Table 6. *Coefficients a_{Ni} in the polynomial for n_{Nk}, equation (2), for primitive hypercubic lattices (Z^N)*

N	a_{N0}	a_{N1}	a_{N2}	a_{N3}	a_{N4}	a_{N5}	a_{N6}	a_{N7}	a_{N8}	a_{N9}
1	2									
2	0	4								
3	2	0	4							
4	0	16/3	0	8/3						
5	2	0	20/3	0	4/3					
6	0	92/15	0	16/5	0	8/15				
7	2	0	392/45	0	28/9	0	8/45			
8	0	704/205	0	352/45	0	64/45	0	16/315		
9	2	0	3272/315	0	76/15	0	8/15	0	4/315	
10	0	2252/315	0	5744/567	0	2408/945	0	160/945	0	4/1575

computer difficult for more-complicated higher-dimensional structures.

References

BRUNNER, G. O. (1979). *J. Solid State Chem.* 29, 41–45.
CONWAY, J. H. & SLOANE, N. J. A. (1988). *Sphere Packings, Lattices and Groups.* New York: Springer-Verlag.
FISCHER, W. & KOCH, E. (1985). In *International Tables for Crystallography*, Vol. A, edited by T. HAHN. Dordrecht: Kluwer.
GOETZKE, K. & KLEIN, H.-J. (1991). *J. Non Cryst. Solids*, 127, 215–220.
JANSSEN, T. (1986). *Acta Cryst.* A42, 261–271.
MARIANS, C. S. & HOBBS, L. W. (1990). *J. Non Cryst. Solids*, 124, 242–253.
MEIER, W. M. & MOECK, H. J. (1979). *J. Solid State Chem.* 27, 349–355.
O'KEEFFE, M. (1991a). *Z. Kristallogr.* In the press.
O'KEEFFE, M. (1991b). *Chem. Mater.* 3, 332–335.
O'KEEFFE, M. & HYDE, B. G. (1980). *Philos. Trans. R. Soc. London Ser A*, 295, 553–623.
STIXRUDE, L. & BUKOWINSKI, M. S. T. (1990). *Am. Mineral.* 75, 1159–1169.

Methodology Employed for the Structure Determination of Tumour Necrosis Factor, a Case of High Non-Crystallographic Symmetry

By E. Y. JONES

Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

N. P. C. WALKER

BASF AG, Ludwigshafen, Germany

AND D. I. STUART

Laboratory of Molecular Biophysics, The Rex Richards Building, South Parks Road, Oxford OX1 3QU, England

Abstract

The structure of the protein tumour necrosis factor (TNF) was determined from crystals of space group $P3_121$ which contain six copies of the TNF monomer per crystallographic asymmetric unit [Jones, Stuart & Walker (1989). *Nature* (*London*), 338, 225–228]. The nature of these crystals (relatively high crystallographic symmetry coupled with multiple copies of the protein in the asymmetric unit) led to some peculiarly challenging problems at several points in the structure determination. In particular, (1) self-rotation function calculations failed to yield clearly interpretable solutions, (2) the analysis of difference Patterson maps for heavy-atom derivatives required the development of a Patterson search program suite GROPAT. The redundancy in the asymmetric unit allowed refinement of poor-quality isomorphous phases at 4 Å resolution and phase extension from 4 to 2·9 Å resolution using real-space symmetry

averaging and solvent flattening in the absence of any isomorphous phase information. Despite further difficulties caused by structural differences between the six independent copies of the monomer the resultant electron density map was of high quality and proved to be easily interpretable.

Introduction

Tumour necrosis factor (TNF) is an important polypeptide mediator of inflammation and the cellular immune response (Old, 1987; Beutler & Cerami, 1988). Mature human TNF is an unglycosylated protein of 157 amino acids (relative mass 17 350 Da). In solution three such TNF subunits are tightly associated to form the biologically active trimer.

Rod-like crystals ($0\cdot5 \times 0\cdot5 \times 1\cdot5$ mm) of recombinant human TNF were grown from ammonium sulfate solution by vapour diffusion (Walker, Marcinowski, Hillen, Machtle, Jones & Stuart, 1990). The crystals belong to the trigonal space group $P3_121$ with unit-cell dimensions $a = b = 165\cdot9$, $c = 93\cdot1$ Å. The crystallographic asymmetric unit contains two independent TNF trimers (*i.e.* six subunits) and the fractional volume of solvent in the crystal is 65%. The crystals diffract to a resolution limit of $2\cdot9$ Å.

