

Crystal Structure of the Complex of Porcine Trypsin with Soybean Trypsin Inhibitor (Kunitz) at 2.6-Å Resolution[†]

R. M. Sweet,[‡] H. T. Wright,[§] J. Janin,[#] C. H. Chothia,[#] and D. M. Blow^{*}

ABSTRACT: The complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) was crystallized from 17% ethanol at pH 7. The crystals have $P2_12_12_1$ symmetry with $a = 59.0$, $b = 62.2$, and $c = 150.5$ Å. They contain one molecule of the 1:1 complex in the crystallographic asymmetric unit. Three-dimensional X-ray diffraction data were measured for these crystals and for crystals soaked with 0.3 mM K_2PtCl_4 , 1 mM $K_3UO_2F_5$, 5 mM mercury salicylate, 0.2 mM $(NH_4)_3IrCl_6$, and 5 mM mercury acetate. All strong reflections to 2.6 Å were measured, including almost all reflections to 3.8 Å and 56% of the reflections between 3.0 and 2.6 Å. Beyond 3.8 Å only the $K_3UO_2F_5$ and salicylate derivatives were useful. In the resulting electron density maps the trypsin molecule was easily identifiable, and although some parts of the inhibitor were not clear, a precise interpretation of residues 1'–93' was made, including the contact region. Coordinates for this part of the inhibitor, and for adjacent parts of the trypsin molecule, were refined by the real-

space refinement technique. Interactions made by soybean trypsin inhibitor with trypsin at P_3 , P_2 , P_1 , P_1' , and P_2' are extremely similar to interactions observed in a crystallographic study of the complex of pancreatic trypsin inhibitor (Kunitz) with bovine trypsin (Rühlmann, A., Kukla, D., Schwager, P., Bartels, K., and Huber, R. (1973), *J. Mol. Biol.* 77, 417) and to the interactions which a normal polypeptide substrate is believed to make. These interactions account for two-thirds of about 300 interatomic contacts and for 13 of the 18 probable hydrogen bonds between the inhibitor and the enzyme. Interatomic distances at the active site show the complex to be in the form of a tetrahedral adduct of the scissile bond to the active serine. The strong binding energy of the inhibitor and the stabilization of the tetrahedral form result from the nature of the active site of the enzyme, which is designed to stabilize the transition state of peptide hydrolysis.

Work on the structure and activity of chymotrypsin and its homologs gave information about the conformation of groups at the active center of the enzyme, and about the mode of binding of small substrates on the amino-terminal side of the bond to be cleaved (for a review see Blow, 1971). This led to a detailed, hypothetical model for the steps of a chymotrypsin-catalyzed hydrolysis (Henderson *et al.*, 1971). The interactions made by a substrate on the carboxyl-terminal side of the scissile bond could not be studied directly, because no suitable substrate analog is known, and because it is difficult to investigate a true substrate before it is hydrolyzed by the enzyme. Soybean trypsin inhibitor (STI)¹ can be reversibly hydrolyzed at a unique site by trypsin, and this reaction is strictly analogous to the enzyme-catalyzed hydrolysis of a normal substrate (for reviews, see Laskowski and Sealock, 1971; Laskowski *et al.*, 1971). We began the experiments described below in the hope that crystals of the STI–trypsin complex would show how the polypeptide chain is oriented at the active site, and that the ori-

entation of the scissile bond proposed in the hypothetical model could be confirmed. The chemical state of the scissile bond in the complex was unknown.

We selected porcine trypsin for crystallization experiments because of its greater stability and homogeneity (Viyathil *et al.*, 1961). In addition, the structure of bovine trypsin was already well advanced, and we felt that further evidence about the conformational homology of the trypsins would be useful.

We learned that Huber and his colleagues were conducting a parallel investigation on the structure of the complex of basic pancreatic trypsin inhibitor (Kunitz) (PTI) with bovine trypsin. The results of this study have now been published (Rühlmann *et al.*, 1973). Prior to this, however, the structure of free PTI (Huber *et al.*, 1970) allowed our two groups to put together all the structural data and predict a structure for a PTI complex which turned out to be correct in its major features (Blow *et al.*, 1972). A similar prediction was made independently by Stroud *et al.* (1971).

The crystallographic study of Rühlmann *et al.* (1973) vindicated these predictions, but added a most unexpected feature. The complex existed neither as a Michaelis complex, nor as an acyl-enzyme intermediate, but as a tetrahedral adduct. We shall present evidence that the STI complex with porcine trypsin exists as a similar adduct, and that both inhibitors bind in an extremely similar fashion. The polypeptide chains adjacent to the scissile bond probably bind exactly as a good polypeptide substrate would bind. We shall compare the similarities between the two inhibitors, and use these similarities to obtain as clear an understanding as possible of the strong, specific binding which they show toward trypsin. In neither case is the area of contact between the two molecules extensive, but it is sufficient to supply 15–20 kcal of binding energy.

These results give excellent confirmation of the existing model for the mode of action of the trypsin family of enzymes,

[†] From the Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England. Received April 11, 1974. During the course of this work, R. M. S. was supported by the Damon Runyon Foundation and the European Molecular Biology Organization, H. T. W. by a National Institutes of Health Fellowship, and J. J. by a NATO Fellowship.

[‡] Present address: Department of Chemistry, University of California, Los Angeles, Calif. 90024.

[§] Present address: Frick Chemical Laboratory, Princeton University, Princeton, N. J.

[#] Present address: Service de Biochimie Cellulaire, Institut Pasteur, 75015-Paris, France.

¹ Abbreviations used are: STI, soybean trypsin inhibitor (Kunitz); PTI, basic pancreatic trypsin inhibitor (Kunitz). Amino acids of trypsin are identified by the residue number of the homologous amino acid in the chymotrypsinogen sequence. Residue numbers in STI are indicated by primes.

and give further insight into the role of the tetrahedral intermediate in the reaction mechanism. They also emphasize the importance of secondary specificity in the hydrolysis of protein and peptide substrates. Finally, they suggest that sufficient binding energy to ensure a stable complex in an intracellular environment can be achieved if a dozen or so amino acids make contact. A short preliminary account of this work has been published (Blow *et al.*, 1974).

Materials and Methods

Crystallization. The first samples of porcine trypsin were a gift from Novo Biochemicals. Later samples were purchased from Koch-Light. We were grateful for a gift of purified soybean trypsin inhibitor complex from Professor M. Laskowski, Jr. Later samples were purchased from Worthington.

Six milliliters of buffer (0.1 M Tris-Cl (pH 8.1)–0.1 M NaCl) was used to dissolve 210 mg of soybean trypsin inhibitor. To this solution was added an equal volume of buffer containing 190 mg of pork trypsin. The solution was left for 3 hr at 4°, filtered, and passed through a column of Sephadex G75 equilibrated with the same buffer. About 250 mg of complex preceded impurities from the column. The fractions containing the complex were concentrated and exchanged against 0.1 M Tris-Cl (pH 7.0) by ultrafiltration on an Amincon PM30 membrane under nitrogen pressure. The volume was adjusted to 7.5 ml and divided into 0.9-ml samples in small dishes.

For crystallization, each sample of protein solution was cooled to approximately –2° and cold ethanol was added slowly with stirring until persistent precipitation occurred, usually after the addition of roughly 0.18 ml. The precipitate dissolved as the dishes were warmed to room temperature. Each dish was seeded with several (from one to a dozen) small crystals or crystal fragments and was left covered at 20° to crystallize. Usable crystals often formed in 3–12 weeks.

The orthorhombic crystals are blocks elongated along the *a* dimension and shortest along *b*. The faces {010}, {001}, {101}, and {10 $\bar{1}$ } are developed, so that they appear six-sided when viewed along *b*. For diffraction experiments crystals were transferred successively into 10 mM Tris-Cl (pH 7.0) solutions containing 20% ethanol and 20% 2-methyl-2,4-pentanediol, 40% 2-methyl-2,4-pentanediol, and finally 50% 2-methyl-2,4-pentanediol. For X-ray diffraction experiments, crystals typically 0.6 × 0.3 × 0.2 mm in size were mounted in tightly fitting glass capillaries, the *a* axis along the length of the capillary, and held in place with cotton fibers in order to prevent movements of the crystal during measurement. The crystal was completely surrounded by the 50% 2-methyl-2,4-pentanediol solution.

Crystallographic Data. Symmetry and systematic absences indicate that the space group is $P2_12_12_1$ (orthorhombic). The diffraction angles of 12 high-order reflections measured with a Hilger-Watts diffractometer and a 0.8-mm focal spot width gave the parameters $a = 59.0$, $b = 62.2$, and $c = 150.5$ Å.

The asymmetric unit (one-quarter of the cell) has a volume of 138,100 Å³, and if one assumes that it contains one molecule each of trypsin (mol wt 23,400) and of inhibitor (20,100), the fraction of volume occupied by protein can be estimated to be 38%, a relatively low value (Matthews, 1968).

Collection of Intensity Data. A computer-controlled Hilger-Watts four-circle diffractometer was used for collection of three-dimensional diffraction intensities. The specimen was 275 mm from the X-ray source, a Philips fine-focus tube with a Cu target and a Ni filter, and 330 mm from the Xe-filled proportional counter. The crystal was mounted so that its *a* axis coincided with the ϕ -spindle, and apertures, chosen to admit

the full integrated diffraction intensity and to minimize background, were placed immediately in front of the specimen and the counter.

A first set of 5-Å data was collected using ω scans of 50 0.01° steps, 1.2 sec each, from which the largest sum of 25 contiguous counts was taken as raw intensity and the sum of the remaining 25 taken as background intensity (Watson *et al.*, 1970). High-resolution data were collected by the integrating geometry-peak scan technique (Wyckoff *et al.*, 1970) in order to increase the speed of measurement. The apparent size of the X-ray source was set to 0.4 (height) × 0.8 (width) mm by increasing the take-off angle to about 6°. Eight counts of 3–8 sec were made at intervals of 0.02° in ω and the largest sum of four contiguous counts was taken as raw intensity. The background was found to be adequately represented as a spherically symmetrical function of $\sin \theta$, which was estimated by observing the intensity of scattered radiation when the crystal and counter were oriented midway between positions for Bragg reflections. The diffracted intensities at settings corresponding to indices ($h/2, k/2, 0$), ($1/2, k/2, 0$) and ($1/2, 1/2, l$) were measured and a smoothed function of $\sin \theta$ was derived from these measurements for each crystal.

Corrections for X-ray absorption by the specimen were made according to Furnas (1951); the axial reflection (2,0,0) was scanned at intervals of 11.25° of the ϕ angle throughout 360° rotation. Possibly due to focusing effects by crystals having experienced mechanical distortions, systematic differences sometimes occurred between measurements made 180° apart. When such differences were large the crystal was rejected. The maximum absorption correction derived by this technique was typically 5–10%.

The number of reflections which could be measured from each crystal was limited by X-ray-induced disorder, which was monitored by periodic measurement of four standard reflections at 4.5-Å spacing. Crystals were changed when the intensity of the standards had dropped by about 10%, which allowed about 8000 reflections to be measured from parent crystals, but only 3000–5000 from heavy atom derivatives. Complete data sets therefore required the use of several crystals, and the following procedure was established. (a) A complete set of parent intensities with positive h, k, l between 20- and 2.6-Å resolution was obtained from crystals of the complex. (b) Reflections more intense than twice their standard deviations were selected and sorted in ten shells of increasing $\sin \theta$. They included almost all of the 3.8-Å data, 87% of the 3–3.8-Å data, and 56% of the 3–2.6-Å data. (c) Native and derivative intensities for the reflections selected in (b) were measured (for longer times) shell by shell, alternating blocks of fifty (h, k, l) with the corresponding ($\bar{h}, \bar{k}, \bar{l}$).

