Acta Cryst. (1977). B33, 1824–1837

Difference Fourier Refinement of the Structure of DIP-Trypsin at 1·5 Å with a Minicomputer Technique*

BY JOHN L. CHAMBERS†§ AND ROBERT M. STROUD‡§

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

(*Received* 18 *June* 1976; *accepted* 22 *November* 1976)

The X-ray crystal structure of diisopropyl fluorophosphate-inhibited bovine trypsin has been refined to a resolution of 1·5 Å with the use of a constrained difference Fourier technique. The refinement was carried out almost entirely on a minicomputer with a set of programs written for this study to largely automate the refinement procedure. Use of the minicomputer has allowed the refinement to progress efficiently at an extremely low cost, from a starting R value of 47·2% at 2·7 Å resolution to 27·7% at 1·5 Å resolution. Two cycles of a least-squares refinement procedure with a first-order gradient minimization have since reduced R to 23·5% at 1·5 Å resolution. The refined model and electron density maps are substantially improved over the starting ones. The methods described may be very attractive when computing funds or access to a large computer are limited.

Introduction

During the past few years a number of protein structures determined by X-ray crystallography have been refined to 2 Å, or better, resolution with a variety of techniques (Watenpaugh, Sieker, Herriot & Jensen, 1973; Sayre, 1974; Diamond, 1974; Huber, Kukla, Bode, Schwager, Bartels, Deisenhofer & Steigemann, 1974; Freer, Alden, Carter & Kraut, 1975; Moews & Kretsinger, 1975; Deisenhofer & Steigemann, 1975; Bode & Schwager, 1975*a*). The refinement programs written by Diamond (1966, 1971) are particularly effective and are used extensively. However, the programs are large and require significant amounts of storage and central processing unit (CPU) time on a large computer. Several of these workers have successfully employed difference Fourier (ΔF) refinement techniques, both alone and in conjunction with Diamond's programs. The ΔF method is economical in terms of programming complexity and cost, and is readily adaptable for use on a small computer.

This paper describes the methods developed and used for ΔF refinement of the structure of diisopropyl fluorophosphate (DFP)-inhibited bovine trypsin originally determined by Stroud, Kay & Dickerson (1971, 1974). A set of refinement programs has been written and implemented almost entirely on a Data General NOVA 800 minicomputer. Use of the minicomputer has enabled the refinement to progress efficiently at a very low cost, and the methods described here may therefore be attractive when large amounts of time and a very low rate of cost on a large computer cannot be obtained.

The refinement has resulted in a model which is greatly improved over the starting one, with substantial (*i.e.* several ångströms) changes in some areas. This study thus indicates, as do those cited earlier, that the improvement to be gained by refinement warrants the routine use of some refinement procedure in the determination of protein crystal structures.

The refined model is described only as much as necessary to illustrate the application of the methods. A detailed description of this structure will be presented in a future paper. The methods described here have also been used successfully in a preliminary refinement of the bovine trypsinogen structure (Kossiakoff, Chambers, Kay & Stroud, 1976).

Methods

DIP-trypsin crystals

The conditions for crystallization of diisopropylphosphoryl (DIP)-trypsin are described by Stroud *et al.* (1974). The crystals are orthorhombic, space group $P2_12_12_1$, with unit-cell dimensions a = 54·84, b = 58·61, c = 67·47 Å, and one molecule per asymmetric unit. Crystals of a size approximately 0·6 × 0·5 × 0·5

* Contribution No. 5352. Supported by National Institutes of Health Grant GM-19984 and National Science Foundation Grant BMS75-01405.

† National Institutes of Health Predoctoral Trainee.

‡ National Institutes of Health Career Development Awardee, US Public Health Service Grant No. GM-70469, and Sloan Foundation Fellow.

§ Present address: Department of Biochemistry and Biophysics, School of Medicine, University of California, San Francisco, California, 94143, USA.

mm were used for recording the high-resolution (1·5 Å) data; those used for data to 2 Å were slightly smaller. Although these crystals contain about 40% α-trypsin (cleaved between Lys 145 and Ser 146) and 60% β-trypsin (uncleaved), they do not appear to be significantly less well ordered than the crystals of benzamidine-inhibited trypsin described by Fehlhammer & Bode (1975), which contain about 80% β-trypsin. The diffraction pattern from the DIP-trypsin crystals extends to at least 1·1 Å resolution.

Data collection and reduction

Diffracted intensities from DIP-trypsin crystals were measured with a Syntex $P\bar{1}$ diffractometer equipped with a graphite monochromator and a helium-filled tube between the crystal and detector. The tube voltage and current for the X-ray source were 40 kV and 20 mA respectively. Background corrections were made with an interpolation technique which accounts for variations in the background with the setting angles, φ, χ, and 2θ (Krieger, Chambers, Christoph, Stroud & Trus, 1974). The remaining data reduction and scaling together of data collected from different crystals were accomplished according to the procedure of Stroud *et al.* (1974). In all, eight crystals were required to provide a full set of intensities to 1·5 Å resolution ($2\theta = 62°$).

4072 intensities to 1·5 Å were measured from different crystals for scaling purposes. The average value of $R = \Sigma |F_1 - F_2|/\Sigma F_1$ (where F_1 is the current value of the structure factor in the master data set, and F_2 is the value for the crystal data to be merged into the master set) was 4·6% for these multiply-measured reflections. Of the 35 566 independent reflections in the 1·5 Å sphere, those 22 117 reflections having structure factors greater than three times their standard deviation were included in the final data set. To 2·1 Å, 88·5% of the 13 261 possible data were observed. Between 2·1 and 1·76 Å the number observed was 4784 of a possible 8979, or 53·3%, and between 1·76 and 1·5 Å, 5602 of 13 326 possible reflections were observed, or 42·0%.

Starting model

The starting coordinates for the refinement were measured from a wire model constructed in an optical comparator (Richards, 1968) with reference to a 2·7 Å MIR-phased map. This map was more recent than that described by Stroud *et al.* (1974), differing in that the 2·7 Å data from the Ag^+ derivative (Chambers, Christoph, Krieger, Kay & Stroud, 1974) had been incorporated into the phase refinement. Atomic coordinates from the model were measured with an electronically operated device, similar in principle to the one described by Salemme & Fehr (1972).

Table 1. *Computing system configuration*

NOVA 800 (Data General Corporation)

32 K (16 bit) words core
Floating-point hardware*
12-½ i.p.s. magnetic-tape drive
25 i.p.s. magnetic-tape drive
2·5 M byte moving-head disk
Keyboard printer
Matrix-line printer/plotter

Typical execution times for Fortran coded instructions

Operation	Time (μ s)
Integer add	2
Integer multiply	80
Floating-point add	20
Floating-point multiply	25
Floating-point divide	50
SIN(X)	400

*Syntex Analytical Instruments.

Minicomputer system

The components of the minicomputer system used in this study are listed in Table 1, with typical execution times for certain types of operations and function evaluations. All computations, except calculation and plotting of electron density maps, were carried out with this system. An IBM 370/158 computer was employed for the electron density syntheses.

Difference Fourier method

Detailed descriptions of the difference Fourier method have been given elsewhere (Cochran, 1951; Stout & Jensen, 1968). The method is based on the principle that an electron density map computed with coefficients $(F_o - F_c)\exp(i\varphi_c)$ (where F_o is the observed structure factor amplitude, and F_c and φ_c are the structure factor amplitude and phase computed from the current model) contains positive peaks in positions where additional electron density should be added to the model, and negative peaks in positions where it should be subtracted, in order to produce better agreement between the model and the observed data. The position of an atom can thus generally be improved by computing the density gradient in the ΔF map at the atomic center and moving the atom toward higher density along this gradient. For an atom which is correctly positioned, the density at the atomic center gives an indication of the shift in the temperature factor for that atom. The density at atomic centers is also strongly influenced by the overall scale and temperature factor between the observed and calculated structure factors.

