MECHANISMS OF ZYMogen ACTIVATION

Robert M. Stroud, Anthony A. Kossiakoff, and John L. Chambers

Norman W. Church Laboratory of Chemical Biology, California Institute of Technology, Pasadena, California 91125

It is now well recognized that proteolytic enzymes play key roles in the regulation of or control over the action of other proteins (1–3). Such enzymes can be found in all species from bacteria to man and in control of diverse systems, which include hormone production, bacteriophage assembly (4), development, fertilization (5), digestion, defense against invading organisms (6), and tissue repair (7, 8). In most cases the proteolytic enzymes are known to be synthesized as inactive precursor proenzymes, or zymogens (9). They are activated by proteolytic cleavage of a single peptide bond in the proenzyme and so become catalytically active. Further control over the degree of specificity for a target molecule or molecules is determined by the degree of specificity inherent to the enzyme. Further control over the time and location of action is often carried out by protein inhibitors of the requisite specificity.

There are several different classes of proteolytic enzymes, which were classified first according to their susceptibility toward different inhibitors (10) and later according to their homologous amino acid sequences or structures within each class (11). There are many excellent papers and reviews that deal with all aspects of the collected information of the proenzymes of proteases (e.g. 1–3, 9, 11). Therefore, in this review we consider the single question of why the zymogens of the serine proteases—the enzyme class about which most is known—are relatively inactive. We consider this in the light of most recent information about the known structures of two proenzymes, chymotrypsinogen (12) and trypsinogen (13), and their activated counterparts, α-chymotrypsin (14) and trypsin (15).

INACTIVITY OF THE PROENZYME

Several reasons have been proposed for the relative inactivity of these proenzymes. The question became particularly intriguing when it was found that the active site in chymotrypsinogen appeared to be preformed and accessible, much as it is in chymotrypsin (12; see Figures 1 and 2). Finding similar structures at the active

1Present address: Brookhaven National Laboratories, Upton, Long Island, New York 11973.
center would suggest that the ubiquitous basis for the unusual reactivity of the catalytic groups in the enzyme should be present in the proenzyme, that the proenzyme should be catalytically active. Indeed, some reactivity does exist in the proenzyme (9, 16), although the catalytic rates for hydrolysis of simple esters are at least \(10^4-10^7\) times lower than for the enzyme. Attention was focused (12) onto the substrate binding site, which appeared to be incompletely formed in chymotrypsinogen, although in trypsinogen the binding site for substrate side chains is open and quite large enough to accept the side chain of a normal substrate of trypsin (13; see Figures 3 and 4). It is not at all obvious why a trypsin substrate should not bind to trypsinogen in a mode similar to that found for trypsin, although it is clear that substrate binding is impaired.

Birktoft, Kraut & Freer (17), as the result of a closer look at the comparison between active center structure in the chymotrypsinogen-chymotrypsin pair, proposed that the development of a high degree of strain in the hydrogen bond between His 57 and Ser 195\(^2\) in the enzyme (Figure 2) but not in the proenzyme (Figure 1) may be a further contributor to the reactivity of the enzyme. Yet in the trypsin-trypsinogen comparison, such differences are certainly much less apparent if they are present at all (13; see Figures 3 and 4). There are two hydrogen bond donors in the enzyme that may stabilize the high-energy transition states during substrate hydrolysis. These are the N–H groups of Gly 193 and Ser 195 (18). The N–H group of Gly 193 is not available for this role in chymotrypsinogen, whereas in trypsinogen the N–H of Gly 193 still points in the same direction as in trypsin but is moved by 1.9\(\AA\) from its site in the enzyme.

\(^2\)Numbering scheme used throughout the review is that of chymotrypsinogen.
Together, it seems that the contribution of correct substrate binding and the ability to stabilize the intermediates in catalysis are implicated most strongly as prerequisites for the enzyme to work efficiently. This leaves open the question of how much, if any, small changes in or around the groups that participate most directly in covalent catalysis contribute to the difference in reactivity of proenzyme and enzyme.