The structure amplitudes of symmetry-related reflections agreed to about 5% between 20- and 3.5-Å resolution, 6.5% from 3.5 to 3 Å, and 11% from 3 to 2.6 Å. The reflections from two planes, ($h, 1, l$) and ($h, 2, l$), were measured on a single parent crystal in order to provide a scaling set of data, overlapping with all shells. The radial distribution of F shows a minimum at 6-Å resolution and a maximum at 4.5 Å. Between 4.4 and 3 Å it follows a Debye-Waller distribution with a temperature factor $B \sim 25$ Å². The parent structure amplitudes were scaled together by a full-matrix least-squares procedure. Scale factors were determined for each shell of derivative data by least-squares fit to the corresponding parent amplitudes.

Preparation of Heavy-Atom Derivatives. Heavy-atom derivatives were prepared by the soaking of crystals for 2–4 weeks in 50% 2-methyl-2,4-pentanediol solutions with various heavy metal compounds added. A preliminary survey of intensity dif-

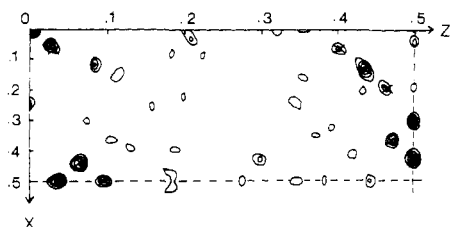


FIGURE 1: (*h0l*) difference Patterson projection of mercury salicylate derivative at 3.2-Å resolution. Crosses indicate heavy atom vectors.

ferences was made by comparison of 15° precession photographs of the (*h0l*) zone. Three-dimensional data to 5-Å resolution were collected as described on crystals soaked in 0.3 mM K₂PtCl₄, 1 mM K₃UO₂F₅, and 5 mM mercury salicylate. High-resolution data eventually included the last two derivatives as well as (NH₄)₃IrCl₆ (0.2 mM) and mercury acetate (5 mM, pH neutralized before addition).

Determination of Heavy-Atom Positions. Difference Patterson projections of the (*h0l*) zone from 15° precession photographs suggested that mercury salicylate (Figure 1) and K₂PtCl₄ might both provide acceptable isomorphous substitutions. A three-dimensional 5-Å difference Patterson map of the K₂PtCl₄ derivative was calculated, and a consistent set of peaks could be identified on the three Harker sections, $x = \frac{1}{2}$, $y = \frac{1}{2}$, and $z = \frac{1}{2}$, from which the position of one Pt atom was derived. Single isomorphous replacement phases were calculated from this position, and the difference electron density map calculated with these phases revealed the presence of at least one other Pt site. Single isomorphous replacement phases from these two sites were then used to calculate a difference electron density map for the K₃UO₂F₅ derivative, and a tentative solution of this derivative was used for further interpretation of the K₂PtCl₄ data. Six Pt and five U sites were eventually identified by extension of this procedure.

At this stage, diffractometer data were collected to 5-Å resolution on the mercury salicylate derivative. A difference electron density map phased with both K₂PtCl₄ and K₃UO₂F₅ confirmed the presence of two clean Hg sites in this derivative. The same technique was then used to identify one Ir site in the (NH₄)₃IrCl₆ derivative and three Hg sites (different from the mercury salicylate sites) in the mercury acetate derivative.

Refinement of Heavy-Atom Positions. Refinement of heavy atom positions, occupancies, and scale factors for derivative data was first attempted by a procedure minimizing

$$\sum_{hkl} [(|F_P| - |F_{PH}|)^2 - (2/\pi) |f_H|^2]$$

(Rossmann, 1960) in the absence of known phases. When preliminary phases became available, refinement procedures minimizing $\sum_{hkl} m_{hkl} (|F_P + f_H| - |F_{PH}|)^2$ were used (Dickerson *et al.*, 1961). Refinement was carried out independently for the shell of data from each individual derivative crystal, and showed in several cases that the occupancies of individual sites varied from crystal to crystal. In each case, an overall scale factor and a temperature factor were refined as well as the occupancy and coordinates of each site. All reflections with a figure of merit m_{hkl} larger than 0.2 were included in the calculation. Data collection from individual crystals often covered a limited range of resolution, and meaningful refinement of atomic temperature factors was impossible. Occupancies were refined with the Debye-Waller factor B arbitrarily fixed at 20 Å² for all atoms.

In order to avoid the bias toward the starting values that calculated phases introduce in the refinement of heavy atom pa-

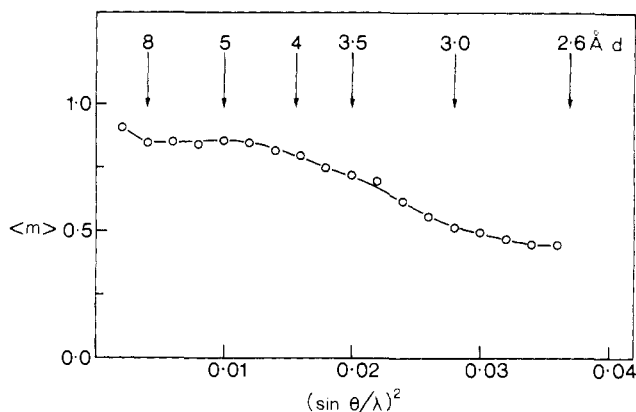


FIGURE 2: Variation of mean figure of merit, m , with Bragg spacing, d .

rameters, phases calculated with all derivatives but one were used (a) to calculate a difference electron density map with data from individual crystals of the derivative omitted in the phasing and (b) to refine the corresponding atomic positions. This procedure eliminates spurious sites and leads to rapid convergence at the correct positions (Blow and Matthews, 1973). However, it tends to give low values for the occupancies. The refinement was therefore done in the last stages using phases from all derivatives. Finally, the anomalous occupancies were refined by a procedure minimizing

$$\sum_{hkl} m_{hkl} [g_H - (\Delta F/2)]^2$$

where g_H is the projection on ($F_P + f_{PH}$) of the calculated anomalous scattering contribution and ΔF is the observed Bijvoet difference ($|F_{hkl}| - |F_{\bar{h}\bar{k}l}|$).

Phasing and Electron Density Maps. Phase information from K₂PtCl₄, K₃UO₂F₅, (NH₄)₃IrCl₆, and mercury salicylate derivatives to a resolution of 5 Å was first used to calculate an electron density map of a quarter of the orthorhombic cell. The mean figure of merit was 0.70 for 2791 reflections.

The K₂PtCl₄ derivative was abandoned for phasing after the initial 5-Å investigation because of lack of isomorphism and irregular substitution of the main Pt sites. Only partial data were collected on the (NH₄)₃IrCl₆ and mercury acetate derivatives: the phasing of the 3.8–2.6-Å data relies entirely on the K₃UO₂F₅ and mercury salicylate derivatives. Table I summarizes the refined parameters which were used. Although the two Hg sites of mercury salicylate have similar and reproducible occupancies, those of the U sites vary. Minor sites were omitted from refinements when their inclusion led to no significant decrease of the lack of closure. Despite changes in occupancy, the refined positions are practically the same in all crystals of each derivative. The lower quality of the high-resolution data limits the use of anomalous dispersion in phasing, as can be seen from the relative values of the average calculated anomalous contribution f_H'' and of corresponding lack of closure E'' . The mean figure of merit of 12,127 reflections between 20- and 2.6-Å resolution was 0.67, but the phases are poorly determined between 3.5 and 2.6 Å (Figure 2).

An electron density map using only part of these data and limited to 3.0-Å resolution was computed at an intermediate stage of the investigation. In this map, the trypsin molecule could be interpreted with few difficulties, but the polypeptide chain could not be traced in the STI molecule.

In parallel with extension of the isomorphous replacement data, the interpretation of the trypsin molecule was used to

TABLE I: Heavy Atom Parameters.

Derivative	Crystal No. ^a	Resolution (Å)	Site	f^b	f'^b	x	y	z	$ \bar{F}_H ^c$	E^d	$ \bar{F}_H '^c$	E''^d	No. of Reflections
(NH ₄) ₂ IrCl ₆	1	20-3.5	Ir1	39		0.178	0.595	0.2765	37	57			6524
Hg salicylate	2	20-5	Hg1	46	3.3	0.466	0.317	0.453	87	45	7	20	2600
			Hg2	38	3.5	0.404	0.322	0.482					
Hg salicylate	3	5-3.8	Hg1	42	2.0	0.468	0.316	0.453	74	55	4	21	2335
			Hg2	37	2.3	0.397	0.314	0.4835					
Hg salicylate	4	5-3.5	Hg1	42	1.6	0.467	0.316	0.453	74	62	3.5	23	2325
			Hg2	37	2.0	0.397	0.315	0.484					
Hg salicylate	5	3.8-3	Hg1	30	1.0	0.467	0.314	0.453	47	37	1.5	19	3221
			Hg2	31	0.8	0.402	0.319	0.4825					
Hg salicylate	6	3-2.6	Hg1	40		0.467	0.310	0.4525	48	44			3659
			Hg2	25		0.402	0.321	0.483					
K ₃ UO ₂ F ₅	7	20-5	U1	37	3.5	0.108	0.499	0.0435	83	47	9	13	1931
			U2	30	3.1	0.167	0.688	0.394					
			U3	22	3.0	0.204	0.699	0.417					
			U4	12	1.2	0.179	0.637	0.2895					
			U5	15	1.7	0.040	0.600	0.4855					
K ₃ UO ₂ F ₅	8	5-3.8	U1	20	1.7	0.107	0.506	0.043	58	45	4	16	2520
			U2	27	1.7	0.168	0.690	0.393					
			U3	20	1.5	0.209	0.698	0.417					
			U4	18	1.1	0.180	0.633	0.2895					
K ₃ UO ₂ F ₅	9	5-3.8	U1	23	1.7	0.106	0.511	0.044	111	97	8	25	2489
			U2	59	4.2	0.166	0.694	0.392					
			U4	53	3.7	0.189	0.636	0.2905					
K ₃ UO ₂ F ₅	10	3.8-3.5	U1	25	1.6	0.111	0.501	0.044	63	63	5.5	25	1404
			U2	24	2.0	0.164	0.692	0.3945					
			U3	28	2.0	0.202	0.703	0.416					
K ₃ UO ₂ F ₅	11	3.8-3.2	U4	23	2.8	0.183	0.635	0.2915					
			U1	16	1.3	0.108	0.500	0.043	48	40	4	19	2769
			U2	17	1.5	0.163	0.687	0.395					
K ₃ UO ₂ F ₅	12	3.5-3	U3	30	2.5	0.206	0.701	0.4175					
			U1	16	0.0	0.108	0.501	0.043	46	41	2.5	24	2404
			U2	21	1.7	0.164	0.687	0.395					
K ₃ UO ₂ F ₅	13	3-2.6	U3	30	2.0	0.206	0.701	0.418					
			U1	15		0.107	0.502	0.042	45	39			2931
			U2	34		0.164	0.694	0.3935					
			U4	27		0.183	0.635	0.291					
Hg acetate	14	20-3.8	Hg3	21	2.1	0.899	0.791	0.221	85	84	6	21	3950
			Hg4	18	0.7	0.997	0.900	0.442					
			Hg5	52	3.3	0.193	0.785	0.2835					

^a As explained in the text, independent refinements were carried out for each shell of intensity data corresponding to an individual heavy atom crystal. ^b The scattering factor, f , and anomalous scattering factor, f' , associated with each heavy atom site are expressed on an arbitrary scale. On this scale the mean protein structure amplitude $|\bar{F}|$ is 500. f and f' were calculated using isotropic form factors $\exp(-B \sin^2 \theta / \lambda^2)$ with $B = 20 \text{ Å}^2$, except for site Ir1, for which a refined value $B = 66 \text{ Å}^2$ has been employed. ^c $|\bar{F}_H|$ is the mean calculated heavy atom structure amplitude and $|\bar{F}_H|'$ is the mean calculated anomalous heavy atom structure amplitude. ^d E and E'' are the root-mean-square lack of closure for isomorphous and anomalous differences, for the "best" phase angle.

