural reorganization which occurs around pH 5.5.

In rare cases, DIP-trypsin crystals grew as trigonal-orthorhombic twins at about pH 5.5. These twins, in which the orthorhombic a axis was parallel to the trigonal a axis, suggested that the two crystal forms had some common symmetry element. The packing diagrams obtained from the structure solutions show that molecules are similarly disposed with respect to 2_1 screw axes in the two crystal forms. The 2_1 axes are parallel to the a axes in both the trigonal and orthorhombic crystals, as indicated in Figure 1. The unit-cell dimensions parallel to the 2_1 axes in the two crystal forms are 54.84 Å in orthorhombic DIP-trypsin and 55.17 Å in trigonal trypsinogen.

A vector parallel to the c axis of the orthorhombic cell in the trypsin molecule is tilted by a 13.49° rotation around the crystallographic 2_1 axis in the trigonal structure (eq 1, below).

Derivative Preparation. Isomorphous derivatives were prepared by placing the crystals into ethanol-water solutions containing heavy-atom salts. Optimum-diffusion conditions and soaking times were established by varying the pH from 6.0 to 8.0 and determining photographically the time necessary for a stable and reproducible diffraction pattern change to occur. Several of the derivatives showed a variation of pattern over an extended period of time.

The initial study of the HgI_4^{2-}, $Pt(NO_2)_4^{2-}$, and mersalyl derivatives showed that pattern stability had been reached after 1 to 2 weeks of crystal soaking. However, these same crystal soaks showed a further distinct change in diffraction pattern after 2 years. When the derivatives were solved, it was found that these changes were caused by appearance of several new minor sites, and fluctuation in the occupancies of several of the original sites. The overall differences between the short- and long-soak derivatives were large enough to allow them to be treated as independent derivatives in the phase refinement.

Data Collection. 1.8 Å data for native trypsinogen and 2.6 Å data for six derivatives were collected on a Syntex $P\bar{1}$ automated diffractometer using graphite-monochromatized Cu$K\alpha$ radiation. The detector was located 40.6 cm from the crystal, and a helium-filled tube was placed between the detector and the crystal to limit air absorption losses of the diffracted beam. Intensity data were collected using the Wyckoff step-scanning method (Wyckoff et al., 1970). An optimal scan width was chosen for each crystal separately, based on the criterion that the first steps outside of the usable peak should not be less than 95% of the peak step counts. Typically, the measurement of each reflection consisted of five steps centered on the diffrac-

[2] Abbreviations used are: DIP, diisopropylphosphoryl; DIPT, LpT, Tg, CT, and Cg refer to the crystallographically determined structures of diisopropylphosphoryl trypsin at ~pH 7.5 from orthorhombic crystals (Chambers and Stroud, 1976), low pH (~5.0) DIP-trypsin from trigonal crystals, trypsinogen (~pH 7.5) from trigonal crystals isomorphic with LpT crystals, tosyl-α-chymotrypsin (Birktoft and Blow, 1972), and chymotrypsinogen (Freer et al., 1970; Birktoft et al., 1976), respectively; NPGB, p-nitrophenyl-p'-guanidinobenzoate; BPTI, the basic pancreatic trypsin inhibitor.

tion maxima, taken over an omega range of 0.10°. The highest sum of three, or of four, adjacent steps was then treated as the integrated intensity. Scan rates ranged from 0.15° to 0.25°/min. In most cases, the native and derivative crystals were stable enough to collect a full 2.6 Å data set from a single crystal before any of the standard reflection intensities had changed by more than 10%.

The method used for data reduction and absorption correction was similar to that described by Stroud et al. (1974), but was altered for four-circle geometry. Individual background measurements were approximated by interpolation from empirical background curves, which were calculated as a function of 2θ, χ, and φ by the method of Krieger et al. (1974a).

Phase Analysis

Difference Patterson Methods. Initial phases for the trypsinogen $F_{(hkl)}$ were determined using the multiple-isomorphous-replacement method. Three-dimensional difference Patterson syntheses (Blow, 1958) were calculated for each one of the derivatives. However, of the ten derivatives examined, not one of the difference Patterson maps could be easily solved. The reasons for this difficulty became apparent only later when it was realized that there were no fewer than 12 sites per asymmetric unit in each one of the derivatives. A difference Patterson map for such a derivative contains in excess of 2500 vectors. Typically, the trypsinogen maps contained 800–1400 peaks of sufficient height to warrant investigation. Even with this large number of peaks, the ratio of the highest peaks on the map to the origin peak encouraged attempts to find at least the major sites by deconvolution of the Patterson map. Therefore, vector superposition methods were applied in attempts to solve the difference syntheses. In no case was an unambiguous solution to any derivative found. The primary source of difficulty was that a large number of cross vectors between nonequivalent "A–B" atom sites mimic the symmetry of "A–A" type vectors between symmetry-equivalent atoms around the threefold axes at $u = 1 - v = \frac{1}{3}, \frac{2}{3}$. As a result, they have double weight and so frustrate attempts at solution.

Solution of the Derivatives. At this stage, heavy-atom derivatives of the isomorphous low pH crystals of DIP-trypsin were investigated. A silver ion derivative was prepared, since a silver derivative of the neutral pH orthorhombic DIP-trypsin had previously shown a simple substitution pattern (Chambers et al., 1974). The trigonal trypsin derivative was unambiguously solved by the superposition programs and contained only two major sites and five minor sites. Since this form of trypsin is highly isomorphous with trypsinogen, an initial set of single isomorphous phases was obtained and they were used directly to determine the heavy atom positions in the trypsinogen derivatives.

Difference Electron Density Maps and Refinement. Three-dimensional single isomorphous replacement phases derived from the major Ag^+ site in the isomorphous DIP-trypsin derivative were used first to locate the major sites in the HgI_4^{2-} and mersalyl derivatives of trypsinogen. The maps were computed using $\Delta F = |F_{PH}| - |F_P|$ for each derivative, and phases from the Ag^+ (trypsin) site structure factors. Three HgI_4^{2-} sites and one mersalyl site were found using this procedure. Heavy-atom parameters were refined using a minicomputer (Data General Corp., NOVA 800), by the method of Dickerson et al. (1961, 1968) in alternating cycles of phase calculation and refinement of parameters. The positional parameters and occupancies of the two trypsinogen derivatives were refined along with those of the Ag^+ trypsin derivative. At a later stage of the refinement, the Ag^+ trypsin derivative was discarded to prevent bias in the phases towards a trypsin-like structure.

Phases from these derivatives were used to find the primary sites in the other trypsinogen derivatives. Secondary sites were located using a double-difference or error synthesis, which uses as structure factors the lack-of-closure error ϵ computed during phase refinement, with the calculated phases of the derivative data

$$(|F_{PH}| - |F_P + f_H|) \exp(i\varphi_{PH})$$

(Blake et al., 1963; Stroud et al., 1974), where f_H is the current value for the calculated heavy atom structure factor. In most derivatives, several minor sites were not included in the phasing calculations as their contributions were below the errors in measurement. Finally, anisotropic temperature factors for the sites of higher occupancy were included in the refinement. The correctness of each derivative solution was checked by permutation of pairs of derivatives which were corefined together, to generate phases which were then used to calculate difference maps for the other derivatives. Summary statistics for the six derivatives used are listed in Table I.