Since there is no clear answer to the question of the basis for such a significant change in the reactivity of proenzyme upon activation (a change crucial to the function of the many related serine proteases in a general sense), the status of the structures and their accuracy and the conditions under which the structures for comparison were determined become particularly important. A summary of these
factors is included in Table 1. At this stage the trypsin and trypsinogen structures have been subjected to more complete refinement and at a higher resolution than have the corresponding chymotrypsin pair. Indeed, the trypsin structure is probably the best determined of any enzyme structure (15, 19). A portion of the most recent electron density map around the active center of DIPT, obtained by Chambers & Stroud (15), is shown in Figure 5.

The pH optimum of the enzymes lies between pH 7 and 9. The structures of the trypsin pair were determined at pH ~7, whereas the α-chymotrypsin structure was initially determined at pH ~4.5 where the molecule has low activity and exists mainly as a dimer (20). However, the α-chymotrypsin structure has since been studied up to pH ~8.3 by Mavridis, Tulinsky & Liebman (21), who reported a number of significant pH-dependent changes that occur between pH ~4.5 and 7.5. The fact that the α-chymotrypsin structure crystallized as a dimer is a further factor that must be taken into account in a detailed structure comparison since Tulinsky et al (22) also found the two molecules that form the dimer to be of somewhat different structure within the dimer interface. This is a critical region for this analysis since it includes the reactive center of α-chymotrypsin.

Therefore, we consider the trypsin-trypsinogen pair of structures most suitable as the primary basis of comparison between enzyme and proenzyme for the purpose

\(^\text{3}\)Abbreviations: DIP, diisopropylphosphoryl; DIPT, LpT, Tg, CT, and Cg refer to the crystallographically determined structures of diisopropylphosphoryl trypsin at pH 7.5, from orthorhombic crystals (15); low pH (~5.0) DIP-trypsin from trigonal crystals (13); trypsinogen (pH 7.5) from trigonal crystals isomorphic with LpT crystals (13); α-chymotrypsin (14); and chymotrypsinogen (12, 17), respectively. NPGB, p-nitrophenyl-p'-guanidino benzoate; BPTI, the basic pancreatic trypsin inhibitor.
Table 1 Summary of corresponding zymogen and enzyme structures used for comparison in the text

<table>
<thead>
<tr>
<th>Structures</th>
<th>Resolution (Å)</th>
<th>State of refinement</th>
<th>Refinement scheme</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Chymotrypsin&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2.0</td>
<td>R ~ 43%</td>
<td>Model building; real space and energy refinement</td>
<td>14</td>
</tr>
<tr>
<td>Chymotrypsinogen&lt;sup&gt;c&lt;/sup&gt;, pH 6.3</td>
<td>2.5</td>
<td>R ~ 43%</td>
<td>Calculated phases; rebuilt model</td>
<td>12, 17</td>
</tr>
<tr>
<td>Trypsin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIPI&lt;sup&gt;c,d&lt;/sup&gt;, pH 5.0</td>
<td>2.1</td>
<td></td>
<td>Difference Fourier versus trypsinogen</td>
<td>13</td>
</tr>
<tr>
<td>DIP&lt;sup&gt;d&lt;/sup&gt;, pH 7.5</td>
<td>1.5</td>
<td>R = 23%</td>
<td>Constrained difference Fourier refinement</td>
<td>15</td>
</tr>
<tr>
<td>Benzamidine&lt;sup&gt;b&lt;/sup&gt;, pH 7.0</td>
<td>1.8</td>
<td>R = 23%</td>
<td>Constrained difference Fourier refinement</td>
<td>19</td>
</tr>
<tr>
<td>Trypsinogen&lt;sup&gt;c&lt;/sup&gt;, pH 7.5</td>
<td>1.9</td>
<td>R = 31%</td>
<td>Constrained difference Fourier refinement</td>
<td>13</td>
</tr>
</tbody>
</table>

<sup>a</sup>State of refinement is indicated as the value of R = Σ |F<sub>o</sub> - F<sub>c</sub>| / ΣF<sub>o</sub>, representing the agreement between observed and calculated structure factors to the resolution indicated.
<sup>b</sup>Crystallized from ammonium sulfate.
<sup>c</sup>Crystallized from ethanol.
<sup>d</sup>Crystallized from magnesium sulfate.

of this review. Further, the comparison between chymotrypsinogen 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 structures of the two proenzymes. Taking for granted that the enzymes themselves are catalytically active (something that would not be obvious if the structures alone were viewed in

Figure 5 A portion of the electron density map for DIPT in the region of the active center (15).
the absence of such knowledge), we compare the two proenzymes emphasizing the question of 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 (23).

