Table 5. *Dihedral angles between all-anti parts in the* C_{15} *ring* (°)

E.s.d.'s are in parentheses.

C(12)···C(15),C(8)···C(12)	87·7 (4)
C(8)···C(12),C(5)···C(8)	88·8 (4)
C(5)···C(8),C(1)···C(5)	88·4 (4)
C(1)···C(5),C(12)···C(15)	88·8 (4)
C(1)···C(5),C(8)···C(12)	1·8 (5)
C(5)···C(8),C(12)···C(15)	21·9 (5)

The most striking aspect of the molecular packing is the H-bonding between molecules related by a centre of symmetry. This interaction, which certainly contributes to the existence of an ordered large-ring structure, is illustrated in Fig. 3.

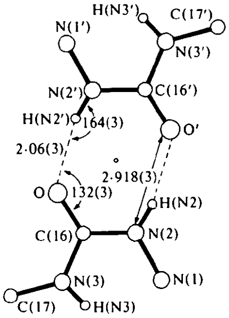


Fig. 3. Hydrogen-bonding scheme. Some distances (Å) and angles (°) are shown, with their e.s.d.'s in parentheses.

Wc thank Dr H. J. T. Bos and Mr L. J. de Noten for synthesizing and crystallizing the compound, Dr A. J. M. Duisenberg for collecting the data, and Dr P. Groth, University of Oslo, for valuable discussions.

References

Bixon, M. & Lifson, S. (1967). *Tetrahedron*, **23**, 769–784.
Cromer, D. T. & Mann, J. B. (1968). *Acta Cryst.* A**24**, 321–324.
Dale, J. (1973). *Acta Chem. Scand.* **27**, 1115–1129.
Groth, P. (1974). *Acta Chem. Scand. Ser. A*, **28**, 294–298.
Groth, P. (1976). *Acta Chem. Scand. Ser. A*, **30**, 294–296.
Hendrickson, J. B. (1969). *J. Am. Chem. Soc.* **89**, 7036–7043.
Johnson, C. K. (1965). *ORTEP*. Report ORNL-3794. Oak Ridge National Laboratory, Tennessee.
Main, P., Lessinger, L., Woolfson, M. M., Germain, G. & Declercq, J. P. (1977). *MULTAN. A System of Computer Programs for the Automatic Solution of Crystal Structures for X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.
Stewart, J. M. (1976). The XRAY 76 system. Tech. Rep. TR-446. Computer Science Center, Univ. of Maryland, College Park, Maryland.
Stewart, R. F., Davidson, E. R. & Simpson, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.
Wiberg, K. B. J. (1965). *J. Am. Chem. Soc.* **87**, 1070–1078.

Acta Cryst. (1979). B**35**, 1861–1874

The Accuracy of Refined Protein Structures: Comparison of Two Independently Refined Models of Bovine Trypsin

By John L. Chambers and Robert M. Stroud

University of California, San Francisco School of Medicine, Department of Biochemistry and Biophysics, San Francisco, California 94143, *USA*

(*Received* 8 *December* 1978; *accepted* 24 *April* 1979)

Abstract

The structure of diisopropyl-fluorophosphate-inhibited bovine trypsin has been refined to a standard crystallographic residual of $R = 0{\cdot}157$ at 1·5 Å resolution for a constrained model (C&S coordinates). Benzamidine-inhibited bovine trypsin has also been independently refined ($R = 0{\cdot}229$ at 1·8 Å) by Bode & Schwager [*J. Mol. Biol.* (1975), **98**, 693–717] (B&S coordinate set). Comparison of these structures after suitable correction for the different inhibitors and consequent structural differences permits an experimental determination of the differences in structure, which places an upper limit on the errors in the refined coordinate sets. The models are remarkably similar in well determined regions. The average positional difference between internal main-chain atoms is 0·146 Å (0·163 Å r.m.s.); however, there are some differences as large as 3–9 Å associated mostly with poorly determined external side chains. The magnitude of the deviations is strongly dependent on the region of the structure compared, and is closely related to refined thermal parameters in our analysis.

0567-7408/79/081861-14$01.00

Estimated errors in the C&S atom positions average 0·15–0·20 Å overall, and range from about 0·10 Å in well determined areas to 1 Å or more in poorly determined regions. Errors in the B&S positions are expected to be somewhat larger. Comparison of structure factors computed from the B&S coordinates [$F_{c(\mathrm{BS})}$] gives a residual with the C&S observed data of $R = 0{\cdot}238$ at 1·8 Å (very close to their own residual of 0·229), indicating that the contribution from real differences between the two 'true' structures or from errors in either observed data set is small compared with errors in the refined coordinate sets. The residual between $F_{c(\mathrm{BS})}$ and structure factors computed from the C&S model was 0·253 at 1·5 Å, suggesting that the two models do not differ from the 'true' structure in a systematically identical fashion, and that further refinement should improve either model. The similarity of the 0·253 value to their own residual of 0·229 at 1·8 Å also suggests that the differences between the C&S and B&S models are about as large as the differences between the B&S structure and the hypothetical 'true' structure. The differences between the two models thus give a reasonable estimate of the kinds of errors which can be expected in a structure with refinement statistics similar to those of Bode & Schwager.

Introduction

The structure of diisopropylphosphoryl (DIP)-inhibited bovine trypsin has been refined at 1·5 Å resolution to an R factor of 0·157 for a highly constrained model (Chambers & Stroud, 1977*a*,*b*). An independent refinement of bovine trypsin has also been carried out (Bode & Schwager, 1975). These independently refined structures make possible a unique comparison, a comparison between essentially the same protein structure refined completely independently in different laboratories by different procedures.

One aim of the following comparison is to obtain reasonable estimates of the errors in protein structures at the levels of refinement of these two trypsin models. Use of protein coordinates for studies of protein folding, energetic contributions of strain, *etc*., or enzyme mechanism depends on an understanding of the limitations of accuracy in the coordinate sets. In most cases, the accuracy of atomic positions quoted by the crystallographer are theoretical estimates. Bode & Schwager (1975) estimated standard deviations of their coordinates to be less than 0·1 Å, which in view of the comparison made here significantly underestimates the real differences in position between the two trypsin structures, except for the best-determined atoms.

A second objective of the comparison is to gain insight into the probable sources of error in refined protein structures and so determine where the limitations of current assumptions, or methods, lie. The most common index used for the state of refinement of a crystallographically determined structure is a direct comparison of the observed diffraction amplitudes (F_o) with those predicted by the refined model (F_c). The error is expressed as a residual $R = \sum_{hkl} |F_o - F_c| / \sum F_o$ and for most of the refined high-resolution structures to date this residual has converged to around $R = 0{\cdot}20$–$0{\cdot}25$.

Although the model and the residual are considerably improved over those at the start of refinement, residuals in this range are high compared with those obtained for smaller structures. Because $\mathbf{F}_o$ and $\mathbf{F}_c$ are vectors, the residual represents only about $1/\sqrt{2}$ of the average magnitude of the difference vector, $|\mathbf{F}_o - \mathbf{F}_c|$. Thus, a residual of 0·25 implies that about one-third of the structure factor has been unaccounted for. Are these residuals relatively high because of inadequacy in the method of representing the structure (isotropic vibration, unique position for each atom, absence of hydrogen atoms, *etc*.) or because of poor accuracy of the observed data, or should continued refinement of such a model be expected to bring about an additional major improvement in this residual?

The coordinates for both our DIP-trypsin structure and those of trypsin complexed to pancreatic trypsin inhibitor, determined by Huber, Kukla, Bode, Schwager, Bartels, Deisenhofer & Steigemann (1974) [the starting model for the refinement of Bode & Schwager (1975)] have previously been compared with those of α-chymotrypsin (Birktoft & Blow, 1972) with similar results in terms of the mean deviation between atom coordinates for homologous sites in the structures (Kossiakoff, Chambers, Kay & Stroud, 1977). However, the comparison of the two trypsin structures yields new information about the accuracy of protein structures *per se*; therefore we first summarize the current status of, and procedures used for, each of the structures under comparison.

Experimental procedures

The refined structures of bovine trypsin

(1) *Chambers & Stroud* (*C&S structure*). The structure of bovine trypsin was determined initially by the multiple-isomorphous-replacement method (Stroud, Kay & Dickerson, 1971, 1974). Crystals of the DIP-inhibited enzyme were grown from 7% $MgSO_4$ solutions at pH 6·8. The crystals were found to contain approximately 60% bovine β-trypsin (uncleaved) and 40% α-trypsin (autolytically cleaved between Lys 145 and Ser 146).