Electron-Density Maps. An electron density map was generated using multiple isomorphous replacement phases (MIR map) obtained from the derivatives listed in Table I. The first map was calculated using (hkl) data to a minimum interplanar spacing of 5 Å (figure of merit, FM = 0.88), and revealed the location and conformation of each molecule in the unit cell. The second MIR map was computed using data to 2.6 Å, and a model was built from Kendrew–Watson components to match the density map using an optical comparator (Richards, 1968). The mean figure of merit for all 5700 reflections with $F_{hkl} > 3\sigma$ from ∞ to 2.6 Å was FM = 0.71 (0.91 for centric reflections) when calculated according to Dickerson et al. (1968). A histogram shown in Figure 2 indicates the distribution of reflections according to their figures of merit.

The map was easily interpreted in most areas, although difficulties were encountered in the regions between residues 10–15, 143–151, and 186–188, all of which lie on the surface of the molecule.

Coordinates of all the reliably determined amino acid residues in trypsinogen were measured using an automated coordinate-measuring device which operates on the principle used by Salemme and Fehr (1972). Amplitudes F_{calcd} and phases φ_{calcd} were calculated from the measured coordinates, and a difference map was computed using terms:

$$\Delta F = |F_o| - |F_c|, \varphi_c$$

The well-determined parts of the molecule (90% of the structure) were adjusted using an automated procedure for constrained difference Fourier refinement and model building, which was developed to operate entirely on a minicomputer (Chambers and Stroud, 1976). Individual temperature factors were assigned to each heavy atom or ion (C, N, O, S, Ca^{2+}) based on electron density at atomic centers. No solvent molecules were included. Structure factors were calculated a second time and the procedure was repeated once more. At this stage, hkl data to 2.1 Å resolution were incorporated and an electron density map was calculated using terms $|2F_o - F_c|, \varphi_c$. Areas of structure which were ambiguous in the 2.6 Å MIR map were rebuilt using this map. Data were included to 1.9 Å resolution, and the molecule was subjected to several cycles of difference Fourier refinement. Isotropic temperature factors were refined so as to account for electron density at atomic centers in the

TABLE I: Statistical Analysis of Derivatives.

Derivative	No. of Sites	Occupancy[a]	$\langle\lvert\Delta F\rvert\rangle$[d]	ϵ[b]	R_c[c]	Resolution (Å)	No. of Reflections
HgI_4^{2-}	25	62 61 49 43	99 (242-81)	68 (93-61)	0.67 (0.37-0.84)	2.6	5450
$Pt(NO_2)_4^{2-}$	14	108 78 61 54	91 (211-64)	64 (87-53)	0.64 (0.35-0.91)	2.6	4213
Mersalyl 1	14	74 53 52 31	103 (195-84)	57 (87-55)	0.57 (0.41-0.65)	3.6	4908
Mersalyl 2	9	65 42 32 23	82 (130-70)	44 (57-42)	0.57 (0.43-0.61)	3.6	2184
Mersalyl 3	13	59 39 38 36	110 (217-98)	78 (74-79)	0.70 (0.33-0.72)	2.6	5043
Ag^+	15	76 49 45 22	98 (180-74)	61 (75-53)	0.60 (0.42-0.72)	2.6	5102

[a] Occupancy of the four most major sites. Scale is arbitrary (approximately 1.4 × absolute scale). [b] $\epsilon = \langle\lvert F_{PH}\rvert - \lvert F_P + f_H\rvert\rangle$ for all reflections. [c] $R_c = (\sum_{h0l}\epsilon_{hkl})/(\sum_{h0l}\Delta F_{hkl})$. [d] Numbers in parentheses give the average values of parameters at the lowest and highest values of the scattering angle 2θ used.

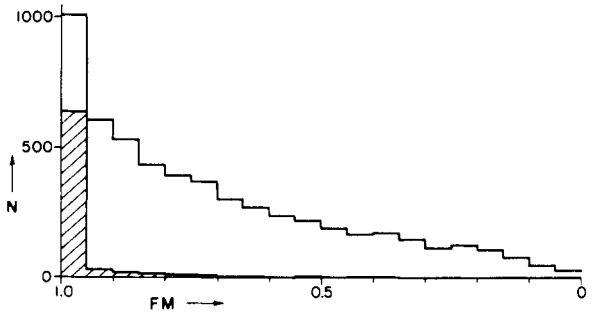

FIGURE 2: Histogram showing distribution of figures of merit for reflections. Histogram for the centrosymmetric reflections alone is shaded.

current $|F_o - F_c|$, φ_c synthesis. Therefore, large temperature factors for individual atoms give a measure of the possible error, or thermal motion of each group. The R_F value at this stage was 31% for all reflections from ∞ to 1.9 Å, where

$$R_F = \frac{\sum|F_o - F_c|}{\sum|F_o|}$$

While most of the molecule is well defined in the electron density (coordinates are probably defined to within about ±0.4 Å), there are small regions between residues 10–14, 143–150, and 186–194 where there is clearly a large degree of libration. These are external and solvent-accessible chains in the molecule. The electron density in the $2F_o - F_c$ synthesis is diffused and low in comparison with other regions, both when these residues were left out of the structure factor calculation and when included in the calculation. In addition, a density map for isomorphous trypsin, computed using trypsinogen phases, with amplitudes for trigonal trypsin was clearly interpretable for every residue in the region of residues 143–150 (10–14 is

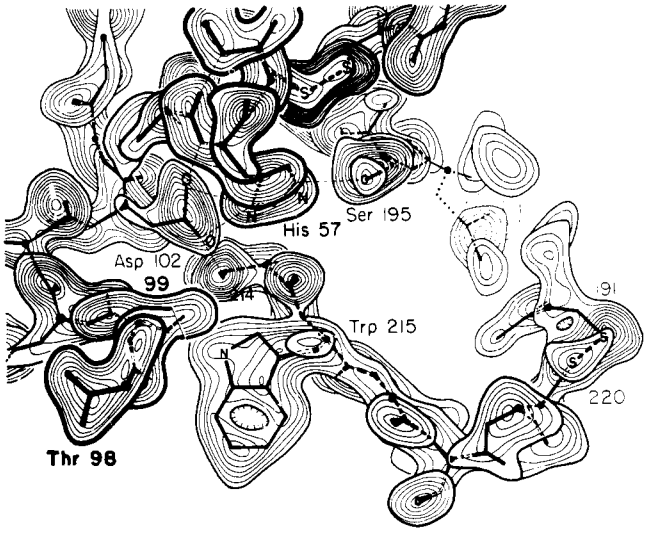

FIGURE 3: The structure around Asp-102, His-57, and Ser-195 in trypsinogen is shown superposed onto the electron-density map at 1.9 Å resolution.

no longer present), and at Ile-16–Val-17. Thus, there is clearly and unequivocally variability in the orientations of the trypsinogen residues 10–14, 143–150, and 186–194.