This question cannot yet be answered simply. Some of the smallest alterations in structure may be crucial for the answer, whereas others may merely be consequences of crystal packing forces, etc. Therefore, we first describe the conditions under which the trypsinogen and trypsin structures were determined. It must be noted that both trypsin and α-chymotrypsin were crystallized from high salt solutions, whereas trypsinogen and chymotrypsinogen were crystallized from ethanol solutions. Trypsin can also be crystallized from ethanol solution (13).

**Trypsinogen Structure**

Crystals of bovine trypsinogen were obtained by vapor diffusion of a 3% solution of protein against 30% ethanol/water (13). 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 (24). 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 were trigonal, space group P3$_1$21 with cell parameters $a$=55.17 Å and $c$=109.25 Å. The unit cell contains one molecule per asymmetric unit (13).

**Trypsin Structure**

At low pH, DIP-trypsin crystals isomorphous with the trypsinogen crystals were obtained. These crystals could be grown at pH 4–5.5 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$_1$2$_1$2$_1$ (15, 25, 26). The trigonal crystals were found to be unstable above pH 5.5, which suggests that there is some type of structural reorganization that occurs around pH 5.5.

Thus, there are three trypsin-related structures used as the basis for comparison between proenzyme and enzyme. The neutral pH, orthorhombic form of DIPT (the most highly refined structure) is compared with the neutral pH, trigonal form of trypsinogen. The fact that DIP-trypsin crystallizes at low pH, in a form isomorphous with Tg (LpT), is useful in that it provides an accurate means of assay for structure change between the two molecules by difference Fourier map techniques. These methods can pinpoint differences in structure that are on the same scale or smaller than the average expected errors in atomic coordinates of the refined structures from different crystal forms. Thus, there is an independent means of assay for structure change, which is almost independent of the absolute accuracy of the orientation of, for example, a single side chain that may not be accurately placed by the constrained difference Fourier refinement procedures in either structure alone.
There is a further complexity to this comparison, which may be used to advantage. There is a strong possibility that LpT has a different structure from the neutral pH form (DIPT), as does chymotrypsin (21, 22). Further, there is the possibility that the 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 only upon a single ionization between pH 2 and 8, and this has been identified as the ionization of Asp 102 at the active site (27). Therefore, any pH-dependent structure change occurring between pH 7 and 5 in trypsin is not a change reflected in the inherent activity of the enzyme, unless it can be associated directly with Asp 102. Thus, the changes between the isomorphous structures (LpT and Tg) are more likely to pick out the important changes for the change in catalytic activity, the subject of this review, in a direct way.

**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 may be due to errors in the model (13). The histogram shown in Figure 6a points up the distance between all α-carbons for residues common to both structures after rotation, and it clearly identifies the regions of greatest structural change. The mean deviation between equivalent α-carbon atoms in the two structures, calculated by leaving out the 32 residues with obviously different structure, was \( \Delta d_{ca} = 0.4 \text{ Å} \). The difference in conformation of the main chain for all these residues is about the same as for the α-carbons of the best region (the C-terminal α-helix) of the Cg/CT comparison (12). The mean deviation of position between all atoms in the main chain and side groups for the same residues of Tg and DIPT was \( \Delta d = 0.45 \text{ Å} \).

Since the two structures are so similar in these regions, the distances presumably represent the sum of errors in the two coordinate sets, coupled with real differences in structure. Most of the differences in side-chain orientation for structurally homologous regions occur on the surface of the protein and are probably not of primary importance for our purpose here.

**The Active Center**

The structure around the active center of Tg is shown in Figure 3. Crystallographic studies of chymotrypsinogen (12) showed that the orientation and immediate environment of the active site residues, Asp 102, His 57, and Ser 195, were very similar to those 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 and Ser 214 (\( \Delta d_{ca} = 0.27 \text{ Å}, \Delta d_{ca} = 0.3 \text{ Å} \),
Figure 6  (a) Histogram showing the distance Δd (Å) between equivalent α-carbon atoms in Tg and DIPT after the two coordinate sets have been rotated and translated to achieve optimum superposition of the two coordinate sets. (b) Histogram showing the average positional difference of all atoms within each residue of the Tg/DIPT comparison.
respectively) 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, although the Oγ of Ser 195 is still hydrogen bonded to the εN of His 57 (Figure 3) as it is in the benzamidine trypsin structure (28). 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 in part be 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 DIPT (28). The fact that Bode & Schwager (19) 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 DIPT when compared with benzamidine trypsin.