Crystallographic data for the current structure were collected to a resolution of 1·35 Å ($2\theta = 69{\cdot}6°$, Cu $K\alpha$ radiation; 48 492 independent *hkl* reflections) on a full-circle diffractometer modified to reduce background levels (Krieger, Chambers, Christoph, Stroud & Trus, 1974; Krieger & Stroud, 1976) and data were corrected as described by Stroud *et al.* (1974).

The structure was refined by a difference Fourier method coupled with idealization of molecular geometry (Chambers & Stroud, 1977*a*,*b*). At the time of this comparison the constrained structure had been refined to a standard crystallographic residual ($R = \sum |F_o - F_c|/\sum F_o$) of 0·157 at 1·5 Å resolution. Since reflections with $F_o \gg F_c$ are likely to be poorly phased in the difference Fourier syntheses (Stout & Jensen, 1968) terms with $(F_o - F_c)/0{\cdot}5(F_o + F_c) > 1{\cdot}2$ (currently 91 of the 22 117 reflections with $F_o > 3\sigma$ to 1·5 Å resolution) were omitted from the refinement. Difference terms with $F_c \gg F_o$ have a much higher probability of being correctly phased, and were included in the C&S calculations. For purposes of comparison the low-angle reflections in the resolution range ∞–7 Å were left out of the calculations of all residuals cited here since they were omitted by Bode & Schwager (1975) in their analysis.

An individual isotropic temperature factor, *B*, was refined for each atom along with the atomic coordinates *x*, *y*, and *z*. Approximately 120 ordered solvent molecules have been located in the current structure. A recent set of coordinates is available from the authors, or from the Brookhaven Protein Data Bank, Department of Chemistry, Brookhaven National Laboratory, Upton, Long Island, New York 11973.

(2) *Bode & Schwager* (*B&S structure*). The structure of bovine trypsin has been independently refined by Bode & Schwager (1975) using the real-space refinement program of Diamond (1971, 1974) alternating with calculation of new electron density maps phased on the model, and with inspection of difference Fourier maps for large errors (Deisenhofer & Steigemann, 1975).

Crystals of the enzyme (benzamidine-inhibited) were grown from $(NH_4)_2SO_4$ solutions at pH 7·0 and contained about 80% β-trypsin and 20% α-trypsin. [Crystals of benzamidine-inhibited trypsin are isomorphous with the pH 6·8 DIP-trypsin crystals (Krieger, Kay & Stroud, 1974).] Crystallographic data to 1·8 Å resolution were recorded on film by the rotation method. The trypsin coordinates of Huber *et al.* (1974) for the trypsin–pancreatic trypsin inhibitor complex were rotated to match the benzamidine-trypsin data (Fehlhammer & Bode, 1975). The orientation was the same as that determined by Stroud *et al.* (1971).

The residual between calculated and observed structure factors for the refined structure of Bode & Schwager (1975) was $R_{\rm BS} = 0{\cdot}229$ to 1·8 Å resolution. Reflections with $F_o \gg F_c$ *or* $F_c \gg F_o$ were omitted from their refinement and from their calculation of *R* according to the criterion $|F_o - F_c|/0{\cdot}5(F_o + F_c) > 1{\cdot}2$, as were the low-angle reflections from ∞–6·8 Å resolution. Thermal parameters, *B*, were tentatively assigned to each atom on the basis of the atomic radii obtained by real-space refinement; however, the residual was not significantly improved by this procedure, and a constant overall temperature factor was assumed. About 50 ordered solvent molecules were located in the refined structure. The coordinates for this comparison were obtained from the Brookhaven Protein Data Bank (entered on 24 January 1977).

Information concerning the two independently refined structures is summarized in Table 1.

The relative states of refinement

The positional differences (Δr) between atoms in the two models contain contributions from the errors in each structure. In the case where random coordinate errors are the major component of Δr (as appears to be true in this study), the expected relationship is $\langle \Delta r\rangle_{\rm r.m.s.} \simeq (\langle\sigma^2_{\rm CS}\rangle + \langle\sigma^2_{\rm BS}\rangle)^{1/2}$, where $\langle\sigma^2_{\rm CS}\rangle$ and $\langle\sigma^2_{\rm BS}\rangle$ represent the mean variance in atomic position in the two coordinate sets. The Δr values will thus be dominated by the structure with larger errors, and some knowledge of the relative magnitudes of $\langle\sigma^2_{\rm CS}\rangle$ and $\langle\sigma^2_{\rm BS}\rangle$ would be useful for estimating the errors from the Δr values.

The refinement statistics indicate that the standard deviations of the C&S coordinates should be less than those for the B&S model. Nevertheless, the value of the crystallographic residual depends on factors in addition to the correctness of the model, such as the effective number of parameters included in the refinement (Moews & Kretsinger, 1975). Comparison of *R* factors

Table 1. *Summary of refinement statistics for the DIP-trypsin* (*Chambers & Stroud,* 1977*a*,*b*; *C&S*) *and benzamidine-trypsin* (*Bode & Schwager,* 1975; *B&S*) *structures*

	C&S	B&S*
Unit-cell dimensions	$a = 54{\cdot}84$ Å $b = 58{\cdot}61$ $c = 67{\cdot}47$	$a = 54{\cdot}89$ Å $b = 58{\cdot}52$ $c = 67{\cdot}63$
Space group	$P2_12_12_1$	$P2_12_12_1$
Molecules per asymmetric unit	1	1
Crystallization conditions	7% $MgSO_4$ Tris buffer pH 6·8	2·4 *M* $(NH_4)_2SO_4$ Phosphate buffer pH 7·0 (10^{-4} *M* $CaCl_2$, 10^{-2} *M* benzamidine)
Data collection	Diffractometer	Film (rotation method)
Resolution of current structure	1·5 Å	1·8 Å
Number of independent reflections to indicated resolution	35 566	20 853
Threshold for unobserved reflections	$\lvert F_o\rvert > 3\sigma$	$\lvert F_o\rvert > 2\sigma$
Number of reflections above threshold	22 117	16 600
Refinement method	Difference Fourier	Real-space†
Residual $R = \sum \lvert F_o - F_c\rvert/\sum F_o$	15·7%	22·9%

* Also Fehlhammer & Bode (1975).

† Diamond (1971); Deisenhofer & Steigemann (1975).

for two different determinations thus does not necessarily provide a reliable assessment of their relative accuracy (Lipson & Cochran, 1957*a*).

In the C&S structure individual temperature factors have been refined and more ordered solvent molecules have been located than in the B&S structure. Even so, $R_{\rm CS}$ computed with a constant temperature factor for all atoms and including only the 50 best-determined solvent molecules was $R'_{\rm CS} = 0{\cdot}199$ at 1·5 Å (22 117 observations), compared with $R_{\rm BS} = 0{\cdot}229$ at 1·8 Å (16 600 observations). Some improvement in $R'_{\rm CS}$ would be expected if atomic positions were actually refined subject to these restrictions, although the resulting coordinates would presumably be less accurate. The difference between the residuals is thus not solely a result of inclusion of these parameters in the C&S refinement.

The effect on the residuals of the different methods used to maintain 'ideal' molecular geometry in the C&S and B&S refinement schemes is more difficult to determine. The procedure used by Bode & Schwager (1975) is predominantly a rigid-group refinement (Diamond, 1971, 1974), although flexibility was allowed in the $N-C_\alpha-C$ bond angles (τ) and in the dihedral angles (ω) determining the planarity of peptide amides. The C&S procedure is analogous to energy minimization (Levitt & Lifson, 1969; Levitt, 1974), simultaneously minimizing deviations from 'ideal' bond lengths, bond angles, and dihedral angles, and movements of the atoms from optimal positions determined in difference Fourier syntheses. Average deviations of the C&S coordinates from 'ideal' values [obtained from Marsh & Donohue (1967), and references therein] were 0·022 Å for bond lengths, 2·7° for bond angles, and 4·6° for dihedral angles. The corresponding mean deviations of the B&S coordinates from these same 'ideal' values were 0·016 Å, 1·6°, and 7·7° respectively. The deviations in bond lengths and, to a lesser extent, bond angles for the B&S coordinates probably arise largely from the slightly different sets of 'ideal' values used in the two refinement schemes. The mean deviation of the C&S bond lengths from ideality is comparable to the differences in bond lengths observed between different structure determinations of the same amino acid. For many of the amino acid structures used as standards by Chambers & Stroud (1977*a*,*b*) or by Diamond (1974), standard deviations of atomic positions are in the range 0·01–0·02 Å (Marsh & Donohue, 1967). Requiring thc C&S coordinates to conform much more closely to a set of standard groups might thus introduce small systematic errors, and it is felt that the present degree of rigidity at this stage of refinement represents a good compromise between freedom from such errors and maintenance of very reasonable stereochemistry.