The course of a typical cycle of ΔF refinement performed in this study is schematized in Fig. 1. Each cycle was begun with calculation of structure factors

based on the current model of the structure. The calculated structure factors were then scaled to the observed ones, which in turn had been placed on an absolute scale early in the refinement, and were then used to generate a new electron density map. This map was either a difference map (computed using coefficients $\Delta F = F_o - F_c$ and phases φ_c), which shows the differences between the current model and an improved one, or a map computed with terms $[nF_o - (n-1)F_c]\exp(i\varphi_c)$ which superimposes n difference maps on the current structure, showing the improved structure in its entirety. Because peaks resulting from a difference synthesis are about half their theoretical height (Henderson & Moffat, 1971), $n = 2$ was used for most of these maps. Empirically, $n = 2$ usually produced the best compromise between freedom from noise and appearance of new information.

Following generation of the density map, the indicated corrections to atomic positions were derived and applied to the structure. Three techniques were used to determine atomic shifts (Fig. 1). The simplest and most accurate procedure was to estimate shifts from the difference maps automatically, with a program written for the NOVA 800 (the middle path in the flow chart of Fig. 1). The automated technique was most useful where relatively small movements from the input structure were required. If the map indicated that gross structural changes were needed, one of the other two branches was taken. The branch to the left in Fig. 1, which was the simpler of the two, involved manual adjustment of atomic positions as indicated by a map made from 12 inch square Plexiglas® sections upon which were drawn both density contours and the atomic positions. It was very difficult to preserve accurate bond lengths and angles when making shifts in these small-scale maps, and incorporation of shifts derived in this manner was relatively tedious. With this technique, atomic shifts could be made at a rate of about 1800 atoms in two days.

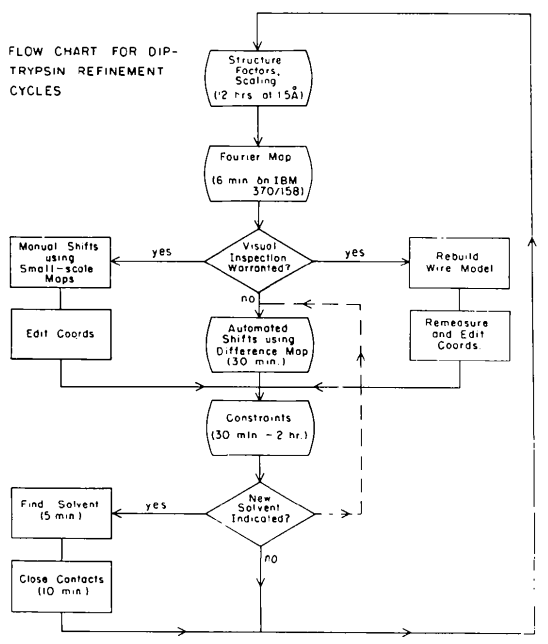

Fig. 1. Flow chart outlining the major steps involved in a single refinement cycle, using the technique described in the text. Alternate branches are indicated in the center of the figure. The main sequence was through automated shifts in difference maps. The branch toward small-scale maps was used for major reorientation of parts of the structure. The wire model was rebuilt once during the refinement as an additional check on its progress, as described in the text.

Once during the refinement the wire model was rebuilt and the coordinates remeasured (the branch to the right in Fig. 1). Although this procedure was useful for interpreting interactions involving long regions of secondary structure and hydrogen-bonding possibilities, making changes in this manner was very tedious and the resulting coordinates were not typically of very good quality.

An automated constraints procedure was used to restore bond lengths and angles to physically reasonable values after the shifts in atomic positions had been applied. Occasionally, additional rounds of atomic shifts and constraints carried out before computing a new map resulted in further improvement, especially when $nF_o - (n-1)F_c$ syntheses were used. The best results with difference syntheses were obtained when a new map was computed after each round.

If indicated in the calculated maps, new solvent molecules were added to the structure, or old ones deleted based on their occupancy and close approach distances to other atoms in the structure. A new refinement cycle was then initiated.

Refinement software

The refinement software and algorithms employed in the individual steps mentioned above are described below. Except where noted, all programs were written in Fortran IV, and execution times refer to execution on the NOVA 800.

(*a*) *Structure factor calculation.* Structure factors were calculated by conventional Fourier transformation of atomic coordinates, with expressions specific for space group $P2_12_12_1$. The assembly language code output by the Fortran compiler for the inner loops in the program was optimized to increase execution speed. The contribution from each atom was added into every reflection before proceeding with the next atom, enabling the use of recursive multiple-angle relations for rapid evaluation of the trigonometric functions. Scattering factors for the individual atoms were approximated by the five-parameter analytical expressions of Forsyth & Wells (1959). Atoms within the

coordinate list were sorted into groups which contained identical scattering factors and individual isotropic temperature factors, requiring only one evaluation of these parameters per group for each reflection. A provision was made in the program for deleting small groups of atoms from, or adding them to, the current set of computed structure factors. Such an option is essential if the entire set of F's is not to be recalculated whenever changes are made over a relatively small portion of the structure.

Execution time for this program was $6{\cdot}75 \times 10^{-4}$ s per atom-reflection. For the DIP-trypsin coordinate file, containing about 1800 atoms per asymmetric unit, total execution time was approximately 4 h at 2·1 Å (11 713 reflections), 7·5 h at 1·76 Å (22 309 reflections), and 12 h at 1·5 Å (35 566 reflections).

(*b*) *Scaling and R calculations.* A number of authors have noted that, at low diffraction angles, the average intensities observed from protein crystals are significantly reduced from their expected values (Watenpaugh *et al.*, 1973; Moews & Kretsinger, 1975). This effect has been attributed to the relatively large amount of disordered solvent present in protein crystals. Such an effect is seen in the scaling plot for DIP-trypsin shown in Fig. 2. This behavior was accounted for by least-squares refinement of a four-parameter function to the scaling plot. The function was chosen to represent a linear fit at high angles with an exponential fall-off at low angles:

$$y = A_1 \exp(-A_2 x) + A_3 x + A_4.$$

This procedure is similar to one described by Moews & Kretsinger (1975). The least-squares fits for both the linear and non-linear functions are shown in Fig. 2. The linear fit for reflections at small values of $\sin^2\theta/\lambda^2$ was totally inadequate, and inclusion of these reflections scaled in this manner caused large low-frequency ripples in the resulting electron density maps.

The improvement in agreement between F_o and F_c for these reflections is shown in Fig. 3. For the linear fit, the residual, R ($R = \Sigma|F_o - F_c|/\Sigma F_o$), increased sharply at low angles, reaching over 75%; however, when properly scaled, with the non-linear expression, the agreement for these reflections was up to three times better. It should be emphasized that, while the technique accounts well for the difference in scale of the low-angle data, the expression used is an empirical one and

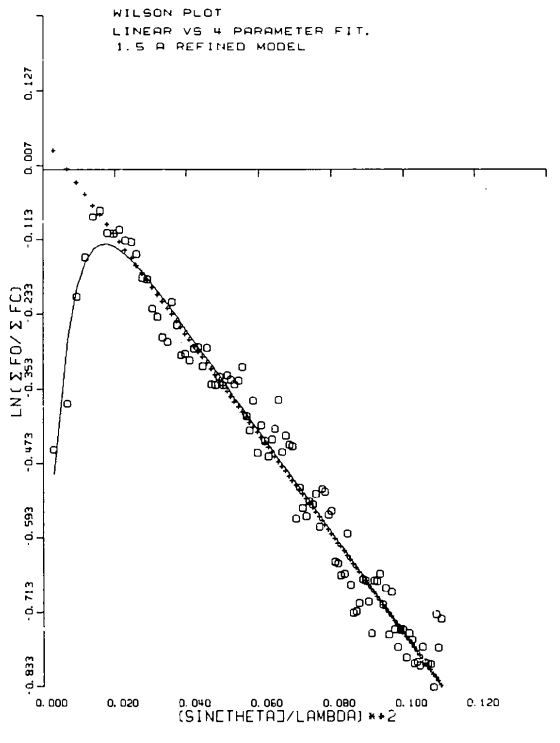

Fig. 2. Scaling plot for the 1·5 Å data. The observed F's for DIP-trypsin fall off rapidly with respect to the calculated F's at low angles. $\Sigma F_o/\Sigma F_c$ was computed for 100 zones in $\sin^2\theta/\lambda^2$. The solid line drawn through these computed points (O) represents the four-parameter fit; points calculated from the linear fit are symbolized by (+). Unit weights were used in this case. For the linear fit, points inside 5 Å resolution were neglected.

