Acta Cryst. (1975). B31, 238

Crystallographic Refinement of the Structure of Bovine Pancreatic Trypsin Inhibitor at 1·5 Å Resolution*

BY JOHANN DEISENHOFER† AND WOLFGANG STEIGEMANN

Max-Planck-Institut für Biochemie und Physikalisch-Chemisches Institut der Technischen Universität, München, Germany (BRD)

(*Received* 14 *June* 1974; *accepted* 19 *August* 1974)

The model of pancreatic trypsin inhibitor determined by Huber, Kukla, Rühlmann, Epp & Formanek [*Naturwissenschaften* (1970). **57**, 389–392] was fitted to an electron-density map of 1·9 Å resolution calculated with isomorphous-replacement phases by application of the real-space refinement procedure of Diamond [*Acta Cryst.* (1971). A**27**, 436–452]. Further refinement was done at 1·5 Å resolution by cyclic real-space refinement against electron-density maps computed with observed structure-factor amplitudes and phases calculated from the preceding model. Five difference Fourier maps helped to correct severe errors in the model, and to locate solvent molecules. After 15 cycles of constrained crystallographic refinement the R value for all 8079 independent reflexions was 0·225. The R value for 6893 reflexions with intensities above the 2σ significance level determined from counting statistics within the range 7·0 Å $> d >$ 1·5 Å was 0·197. Details of the refined model are presented in terms of hydrogen bonds and conformational angles.

1. Introduction

The bovine pancreatic trypsin inhibitor (PTI) is one of the numerous natural proteinase inhibitors, whose physiological function is to inhibit the digestive enzymes. PTI is a polypeptide of 58 amino acids with a fairly high content of arginine and lysine which makes it a basic protein (Kassell, Radicevic, Ansfield & Laskowski, 1965; Anderer & Hörnle, 1966). The polypeptide chain of PTI is cross-linked by three disulphide bridges which may account for its high stability against denaturation (Vincent, Chicheportiche & Lazdunski, 1971). The tertiary structure of PTI has been determined by Huber, Kukla, Rühlmann, Epp & Formanek (1970*a*) on the basis of an electron-density map at 2·5 Å resolution calculated from observed structure-factor amplitudes and multiple-isomorphous-replacement phases

It is well known that in protein crystallography the isomorphous-replacement phases usually are very inaccurate. Electron-density maps calculated with such phases do not allow one to make precise statements about structural details which are essential for the investigation of protein function, *e.g.*, the exact location of hydrogen bonds and of distorted peptide geometry in the molecule. At present, the only way to obtain more accurate protein models from crystallographic work is to use the experimentally determined phases only as starting information, and to replace them by calculated phases in a following refinement step.

The first successful approach of this kind was made on rubredoxin by Watenpaugh, Sieker, Herriott & Jensen (1971, 1973), who refined their model using difference-Fourier and least-squares methods. Later on, direct least-squares refinement of phases was done on the same structure (Sayre, 1972, 1974).

The purpose of this paper is to describe the constrained crystallographic refinement of the PTI model at 1·5 Å resolution using Diamond's (1971) real-space refinement procedure together with electron-density maps obtained from observed structure-factor amplitudes and calculated phases. The method is discussed in terms of the results, and details of the refined model are described. A preliminary account of the PTI refinement has been given earlier (Deisenhofer & Steigemann, 1973).

2. Experimental

A summary of PTI crystal parameters as determined by Huber *et al.* (1970*a*) is given in Table 1. The native data of Huber *et al.* (1970*a*) were expanded to 1·5 Å resolution, and the set from the heavy-atom derivative (PCMB) to 1·9 Å resolution. A computer-controlled four-circle diffractometer working with the $\omega/2\theta$ scan technique was used. The X-ray tube with a Cu anode was operated at 40 kV, 18 mA; the focus size was 0·4 × 4 mm with a take-off angle of 6°. Cu $K\alpha$ radiation ($\lambda = 1{\cdot}5418$ Å) was selected with a Ni filter. The distances between focal spot and crystal and between crystal and scintillation counter were both 360 mm. The intensity profile of each reflexion was scanned twice in steps of 1/100° in ω; the background was counted on each side of the peak for half as long as the peak itself. The counting time per step was set inversely proportional to the intensity at the peak of the reflexion; the upper limit was set at 3 s/step.

* The work reported in this paper was from the Dissertation of J. Deisenhofer, Munich, 1974.

† To whom correspondence should be addressed at the Max-Planck-Institut für Biochemie, 8033 Martinsried, Germany.

Table 1. *PTI crystal parameters*

Orthorhombic unit cell: $a=43{\cdot}1$ Å, $b=22{\cdot}9$, $c=48{\cdot}6$, $a\times b\times c\simeq 48000$ Å^3
Space group: $P2_12_12_1$
(one molecule per asymmetric unit)
Buffer: 2·25 *M* K_2HPO_4, *p*H 10

Many crystals showed cracks perpendicular to the *c* axis and a scan range of 0·7° in ω was necessary to sample the whole diffracted intensity for the outermost reflexions; the maximum time for measuring a reflexion was 10 min. Reflexions above the 2σ significance level determined from counting statistics were scanned a second time.

Loss of intensity due to radiation damage or movement of the crystal was monitored by measuring the intensity of one reference reflexion at intervals of 15 reflexions. In addition, 30 strong control reflexions were measured every second day. Crystals were used until the intensity of the control reflexions had decreased by more than 10%. The complete set of native data out to 1·5 Å resolution contained 15144 measurements from nine crystals for 8079 independent reflexions; 6988 of these reflexions had intensities above 2σ.*

The intensities from each crystal were corrected for absorption effects by an empirical method (Huber & Kopfmann, 1969). The absorption-corrected data from the individual crystals were scaled together using a method developed by one of us (Steigemann, 1974). This method determined for each crystal both a linear scale factor and a relative temperature factor by means of the commonly measured reflexions. In this way it is possible to take into account variations of crystalline order.

3. Method

3.1. *Diamond's real-space refinement procedure*

The real-space refinement procedure (Diamond, 1971) constitutes the central part of the crystallographic refinement described in this paper, and therefore we shall briefly describe its principles of operation (Diamond, 1974). This method uses a model electron density $\varrho_m(\mathbf{r})$ which is assembled from spherical Gaussian atoms with weights Z_i and radii a_i at the positions r_i:

$$\varrho_m(\mathbf{r})=K\sum_i Z_i G(a_i,\mathbf{r}-\mathbf{r}_i)+b$$

with

$$G(a,\mathbf{r})=a^{-3}\exp(-\pi r^2/a^2)\,.$$

K here is an overall scale factor and b is a background level. The procedure then adjusts the parameters K, $Z_i, a_i, \mathbf{r}_i$ and b in order to minimize the volume integral (residual) $\int(\varrho_0-\varrho_m)^2 dv$ (ϱ_0 is the observed electron density). In a flexible chain the atomic positions $\mathbf{r}_i$ are interconnected and the corresponding free parameters are the conformational angles of the polypeptide. The procedure operates on a 'molten zone' of some consecutive residues. The zone is then shifted along the chain by one residue after each set of adjustments. To conserve chain continuity only those combinations of rotations are selected which do not move the two ends of the molten zone. All adjustments are controlled by the appropriate setting of filter levels for indicating those combinations of parametric shifts (eigenshifts) which introduce large deformations with little decrease of the residual. Undesirable shifts are characterized by a correspondingly small eigenvalue of the normal matrix. Eigenshifts belonging to eigenvalues below the threshold $\lambda_{\min}$ (set either explicitly or implicitly by the filter ratio $\lambda_{\min}/\lambda_{\max}$) are therefore ignored. With this facility it is also possible to take into account the differing flexibilities of the various conformational parameters by assigning to each kind of parameter a weight inversely proportional to its stiffness. In this way, more flexible parameters with high weights are altered in preference to stiff parameters with low weights.

A detailed description of the real-space refinement was given by its author (Diamond, 1971).