The change in position of Cα Ser 195 seen in Tg may be of some importance for the relative inactivity of trypsinogen. It lies at the end of the sequence of residues 186–195, all of which occupy different positions in Tg and DIPT as they do in the comparison of Cg with CT. These residues are loosely organized and adopt several slightly different conformations in Tg (13).

Birktoft, Kraut & Freer (17) observed a striking difference between the structures of CT and Cg in the active center. They defined an “ideal” site for the Oγ of Ser 195 relative to the side chain of His 57. This was chosen to lie 2.8 Å away from the εN of His 57 and in the direction of the sp² orbital of the εN. Although they found that the Oγ of Ser 195 was within 0.7 Å of this site in Cg, it could not be brought closer than about 2.5 Å to this site in CT by any manipulation of these side-chain orientations (17). They raised the question of whether a strained hydrogen bond at this site in the enzyme could contribute to enhanced catalytic efficiency. The Oγ of Ser 195 in Tg and in DIPT (where the Oγ is covalently bonded to the DIP group) are not so dramatically different. Although the actual position of the Oγ differs by almost 1 Å between these structures, the Oγ lies at 0.7 and 1.1 Å, respectively, away from the ideal position; it is not clear that this difference is enough to account for any significant contribution of distortion at this site to the catalytic mechanism. Birktoft, Kraut & Freer (17) have also identified other quite large differences between CT and Cg in this region of interface between the two “structural cylinders” (14) of CT. We see no corresponding difference in the Tg/DIPT comparison. Perhaps the fact that the CT structure existed as a dimer is in part responsible for alteration in the structure of α-chymotrypsin, relative to the neutral pH form of the enzyme?

This does not rule out the possibility that even a small change in the active center structure could contribute to the generation of enzyme activity; however, at present there is no common factor that can be identified.

*The Side-Chain Binding Pocket*

Freer et al (12) questioned whether the absence of a specific binding pocket could explain the inactivity of Cg. Trypsinogen has a large and accessible binding pocket.
(see Figure 3), although the structure of the cavity is different from that found in DIPT or in benzamidine trypsin (see Figure 4).

Benzamidine is a competitive inhibitor of trypsin \(K_i = 1.8 \times 10^{-5} \text{ M}\) (29), which binds inside the side-chain-specific binding pocket of the enzyme (19, 28). Benzamidine was present at 10 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 (30). The binding site seen in the Tg structure is large enough to accept a benzamidine molecule, and the fact that none was found implies that the differences between zymogen and enzyme structures (as shown in Figures 3 and 4) 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 only partially formed in Cg (12). There are much smaller but significant changes between Tg and DIPT (or LpT) in the chain from residues 188A–195. 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 \(\alpha\)-carbon of Gln 192 in Tg moves by 0.9 \(\text{Å}\) on activation. The side chain of Gln 192 moves by about 4 \(\text{Å}\). The side chain of Asp 189 inside the specific binding pocket of trypsin and responsible for the substrate specificity of trypsin lies within 0.9 \(\text{Å}\) of the position found in Tg, \(\Delta d_{\text{ca189}} = 0.9 \text{ Å}\). 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 \(10^2\) times worse than for trypsin (31). 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 (32–35). In Tg the chain structure differs slightly from trypsin, as shown in Figure 3. The C–O of Ser 214 (\(\Delta d = 0.3 \text{ Å}\)) and the N–H (\(\Delta d = 0.5 \text{ Å}\)) and C–O (\(\Delta d = 1.4 \text{ Å}\)) 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 N–H and C–O of Gly 216 would be unable to participate in substrate binding (12). This and the other differences in binding site structure between Cg and Tg are not inconsistent with the fact that the BPTI will form a stoichiometric, 1:1 complex with trypsinogen and with trypsin or chymotrypsin but will not so complex with chymotrypsinogen (36).
There was an error in the original placement of the side chain of Trp 215 in DIPT (Figure 8 of reference 28). Refinement of that structure revealed that the indole ring is reversed, such that the εN points in toward the center of the molecule (15). The same configuration is found in Tg. This orientation is the same as that found in CT (14). Chemical modification of Trp 215 in trypsinogen blocks the binding of BPTI by trypsinogen (37).

