References

CHUNG, S. J. & HAHN, TH. (1975). *Acta Cryst.* A**31**, S1.
CHUNG, S. J. & HAHN, TH. (1976). *Z. Kristallogr.* **144**, 427.
CHUNG, S. J., HAHN, TH. & KLEE, W. E. (1983). *Z. Kristallogr.* **162**, 51–53.
FIGUEIREDO, M. O. (1982). *Fortschr. Mineral. Beih.* **60**, 26–27.
FISCHER, W. (1973). *Z. Kristallogr.* **138**, 129–146.
FISCHER, W. (1982). *Fortschr. Mineral. Beih.* **60**, 23–24.
HARARY, F. (1969). *Graph Theory.* Reading, MA: Addison-Wesley.
HEESCH, H. & LAVES, F. (1933). *Z. Kristallogr.* **85**, 443–453.
LIMA-DE-FARIA, J. & FIGUEIREDO, M. O. (1976). *J. Solid State Chem.* **16**, 7–20.
MEIER, W. M. & VILLIGER, H. (1969). *Z. Kristallogr.* **129**, 411–423.
SMITH, J. V. (1977). *Am. Mineral.* **62**, 703–709.
SMITH, J. V. & BENNETT, J. M. (1981). *Am. Mineral.* **66**, 777–788.
WELLS, A. F. (1977). *Three-dimensional Nets and Polyhedra.* New York: Wiley.
WELLS, A. F. (1979). *Further Studies of Three-Dimensional Nets. ACA Monogr.* No. 8. Pittsburgh: Polycrystal Book Service.

Acta Cryst. (1984). A**40**, 50–57

Crystallographic Refinement of Macromolecules having Non-crystallographic Symmetry

BY T. ALWYN JONES AND LARS LILJAS

Department of Molecular Biology, Wallenberg Laboratory, Box 562, Uppsala, Sweden

(*Received* 22 *April* 1983; *accepted* 30 *August* 1983)

Abstract

Some very large biological macromolecules, such as viruses, exhibit a high degree of non-crystallographic symmetry. A method is described to refine crystallographically such large structures using a combination of molecular averaging in real space, automatic real-space refitting and interactive refitting using computer graphics. The method has been successfully applied to a small plant virus, satellite tobacco necrosis virus, containing 11 700 amino acids in the crystallographic asymmetric unit. The starting model for the refinement was built with 3·7 Å phases. These have been refined and the resolution extended to 2·5 Å.

Abbreviations used in the text: STNV satellite tobacco necrosis virus, SBMV southern bean mosaic virus, TBSV tomato bushy stunt virus, FFT fast Fourier transform, n.c.s., non-crystallographic symmetry.

Introduction

The first macromolecules to be crystallographically refined were small proteins containing about fifty amino acids in the asymmetric unit. Rubredoxin was refined using difference-Fourier and reciprocal-space least-squares methods (Watenpaugh, Sieker, Herriot, & Jensen, 1973), and pancreatic trypsin inhibitor by cyclic application of real-space refinement (Deisenhofer & Steigemann, 1975; Diamond, 1971). Since then reciprocal-space refinement has been improved by introducing extra observations in the form of restraints to bond lengths, angles *etc.* (Konnert, 1976; Hendrickson & Konnert, 1980), by using constraints and elastic restraints (Sussman, Holbrook, Church & Kim, 1977), and by the use of FFT methods (Agarwal, 1978; Jack & Levitt, 1978). These methods are not automatic and require manual intervention at various stages. Fortunately, this aspect has been greatly simplified by the use of computer graphics systems such as *FRODO* (Jones, 1982) and *BILDER* (Diamond, 1982). It is now possible to refine successfully protein molecules containing 750–800 residues in the asymmetric unit. However, it is still very difficult and time consuming to build and refine macromolecules starting from maps that are poorly phased with data extending to 3·0–3·5 Å resolution. This requires a lot of manual intervention and many cycles of refinement [*e.g.* the immunoglobin Fc fragment refined by Deisenhofer (1981)].

Many interesting macromolecules contain multiple copies of a protein subunit. Viruses in particular contain many copies of a single polypeptide chain. Three spherical plant viruses have been extensively studied by X-ray crystallography, STNV, TBSV and SBMV (Liljas *et al.*, 1982; Harrison, Olson, Schutt, Winkler, & Bricogne, 1978; Abad-Zapatero *et al.*, 1980). All three have icosahedral symmetry. In the classification of Caspar & Klug (1962) STNV is of the simplest $T = 1$ type, possessing exact icosahedral symmetry and therefore having 60 identical subunits in its protective coat. It turns out that when STNV crystallizes the asymmetric unit is the complete virus particle. Both TBSV and SBMV are $T = 3$ particles with 180 subunits making up the particle. When they crystallize some of the icosahedral symmetry elements become space-group symmetry elements so that TBSV has 15 and SBMV 30 protein subunits in the crystallographic asymmetric unit.