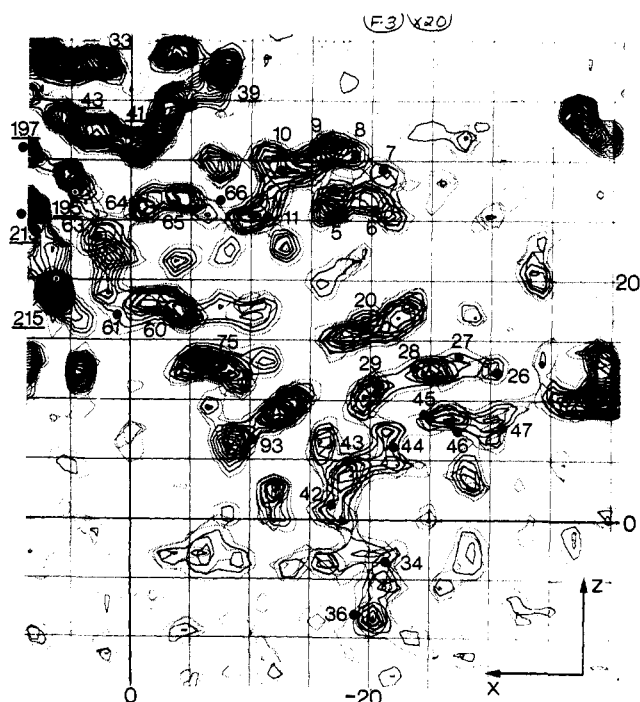


FIGURE 3: Section $x = -\frac{1}{4}$ to $-\frac{3}{4}$ of the "combined" electron density map. The high contours at the upper left are part of the trypsin molecule (cf. Figure 6). C^α positions which appear in these sections are indicated. C^α positions for trypsin are underlined.

provide further phase information. The underlying principle of the technique is that if a substantial part of a structure is known accurately, the phase angles of the scattering of this part of the structure will provide a useful estimate of the phase angles of the structure factors of the total structure. The accuracy of this approach has been analyzed by Sim (1959). Rossmann and Blow (1961) suggested a technique for combining it with isomorphous replacement data (see also Hendrickson and Lattman, 1970), but this has never been seriously used.

Rough coordinates for the trypsin molecule in the STI complex were generated by appropriate rotation and translation of preliminary bovine trypsin coordinates kindly provided by Dr. R. M. Stroud. The orientation of the porcine trypsin molecule in the STI complex is given in Table II. Suitable editing of these coordinates, coupled with the Diamond (1966) model building technique, were used to convert the sequence to that of porcine trypsin. Real space refinement was then used to fit the coordinates to the electron density map (see next section). It

TABLE II: Orientation of Porcine Trypsin Molecule in the STI Complex.^a

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 0.4306 & -0.2285 & 0.8732 \\ -0.0430 & 0.9611 & 0.2727 \\ -0.9015 & -0.1550 & 0.4040 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} -9.05 \\ -6.25 \\ 36.25 \end{pmatrix}$$

^a The transformation matrix relates the coordinates (x', y', z') of a point in the porcine trypsin molecule centered near (10,0,30) in the STI complex to an equivalent position (x, y, z) in the coordinate system for α -chymotrypsin (Birktoft and Blow, 1972). The coordinates are in angstroms in each case, and in the STI complex they are measured from the crystallographic origin. The transformation relates homologous atoms with a standard deviation of about 2.5 Å, but is more accurate in the active center region.

TABLE III: Analysis of a Sample of Trypsin "Heavy Atom" Phases.

Fig of Merit of Phase Determ. by Isomorphous Replacement	Mean Phase Angle Discrepancy (Deg) between Isomorphous Replacement and Trypsin "Heavy Atom" Phases ^a
0.9-1.0	54
0.8-0.9	51
0.7-0.8	64
0.6-0.7	57
0.5-0.6	70
0.4-0.5	67
0.3-0.4	93

^a Mean difference of phase angle between the "most probable phase" derived from the complete isomorphous results and the phase of the calculated trypsin structure factor, from a sample of 500 reflections of moderate intensity.

was found that a considerable amount of guidance was needed to keep the chains in the density of the 3.0-Å map.

Model electron density generated by the real space refinement program was transformed into a set of trypsin structure factors in the appropriate unit cell by the use of a fast Fourier transform program (Ten Eyck, 1973). These phases were used to compute electron density maps phased on trypsin, or combined with the complete isomorphous replacement phases (Rossmann and Blow, 1961). We are grateful to Mr G. Brice for his help in these calculations.

As is usual in such maps, the electron density of parts of the structure used for phasing is greatly enhanced (Figure 3). Questions about whether the "combined" map is more suitable for structural interpretation can only be answered in a subjective way, and the enhanced density of the trypsin molecule is certainly a disadvantage, on the subjective level. It also makes the "combined" map unsuitable for use in real space refinement. In practice, the final isomorphous replacement map was found to be readily interpretable in most regions, and we only referred to the "combined" map in regions of difficulty. In the majority of difficult parts of the inhibitor, the "combined" map was no easier to interpret, probably indicating that most difficulties are not due to inaccuracies of phasing. The "combined" map was definitely useful in regions close to the heavy atom sites, where there tend to be disturbances in the map obtained by isomorphous replacement.

In order to assess the performance of the heavy atom method, the phase angle data for a small number of reflections have been analyzed in detail. A summary of this analysis is given in Table III, which suggests that the mean phase error in the trypsin phases is about 50°.

Model Building and Refinement. An electron density map using all isomorphous and anomalous scattering data was computed at intervals of $a/64$, $b/64$, and $c/160$. Contours traced on transparent sheets as sections of constant y at a scale of 2.0 cm/Å were the basis for model building in a half-silvered mirror system (Richards, 1968). Model components were purchased from Cambridge Repetition Engineers.

Coordinates for parts of trypsin close to the active site and for residues 1'-93' of the inhibitor were recorded from the model using a plumb-line, and these were adjusted to conform to standard bond lengths and angles by a model building proce-

ture (Diamond, 1966). These coordinates were refined to the electron density by the real space refinement procedure (Diamond, 1971, 1974). In a first cycle of real space refinement, atomic radii were fixed, and only the overall scale factor, K , and the atomic positions were refined. A radius of 1.6 Å was chosen from preliminary trials. In a second cycle of real space refinement atomic radii were allowed to vary up to a maximum of 3.0 Å. Following the usual practice, the angle τ (C^α) was not constrained in either the model building or real space refinement calculations.

The coordinates resulting from the second cycle of refinement are given in Table IV. These coordinates contain a few τ (C^α) angles up to 30° from ideal values, the geometry of the disulfide bridges is imperfect, and some unreasonable atomic contacts exist, especially at the interaction of Lys-60 and Asp-1'. In the best defined parts of the map, however, including the active-site region, atomic positions are probably good to better than 0.4 Å. This impression is strengthened by the comparison, discussed below, of certain interatomic distances and conformational angles with those given for the PTI complex (see Tables VIII and IX).

Results

The three-dimensional structure of this complex can be considered at several levels of molecular organization. The structure will be described first as a crude dimeric aggregate and second as two complicated but independent molecules. Finally, the contact region between the enzyme and inhibitor molecules will be described in terms of atomic interactions, and we will present the evidence for our interpretation of the molecular conformation in the enzyme's active site.

Low-Resolution Structure. The 5-Å electron density map clearly showed two globular regions of high density in the asymmetric unit of the orthorhombic cell, one of which was recognized from its shape to be the trypsin molecule. The other more spherical molecule made contact over small areas with two trypsin molecules. One of the contact zones could be identified as the active site of the enzyme. The molecular complex between soybean inhibitor and trypsin, as demonstrated by the 5-Å map, is represented in Figure 4.

Porcine Trypsin Molecule. Interpretation of the trypsin part of the electron density map was facilitated by the obvious analogy with the known structures of bovine chymotrypsin (Birktoft and Blow, 1972), porcine elastase (Shotton and Watson, 1970), and bovine trypsin (Stroud *et al.*, 1974). The amino acid sequence of porcine trypsin (Hermodson *et al.*, 1973), which was kindly made available to us before publication by K. A. Walsh and H. Neurath, shows a high degree of homology with the bovine enzyme, and our studies confirm that the same disulfide linkages are made. The most significant differences between porcine and bovine trypsin, so far as the complex with soybean inhibitor is concerned, are the substitution of Tyr-39 in the bovine enzyme for Ser, and the substitution of Ser-217 in the bovine enzyme for Tyr in porcine trypsin.

The structure of the porcine trypsin molecule in the STI complex has not yet been critically compared to the other structures which are known. We have concentrated our efforts so far on the part of the molecule involved in binding the inhibitor. Even in this region, only qualitative comparisons with the other known structures can be reported at present.

There are two differences between trypsin and chymotrypsin for which these studies give further information. In chymotrypsin the buried residue Asp-102, involved in the charge relay system, is hydrogen bonded to His-57 and to the hydroxyl of Ser-214 through one of its carboxyl oxygen atoms. The other



FIGURE 4: Model of the STI complex, made from an electron density map at 5-Å resolution. The part representing trypsin is shaded less strongly.

carboxyl oxygen is hydrogen bonded to the main chain NH of His-57. In trypsin, Stroud *et al.* (1974) observed this carboxyl group to be oriented so that $N^{\delta 1}$ -(His-57) lies between the two carboxyl oxygen atoms, so that it is presumably equally hydrogen bonded to each of them. A similar situation has been noted by Huber *et al.* (1974) in the PTI-bovine trypsin complex, and we find the same in the STI-porcine trypsin complex. Since this point was not mentioned by Vandlen and Tulinsky (1973) in their studies on the chymotrypsin structure as pH is varied, it seems to be a difference between trypsin and chymotrypsin and not an effect of pH.

In trypsin there is a deletion at residue 218, which lies at the bottom of the substrate binding pocket in chymotrypsin. In the STI complex of pork trypsin there is rather clear evidence that the chain 216-217-219-220 forms a β bend of type I (Venkatachalam, 1968) in which CO(216) is hydrogen bonded to NH(220). In chymotrypsin, CO(216) is directed toward the solvent, almost 180° from its orientation in the β bend, and is implicated in secondary binding of a polypeptide substrate. Both Krieger *et al.* (1974), illustrating the binding of benzamide to bovine trypsin, and Huber *et al.* (1974), illustrating the binding of pancreatic trypsin inhibitor, show this carbonyl group in an intermediate position, about 90° away from either.