The course of the structure determination was dictated by the presence of six copies of the TNF subunit in the crystallographic asymmetric unit. The orientation of these six subunits was unknown, which rendered the solution of heavy-atom derivatives problematic. Thus the search for heavy-atom derivatives was guided by a need to minimize the number of heavy-atom sites to facilitate the interpretation of difference Patterson maps. Various methods were employed to assess putative heavy-atom derivative data and an automatic Patterson search program, *GROPAT* (Jones & Stuart, in preparation), was developed. Conversely, the oversampling of the molecular transform allowed the initial poor-quality heavy-atom-derivative phases to be refined, and phase extension from 4 to $2\cdot9$ Å was achieved solely by solvent flattening and averaging techniques.

The resultant structure of TNF at $2\cdot9$ Å resolution has been reported (Jones, Stuart & Walker, 1989) and the refined coordinates have been deposited with the Brookhaven Protein Data Bank. The individual TNF molecule has the classic topology of the eight-stranded flattened β-barrel, a motif previously observed in the viral coat proteins of small RNA plant viruses and picornaviruses. The subunit β-barrels associate with their β-strands running parallel to the threefold axis to form the trimer. The model has been refined using the program *XPLOR* (Brunger, Kuriyan & Karplus, 1987) to a current crystallographic R factor of 20·4% (on all data from 6 to $2\cdot9$ Å with no model for the solvent structure). On average the B factors are high as expected from the rapid fall off

in diffraction intensity with resolution. Both trimers exhibit significant deviations from strict threefold symmetry.

This paper draws on novel aspects of the TNF structure determination. Points addressed are (*a*) the failure of self-rotation function calculations to yield the non-crystallographic threefold orientations of the two trimers, (*b*) the development of protocols for analysis of heavy-atom derivatives, (*c*) the refinement of poor initial single isomorphous replacement (SIR) phases, (*d*) the use of real-space symmetry averaging and solvent flattening in phase extension from 4 to $2\cdot9$ Å.

Data collection and quality

Initial data collection on CEA X-ray film used synchrotron radiation (wavelength $0\cdot88$ Å limited by a $0\cdot3$ mm collimator) at station 9.6 of the SERC Synchrotron Radiation Source (Daresbury, England). Native data to an effective resolution of $2\cdot9$ Å were collected at 293 K as 1° oscillation photographs with a crystal-to-film distance of $17\cdot0$ cm for the first data set and $18\cdot5$ cm for a subsequent set. A rapid survey of ten putative heavy-atom derivatives was carried out at lower resolution ($3\cdot4$ Å) using a 2° oscillation range with a crystal-to-film distance of 22 cm. The crystals were rotated about the c/c^* axis except during collection of the native blind region for which c/c^* was perpendicular to the rotation axis. Translation along the c/c^* axis to two or three fresh positions resulted in only one crystal being required for each data set (taken as a total rotation of 30° about c/c^*). Typical exposures of 100 s for a 2° oscillation allowed collection of a relatively complete $3\cdot4$ Å data set in 40 min. The films were digitized using a Joyce–Loebl Scandig microdensitometer on a 50 μm raster and processed with standard in-house programs (Stuart, Levine, Muirhead & Stammers, 1979).

A Xentronics area detector on a Rigaku RU200H rotating-anode X-ray generator (apparent source dimensions $0\cdot3 \times 0\cdot3$ mm, operated at 50 kV, 60 mA with a graphite monochromator, 1 mm primary and $0\cdot5$ mm secondary collimation) afforded the ideal means for screening a further extensive range of potential heavy-atom derivatives inhouse at Oxford. Eleven data sets were collected at 298 K to $5\cdot0$ Å resolution (detector swing angle 0°) or $3\cdot5$ Å resolution (detector swing angle 10°) with a crystal-to-detector distance of 18 cm, an oscillation frame width of $0\cdot2$° and an exposure time per frame of 200 s. The crystals were rotated over a 60° range about an axis at 45° to c/c^*; a further 60° of data were collected after rotating by 90° on φ (axis of rotation parallel to c/c^*). The data were processed using the *XENGEN* program package (A. Howard, *XENGEN* manual). Complete data sets were collected and processed within 48 h with little crystal decay and minimal manual intervention.