The deviations of the B&S dihedral angles from ideality result from nonplanarity of the peptide amides. The fact that this deviation (7·7°) is larger than that for the C&S coordinates (4·6°) suggests that a greater degree of flexibility is required in the smaller number of geometrical parameters used by Bode & Schwager (1975) than in the corresponding C&S parameters, in order to best fit the electron density. This idea is supported by the significantly larger mean deviation of the $N-C_\alpha-C$ bond angle, τ, from ideality ($\tau_{\rm ideal} = 111{\cdot}0°$; Marsh & Donohue, 1967) in the B&S structure (5·5°) compared with that in the C&S structure (2·8°).

Thus, while more geometrical variables are allowed in the C&S structure, determination of the effective number of parameters in each refinement, taking into account the degree of flexibility in the various restraints, is not a simple matter, and a full treatment of the problem is not suitable here. In terms of the accuracy of the structure, the important consideration is not the number of parameters in the refinement, but whether introduction of these parameters brings about an improvement in the model. As discussed below, the individual thermal parameters are physically reasonable and are good indicators of the reliability of the structure in a given region. Likewise, the C&S idealization procedure allows strain to be distributed in an energetically reasonable fashion.

Because of the problems in quantitatively determining the relative degree of accuracy of two structure determinations based on their refinement statistics, the contribution of each structure to the $\langle \Delta r \rangle$ values will be assumed to be roughly equal in the subsequent calculations. The estimate of $\langle \sigma^2_{\rm CS} \rangle$ should probably be slightly smaller than the resulting values; that for $\langle \sigma^2_{\rm BS} \rangle$ should probably be somewhat larger.

Results and discussion

Comparison of the coordinates

In order to compare the two structures, the B&S coordinates *x*,*y*,*z* listed in Å, were transformed to match the C&S coordinates in Å according to the equation

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} (54{\cdot}89/2)\ \text{Å} \\ (58{\cdot}52/2)\ \text{Å} \\ 0{\cdot}0 \end{pmatrix}.$$

Orientations of groups which are symmetric with respect to a 180° rotation were made equivalent in the two structures (*e.g.* Asx and Glx carboxyl or amide groups, and Tyr and Phe rings). Regions of the molecules where there are differences between the DIP and benzamidine structures (Krieger, Kay & Stroud, 1974) were left out of the Δr calculations below.

Distances Δr were computed between all corresponding atomic positions in the two structures and are distributed as shown in Fig. 1. For the case where the errors in the atomic coordinates x,y,z are random, the vector differences $\mathbf{\Delta r}$ are expected to follow a normal distribution. However, the distribution in Fig. 1 is that of the scalar differences, $\Delta r = |\mathbf{\Delta r}|$, and the expected form of the probability distribution of $|\mathbf{\Delta r}|$ is thus:

$$P(|\mathbf{\Delta r}|) = \left(\frac{2}{\pi}\right)^{1/2} \frac{(\Delta r)^2}{\sigma_x^3} \exp[-(\Delta r)^2/(2\sigma_x^2)],$$

where σ_x is the overall standard deviation in the x coordinate, and it has been assumed that $\sigma_x = \sigma_y = \sigma_z$. In terms of the radial standard deviation in position, σ_r, this expression becomes:

$$P(|\mathbf{\Delta r}|) = \left(\frac{54}{\pi}\right)^{1/2} \frac{(\Delta r)^2}{\sigma_r^3} \exp[-(\Delta r)^2/(\tfrac{2}{3}\sigma_r^2)]. \quad (1)$$

These expressions are analogous to that for the speed distribution of randomly moving particles in three dimensions (Maxwell, 1860).

Most of the differences fall into the 0–1 Å range and appear to be distributed as expected for random errors. A distribution curve in the form of equation (1) refined to the differences in this range yielded a value of $\sigma_r = 0{\cdot}22$ Å. The r.m.s. value of Δr for all atoms where $\Delta r \leq 1$ Å was 0·26 Å. The range 1–10 Å contains systematic differences between the two structures. There are several differences as large as 3–9 Å which are associated mostly with atoms flagged by Bode & Schwager (1975) as not visible in their electron density map. All of the atoms so flagged (about 5% of all protein atoms, or 20% of all external side-chain atoms) are in side chains on the surface of the molecule and were not included in their structure factor calculations, or in their refinement. The maximum differences excluding these atoms were in the 2–4 Å range (see Fig. 1).

Since the flagged atoms were not visible in the electron density maps or refined in the B&S structure they are not as useful for estimation of random errors in the two structures. At the higher stage of refinement

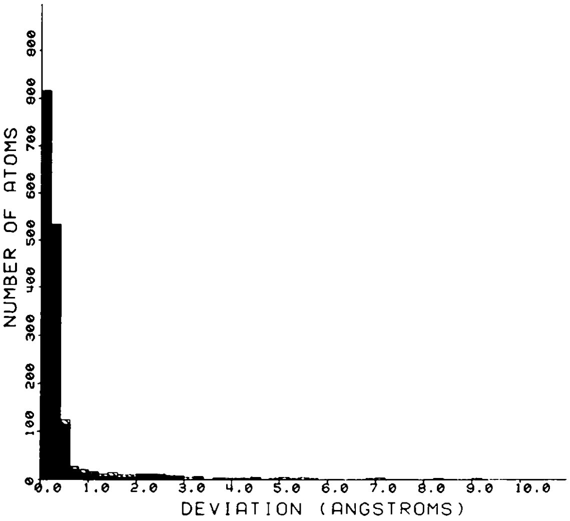


Fig. 1. Histogram indicating the distribution of the deviations between the C&S and B&S coordinate sets. The range 0–1 Å contains random errors, distributed with a standard deviation $\sigma = 0{\cdot}22$ Å. The range 1–10 Å contains systematic differences between the two structures. The striped bars represent the atoms flagged by Bode & Schwager (1975) as not visible in their density map.

Table 2. *Average positional differences, Δr (Å), between the C&S and B&S atomic coordinates*

	Internal only(a)			External only(a)			All		
	$\langle\Delta r\rangle_{\rm r.m.s.}^{(b)}$	$\langle\Delta r\rangle^{(c)}$	$\langle B\rangle^{(d)}$	$\langle\Delta r\rangle_{\rm r.m.s.}$	$\langle\Delta r\rangle$	$\langle B\rangle$	$\langle\Delta r\rangle_{\rm r.m.s.}$	$\langle\Delta r\rangle$	$\langle B\rangle$
(*A*) Including flagged atoms									
α-Carbon	0·153	0·139	6·9	0·286	0·224	11·3	0·254	0·198	10·0
Main chain	0·163	0·146	8·0	0·308	0·226	12·5	0·273	0·202	11·1
All atoms	0·420	0·223	8·9	0·973	0·460	15·0	0·835	0·388	13·1
Side chain	0·614	0·326	10·1	1·468	0·790	18·6	1·202	0·611	15·4
(*B*) Without flagged atoms									
α-Carbon	0·153	0·139	6·9	0·286	0·224	11·3	0·254	0·198	10·0
Main chain	0·163	0·146	8·0	0·308	0·226	12·5	0·273	0·202	11·1
All atoms	0·420	0·223	8·9	0·454	0·291	13·4	0·462	0·277	11·9
Side chain	0·614	0·326	10·1	0·636	0·408	15·0	0·635	0·380	13·0

(*a*) External segments (*i.e.* those in contact with the solvent regions surrounding the surface of the enzyme) included residues G18–P28, N34–H40, N48–Q50, C58–R65A (except V65 side chain), D71–N101, K107–L137, N143–V154, K156 side chain, K159–P161, L163–N179, Y184A–K188A, C201–K204, W215–G216, A221–P225, K230 side chain, C232–W237, K239–Q240 and A243–N245. The side chains of residues H40, F41, C58, I63, I73, I89, Y94, L99, V118, I121, L123, L137, L163, Y184A, W215, Y234, W237 and I242 are only partly accessible and were omitted from the calculations for either internal or external segments. Small conformational changes between DIP- and benzamidine-trypsin occur for H57, D189–S195, and S217–C220 (Krieger, Kay & Stroud, 1974) and these residues were therefore excluded from the above calculations. The remainder of the structure (about 30% of all atoms) was classified internal.

(*b*) $\langle\Delta r\rangle_{\rm r.m.s.}$ is the root-mean-square positional difference in Å.

(*c*) $\langle\Delta r\rangle$ is the average positional difference in Å.