3.2. *Refinement against* 1·9 Å *map*

The atomic coordinates which had been obtained from a Kendrew skeleton model of the PTI (Huber *et al.*, 1970*a*) were adjusted to standard bond lengths and bond angles using Diamond's mathematical model-building procedure (Diamond, 1966). The specifications used for model building are listed in Table 2. The r.m.s. error between coordinates before and after model building was 0·37 Å. [A list of the sources of the amino acid groups used in model building is given by Diamond (1974), Table 1.]

Table 2. *Specifications for model building*

Probe	Lengths (residues)	Filter constants* C_1	C_2	Flexible τ(NC$^\alpha$C)
1	1	0·5	10^{-4}	no
2	2	0·1	10^{-4}	no
3	3	0·1	10^{-4}	no
4	3	0	10^{-2}	yes

* Eigenshifts are permitted if either $\lambda \geq C_1 \,.\, \varepsilon^2$ or $\lambda \geq C_2 \,.\, \lambda_{\max}$, where λ are eigenvalues of normal matrix; ε^2 is the residual.

The crystallographic R value ($R=\sum||F_o|-|F_c||/\sum|F_o|$, where F_o are the observed, and F_c the calculated structure factors) for the corrected model was 0·52 at 2·5 Å resolution. A difference Fourier map showed very few interpretable details.

For this reason an electron density map of 1·9 Å resolution was calculated using multiple-isomorphous-replacement phases for $d>2{\cdot}5$ Å and single-isomor-

* Tables of coordinates and structure factors have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 30648 (63 pp., 2 microfiches). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 13 White Friars, Chester CH1 1NZ, England.

phous-replacement phases for 2·5 Å > d > 1·9 Å from the PCMB derivative (d is the Bragg spacing). All reflexions with a figure of merit of $m > 0{\cdot}2$ were given unit weight; those with $m < 0{\cdot}2$ were set to zero. The map was calculated with a grid of 0·75 Å.

Because of the higher resolution the 1·9 Å map showed more details than the 2·5 Å map of Huber *et al.* (1970*a*), especially in the backbone. For seven large side chains (Glu 7, Lys 15, Arg 17, Lys 26, Lys 41, Arg 42, Lys 46) which protrude into the solvent region the map contained no significant continuous density. The region near the C-terminus, which is a preferred binding site for heavy-atom compounds, was severely disturbed.

The PTI model was fitted to the 1·9 Å map using Diamond's (1971) real-space refinement procedure. In this step the free parameters of the model were the conformational angles $\varphi, \psi, \chi^{1,2,3,4}$ with weights of 100, and χ^5 with a weight of 1·0. [Names of atoms and parameters in this paper follow the IUPAC-IUB Commission on Biochemical Nomenclature (1970*a*, *b*); further definitions are given by Diamond (1974)]. A filter ratio of $\lambda_{min}/\lambda_{max} = 0{\cdot}0001$ was used.

The atomic radii were held constant at 1·2 Å; scale factor and background level were redetermined in each refinement run. The rise of the scale factor indicated the movement of atoms into higher density. The refinement was started with the fitting of electron densities of only backbone and C^β atoms. In the following runs those remaining side-chain atoms which could be located in the electron-density map were included. The inspection of the F_o map together with the refined atomic positions showed that eight side groups had moved into wrong density. This had happened because of errors in the model based upon the 2·5 Å density map. Proper convergence could easily be achieved by rotating the side groups by angles between 120 and 180°. After eight cycles of real-space refinement and a r.m.s. shift of 1·3 Å for all atoms and 0·5 Å for backbone atoms, the R value at 1·5 Å resolution had decreased to 0·45.

3.3. *Constrained crystallographic refinement of the PTI model*

A flow chart of the cyclic crystallographic refinement of the PTI model is shown in Fig. 1.

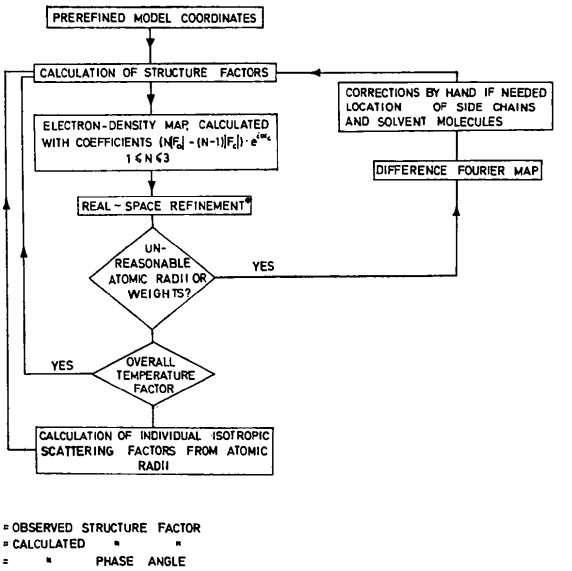


Fig. 1. Flow chart of cyclic crystallographic refinement of the PTI model. Refinement was stopped when no further significant improvement of the model could be expected. F_o = observed structure factor, F_c = calculated structure factor, α_c = calculated phase angle. [Real-space refinement: Diamond (1971).]

Table 3. *Course of PTI refinement*

Cycle	ΔF map	R value	Observations per parameter*	Comment
1		0·425	13·1	2·0 Å resolution
2		0·405	10·6	τ flexible
3		0·396	23·6	1·5 Å resolution
4		0·384	23·6	
5	+	0·361	23·6	
6		0·341	8·7	17 H_2O included, individual B's refined
7	+	0·316	8·7	
8		0·308	7·7	16 H_2O added
9		0·289	7·7	
10	+	0·260	7·3	ω flexible
11		0·253	6·8	12 H_2O added
12	+	0·240	6·8	
13	+	0·235	6·5	10 H_2O added
14		0·232	13·0	8 H_2O removed, no translational and conformational refinement
15		0·225† 0·197‡	6·7	Full set of parameters

* 6988 significant observations divided by number of free model parameters in current cycle.
† 8079 independent reflexions.
‡ 6893 significant observations with 7·0 Å > d > 1·5 Å.

Structure factors, calculated from the model, were combined with the observed structure-factor amplitudes to derive a new electron-density map. The model was fitted to this map using Diamond's real-space refinement procedure. In this step constraints were imposed on the PTI model by shifting conformational angles instead of independent atomic positions. Bond lengths and most of the bond angles in the protein, which were established during model building (Diamond, 1966), were held constant. In an advanced state of the refinement, the overall temperature factor was replaced by individual atomic scattering factors derived from the refined atomic radii. A new cycle started with a structure-factor calculation from the refined model. 15 of these cycles were carried out.

For the correction of serious errors in the model which led real-space refinement into side minima (*e.g.* a side chain orientation mis-set by 180°) five difference Fourier maps were calculated. These maps were also used to locate solvent molecules.

All steps of the refinement process, except for the inspection of difference Fourier maps, were carried out on a computer. The course of the refinement is shown in Table 3. A description of each refinement step will be given in the following.

3.4. *Structure-factor calculation*

For the structure-factor calculations of cycles 1 to 5 atomic form factors (Forsyth & Wells, 1959) were used together with an overall temperature factor of 11·4 Å^2 as determined from a Wilson plot. In cycles 6 to 13 individual Gaussian atomic scattering factors (Diamond, 1971) were derived from the atomic radii:

$$f_i = Z_i \exp\left[-4\pi(ca_i)^2 \sin^2\theta/\lambda^2\right]$$

where a_i (Å) is the radius of the ith atom, Z_i (e) is the occupancy of the ith atom, θ is the scattering angle, λ is the X-ray wavelength, and c is a global factor which compensates for a systematic error in the radii. If the refined radii are systematically too small or too great, the ratio $\langle|F_o|^2\rangle/\langle|F_c|^2\rangle$ decreases or increases with resolution. In our case a value of $c=0{\cdot}95$ was used.

In cycles 7 to 11 the refined radii were corrected by a simple procedure: a mean radius a_0 was calculated for the main-chain atoms of each amino acid. The radii of the corresponding side-chain atoms were made to increase linearly with the distance x_i from C$^\alpha$ according to $a_i = a_0 + kx_i$ where k was determined for each side chain by a least-squares procedure. By this correction large variations between radii of neighbouring atoms were avoided; from cycle 12 on, this was no longer necessary.