A region of the 1.9 Å electron density map around the active center is shown in Figure 3. The structure in this region is superposed on the density map. Most the carbonyl groups are clearly identified in the map and six-membered rings often have dimpled density in their centers, as, for example, does the side chain of Trp-215. The sulfur atoms in disulfide bridges are clearly resolved from each other. Some 10–20 ordered solvent molecules are contained within the molecular boundary, and, in cases where they exist as isolated molecules, these

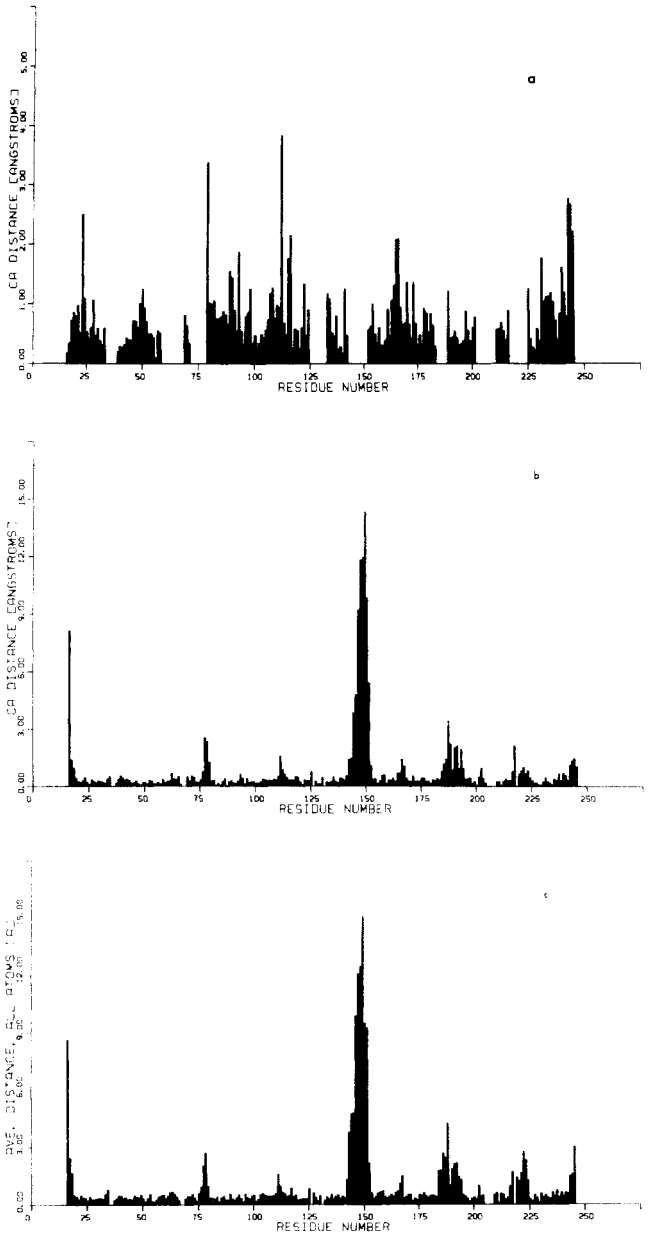

FIGURE 4: (a) Histogram showing the difference (Δd) between α-carbon positions in DIPT and CT. (b) Histogram showing the difference (Δd) between α-carbon positions of the Tg and DIPT structures after suitable rotation. Large differences in structure are seen around residues 16–19, 142–152, 186–194, and 215–220. (c) Histogram showing the mean difference (Δd) in position for all atoms in each residue between Tg and DIPT.

can usually be resolved. Solvent in the cavity which becomes occupied by Ile-16 upon activation was generally not resolved into discrete peaks.

Results and Discussion

The coordinates for trypsinogen are regarded as only preliminarily refined (R_F = 31%). Typically, wire-model coordinates of a protein give rise to R_F values of ~40–45%, and can often contain occasional large errors of several angstroms. As refinement continues, the R_F value decreases and a value of <25% for a constrained structure often seems to represent convergence onto a respectable set of coordinates and isotropic temperature factors which account for thermal motion. Approximately, ten well-ordered solvent molecules were included in the later phase calculations. The model is clearly accurate for nearly all of the structure to within about 0.4 Å.

Comparison of Structures. The trypsinogen structure (Tg) was compared both with the orthorhombic DIP-trypsin structure (DIPT) and with the trigonal low-pH DIP-trypsin structure (LpT). The trypsinogen structure will be described in reference to the DIP-trypsin structure. The Tg and DIPT structures were determined and refined completely independently. The differences in structure between Tg and DIPT will, at the same time, be compared with the differences between the structures of chymotrypsinogen (Cg; Freer et al., 1970) and tosyl-α-chymotrypsin (CT; Birktoft and Blow, 1972).

Current Status of Trypsin Studies. The DIPT structure determined initially by Stroud et al. (1971, 1974) has since been refined using an automated procedure for constrained difference Fourier refinement (Chambers and Stroud, 1976). The agreement index was R = 23.5% for all reflections observable to a resolution of 1.5 Å at the time this comparison was made. The crystals were grown from 5.8% $MgSO_4$ solutions at about pH 7.5. X-ray data are observable to 1.1 Å resolution and, so far, recorded to 1.5 Å (22 117 observable reflections out of 35 566 independent reflections). There are well-defined and small conformational changes between this structure and that of benzamidine-trypsin (Krieger et al., 1974b). These changes have been shown by recomputing the maps with more recent phases not to be an artificial feature of the difference map, as suggested by Bode and Schwager (1975a), based on a comparison of their native trypsin structure with the 1.8 Å refined structure of benzamidine-inhibited trypsin.

The orthorhombic DIP-trypsin structure (Chambers and Stroud, 1976) is now among the most highly refined of any enzyme structure (R = 23.5% to 1.5 Å resolution), as is the more recently determined structure for benzamidine-trypsin of Bode and Schwager (1975a) (R = 22.9% to 1.8 Å resolution). No comparison between these two independently refined trypsin structures has yet been made. The comparison of Huber et al. (1974) was made between their refined structure and our initial unrefined wire-model coordinates for DIP-trypsin derived from the first 2.7 Å MIR density map. Wire-model coordinates such as these are inadequate for detailed comparison, even when the map is of the highest quality (Chambers and Stroud, 1976).

A comparison between our trypsin structure and α-chymotrypsin (Birktoft and Blow, 1972), similar to that made by Huber et al. (1974) between their trypsin structure and α-chymotrypsin, was carried out using a rigid-body least-squares computer program to superpose the two structures. Atoms compared in this manner included all α-carbon atoms, except in regions where amino acid deletions or insertions occur. The residue numbers included were thus the same as those used by Huber et al. (1974).

The average positional difference for α-carbon atoms of these residues was 0.80 Å, a value similar to that of 0.75 Å obtained by Huber et al. (1974) in their comparison of all main chain atoms for these residues (see Figure 4a). The similarity between this result and that of Huber et al. (1974) suggests that the DIP-trypsin refinement has resolved the discrepancies reported by those authors between the 2.7 Å unrefined DIP-trypsin structure and their 1.9 Å refined trypsin-PTI structure.

The LpT structure crystallizes preferentially below pH 5.5 and the crystals are isomorphous with those of Tg. This structure was compared directly with Tg by computing difference maps with amplitudes $\Delta F = |F_{Tg}| - |F_{LpT}|$, and phases for Tg to a resolution of 2.1 Å. This difference map is

a more sensitive test of small conformational changes than comparison of coordinates, and is correspondingly less prone to interpretative error. Such a map can clearly show changes which are smaller than the estimated error in the current set of atomic positions. Thus, "identical position" in the comparison of atomic positions between Tg and LpT probably implies identity within ±0.1 Å (which for a single carbon atom would result in a pair of peaks of 5σ above the noise level in the map), whereas the same phrase applied to the Tg/DIPT comparison implies equivalence to within about ±0.3 Å.