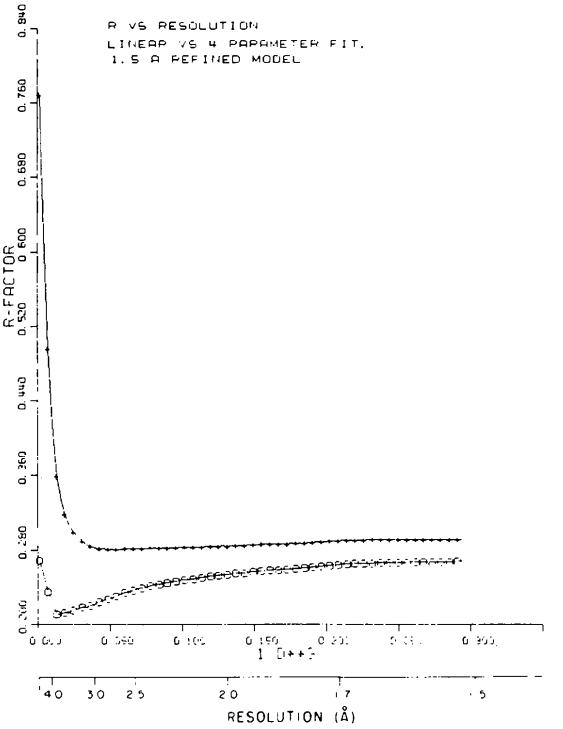

Fig. 3. Graph of overall R value *vs* resolution showing the improvement in agreement between F_o and the scaled F_c for low-angle reflections when using the four-parameter scale. The R value was computed for each point with all reflections to the indicated resolution. (+) linear fit; (O) four-parameter fit.

is not meant to be a mathematical description of a physical phenomenon. The actual function used, as long as it conforms well to the shape of the scaling plot, does not appear to be critical to either the appearance of the density maps, or the value of R.

Reflections for which $(F_o - F_c)/[0{\cdot}5(F_o + F_c)] > 120\%$ were left out of the calculation of the residual (about 700 reflections to 1·5 Å). These reflections were also omitted from the calculation of density maps, since the ΔF term which results for the case where $F_o \gg F_c$ is likely to be poorly phased. Difference terms for those reflections where $F_c \gg F_o$ are probably correctly phased and were included in the R and Fourier map calculations (Stout & Jensen, 1968). Unobserved reflections were included in the scaling with F's equal to twice the average standard deviation for unobserved F's in the appropriate zone of $\sin\theta$. They were not included in the Fourier or R calculations.

Absolute scaling was accomplished with the use of a standard Wilson plot technique, neglecting the reflections inside 5 Å resolution. The overall temperature factor thus obtained was 14 Å^2 for the 1·5 Å data. All atoms included in the structure factor calculation (*i.e.* protein and ordered solvent) were included in the absolute scaling.

(*c*) *Fourier map calculation.* Generation of electron density maps was the only step of the refinement carried out on the IBM 370/158, with a Fourier program originally written by G. N. Reeke. CPU time required to generate a map of approximately 100 000 grid points from 22 117 reflections to 1·5 Å resolution was 6·5 min. A grid spacing of approximately 1·25 Å along each of the crystallographic axes was used throughout most of the refinement. The coarse grid produced satisfactory results with 2·7 and 2·1 Å data, and was economical in terms of computing cost and map storage requirements. The fact that the grid size exceeded the interval corresponding to the Nyquist frequency (*i.e.* was greater than half the resolution of the data) at 2·1 Å did not appear to affect the course of the refinement adversely. At 1·76 Å and higher resolution, better results were obtained with a finer grid, and an interval of 0·75 Å was therefore used in later stages.

(*d*) *Atomic shifts.* Because of the iterative nature of the refinement procedure, small errors introduced by making approximations in order to simplify programming or computation can be corrected in subsequent cycles. A number of such approximations were made in the atomic shifts program.

Electron density at an atomic position, ρ, was interpolated from a three-dimensional array of either 8 or 27 grid points (depending on the coarseness of the map grid) surrounding the atom, according to the scheme

$$\rho = \frac{\sum_i \rho_i w_i}{\sum_i w_i}$$

where ρ_i are densities at each of the grid points, and w_i are weights for each grid point in the sum, which fall off exponentially [$w_i = \exp(-ad_i^2)$] as the distance d_i between the grid point and the atom increases. Values for the constant a were in the range 0·25–1·0 Å^{-2}. Gradients were computed by evaluation of the density 0·1 Å on either side of the atom along each coordinate axis. In more recent cycles, the electron density was approximated with a three-dimensional polynomial, with coefficients (a_1–a_7) determined from seven grid points around the atom center.

$$\rho = a_1\Delta x^2 + a_2\Delta y^2 + a_3\Delta z^2 + a_4\Delta x + a_5\Delta y + a_6\Delta z + a_7.$$

This expression allows analytical evaluation of the gradients, resulting in faster execution. Both schemes produced satisfactory shifts.

The correction to an atomic position indicated in a ΔF map for non-centrosymmetric structures is $\Delta\mathbf{x} = -2\mathbf{g}/C$, where $\Delta\mathbf{x}$ is the change in position, $\mathbf{g}$ is the difference density gradient at the atomic site, and C is the curvature at that site in the corresponding F_o synthesis (Cochran, 1951; Stout & Jensen, 1968). In practice, the curvature has been approximated with empirically determined values for each type of atom or ion (Watenpaugh *et al.*, 1973; Freer, Alden, Carter & Kraut, 1975; Moews & Kretsinger, 1975).

A number of factors in addition to the atom type influence the actual values to be used, including resolution of the data, the overall scale and temperature factors, and the scale of the densities written at the grid points of the map. To account for variations in these parameters over the course of the refinement, and to give the program greater flexibility, the curvature was taken to be proportional to the atomic number, Z, and shifts were scaled by a specified constant, K, chosen to produce reasonable values for the shifts: $\Delta\mathbf{x} = K\mathbf{g}/Z$. A value of K was used which produced positional changes approximately equal in magnitude to those estimated from $2F_o - F_c$ maps (averaging about 0·4 Å in early cycles and 0·1 Å in later stages). After a shift was computed, the density at the prospective new atomic site in the map used to derive the shift was evaluated to make certain that the atom had moved into more positive density. If not, the program repeatedly cut the size of the shift by half until this criterion was met. In addition, any shift exceeding a specified maximum, Δx_{max} (usually 0·4 Å), was reset to this value, maintaining the original direction. The above expression enabled shifts to be readily evaluated from MIR maps and $nF_o - (n-1)F_c$ maps, as well as from difference Fourier syntheses.

The electron density (ρ) for an atom can be approximated by a spherical Gaussian of the form,

$$\rho \simeq (B/4\pi)^{-3/2} Z \exp[-\pi r^2/(B/4\pi)],$$

where r is the distance from the atomic center, Z is the number of electrons, and B is the individual isotropic temperature factor for the atom (Diamond, 1971). From this expression, the expected change in temperature factor (ΔB) corresponding to a small change in electron density at the center of the atom ($\Delta\rho_0$) is,

$$\Delta B = \frac{-\Delta\rho_0}{12\pi^{3/2}Z} B^{5/2}.$$

This expression is in practice effective only for atoms which are correctly positioned, and is also affected by scaling errors. The equation was therefore modified empirically to provide more satisfactory results, and the expression used to compute corrections to the individual isotropic temperature factors was,

$$\Delta B = \frac{-K_2(\rho_\Delta - \overline{\rho_\Delta})}{Z} \sqrt{B} \exp(-\Delta x/\Delta x_{\max})$$

where ΔB is the shift, ρ_Δ is the difference electron density at the atomic site, $\overline{\rho_\Delta}$ is the expected average difference electron density for all atomic sites, and K_2 is a scaling constant similar to the one for positional shifts.