0108-7673/84/010050-08$01.50 © 1984 International Union of Crystallography

The STNV structure determination of Liljas *et al.* (1982) gave a map calculated with low-resolution phases, which resulted in a relatively poor atomic model with 11 700 amino acids in the asymmetric unit but with high-resolution data available to 2·5 Å. If we wish to produce a refined structure of STNV the problem is intractable with the present generation of computers unless we rigidly constrain the model to maintain strict icosahedral symmetry. Of course, any deviation from this n.c.s. results in loss of signal in the electron density maps.

Possible methods

Structure-factor calculations can be made using coordinates having exact n.c.s. At least three refinement methods readily suggest themselves as possible candidates:

1. *Automated difference Fourier*

Freer, Alden, Carter & Kraut (1975) have described the use of this method for proteins where individual atoms are shifted according to the gradient in the difference Fourier calculated with model phases. In our case a difference Fourier calculated from model phases will not possess exact n.c.s. However, the map can be averaged according to the n.c.s. and then used to calculate shifts of the atoms in the n.c.s. unit. Since one starts with poor coordinates it would be sensible to evaluate the shifts for groups of linked atoms. After applying the shifts, the distorted structure can be regularized to maintain correct stereochemistry. In this method the coordinates are changed just once in the cycle.

2. *Jack & Levitt method*

This method uses Cochran's (1948) observation that the shifts computed from a difference map are equivalent to those from diagonal least squares except that each term in the normal matrix is weighted by the inverse of the atomic form factor. Jack & Levitt (1978) obtained the first and second derivatives of the crystallographic residual from a difference Fourier which again could be an averaged map. Scaled shifts can be applied directly to the coordinates or the crystallographic residual and its derivatives passed as a pseudo-energy term to Levitt's energy minimization program (Levitt, 1974).

3. *Real-space refinement*

The method described by Diamond (1971) attempts to minimize the function

$$R = \int (\rho_{\text{obs}} - \rho_{\text{model}})^2 \, dv$$

whilst maintaining strict stereochemical constraints (except for the τ angle). The program was extensively used for crystallographic refinement by improving the fit of the coordinates to electron density maps based on calculated phases (Deisenhofer & Steigemann, 1975). A number of cycles of refitting can be made into the same map without making a new structure-factor calculation. Clearly, for structures with n.c.s. the map would be an averaged $2F_o - F_c$ map.

For all three methods, a cycle of refinement consists of a structure-factor calculation, a Fourier calculation, averaging, and then the making of a new model. Probably 4–5 cycles of refinement are possible before manual intervention on a computer graphics system (Deisenhofer, 1981; Remington, Wiegand & Huber, 1982). These methods suffer from the fact that the model phases have severe errors in them, which hopefully get reduced during the course of the refinement. The presence of n.c.s. allows us to reduce these phase errors at any stage in the refinement by using the cycling process shown in Fig. 1 which is a phase refinement method. This results in an improved electron density map to which the atomic model can be refitted. We therefore decided to use real-space methods for the refinement of STNV.

STNV refinement cycle

Step 1

In this step we generate the electron density distribution of the n.c.s. unit from a list of atomic coordinates. For STNV the icosahedral asymmetric unit was taken as the triangular wedge enclosed by two fivefold and one threefold axes. Parts of three polypeptide chains are contained in this wedge. We use the program described by Ten Eyck (1977), where each atom

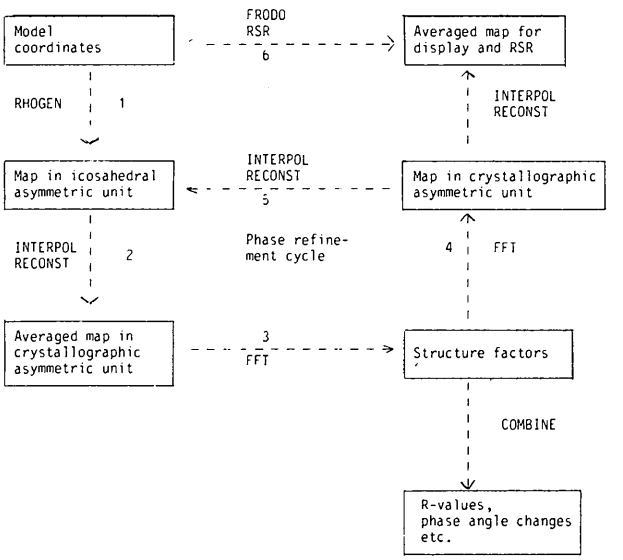

Fig. 1. STNV refinement cycle, combining cyclic phase refinement and real-space refitting.

is described as the sum of three Gaussian functions which take into account the atomic form factor and temperature factor. We reduce the aliasing problem of the FFT by adding an additional temperature factor of 20 Å^2. The effect of this extra temperature factor is removed in step 3.

Step 2

The map in the n.c.s. unit is expanded into the crystallographic asymmetric unit, using the program system of Bricogne (1974, 1976). A program equivalent to *GENERATE* with option 5 is used to generate indices for Bricogne's *INTERPOL* program. These indices give for all map points inside an envelope in the crystallographic asymmetric unit the corresponding point in the n.c.s. unit after the appropriate rotations and translations. After interpolation in the subunit map, the *INTERPOL* output is sorted and made into a crystallographic asymmetric unit map by *RECONST*. This map had a grid spacing of 1·05 Å throughout the refinement. All lists of indices from *GENERATE* were kept on magnetic tape, while all maps and interpolated densities were stored on disk.