In cycles 14 and 15 atomic form factors were combined with individual isotropic temperature factors derived from atomic radii according to

$$B_i = 4\pi[(ca_i)^2 - a_0^2]$$

where a_0 are 'zero temperature radii' which were obtained by fitting Gaussian curves to the form factor curves of C, N, O and S atoms. The values of a_o used were 0·95 Å for C, 0·90 Å for N, 0·83 Å for O and 0·90 Å for S; c had to be set to 1·04 in this case. These numbers were determined in cycle 14; no translational or conformational refinement was done in this cycle.

PTI atoms were always weighted with the neutral electron number; solvent molecules were weighted with

Table 4. *Summary of real-space refinement on the PTI model*

Cycle	n: Type of electron density*	Resolution (Å)	K	Z	Refined quantities $a\,\varphi,\psi,\chi^{1,2,3,4}$†	$\chi^5,\theta_{1,2,3}$†‡	τ†	ω†	Filter constants KDAZ§ λ_{min}	TRANS, BEND‖ $\lambda_{min}/\lambda_{max}$	R.m.s. shift (Å) (real atoms)
1	1	2·0	+		100	1	0	0	10^{-1}	10^{-4}	0·12
2	1	2·0	+		100	1	1	0	10^{-1}	10^{-4}	0·12
3	1	1·5	+		100	1	1	0	10^{-1}	10^{-4}	0·07
4	2	1·5	+		100	1	1	0	10^{-1}	10^{-4}	0·10
5	3	1·5	+		100	1	1	0	10^{-1}	10^{-2}	0·20
6	2	1·5	+		+ 100	1	1	0	10^{-1}	10^{-4}	0·13
7	3	1·5	+		+ 100	1	1	0	10^{-1}	10^{-3}	0·34
8	1, 2	1·5	+		+ 100	1	1	0	10^{-1}	10^{-4}	0·14
9	2	1·5	+		+ 100	1	1	0	10^{-1}	10^{-4}	0·21
10	1, 2	1·5	+		+ 100	1	1	10	10^{-1}	10^{-4}	0·13
11	1, 2	1·5	+		+ 100	1	1	10	10^{-1}	10^{-4}	0·08
12	2	1·5		+	+ 10	10	10	10	1·0	10^{-4}	0·05
13	2	1·5		+	+ 10	10	10	10	1·0	10^{-4}	0·10
14	2	1·5	+		+ 0	0	0	0	10^{-1}	—	—
15	2	1·5	+		+ 10	10	10	10	10^{-1}	10^{-4}	0·08

* Electron density maps were calculated with coefficients $[n|F_o| - (n-1)|F_c|]$. exp $(i\alpha_c)$.
† Numbers in columns are parametric weights in conformational refinement; a parameter is rigid if the weight is zero.
‡ $\theta_{1,2,3}$ occur in prolines
$\theta_1 = \theta(C_{i-1}\text{–}C_i^\alpha\text{–}N_i\text{–}C_i^\beta)$
$\theta_2 = \theta(C^\alpha\text{–}N\text{–}C^\beta\text{–}C^\delta)$
$\theta_3 = \theta(N\text{–}C^\beta\text{–}C^\delta\text{–}C^\gamma)$.
§ In KDAZ combinations of scale factor K, background d, radii a_i, and occupancies Z_i are refined.
‖ Translational and conformational refinement.

refined occupancies (only molecules with $Z_i > 2{\cdot}5$ e were included).

After each structure-factor calculation R values and absolute scale factors were calculated by shells for all reflexions with $d > 1{\cdot}5$ Å. The shells were of approximately equal volume.

3.5. *Electron-density maps*

The electron-density maps were calculated for one asymmetric unit. In the first two cycles the resolution was 2·0 Å; in all the following cycles the full available resolution of 1·5 Å was used. The distance between grid points was 0·6 Å in cycles 1 to 11, and 0·4 Å for the following maps. The smaller grid distances improved the results of real-space refinement, especially the atomic radii. A scale factor of 100 was always applied after absolute scaling. The coefficients of the Fourier summation had the general form

$$[n|F_o| - (n-1)|F_c|] \exp(i\alpha_c)$$

with $1 \leq n \leq 3$ (α_c is the calculated phase angle). $3F_o - 2F_c$ maps were used with good results in cycles 5 and 7 to accelerate the automatic correction of errors in several peptide orientations, and in the disulphide bridge, Cys 14–38. In a number of cases, however, those maps led to overestimated shifts so that mainly maps with $n=2$ were used to speed up the convergence of real-space refinement. In cycles 8, 10, and 11 the atomic radii were refined in F_o maps, and the conformational angles in $2F_o - F_c$ maps (see Table 4).

3.6. *Real-space refinement*

A summary of real-space refinement on the PTI model is presented in Table 4.

During the first 11 cycles the conformational angles were given different weights according to their assumed flexibility. No attempt was made to optimize these weights. In the last cycles all weights of conformational parameters were made equal because the electron-density maps were essentially correct and there was no further reason to select between different types of parameter.

In cycles 1 to 5 the atomic radii were set to 1·2 Å. In most of the following cycles PTI radii were refined together with scale factor K and background level b. Simultaneous refinement of radii a_i, occupancies Z_i and background was done in cycles 12 and 13 with a filter level of $\lambda_{min} = 1{\cdot}0$. In general the results were comparable to those of K,b,a_i refinement. In several cases, however, atoms with a high temperature factor were given small radii and small occupancies, contrary to the desired behaviour so that finally K,b,a_i refinement was preferred.

Besides translational refinement, K,b,a_i refinement was done on the solvent molecules in two cycles (6, 7); the occupancies, which were used in the structure-factor calculation, were refined in cycles 8–15. The resulting Z_i values indicated that all located solvent molecules are water.

In the refinement of combinations of K,b,a_i, and Z_i the filter level λ_{min} was set as a lower limit for the eigenvalues of the normal matrices, whereas in translational and conformational refinement the filter ratio $\lambda_{min}/\lambda_{max}$ was used. Most of the real-space refinements were done with a molten zone of six residues. The r.m.s. movements listed in Table 4 are those of all PTI atoms whose parameters have been refined; movements of solvent molecules and corrections derived from ΔF maps are not included.

3.7. *Difference Fourier maps*

Difference Fourier maps were calculated with coefficients $(|F_o| - |F_c|) \exp(i\alpha_c)$ during cycles 5, 7, 10, 12 and 13. Contour lines were plotted together with the atomic positions which allowed a quick inspection of the maps.

Conformational errors in the atomic model, which could not be corrected by real-space refinement, were detected in the side chains of Val 34, Arg 39, Lys 46, Asp 50, Arg 53, and Cys 38, and in Gly 56. With improved phases the ΔF maps also showed details of the side chains which could not be located in the 1·9 Å isomorphous electron-density map. The positions of the side chains of Glu 7, Arg 17, and Arg 42 were found completely. The positions of the following atoms which are mobile are unknown: Lys 15: CE, NZ; Lys 26: CE, NZ; Lys 41: NZ; Gly 57: C, O; all atoms of Ala 58.

In most cases the ΔF maps were easy to interpret and the necessary rotation angles were estimated with the help of amino-acid models. Because of the large convergence radius of the real-space refinement procedure these estimates did not require great accuracy. When unique interpretation could not be made, the atoms with uncertain positions were converted into dummy atoms, and did not contribute to the F_c calculations until they could be located correctly in one of the next ΔF maps.

In the 1·9 Å isomorphous map 24 peaks were interpreted as water molecules. Holes in the ΔF maps showed that seven of them had to be removed. In total, four water molecules were found in the interior of the PTI molecule, and 43 of them close to the protein surface forming single hydrogen bonds with the protein or bridges between adjacent protein molecules (see §4.2).

In large gaps between protein molecules the ΔF maps showed continuous electron density caused by disordered solvent. This diffuse density was almost completely removed in the last ΔF map by ignoring the 93 reflexions with $d > 7{\cdot}0$ Å. Also, eight of the peaks which had been interpreted as localized solvent molecules were absent in this map, and were no longer included in structure-factor calculations.