(*d*) $\langle B\rangle$ is the average C&S individual temperature factor for the atoms compared in Å^2.

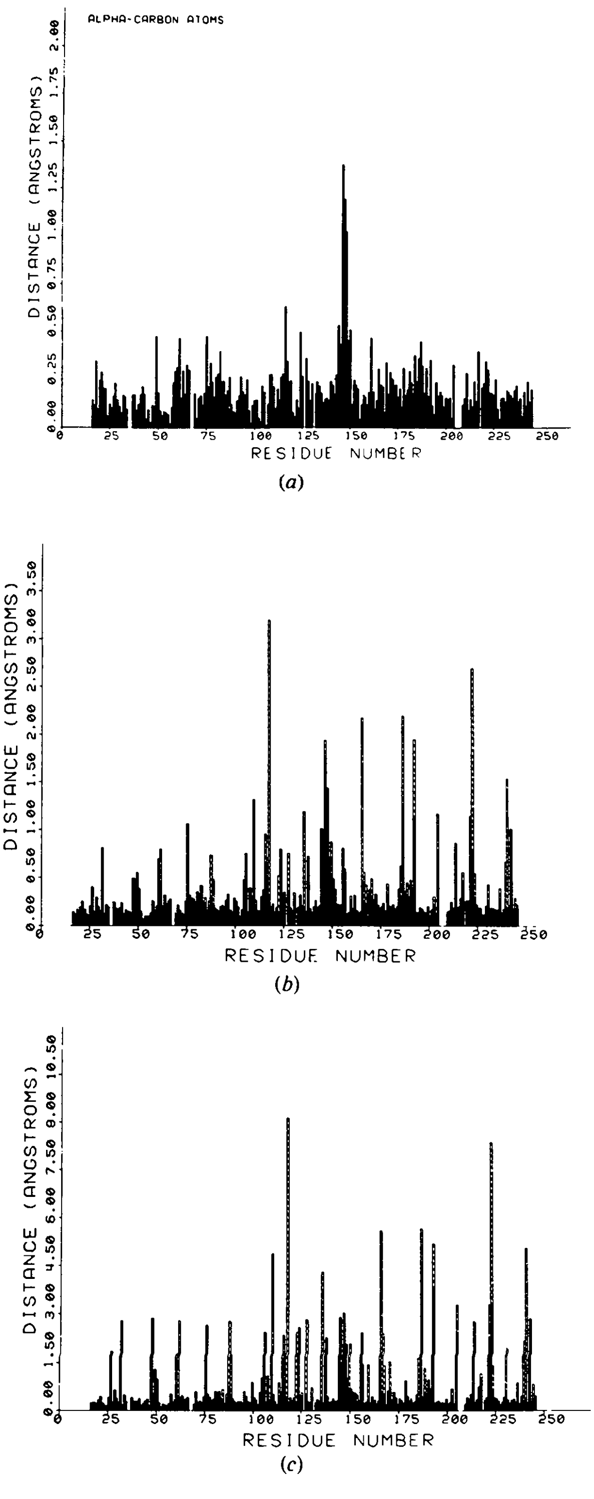
of the C&S structure, and with the use of individual temperature factors, there is quantitative evidence on their position and positional accuracy. Thus, Table 2 and Figs. 1 and 2 reflect the statistics both with and without inclusion of the flagged atoms. Without the flagged atoms, the good agreement over most of the structure is best represented and expected errors are reasonably predicted by equation (1). With inclusion of the flagged atoms, the differences represent the likely overall discrepancies or the reliabilities which are to be expected for similar structures at that stage of refinement.

The positional differences, summarized in Table 2, depend upon the location of the α-carbon, main-chain, or side-chain atoms in the structure. As expected, the differences are smallest for α-carbon atoms in the center of the molecule ($\langle \Delta r\rangle_{\rm r.m.s.} = 0{\cdot}153$ Å), greater for all main-chain atoms ($\langle \Delta r\rangle_{\rm r.m.s.} = 0{\cdot}273$ Å) and greatest for external side chains ($\langle \Delta r\rangle_{\rm r.m.s.} = 1{\cdot}47$ Å; 0·636 Å excluding flagged atoms, see Fig. 2). These differences closely follow the distribution of refined thermal parameters in the C&S structure as shown in Fig. 3. Thus, it appears that the uncertainty in placement of an atom is related to its refined individual temperature factor. Such a relationship is expected from theoretical considerations, as, for example, in the formula given by Cruickshank (1949) for estimating standard deviations of atomic coordinates.

Thirty eight of the ordered solvent molecules or ions in the C&S structure were also located in the B&S structure. The value of $\langle \Delta r\rangle_{\rm r.m.s.}$ for these atoms was 0·34 Å and their average (C&S) temperature factor was 14·7 Å^2. This positional difference is significantly smaller than those in Table 2 for atoms in the protein with temperature factors near this value, probably because these solvent-atom positions are relatively free from systematic errors imposed by misinterpretation or constraints. The higher-than-average temperature factor for these atoms is consistent with the fact that their r.m.s. positional difference is greater than the overall value derived from Fig. 1, for the component arising from random errors alone ($\sigma_r = 0{\cdot}22$ Å).

In a few cases it is clear that the differences are due to incorrect positioning of atoms in the B&S structure, where, for example, the internal Val 213 side chain is rotated 160° with respect to the position in the C&S structure where it is very well defined, as shown in Fig. 4. [Bode & Schwager (1975) mentioned that the Val 213 side chain might not be optimally positioned in their model because they found non-bonded close contacts with the carbonyl O atoms of residues 194 and 197.] However, to this point there has been minimal feedback between one set of coordinates and the other: in no case has one structure been modified in the light of the other.

One of the most common causes of the larger discrepancies was 120° or 180° rotation about χ_1 (the



(*a*)

(*b*)

(*c*)

Fig. 2. (*a*) Histogram of the deviation of α-carbon atom positions between the two structures according to their position in the sequence. (*b*) Histogram similar to (*a*), except that the deviation shown is the average for all atoms in the residue. The striped portions of the bars represent the contribution from the flagged atoms. (*c*) Deviations shown in this histogram are the maximum for each residue. The striped areas again represent flagged atoms.

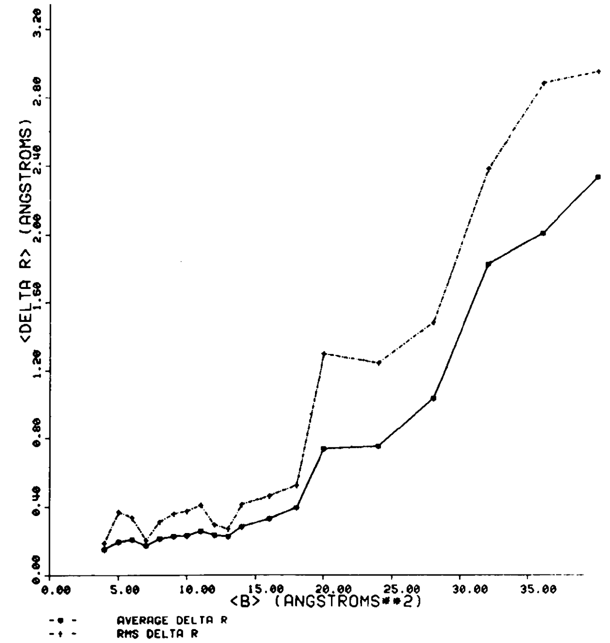


Fig. 3. Relation between $\langle \Delta r \rangle$ and the refined C&S temperature factors, B. The mean value of Δr is represented by the solid line; the broken line represents the root-mean-square Δr.

C_α—C_β bond) in valine and threonine residues, or about χ_2 (the C_β—C_γ bond) in leucine. The residues thus affected were T26,[*] V31, V75, L105, L123, L137, T144, T149 (flagged by Bode & Schwager), L155, L185 (flagged) and V213. Also common were large χ_1 rotations in serine residues, S61, S122, S127, and S166, although these in all cases involved atoms flagged by Bode & Schwager (1975). In longer side chains where the differences were large, Δr usually became larger proceeding from the α-carbon to the end of the side chain, resulting in a completely different position. This was the case with K87CE-NZ, K109CG-NZ, D165CB-OD2, E186CG-OE2, K204CG-NZ, Q221CG-NE2, K222CG-NZ, K239CD-NZ, and Q240CD-NE2. The largest difference in the main chain (2·12 Å for unflagged atoms) occurs at Ser 146. This is the site of the α–β trypsin autolytic cleavage, and there is some statistical disorder in this region in both structures. There are also large differences for the side chain of I47 where there is a large discrepancy (~160°) in χ_2, and for that of I242, where the entire side chain is rotated ~180° about the α–β bond (χ_1). The magnitudes of these differences are shown in Fig. 2(*a*)–(*c*).