Since atoms at the beginning of the refinement tend to reside in negative difference density because they are not correctly positioned, there is a tendency for the temperature factors to become too large in early cycles, providing they are allowed to vary. The terms involving $\overline{\rho_\Delta}$ and $\exp(-\Delta x/\Delta x_{\max})$ enable this interaction between positions and temperature factors to be partially decoupled. The former term also compensates for errors in scaling of calculated F's to the observed ones, which can also greatly influence ΔB. The latter term reduces the shift in the temperature factor according to the size of the gradient at the atomic site; thus, the full shift is applied only when an atom is close to its correct position. The equation used produced satisfactory values for ΔB, which resulted in rapid improvement in flatness of the ΔF maps at atomic centers, and reasonable values for the temperature factors themselves (values of 10, 12, 14, 16, 18, 23, 28, 38, and 58 Å^2 were used). Larger B factors were generally associated with external side chains.

Execution time for this program for a coordinate list of about 1800 atoms was approximately 30 min.

(*e*) *Constraints*. Constraints were applied to the coordinates by minimization of a residual composed of sums of squares of the deviations of the atoms from 'ideal' relative positions. Four types of differences contributed to the residual. The first was the difference between the positions of the atoms in the starting structure and those in the constrained structure. This term was given unit weight throughout the constraints procedure. The other three terms contained deviations from standard values of the bond lengths, bond angles, and dihedral angles to be constrained. Each of these three terms was given a relative weight which was increased on each pass through the atom list, allowing greater movements from the starting structure and a closer approach to standard constraints in each subsequent pass. The procedure was terminated either after a fixed number of passes, or when r.m.s. deviations of bond lengths and angles had fallen below selected values. Similar procedures have been described by Hermans & McQueen (1974) and Freer *et al.* (1975). More recently, similar techniques have been independently developed for constraining bond lengths, angles, and dihedral angles by minimization of deviations from standard interatomic distances only (Dodson, Isaacs & Rollett, 1976; Ten Eyck, Weaver & Matthews, 1976).

Standard values for bond lengths and angles were obtained from Marsh & Donohue (1967), and the references cited therein. Standard values for the dihedral angles constrained were chosen in order to make the peptide amide groups *trans* and planar. In later stages, planarity of rings and side-chain amide, carboxyl, and guanidino groups was enforced by constraining dihedral angles, since bond length and bond angle constraints alone were not in practice sufficient to make these groups planar. In the case of side-chain amide and carboxyl groups, dummy dihedral angles were constrained [*e.g.* $\theta(C_\beta, C_\gamma, O_{\delta_1}, O_{\delta_2})$ in Asp].

Constraints were applied to a zone of one to five residues. A small portion of the peptide chain on either side of the zone was used as a fixed boundary, which was necessary for convergence of the procedure at the interface between zones. A simple steepest-descent gradient minimization technique was used, and was very reliable in its ability to produce converging shifts even when atomic coordinates were far removed from ideality; it was also fast and satisfactory in its rate of convergence. (This was in contrast to the gradient-curvature techniques, *e.g.* Newton–Raphson minimization tried in early versions of the program, which converged rapidly in the neighborhood of the minimum, but often produced unreasonable shifts for regions which were poorly constrained in the starting structure, and required substantially more computing time per cycle.)

Since the molecular parameters in real proteins, particularly the planarity of peptide amides (Ramachandran, Lakshminarayanan & Kolaskar, 1973; Huber *et al.*, 1974) may deviate significantly from 'ideal' values, the fact that the constraints can be given any desired degree of flexibility with this procedure is highly advantageous. It is moreover helpful to keep the constraints rather flexible when performing alternate cycles of atomic shifts and constraints in order to prevent oscillation between a constrained and an unconstrained structure.

Four passes through the coordinate file were generally run after atomic shifts had been applied in order to achieve the desired rigidity in the constraints.

Weights for the contributions to the residual of bond lengths, angles, and dihedral angles were initially set at 5, 1 and 1 for lengths in ångströms and angles in radians, and were multiplied by 3, 6 and 6, respectively, at the beginning of each new pass. Resulting overall r.m.s. deviations from ideality were typically 0·07 Å in bond lengths, 3·5° in bond angles, and 4·5° in dihedral angles. Maximum deviations were usually about four times the overall r.m.s. deviations. Execution time was about 30 min for each pass through the coordinate file containing 223 residues.

(*f*) *Peak search and close contacts.* The peak search and close-contacts programs were used primarily for selection of solvent molecules from difference maps. The peak search program locates prospective solvent straightforwardly from one asymmetric unit of the map, and runs in about 5 min for 100 000 grid points.

Close contacts between nonbonded atoms were tabulated by first sorting the atoms based on their position in the asymmetric unit, requiring comparison between a relatively small number of atoms and their symmetry equivalents to find all distances less than a specified value. Approximately 20 min was required to tabulate all close contacts between atoms in an 1800-atom-coordinate file. Prospective solvent molecules which were closer than 2·5 Å to atoms already present in the structure, or solvent which had moved within this distance of other atoms in the protein, were generally removed from the data set. In a few cases, particularly when close contacts existed between groups of the protein itself, corrections were made manually. The current structure contains 174 solvent molecules, most of them hydrogen-bonded to groups on the surface of the protein and to other solvent. About 30 of them can be considered internal.

Results

Progress of the refinement

The graph in Fig. 4 indicates the progress of the refinement to date. R at 2·7 Å computed for the starting structure (revision 1) was 47%, rising to 50% upon incorporation of the 2·1 Å data. Manual adjustment of atomic positions with the use of small-scale $2F_o - F_c$ maps reduced the residual to 40·2%.

The automated atomic shifts program was written at this stage. The improvement in R was very rapid, and after three cycles of shifts determined from ΔF maps, R was 34%. The resulting r.m.s. deviations from standard bond lengths and angles were, however, very poor – 0·5 Å and 22° respectively. Moreover, 33 α-C atoms and 10 Ilu and Thr β-carbons had adopted the wrong sense. The constraints program was subsequently implemented, and several cycles of automated shifts with a $2F_o - F_c$ map followed by constraints were performed. A number of the wrong-handed asymmetric C atoms were corrected manually with small-scale maps. The resulting overall r.m.s. deviations from ideality for bond lengths and angles were 0·15 Å and 6·5°; at the same time, there was only a small decrease in the average value of electron density at atomic centers. R for this constrained structure was 38·8% at 2·1 Å.

Alternate cycles of shifts and constraints were then applied nine times, using seven difference Fourier maps (revisions 6–12 in Fig. 4). During these cycles, solvent atoms were added and deleted where indicated. At revision 12 of the coordinates (see Fig. 4), the atom file contained 220 solvent molecules. Deviations from ideal bond lengths, angles, and dihedral angles were 0·1 Å, 2·4 and 3·8° (r.m.s.) respectively. The sense of all asymmetric atoms had been corrected and the residual

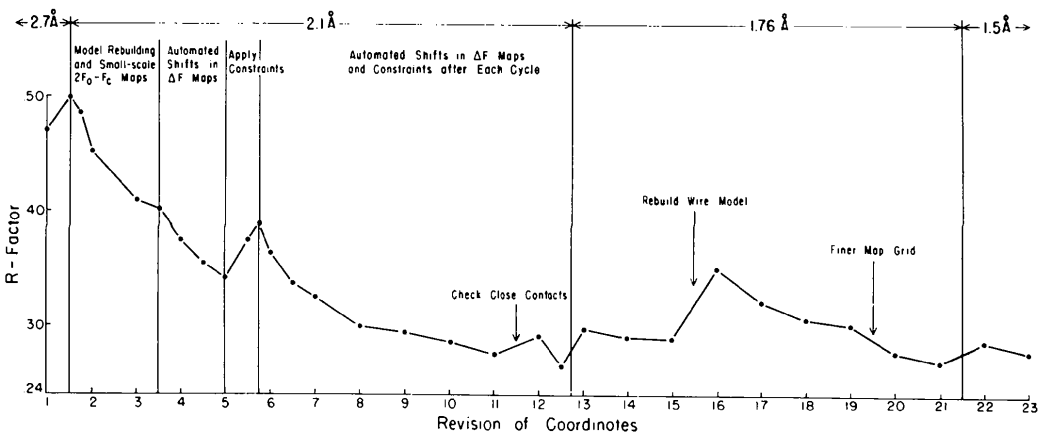

Fig. 4. Summary of DIP-trypsin refinement using the scheme outlined in Fig. 1, indicating the improvement in agreement between F_o and F_c. The decrease in R over the course of the refinement was from 47·2% at 2·7 Å to 27·7% at 1·5 Å. The sharp increase in R at revision 16, after rebuilding the wire model to a 1·76 Å refined map, resulted from errors in the wire-model coordinates, as described in the text.

had been reduced to 26·6% at 2·1 Å. The 1·76 Å data were included, and two refinement cycles produced a 1·76 Å R of 28·8% (231 solvent).