It must be noted that there is a strong possibility that the low-pH form of DIP-trypsin has a different structure from the neutral-pH form, as do low- and neutral-pH α-chymotrypsin (Tulinsky et al., 1973; Mavridis et al., 1974). Further, there is the possibility that the isomorphic low-pH trypsin structure is more like trypsinogen in some ways, and thus the difference map procedure would fail to identify any difference between the two proteins at such sites. In fact, the difference map is very clean in many areas where the neutral pH structures appear to be slightly different. However, the *activity* of the enzyme depends on only a single ionization between pH 2 and 8, and this has been identified as the ionization of Asp-102 at the active site (Koeppe and Stroud, 1976). Therefore, any pH-dependent structure change which does occur between pH 7 and 5 in trypsin is not a change which is reflected in the inherent activity of the enzyme, unless it can be associated directly with ionization of Asp-102. Thus, the changes between the isomorphous structures as seen in the difference Fourier are more likely to pick out the important differences for the change in catalytic activity in a direct way. This adds emphasis to this comparison as an assay for the important catalytic change.

Similarities in Structure. The atomic coordinates for DIPT were rotated to match those of Tg by least-squares minimization of the distances between equivalent atoms in the two molecules. Regions of major change in structure, i.e., residues 16–19, 143–152, 186–195, and 215–220, were left out of the rotation calculation, as were residues 76–80 where the differences are most probably due to errors in the model (see below).

The refined equation, which described rotation of the DIPT coordinates x, y, z (listed in angstroms with respect to the a, b, c axes) to match the trypsinogen set $(x', y', z')_{Tg}$ (listed in angstroms with respect to an orthogonal axial set, such that x' is perpendicular to the crystallographic b, c axes; y' and z' are parallel to the b and c axes, respectively), is:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} -0.0017 & 0.9724 & -0.2333 \\ 1.0000 & 0.0028 & 0.0046 \\ 0.0052 & -0.2333 & -0.9724 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 1.82 \\ -11.06 \\ -11.47 \end{pmatrix} \quad (1)$$

This rotation matrix closely approximates the form of a simple rotation of 13.5° about 2_1 axes in the two structures

$$\sim \begin{pmatrix} 0 & 1 & -0.23 \\ 1 & 0 & 0 \\ 0 & -0.23 & -1 \end{pmatrix}$$

It therefore explains the formation of the orthorhombic-trigonal twinned crystals and the congruence of the two a axial dimensions which are within 0.7% of each other (see Figure 1, and Experimental Section).

The histogram shown in Figure 4b indicates the distance between all α carbons for residues common to both structures after rotation, and clearly identifies the regions of greatest structural change. The mean deviation between equivalent α-carbon atoms in the two structures was calculated by leaving out the 32 residues with obviously different structure, and was $\Delta d_{C\alpha} = 0.38$ Å. Thus, the difference in conformation of the main chain for all these residues is less than for the α carbons of the best region (the C-terminal α helix) of the Cg/CT comparison (Freer et al., 1970).

The mean deviation of position between all atoms in the main chain and side groups for the same residues was $\Delta d =$ 0.45 Å (Figure 4c). Since the two structures are so similar in these regions, these distances presumably represent the sum of errors in the two coordinate sets, coupled with real differences in structure. The errors in the trypsinogen coordinates are clearly much greater than those of the more refined DIPT set. Real differences in structure could be due to (1) crystal packing forces, (2) different solvent or counterion effects, and (3) to genuine consequences of the difference in covalent structure of Tg and DIPT. Most of the differences in side-chain orientation for structurally homologous regions do occur on the surface of the protein.

It is notable that DIPT and Tg are much more homologous in structure than are CT and Cg. In a recent analysis, Birktoft et al. (1976) identified several regions of large difference between CT and Cg, associated with a relative movement of the two structural cylinders which are formed by the first and second halves of the chymotrypsin sequence. There is no corresponding difference between DIPT and Tg. However, it must be noted that the CT structure was determined at low pH, ~4.5, where the enzyme has different structure from the neutral-pH form (Tulinsky et al., 1973), and crystallized as a dimer, where the active center lay in the dimer interface.

The Active Center. The structure around the active center of Tg is shown in Figure 5a. Crystallographic studies of chymotrypsinogen (Freer et al., 1970) showed that the orientation and immediate environment of the active-site residues, Asp-102, His-57, and Ser-195, were very similar to that found in chymotrypsin. However, a small change in the orientation of Ile-99 permitted limited access of Asp-102 and His-57 to solvent in CT, and the hydroxyl of Ser-214 was not hydrogen bonded to Asp-102 in Cg, raising a question as to the importance of subtle changes in this region for zymogen activation. Similar features are not seen in Tg, where both Leu-99 ($\overline{\Delta d}_{99} = 0.5$ Å) and the Oγ of Ser-214 ($\Delta d = 0.2$ Å) are in identical positions in the zymogen and enzyme; therefore, they cannot be common features of the mechanism of activation.

The α carbon of Ser-195 differs by only 0.4 Å between Tg and DIPT, and the Oγ of Ser-195 in Tg is hydrogen bonded to the εN of His-57 (Figure 5a) as it is in the benzamidine-trypsin structure (Kreiger et al., 1974b). The Tg–LpT difference map also indicates a small shift in the orientation of the imidazole of His-57 between the two structures. The presence of a DIP group at Ser-195 may be, in part, responsible for the observed shift of 0.3 Å in the ring. This movement of the imidazole ring is similar to the shift observed previously in a comparison of benzamidine-trypsin and DIP-trypsin (Krieger et al., 1974b). It is also clear from the latter comparison, and from more recently computed difference maps using improved phases, that there are several other small differences between these two structures. The fact that Bode and