Table 1. *Statistics for TNF data sets*

Data set	Resolution (Å)	Redundancy	Unique reflections (% complete)	R_{merge}‡ (all data)	Number of crystals
Native (high-resolution film)	2·9	1·0	24 385 (76%)	12·5*	2
Native (low-resolution film)	3·4	2·3	16 031 (78%)	7·5	2
Native (low-resolution Xen)	5·0	4·5	5 582 (85%)	5·6	1
Merged native 1	2·9	3·3	26 662 (83%)	8·3	–
Native 2 (high-resolution film)	2·9	2·0	22 934 (71%)	8·4*	1
Merged native 2	2·9	4·5	29 725 (93%)	7·2	–
$K_2Pt(CN)_4$ (low-resolution film)	3·4	1·0	7 536 (37%)	7·0	1
$K_2Pt(CN)_4$ (low-resolution Xen)	5·0	4·2	6 314 (96%)	7·7	1
Merged $K_2Pt(CN)_4$†	3·4	3·5	11 944 (58%)	8·0	–
K_2HgI_4 (low-resolution Xen)	5·0	4·1	5 977 (91%)	9·5	1
$Hg(Ac)_2$ (low-resolution Xen)	5·0	4·3	5 888 (90%)	7·7	1
$Hg(Ac)_2$ (high-resolution Xen)	3·5	3·3	14 513 (70%)	9·9	1
Merged $Hg(Ac)_2$	3·5	4·9	14 779 (80%)	7·3	–

* Including partials.

† 10% contribution from Enraf–Nonius FAST data.

‡ $R_{merge} = \sum_j\sum_h|I_{h;j} - \langle I_h\rangle|/\sum_j\sum_h\langle I_h\rangle$.

The statistics for some typical TNF data sets are shown in Table 1. No attempt was made to measure the weak data beyond 3·5 Å on the area detector; however, to this resolution Daresbury and area detector data sets could be merged satisfactorily revealing no systematic biases (Table 1 and data not shown). The completeness *versus* resolution of the native data set (native 1 in Table 1) used for the initial structure determination is illustrated in Fig. 1. Further high-resolution data were processed and included during the structure refinement to give a current native data set (native 2 in Table 1) which remains over 90% complete in the 3·5 to 3·1 Å range. The higher-resolution data are relatively very weak so that to record measurable data to 2·9 Å resolution much increased exposure times were required (150 s for a 1° oscillation at the Daresbury SRS station 9.6, more than double the exposure time used for the lower-resolution data sets). Inspection of the linear region (4 to 2·9 Å) of a Wilson plot for the data indicates that the effective mean B exceeds 80 Å².

Self-rotation function and native Patterson

Self-rotation functions were calculated using the program *POLARRFN* [based on code by W. Kabsch working in spherical polar angles (Tanaka, 1977)];
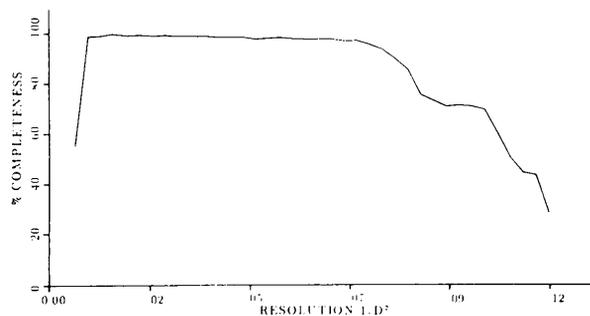
Fig. 1. Completeness of TNF native data set as used for the initial structure determination.

some duplicate calculations were performed as checks using *ALMN* (Dodson, 1985) and, in retrospect, *MERLOT* (Fitzgerald, 1988). All three programs are based on the fast-rotation-function algorithm of Crowther (1972). A rotation may be described in terms of the angles ω and φ which respectively define the latitude (taken as angle from the pole) and longitude (angle around the equator) of the rotation axis and κ which is the angle of rotation about this axis. The significance of any peaks in the self-rotation function was gauged by reference to the behaviour of a parallel series of rotation-function calculations performed on randomized structure-factor amplitudes. These control data sets were generated from real TNF data sets by randomizing the match of the amplitudes to the indices *hkl*. Since the crystallographic asymmetric unit was believed to contain two TNF trimers, it was expected that two non-crystallographic symmetry axes would be apparent on the $\kappa = 120°$ section of the self-rotation function. However, despite many careful calculations using different resolution ranges, various integration radii, independent data sets, sharpened data and Patterson origin removal, no convincing peaks were observed on the $\kappa = 120°$ section.