Step 3

The structure-factor calculation is made with a space-group-specific FFT program of Ten Eyck (1977). The observed and calculated structure factors are scaled and weighted with the program *COMBINE* of Bricogne which also removes the effect of the extra temperature factor introduced in step 1. We modified the program to calculate local scale factors within intervals of 0·0005 Å^{-2} in $(\sin\theta/\lambda)^2$.

Step 4

Fourier calculation with the FFT (Ten Eyck, 1973) using $2F_o - F_c$ amplitudes and model phases. The resulting map does not possess n.c.s.

Step 5

The map is averaged using Bricogne's program system of *INTERPOL*, sort, and *RECONST*. The indices for *INTERPOL* are produced by *GENERATE* with option 2, where for all points inside an envelope covering 1/60th of the virus particle the 60 corresponding points in the map of the whole virus particle are calculated. The envelope used by *GENERATE* was the same as that used in our earlier work (Liljas *et al.*, 1982). After interpolation and reconstruction, the averaged map is in the asymmetric unit produced by step 1.

Steps 2–5 constitute the phase-refinement stage and may be repeated a number of times. Finally, an averaged map is reformatted for the refitting process, using the program *MAPPAGE*, written by T. A. Jones. In this format, the map is split into non-overlapping three-dimensional bricks of scaled density in a way similar to the pre-contoured electron density maps used by the early versions of *FRODO* (Jones, 1978).

Table 1. *Dictionary of amino acids*

Ala	*CB*	
Arg	*CB, CG, CD*	*NE, CZ*, NH1, NH2
Asn	*CB, CG*, OD1, ND2	
Asp	*CB, CG*, OD1, OD2	
Cys	*CB*, SG	
Gln	*CB*	*CG, CD*, OE1, NE2
Glu	*CB*	*CG, CD*, OE1, OE2
His	*CB, CG*, ND1, *CD2*, NE2, *CE1*	
Ile	*CB*, CG1, *CG2*	*CD1*
Leu	*CB, CG*, CD1, *CD2*	
Lys	*CB, CG, CD*	*CE, NZ*
Met	*CB, CG*	*SD, CE*
Phe	*CB, CG*, CD1, *CD2*, CE1, *CE2, CZ*	
Pro	*CB, CG, CD*	
Ser	*CB*, OG	
Thr	*CB*, OG1, *CG2*	
Trp	*CB, CG*, CD1, *CD2*, NE1, *CE2*, CE3, CZ2, CZ3, CH2	
Tyr	*CB, CG*, CD1, *CD2*, CE1, *CE2, CZ*, OH	
Val	*CB*1, *CG*1, *CG2*	

Our present version of *FRODO* runs on a Vector General 3404 display and is controlled by a Digital Equipment Vax750. This version can access the map for real-time contouring (Jones, 1982) or for interactive real-space refitting (this we call finger-tip refinement).

Step 6

The refitting process can be entirely interactive using *FRODO* or with a new non-interactive real-space fitting program written by T. A. Jones. Normally we run a number of non-interactive refits and then touch up the results using *FRODO*. In these programs each atom is assumed to have a Gaussian profile of the type

$$\rho(r) = \frac{Z}{a^3}\exp\left(-\frac{\pi r^2}{a^2}\right),$$

where Z is the atomic number of the atom and a the atomic radius. In the non-interactive program the user must define a zone of residues which are to be fitted and the number of grid points around each atomic centre within which to build up the model density. The coordinates are kept in the *FRODO* coordinate file format, and all the density covering the desired atoms is read in from the map data set. After building up the density for the current model, the observed and model densities are scaled together and their differences are stored in computer core. The fitting is made on atomic fragments. Each residue is described in a dictionary as a number of rigid fragments which may consist of any number of atoms. The dictionary presently used for amino acids is shown in Table 1. We have treated the main chain as a fragment CA, C, O and the N of the next residue. The side-chain fragments are chosen to maximize the number of atoms in a rigid group. The fragments are fitted sequentially, first according to the sequence order, and then according to the order in the dictionary.

When the program attempts to fit a particular fragment, it first adds its model density back into the stored difference density and then either attempts to maximize the convolution product

$$\sum \rho_{obs}\rho_{model}$$

or tries to minimize the grid sum

$$\sum (\rho_{obs} - \rho_{model})^2,$$

where the sums are taken over all grid points containing the fragment model density. This method of storing the observed density prevents the fragment from moving into density belonging to other atoms.

We provide two fitting options because although the convolution product biases the fit to place atomic centres on grid points, it is less sensitive to errors produced by scaling the model to the observed density. It can therefore be used when the fit to the observed density is poor. The effect of grid biasing is most severe when fitting single atoms and is reduced by fitting atomic fragments. The program uses the Hooke and Jeeves optimization algorithm as described by Walsh (1975). This algorithm requires a step size for the search and since this is not the time-critical part of the refinement cycle we normally do five searches with decreasing translational steps of 0·5, 0·4, 0·3, 0·2 and 0·1 Å. The step size for the rotational search around the centre of gravity of the fragment is chosen such that at the edge of the fragment the atoms move a distance equal to the translation step.