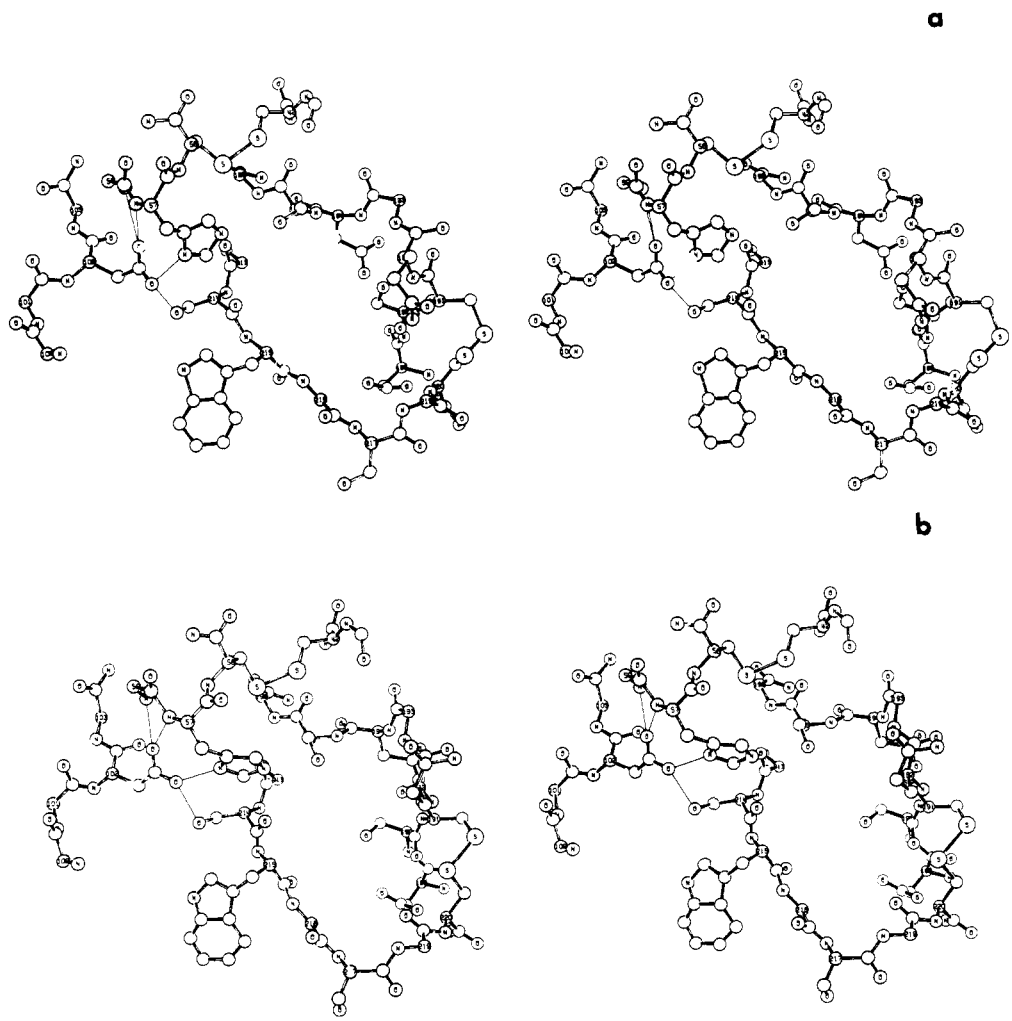

FIGURE 5: Structure of the active site in trypsinogen (a) is compared with structure for the same region of trypsin (b).

Schwager (1975a) did not find any such differences between benzamidine-trypsin and native trypsin suggests that the presence of the covalently bonded DIP group may be responsible for small differences in the DIP structure. Nevertheless, no significant movement of the α carbon of Ser-195 was observed in DIP-trypsin when compared with benzamidine-trypsin.

Birktoft et al. (1976) found that the hydrogen bond between the $N\epsilon_2$ of His-57 and the $O\gamma$ of Ser-195 was strained in CT, while it was relatively normal in Cg. Our present results with the trypsin system differ in that the $O\gamma$ of Ser-195 (even though attached to the DIP group) in DIP-trypsin is still only 1.15 Å from the ideal position (described by Birktoft et al., 1976) and only 0.67 Å from the ideal position in trypsinogen. This is to be compared with the corresponding values of about 2.5 Å for α-chymotrypsin and 0.7 Å for chymotrypsinogen.

Since the hydrogen-bond distances between His-57 and Ser-195 in trypsinogen are normal (2.5 Å) and can be made normal in trypsin without the DIP group (2.8 Å in DIP-trypsin), it does not seem that significant distortion at this site can be considered as a common feature of the activation mechanism of serine proteases.

The Side-Chain Binding Pocket. Freer et al. (1970) questioned whether the absence of a specific binding pocket could explain the inactivity of Cg. Trypsinogen has a large and accessible binding pocket, although the structure and stability of the cavity is different from that found in DIPT (or in benzamidine-trypsin (Krieger et al., 1974b)).

The region of chain between 186 and 194 is loosely organized and adopts several different conformations in Tg. The present structure for this chain was determined carefully. Evidence for this structure as the most predominant conformation depended on (1) the isomorphous-replacement density map; (2) $2F_o - F_c$ or $F_o - F_c$ electron-density syntheses using calculated phases for a model from which this region was omitted; (3) the LpT–Tg difference map, which gives the most direct measure of the loose organization of the 186–194 chain. This difference map shows unequivocally that the chain structure is variable and that the low density is not due to errors in the phases.

Two hydrogen bonds present in DIPT, between Val-17 NH and the main-chain carbonyl of Asp-189, and between the Val-17 C=O and Asp-189 NH, are absent in Tg. These may contribute relative stability to the chain in DIPT.

Benzamidine is a competitive inhibitor of trypsin (K_i = 1.8 $\times 10^{-5}$ M) (Mares-Guia and Shaw, 1965) which binds inside the side-chain specific binding pocket of the enzyme (Krieger et al., 1974b). Benzamidine was present at 20 mM concentration (at least 10^3 times the K_i for trypsin) in the trypsinogen-crystallizing solution; yet, there was no benzamidine bound to the proenzyme in the Tg structure. This emphasizes the fact that the structure for the binding site, as seen in the crystals, represents a functionally significant impediment to the binding capabilities of the site for benzamidine, which is an analogue

of the side chain of a substrate for trypsin. The binding constant for benzamidine binding to trypsinogen in solution was found to be about 10^5 times larger than for binding to trypsin (Gertler et al., 1974). The binding site seen in the Tg structure is large enough to accept a benzamidine molecule, and the fact that none was bound implies that the differences between zymogen and enzyme structures (as shown in Figure 5) are quite adequate to account for a large change in specificity.

A relatively large movement of the backbone chain from residues 187 to 194 in the Cg/CT comparison was correlated with the formation of the specific binding site for substrate side chains in CT, a cavity which was only partially formed in Cg (Freer et al., 1970). There are much smaller but significant changes between Tg and DIPT (or LpT) in the chain from residues 188A to 195 (Figure 5). The migration of Met-192 from a completely buried position (in the site occupied by the ion pair between Ile-16 and Asp-194) in Cg to the outside of the CT molecule also has no comparable counterpart in the Tg/LpT comparison. The α carbon of Gln-192 in Tg moves by ~1 Å on activation. The side chain of Asp-189 inside the specific binding pocket of trypsin and responsible for the substrate specificity of trypsin lies within ~1 Å of the position found in Tg, $\Delta d_{C\alpha 189} = 0.9$ Å. Thus, it would seem that modification, rather than generation of the specific cavity, must be considered as a possible reason for zymogen inactivity in general.

Studies of the residual activity of trypsinogen have generally been carried out on simple esters where hydrogen bonds to the main chain of 214–219 are not made. For NPGB hydrolysis by Tg, the binding constant was found to be at least 10^3 times worse than for trypsin (Robinson et al., 1973). Changes in the binding-pocket structure could well be responsible for some or all of this poorer binding.

Residues 214–220. In Cg, changes occurred in the main chain between Ser-214 and Cys-220, a region known to be involved in hydrogen bonding to a peptide substrate (Segal et al., 1971; Sweet et al., 1974; Huber et al., 1974; Blow, 1974b). In Tg, the chain structure differs slightly from trypsin as shown in Figure 5. The C=O of Ser-214 ($\Delta d = 0.3$ Å), and the NH ($\Delta d = 0.5$ Å) and C=O ($\Delta d = 1.4$ Å) of Gly-216 still point in the same direction. The peptide bond between Gly-219 and Cys-220 is apparently rotated by 180° with respect to trypsin at neutral pH (DIPT). There is a difference in residues 214–217 of Cg in that the NH and C=O of Gly-216 would be unable to participate in substrate binding (Freer et al., 1970). This and the other differences in binding-site structure between Cg and Tg are not inconsistent with the fact that the basic pancreatic trypsin inhibitor (BPTI) will form a stoichiometric 1:1 complex with trypsinogen and with trypsin or chymotrypsin, but will not so complex with chymotrypsinogen (Dlouha and Keil, 1969).

There was an error in the original placement of the side chain of Trp-215 in DIPT (Figure 8 of Krieger et al., 1974b). Refinement of that structure revealed that the indole ring is reversed, such that ϵN points in toward the center of the molecule (Chambers and Stroud, 1976). The same configuration is found in Tg. This orientation is now the same as that found in CT (Birktoft and Blow, 1972). Modification of Trp-215 in trypsinogen blocks the binding of BPTI by trypsinogen (Keil-Dlouha and Keil, 1972).