3.8. *Computer-time*

The computation times given in the following are valid for the program versions used in April 1974 on a Siemens 4004/150 computer (cycle time 0·75 μ sec).

(1) Structure factor calculation (program written by T. A. Jones and W. Steigemann): 35 min.
(2) Fourier summation (W. Steigemann): 15 min.
(3) Real-space refinement Mark 3 (R. Diamond): 150 min. Together with ΔF maps and service programs the mean computer time for one cycle was about four hours.

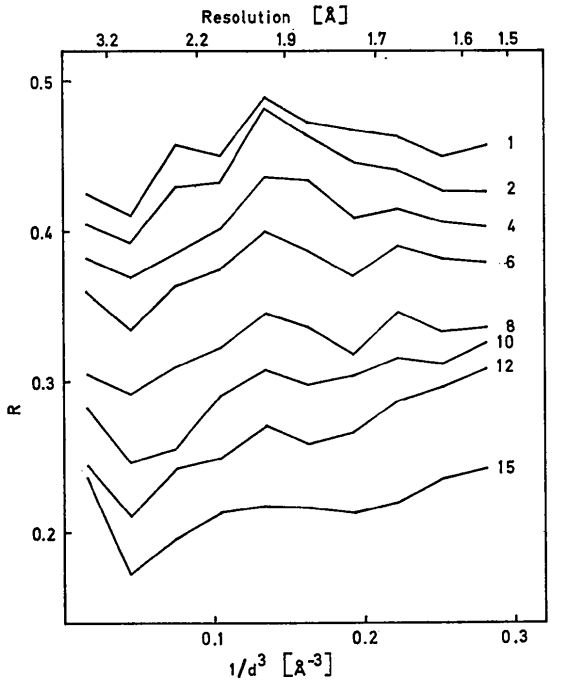


Fig. 2. Decrease of the R value during cyclic refinement. Curves represent R values from successive shells with equal numbers of reciprocal-lattice points. All measured reflexions are included. Cycle numbers are written at right.

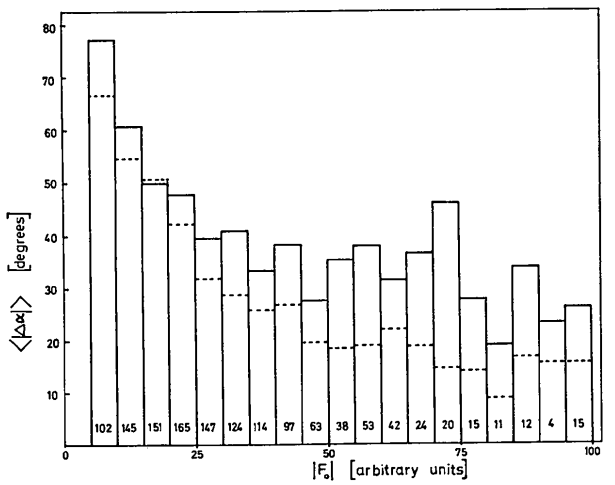


Fig. 3. Mean phase differences $\langle|\Delta\alpha|\rangle$ between final phases and phases from multiple isomorphous replacement (full lines) and first calculated phases (dashed lines) plotted *versus* the observed structure factor amplitude $|F_o|$. Only significantly measured non-centric reflexions with $d > 2{\cdot}5$ Å are included. The number of reflexions within each interval of $|F_o|$ is written at the bottom.

4. Results and discussion

4.1. *The refinement procedure*

The crystallographic refinement of the PTI model resulted in a mean R value decrease of 0·015 per cycle ending up at $R = 0{\cdot}225$ for all 8079 independent reflexions. One common absolute scale factor was applied. When 1091 reflexions with intensities below the 2σ significance level and 95 reflexions with $d > 7{\cdot}0$ Å were neglected the final R value was reduced to 0·197. The number of parameters in the final model was 1041, namely 358 conformational angles, three degrees of freedom each for translation and rotation of the whole PTI molecule, 489 temperature factors (one for each localized non-hydrogen atom), and 141 positional parameters and 47 occupancies of solvent molecules.

Fig. 2 shows R values calculated for reflexions within ten successive shells of equal volume in reciprocal space. The curves are very similar to those obtained by Watenpaugh *et al.* (1971, 1973) during the refinement of rubredoxin. The high R values for reflexions with $d > 4{\cdot}1$ Å can be attributed to the contribution of the solvent continuum. The r.m.s. deviation of main-chain atom coordinates between the improved starting model and the final model is 0·45 Å. The small shifts of model parameters and the small decrease of the R value during the last cycles of refinement together with the smoothness of the final ΔF map indicated that no further significant improvement of the model by real-space refinement could be expected with the set of free parameters given above.

One reason the R value could not be reduced further may be that systematic errors were introduced when data were collected from nine different crystals. Another reason is the limited accuracy of temperature factors obtainable with the method described. Small but significant errors of individual temperature factors account for most of the 56 peaks and 36 holes in the last difference Fourier map. These features are as large as 0·4 e Å^{-3} in some cases. (The 18 highest peaks (0·4–0·6 e Å^{-3}) are due to weakly occupied water molecules which had been excluded from the calculations as pointed out above. For comparison, the height of a representative oxygen atom in the Fourier map is 3 e Å^{-3}.)

The R value is also affected by the constraints in the model. Significant variations of the given standard bond lengths are improbable because they would cause a large increase in the potential energy (Ramachandran, 1969). The same, however, does not follow for the standard bond angles, but the limitation of angular flexibility to the interbond angle $\tau(NC^{\alpha}C)$ and the torsion angle $\omega(C^{\alpha}_{i-1}C_{i-1}N_iC^{\alpha}_i)$ is clearly arbitrary. If strain is present in a protein molecule it is certainly distributed over many bond angles rather than concentrated around this point.

The ΔF maps, however, indicated that significant changes could not be expected by increasing the flexibility of the model. The redistribution of strain and similar subtle changes are considered to be a field for energy refinement (Scheraga, 1968; Levitt & Lifson, 1969; Karplus, personal communication).

A list of the refined values of conformational angles of the PTI model is given in Table 5.

To test the crystallographic significance of the ω angles, model building (Diamond, 1966) was carried out with the refined atomic coordinates of cycle 14 which restored planar peptide groups. Although the r.m.s. shift of atomic positions was only 0·08 Å the R value increase of 0·018 showed that flexible ω angles improve the model even at a 0·005 significance level (Hamilton, 1965).

Unlike least-squares refinement in reciprocal space, the refinement method described here does not allow the direct calculation of standard deviations for the model parameters. However, it is possible to estimate

Table 5. *Conformational angles* (°) *of PTI*

*S*1, *S*2, *S*3, *S*4, *S*5 are angles $\chi^{1,2,3,4,5}$, except for prolines where *S*1, *S*2, *S*3 are angles $\theta_1, \theta_2, \theta_3$.