For the trigonal space group $P3_121$ there is a peak on the $\kappa = 120°$ section (Fig. 2a) corresponding to the crystallographic threefold screw axis which could potentially obscure any non-crystallographic threefold rotation axes. However, the alignment of a non-crystallographic threefold axis with a crystallographic threefold screw axis preserves the relative orientation of molecules at $z = 0$, $\frac{1}{3}$ and $\frac{2}{3}$ (Fig. 3). Thus a strong peak should occur in the native Patterson map on the $\frac{1}{3}$ and $\frac{2}{3}w$ Harker sections. No such feature is visible in the TNF native Patterson map. Indeed, the native Patterson map calculated over a range of resolution shells (including some limited to very low resolution) lacked any dominant features. Since it was expected that an angular divergence between crystallographic and non-crystallographic threefold axes sufficient to

(a)   (b)

(c)   (d)

(e)   (f)
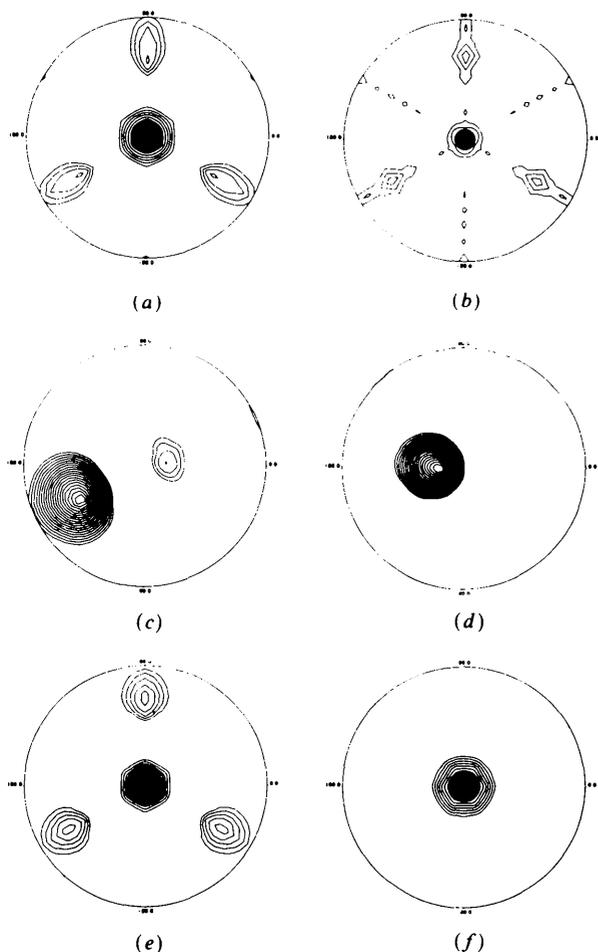
Fig. 2. $\kappa = 120°$ sections for the self-rotation function calculated using *POLARRFN* for all data to 3·7 Å with an integration radius of 21 Å. In these projections ω varies radially and φ varies circumferentially. The orientation of the crystallographic axes is such that **a** is horizontal and **c** is normal to the paper. (a) Native TNF data. (b) Randomized data. (c) Model data for trimer 1 in $P1$ cell. (d) Model data for trimer 2 in $P1$ cell. (e) Model data for trimer 1 in $P3_121$ cell. (f) Model data for trimer 2 in $P3_121$ cell.
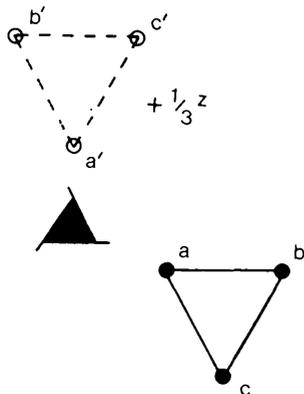
Fig. 3. For a trimer of identical subunits a, b, c with threefold axis aligned with the crystallographic threefold screw axis the Harker vectors **ab′ bc′ ca′** are all equivalent.

eliminate any peaks on the native Patterson map would resolve the crystallographic and non-crystallographic peaks on the self-rotation function this explanation for the lack of non-crystallographic peaks appeared to be ruled out.