After making the best fit, the new fragment model density is subtracted from the map to form an updated difference array. After a pass through the complete zone the model coordinates are regularized using *FRODO* in batch mode. This is the restrained regularization described by Hermans & McQueen (1974). The refitting and regularization steps may be repeated a number of times in the same map until the r.m.s. shift is suitably small. To demonstrate the convergence properties of the algorithms we took the best STNV model and introduced random errors of 0·5 Å into a fifty-residue zone of coordinates. These were regularized and then automatically refitted. The r.m.s. difference from the starting coordinates was 0·12 Å when using a minimum translation shift length of 0·1 Å.

FRODO has three new screen menu options for use with real-space fitting. Just like the non-interactive program, these options work with atomic fragments but in this case they are defined using *FRODO*'s bond making and breaking options. One command, RSCL, merely calculates a new constant to scale the fragment's model density onto the observed density. The second, RSR, does a translation and rotation search to improve the fit using the convolution option described above (the necessary constants are defined with a new CHAT option called RSR). The third command, RSFX, does just a rotational search around the identified atom. For the two fitting options, all other atoms on the screen have their contributions subtracted out of the observed density before starting the fitting process. The fragment chosen can be anything from a single water molecule to a whole α-helix. The interactive fitting option is most useful at the beginning of the refinement project, but since the method is non-deterministic it is also used to move fragments out of false maxima.

Fig. 2 shows the remodelling of residues 25 to 27 (Glu, His, Lys). The starting model *M*1 is shown in Fig. 2(*a*) with the final refined 2·5 Å electron density map. These coordinates have severe errors, particularly in the side chain of Glu25 and in the three peptide planes. The errors in the side chain of Glu25 and in the peptide plane between 25 and 26 (made up of the fragment C*A*, C, O of 25 and N of 26) are too large to be corrected automatically by RSR. The strategy employed is to use the *FRODO* command FBRT to move the necessary fragments into the desired density and then to make an exact fit with RSR. In Fig. 2(*b*), the O*E*1, O*E*2, C*D*, C*G* fragment of Glu25, the peptide plane unit between 25 and 26 and the N*Z*, C*E*, C*D* tip of Lys27 have been shifted as groups close to their density. The single atom C*B* of Glu25 has also been manually placed into density. We decided that the peptide plane between 26 and 27 although seriously in error would be fitted correctly by RSR. Each fragment shown in Fig. 2(*c*) has been fitted with RSR which corrected the 26–27 peptide plane without problems. Since there is no attempt to restrain distances and angles between groups during fitting it is necessary to regularize these coordinates. At this stage one can decide if certain atoms should be fixed during the idealization step. Fig. 2(*d*) shows the coordinates obtained with no fixes.

At present *FRODO* will not automatically display non-crystallographically related atoms unless they exist in the coordinate data set as explicit atoms. However, related chains can be put in the data set and be differentiated by extra characters in the residue name.

Results on STNV

Some details of the refinement are given in Table 2. All maps used in the averaging and refitting step were calculated with $2F_o - F_c$ amplitudes. Fig. 3 shows crystallographic *R* factors as a function of resolution for various models and phase-refined averaged maps. The plot marked σ is the σ factor and shows the variation of the function $\sum \sigma / \sum F_o$ with resolution, where σ is our e.s.d. of F_o, the observed structure factor. The nomenclature in the figure specifies the model number and number of cycles of phase refinement. The initial model *M*1 was built with *FRODO* into the map *M*0*P*20 as described by Liljas *et al.*