		φ	ψ	ω	τ	*S*1	*S*2	*S*3	*S*4	*S*5
ARG	1	0	144	179	116	82	−179	56	82	1
PRO	2	−58	152	177	113	−174	177	−170		
ASP	3	−61	−30	−177	116	−86	5			
PHE	4	−74	−12	169	112	66	71			
CYS	5	−65	−17	176	117	−56	−88	−87		
LEU	6	−92	−5	175	118	−54	−174			
GLU	7	−75	147	175	111	17	177	73		
PRO	8	−67	157	178	110	−178	−165	162		
PRO	9	−64	145	−173	109	178	−175	157		
TYR	10	−126	106	−179	106	178	77			
THR	11	−70	−41	173	111	−77				
GLY	12	94	179	−176	107					
PRO	13	−86	−7	178	121	178	−170	148		
CYS	14	−89	165	−170	111	−73	96	101		
LYS	15	−127	32	163	114	−111	92			
ALA	16	−77	178	−171	109					
ARG	17	−134	88	−163	110	−62	162	−137	113	−12
ILE	18	−125	118	−176	99	−65	−173			
ILE	19	−84	118	178	108	−57	−55			
ARG	20	−126	172	175	111	−64	−77	177	89	−3
TYR	21	−110	146	170	111	−72	77			
PHE	22	−129	151	167	106	76	100			
TYR	23	−80	130	−174	115	−171	−108			
ASN	24	−110	104	−175	100	173	−11			
ALA	25	−65	−28	171	117					
LYS	26	−67	−34	−177	116	−88	−121	105		
ALA	27	−93	−20	−179	113					
GLY	28	81	15	−175	118					
LEU	29	−159	175	−172	110	44	72			
CYS	30	−98	146	−177	111	−68	−122	−90		
GLN	31	−133	161	173	103	−67	178	145		
THR	32	−83	150	170	116	53				
PHE	33	−146	166	179	104	74	114			
VAL	34	−96	120	−175	107	172				
TYR	35	−100	134	178	100	161	53			
GLY	36	−73	−11	170	122					
GLY	37	105	0	174	116					
CYS	38	−146	151	178	119	74	−118	103		
ARG	39	61	40	−178	117	−46	−52	−178	−174	2
ALA	40	−61	152	177	115					
LYS	41	−103	180	−172	95	−88	−165	−175		
ARG	42	−85	−20	167	120	−83	−160	110	179	−8
ASN	43	−77	74	−166	113	−161	7			
ASN	44	−166	107	−175	104	176	−30			
PHE	45	−129	158	−177	112	−56	91			
LYS	46	−88	−7	174	115	−80	160	−142	91	
SER	47	−154	162	179	110	76				
ALA	48	−67	−34	−175	109					
GLU	49	−69	−42	173	112	−73	86	173		
ASP	50	−65	−41	175	112	−97	52			
CYS	51	−59	−50	−179	114	175	−94	−87		
MET	52	−71	−32	176	113	−70	−62	−71		
ARG	53	−63	−42	175	111	178	156	37	−125	7
THR	54	−76	−46	−173	115	−60				
CYS	55	−106	−4	−171	124	−69	−64	−91		
GLY	56	−78	−14	174	115					

the accuracy of the positions of the peaks in the elecfron density by application of Cruickshank's (1949) tormulae:

$$\sigma(x)=2\pi(\sum h^2(|F_o|-|F_c|)^2)^{1/2}/aV|C|$$

where a is the length of the unit cell in the x direction, V is the volume of the unit cell, and C is the central curvature of the electron density at a peak position [equivalent expressions are valid for $\sigma(y)$ and $\sigma(z)$]. Summations over all reflexions within the sphere with $d>1{\cdot}5$ Å were carried out and the result was multiplied by a factor of two to account for a non-centrosymmetric structure. The resulting standard deviations of the peak positions were

$$\sigma(x)\simeq\sigma(y)\simeq\sigma(z)\simeq 0{\cdot}35/|C|\ (\text{Å})\ .$$

The central curvature, which depends on the resolution and the temperature factor, can be determined either from the electron density at a peak in the F_o map (Watenpaugh *et al.*, 1973) or by differentiation of the model density used in real-space refinement. With both methods a value of $C\simeq -10$ e Å^{-5} was obtained for a typical carbonyl oxygen (Cys 14); the resulting standard deviations of the peak position were $\sigma(x)=\sigma(y)=\sigma(z)=0{\cdot}035$ Å, which lead to a radial standard deviation of $\sigma(r)=0{\cdot}06$ Å.

Although the atomic positions in the model fit quite well to the positions of electron-density peaks the estimated standard deviations cannot be taken as a direct measure of the accuracy of the model coordinates. One reason is that Diamond's real-space refinement procedure works with a volume fitting rather than a point-to-point fitting. The second reason is the high correlation between positional errors of neighbouring atoms in a constrained model. A mathematical analysis for such a situation does not yet exist. The impression of the authors is that the constraints produce, in regions of high B, a model better than that suggested by the estimated standard deviations, whereas for atoms with very sharp density peaks a perfect fit may be slightly inhibited by the rigidity of the model. The estimated standard deviations can be compared with the deviations in the positions of the sulphur atoms in the three disulphide bridges, since the refinement procedure gives two estimates of these coordinates (Diamond, 1971). (The S–S cysteine bridges are

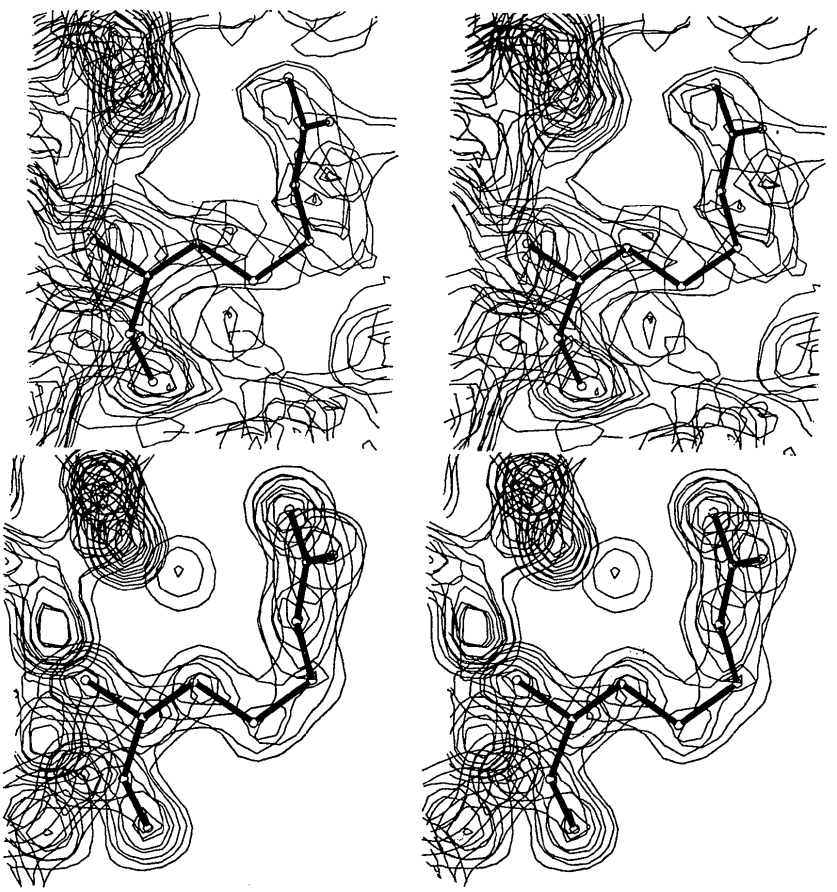


Fig. 4. Stereo pairs showing identical sections around Arg 53 of the 1·9 Å electron-density map calculated with isomorphous-replacement phases (top) and of the 1·5 Å resolution map obtained with refined phases (bottom). Contour lines in the 1·9 Å map are drawn at intervals of 0·3 e Å^{-3} starting with the 0·3 e Å^{-3} line. Contour lines in the 1·5 Å map start at 0·5 e Å^{-3} with intervals of 0·5 e Å^{-3}. The refined model of Arg 53 is drawn into both stereo pairs.

represented twice, as side chains of the two linked main chains, and are independently fitted.) The mean distance between equivalent sulphur atoms in the final model is 0·033 Å and agrees well with $\sigma = 0{\cdot}027$ Å estimated with Cruickshank's method. The mean difference between equivalent χ^3 angles is three degrees.