The only other feature present on the $\kappa = 120°$ section for TNF data also appears in the control self-rotation functions calculated using randomized data sets (Fig. 2a, b). This peak lies on and is smeared along $\varphi = 90, 210$ and $330°$; the directions corresponding to the Harker section arising from the crystallographic twofold axis. Its presence in self-rotation calculations for randomized data appears to arise from the accumulation of multiply weighted noise peaks along Harker directions; thus this peak in the native TNF self-rotation was viewed with particular caution and no firm evidence for the position of the two non-crystallographic threefold axes appeared to be available.

In order to investigate the reasons for the failure to detect the non-crystallographic threefold axes by the use of self-rotation functions, structure factors have been calculated from the refined TNF model (F_{calc} s) and these have been used to perform a series of model calculations. The rotation functions were calculated for a variety of resolution ranges but showed little.variation; a representative series using model data to 3·7 Å is shown in Fig. 2. Figs. 2(c) and (d) illustrate the $\kappa = 120°$ sections based on F_{calc} s for each trimer independently isolated in a $P1$ cell. The peak in both cases is extremely blurred and extends over κ sections from 60 to 180°. The nature of the TNF molecule, which consists entirely of β-strands approximately aligned with and arrayed around the non-crystallographic threefold axis, does not lead to a clean solution. On repeating the calculations for each trimer in the correct high-symmetry space group, $P3_121$, the results are in agreement with those obtained using the experimentally observed structure-factor amplitudes (Figs. 2e, f). The low peak corresponding to one of the trimers lies on the Harker section. The second peak, although 20° in ω from the crystallographic threefold axis, is still sufficiently close, because of its width, that it blurs completely into the wings of the crystallographic peak merely contributing a slight broadening effect. For the actual situation (two trimers in space group $P3_121$) the occurrence of two diffuse non-crystallographic peaks multiplied up to nine peaks by the crystal symmetry plus the presence of the crystallographic threefold axis on the $\kappa = 120°$ section simply results, even using error-free synthetic data, in a high background with only slight traces of the peaks due to the non-crystallographic threefold axes.

In summary, the structure of the TNF trimer, consisting as it does of no α-helices and a somewhat curved β-sandwich aligned approximately along the molecular threefold axis, leads to a vector map domi-

Table 2. *Putative heavy-atom derivatives*

Derivative	R_{merge}‡ (all data)	Average isomorphous difference (∞ to 5 Å)	Correlation* 3σ 10-7 vs 7-5 Å	Correlation† 3σ 10-7 Å	Useful derivative
$K_2Pt(CN)_4$	8·0 (3·4 Å)	16·1	11·8%	43·6%	Yes
WAC	5·4 (5·0 Å)	11·8	3·9%	–	No
K_2HgI_4	9·5 (5 Å)	27·5	27·0%	–	Yes
K_2PtCl_4	9·7 (3·4 Å)	11·2	12·1%	–	No
$KAu(CN)_2$	10·3 (3·4 Å)	18·5	9·7%	–	No
$Hg(Ac)_2$	7·3 (3·5 Å)	19·1	31·9%	99·7%	Yes
Noise	–	–	4·8%	0·0%	–
Perfect	–	–	83·0%	–	–

* Correlation between Patterson maps based on 10-7 Å and 7-5 Å data.
† Correlation between Patterson maps based on independent 10-7 Å data sets.
‡ R_{merge} as defined for Table 1.

nated by a large number of essentially equal-weight vectors aligned in roughly similar directions. When combined with the smearing implicit in the Patterson convolution this situation gives rise to a diffuse 'molecular Patterson'; taken together with the complexity of the space group and the somewhat special directions of the threefold axes this results in a very unfavourable case for analysis using the self-rotation function. The structure determination therefore had to proceed without knowledge of the orientation of the two non-crystallographic threefold axes.