**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 Δd = 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 O₈⁻ 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 (see Figures 7 and 8).

Freer et al (12) 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 (38, 39) 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 carboxyl via water molecules hydrogen bonded to the õN of His 40. The εN 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 (see Figure 9).

*Figure 7* Structure of trypsinogen in the region of His 40. The N₈⁻ of His 40 is hydrogen bonded to the Oγ of Ser 32 as it is in trypsin. The C–O of Gly 193 is moved from its position in trypsin. Asp 194 adopts a similar conformation to that seen in trypsin (13).
Figure 8 The structure of trypsin in the region of His 40 shows the hydrogen bond formed between Nδ-1 of His 40 and the Oγ of Ser 32, and the hydrogen bond between the Ne-1 of His 40 and the C–O of Gly 193. Asp 194 is salt bridged to the amino terminal of Ile 16 (15).

Figure 9 The chymotrypsinogen structure around His 40 and Asp 194 shows a closer interaction between these two groups, which are in quite different orientation than that seen in trypsin or trypsinogen. Coordinates were obtained from the Brookhaven Protein Data Bank (12).
In Tg neither Lys 145 nor His 40 lies close to the side chain of Asp 194, and therefore they cannot be involved in stabilizing Asp 194. The side chain of Lys 145 is on the outside of the molecule and the \(\zeta N\) of Lys 145 is about 10 Å from Asp 194. Indeed, there is no basic side chain 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 (\(\Delta d_{ca} = 0.4\) Å). The \(\zeta N\) of His 40 is hydrogen bonded to the Oy of Ser 32 in both Tg and DIPT (also in CT but not in Cg). Ne–2 of His 40 is hydrogen bonded to the C–O of Gly 193 in DIPT and CT, but it is hydrogen bonded to an external solvent molecule in Tg. This change corresponds to the movement of the Ca of 194 (\(\Delta d = 0.9\) Å) and 193 (\(\Delta d = 2.0\) Å) toward the solvent-filled cavity around Asp 194. Wright (39) 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 (40, 41). His 40 is also not a conserved residue among trypsin-like enzymes; it is Leu in bovine thrombin, Gly in blood-clotting factor X, and Met, Arg, or Leu in bacterial trypsin sequences (41).

**The Autolysis Loop (Residues 142–151)**

The largest conformational difference between Tg and DIPT or LpT occurs between residues 142–151. This sequence was referred to as the autolysis loop (42) since it contains a region where two amino acids are excised in a-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 were to approach the Lys 15–Ile 16 bond in the Tg conformation.

Trp 141 is essentially in the same orientation in Tg as found in DIPT (\(\Delta d_{ca} = 0.6\) Å). The side chain of Asn 143, which was solvent accessible in DIPT, points in toward the cavity around Asp 194 in Tg (\(\Delta d_{ca} = 1.4\) Å). The side chain of Thr 144 (Tg) is moved by several angstroms toward the disulfide Cys 191–Cys 220 (\(\Delta d_{ca} \sim 4.0\) 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 access of the activating enzyme to the Lys 15–Ile 16 bond. The sequences for residues 146–154 are extremely variable in related enzymes (41), 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 Effect of Divalent Cations on Trypsinogen**

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, (43)], is common to both trypsinogen and trypsin. Occupancy of this site stabilizes the protein toward thermal denaturation or autolysis (24). The primary \(Ca^{2+}\) binding site in trypsin was first identified and described in detail by Bode & Schwager (44), based on analysis of an indepen-
dently refined trypsin structure seen at 1.8 Å and on the Ca²⁺ content of their crystals. The difference map calculated between the isomorphous crystals of Tg and LpT showed no electron density in the Ca²⁺ region. Therefore, the primary calcium ion binding site in trypsinogen and low pH trypsin are identical. There are small differences between this Ca²⁺ site in Tg and DIPT (13).