Since it appeared that additional automated cycles would produce little further improvement in R, it was felt that a large-scale visual inspection of the structure in its current state was warranted, and the wire model was rebuilt to a $2F_o - F_c$ map. A number of substantial changes from the starting wire model had taken place, and will be briefly described later. Some additional improvements were made at this step, including major reorientation of six amide planes and addition of seven ordered solvent molecules. Approximately 70% of the coordinates were replaced with coordinates measured from the rebuilt model. The 1·76 Å R index rose sharply to 35·0%.

Examination of the new coordinates in a small-scale map showed that, while the atoms were not far from their intended positions, large pieces of the structure appeared randomly translated by about 0·5 Å from the center of the density, particularly in that direction perpendicular to the plane of the map sections in the Richards box. Five difference Fourier refinement cycles rapidly corrected these translations, R at 1·76 Å dropping to 26·9%. During these cycles a number of chemically unreasonable and low-occupancy solvent molecules were removed, leaving a total of 174. Following these cycles, 1·5 Å data were added, and one subsequent refinement cycle reduced the 1·5 Å R to 27·7% for the constrained structure containing 174 ordered solvent molecules (r.m.s. deviations from ideality for bond lengths, angles, and dihedral angles were 0·07 Å, 3·2 and 4·6° respectively).

Changes from the starting model

A histogram indicating the movements of α-C atoms between the starting and current models is shown in Fig. 5. The average positional change for α-C atoms was 0·9 Å, and for all atoms it was 1·2 Å. The largest movements were in the regions of residues 60–63, 74–80 [the primary calcium binding site, first identified and described in detail by Bode & Schwager (1975*a*,*b*)], 110–121, 144–147 (the α-, β-trypsin autolysis site), 186–187, and particularly in the C-terminal α-helix. All of these residues are on the surface of the enzyme, and were not unambiguously defined in the MIR map. The problem with the terminal α-helix was largely a translational one, since it was the region farthest from the mirror of the Richards box when building the wire model, and was therefore most prone to errors in depth perception.

One type of error commonly encountered over the course of the refinement was the 180° misorientation of an amide plane. At a resolution of 2·7 Å the MIR map did not allow unambiguous assignment in a number of cases; however, the errors were clearly identified in the high-resolution ΔF maps. The regions where these misinterpretations occurred were again all on the surface of the molecule: the amide planes between residues 22–23, 24–25, 34–37, 39–40, 75–76, 84–85, 85–86, 110–111, 117–118, 144–145, 187–188 and

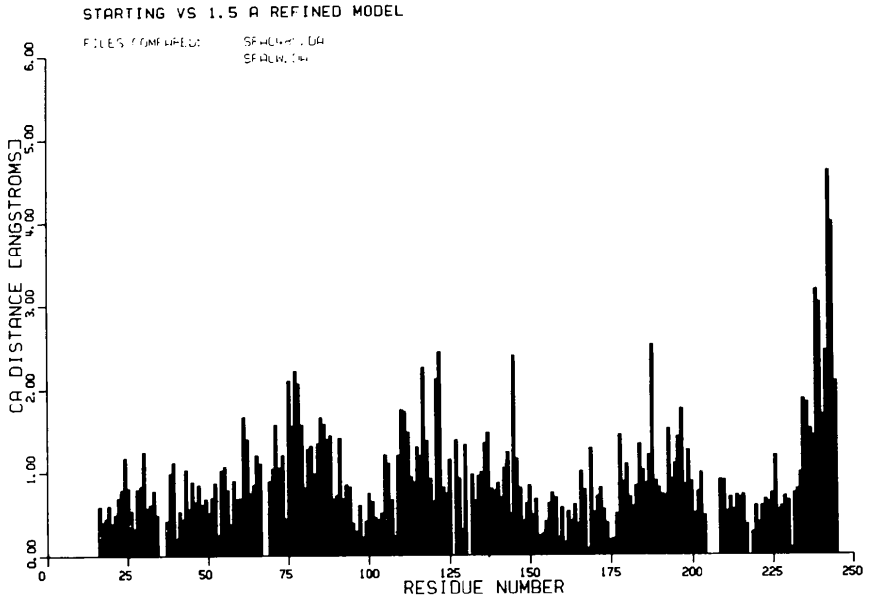

Fig. 5. Histogram showing the distances between α-C atoms in the starting and refined structures. The largest changes all occurred on the surface of the molecule, especially in the C-terminal α-helix.