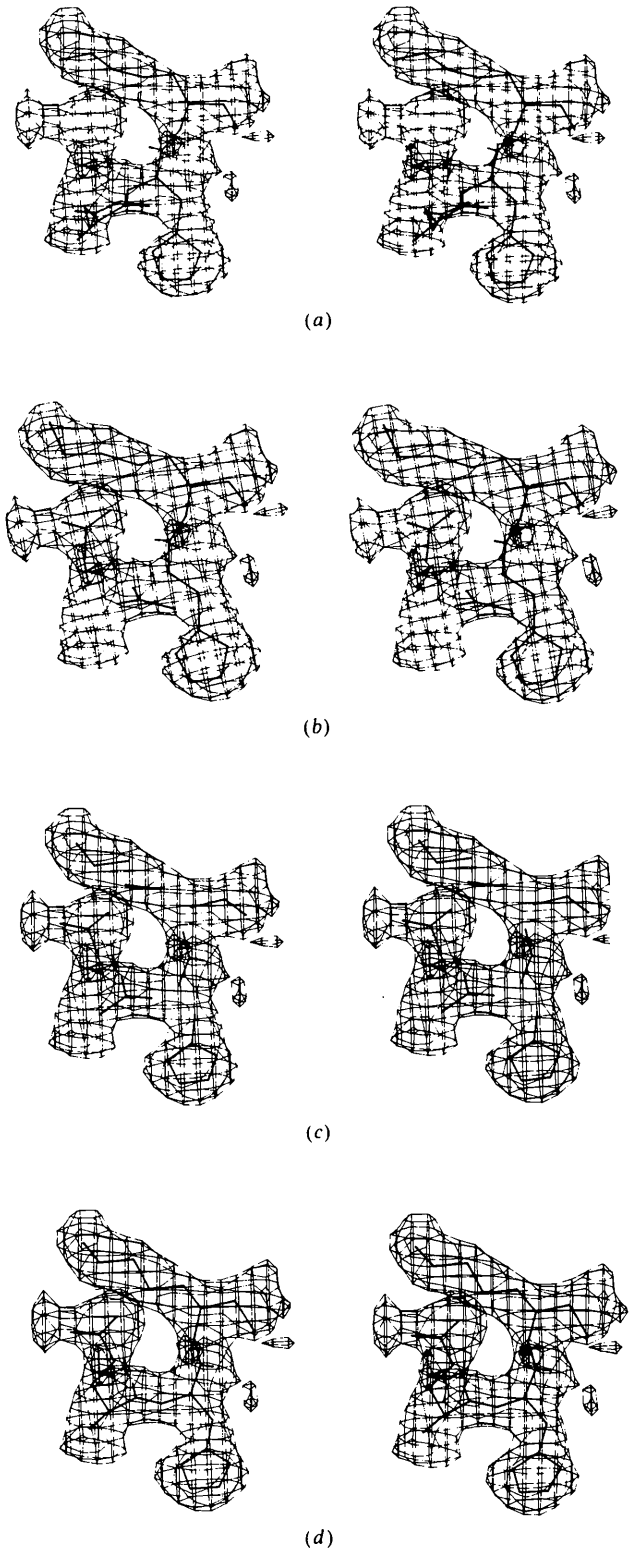

Fig. 2. Interactive refitting with finger-tip refinement. (*a*) Zone 25–27 in our first model. (*b*) A rough manual fit has been made to the fragments most in error. (*c*) Each fragment has been refitted with the RSR menu command. (*d*) The result after model regularization.

(1982). The phases for this map were obtained after 20 cycles of phase refinement averaging out to a resolution of 3·0 Å starting from single isomorphous replacement phases. The *R* factor for this map *M*0*P*20 is shown in Fig. 3(*d*). There is a sharp increase in *R* factor at 3·7 Å which makes it very different from the equivalent curve published by Abad-Zapatero *et al.* (1981) in their work on SBMV. It should be noted, however, that the *R* factors for STNV are better than the equivalents for SBMV for every resolution range. After completing the crystallographic refinement we were able to calculate the mean difference in phases between the reflections used to calculate the *M*0*P*20 map and the last map *M*5*P*2. In the range 15·0–3·7 Å this was 15°, and in the range 3·7–3·0 Å it was 90°, *i.e.* the starting phases in the latter range were random.

Structure factors for the first model *M*1 were calculated to 2·8 Å using an overall temperature factor of 15 Å^2 and using a step 1 grid of 0·7 Å. It is immediately apparent in Fig. 3(*a*) that the *R* factor for this model shows no sharp discontinuity and that it gets gradually worse at high resolution. In four cycles of phase refinement (steps 2–5) the *R* factor distribution changed as shown in Fig. 3(*b*). The map calculated with these phases showed spectacular improvements. Normally, when refining a protein one would be content to locate two or three incorrect peptide planes when inspecting the results on a graphics system. In this second map we flipped over the peptide planes of 30 residues and entirely changed the direction of some side chains. It was also clear that the virus particle has three different metal-ion binding sites. These are probably calcium ions which link n.c.s.-related subunits. One is in a general position and is coordinated by the main-chain carbonyl oxygens of residues 61 and 64, O*D*1 of Asp194 and the O*E*1 of Glu25 from an icosahedral threefold-related subunit. In model *M*1 this region had many errors. In particular, the side chains of Glu25 and Asp63 had been built pointing in the opposite direction in the first map. A second ion is on a special position on the n.c.s. threefold axis and is coordinated by the side-chain carboxylate groups of three Asp55's. The third binding site is located on the fivefold axis, coordinated by the carbonyl oxygens of five Thr138's.

The r.m.s. shifts between coordinates show that the changes made at the first *FRODO* session (changing model *M*1 to *M*1*VG*) are clearly the most significant. At this stage the interactive RSR options were not available, we therefore made three cycles of non-interactive fitting (this was with an earlier not so efficient version of the program) to produce model *M*2. These coordinates were used to extend the phases to 2·6 Å. The step 1 and 5 grids were reduced to 0·65 Å. At the next display session only two peptide planes needed flipping and none have since been flipped. It was already apparent that many water