Asp-194. The side chain of Asp-194 forms a salt bridge with Ile-16 in trypsin. The N terminal of Ile-16 is not available in Tg, since it lies in the zymogen-activation peptide. Asp-194 is in a similar orientation inside the trypsinogen molecule, although the α carbon is moved by $\Delta d = 0.9$ Å. The side chain is surrounded by internal solvent molecules. There appears to be only one intramolecular hydrogen bond to Asp-194 in Tg between the Oδ-1 of Asp-194 and the Oγ of Ser-190. The side chain of Asn-143 (in Tg) points in toward the solvent cavity around Asp-194.

Freer et al. (1970) postulated that interaction between Arg-145 and Asp-194 could contribute to charge stabilization of the carboxyl group of Asp-194 in Cg—possibly through water molecules. In a more detailed analysis, Wright (1973a,b) found that the guanidine group of Arg-145 was too far from Asp-194 to participate directly in this role, but could neutralize the charged carbonyl via water molecules hydrogen bonded to the δ nitrogen of His-40. The ϵ nitrogen of His-40 was found to be 3.5 Å from the carboxyl group of Asp-194, and so was indirectly implicated as providing a possible pathway for proton transfer from the surface of the molecule to the buried carboxyl group.

In Tg, neither Lys-145 nor His-40 lies close to the side chain of Asp-194 (see Figure 6), and, therefore, cannot be involved in stabilizing Asp-194. The side chain of Lys-145 is on the outside of the molecule and the ζN of Lys-145 is about 10 Å from Asp-194. Indeed, there are no basic side chains close enough to contribute directly to charge neutralization of Asp-194. His-40 is found in the same orientation in Tg as in LpT, and almost exactly the same as in DIPT ($\Delta d_{C\alpha} = 0.4$ Å). The Nδ-1 of His-40 is hydrogen bonded to the Oγ of Ser-32 in both Tg and DIPT (and CT, but not Cg). Nϵ-2 of His-40 is hydrogen bonded to the C=O of Gly-193 in DIPT and CT, but is hydrogen bonded to an external solvent molecule in Tg. This change corresponds to the movement of the Cα of 194 (Δd = 0.9 Å) and 193 ($\Delta d = 2.0$ Å) toward the solvent-filled cavity around Asp-194. Wright (1973b) emphasized the apparent conservation of a basic group at residue 145 of the trypsin-like enzymes, although there are now exceptions in the sequences of dogfish trypsin (where the residue is methionine), and in all the bacterial enzymes (Bradshaw et al., 1970; De Haen et al., 1975). His-40 is also not a conserved residue among trypsin-like enzymes, and is Leu in bovine thrombin, Gly in blood-clotting factor X, and Met, Arg, or Leu in bacterial trypsin sequences (De Haen et al., 1975).

The "Autolysis Loop" (Residues 142–151). The largest conformational difference between Tg and DIPT or LpT occurs between residues 142 and 151. This sequence was referred to as the "autolysis loop" by Blow et al. (1969), since it contains a region where two amino acids are excised in α-chymotrypsin. The difference map between Tg and LpT (Tg–LpT) showed clear negative density, which picked out the conformation of the autolysis loop in LpT. There was much less (~50%) positive density in the map for this part of the chain, establishing that the chain is flexible in Tg and adopts a quite different conformation there. (The LpT structure was identical to that of DIPT in this region.) A further change in the structure of this loop or of Lys-15–Ile-16 would be required if an activating enzyme was to approach the Lys-15–Ile-16 bond in the Tg conformation.

Trp-141 is in essentially the same orientation in Tg as found in DIPT ($\Delta d_{C\alpha} = 0.6$ Å). The side chain of Asn-143, which was solvent accessible in DIPT, points in toward the cavity around Asp-194 in Tg ($\Delta d_{C\alpha} = 1.4$ Å). The side chain of Thr-144 (Tg) is moved by several angstroms toward the disulfide Cys-191–Cys-220 ($\Delta d_{C\alpha} \simeq 4.0$ Å), and residues 145–151 are in quite different conformation in the two structures. They are all solvent accessible in both Tg and DIPT. There may be a functional correlation between the loose structure of this chain in the zymogen, and the necessity for

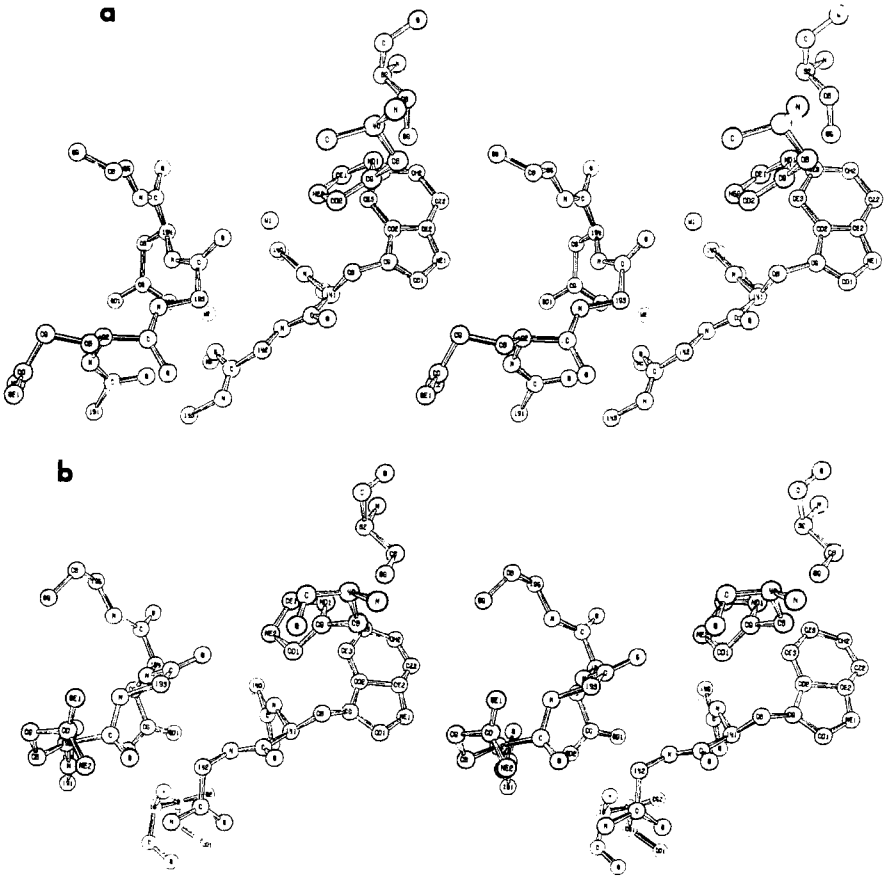

FIGURE 6: (a) Structure around Asp-194 and His-40 in Tg. (b) Structure around Asp-194 and His-40 in DIPT.

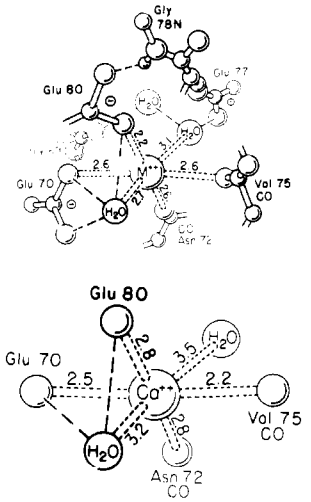

FIGURE 7: A schematic of the coordination of ions at the primary Ca^{2+} ion binding site in trypsin (top) and trypsinogen (bottom). Distances between the ion and chelating ligands are indicated in angstroms.

access of the activating enzyme to the Lys-15–Ile-16 bond. The sequences for residues 146–154 are extremely variable in related enzymes (De Haen et al., 1975), and no specific roles can be assigned to any of the side chains of these residues in trypsinogen, as judged from the Tg structure.