Inhibitor Molecule. The amino acid sequence of soybean trypsin inhibitor (Kunitz) contains 181 residues and is represented in Figure 5 (Koide and Ikenaka, 1973). It is quite different from the other trypsin inhibitors whose sequences are known. It contains two disulfide bridges, Cys-39'-Cys-86' and Cys-138'-Cys-145'. The peptide bond between Arg-63' and

TABLE IV: Coordinates for Residues 1'-93' of STI and the Adjacent Residues of the Trypsin Molecule.^a

[illegible]

	X	Y	Z		X	Y	Z		X	Y	Z		X	Y	Z		X	Y	Z		X	Y	Z
N	-30.2	-2.2	12.1	CZ	-13.2	-13.4	-8.6	CG	-31.8	-5.1	13.2	C	1.7	-3.6	18.5	O	-5.2	2.0	18.1	CG2	-17.7	-18.4	-3.3
CA	-28.9	-2.8	11.9	NE	-14.2	-13.5	-8.8	CB	-31.6	-6.3	14.1	O	2.7	-4.2	18.9					CA	-18.1	-20.1	-1.7
C	-28.2	-4.0	12.6	CD	-15.9	-13.7	-8.3	CA	-30.9	-7.4	13.3									C	-17.3	-21.1	-2.6
O	-28.3	-4.1	13.8	CG	-16.2	-12.3	-8.0	C	-31.2	-8.5	14.8	62' TYROSINE				72' PROLINE				O	-16.1	-20.8	-2.9
				CB	-17.6	-12.5	-7.4	O	-31.4	-8.3	15.6	N	1.1	-2.7	19.3	N	-3.0	2.4	17.7				
27' PHENYLALANINE				CA	-17.8	-11.8	-6.1					OE8	3.0	3.7	20.1	CD	-1.6	2.6	18.1	84' METHIONINE			
N	-27.4	-4.8	11.9	C	-17.7	-13.0	-5.2	50' LEUCINE				CD2	0.5	1.1	20.8	CG	-0.7	2.0	17.0	N	-17.9	-22.2	-3.0
CD2	-26.4	-4.6	13.3	O	-17.7	-14.2	-5.5	N	-31.1	-9.8	13.9	CE2	1.0	2.4	20.4	CB	-1.6	2.0	15.8	CE	-17.8	-27.3	-6.9
CE2	-26.4	-10.0	13.9					CD2	-33.5	-9.8	13.5	CZ	2.8	2.5	20.4	CA	-3.0	1.9	16.3	SD	-18.8	-26.6	-5.6
CD	-26.7	-11.0	13.1	39' CYSTINE				CD1	-34.7	-11.5	14.8	CE1	3.3	1.4	20.6	O	-3.6	0.4	16.3	CG	-17.6	-25.4	-5.0
CB1	-26.9	-10.8	11.7	N	-16.8	-12.6	-4.1	CG	-33.6	-11.3	13.7	CD1	2.7	0.1	20.9		-2.9	-0.5	17.0	CB	-18.3	-28.2	-4.8
CD1	-26.9	-9.5	11.2	SG	-15.1	-16.2	-3.1	CB	-32.3	-11.9	14.2	CG	1.3	0.0	21.0	73' LEUCINE				CB	-17.2	-23.2	-4.8
CB	-26.7	-8.4	12.0	CB	-15.2	-14.4	-3.6	CA	-31.3	-10.9	14.6	CB	0.7	-1.4	21.3	N	-4.7	0.2	15.7	C	-16.4	-22.6	-4.9
CA	-26.7	-7.1	11.4	CA	-16.3	-13.6	-3.1	C	-29.8	-11.4	14.8	CA	1.5	-2.5	20.7	CD2	-8.0	-0.4	18.3	O	-16.6	-22.9	-6.1
CB	-26.7	-5.9	12.4	C	-16.4	-12.2	-2.5	O	-29.5	-12.5	15.0	C	1.3	-3.8	21.4	CD1	-7.5	1.2	16.5				
C	-25.3	-5.5	12.7	O	-15.8	-11.3	-3.0					O	0.3	-4.5	21.2	CG	-7.0	0.0	17.3	85' LEUCINE			
O	-25.0	-4.3	12.9					51' ASPARTATE				CB	-6.7	-1.1	16.3	CB	-6.7	-1.1	16.3	N	-15.4	-21.8	-4.5
				40' PROLINE				N	-29.0	-10.4	14.4	CA	-5.3	-1.1	15.6	C	-5.4	-1.6	14.2	CD2	-13.1	-20.0	-7.6
28' GLYCINE				N	-17.2	-12.1	-1.5	OD2	-25.3	-11.1	17.1	C	-5.4	-1.6	14.2	O	-5.1	-0.8	13.3	CB	-14.9	-18.2	-7.3
N	-24.4	-6.4	12.8	CD	-17.9	-13.2	-0.7	OD1	-25.0	-10.0	15.3	NEH2	7.4	-9.2	19.9					CG	-14.6	-19.7	-7.6
CA	-23.0	-6.1	13.1	CG	-18.2	-12.5	-0.6	CG	-25.7	-10.8	16.0	NEH1	6.3	-9.9	21.8	74' SERINE				CB	-15.3	-20.6	-6.6
C	-22.3	-6.0	11.7	CB	-18.4	-11.0	-0.2	CB	-27.0	-11.3	15.6	CZ	6.6	-9.0	20.8	O	-5.7	-2.8	14.0	CA	-14.5	-21.2	-5.5

BIOCHEMISTRY, VOL. 13, NO. 20, 1974 4219

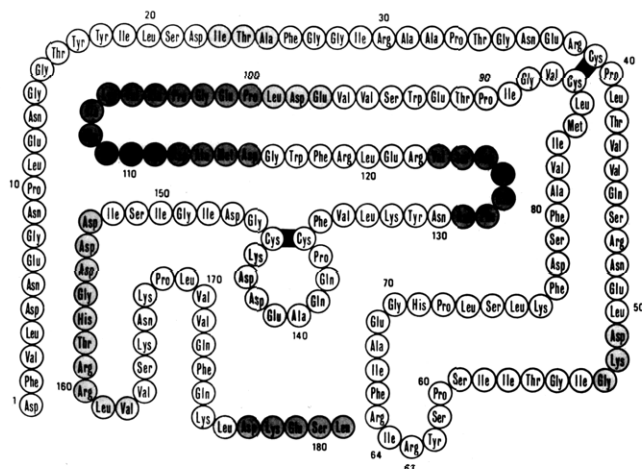


FIGURE 5: Amino acid sequence of STI. The weaker shading shows regions where the electron density is weak, but could be followed. The densely shaded amino acids could not be located (published by permission of T. Ikenaka).

Ile-64' is preferentially and reversibly split by trypsin (Ozawa and Laskowski, 1966).

The sequence around residues 63' and 64' was easily identified in the high-resolution map as an external loop of electron density which enters the active site of trypsin (Figure 6). It provided a starting point from which the chain could be followed in both directions to the Cys-39'-Cys-86' disulfide bridge and to the amino terminus. The identification of the sec-

ond disulfide showed that very little density in the map corresponds to the exposed hydrophilic sequence of the short loop linking Cys-136' to Cys-145'. This disulfide loop is known to be readily available to attack by reducing agents, without an effect on inhibitor activity (Di Bella and Liener, 1969).

There are other parts of the structure where very little electron density is seen for hydrophilic sequences of up to eight residues, especially residues 107'-115', 123'-129', and the terminal sequence 177'-181', which could not be assigned to electron density. The first two sequences form part of an extended region of hydrophilic amino acids 92'-127', which was pointed out by Koide and Ikenaka (1973). They apparently constitute external loops on the inhibitor molecule. The presence of lysine or arginine at residues 106', 111', 122', and 178' suggests that proteolysis of the inhibitor may have occurred during crystallization. Acrylamide gel electrophoresis of reduced samples of crystals available after the electron density maps had been studied showed the existence of some peptide fragments in some crystal specimens, but the results were not reproducible.

The polypeptide chain from 1' to 106' and from 130' to 176' could be traced with confidence. Another short stretch of density, which appears clearly in the map, was tentatively assigned to residues 116'-122'. Figure 7 represents the folding of the chain of the inhibitor molecule.

The overall structure of the soybean inhibitor is that of a sphere about 35 Å in diameter, made of criss-crossing loops wrapped around a core of hydrophobic side chains. These loops are arranged in such a way that the amino-terminal half of the polypeptide chain forms the bottom part of the sphere, while

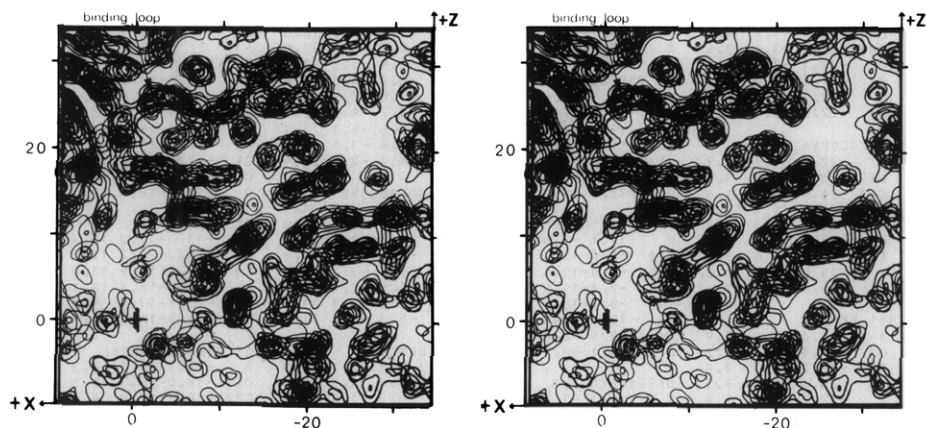


FIGURE 6: Stereoview of part of the electron density map, showing the active site of the STI complex. The main features may be identified by reference to Figure 3, which covers approximately the same region of the map.

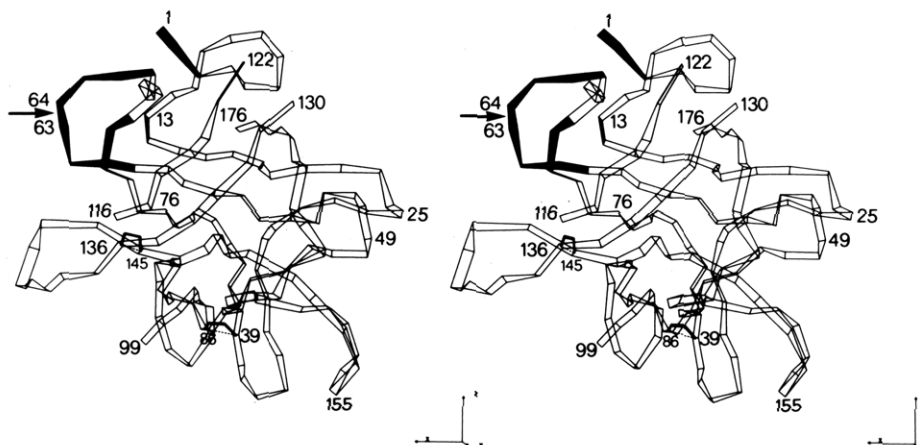


FIGURE 7: The chain folding of STI. Residues which are in contact with trypsin are blackened. The assignment of residues 116'-122' is tentative.

TABLE V: Contacts between Trypsin and STI.^a

AA in Trypsin	Asp-1'	Phe-2'	Asn-13'	P ₄ Pro-60'	P ₃ Ser-61'	P ₂ Tyr-62'	P ₁ Arg-63'	P ₁ ' Ile-64'	P ₂ ' Arg-65'	P ₃ ' Phe-66'	His-71'	Pro-72'	Total Contacts
Ser-39									2	4			6
His-40									2, 5	3			2, 8
Phe-41		13						4	1	2			20
Cys-42								1					1
His-57						21	4	10			1, 8		1, 43
Cys-58								4					4
Lys-60	<i>1*</i> , 20*** ^b												1, 20
Asn-97						<i>1, 8</i>						2	1, 10
Leu-99					3	17*							20
Asp-102						5							5
Tyr-151								1	15				16
Asp-189							2*, 14						2, 14
Ser-190							<i>1*</i> , 19*						1, 19
Cys-191							12						12
Gln-192			<i>1, 9</i>	<i>1, 1</i>		2	14	<i>1, 5</i>	1				3, 32
Gly-193							<i>1, 2</i>	<i>1, 4</i>	1				2, 7
Asp-194							1						1
Ser-195							<i>1, 12</i> ^c	8					1, 20
Val-213							4*						4
Ser-214						4	<i>1, 3</i>						1, 7
Trp-215					10	5	<i>1*</i> , 22*						1, 37
Gly-216					<i>1, 1</i>		<i>1, 9</i>						2, 10
Tyr-217					2								2
Cys-220							1						1
Total	<i>1, 20</i>	13	<i>1, 9</i>	<i>1, 1</i>	<i>1, 16</i>	<i>1, 62</i>	8, 117	2, 37	2, 25	9	<i>1, 8</i>	2	18, 319 contacts

^a The table indicates the number of possible *hydrogen bonds* (in italics) and the number of other interatomic contacts between amino acids of trypsin and amino acids of STI. The hydrogen bonds are itemized in Table VI. An interatomic contact is counted for every pair of atoms which is within 0.5 Å of the theoretical van der Waals' contact distance. For each interatomic distance which is too short by 0.5 Å or more, an asterisk is recorded, indicating that a further adjustment of the coordinates is needed.