Table 2. *Details of refinement*

Model	*R* factor 2·8 Å	2·6 Å	2·5 Å	Number of averaging cycles*	Number of RSR cycles†	R.m.s. shift from preceding model (Å)
*M*1	0·440	0·459	0·475	4		
*M*1*VG*					3	1·13
*M*2	0·332	0·356		1	2	0·59
*M*3		0·335	0·349	3	2	0·52
*M*4			0·300	2	2	0·46
*M*5			0·294	2	2	0·19
*M*6	0·243	0·261	0·271			0·17

* Number of averaging cycles done starting with the existing model before making a new model.
† Number of RSR cycles done in the averaged map before defining a new model.

molecules could be clearly seen in the maps, but we did not add any until the next cycle when we extended the resolution to include all of our measurements to 2·5 Å. Because of limitations in our computer disk capacity we were forced to keep the same grid sizes. In *M*3 we had 70 waters, in *M*4 104, in *M*5 180 and in *M*6 a total of 158 in the n.c.s. unit. The changes made to the protein in these display sessions were minor, and most of the time was spent identifying new water molecules.

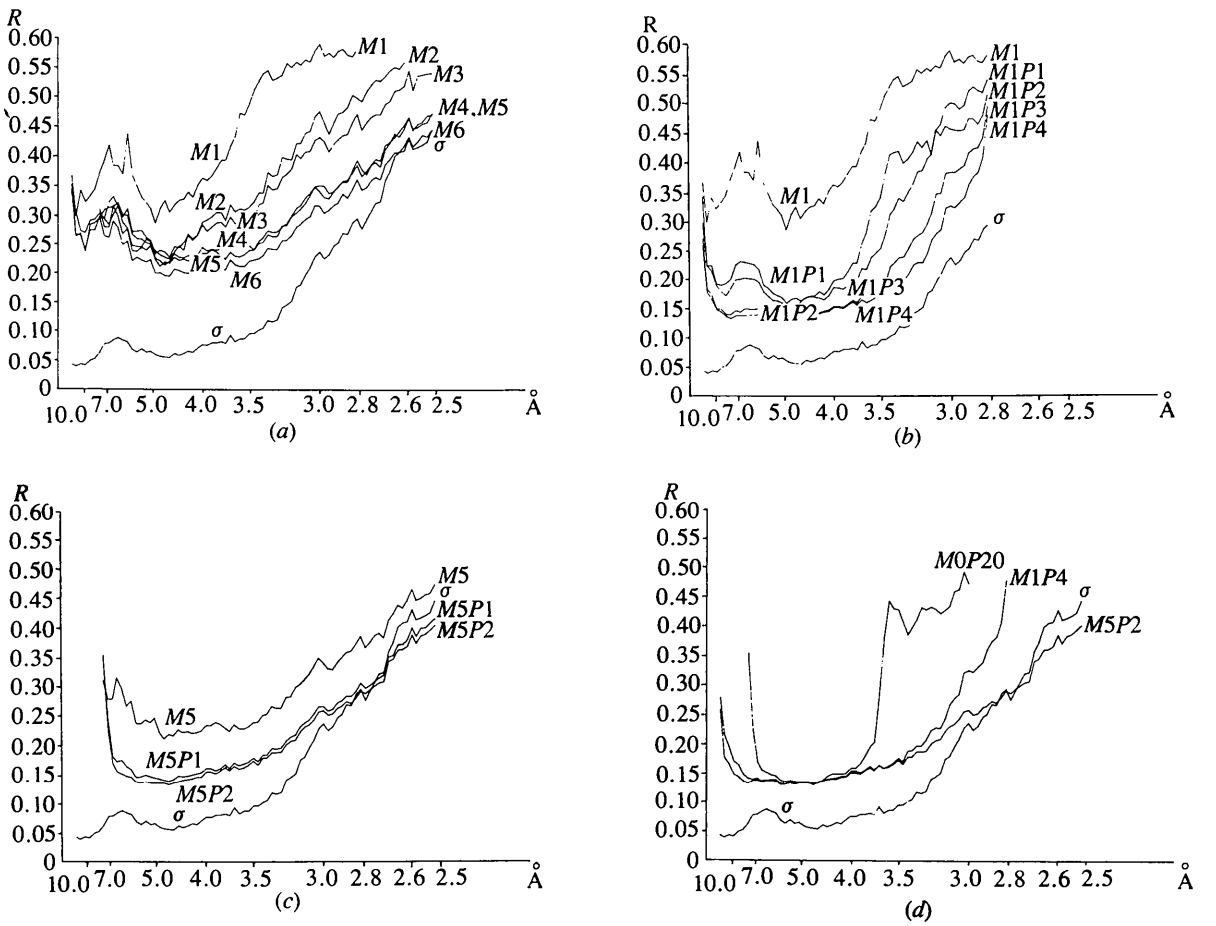

Fig. 3. *R* factors as a function of resolution for various models and electron density maps. (*a*) The improvement in *R* factor for each model. The curve marked σ shows the σ-factor variation (see text for definition) to show the estimated errors in our observed structure-factor amplitudes. (*b*) The improvements made by cyclic phase refinement at the most critical time in the refinement. Four cycles were sufficient to improve the quality of the map to allow us to build an essentially correct model. (*c*) The course of the final phase-refinement cycles. The convergence is rapid and shows that model and averaged map have the same variation, gradually increasing with resolution. The phase refinement of *M*5 was done between 7·5 and 2·5 Å resolution. This is the cause of the sharp increase near 7 Å resolution for *M*5*P*2. (*d*) A plot showing the step nature of our starting *R* factor and how improving the model and extending the resolution finally results in a residual which has no discontinuity, and shows significant improvement in the range 3·5–2·5 Å.