The progress of the crystallographic refinement of the PTI model proved that the phases calculated from the model after real-space refinement against the 1·9 Å isomorphous map were better than the isomorphous phases. This is shown in Fig. 3 where the mean phase differences $\langle|\Delta\alpha|\rangle$ between the final phases and (*a*) multiple isomorphous replacement phases, (*b*) calculated phases after the first cycle are plotted for non-centric reflexions *versus* the observed structure factor amplitudes. In both cases $\langle|\Delta\alpha|\rangle$ decreases with in-

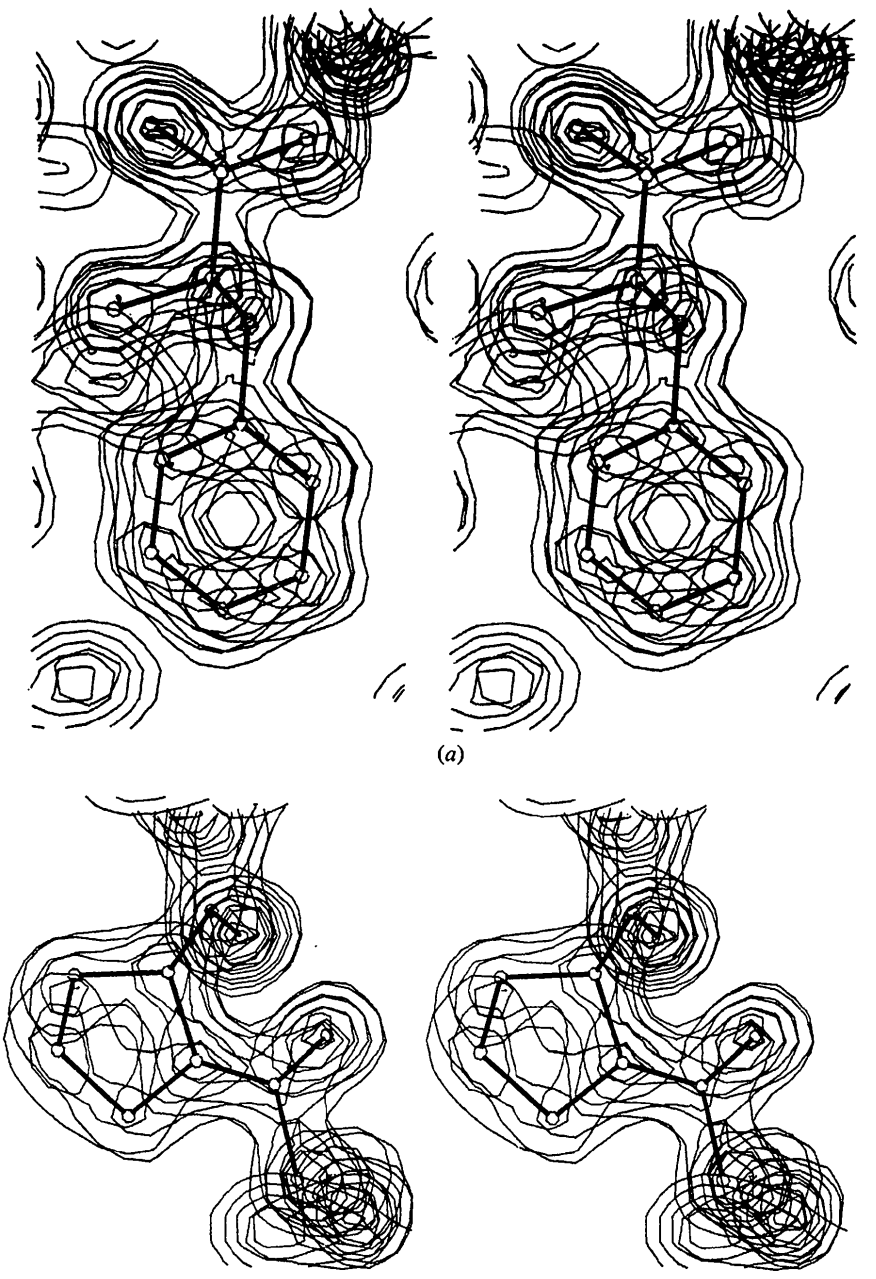
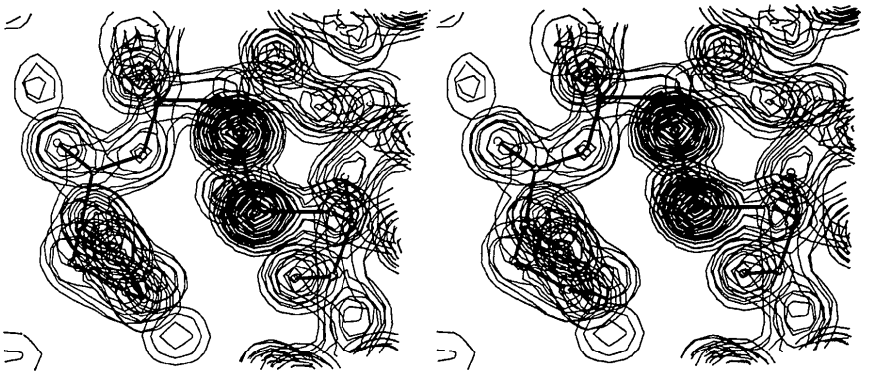


Fig. 5. Stereo pairs showing sections around (*a*) Phe 45, (*b*) Pro 8, (*c*) Pro 13–Cys 14–Cys 38 of the electron-density map calculated with refined phases (1·5 Å resolution). Contour lines are drawn at intervals of 0·5 e Å^3 starting with the 0·5 e Å^{-3} line. Parts of the refined model are drawn into the electron-density sections.



(*c*)

Fig. 5 (*cont.*)

creasing $|F_o|$. The mean differences of the multiple-isomorphous-replacement phases are about 10° greater than the error indicated by the figure of merit.

The effect of phase improvement on the electron density is shown in Fig. 4 with the example of Arg 53, an amino acid with temperature factors above the average B in the side-chain atoms. Refined atomic positions are drawn in identical sections of the 1·9 Å isomorphous map and of the refined electron density (1·5 Å resolution). Stereo pairs with sections around Phe 45, Pro 8, and Pro 13–Cys 14–Cys 38 are shown in Fig. 5(*a*), (*b*), (*c*) as examples of the final map. Fig. 6 shows the peptide Arg 42–Asn 43 as an example of deviation from planarity. The map and the fit are generally of the quality demonstrated in Figs. 4 to 6 except at the C terminus (from Gly 57 on) and at several external side chains.

4.2. *Description of the model*

Hydrogen bonds in the PTI molecule were located by calculating distances between atoms in the final model. A hydrogen bond was counted when the donor–acceptor distance was not greater than 3·1 Å. In Tables 6 and 7 the 29 intramolecular thus detected H bonds between PTI atoms (19 main-chain bonds, seven main-chain–side-chain bonds, and three side-chain bonds) are listed together with the donor–acceptor distance. Many of these bonds have already been postulated by Huber *et al.* (1970*a*), and Huber, Kukla, Rühlmann & Steigemann (1970*b*, 1971).

The network of hydrogen bonds between main-chain atoms (Table 6) shows clearly the two domains of secondary structure: a double stranded antiparallel β structure, composed of residues 16 through 36, which is twisted about 180°, and nearly three turns of a right-handed α helix composed of residues 47 through 56 (Huber *et al.*, 1970*a*, 1970*b*, 1971). An analysis of the conformational angles of residues 17–24 and 29–35 shows that the antiparallel β structure is considerably distorted. The mean value of φ in this region is $-114°$ with a standard deviation of 24°; the mean ψ angle is 139° with a standard deviation of 26°. These values have to be compared with the corresponding standard values of $\varphi=-139°$ and $\psi=135°$ for a regular antiparallel β structure (Arnott, Dover & Elliott, 1967). The right-handed α helix is very regular between residues 48 and 54. Here the mean values of φ and ψ are $-67°$ and $-41°$ respectively, with a standard deviation of only 6° in both cases. The mean values are very

Table 6. *Intramolecular hydrogen bonds*

Main chain	Main chain	Distance (Å)	
Thr 11 C=O···NH	Gly 36	2·9	
Ala 16 NH ···O=C	Tyr 35	3·1	Antiparallel β-sheet
Ile 18 C=O···NH	Tyr 35	2·7	
Arg 20 NH ···O=C	Phe 33	3·1	
Arg 20 C=O···NH	Phe 33	3·0	
Phe 22 NH ···O=C	Gln 31	2·9	
Phe 22 C=O···NH	Gln 31	2·9	
Asn 24 C=O···NH	Gly 28	2·9	
Asn 24 C=O···NH	Ala 27	3·1	
Asn 24 NH ···O=C	Leu 29	3·0	
Tyr 21 NH ···O=C	Phe 45	2·7	
Tyr 21 C=O···NH	Phe 45	2·8	
Ser 47 C=O···NH	Cys 51	3·1	C-terminal helix
Ala 48 C=O···NH	Met 52	2·8	
Glu 49 C=O···NH	Arg 53	3·0	
Asp 50 C=O···NH	Thr 54	3·1	
Cys 51 C=O···NH	Cys 55	2·8	
Met 52 C=O···NH	Gly 56	2·8	