Analysis of isomorphous heavy-atom derivatives

The human TNF molecule contains one disulfide bond, no free cysteine residues or methionines and three histidines per 157 amino acid subunit. A range of standard heavy atoms was tested. Native and derivative data were scaled using an established seven-parameter anisotropic-B-factor method (Stuart, Levine, Muirhead & Stammers, 1979) and a number showed encouraging isomorphous differences. However, with six molecules per asymmetric unit in a $P3_121$ cell even a single heavy-atom binding site per subunit results in over 1000 peaks at close to noise level in the difference Patterson map. In an attempt to render the problem more tractable, heavy-atom compounds with internal threefold symmetry were tested in the hope of binding only two positions per asymmetric unit. These compounds included K_2HgI_4 in excess KI (Petsko, 1985), tetrakis(acetoxymercurio)methane (abbreviated to TAMM) (O'Halleran, Lippard, Richmond & Klug, 1987) and $[\{W_3O_2(O_2CCH_3)_6\}(H_2O)_3](CF_3SO_3)_2$ (abbreviated to WAC) (Ladenstein, Bacher & Huber, 1987). One of these, K_2HgI_4, resulted in a large mean fractional isomorphous difference of 26% but the difference Patterson map was not readily interpretable; the complex pattern of low-level peaks in the map was not consistent with binding of the HgI_3^- group at a single site on the trimeric threefold axes. This derivative did in fact prove useful but the pattern of binding was

much more complex than had been hoped for (see below). A potential target for at least limiting the number of sites to one per subunit was presented by the presence of a single disulfide bond. This disulfide was reduced and reformed in the presence of mercury acetate $Hg(Ac)_2$ in accordance with the method of Ely, Girling, Schiffer, Cunningham & Edmundson (1973). The resultant data set showed a mean fractional isomorphous difference of 21%.

Visual inspection of the difference Patterson maps proved of little efficacy in assessing whether a heavy-atom compound had yielded an isomorphous derivative. For one full-occupancy heavy-atom site per subunit (*e.g.* one Hg atom), the typical mean fractional isomorphous differences may be calculated to be approximately 20% in F and values of about this magnitude were obtained for several putative derivatives (see Table 2). The stability of unit-cell dimensions and minimal variation in mean fractional differences over resolution were taken as preliminary indicators of isomorphous behaviour. To distinguish between difference Patterson maps which consisted purely of random-noise peaks and those containing actual information a simple correlation coefficient

$$\sum_u D_u^1 \times D_u^2 / \sum_u |D_u^1 \times D_u^2|$$

(where D_u^1 and D_u^2 are the values at corresponding pixels u in maps 1 and 2 and the summations are over all pixels above a given threshold) was calculated between pairs of origin-removed difference Patterson maps which had been calculated using independent data. Thus for a putative derivative correlation coefficients were calculated between Patterson maps based on 10 to 7 Å and 7 to 5 Å data (Table 2). The results varied from 32% (on pixels greater than 3σ) for the $Hg(Ac)_2$ to 4% (equivalent to the value obtained for correlating random-noise peaks) for the WAC. For a virtually noise-free Patterson map generated from test data a value of 83% was obtained on the 3σ peaks. Where independent data sets were available for the same derivative Patterson maps were

calculated over 10 to 7 Å using different data sets for the native as well as the derivative. This gave 44% agreement for $K_2Pt(CN)_4$, 99·7% for $Hg(Ac)_2$, 0% in a control calculation between artificial test data and $Hg(Ac)_2$ and, surprisingly, 31% between $K_2Pt(CN)_4$ and K_2HgI_4 (all values are for pixels greater than 3σ). The latter result implied the presence of common heavy-atom sites between $K_2Pt(CN)_4$ and K_2HgI_4. Overall the analysis highlighted $Hg(Ac)_2$ as the most promising derivative. It should be noted that great care had to be taken to use completely independent data in these calculations, the use of a common native set inflated the correlation between $K_2Pt(CN)_4$ and K_2HgI_4 to 66%.

Further analysis of the difference Patterson maps was undertaken using Patterson search techniques incorporated in a set of programs, *GROPAT*, which were developed for the purpose. Origin-removed sharpened Patterson maps were used, generally for the resolution range 10 to 6 Å. A detailed description of the programs and their performance will be reported separately (Jones & Stuart, manuscript in preparation). A grid search over the asymmetric unit first established the top 200 positions for single sites judged to generate Harker peaks in best agreement with the actual Patterson map. These sites were then refined individually on a finer grid to optimize agreement with the Patterson function. All possible pairwise combinations of these 200 sites were then assessed in terms of the agreement of their cross peaks with the actual Patterson and the resulting best 200 pairs were listed. This analysis was carried out for six putative derivatives. The $Hg(Ac)_2$ derivative yielded a clear pattern of interconnected pairwise solutions with all possible combinations of the single-site solutions originally ranked 3, 57, 65, 67 and 124 grouped at the top of the pairs list. Fourier difference maps based on phases from these five sites revealed a sixth site. Sites 3, 65 and 67 obeyed threefold symmetry forming an equilateral triangle with sides of about 11 Å. Sites 57 and 124 with the addition of the new site formed a second similar but independently orientated triangle.