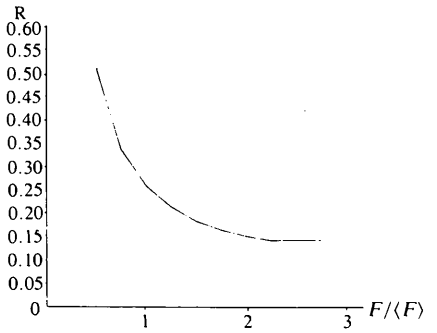

Fig. 4. Variation of R factor with observed structure-factor amplitudes. We have chosen to plot the abscissa as the ratio of the amplitude with the mean value at that resolution.

The curves in Fig. 3(a) show a steady improvement in R factor in going from model to model up to $M5$, by which time the positional refinement had stopped. We made clear gains between $M3$ and $M4$ by using a finer 0·7 Å grid in the *FRODO* and RSR maps rather than the original 1·0 Å. Such effects had already been pointed out by Diamond (1971). By model $M5$, the phase refinement process converged rapidly in one cycle, Fig. 3(c). At this stage, we introduced group temperature-factor refinement. All atoms within each rigid group were constrained to have the same atomic radius, which was allowed to vary to minimize the grid sum difference

$$\sum (\rho_{\text{obs}} - \rho_{\text{model}})^2.$$

Deisenhofer & Steigemann (1975) in their high-resolution real-space refinement of pancreatic trypsin inhibitor assumed that the atomic form factor could be described by a Gaussian function. The transformation from atomic radius to temperature factor is then of the form

$$B = 4\pi[(ca)^2 - a_0^2],$$

where a_0 is the zero-temperature radius and c is a constant chosen to remove systematic differences between F_o and F_c. In their refinement they assigned a_0 values to each atomic type and a single value for c. We have used single values for both constants, $c = 0{\cdot}93$, $a_0 = 0{\cdot}90$ Å. The new model $M6$ shows a marked improvement in R factor over $M5$. The variation of the $M6$ R factor as a function of the ratio of F_o to mean F_o in its resolution range is shown in Fig. 4.

The improvements made between the first and the final electron density maps are shown as R factors in Fig. 3(d) and for a few residues in Fig. 5. For clarity