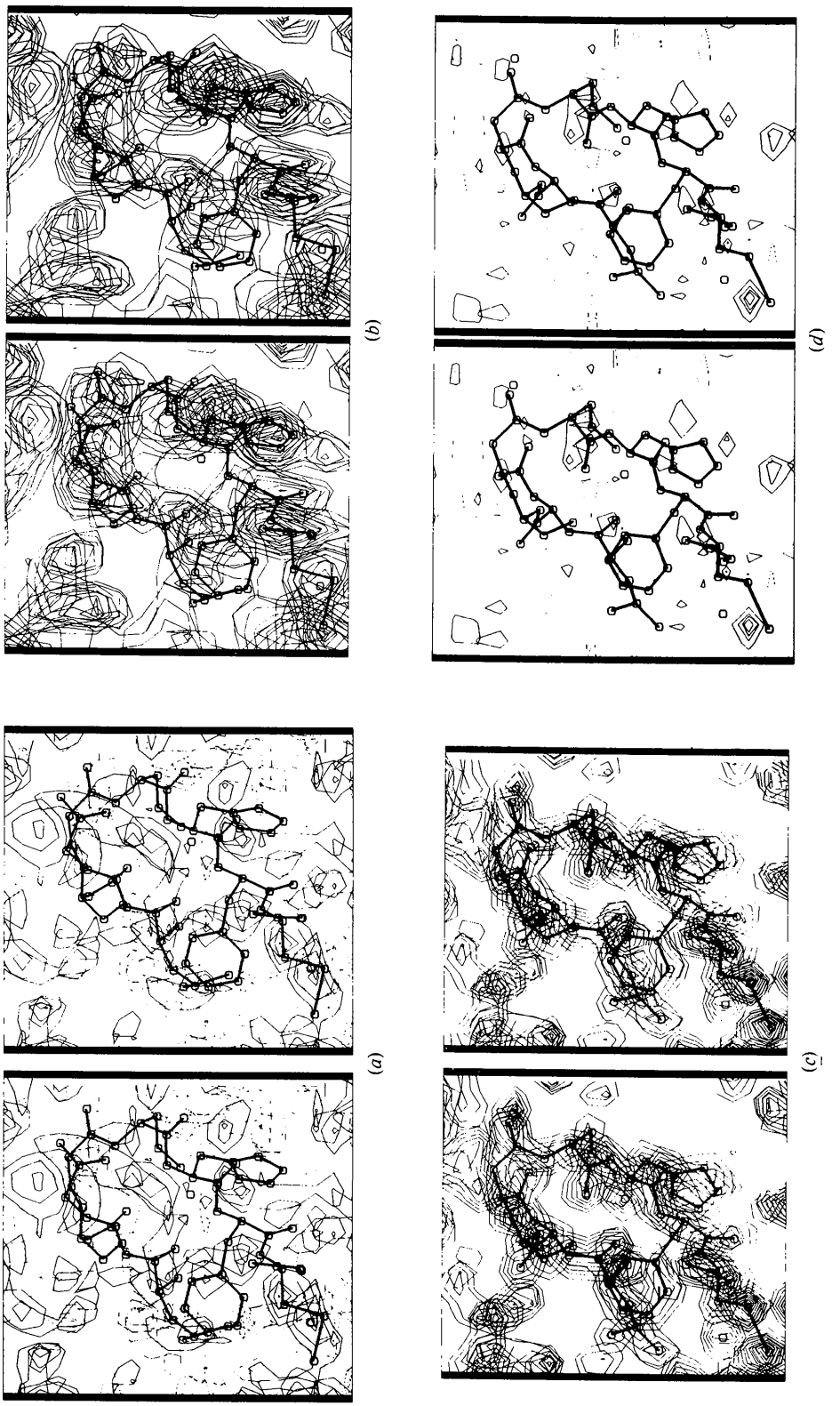

Fig. 6. The fit of the model to the electron density at different stages of refinement in the region around His 40, viewed down the crystallographic *b* axis, pictured in stereodrawings generated with the minicomputer. The portion of the sequence shown is L33, N34, S37, G38, Y39 (the entire ring is not shown), H40, F41, C42 (plus the Sγ of C58), and G43. (*a*) The misorientation of the Y39–H40 (and to a lesser extent the N34–S37) amide is reflected in an early 2·1 Å ΔF map. Contours are from 0·3 e Å^{-3} in increments of 0·15 e Å^{-3}. The dotted contours represent negative density. (*b*) The orientation of these amides is ambiguous in the 2·7 Å MIR map, contoured from 0·375 e Å^{-3} in steps of 0·25 e Å^{-3}. (*c*) The reorientation of these amides allows an excellent fit to the 1·5 Å refined $|F_o \exp(i\varphi_c)|$ map, where the ambiguity no longer exists. Contours are from 0·5 e Å^{-3} in steps of 0·35 e Å^{-3}. (*d*) The 1·5 Å ΔF map is correspondingly clean since the required corrections have been made. Contours are from 0·2 e Å^{-3} in 0·1 e Å^{-3} steps.

222–223 all required reorientation. The region of the molecule around His 40 in one of the early 2·1 Å ΔF maps, showing the misorientation particularly of the Tyr 39–His 40 amide, is pictured in Fig. 6(*a*). Fig. 6(*b*) is a drawing of the 2·7 Å MIR map and starting model in the same region. At this resolution, an unambiguous assignment cannot be made for the Asn 34–Ser 37 and Tyr 39–His 40 peptides. Fig. 6(*c*) and (*d*) are respectively pictures of the current model and 1·5 Å refined map, and the 1·5 Å difference map in this region. The difference map is much flatter since the required corrections have been made.

Although large structural changes did occur in some parts of the molecule, the chemically important residues of the active site and binding pocket did not move greatly upon refinement. Positional changes for these residues resulted from limitations in resolution and clarity of the MIR map, rather than from misinterpretation. The most significant change involving the active site is the tilting of the Asp 102 carboxyl group and the His 57 imidazole into positions where the interaction between them is no longer symmetric as previously suggested (Stroud *et al.*, 1974; Krieger, Kay & Stroud, 1974). The δ-N of His 57 points more toward the lower O atom, Oδ-2, of the Asp 102 carboxyl, as shown in Fig. 7, forming a 2·7 Å hydrogen bond. The distance between the upper O atom, Oδ-1, and His 57-Nδ is too great for favorable hydrogen-bond formation, 3·9 Å. However, this atom does form hydrogen bonds with the main-chain N atoms of Ala 56 and His 57 of 2·7 and 3·2 Å respectively. As can be seen from Fig. 7, the positions of these groups are well defined in the refined map.

Refinement of the structure from residue Asn 72 to Gln 81 indicated a tightly bound positive ion and several well-ordered solvent molecules, which were previously not resolved in the 2·7 Å MIR map. Bode & Schwager (1975*a*,*b*) in their independent refinement of the structure of benzamidine-inhibited bovine trypsin recognized this area as the primary calcium ion binding site. While the refined structure of DIP-trypsin arrived at from the present study is nearly identical in this area to that reported by Bode & Schwager, the density for the positive ion in the present study is about half of that expected for a fully occupied calcium ion. It is possible that this site is partially occupied by Ca^{2+}; however, the ligands do not appear disordered as expected for such a situation. Since the crystallizing solution contained a high concentration of Mg^{2+} (5·8% $MgSO_4$), occupancy by a magnesium ion is another possibility in spite of the difference in atomic radius between Ca^{2+} and Mg^{2+}. Occupancy by a water molecule is not as likely, because of the high density of negative charge surrounding this site.

Discussion

There are a number of criteria which indicate that the refinement has indeed resulted in an improved structure. First, the refined structure fits the observed data better than did the starting structure. The most commonly quoted index of this agreement is the standard crystallographic R, which decreased from 50% at 2·1 Å to 27·7% at 1·5 Å, as just described. This decrease is derived, however, from a number of sources in addition to improvement of atomic positions, including the use of individual temperature factors, inclusion of solvent, use of the four-parameter scaling procedure, and omission of the approximately 700 reflections with $F_o \gg F_c$ from the data set. The contribution of each of these factors to the decrease in R can be estimated by computing the R value for the refined model without its benefit. The respective contributions to R at 2·1 Å are 3, 3, 2 and 1·5%. The R values for the starting and current model at this resolution are 50 and 25%, respectively, leaving about 15 of the 25% decrease attributable to improvement in atomic positions of the atoms originally in the starting model.

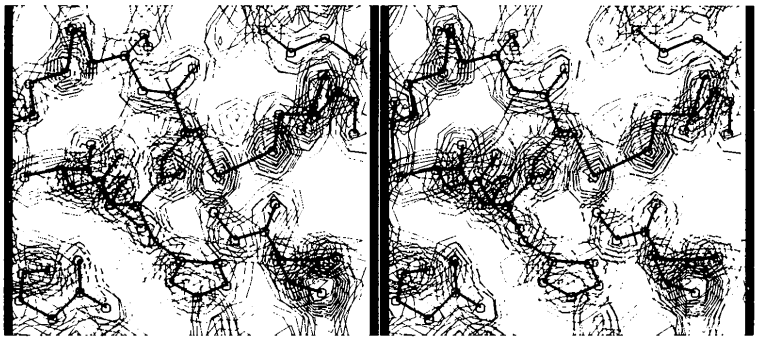

Fig. 7. A view down the *b* axis of the area around the catalytic site residues Asp 102, His 57, and Ser 195, showing the fit of the refined coordinates to the 1·5 Å refined $|(2F_o - F_c)\exp(i\varphi_c)|$ density. The S atoms in the 42–58 disulfide bridge are well resolved. The very large peak at the lower right arises from the P and O atoms of the DIP group. Contours start at 0·5 e Å^{-3} in 0·5 e Å^{-3} steps.

Although R is a useful indicator of the overall progress of the refinement, it was noted in this study as well as by Huber *et al.* (1974) that small changes in isolated portions of the structure, such as a flipped amide plane, are not strongly reflected in the value of R. A much more sensitive device for the detection of these local errors is the ΔF map. The average unsigned difference electron density dropped from 0·31 e Å^{-3} for the 2·1 Å ΔF map at the start of the refinement to 0·08 e Å^{-3} for the 1·5 Å refined ΔF map, where the integration was performed over an entire asymmetric unit. The present 1·5 Å difference map contains few interpretable features, with no density above 0·54 e Å^{-3} in magnitude. These numbers can be compared with 2·5 e Å^{-3} for a well-defined carbonyl O in the refined $2F_o - F_c$ map.