^b The side chains of Lys-60 and Asp-1' were drawn unreasonably close together by the real-space refinement procedure, resulting in four unacceptably short contacts. There are probably only three or four true contacts. ^c Covalent bond between O^γ(195) and C(63'). The neighbor and next-nearest neighbor interactions have not been included in the number of contacts.

the carboxyl-terminal section is mostly on the top. No α helix can be found, in agreement with optical studies (Jirgensons, 1965; Ikeda *et al.*, 1968). Most of the chain is involved in approximate β -pleated sheet structures, but little regular sheet is formed. The disulfide bridges are both on the surface; Cys-39'-Cys-86' is on the bottom and Cys-138'-Cys-145' is on the top of the molecule. Both bridges are more than 15 Å from the region of the inhibitor which binds to the active site of trypsin.

Region of Interaction. The high-resolution map shows that only about 12 amino acids out of the 181 which compose the inhibitor make contact with the trypsin molecule. These residues are indicated by blackening in Figure 7. Most of the contacts involve the sequence Ser-61'-Tyr-62'-Arg-63'-Ile-64'-Arg-65'-Phe-66' which forms a curved loop protruding from the main body of the inhibitor molecule. This "binding loop" appears extremely well defined in the high-resolution map, which indicates that it is strongly immobilized by its interactions with the enzyme (Figure 8). It includes the bond Arg-63'-Ile-64' shown to be hydrolyzed by the trypsin-catalyzed conversion of the virgin inhibitor to its modified form (Ozawa and Laskowski, 1966). A summary of all the contacts made between enzyme and inhibitor is given in Table V, and Figure 9 is

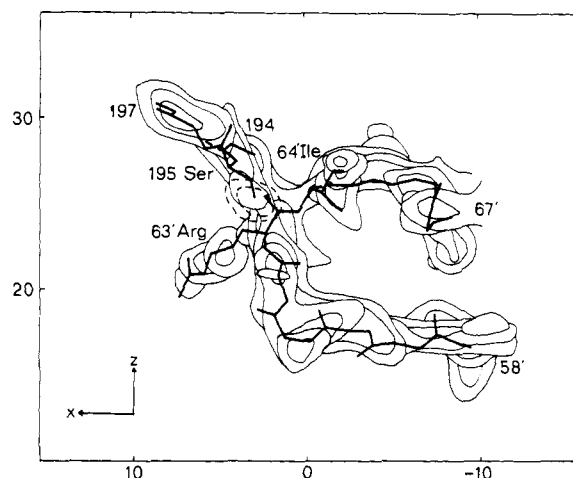


FIGURE 8: The "binding loop," as it appears in the electron density map, showing how the main chain of residues 194-197 and 58'-67' fit to the observed density. The side chains of Ser-195, Arg-63', and Ile-64' are shown. This feature can be seen in the upper left quadrant of Figure 6. The dashed lines show the density of the imidazole of His-57, which lies above Ser-195.

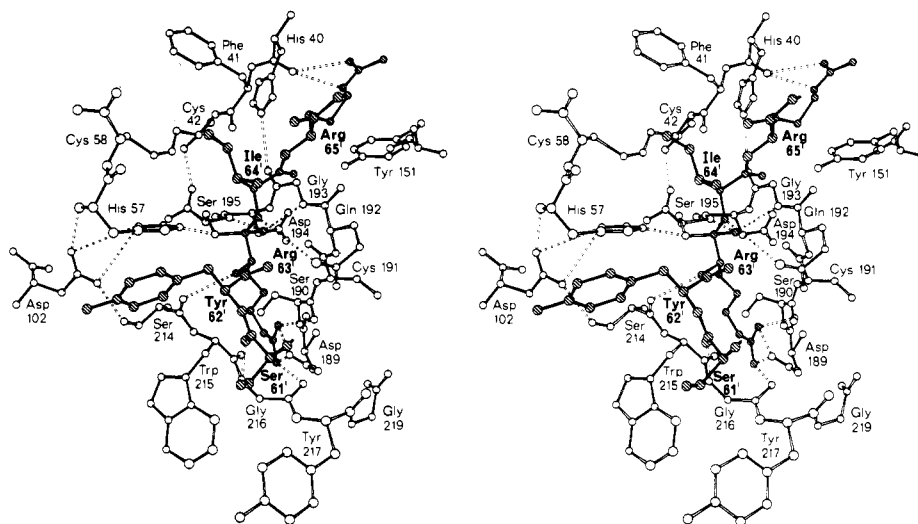


FIGURE 9: Schematic drawing of atoms at the active site of the STI complex. Atoms of the inhibitor molecule are shaded, and the carbonyl carbon of the scissile peptide bond 63'-64' is shown tetrahedrally coordinated. Trypsin residues Asn-97 and Leu-99 which lie in front of Asp-102 and touch the distal part of Tyr-62' have been omitted for clarity. The orientation of this diagram corresponds to published views of the active sites of α -chymotrypsin and trypsin (Birktoft and Blow, 1972; Stroud *et al.*, 1974). The view is toward the trypsin molecule from inside the inhibitor.

a schematic drawing showing the main interactions. Table VI summarizes the polar interactions.

The side chain of Arg-63' occupies its expected position in the primary specificity pocket of the trypsin molecule, as predicted from the formyl-L-tryptophan complex of chymotrypsin (Steitz *et al.*, 1969), and from the benzamidine complex of trypsin (Krieger *et al.*, 1974). The two N^H atoms of the guanidinium group are both within hydrogen-bonding distance of the carboxylate of Asp-189. One of them probably also forms a hydrogen bond to the carbonyl of Gly-216, and the other to the carbonyl of Ser-190. The peptide NH of Arg-63' makes a hydrogen bond to CO(214). Because of the deep invagination of the primary binding pocket, and its snug fit to the arginine side chain, Arg-63' makes about 40% of the interatomic contacts counted in Table V.

Residue Tyr-62' occupies position P_2 (notation of Schechter and Berger, 1967). The phenolic part of the side chain lies between the side chains of Leu-99 and His-57, parallel to the imidazole ring of the latter. It also makes contact with the side chain of Asn-97, to which a hydrogen bond is formed. These interactions are likely to provide one contribution to the difference spectrum observed on binding soybean inhibitor to trypsin (Edelhoc and Steiner, 1965). The possibility of an interaction of this kind by a bulky group in P_2 has been noted by Segal *et al.* (1971).

The carbonyl of Ser-61' (P_3) forms a hydrogen bond to NH (Gly-216). A similar interaction was found by Segal *et al.* (1971) for acetylalanylalanylphenylalanyl chloromethyl ketone with γ -chymotrypsin. In γ -chymotrypsin the amido group in P_3 was found to make a hydrogen bond to CO(216). This is not possible in the STI complex (and probably in other trypsin complexes) because of the different orientation of the peptide bond 216-217, discussed in a previous section.

Before dealing with the conformation of the scissile bond itself, the interactions made on the leaving group side will be discussed. Ile-64' (P_1') is in contact with the disulfide bridge Cys-42-Cys-58. The longer branch of its side chain makes contact with both sulfur atoms, forming a hydrophobic region which also includes Phe-44, Phe-2', Pro-10', and Phe-66'. In the complex with bovine trypsin, this hydrophobic region would include further interactions with Tyr-39, but in porcine trypsin residue 39 is Ser. Arg-65' (P_2') lies in a shallow groove in the surface

of the trypsin molecule with its guanidinium group lying on top of Tyr-151, and forming hydrogen bonds to CO (Phe-41). The interaction with Tyr-151 is probably a charge transfer interaction, which will have a considerable effect on the binding spectrum (Edelhoc and Steiner 1965) though it is unlikely to increase the binding energy greatly. There is no charged group on the enzyme to balance the charge on Arg-65', which also interacts with Glu-12'.

It is interesting that an almost identical interaction is found between Tyr-42 α_1 and Arg-40 β_2 of hemoglobin. Dr. M. F. Perutz (personal communication) has shown that the difference spectrum observed on association of hemoglobin dimers has a maximum at 298 nm and is extremely similar to the difference spectrum shown on association of STI with trypsin.

A unique conformation for Gln-192 cannot be assigned from the electron density map which shows a strong rod of density only 3 Å long for the side chain. The homologous residue Met-192 in chymotrypsin was found to have a variable side-chain conformation, depending on substrate binding interactions. There is little room for the Gln side chain in the STI complex, and the most satisfactory conformation appears to be with χ_2 near 0°, so that C^β and C^α are almost eclipsed. The most probable hydrogen bonds for the Gln-192 side chain are those given in Table V—a hydrogen bond from the side-chain NH_2 to CO(60'), and to the side-chain CO from the side chain of Asn-13'. This side-chain CO(192) could also possibly accept a hydrogen bond from the NH of the scissile bond, NH(64').

The orientation of the peptide bond between Ile-64' and Arg-65' presents a problem. It seems to be clearly indicated in the map, but in the orientation given there is no suitable hydrogen bond for NH(65'). Rotation of this peptide group through 90° makes possible hydrogen bonds to CO(64') from the amide group of Asn-13' and from NH(65') to CO(41). In the given orientation CO(41) makes no hydrogen bonds.

Outside the binding loop (Ser-61'-Phe-66'), two interactions made by the inhibitor deserve special mention. The carboxyl group of Asp-1' forms an ion pair with the ϵ -amino group of Lys-60. This is the only ion pair interaction between the two molecules outside the primary binding site. $N^{\epsilon}H$ -(His-71') forms a hydrogen bond to CO-(His-57). Chemical and proton magnetic resonance data (Mattis and Laskowski, 1973; Markley, 1973) indicate that cleavage of the bond between Arg-63'

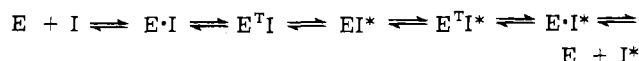
and Ile-64' affects the pK of a histidine residue on the inhibitor. This must be His-71', which is 6 Å from the scissile bond in the complex, as the only other histidine residue is on the other side of the molecule.

The binding loop makes few interactions with the rest of the inhibitor molecule. In the absence of trypsin, all side chains in the binding loop would be exposed to solvent. Only one hydrogen bond can be assigned linking the main chain of this loop to the rest of the inhibitor molecule, between CO-(Tyr-62') and N^δ-(Asn-13'). The ionic interaction mentioned above, between Glu-12' and Arg-65', may not persist in the free inhibitor, as the arginine side chain might be able to rotate.

There is only one possible interaction between trypsin and the parts of the inhibitor which we have not been able to interpret with confidence. This is a probable contact between the side chain of Tyr-217 and a residue tentatively identified as Gly-116'. In bovine trypsin residue 217 is serine and the contact could not be made.