Comparison of structure factors computed from the two models: origin of the positional differences

In order to investigate further the origin of the differences between the two models, calculated structure factor amplitudes and phases were generated for the C&S [$F_{c(CS)}$, $\alpha_{c(CS)}$] and B&S [$F_{c(BS)}$, $\alpha_{c(BS)}$] structures. In the case of the B&S coordinates, the benzamidine molecule and the anion bound at the catalytic site (702 OH in the notation of Bode & Schwager) were removed from the list and the coordinates of the DIP-group were inserted from the C&S structure, for comparison with our observed DIP-trypsin data [$F_{o(CS)}$]. The residual $R_3 = \sum |F_{o(CS)} - F_{c(BS)}|/\sum F_{o(CS)}$ between our observed amplitudes and those computed from this modified B&S coordinate set was 0·243 at 1·5 Å and 0·238 at 1·8 Å with the flagged atoms omitted from the calculation (with these atoms included the respective values were 0·254 and 0·251). For comparison the approximately 250 reflections with worst agreement according to the criterion of Bode & Schwager (1975) were excluded from this calculation. The closeness of R_3 to the residual quoted by Bode & Schwager with their own data ($R_{BS} = 0{\cdot}229$ at 1·8 Å) suggests that both the errors in the observed data and any real differences between the 'true' structures as they exist in the crystals [aside from the different inhibitors and small inhibitor-induced conformation changes (Krieger, Kay & Stroud, 1974)] are

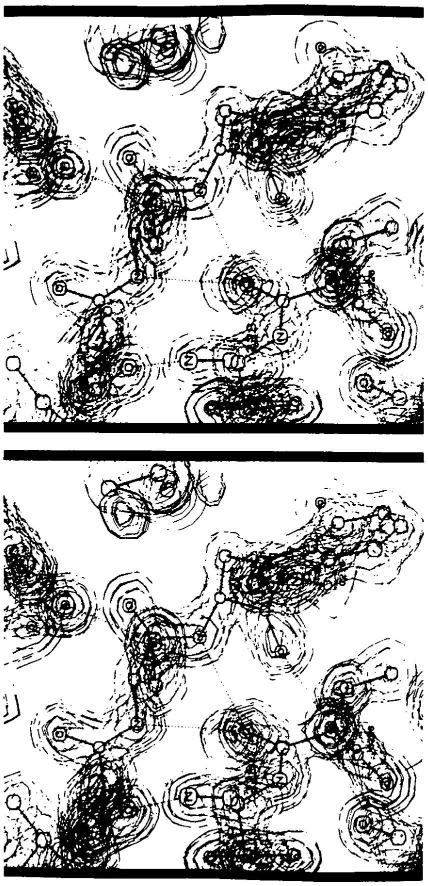


Fig. 4. The fit of the refined C&S model to an electron density map computed with coefficients $(2F_o - F_c)e^{i\alpha_c}$ is shown for the region from Val 213 to Trp 215. Such a synthesis reveals errors more clearly than an $F_oe^{i\alpha_c}$ synthesis. In the B&S structure the internal Val 213 side chain is rotated by 160° from the position shown here for the C&S structure. Contours are from 0·6 e Å^{-3} in 0·6 e Å^{-3} steps. The grid interval is 0·3 Å.

[*] The single-letter amino acid code is as listed in Table 1 of Stroud *et al.* (1971).

small compared with the lack of fit of the models to the observed data.

R_3 was also computed for the B&S coordinates after assigning each atom the same individual isotropic temperature factor as in the refined C&S structure. For this model R_3 fell to 0·233 at 1·5 Å (0·239 including flagged atoms). The fact that this decrease (ΔR_3 = 1·5%, when flagged atoms are included) was not as large as the increase in $R_{\rm CS}$ when individual temperature factors were omitted from the calculation ($\Delta R_{\rm CS}$ = 3%; Chambers & Stroud, 1977*a*) probably stems from the strong dependence of the refined temperature factors on atomic position.

The residual $R_4 = \sum |F_{c({\rm CS})} - F_{c({\rm BS})}|/\sum F_{c({\rm CS})}$ = 0·253 at 1·5 Å (excluding flagged atoms from the B&S model) provides important information about the characteristics of the differences in structure. If the differences between the C&S structure and the 'true' structure were uncorrelated with those between the B&S structure and the 'true' structure R_4 would have a value near 0·278 $\{=[(0{\cdot}157)^2 + (0{\cdot}229)^2]^{1/2}\}$. A value for R_4 of 0·072 representing the minimum possible difference between the C&S ($R_{\rm CS}$ = 0·157) and B&S ($R_{\rm BS}$ = 0·229) structures would suggest that both structures are about as close to the 'true' structure as allowed by the assumptions of unique positions for all atoms, isotropic thermal vibration and discreet solvent molecules, and by omission of the H atoms from the structure factor calculations. This is not the case, however, since this residual was R_4 = 0·253. Moreover, the terms $\Delta F_{\rm BS} = F_{o({\rm CS})} - F_{c({\rm BS})}$ were of opposite sign from $\Delta F_{\rm CS} = F_{o({\rm CS})} - F_{c({\rm CS})}$ for 40% of the reflections (a 50% value would indicate uncorrelated errors). The systematic component arising from the means of representing the structure is thus relatively small, suggesting that the major component of the disagreement between F_o's and F_c's can be accounted for by further refinement, even with the *same* assumptions of isotropic vibration, *etc.* It should be noted, however, that these figures apply only to the data in the 7–1·5 Å range. The correlation between $\Delta F_{\rm CS}$ and $\Delta F_{\rm BS}$ is very high in the ∞–7 Å range where the F_c's are in both cases considerably larger on average than the $F_{o({\rm CS})}$'s, since no representation of the solvent continuum was included in either model.

Table 3. σ_r (Å) *for all atom types in the refined C&S structure, computed from the formula of Cruickshank* (1949)

These values are underestimates of the real errors, as described in the text. The best estimates of the errors in the C&S structure are about twice the values in this table. The average temperature factor for all atoms in the C&S structure (excluding solvent) was 12 Å^2.

B (Å^2)	C	N	O	P	S	Ca^{2+}
4	0·076	0·060	—	—	—	—
5	0·081	0·063	0·051	—	—	—
6	0·087	0·068	0·055	—	—	—
7	0·093	0·072	0·058	—	0·030	—
8	0·099	0·077	0·062	—	—	—
9	0·105	0·082	0·066	—	0·034	—
10	0·111	0·087	0·070	—	0·036	—
11	0·118	0·092	0·075	0·040	0·038	—
12	0·125	0·098	0·079	—	0·041	—
13	0·133	0·104	0·084	—	0·043	—
14	0·140	0·110	0·089	—	0·046	—
16	0·157	0·123	0·100	—	—	0·038
18	0·175	0·138	0·111	—	—	—
20	0·195	0·153	0·124	—	—	—
24	0·240	0·190	0·153	—	—	—
28	0·292	0·230	0·187	—	—	—
32	0·352	0·278	0·227	—	—	—
36	0·419	0·332	0·272	—	—	—
40	0·496	0·393	0·323	—	—	—

The similarity of R_4 = 0·253 and $R_{\rm BS}$ = 0·229 also suggests that the differences between the C&S and B&S structures are about as large as the differences between a structure such as the B&S structure with a residual of about 23%, and the hypothetical 'true' structure. The Δr values therefore give a reasonable estimate of the kinds of errors which can be expected in a structure with refinement statistics similar to those of Bode & Schwager (1975).

Factors contributing to the differences in structure

(1) *Random and small systematic errors in the refined coordinates*

Errors in the refined coordinates are a major source of the differences between the two structures. Bode & Schwager (1975) estimated the standard deviations of their atomic coordinates to be $\sigma_x = \sigma_y = \sigma_z$ = 0·06 Å, yielding a radial standard deviation σ_r = 0·10 Å. Their estimate was based on calculations made for a representative carbonyl oxygen, according to the formula of Cruickshank (1949)

$$\sigma_x = \frac{-2\pi n}{aVC_o}\left[\sum_{hkl} h^2(F_o - F_c)^2\right]^{1/2}. \qquad (2)$$

In this expression $n = 2$ for noncentrosymmetric structures, a is the a axial length in Å, V the unit-cell volume in Å^3, and C_o the central atomic curvature.