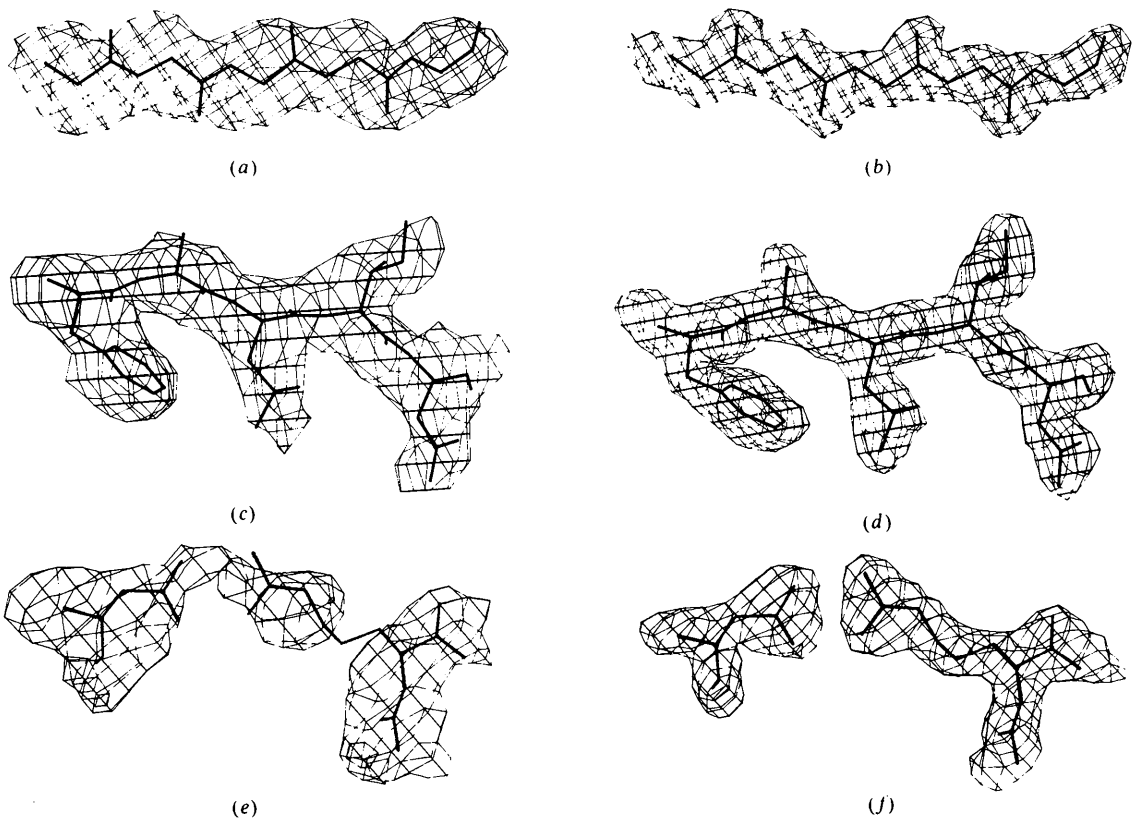

Fig. 5. Improvement of electron density maps as a result of the refinement. (a) and (b) The improvement in density for carbonyl oxygens, and (c) and (d) for the equivalent side chains (Phe, Ala, Leu, Ile, Asn). (e) and (f) The improvements for one of the subunit–subunit salt bridges between Asp 92 and Arg 96 of a twofold-related molecule.

we have used envelope contouring around just these residues. We have superimposed the final model *M6* on maps *M0P20* and *M5P2*. Figs. 5(*a*)–(*d*) are of the β strand Phe29 to Asn33. In Figs. 5(*a*), (*b*) only the main-chain atoms are drawn to demonstrate the improvement in the quality of the density for carbonyl oxygens, and in Figs. 5(*c*), (*d*) for the side chains. Figs. 5(*e*), (*f*) show the salt bridge between two two-fold-related molecules at residues Asp92 and Arg96. Although no distance restraints have been applied to this interaction the H bonds turn out to be 2·76 and 3·00 Å.

Discussion

We found it surprisingly easy to refine STNV. The programs described above were being continuously improved during the refinement so that we could do it today with less cycles. The computationally expensive phase averaging cycles are straight-forward once everything has been set up. We used up 12 cycles of phase refinement and could have made do with 8–10 if we had extended the resolution to 2·5 Å with phases from model *M2*. Four cycles of phase refinement would then have produced phases comparable in quality to *M5P2*. We suspect that the method we have described is as computationally efficient as the alternatives listed above.

We are not clear why the initial phase refinement produced nonsense between 3·7 and 3·0 Å and we do not know if better initial phases would have brought us over the barrier. This is the resolution range where side-chain groups normally start appearing in maps but where main-chain carbonyl groups are still difficult to observe. Fortunately, main-chain carbonyl groups were usually at least logically correctly placed in the first map because the STNV subunit contains a large amount of β-sheet structure and some helix. This effect may have been an unexpected bonus in getting the improved map *M1P4*.

We have not changed our n.c.s. parameters during the refinement. The quality of our final map suggests that in the N-terminal region residues 1–11 do not obey the icosahedral symmetry and that from 12–22 it improves until by the top of the helix at residue 24 it is essentially exact.

The error in our final model is difficult to estimate. The method of Luzzati (1952) assumes that the reflection data is error free and this is clearly not true in our case. The quality of our reflection data is worse than the equivalent data for a protein containing only 195 amino acids. Of our 500 000 unique reflections to 2·5 Å resolution, 245 000 are single measurements. The σ-factor curve in Fig. 3 shows that the data gets worse at high resolution as we would expect because of the large number of weak reflections. The appearance of the curve is very similar to the later model and density *R* factors and suggests at the very least that a large part of the discrepancy in the *R* factor is due to experimental errors.

We would like to thank Professor Bror Strandberg for encouraging us during this work, the Swedish Natural Science Research Council for its continual support and the Wallenberg Foundation for buying us a computer graphics system.

References

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1980). *Nature (London)*, **286**, 33–39.

ABAD-ZAPATERO, C., ABDEL-MEGUID, S. S., JOHNSON, J. E., LESLIE, A. G. W., RAYMENT, I., ROSSMANN, M. G., SUCK, D. & TSUKIHARA, T. (1981). *Acta Cryst.* B**37**, 2002–2018.

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

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

DEISENHOFER, J. (1981). *Biochemistry*, **20**, 2361–2370.

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

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

DIAMOND, R. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 318–325. Oxford Univ. Press.

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

HARRISON, S. C., OLSON, A. J., SCHUTT, C. E., WINKLER, F. K. & BRICOGNE, G. (1978). *Nature (London)*, **276**, 368–373.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01–13.25. Bangalore: Indian Academy of Sciences.

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

JACK, A. & LEVITT, M. (1978). *Acta Cryst.* A**34**, 931–935.

JONES, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.

JONES, T. A. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 303–317. Oxford Univ. Press.

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

LEVITT, M. (1976). *J. Mol. Biol.* **104**, 59–107.

LILJAS, L., UNGE, T., JONES, T. A., FRIDBORG, K., LÖVGREN, S., SKOGLUND, U. & STRANDBERG, B. (1982). *J. Mol. Biol.* **159**, 93–108.

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

REMINGTON, S., WIEGAND, G. & HUBER, R. (1982). *J. Mol. Biol.* **158**, 111–152.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800–804.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.

TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486–492.

WALSH, G. R. (1975). *Methods of Optimization.* New York: Wiley.

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