Active Site. There is good evidence that STI acts like a true substrate of trypsin (Ozawa and Laskowski, 1966; Finkenshtadt and Laskowski, 1967; Niekamp *et al.*, 1969; Hixson and Laskowski, 1970). Trypsin inhibitors differ from normal substrates in their exceptionally high binding constant, and because in many cases the catalytic reaction is blocked before hydrolysis of the inhibitor occurs. In the case of STI, however, the bond between Arg-63' and Ile-64' is cleaved quite rapidly. Trypsin catalyzes the hydrolysis of this bond in the virgin inhibitor (I) and its synthesis in the "modified" inhibitor (I*), under conditions where the cleavage is thermodynamically reversible.

The crystalline complex can therefore be either a nonbonded association between the enzyme and inhibitor molecules (Michaelis complex), E·I, or between the enzyme and the modified inhibitor, E·I*, or an intermediate in the chemical steps of catalysis. The generally accepted mechanism of action of serine proteases involves an acyl-enzyme, in which the carbonyl group of the scissile bond forms an ester linkage with the side chain of the catalytically essential residue Ser-195. Each step of the reaction is assumed to proceed by a reaction in which this carbonyl carbon and its three ligands are attacked, approximately perpendicular to their plane, by a nucleophile, so that there is an intermediate state in which this carbon atom has four ligands, approximately tetrahedrally arranged. The sequence of events could be written



where I* is written when the peptide bond is broken and superscript T signifies a tetrahedral intermediate. (The "deacylation" steps to the right of EI* involve attack by a water molecule, which is not indicated explicitly in the above scheme.) A stereochemically reasonable model for the way in which these steps might proceed was given by Henderson *et al.* (1971). Kinetic analysis involving model substrates has given no convincing evidence that the tetrahedral intermediate is a kinetically significant species, in reactions involving the trypsin family of enzymes. However, the crystallographic work of Rühlmann *et al.* (1973) led to the surprising conclusion that the dominant species present in the crystalline complex of trypsin with PTI was a tetrahedral adduct.

At a resolution of 2.6 Å one does not expect to be able to assign individual positions for every atom in the structure; on the other hand, as inspection of Figure 8 will confirm, there can be clear evidence about the conformation of the polypeptide chain, and of side chains. This imposes especially strong re-

TABLE VI: Polar Interactions between Trypsin and Inhibitor.

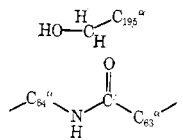
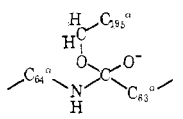
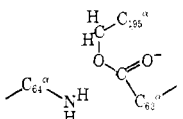
Amino Acids Involved		Interacting Groups ^a		Note
Inhibitor	Trypsin	Inhibitor	Trypsin	
Asp-1	Lys-60	-COO ⁻ ...	+H ₃ N-	
Asn-13	Gln-192	-CONH ₂ ...	OCNH ₂ -	
Pro-60	Gln-192	CO...	H ₂ NCO-	c
Ser-61	Gly-216	CO...	HN	
Tyr-62	Asn-97	-OH...	OCNH ₂ -	
		NH ₂ ...	O	
Arg-63	Asp-189	-C	+ - C-	
		NH ₂ ...	O	
	Ser-190	Guanidinium ⁺	NH ₂ ...OC	b
	Gly-193	CO...	HN	
	Ser-195	CO...	HN	
	Ser-214	NH...	OC	
	Trp-215	Guanidinium ⁺	NH...OC	b
	Gly-216	Guanidinium ⁺	NH ₂ ...OC	c
Ile-64	Gln-192	NH...	OCNH ₂ -	c
	Gly-193	CO...	HN	b
Arg-65	His-40			

^a For typographical convenience, main chain groups are recorded simply as NH, CO. ^b Short distance between these groups, but bond directions make hydrogen bond doubtful. ^c Groups are well directed, but coordinates indicate distances which are slightly too long for a favorable hydrogen bond. ^d Charge-transfer interaction.

strictions on the positions of the C^α atoms, being the points where the side chains branch from the main chain, and often on the positions of main chain carbonyl groups, which tend to produce "swellings" of high density in the main chain. The real space refinement procedure, which achieves a best fit to the observed electron density by rotation about dihedral angles, while maintaining bond lengths and the majority of interbond angles at their theoretical values, is designed to make an accurate and objective analysis of an electron density map of this kind. In the main real-space refinement operations, the scissile bond was treated as a standard peptide unit. O^γ-(Ser-195) was located in the position indicated by the electron density, which corresponds roughly to its orientation in "acyl"-enzymes (that is, the orientation observed in indoleacryloyl-chymotrypsin, tosyl-chymotrypsin, tosyl-elastase, and in diisopropylphosphoryl-trypsin).

In order to obtain clear evidence about the dispositions of atoms at the active site, a separate cycle of real space refinement was performed, in which the peptide bond between Arg-63' and Ile-64' was deleted. Under these circumstances there was no constraint on the distance between C^α(63') and C^α(64'). This distance would be chiefly determined by the need to fit the side chains into density, and to maintain continuity with the adjacent parts of the polypeptide chain. The program was run in such a way that the electron density, used to guide

TABLE VII: Interatomic Distances at Different Reaction Steps.

	Expected Distances at Various Reaction Steps (Å)	
	C ^β (195)–C ^α (63')	C ^α (63')–C ^α (64')
Michaelis complex		
	5.2	3.8
Tetrahedral intermediate		
	3.7	3.8
Acyl-enzyme		
	3.7	5.0
	Distance Actually obsd (Å)	
	3.7	3.9

the refinement, is not modified by parts of the structure which have already been refined, so that the three most important amino acids, Ser-195, Arg-63', and Ile-64', were positioned independently of one another.

The resulting distances between C^α(63'), C^α(64'), and C^β(195) are given in Table VII. The latter atom is, of course, fixed entirely by the main chain atoms next to it, and it cannot be disturbed by side-chain rotations. The chief structural consequence of breaking a bond is that nonbonded atoms need to be further apart, usually by more than 1 Å than the covalent bond distance. This difference is large enough to be detected easily from a good 2.6-Å electron density map, and the result appears to show clearly that the observed structure is a tetrahedral adduct. This had already been suspected, because in normal refinement (when the scissile bond was treated as an ordinary peptide bond) its carbonyl C had come within 1.5 Å of O^γ(195), and because of the difficulty of obtaining satisfactory interatomic distances in models for any of the other possibilities.

The evidence that the crystal structure is a tetrahedral adduct is strong, but indirect. At this resolution we have no evidence whether the bond lengths around the tetrahedral carbon

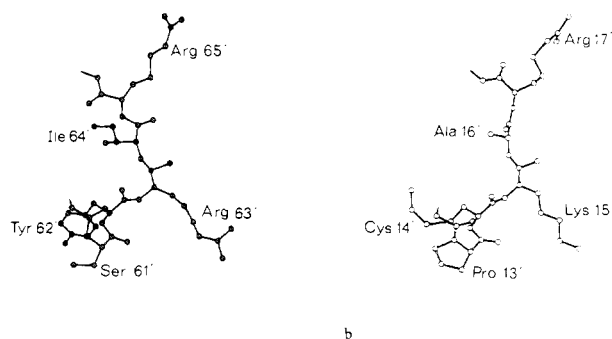


FIGURE 10: Comparison of active-region chains in STI and PTI (Rühlmann *et al.*, 1973). The viewpoint of these diagrams is further to the right than that in Figure 9. If the complex were shown, the enzyme would occupy the right-hand side of the diagram.

TABLE VIII: Main Chain Conformation at the Active Sites of PTI and STI Complexes.

	PTI ^a	Site	STI	
Pro-13'	$\phi = -86$ $\psi = -28$	P ₃	$\phi = -46$ $\psi = -21$	Ser-61'
Cys-14'	$\phi = -64$ $\psi = 152$	P ₂	$\phi = -77$ $\psi = 136$	Tyr-62'
Lys-15'	$\phi = -116$ $\psi^b = 87$	P ₁	$\phi = -89$ $\psi^b = 85$	Arg-63'
Ala-16'	$\phi^b = -135$ $\psi = 166$	P ₁ '	$\phi^b = -135$ $\psi = -115$	Ile-64'
Arg-17'	$\phi = -116$ $\psi = 87$	P ₂ '	$\phi = -175$ $\psi = -179$	Arg-65'

^a R. Huber and collaborators, personal communication.

^b From refinement where the scissile bond was assumed to be a planar peptide bond.

are abnormal. In preparing Table IV the coordinates have been adjusted to give a tetrahedral conformation about C(63') and normal bond lengths. In order to achieve this, bond angles at O^γ(195) and N(64') had to be increased slightly. Huber *et al.* (1974) suggest that the C(63')–O^γ(195) bond is longer than normal, and that the PTI complex is in a state intermediate between the acyl-enzyme and the tetrahedral adduct. We do not feel that our results are sufficiently accurate to give evidence on this point.

In this complex, N^ε(57) is within hydrogen bond distance of O^γ(195). The distance N^ε(57)–N(64') is 4.0 Å, approximately a contact distance when the proton which is presumably localized near N^ε(57) is allowed for. A significant movement of the histidine side chain would be needed to bring them within hydrogen bond distance. Presumably such a movement must occur before the bond C(63')–N(64') breaks. Movements of the required magnitude have been observed (between native and indoleacryloyl-chymotrypsin, for example (Henderson, 1970)).

Discussion

STI and PTI Bind Like Good Substrates of Trypsin. All of the interactions which would be expected by the P₃, P₂, and P₁ amino acids of a good trypsin substrate on the basis of available structural data are found in STI: the hydrogen bond to CO(61') from NH(216), the interactions made by the phenolic side chain of Tyr-62' with His-57, Asn-97, and Leu-99, and all the interactions made by Arg-63'. No structural data are available about P₁' for good substrates, but model building (Blow *et al.*, 1972), originally guided by the PTI structure (Huber *et al.*, 1970), suggested the type of interaction observed for Ile-64', and kinetic data are consistent with this type of interaction (Fersht *et al.*, 1973).

Extremely similar interactions are made by the analogous residues of PTI (Pro-13'–Cys-14'–Lys-15'–Ala-16') despite the lack of any homology between the two segments of sequence. The similarity of conformation is confirmed by comparison of main chain dihedral angles between the two inhibitors (Table VIII), and is illustrated in Figure 10. A change in ($\psi_i - \phi_{i+1}$) rotates the plane of the peptide bond between two residues, but has little effect on the structure as a whole; a change in ($\psi_i + \phi_{i+1}$) alters the orientation of a side chain. The assignment of ($\psi_i - \phi_{i+1}$) is thus subject to considerable error, while ($\psi_i + \phi_{i+1}$) can be determined more precisely.

TABLE IX: Comparison of Various Distances in the STI and PTI Complexes.

Porcine Trypsin-STI ^a	Distance (Å)	Bovine Trypsin-PTI ^b	Distance (Å)
Trypsin Conformation			
N ^ε 2(His-57)-O ^γ (Ser-195)	2.8	N ^ε 2(His-57)-O ^γ (Ser-195)	2.6
N ^δ 1(His-57)-O ^δ 1(Asp-102)	3.5	N ^δ 1(His-57)-O ^δ 2(Asp-102)	3.2
N ^δ 1(His-57)-O ^δ 2(Asp-102)	3.2	N ^δ 1(His-57)-O ^δ 1(Asp-102)	2.9
Catalytic Site			
O ^γ (Ser-195)-C(Arg-63')	1.4 ^c	O ^γ (Ser-195)-C(Lys-15I)	2.3
N ^ε 2(His-57)-N(Ile-64')	3.7	N ^ε (His-57)-N(Ala-16I)	4.1
N(Gly-193)-O ⁻ (Arg-63')	3.0	N(Gly-193)-O ⁻ (Lys-15I)	2.7
N(Ser-195)-O ⁻ (Arg-63')	3.4	N(Ser-195)-O ⁻ (Lys-15I)	2.7
Primary Specificity			
O ^δ (Asp-189)-N ^η 2(Arg-63')	2.2	O ^δ 1(Asp-189)-N ^ε (Lys-15I)	3.3
O ^δ (Asp-189)-N ^η 1(Arg-63')	3.0	O ^δ 2(Asp-189)-N ^ε (Lys-15I)	3.3
Secondary Specificity			
O(Ser-214)-N(Arg-63')	3.2	O(Ser-214)-N(Lys-15I)	3.3
N(Gly-216)-O(Ser-61')	3.3	N(Gly-216)-O(Pro-13I)	3.3

^a From the coordinates of Table IV. ^b Huber *et al.* (1974). ^c Constrained to covalent bond length.