The curvature C_o is the second derivative of electron density at the atomic center, and can be estimated either from electron density maps directly or from the second derivative of the transform of the scattering factor $f_o \exp[-B \sin^2 \theta/\lambda^2]$ for an atom, at its center (Leung, Marsh & Schomaker, 1957; Stout & Jensen, 1968):

$$C_o = \frac{-4\pi^2}{a^2 V} \sum_{hkl} h^2 f_o \exp[-B \sin^2 \theta/\lambda^2]. \qquad (3)$$

Error estimates have been computed from equation (2) for each atom type (*i.e.* having a given number of electrons and individual temperature factor) in the C&S structure (Table 3). The estimates are 0·10 Å or less only for well determined atoms, and are as high as 0·5 Å in poorly determined regions. Even so, the standard deviations computed from this formula are probably underestimates. Lipson & Cochran (1957*b*) have noted that the formula of Cruickshank (1949) is equivalent to

$$\sigma_x = \frac{\langle G_x \rangle_{\text{r.m.s.}}}{|C_o|},$$

where $\langle G_x \rangle_{\text{r.m.s.}}$ is the root-mean-square difference density gradient in the x direction. Thus, σ_x is equal to the estimated r.m.s. shift in x during a *single* cycle of difference Fourier refinement. This is indeed the case for the values in Table 3. They are approximately equal to the (radial) r.m.s. positional shifts in recent difference Fourier refinement cycles for the C&S structure, for atoms of the corresponding type. However, the 'true' atomic positions cannot in practice be reached in a single cycle, and for a structure which is still many cycles from convergence this formula tends to underestimate the deviation of the coordinates from their true position [this problem has been noted by Fermi (1975)].

Comparison of models separated by more than one difference Fourier cycle may thus give a more realistic estimate of the deviation of the coordinates from their true position at this stage of refinement. For the purpose of such an estimate, three recent sets of coordinates for the C&S structure were compared. The present constrained structure, having idealized molecular geometry (referred to as structure *C*), was obtained by idealization of the coordinates resulting from four difference Fourier cycles starting with the preceding idealized model (referred to as structure *A*). The r.m.s. positional difference between all atoms in the two constrained models *A* and *C* was 0·14 Å, and the maximum difference was 0·85 Å. The r.m.s. deviation between the coordinates in structure *A* and the unconstrained model four difference Fourier cycles later (referred to as structure *B*) was 0·21 Å and the maximum was 1·18 Å. The only change between this unconstrained model, *B*, and the current structure, *C*, was the idealization of bond lengths and angles. Comparison of the *B* and *C* structures gave an r.m.s. deviation of 0·20 Å for all atoms, with a maximum difference of 1·20 Å. In all cases, the maximum differences occurred in less well defined external side chains, as expected.

It is therefore apparent that the changes still taking place in the structure are larger than the errors computed from the formula of Cruickshank (1949) at this stage of refinement. On the basis of these changes we estimate random and small systematic errors in the C&S structure to be about 0·15–0·2 Å on average, and up to 1 Å or more for the least well determined regions [*i.e.* about twice the errors estimated from Cruickshank's (1949) formula], keeping in mind that the value varies considerably with the particular area of the structure in question. Errors in the B&S coordinates are expected to be comparable to or greater than these values, and the estimate of 0·1 Å given by those authors could be valid only for the best-determined atoms in their structure; it is misleading in the case of side-chain atoms or in external segments.

Another estimate of the overall mean error in the C&S structure was made according to the method of Luzzati (1952), from the dependence of R_{CS} on sin θ. Interpretation of the results (Table 4) is complicated by the fact that the increase of R_{CS} with resolution is much less than predicted by Luzzati's (1952) theory for a

Table 4. *Estimates of the overall mean error in the C&S atomic positions by the method of Luzzati* (1952)

Number of reflections[(a)]	$\langle \sin^2 \theta/\lambda^2 \rangle$	⟨Resolution⟩	$\lvert s \rvert$[(b)]	$\langle R_{\text{CS}} \rangle$[(c)]	$\langle \delta r \rangle \lvert s \rvert$[(d)]	$\langle \delta r \rangle$[(e)]
1028	0·0094 Å^{-2}	5·15 Å	0·194 Å^{-1}	0·179	0·0747	0·385 Å
1687	0·0181	3·72	0·269	0·145	0·0598	0·222
1945	0·0269	3·05	0·328	0·165	0·0688	0·209
2064	0·0359	2·64	0·379	0·172	0·0718	0·189
2135	0·0448	2·36	0·424	0·155	0·0643	0·152
2083	0·0537	2·16	0·463	0·146	0·0603	0·130
2382	0·0627	2·00	0·500	0·151	0·0625	0·125
1972	0·0716	1·87	0·535	0·148	0·0615	0·115
1965	0·0812	1·75	0·571	0·154	0·0639	0·112
1999	0·0894	1·67	0·599	0·158	0·0656	0·110
1209	0·0984	1·59	0·629	0·164	0·0685	0·109
1126	0·1074	1·52	0·658	0·168	0·0699	0·106

(*a*) Number of reflections $>3\sigma$, excluding lower-angle reflections inside 7 Å resolution.
(*b*) Magnitude of the scattering vector.
(*c*) $\sum |F_{o(\text{CS})} - F_{c(\text{CS})}| / \sum F_{o(\text{CS})}$, summed over those reflections in each zone.
(*d*) Values linearly interpolated from Table 2 of Luzzati (1952).
(*e*) Estimate of the mean error in atomic position.

random error distribution and identical, spherically symmetric atoms. The error estimate is thus highly dependent on the zone of sin θ considered, ranging from 0·38 Å (0·48 Å r.m.s.) for reflections near 5 Å resolution, to 0·17 Å (0·21 Å r.m.s.) for the 3·5 to 2·0 Å reflections, to 0·11 Å (0·14 Å r.m.s.) for the reflections from 2·0–1·5 Å resolution. This effect may arise from the fact that atoms with the smallest temperature factors make the dominant contribution to both $F_{o(\mathrm{CS})}$ and $F_{c(\mathrm{CS})}$ at high angles (*i.e.* in the range 2·0–1·5 Å), and thus the estimate derived from reflections in this range applies mainly to these best-determined atoms. On the other hand, lower-resolution reflections contain an appreciable contribution from all atoms, and the estimate of 0·21 Å derived from the 3·5–2·0 Å data may therefore be more appropriate as an overall estimate of the r.m.s. error. R_{CS} for the reflections in the 5 Å range is still affected somewhat by lack of a representation of the solvent continuum in the model. The overall average value of δr from Table 4 is 0·16 Å (0·20 Å r.m.s.). The smaller increase of R_{CS} with resolution also arises partly from the fact that Fourier refinement methods tend to give higher weight to high-resolution reflections than do least-squares methods (Cochran, 1948). This problem can be overcome by appropriate weighting of the coefficients in the Fourier syntheses, although to date unit weights have been used in the C&S refinement.

The estimates according to Luzzati's (1952) method are thus not inconsistent with those suggested above for the C&S structure. Fermi (1975) and Takano (1977*a*,*b*) have applied Luzzati's method to structures of human deoxyhemoglobin, sperm whale metmyoglobin, and sperm whale deoxymyoglobin, respectively, with satisfactory results. In the study by Fermi (1975) the resulting error estimates compared very well with the differences observed between independent protein molecules in the asymmetric unit (about 0·4 Å r.m.s. overall). However, errors estimated by the method of Cruickshank (1949) were too small (about 0·1 Å) at that stage of refinement, as also appears to be the case in the present study.

Nevertheless, the most useful error estimates are not the overall figures, but those made according to atom type, or region of the structure. In the case of the C&S structure, the best estimate that can be made of the standard deviations in atomic position, as indicated by the various methods above, is thus about twice the error computed according to the method of Cruickshank (1949) [*i.e.* the errors given in Table 3 for each atom type, scaled up to match the estimates computed from the changes still occurring in the structure and from the method of Luzzati (1952)].

(2) *Errors in the observed data*

The close agreement between the published R factor ($R_{\mathrm{BS}} = 0{\cdot}229$) of Bode & Schwager (1975) and that for the B&S structure against our observed data ($R_3 = 0{\cdot}238$ at 1·8 Å) implies that errors in either of the observed data sets probably do not make a major contribution to the differences between the coordinates at this stage. Errors in the observed DIP-trypsin F's have been estimated to average 4·6%, based on comparison of overlapping reflections (*i.e.* reflections which were measured from more than one crystal) (Chambers & Stroud, 1977*a*,*b*).