The Primary Calcium Ion Binding Site. There are two calcium ion binding sites in trypsinogen. The primary site, with higher affinity for calcium ion ($K_D \sim 10^{-3.2}$ M, Abita et al., 1969), is common to both trypsinogen and trypsin. Occupancy of this site stabilizes the protein toward thermal denaturation or autolysis (Northrop et al., 1948). Figure 7 shows a schematic of the primary Ca^{2+} ion site in trypsinogen and in trypsin (Chambers and Stroud, 1976). Coordination of the ion in both models is approximately octahedral and the ligands are identical in the two structures, although there are differences of up to 0.7 Å in the absolute positions of the ligands involved ($\Delta d_{Ca^{2+}} = 0.15$ Å).

It seems likely that these differences may be due to error rather than to any real structure change, since there were no peaks in the Tg–LpT difference map around the ion. The peak of electron density assigned to the Ca^{2+} ion in trypsinogen was the highest peak on the entire Fourier map, even during the first stages of Fourier refinement at 2.1 Å resolution and site occupancy refined to 18 electrons.

The primary Ca^{2+} binding site in trypsin was first identified and described in detail by Bode and Schwager (1975b) based on analysis of an independently refined trypsin structure seen at 1.8 Å and the Ca^{2+} content of their crystals. No comparison of their coordinates with ours has been made, although the structures for the Ca^{2+} site in all three determinations seem to be essentially the same. In our case, no Ca^{2+} was added to any of the crystallizing solutions and no attempt has been made to determine the Ca^{2+} content in the crystals.

The ion found at the primary site in DIPT is probably a magnesium ion present in the crystallizing liquor (5.8% $MgSO_4$). The electron density at this site refined to that expected for a water molecule (10 electrons) or a magnesium ion (Chambers and Stroud, 1976). Since Ca^{2+} accelerates autoactivation of trypsinogen to trypsin, reasonable care was taken to exclude it from the trypsinogen-crystallizing solution. Ca^{2+} was present in the buffer (0.02 M in $CaCl_2$) used during the chromatographic purification, although the solution was

later dialyzed using large volumes of 10^{-3} M HCl. Thus, it was surprising to find that the bound ion at the primary ion site in Tg refined to an occupancy of 18 electrons. We presume this to be a single site, fully occupied by a single, tightly-bound, Ca^{2+} ion, which remained as an integral component of the zymogen throughout.

The difference map calculated between the isomorphous crystals of Tg and LpT showed no electron density in the Ca^{2+} region. Therefore, the primary calcium ion binding sites in trypsinogen, low-pH trypsin, or DIP-trypsin at about pH 7.5, are similar to one another and to the structure described by Bode and Schwager (1975b) in benzamidine-trypsin.

The Second Ca^{2+} Binding Site and the N-terminal Sequence. A secondary Ca^{2+} ion binding site of $K_D \sim 10^{-1.8}$ M exists only on the zymogen. Binding of Ca^{2+} at this site is essential for complete and efficient activation of trypsinogen (Northrop et al., 1948). The effect of binding is to improve the substrate character of the activation peptide, so, favoring tryptic hydrolysis of the Lys-15–Ile-16 peptide bond. (K_m decreases by a factor of three, while k_{cat} remains unchanged (Abita et al., 1969).) Tryptic hydrolyses of the nonapeptide Val-$(Asp)_4$-Lys-Ile-Val-Gly and of the hexapeptide Val-$(Asp)_2$-Lys-Ile-Val have been shown to be Ca^{2+} dependent in a very similar way, while hydrolyses of peptides containing only one aspartate do not depend on Ca^{2+} (Delaage et al., 1967; Abita et al., 1969). Thus, Ca^{2+} ion binding to Asp-13 and Asp-14 together is sufficient to explain the zymogen Ca^{2+} ion effect (Abita et al., 1969). There is no corroborative crystallographic evidence as yet, since Ca^{2+} was excluded from the crystallizing solution.

The overall weak electron density for much of the N-terminal hexapeptide of trypsinogen must be attributed to local disorder or loosely organized structure, rather than to phasing error. This conclusion is implied by the (Tg–LpT) difference map computed using phases for trypsinogen. In regions where there is difference in structure, this map should show a positive-density image for trypsinogen, and a somewhat weaker negative image for trypsin. Yet, the positive density for trypsinogen residues 10–15 was consistently much lower in amplitude than the negative images for trypsin structure.

The loose arrangement of this hexapeptide on the surface of the proenzyme is consistent with the finding that the synthetic nonapeptide Val-$(Asp)_4$-Lys-Ile-Val-Gly is hydrolyzed by trypsin at almost the same rate ($k_{cat} \sim 10^{-3}$ s^{-1}) as the Lys–Ile bond on the same sequence in Tg for which $k_{cat} = 2.5 \times 10^{-3}$ s^{-1} (Abita et al., 1969). Thus, it is clear that the slow hydrolysis of the Lys–Ile bond by trypsin does not depend on any unique tertiary structure of these residues. The four aspartates are all available to modification by carbodiimides, and such modification does not destroy activatibility of trypsinogen (Radhakrishnan et al., 1967). This is in agreement with our finding that the Val-10–Asp-14 sequence is not a determinant of tertiary structure. Thus, the binding of Ca^{2+} to Asp-13 and Asp-14 is probably independent of the protein structure as an ion-specific conformer, and is adequately explained by an induced structural change in the peptide (10–15) alone, or by neutralization of charge at Asp-13 and Asp-14.

There is a clear necessity for trypsinogen among all pancreatic zymogens to develop immunity to autolysis or activation by its own residual activity, since trypsin activates other zymogens. The loosely structured hexapeptide is presumably protected against trypsin more by its negative charge than by tertiary structure. The Lys-15–Ile-16 is not inaccessible, although some reorientation of this chain or of the autolysis loop must accompany enzyme binding to this bond. The fact that this bond is still the most readily hydrolyzed is consistent with the fact that compact native structures are generally more resistant to degradation than are denatured ones. The φ and ψ angles on either side of Lys-15 are not dissimilar to those found for the susceptible Lys-15 bond on the trypsin inhibitor (a good substrate model). Since it is clearly proven that both this chain and the neighboring autolysis loop have little to restrain their configuration, the reorientation necessary for access by the activating enzyme should be energetically easy to accomplish.

Because of disorder, no real density for the side chains of Asp-11 or Asp-12 exists.

The Reasons for Zymogen Inactivity. Freer et al. (1970) and later Wright (1973b) identified four possible factors which might be responsible for inactivity of chymotrypsinogen. We now see that two of them do not seem to be shared by trypsinogen, since (1) the potential source of a hydrogen bond from Gly-216 NH is not excluded, although this group is moved by 0.5 Å from its "substrate-binding" position in the free enzyme, and (2) the side chain of Gln-192 occupies a fairly innocuous position outside the trypsinogen molecule, whereas the corresponding Met-192 of chymotrypsinogen was shown to undergo a complete reorganization upon activation.

The third factor, an incompletely formed binding site in Cg, does have a counterpart in Tg, although the binding pocket is open and accessible in Tg, whereas it was less so in Cg. Nevertheless, alterations in the binding site are sufficient to alter or remove specificity.