Table VIII shows that in the three peptide bonds between P₃ and P₁' ($\psi_i + \phi_{i+1}$) agrees to within 11° between the two inhibitors, which is doubtless at the level of experimental error.

There is a further striking similarity between PTI and STI in the presence of arginine at P₂' (Arg-65' in STI). Although the preceding peptide bond is quite differently oriented in the two inhibitor complexes, the guanidinium group reaches a similar orientation in each. In both complexes there is a hydrogen bond from CO(40) to the guanidinium group, which lies in contact with the phenolic part of Tyr-151 and roughly parallel to it.

Further evidence of the close analogy between the PTI and STI complexes is given in Table IX, where various distances given in published accounts of the PTI complex are compared with analogous distances taken from our coordinates for the STI complex.

Binding Constant of the Complex. Perhaps the most important question raised by the trypsin inhibitors is "Where do they derive the energy for their extremely tight binding?" The answer to this question may lead to some insight into protein-protein interactions in general.

It seems evident that the binding energy does not derive from some new and unforeseen type of interaction, but from the sum of small energy terms derived from many interactions. In view of the limited area of contact observed in both the PTI and STI complexes, it is at first sight surprising that they should amount to so much. The strongest binding constant known for a trypsin inhibitor is that of PTI to trypsin, and corresponds to $\Delta F_{\text{binding}} = -18.5$ kcal/mol (Vincent and Lazdunski, 1972). The dissociation constant of 10^{-11} M for STI (Laskowski and Sealock, 1971) leads to $\Delta F = -15$ kcal/mol. Calorimetric measurements have given a positive enthalpy of +3 kcal/mol for formation of the STI complex from free STI and trypsin (Baugh and Trowbridge, 1972), and thus the $T\Delta S$ contribution to ΔF must be -18 kcal/mol.

This positive enthalpy term is easier to understand in the light of the evidence presented by Rühlmann *et al.* (1973) and here that these complexes exist as tetrahedral adducts. Formation of the covalent bond from the Michaelis complex would normally be expected to absorb energy. Evidently it is stabilized by other forces, of entropic origin, which favor the tetra-

hedral form because the two molecules "fit" each other better. In Table X, we present a rough "account" of these forces, and try to show how they may "balance" the reckoning of these energies. The numbering of the following paragraphs corresponds to numbered entries in the table.

(1) Before we discuss the entropy-driven effects, there is one source of energy which is primarily enthalpic to be considered. There is good evidence that in free PTI, significant strain is associated with peptide bonds at the active site, which could be relieved on forming the complex (Deisenhofer and Steigemann, 1974). This strain energy, estimated as 5 kcal/mol, appears to be a consequence of the folding of the PTI molecule as a whole.

TABLE X: Energetics of Binding Soybean Trypsin Inhibitor to Trypsin.

	Contribution to Binding Energy (kcal/mol)
(1) "Strain" of peptide chain, relieved on binding	0 (-5, PTI)
(2) Interactions at primary binding site, P ₁	-4
(3) Interactions at P ₂ and P ₃	-2
(4) Interactions at P ₁ '	-2.5
(5) Interactions elsewhere	Ca. -4
(6) No loss of internal degrees of freedom	Unknown negative contribn
(7) Tetrahedral ligands at scissile bond	Unknown positive contribn
(8) More favorable charge distribution, including better hydrogen bonding to tetrahedral oxyanion	Unknown negative contribn
Total binding energy	
(entropic)	-18
(enthalpic)	+3
ΔF	-15 (-18.5, PTI)

We have no information about the structure of free STI, but in view of the relatively loose attachment of the binding loop to the rest of the molecule, such an effect seems unlikely to exist.

(2) K_m for a simple trypsin substrate that utilizes only the primary binding site P_1 , like benzoylargininamide, is 2×10^{-3} M (Harmon and Niemann, 1949) corresponding to $\Delta F_{\text{binding}} = -4$ kcal/mol. This binding energy is composed of two components which almost cancel each other. A negative term results from favorable interactions with the enzyme in P_1 ; a positive term results from the loss of rotational and translational freedom when the substrate is bound. The latter term is relatively large for a small substrate, and increases comparatively slowly as its size is increased (Page and Jencks, 1971). This explains why a small net binding energy results from the many interactions in the primary binding site.

(3) The binding energy increases as the P_2 and P_3 sites are occupied. Segal (1972) studied the kinetic constants of chymotrypsin chloroketones which use the P_2 and P_3 sites, and concluded that these probably provide an additional -2 kcal/mol of binding energy. (This would include the formation of one hydrogen bond involving CO(216), which occurs in chymotrypsin-substrate complexes, but not in those with trypsin.)

(4) Fersht *et al.* (1973) have shown that the attack of alanine at P_1' enhances the rate of deacylation of acyl-chymotrypsin, by comparison with water, by a factor of 50–100 times. Assuming that the interactions made by the alanine moiety go to reducing the activation energy of the reaction, and that other interactions are the same for alaninamide or water, this suggests a binding energy $\Delta F = RT \ln (k_{\text{water}}/k_{\text{alaninamide}}) = -2.5$ kcal/mol.

(5) Referring to Table V, it is seen that the interactions so far considered, involving residues 61'–64', account for about two-thirds of all the observed interactions in the complex, and the total free energies noted above may be increased by a further 50% to account in a crude fashion for the interactions made by the rest of the molecule. (For reasons noted under 2, this is an underestimate.)

(6) There is an important effect which leads to underestimates under headings 2, 3, 4, and consequently 5. Under headings 2 and 3 we are using the binding energies of amino acid and peptide substrates to estimate contributions to the binding energy of a firmly folded protein molecule to the same sites. In binding small substrates to these sites *internal rotational* freedom is lost, in addition to the loss of overall translational and rotational freedom. Although, as Page and Jencks (1971) point out, this effect is relatively small for a small molecule reactant, it is certainly not negligible for the acetyl tripeptides involved in 3. Similarly under heading 4, we are comparing a free alaninamide group to a group which is already in the proper position and orientation because it is part of the STI molecule. We make no attempt to estimate the contribution of these orientation effects to the binding energy. If the folded STI molecule were perfectly oriented to make these interactions, and all internal freedoms were completely suppressed, the effect would be a maximum, but there is no evidence about how closely this is realized in practice. Table X shows that only a small contribution is needed from this source to give a satisfactory account of the total binding energy.

(7) The energetics of formation of the tetrahedral adduct itself are discussed in the following section.

Contributions 2–6 will be predominantly entropic in character. We have not attempted to assess the contribution of hydrogen bonds to the binding energy of STI to porcine trypsin, but it is likely to be slight. When the complex forms, none of the potential hydrogen-bond donors or acceptors is buried alone (or

at most one pair; see the section on Region of Interaction above). Since surface polar groups in the isolated molecules are likely to be hydrogen bonded to water molecules, little free energy will be realized by complementary hydrogen-bond formation in the complex (Kauzmann, 1959). However, formation of the complex will reduce the total surface area of the molecules and consequently the volume of the potential hydration sphere. Chothia (1974) has recently estimated this effect to be 0.024 kcal/mol per \AA^2 of protein surface.

Although we do not pretend to account quantitatively for the exact binding energies which have been observed, these calculations suggest that there is sufficient free energy available, from entropy driven effects, to account for the observed interaction energies despite the rather small areas of the two molecules which interact. It seems likely that other cases of strong, specific binding will show similar areas of interaction, since this is sufficient to give an association which is virtually complete under physiological conditions, where 1 molecule/cell is typically 10^{-9} M.

Why Is the Tetrahedral Form Stabilized? The other surprising and unforeseen feature of these structures is the stabilization of the tetrahedral adduct. What this may "cost" in terms of enthalpy cannot be assessed (Table X, line 7), except that it cannot exceed the free energy of activation of the acylation step of hydrolysis of the scissile bond. Fastrez and Fersht (1973) have estimated this activation energy as 16.5 kcal/mol for an amide substrate of chymotrypsin.

The observation of a tetrahedral adduct in both PTI and STI complexes indicates that this form of the total system is a minimum of free energy in both cases. Observation in two different systems rules out any suggestion that this is due to some particular effect of the crystal packing, unrelated to the functions of the different molecules.

The inescapable conclusion is that the tetrahedral adduct is stabilized because this stabilization is the function of the trypsin-like enzymes, lowering the activation energy of the reactions they catalyze. If, for example, the energy of the transition state of the relevant reaction step were lowered by $\Delta(\Delta F^*) = 8$ kcal/mol, the rate of the acylation and deacylation steps would be increased by a factor $\exp(\Delta(\Delta F^*)/RT)$, or about 10^6 .

To the extent that this reduction of energy of the transition state results from binding interactions, we might be able to form some estimate of it by a similar method of accounting as was used in the previous section. If one thinks of the enzyme as a pair of tongs, which grasp the inhibitor on each side of the scissile bond, one can see that no strain can be applied to the scissile bond unless both jaws of the tongs are in use, and that the strain energy cannot exceed the binding energy of the more weakly attached side of the tongs. This point of view emphasizes the importance of interactions at P_1' (and perhaps also at P_2'), which assist the preferential "fitting" of the tetrahedral adduct to the active site. A detailed structural analysis shows that there would be overcrowding around Ser-195, if the P_1' interactions were made in the Michaelis complex (Fersht *et al.*, 1973), and that there is no room to break open the scissile bond while the P_1' interactions are maintained (Table VII).

However, other factors, which have been described previously as *electronic strain* (Blow and Steitz, 1970), contribute to the stabilization of the tetrahedral adduct. They include the further delocalization of the negative charge at the active center. This charge may be thought of as arising from the buried Asp-102 (Blow *et al.*, 1969; Hunkepilliar *et al.*, 1973). In the tetrahedral form, the negative charge is probably transferred largely to the carbonyl oxygen of the tetrahedral carbon (see the scheme given by Henderson *et al.*, 1971). The charge

would be maximally delocalized in intermediates between the tetrahedral form and the Michaelis or acyl forms (Huber *et al.*, 1974). This negative charge will also strengthen the hydrogen bonds between carbonyl oxygen and NH(193) and NH(195), which are better aligned to the tetrahedral form (Henderson, 1970) (Table X, line 8).

Correlation of Structure with Protein Modification Studies. Various protein modifications have been performed on the components of this complex. We can understand many of the results of these modifications in terms of the structure of the complex.