(3) *Differences between the two 'true' structures as they exist in the crystals*

Real differences could arise from the slightly different crystallization conditions, as well as from conformational changes due to the different inhibitors present. The inhibitor in the C&S structure (which now appears to be a monoisopropylphosphoryl group) is covalently attached to Ser 195 O_γ and strongly resembles a negatively charged tetrahedral intermediate in the reaction sequence. Benzamidine, on the other hand, binds non-covalently in the specific binding pocket of the enzyme, and mimics the side chain of a specific substrate. The existence of conformational differences between DIP- and benzamidine-trypsin has been demonstrated crystallographically (Krieger, Kay & Stroud, 1974) and the difference-map features described in that study have been confirmed using refined phases; these differences are small and highly localized. For example, the C191–C220 disulfide bridge in the lower portion of the specificity pocket is moved inward toward the benzamidine binding site in benzamidine-trypsin relative to DIP-trypsin. The magnitude of this movement obtained by comparison of the B&S and C&S coordinates is 0·31 ± 0·15 Å (where the uncertainty was derived from $\langle \Delta r \rangle_{\mathrm{r.m.s.}} = 0{\cdot}153$ Å for all sulfur atoms, excluding C191 and C220) which, although smaller, is not inconsistent with the value of 0·7 ± 0·3 Å derived by Krieger, Kay & Stroud (1974) from their $F_{o(\mathrm{benzamidine})} - F_{o(\mathrm{DIP})}$ difference map.

The imidazole ring of H57 in DIP-trypsin is moved relative to its position in benzamidine-trypsin, where H57 N_ε points more toward S195 O_γ to which it presumably forms a hydrogen bond (Bode & Schwager, 1975). Those regions where changes were observed, residues H57, D189–S195 and G217–C220, were omitted from the average positional difference calculations presented in Table 2. In view of the results of the earlier comparison (Krieger, Kay & Stroud, 1974) and the structure-factor comparison above, the contribution of the real differences in structure to the deviations between the coordinate sets, excluding the groups where changes were observed, is probably small. Nevertheless, a difference map computed between $F_{o(\mathrm{BS})}$ and $F_{o(\mathrm{CS})}$ would be the most useful indicator of the real differences between the two structures.

An interesting possibility is the existence of more than one preferred orientation for some of the more loosely determined parts of a protein structure; for example, external side chains. This kind of phenomenon, if it were shown to exist, would make the current assumptions of unique atomic positions inadequate for further refinement of those regions. Such a situation has not been observed at a high level of confidence in our difference maps to date, however.

(4) *Systematic differences arising from misinterpretation of Fourier maps*

One likely cause of the larger deviations (3–9 Å) in Fig. 1 is misorientation of parts of the structure which could not be unambiguously placed in the Fourier maps. Difference Fourier maps were used by both groups to correct this type of error. In spite of the relatively low residual for the C&S structure, recent difference maps still indicate significant changes, although these are now confined mostly to external side chains and solvent molecules. Bode & Schwager (1975) found their final difference map almost featureless. It may be that refinement of individual thermal parameters as in the C&S case can reduce the apparent noise level in the difference maps such that some of the weaker features arising from positional errors can be better recognized.

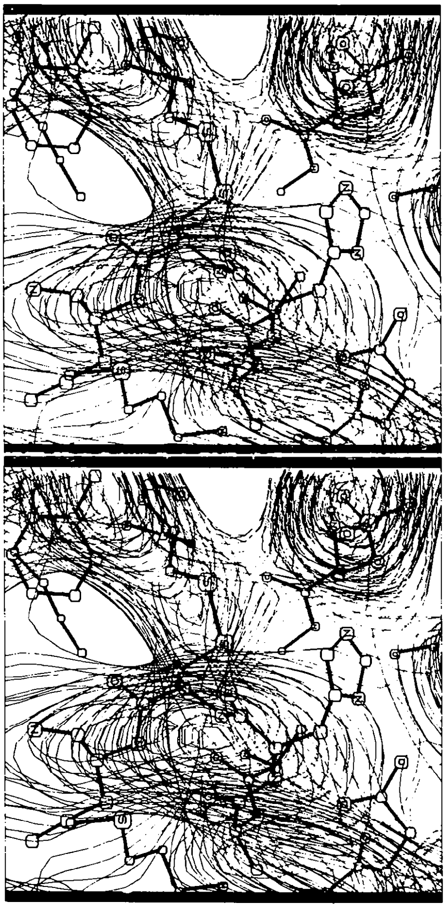
Although the two refined models are closely similar, it is surprising that systematic differences as large as 3–9 Å can be found. Huber *et al.* (1974) had previously noted differences of this same order when they compared their structure of trypsin complexed with pancreatic trypsin inhibitor with the first wire-model coordinates for DIP-trypsin. In this case they assigned these differences to errors in the initial wire-model coordinates. First sets of coordinates such as these are clearly unsuitable for the kind of detailed comparison made by those authors. The errors can be quite high, particularly when the resolution of a newly determined structure (limited by contributions from the heavy-atom derivatives) is often restricted to 2·5–3 Å, and the differences which were found in that comparison are probably typical of those expected from reasonably-high-quality density maps at that stage.

The errors introduced in the process of building a wire model of the usual type (Kendrew–Watson components on a scale of 20 mm ≡ 1 Å) with an optical comparator (Richards, 1968) were previously evaluated to be of the order of 10 mm, or 0·5 Å (Chambers & Stroud, 1977*a*,*b*). This value was based on the positional change between (*a*) coordinates measured (to within ±1 mm, or ±0·05 Å) from a wire model carefully built with reference to a 1·76 Å Fourier map computed with refined phases, and (*b*) the coordinates after optimal positioning by the difference Fourier method. Thus, 0·5 Å is probably a 'best-case' estimate. A standard deviation of 1·0 Å for errors of this kind is probably realistic for any wire model built to a 2·5 Å MIR map.

In the case of lower-resolution maps or maps computed with less accurate phases, larger errors can occur in construction of the model. The series of stereopictures in Fig. 5 are density maps of the DIP-trypsin catalytic site computed at different resolutions with observed amplitudes F_o, and phases calculated from the refined C&S structure. The refined model is superposed on the density map in each case, and the series shows some of the difficulties encountered in interpretation of even a very-high-quality low-resolution map. For example at 6 Å resolution (minimum interplanar spacing) (Fig. 5*a*) the disulfide bridge (Cys 42–Cys 58) is only in very weak density (the density in the picture is well behind the two sulfur atoms), while at 4·5 Å (Fig. 5*b*) the density for the bridge, although visible, is significantly removed from the refined atomic positions causing the bridge to appear miscentered in the density. A similar effect can be seen in the 3 Å map (Fig. 5*c*) for the β-carbon of His 57. In all the lower-resolution images overlap or proximity of the density from neighboring groups influences the appearance of the map. Thus, significant systematic errors can be incorporated into a structure though the atoms are positioned in the highest density. A tendency toward this type of error in wire-model coordinates was noted by Diamond (1974). At resolutions corresponding to interplanar spacings of 2 Å or less (Fig. 5*d*), where atoms begin to be resolved, the placement of groups is relatively free from such problems. These errors are readily corrected by structure refinement.

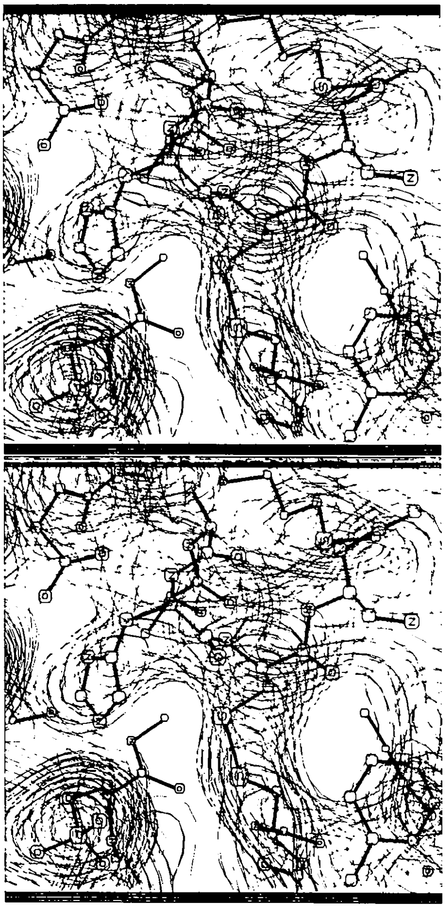
Systematic errors which are considerably more difficult to correct with the use of an automated refinement procedure can result from misinterpretation of the density. Examples are 180° misorientation of an amide plane or a valine side chain. Misinterpretations such as these place the group in a local minimum, making refinement to the correct orientation by any of the currently used types of refinement programs highly unlikely. Visual inspection of difference Fourier maps is by far the most useful method for detection of these errors. The importance of such a visual inspection at various stages of the refinement cannot be overstressed.