Table 7. *Intramolecular hydrogen bonds*

Main chain	Side chain	Distance (Å)
Glu 7 C=O···NH_2	Asn 43	2·8
Tyr 23 C=O···NH_2	Asn 43	3·1
Tyr 23 NH ···O=C	Asn 43	2·8
Val 34 C=O···OH	Thr 11	3·0
Cys 38 NH ···OH	Tyr 35	3·0
Asp 50 NH ···OH	Ser 47	3·1
Cys 55 C=O···NH	Arg 1	2·8

Side chain	Side chain	Distance (Å)
Arg 20 NH_2·····O=C	Asn 44	2·8
Asn 24 O=C−NH_2···O=C−NH_2	Gln 31	2·8
Asp 50 C=O·····NH_2	Arg 53	2·9

close to the standard values for a right-handed α helix of $\varphi = -57°$ and $\psi = -47°$ (Arnott & Dover, 1967).

In addition to the hydrogen bonds listed in Tables 6 and 7, 12 intramolecular H bonds are contributed by four internal water molecules. Fig. 7 shows one of them, which lies in a pocket near the 14–38 disulphide bridge. It is hydrogen-bonded to the carbonyl oxygens of Thr 11 and Cys 38 with donor–acceptor distances of 2·7 and 2·8 Å respectively. Two additional H bonds exist between the water molecule and the nitrogens of Cys 14 and Cys 38; the distances of 3·1 and 3·2 Å are quite large. The four nearest neighbours of this water molecule lie at the corners of a distorted tetrahedron.

The other three internal water molecules can be seen in Fig. 8. They occupy the space in a cleft between residues 8 to 10 and residues 40 to 44. The end of the cleft is marked by the benzene ring of Phe 33. The two inner water molecules are tetrahedrally coordinated. The distances of the hydrogen-bonded atoms fall between 2·7 and 3·0 Å. All four internal water molecules seem to be part of the structure of the native PTI molecule; they can be expected at the same positions in solution.

The 43 external water molecules which could be located during the refinement fill about one third of the crystal volume between PTI molecules. 42 H bonds between polar groups of the protein and ordered water,

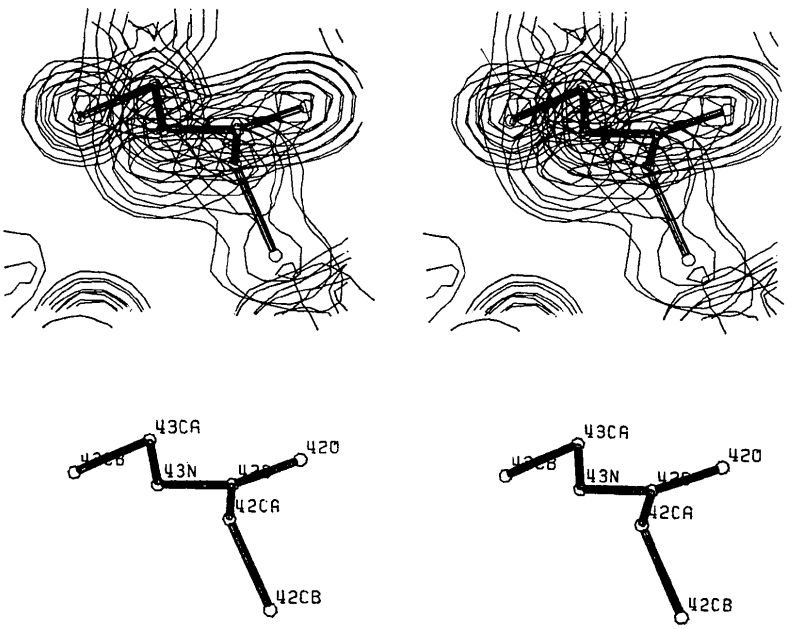


Fig. 6. Stereo pairs of peptide Arg 42–Asn 43 with $\omega = 167°$. Top: refined model with refined electron density (1·5 Å resolution; contours as in Fig. 5); bottom: refined model with labelled atoms.

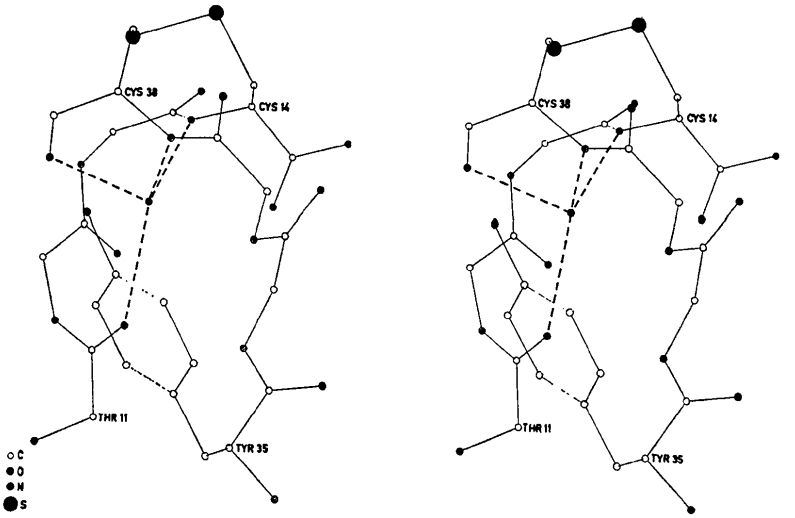


Fig. 7. Stereo pair showing parts of the refined PTI model around one internal water molecule (Thr 11 to Cys 14 and Tyr 35 to Cys 38). Protein–water hydrogen bonds are dashed.

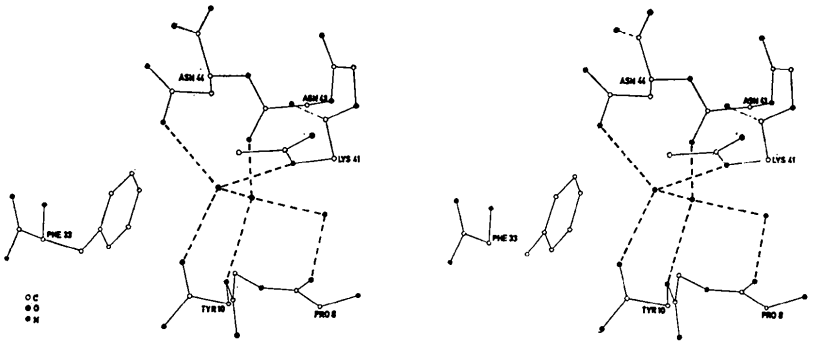


Fig. 8. Stereo pair showing three internal water molecules and neighbouring PTI chains (Pro 8 to Tyr 10, Ala 40 to Asn 44, and Phe 33). Hydrogen bonds are dashed.

and 12 water–water H bonds were detected. These bonds involve all but two of the localized water molecules. The degree of order in the water structure is greatest where neighbouring PTI molecules in the crystal come close together. In such regions five water bridges consisting of up to five molecules are formed. The mean donor–acceptor distance of H bonds in which external water molecules are involved is 2·9 Å with a standard deviation of 0·17 Å.

One close contact was found in the final model between the carbonyl oxygen of Phe 4 and C^{γ} of Glu 7 which are only 2·4 Å apart. The high temperature factor of the C^{γ} and the difference Fourier maps indicate that this atom occupies two alternative positions which are both more than 3 Å away from the oxygen. The model position is midway between these alternatives.