In anhydrotrypsin, in which Ser-195 is modified to dehydroalanine (Weiner *et al.*, 1966), the enzymatic mechanism is disabled and a tetrahedral adduct cannot form. The observation that inhibitors bind strongly to anhydrotrypsin (Ako *et al.*, 1974; Lazdunski *et al.*, 1974) was taken as evidence that the chemical steps of tryptic hydrolysis were not involved in inhibitor binding. We can now see that the removal of O^γ(195) in anhydrotrypsin relieves the overcrowding at the active site, and allows strong binding to occur without formation of a chemical bond. We cannot tell whether the overcrowding is completely eliminated, or whether the fit of the unmodified inhibitor to anhydrotrypsin is as exact as the fit which the natural complex exhibits. The binding of PTI and STI to anhydrotrypsin is slightly weaker, $\Delta F_{\text{binding}}$ being altered by about 1.4 kcal/mol for PTI (Lazdunski *et al.*, 1974) and about 3 kcal/mol for STI (Ako *et al.*, 1974). Lima bean inhibitor binds more strongly to anhydrotrypsin than to trypsin (Ako *et al.*, 1974).

Liener and coworkers have studied the availability of tyrosine and arginine side chains to modifying reagents in STI, bovine trypsin, and the complex (Papaionnou and Liener, 1970; Delanco and Liener, 1973). The two tyrosines of bovine trypsin which are buried in the complex are likely to be Tyr-151 and Tyr-39 (which is Ser in porcine trypsin), as they suggest (Delanco and Liener, 1973). Tyr-228 is close to the binding pocket and is completely buried in the complex, but is unavailable for reaction even in the free enzyme (Kenner *et al.*, 1968; Holeyšvský *et al.*, 1969). In the porcine trypsin complex Tyr-217 becomes buried, but this is Ser in bovine trypsin. Papaionnou and Liener (1970) also found two tyrosines of STI buried in the complex. The structure shows that one of these will be Tyr-62'. Kowalski *et al.* (1974) confirmed the identify of Tyr-62' in the contact region, but found no other tyrosines to be buried upon formation of the complex.

Delanco and Liener (1973) found that 2.5 fewer arginines of STI and one fewer arginine of trypsin were available in the complex. Two of these are certainly Arg-63' and -65' as suggested. A further contribution might possibly come from Arg-129' or -132', which we have not located. We cannot suggest which arginine of trypsin would be buried, since both Arg-65A and Arg-117 are far from the inhibitor.

Kowalski *et al.* (1974) performed a number of chemical modifications on STI in which Arg-63' and Ile-64' are deleted, substituted by other amino acids, or have extra substituents added to them. The inhibitory activities of the resulting modified STI molecules correlate well with structural results. These workers found that (Lys-63')-STI behaves like native (Arg-63')-STI. (Trp-63')-STI is a good *chymotrypsin* inhibitor, and (Phe-63')-STI behaves much like native STI. They also showed that Ile-64' can be replaced by Gly, Ala, or Leu with essentially complete retention of activity, and that absence of Ile-64' or addition of any substituent to Arg-63'-COOH (such as glycynamide) or to H₂N-Ile-64' (carbamyl, acetamido, or citraconyl groups) destroys binding of the inhibitor to trypsin. The retention of activity following the substitutions at 63' and

64' are consistent with utilization of specificities at P₁ and P₁'. Loss of activity on deletion of Ile-64' and on addition of substituents in the scissile bond region reflects a finely tuned binding between fairly rigid molecules.

Acknowledgment

We are grateful to K. A. Walsh and H. Neurath for providing sequence information about porcine trypsin prior to publication, and to R. Huber for much information about the structure of PTI and the PTI complex.

References

- Ako, H., Foster, R. J., and Ryan, C. A. (1974), *Biochemistry* 13, 132.
- Baugh, R. J., and Trowbridge, C. G. (1972), *J. Biol. Chem.* 247, 7498.
- Birktoft, J., and Blow, D. M. (1972), *J. Mol. Biol.* 68, 187.
- Blow, D. M. (1971), *Enzymes*, 3rd Ed., 3, 185.
- Blow, D. M., Birktoft, J. J., and Hartley, B. S. (1969), *Nature (London)* 221, 337.
- Blow, D. M., Janin, J., and Sweet, R. M. (1974), *Nature (London)* 249, 54.
- Blow, D. M., and Matthews, B. W. (1973), *Acta Crystallogr., Sect. A* 29, 56.
- Blow, D. M., and Steitz, T. A. (1970), *Annu. Rev. Biochem.* 39, 63.
- Blow, D. M., Wright, C. S., Kukla, D., Rühlmann, A., Steigemann, W., and Huber, R. (1972), *J. Mol. Biol.* 69, 137.
- Chothia, C. (1974), *Nature (London)* 248, 338.
- Deisenhofer, M., and Steigemann, W. (1974), Second International Research Conference on Proteinase Inhibitors (Grosse Ledder, October 1973), Heidelberg, Springer Verlag (in press).
- Delanco, J. E., and Liener, I. E. (1973), *Biochim. Biophys. Acta* 303, 274.
- Diamond, R. (1966), *Acta Crystallogr.* 21, 253.
- Diamond, R. (1971), *Acta Crystallogr., Sect. A* 27, 436.
- Diamond, R. (1974), *J. Mol. Biol.* 82, 371.
- Di Bella, F. P., and Liener, I. E. (1969), *J. Biol. Chem.* 244, 2824.
- Dickerson, R. E., Kendrew, J. C., and Strandberg, B. E. (1961), in *Computing Methods and the Phase Problem in X-Ray Crystal Analysis*, Oxford, Pergamon Press, pp 236-251.
- Edelhoc, H., and Steiner, R. F. (1965), *J. Biol. Chem.* 240, 2877.
- Fastrez, J., and Fersht, A. R. (1973), *Biochemistry* 12, 2025.
- Fersht, A. R., Blow, D. M., and Fastrez, J. (1973), *Biochemistry* 12, 2035.
- Finkensadt, W. R., and Laskowski, M., Jr. (1967), *J. Biol. Chem.* 242, 771.
- Furnas, T. C. (1951), *Single Crystal Orienter Instruction Manual*, Milwaukee, Wis., General Electric Co.
- Harmon, K. M., and Niemann, C. (1949), *J. Biol. Chem.* 178, 743.
- Henderson, R. (1970), *J. Mol. Biol.* 54, 341.
- Henderson, R., Wright, C. S., Hess, G. P., and Blow, D. M. (1971), *Cold Spring Harbor Symp. Quant. Biol.* 36, 63.
- Hendrickson, W. A., and Lattman, E. E. (1970), *Acta Crystallogr., Sect. B* 26, 136.
- Hermanson, M. A., Ericsson, L. H., Neurath, H., and Walsh, K. A. (1973), *Biochemistry* 12, 3146.
- Hixson, H. F., Jr., and Laskowski, M., Jr. (1970), *J. Biol. Chem.* 245, 2027.

- Holeyšovský, V., Keil, B., and Šorm, F. (1969), *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 3, 107.
- Huber, R., Kukla, D., Rühlmann, A., Epp, O., and Formanek, H. (1970), *Naturwissenschaften* 57, 389.
- Huber, R., Kukla, D., Steigemann, W., Deisenhofer, J., and Jones, A. (1974), Second International Research Conference on Proteinase Inhibitors (Grosse Ledder, October 1973), Heidelberg, Springer Verlag (in press).
- Hunkepilliar, M., Smallcombe, S., Whitaker, D. R., and Richards, J. H. (1973), *J. Biol. Chem.* 248, 8306.
- Ikeda, K., Hamaguchi, K., Yamamoto, M., and Ikenaka, T. (1968), *J. Biochem. (Tokyo)* 63, 521.
- Jirgensons, B. (1965), *J. Biol. Chem.* 240, 1064.
- Kauzmann, W. (1959), *Advan. Protein Chem.* 14, 1.
- Kenner, R. A., Walsh, K. A., and Neurath, H. (1968), *Biochem. Biophys. Res. Commun.* 33, 353.
- Koide, T., and Ikenaka, T. (1973), *Eur. J. Biochem.* 32, 417.
- Kowalski, D., Leary, T. R., McKee, R. E., Sealock, R. W., Wang, D., and Laskowski, M., Jr. (1974), Second International Research Conference on Proteinase Inhibitors (Grosse Ledder, October 1973), Heidelberg, Springer Verlag (in press).
- Krieger, M., Kay, L. M., and Stroud, R. M. (1974), *J. Mol. Biol.* 83, 209.
- Laskowski, M., Jr., Duran, R. W., Finkenshtadt, W. R., Herbert, S., Hixson, H. F., Jr., Kowalski, D., Luthy, J. A., Mattis, J. A., McKee, R. E., and Niekamp, C. W. (1971), in *Proceedings of the International Research Conference on Proteinase Inhibitors*, Fritz, H., and Tschesche, H., Eds., Berlin, Walter de Gruyter and Co., pp 117-134.
- Laskowski, M., Jr., and Sealock, R. W. (1971), *Enzymes*, 3rd Ed., 3, 375.
- Lazdunski, M., Vincent, J.-P., Schweitz, H., Peron-Renner, M., and Pudles, J. (1974), Second International Research Conference on Proteinase Inhibitors (Grosse Ledder, October 1973), Heidelberg, Springer Verlag (in press).
- Markley, J. L. (1973), *Biochemistry* 12, 2245.
- Matthews, B. W. (1968), *J. Mol. Biol.* 33, 491.
- Mattis, J. A., and Laskowski, M., Jr. (1973), *Biochemistry* 12, 2239.
- Niekamp, C. W., Hixson, H. F., Jr., and Laskowski, M., Jr. (1969), *Biochemistry* 8, 16.
- Ozawa, K., and Laskowski, M., Jr. (1966), *J. Biol. Chem.* 241, 3955.
- Page, M. I., and Jencks, W. P. (1971), *Proc. Nat. Acad. Sci. U.S.* 68, 1678.
- Papaionnou, S. E., and Liener, I. E. (1970), *J. Biol. Chem.* 245, 4931.
- Richards, F. M. (1968), *J. Mol. Biol.* 37, 225.
- Rossmann, M. G. (1960), *Acta Crystallogr.* 13, 221.
- Rossmann, M. G., and Blow, D. M. (1961), *Acta Crystallogr.* 14, 641.
- Rühlmann, A., Kukla, D., Schwager, P., Bartels, K., and Huber, R. (1973), *J. Mol. Biol.* 77, 417.
- Schechter, I., and Berger, A. (1967), *Biochem. Biophys. Res. Commun.* 27, 157.
- Segal, D. M. (1972), *Biochemistry* 11, 349.
- Segal, D. M., Powers, J. C., Cohen, G. H., Davies, D. R., and Wilcox, P. E. (1971), *Biochemistry* 10, 3728.
- Shotton, D. M., and Watson, H. C. (1970), *Nature (London)* 225, 811.
- Sim, G. A. (1959), *Acta Crystallogr.* 12, 813.
- Steitz, T. A., Henderson, R., and Blow, D. M. (1969), *J. Mol. Biol.* 46, 337.
- Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1971), *Cold Spring Harbor Symp. Quant. Biol.* 36, 125.
- Stroud, R. M., Kay, L. M., and Dickerson, R. E. (1974), *J. Mol. Biol.* 83, 185.
- Ten Eyck, L. F. (1973), *Acta Crystallogr., Sect. A* 29, 183.
- Vandlen, R. L., and Tulinsky, A. (1973), *Biochemistry* 12, 4193.
- Venkatachalam, C. M. (1968), *Biopolymers* 6, 1425.
- Vincent, J. P., and Lazdunski, M. (1972), *Biochemistry* 11, 2967.
- Viyathil, A. J., Buck, F., Biev, M., and Nord, F. F. (1961), *Arch. Biochem. Biophys.* 92, 532.
- Watson, H. C., Shotton, D. M., Cox, J. M., and Muirhead, H. (1970), *Nature (London)* 225, 806.
- Weiner, H., White, W., Hoare, D. G., and Koshland, D. E., Jr. (1966), *J. Amer. Chem. Soc.* 88, 3851.
- Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., and Richards, F. M. (1970), *J. Biol. Chem.* 245, 305.