Which of the larger systematic differences between the C&S and B&S structures are due to misinterpretation of density maps in poorly defined areas, and which arise from real differences cannot be determined at this stage. It is clear from the character of coordinate differences (in large part random errors) that continued refinement should improve either structure, and might allow the causes of the larger discrepancies to be determined.



(a)

Fig. 5. Stereopictures of the DIP-trypsin catalytic site. The density maps were computed with coefficients $F_o e^{i\alpha_c}$, and the refined C&S model is superposed. (a) Coefficients to 6 Å resolution were used. The disulfide bridge Cys 42–Cys 58 no longer lies in the density.



(b)

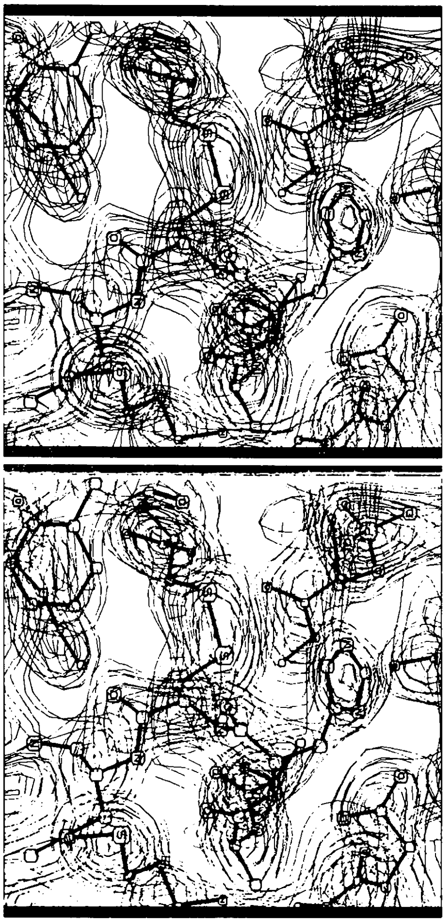
Fig. 5 (cont.) (b) At 4·5 Å resolution, overlap of density from neighboring groups causes the disulfide bridge to appear uncentered in the density.

Summary and conclusion

The comparison between these two structures has shown them to be very similar overall, as expected. We estimate random and small systematic errors in the C&S structure to be 0·15–0·20 Å over much of the structure, increasing to over 1 Å for the least well determined regions.

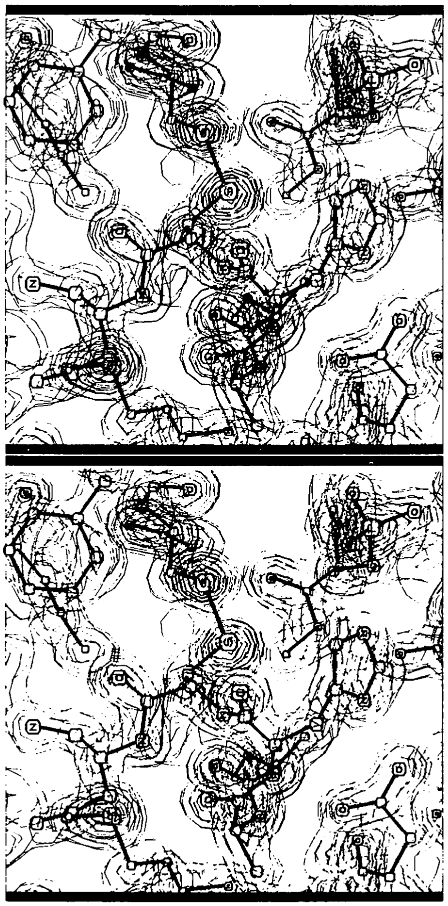
An important aspect of the differences between the structures, and hence the error estimates, is that they vary widely over different parts of the molecule. In the case of the trypsin structures it is fortunate that the catalytically important residues are in general very well defined (see Figs. 4 and 5). However, this will not necessarily be the case for all proteins, and some assessment of the reliability of different areas of the molecule is thus valuable to a biochemist working with crystallographically determined coordinates. Estimates for different atom types based on their refined atomic curvature (*i.e.* individual temperature factor and number of electrons) appear to be very useful in this regard, and are also good indicators of regions where interpretative errors of the kind reflected in the 3–9 Å range of Fig. 1 are likely to occur. This type of information can also be provided by a detailed description of the electron density map (such as that given by Birktoft & Blow, 1972), or to some extent by flagging the atoms in poorly determined regions, as done by Bode & Schwager (1975). Overall error estimates, or those made based on a 'typical' atom are less useful and can be misleading in poorly determined areas.

It is apparent from the remarkable agreement over much of the molecule that both the B&S and C&S models have benefited considerably from refinement. The errors in these models are far below those which would be expected in a set of unrefined wire-model coordinates. However, the structure factor comparisons suggest that errors in the models still represent the major component of the R factors at these stages of refinement, and that continued refinement should improve either structure. These results imply that residuals of around $R = 0{\cdot}20$ for protein structures cannot by any means be taken to imply an optimally refined, or converged solution of the molecular structure. Nevertheless, the advances currently being made



(c)

Fig. 5 (*cont.*) (*c*) At 3 Å resolution, a similar phenomenon to Fig. 5(*b*) is seen for His 57 C_β, which appears miscentered due to overlap of density from neighboring groups.



(d)

Fig. 5 (*cont.*) (*d*) At 1·5 Å resolution individual atoms have begun to be resolved, making placement of groups more straightforward.

in the area of protein structure refinement should allow refinement to be carried out routinely and efficiently to convergence with minimal cost.

We are grateful to R. E. Marsh for valuable discussions. This work was supported by National Institutes of Health Grant GM-24485, by National Science Foundation Grant BMS75-04105, and by funding from the Research Evaluation and Allocation Committee, University of California, San Francisco. RMS is a Sloan Foundation Fellow.

References

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

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

Chambers, J. L. & Stroud, R. M. (1977*a*). *Acta Cryst.* B**33**, 1824–1837.

Chambers, J. L. & Stroud, R. M. (1977*b*). Am. Crystallogr. Assoc. Meeting, Michigan State Univ., East Lansing, Michigan. Abstracts, Vol. 5(2), p. 59.

Cochran, W. (1948). *Acta Cryst.* **1**, 138–142.

Cochran, W. (1951). *Acta Cryst.* **4**, 408–411.

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

Deisenhofer, J. & Steigemann, W. (1975). *Acta Cryst.* B**31**, 238–250.

Diamond, R. (1971). *Acta Cryst.* A**27**, 436–452.

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

Fehlhammer, H. & Bode, W. (1975). *J. Mol. Biol.* **98**, 683–692.

Fermi, G. (1975). *J. Mol. Biol.* **97**, 237–256.

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

Kossiakoff, A. A., Chambers, J. L., Kay, L. & Stroud, R. M. (1977). *Biochemistry*, **16**, 654–664.

Krieger, M., Chambers, J. L., Christoph, G. G., Stroud, R. M. & Trus, B. L. (1974). *Acta Cryst.* A**30**, 740–748.

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

Krieger, M. & Stroud, R. M. (1976). *Acta Cryst.* A**32**, 653–656.

Leung, Y. C., Marsh, R. E. & Schomaker, V. (1957). *Acta Cryst.* **10**, 650–652.

Levitt, M. (1974). *J. Mol. Biol.* **82**, 393–420.

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

LIPSON, H. & COCHRAN, W. (1957*a*). *The Determination of Crystal Structures*, p. 146. London: Bell.

LIPSON, H. & COCHRAN, W. (1957*b*). *The Determination of Crystal Structures*, p. 308. London: Bell.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

MARSH, R. E. & DONOHUE, J. (1967). *Adv. Protein Chem.* **22**, 235–256.

MAXWELL, J. C. (1860). *Philos. Mag.* **19**, 31. Cited in MOORE, W. J. (1972). *Physical Chemistry*, pp. 133–141. Englewood Cliffs, New Jersey: Prentice-Hall.

MOEWS, P. C. & KRETSINGER, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.

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

STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination: A Practical Guide*, p. 404. New York: Macmillan.

STROUD, R. M., KAY, L. & DICKERSON, R. E. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 125–140.

STROUD, R. M., KAY, L. & DICKERSON, R. E. (1974). *J. Mol. Biol.* **83**, 185–208.

TAKANO, T. (1977*a*). *J. Mol. Biol.* **110**, 537–568.

TAKANO, T. (1977*b*). *J. Mol. Biol.* **110**, 569–584.