The correctness of the refined structure is also indicated by the representative result shown in Fig. 8, which is a picture of the 1·5 Å $2F_o - F_c$ map in the same area as that of Fig. 6. In this case, however, all the atoms shown (N34 C=O, S37, G38, Y39 main chain, H40, F41, C42, and G43) have been omitted from the structure factor calculation. The density for these atoms is nevertheless very clear and, although the resolution suffers somewhat, the structural information is clearly contained in the observed F's. This result serves to emphasize that the $2F_o - F_c$ map is not merely a reflection of the atoms included in the phase calculation.

The second criterion indicating the improvement of the refined model is that it has become chemically more reasonable. Larger regions of the protein fit into classical patterns of secondary structure. In a number of regions formerly lacking good hydrogen bonding, new hydrogen bonds were formed, many of them through solvent molecules. Forty-nine close contacts between nonbonded atoms, shorter than the extreme limits of Ramachandran & Sasisekharan (1968), have been eliminated. The wire model rebuilt from the 1·76 Å refined map contains no large empty regions within the interior (except the specific binding pocket), whereas several of these areas were present in the starting model.

The MIR map also serves as a check on the correctness of the refined model, and was a very useful aid in regions where the calculated maps were noisy. The major changes made in the structure resulted primarily from difficulties in interpretation of the MIR map (except in the terminal α-helix); however, these changes have not generally resulted in contradictions with this map. The refined coordinates fit the MIR electron density considerably better than did the starting coordinates.

The mean change in phase angle, $\Delta\varphi$, between the 'best' MIR and refined phases was 62°, and between the phases calculated from the starting model and refined phase it was 58°, for all reflections to 2·7 Å. These are large average changes, but are similar to those reported for other protein structures (Watenpaugh *et al.*, 1973; Huber *et al.*, 1974; Moews & Kretsinger, 1975; Deisenhofer & Steigemann, 1975). However, the appearance of the Fourier map is ultimately determined by contributions from a large number of reflections, and because $\Delta\varphi$ contains no information concerning the interrelations between the various reflections, it is not a very meaningful number when considering how different the MIR and refined maps really are.

In order to visualize the kinds of differences involved between the MIR and refined maps, a difference synthesis was computed with coefficients $F_o \exp(i\varphi_{MIR}) - F_o \exp(i\varphi_c)$. While substantial difference density ($\Delta\rho$) was present over most of the map, this density was considerably smaller at atomic positions than the corresponding density (ρ_0) in an $F_o \exp(i\varphi_c)$ map computed at the same resolution. This result reflects the strong

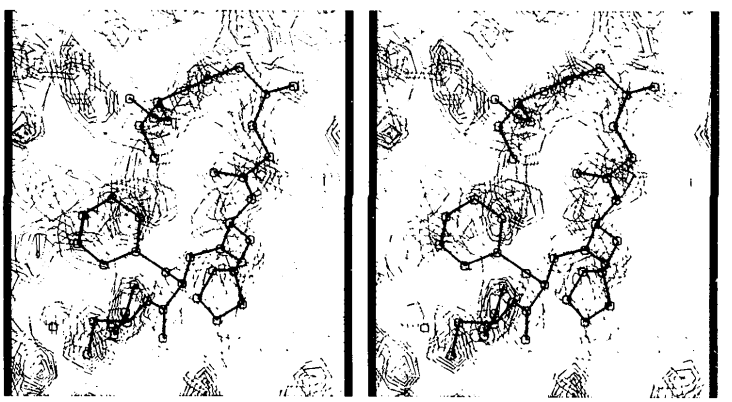

Fig. 8. View of the same area as shown in Fig. 6 of a 1·5 Å $(2F_o - F_c)\exp(i\varphi_c)$ map. All the atoms shown in this drawing were omitted from the phasing, indicating that the structural information is indeed present in the observed F's. Contours are from 0·5 e Å^{-3} in 0·4 e Å^{-3} steps.

similarity between the MIR and 2·7 Å refined maps. An index of the disagreement between these maps can be computed from the r.m.s. densities in the difference and F_o syntheses:

$$R_\varphi = \frac{\langle \Delta\rho \rangle_{\text{r.m.s.}}}{\langle \rho_0 \rangle_{\text{r.m.s.}}}.$$

Averaged over the entire asymmetric unit, R_φ was 86%. This number is smaller than the R_φ of 106% estimated from the 62° value of $\Delta\varphi$, indicating that the differences between the MIR and refined phases tend to be smaller for the larger reflections. Averaged over only those grid points surrounding the atoms in the refined structure, R_φ fell to 28%.

The average change of 1·2 Å between the starting wire-model coordinates and the current set is quite substantial, and arises from two major sources of error. The first source involves errors in interpretation caused primarily by the lack of resolution and clarity, and phase errors in the MIR map. These changes result in movements of up to several ångströms, but occur over relatively small portions of the molecule. The other major source of error is in the actual construction of the wire model, particularly due to inaccurate depth perception, irregularities in the mirror, and mechanical difficulties in fitting residues together, placement of supports, and bent parts. These errors are comparatively small, but are distributed over the entire molecule. Errors arising from measurement of wire-model coordinates were very small – within 0·5 mm, or 0·025 Å.

A reasonable estimate of the size of the non-interpretative errors in the wire-model coordinates can be derived from the positional differences between the coordinates of the wire model rebuilt at 1·76 Å (revision 16 in Fig. 5) and those obtained from the five cycles of refinement following this rebuilding, since no gross reinterpretations were involved during these cycles. The mean change in atomic positions between these two sets was 0·6 Å. There were no overall systematic translations between the data sets compared. Correction of these errors led to a substantial decrease in R, from 35 to 27% at 1·76 Å, the R value being sensitive to small translations of large pieces of the structure. Similar types of errors, although considerably smaller in magnitude (and dependent upon the coarseness of the map grid), appeared when performing shifts manually in the small-scale maps. For this reason it is felt that these approaches are inadequate to provide an optimal fit to the electron density: some type of computerized scheme for atomic shifts is essential for an optimal fit with a correspondingly lower value of R.

Several early problems resulted from the high content of ordered and disordered solvent. In retrospect, the presence of ordered solvent was one of the greatest factors in misinterpretation of the MIR map. At 2·7 Å resolution little of the ordered solvent was separated from neighboring residues, leading to inaccurate placement of these residues in the model.

Although it was found that inclusion of ordered solvent and use of individual temperature factors could produce significant reduction in the R value in early stages, it is felt that their application in the initial stages of refinement is a mistake. Difference Fourier maps at the start were relatively noisy, making selection of peaks corresponding to real solvent difficult. Solvent added in these early stages tended to move toward the atoms of the protein, influencing the gradients and corresponding atomic shifts. In these stages it is best to add solvent only at a very high confidence level (the MIR map can be very helpful), and preferably not until the ΔF maps have become easily interpretable. Because of the strong coupling between the atomic positions and the shifts indicated for individual temperature factors, individual temperature factors should also be left largely untouched at the beginning of the refinement.

Unfortunately, the fact that the ΔF maps were very noisy in early cycles made location of parts of the structure which had been omitted from the phasing, or which were very far from the correct position, very difficult. The technique which produced the best result in this situation was to perform several cycles of refinement in order to optimize the fit for the known parts of the structure. This improved the phases to the point where the missing features could often be recognized, and made the ΔF maps considerably less noisy. One method which was not employed in this study but may be very useful in cases where the early calculated maps are very noisy is the use of a weighting factor such as the one described by Sim (1959) to reduce the contribution of terms which are likely to be poorly phased.

Although the R value and the flatness of the ΔF map indicate that the refined structure is generally correct, there are a few regions which would benefit from further refinement. The first is the area around Lys 145, where the α/β-trypsin autolytic cleavage occurs. There is some disorder in this region arising from the fact that the crystals used contain about 40% α-trypsin, and that this region forms an external loop, with some thermal motion probably present. Bode & Schwager (1975*a*) in their refinement of the benzamidine–trypsin structure, which was analyzed as 80% β-trypsin, also suggest that there is some mobility in this region. Another area requiring further work is the C-terminal α-helix between residues 238–244. The slowness of convergence in this region is primarily a result of the large errors in the starting model.