A secondary Ca²⁺ binding site of \( K_D \approx 10^{-1.8} \) M exists only on the zymogen. Binding of Ca²⁺ at this site is essential for complete and efficient activation of trypsinogen (24), and this site is more pertinent to the subject of this review. 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 whereas \( k_{cat} \) remains unchanged (43).] Tryptic hydrolysis 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 Ca²⁺ dependent in a very similar way, whereas hydrolyses of peptides containing only one aspartate do not depend on Ca²⁺ (43, 45). Ca²⁺ binding to Asp 13 and Asp 14 together, thus, is sufficient to explain the zymogen Ca²⁺ effect (43). There is as yet no corroborative crystallographic evidence for binding at this site since Ca²⁺ was excluded from the trypsinogen 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 \( (k_{cat} \approx 10^{-3} \text{ sec}^{-1}) \) as the Lys-Ile bond on the same sequence in Tg for which \( k_{cat} = 2.5 \times 10^{-3} \text{ sec}^{-1} \) (43). Thus, it is clear that the slow hydrolysis of the Lys-Ile bond by trypsin, a factor important in preventing premature activation of all trypsinogens, 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 activatability of trypsinogen (46). This is in agreement with our finding that the Val 10-Asp 14 sequence is not a determinant of tertiary structure (13). The binding of Ca²⁺ 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 it activates other zymogens. The loosely structured hexapeptide is presumably protected against trypsin more by its negative charge than by tertiary structure. The hexapeptide sequence also provides a good substrate character for the physiological activator, enterokinase (47). The hexapeptide, with its dual role as most favored substrate for enterokinase and a poor but still most favored substrate site for activation by trypsin, is highly conserved throughout all species (41). The Lys 15–Ile 16
bond is not inaccessible, although some reorientation of this chain or of the autolysis loop must accompany enzyme binding to this site. The fact that this bond is still the most readily hydrolyzed by trypsin in the presence of Ca\(^{2+}\) is consistent with the fact that compact native structures are generally more resistant to degradation than are denatured ones. The \(\phi\) and \(\psi\) 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.

In many of the trypsin-like enzymes involved in biological control mechanisms, the primary activating cleavage occurs at a site analogous to the Lys 15–Ile 16 bond with respect to the catalytic component of the protein. This split may be far from the N terminal of the whole protein since the activation peptides, or proteins released, often have separate functions (11). In prothrombin, for example, the activating cleavage occurs between residues Arg 323 and Ile 324 in the sequence. The B-chain (residues 324–582) is highly homologous with trypsin, whereas the A-chain, which remains attached to the B-chain via a disulfide bridge, is thought to have an important role in the blood-clotting mechanism (48).

The Reasons for Zymogen Inactivity

Freer et al (12) and later Wright (39) identified four possible factors that might be responsible for inactivity of chymotrypsinogen. We now see that three of them do not seem to be shared by trypsinogen, since (in Wright’s notation): (i) the binding site of trypsinogen is open and accessible whereas it is not quite so open in chymotrypsinogen, (iii) the potential source of a hydrogen bond from Gly 216 NH is not excluded, although this group is moved by 0.5 \(\AA\) from its substrate binding position in the free enzyme; and (iv) 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 remaining reason was the lack of an important hydrogen-bonding group in the N–H of Gly 193. It has been pointed out that the N–H groups of Gly 193 and of Ser 195 could play an important catalytic role in stabilization of reaction intermediates (18, 49, 50). In chymotrypsinogen the N–H of Gly 193 points away from the position for the carbonyl oxygen of the substrate (12). In trypsinogen the N–H 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 (18) is removed by 1.9 \(\AA\) from the position of the site in trypsin. Thus, oxyanion stabilization is possible in trypsinogen but not at the ideal site. The 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, which is subsequently occupied by Ile 16 upon activation, when Ile 16 forms the salt bridge to Asp 194. There is also plausible connection between the small changes in the binding pocket and the location of Ile 16. Similar change 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 is 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 ε N of His 57) and the oxyanion binding site formed by the N–H groups of Gly 193 and Ser 195 are too far apart to cooperate in substrate hydrolysis. Even if the proenzyme structure were 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 N–H 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 Melvin Jones and Rebecca Price for assistance with calculations used in the comparison of structures. We gratefully acknowledge the support received from research grants from the National Institutes of Health and the National Science Foundation. One of us (RMS) is the recipient of a National Institutes of Health Career Development Award, another (AAK) was the recipient of a National Institutes of Health Postdoctoral Fellowship, and another (JLC) holds a National Institutes of Health Predoctoral Traineeship.

This is contribution No. 5445 from the Norman W. Church Laboratory of Chemical Biology, California Institute of Technology.

Literature Cited