The remaining factor was the lack of an important hydrogen-bonding group in the NH of Gly-193. Steitz et al. (1969), Henderson (1970), and later Robertus et al. (1972) postulated that the NH groups of Gly-193 and of Ser-195 could play an important catalytic role in stabilization of reaction intermediates. In chymotrypsinogen, the NH of Gly-193 points away from the position for the carbonyl oxygen of the substrate (Freer et al., 1970). In trypsinogen, the NH of Gly-193 and of Ser-195 both point toward a site not dissimilar to that found in trypsin and chymotrypsin. Nevertheless, this site, termed "the oxyanion hole" by Robertus et al. (1972), is removed by 1.9 Å away from the position of the site in trypsin. Thus, oxyanion stabilization is possible in trypsinogen, but not at the ideal site. Peaks in the (Tg–LpT) difference map for this region are further testimony to the differences between these two molecules and also indicate a shift of about 2.0 Å. Loss of this component of the catalytic apparatus may well account for the slower reaction rate of the zymogen. There is a plausible connection between the movement of this site away from the cavity around Asp-194 and the formation of the salt bridge to Ile-16 upon activation. There is also plausible connection between the small changes in the binding pocket and the relocation of Ile-16. Similar changes in the structure of the binding site could account for the poorer binding of substrate by the zymogens in general.

Conclusion

The structure of trypsinogen is generally much closer to that of trypsin than in chymotrypsinogen to chymotrypsin. The structure of trypsinogen does not exclude the possibility of substrate binding in a mode similar to that found for trypsin, although changes in the structure of this region contribute to an impaired or altered substrate binding mode—certainly for benzamidine and, most probably, for a substrate side chain. If the proenzyme is considered to be rigid, then the general base catalyst (the ϵN of His-57) and the oxyanion-binding site formed by the NH groups of Gly-193 and Ser-195 are too far

apart to cooperate in substrate hydrolysis. Even if the proenzyme structure was to change on substrate binding, as seems likely, nonproductive binding may provide another important component for zymogen inactivity. Such binding might still leave too great a distance, or an unfavorable interaction between the substrate and elements of the catalytic center. The altered position of the chain between Lys-188A and Ser-195 and the main chain between Trp-215 and Ser-217 could be responsible for competitive, yet nonproductive, substrate binding, as could the NH group of Gly-193, since they are normally involved in orienting the substrate. These possibilities remain as prime candidates in a universal scheme for inactivity of the trypsinogen-like zymogens.

Acknowledgments

We thank M. O. Jones and R. Price for assistance with calculations for the comparisons of structures.

References

Abita, J. P., Delaage, M., Lazdunski, M., and Savrda, J. (1969), *Eur. J. Biochem. 8,* 314–324.

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

Birktoft, J. J., Kraut, J., and Freer, S. T. (1976), *Biochemistry, 16,* 4481.

Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C., and Poljak, R. J. (1963), *Acta Crystallogr. 16,* A77.

Blow, D. M. (1958), *Proc. R. Soc. London, Ser. A 247,* 302–336.

Blow, D. M. (1974a), *Isr. J. Chem. 12,* 483–494.

Blow, D. M. (1974b), *Bayer-Symp. 5,* 473–483.

Blow, D. M., Birktoft, J. J., and Hartley, B. S. (1969), *Nature (London) 221,* 337–340.

Bode, W., and Schwager, P. (1975a), *J. Mol. Biol. 98,* 693–717.

Bode, W., and Schwager, P. (1975b), *FEBS Lett. 56,* 139–143.

Bradshaw, R. A., Neurath, H., Tye, R. W., Walsh, K. A., and Winter, W. P. (1970), *Nature (London) 226,* 237–239.

Chambers, J. L., Christoph, G. G., Krieger, M., Kay, L., and Stroud, R. M. (1974), *Biochem. Biophys. Res. Commun. 59,* 70–74.

Chambers, J. L., and Stroud, R. M. (1976), *Acta Crystallogr.* (in press).

Davie, E. W., and Neurath, H. (1955), *J. Biol. Chem. 212,* 515–529.

De Haen, C., Neurath, H., and Teller, D. C. (1975), *J. Mol. Biol. 92,* 225–259.

Delaage, M., Desnuelle, P., Lazdunski, M., Bricas, E., and Savrda, J. (1967), *Biochem. Biophys. Res. Commun. 29,* 235–240.

Dickerson, R. E., Kendrew, J. C., and Strandberg, B. E. (1961), *Acta Crystallogr. 14,* 1188–1195.

Dickerson, R. E., Weinzierl, J. E., and Palmer, R. A. (1968), *Acta Crystallogr., Sect. B 24,* 997–1003.

Dlouha, V., and Keil, B. (1969), *FEBS Lett. 3,* 137–140.

Freer, S. T., Kraut, J., Robertus, J. D., Wright, H. T., and Xuong, Ng. H. (1970), *Biochemistry 9,* 1997–2009.

Gertler, A., Walsh, K. A., and Neurath, H. (1974), *Biochemistry 13,* 1302–1310.

Hartley, B. S. (1970), *Philos. Trans. R. Soc. London, Ser. B 257,* 77–87.

Henderson, R. (1970), *J. Mol. Biol. 54,* 341–354.

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

Keil-Dlouha, V., and Keil, B. (1972), *J. Mol. Biol. 67,* 495–505.

Koeppe, R. E., II, and Stroud, R. M. (1976), *Biochemistry 15,* 3450–3458.

Krieger, M., Chambers, J. L., Christoph, G. G., Stroud, R. M., and Trus, B. L. (1974a), *Acta Crystallogr. Sect. A 30,* 740–748.

Krieger, M., Kay, L. M., and Stroud, R. M. (1974b), *J. Mol. Biol. 83,* 209–230.

Mares-Guia, M., and Shaw, E. (1965), *J. Biol. Chem. 240,* 1579–1585.

Maroux, S., Baratti, J., and Desnuelle, P. (1971), *J. Biol. Chem. 246,* 5031–5039.

Mavridis, A., Tulinsky, A., Liebman, M. N. (1974), *Biochemistry 13,* 3661–3666.

Mikes, O., Tomasek, V., Holeysovsky, V., and Sorm, F. (1967), *Collect. Czech. Chem. Commun. 32,* 655–677.

Neurath, H. (1975), in Cold Spring Harbor Symposium on Proteases and Biological Control, *1975,* Reich, E., Rifkin, D. B., and Shaw, E., Ed., Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory, pp 51–64.

Neurath, H., and Dixon, G. H. (1957), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 16,* 791–801.

Northrop, J. H., Kunitz, M., and Herriot, R. M. (1948), Crystalline Enzymes, New York, N.Y., Columbia University Press, pp 133–136.

Radhakrishnan, T. M., Walsh, K. A., and Neurath, H. (1967), *J. Am. Chem. Soc. 89,* 3059–3061.

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

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

Robinson, N. C., Neurath, H., and Walsh, K. A. (1973), *Biochemistry 12,* 420–425.

Salemme, F. R., and Fehr, D. G. (1972), *J. Mol. Biol. 70,* 697–700.

Schroeder, D. D., and Shaw, E. (1968), *J. Biol. Chem. 243,* 2943–2949.

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

Sigler, P. B., Blow, D. M., Matthews, B. W., and Henderson, R. (1968), *J. Mol. Biol. 35,* 143–164.

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

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

Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1974), *J. Mol. Biol. 83,* 185–208.

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

Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Mani, N. V., and Wright, L. H. (1973), *Biochemistry 12,* 4185–4192.

Walsh, K. A., and Neurath, H. (1964), *Proc. Natl. Acad. Sci. U.S.A. 52,* 884–889.

Wright, H. T. (1973a), *J. Mol. Biol. 79,* 1–11.

Wright, H. T. (1973b), *J. Mol. Biol. 79,* 13–23.

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