Histograms of the refined χ^1 and χ^2 angles of side chains are shown in Fig. 9. For the χ^1 angles the values which correspond to a staggered conformation (−60, 60, and 180°) are clearly preferred; only Lys 15 is in an eclipsed conformation. For the χ^2 angles also the staggered positions are preferred but the number of exceptions is much greater.

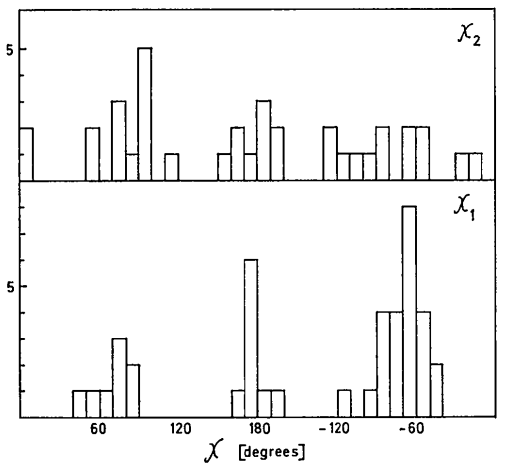


Fig. 9. Histograms of refined χ^1 and χ^2 angles of PTI (Glu 7 not included).

Three of the four prolines (8, 9, and 13) have the conformation C^{γ}-*endo* (Ramachandran, Lakshminarayanan, Balasubramanian & Tegoni, 1970), whereas the ring of Pro 2 has C^{γ}-*exo* (see Table 5).

Histograms of $\tau(NC^{\alpha}C)$ angles and ω angles are shown in Fig. 10 and Fig. 11. The mean value of $\tau(NC^{\alpha}C)$ is 111·8° with a standard deviation of 5·8°; the mean of ω is 179·3° with 7·1° standard deviation. The fact that the ω distribution is broader than the τ distribution fits the assumption that the potential energy function is less steep for an ω torsion than for a change in τ (Ramachandran, 1969; Winkler & Dunitz, 1971).

As indicated in Fig. 11, the peptides at the binding site of the PTI to trypsin, especially the susceptible Lys 15–Ala 16 bond, are particularly non-planar. As a result of the ω torsion, the carbonyl oxygen of Lys 15 is shifted part of the way towards the position it would occupy if the carbonyl carbon were tetrahedrally coordinated, as has been observed in the trypsin–PTI complex (Huber, Kukla, Steigemann, Deisenhofer & Jones, 1973).

5. Conclusion

The refinement of PTI has shown that the method described in this paper can yield an accurate protein model in which information from small-structure X-ray work is conserved as much as possible. The reduction in the number of model parameters by constraint suggests the application of the method to protein structures also in cases where the resolution is considerably lower than 1·5 Å.

Refinement similar to that described here on the trypsin–PTI complex at 1·9 Å resolution (Huber, Kukla, Bode, Schwager, Bartels, Deisenhofer & Steigemann, to be published) has been completed. The refinement of Bence–Jones variable-part dimer at 2·5 Å resolution (Epp *et al.*, personal communication) is in an advanced state with promising results.

We thank Dr R. Huber for helpful and stimulating discussions during all stages of this work. We are greatly indebted to Dr R. Diamond who kindly pro-

vided us with his real-space refinement program. We also thank the referee, Dr P. M. Colman, and Dr E. E. Lattman for critical comments on the manuscript. Frau K. Epp and Frau E. Preuss helped us with the drawings.

This work was supported by Deutsche Forschungsgemeinschaft and Sonderforschungsbereich 51.

References

ANDERER, F. & HÖRNLE, S. (1966). *J. Biol. Chem.* **241**, 1568–1572.

ARNOTT, S. & DOVER, S. D. (1967). *J. Mol. Biol.* **30**, 209–212.

ARNOTT, S., DOVER, S. D. & ELLIOTT, A. (1967). *J. Mol. Biol.* **30**, 201–208.

CRUICKSHANK, D. W. J. (1949). *Acta Cryst.* **2**, 65–82.

DEISENHOFER, J. & STEIGEMANN, W. (1973). In *2nd International Research Conference on Proteinase Inhibitors* (*Bayer Symposium* V), edited by H. FRITZ, H. TSCHESCHE, L. J. GREENE & E. TRUSCHEIT. New York, Heidelberg, Berlin: Springer Verlag. In the press.

DIAMOND, R. (1966). *Acta Cryst.* **21**, 253–266.

DIAMOND, R. (1971). *Acta Cryst.* A**27**, 435–452.

DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.

FORSYTH, J. B. & WELLS, M. (1959). *Acta Cryst.* **12**, 412–415.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

HUBER, R. & KOPFMANN, G. (1969). *Acta Cryst.* A**25**, 143–152.

HUBER, R., KUKLA, D., RÜHLMANN, A., EPP, O. & FORMANEK, H. (1970*a*). *Naturwissenschaften*, **57**, 389–392.

HUBER, R., KUKLA, D., RÜHLMANN, A. & STEIGEMANN, W. (1970*b*). In *Proceedings of the International Conference on Proteinase Inhibitors, Munich* 1970, edited by H. FRITZ & H. TSCHESCHE, pp. 56–65. Berlin: Walter de Gruyter.

HUBER, R., KUKLA, D., RÜHLMANN, A. & STEIGEMANN, W. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 141–150.

HUBER, R., KUKLA, D., STEIGEMANN, W., DEISENHOFER, J. & JONES, T. A. (1973). In *2nd International Research Conference on Proteinase Inhibitors* (*Bayer Symposium* V), edited by H. FRITZ, H. TSCHESCHE, L. J. GREENE & E. TRUSCHEIT. New York, Heidelberg, Berlin: Springer Verlag. In the press.

IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (1970*a*). *J. Biol. Chem.* **245**, 6489–6497.

IUPAC-IUB COMMISSION ON BIOCHEMICAL NOMENCLATURE (1970*b*). *Biochemistry*, **9**, 3471–3479.

KASSELL, B., RADICEVIC, M., ANSFIELD, J. & LASKOWSKI, M. SR (1965). *Biochem. Biophys. Res. Commun.* **18**, 255–258.

LEVITT, M. & LIFSON, S. (1969). *J. Mol. Biol.* **46**, 269–279.

RAMACHANDRAN, G. N. (1969). *Protein Res.* **1**, 5–17.

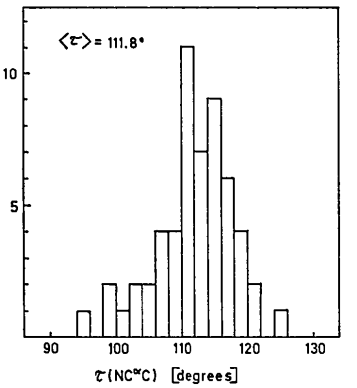


Fig. 10. Histogram of refined $\tau(NC^{\alpha}C)$ angles of PTI.

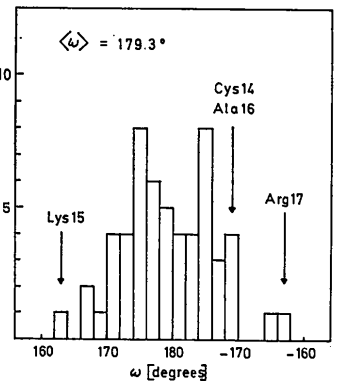


Fig. 11. Histogram of refined ω angles of PTI. Non-planar peptides at the binding site of PTI to trypsin are marked.

RAMACHANDRAN, G. N., LAKSHMINARAYANAN, A. V., BALASUBRAMANIAN, R. & TEGONI, G. (1970). *Biochim. Biophys. Acta*, **221**, 165–181.

SAYRE, D. (1972). *Acta Cryst.* A**28**, 210–212.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.

SCHERAGA, H. A. (1968). *Advanc. Phys. Org. Chem.* **6**, 103–184.

STEIGEMANN, W. (1974). PhD thesis, Technische Univ. München.

VINCENT, J. P., CHICHEPORTICHE, R. & LAZDUNSKI, M. (1971). *Eur. J. Biochem.* **23**, 401–411.

WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.

WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.

WINKLER, F. K. & DUNITZ, J. D. (1971). *J. Mol. Biol.* **59**, 169–182.