There are several improvements which can be made in the present refinement system. Although a refinement cycle can presently be carried out in a single day, it would be advantageous to shorten the time required. The rate-limiting step is presently the calculation of structure factors. It may be possible to greatly shorten

this step by use of a three-dimensional fast Fourier transform (FFT) program such as that described by Ten Eyck (1973). Generation of sampled electron density from a list of atomic coordinates, suitable for transformation to obtain 1·5 Å structure factors (*i.e.* one asymmetric unit of the structure on a 0·75 Å grid) currently requires about 20 min on the minicomputer. A three-dimensional FFT program would also be useful for generating electron density maps, and implementation on the NOVA 800 would eliminate the need for use of a large computer for any step in the cycle. The fast Fourier algorithm would be essential for minicomputer refinement of a much larger structure, or for extending the resolution of the present structure to the limits imposed by the diffraction pattern.

The treatment of low-angle reflections, currently accomplished through use of the nonlinear scaling expression, might be improved by including some physically reasonable model for the disordered-solvent continuum in the structure factor calculation.

An improvement which is expected to facilitate convergence of the refinement is the incorporation of information from the electron density map into the residual minimized in the constraints program, to yield a constrained real-space refinement program. The modifications required are very simple, and the results obtained to date in tests of this program indicate that it is considerably more efficient than alternate cycles of atomic shifts and constraints. This program can also be readily adapted for energy minimization of the type described by Levitt (1974), as suggested by Hermans & McQueen (1974). Konnert (1976) has used a similar technique in reciprocal space to perform constrained structure factor least-squares refinement. One further capability which can be added to the constraints program involves the configuration at asymmetric atoms. Currently, any inversion of configuration brought about by allowing atomic shifts to become too large between applications of constraints must be corrected manually. Nevertheless, it seems likely that inclusion of information about the configuration, such as the scalar triple product of the bonds to the atom, could prevent or correct these errors.

The computing cost of the DIP-trypsin refinement to date has been approximately $650, almost entirely for calculation and plotting of Fourier maps on the IBM 370/158. About 320 h of CPU time on the NOVA 800 have been required over the course of the refinement for programs in which computation is rate-limiting, excluding program development. Considering the sevenfold decrease in CPU time that we have generally observed for computer-bound programs on the IBM 370/158, this translates into about $25 000 for equivalent time on the larger machine at the current overall average rate of $540 per CPU hour (including use of peripherals, *etc.*). It must be kept in mind that, had the refinement been carried out on a large machine, the programs would have been written to take advantage of the greater core storage available with a corresponding significant decrease in execution time. However, since the figure above is exclusive of program development and neglects the very large number of smaller programs executed during the refinement, it is felt that this figure represents a realistic estimate of the equivalent cost on the IBM 370. The savings in computing costs with the use of the minicomputer have therefore been substantial, and considering the potential cost for refinement of a protein structure, these savings can rapidly begin to offset the initial investment for the smaller machine.

Since this paper was first submitted for publication, two cycles of a least-squares refinement procedure using a first-order gradient minimization reduced the 1·5 Å R to 23·5%. The mean change in atomic position was small – about 0·07 Å during each cycle. Details will be described in a later paper.

Note added in proof:– Six more difference Fourier refinement cycles have since reduced R to 21·5% at 1·5 Å.

We wish to thank Drs Monty Krieger and Anthony Kossiakoff for valuable discussions.

References

BODE, W. & SCHWAGER, P. (1975*a*). *J. Mol. Biol.* **98**, 693–717.

BODE, W. & SCHWAGER, P. (1975*b*). *FEBS Lett.* **56**, 139–143.

CHAMBERS, J. L., CHRISTOPH, G. G., KRIEGER, M., KAY, L. M. & STROUD, R. M. (1974). *Biochem. Biophys. Res. Commun.* **59**, 70–74.

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

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

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

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

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

DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 311–315.

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

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

FREER, S. T., ALDEN, R. A., CARTER, C. W. JR & KRAUT, J. (1975). *J. Biol. Chem.* **250**, 46–54.

HENDERSON, R. & MOFFAT, J. K. (1971). *Acta Cryst.* B**27**, 1414–1420.

HERMANS, J. JR & MCQUEEN, J. E. JR (1974). *Acta Cryst.* A**30**, 730–739.

HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., BARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

KOSSIAKOFF, A. A., CHAMBERS, J. L., KAY, L. M. & STROUD, R. M. (1976). *Biochemistry*. In the press.

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.
LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.
MARSH, R. E. & DONOHUE, J. (1967). *Advanc. Protein Chem.* **22**, 235–256.
MOEWS, P. C. & KRETSINGER, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
RAMACHANDRAN, G. N., LAKSHMINARAYANAN, A. V. & KOLASKAR, A. S. (1973). *Biochem. Biophys. Acta*, **303**, 8–13.
RAMACHANDRAN, G. N. & SASISEKHARAN, V. (1968). *Advanc. Protein Chem.* **23**, 284–438.
RICHARDS, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.
SALEMME, F. R. & FEHR, D. G. (1972). *J. Mol. Biol.* **70**, 697–700.
SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination: A Practical Guide*, pp. 356–385. New York: Macmillan.
STROUD, R. M., KAY, L. M. & 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.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
TEN EYCK, L. F., WEAVER, L. H. & MATTHEWS, B. W. (1976). *Acta Cryst.* A**32**, 349–350.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.

Acta Cryst. (1977). B**33**, 1837–1840

Structure Cristalline du Tétraméthyl-2,2,6,6 (Hydroxyimino)-4 Pipéridine Oxyl-1

PAR D. BORDEAUX ET J. LAJZÉROWICZ

Laboratoire de Spectrométrie Physique (associé au CNRS), Université Scientifique et Médicale de Grenoble, BP n° 53, 38041 Grenoble Cédex, France

(*Reçu le 3 novembre* 1976, *accepté le 24 novembre* 1976)

2,2,6,6-Tetramethyl-4-(hydroxyimino)piperidine-1-oxyl crystallizes in space group $Cmc2_1$, with $Z = 4$, $a = 13{\cdot}634$ (5), $b = 9{\cdot}158$ (5), $c = 8{\cdot}292$ (2) Å. The structure is a racemic solid solution with static disorder; rigid-molecule refinement and **TLS** constraints have been used. The nitroxide free-radical molecules have the 'chair' conformation and are linked by hydrogen bonds.

Introduction

Le radical nitroxyde tétraméthyl-2,2,6,6 (hydroxyimino)-4-pipéridine oxyl-1 (que nous appellerons tanoxime dans la suite de l'exposé) a été synthétisé et étudié au Laboratoire de Chimie Organique Physique du Centre Nucléaire de Grenoble. Les valeurs des écarts hyperfins du ^{13}C et de la largeur de raie électronique permettent d'envisager une forme croisée pour les molécules de tanoxime en solution (Brière, Lemaire & Rassat, 1965); les molécules sont alors en inversion rapide. Par ailleurs, l'étude des propriétés magnétiques sur poudre a conclu à l'existence d'un ordre magnétique à basses températures (Veyret, 1975).

La détermination de la structure était intéressante pour confirmer et essayer d'interpréter ces résultats d'autant plus que les symétries des groupes d'espace trouvés posaient des problèmes pour la structure.

Partie expérimentale

Données cristallographiques – groupe spatial

A partir d'une solution d'alcool méthylique, nous obtenons des cristaux rouges de forme bipyrimidale [$a = 13{,}634$ (5), $b = 9{,}158$ (5), $c = 8{,}292$ (2) Å; $Z = 4$; $V = 1035$ Å^3].

Les extinctions conduisent aux trois groupes: *Cmcm*, les positions spéciales d'ordre 4 ont la symétrie *2mm*; *Ama2*, les positions d'ordre 4 ont la symétrie 2 (paramètres de maille: $a = 8{,}292$, $b = 9{,}158$, $c = 13{,}634$ Å); et $Cmc2_1$, les positions d'ordre 4 ont la symétrie 2 ou *m*.

Les tests de symétrie sur les facteurs de structure normalisés conduisant à un groupe noncentrosymétrique, nous avons éliminé le groupe *Cmcm*.

Le cycle pipéridine peut avoir: (1) Soit une conformation croisée: conformation admise en solution