The parameters describing the six sites of substitution for the $Hg(Ac)_2$ derivative were refined using the 'phase refinement' method [programs *PHASE* and *REFINE* coded by Dr P. Shaw (unpublished), based on the method of Muirhead, Cox, Mazzarella & Perutz (1967) with optional eigenvalue filtering coded by D. I. Stuart after Diamond (1971)]. The initial refinement was based on the centric reflections from 20 to 6 Å resolution with figures of merit (FM) greater than 0·5. Centroid phases were used throughout and starting values for the magnitudes of the mean isomorphous error of closure $|E|$, the occupancies and the B factors of the sites were selected on the basis of trial and error taking care that the majority of the data were initially included in the refinement process.

The refinement, with eigenvalue filtering, used $(FM)^2$ weighting. After several rounds of phasing and refinement of occupancy, position and scale factor general reflections were included at half weight. Rounds of phasing and refinement then iterated between refinement on centric reflections for occupancies and refinement on general reflections for position and scale factor. B factors were not stable when refined individually on data to 6 Å and so were simply optimized at a common value using a step search procedure. Further sites were checked for by the calculation of difference and double difference maps. Data to 4·5 Å resolution with no low-resolution cut off were then included and the minimum FM for refinement gradually reduced from 0·5 to 0·3. Anisotropic B factors could now be refined in alternate steps between refinement of heavy-atom positions and an overall B factor (describing the difference in order between the native and derivative data sets) and refinement on centrics only of the occupancies. The heavy-atom-refinement program was modified to allow the non-crystallographic threefold symmetry to be imposed on the heavy-atom sites by constraining the sites to form equal-sided triangles. Similarly, the occupancy was constrained to be equal for all six sites. These modifications greatly improved the stability of the refinement. The final mean FM on all data to 4·5 Å was 0·38.

The refined phases from the $Hg(Ac)_2$ derivative (after several cycles of solvent modification at 6 Å resolution as detailed in the next section) were used in the calculation of difference Fourier maps for the remaining putative derivatives. Of these, two, K_2HgI_4 and $K_2Pt(CN)_4$, were revealed as multiple site derivatives of rather poor quality. They had several sites in common (Table 3) as had been suspected from the Patterson map correlations. In particular, both these derivatives had occupied a site on the threefold axes of the two trimers, indeed for the K_2HgI_4 derivative these constituted the highest occupancy sites (as hoped the HgI_3^- group had bound at sites with threefold symmetry but unfortunately not exclusively). None of the sites in either derivative could be grouped into triplets consistent with the non-crystallographic symmetry. The derivatives were refined separately using data to 6 Å resolution [initially on centric reflections, then including all reflections following a similar strategy to that adopted for the $Hg(Ac)_2$ derivative] but were not of particularly good quality (Table 4). The inclusion of eigenvalue filtering was essential in the refinement of the crystallographic twofold site. The final mean FM on all data to 6·0 Å was 0·34 for both these derivatives.

In retrospect the major site in the $K_2Pt(CN)_4$ derivative, on a crystallographic twofold axis, and one of the minor sites were picked up by *GROPAT* and ranked 4 and 117 respectively in the list of best possible single sites for this derivative. For the K_2HgI_4

Table 3. *Heavy-atom sites*

Derivative	Occupancy	x	y	z	B (Å^2)	Site	Comments
Hg(Ac)$_2$	0·028	1·1402	0·5483	0·1236	95·0	$A1$	G
	0·028	1·1018	0·5559	0·0955	95·0	$A1$	G
	0·028	1·1396	0·5607	0·0364	95·0	$A1$	G
	0·028	0·7284	0·1986	0·1215	95·0	$A2$	G
	0·028	0·6752	0·1417	0·1109	95·0	$A2$	G
	0·028	0·7386	0·1418	0·1483	95·0	$A2$	–
K$_2$Pt(CN)$_4$	0·027	0·6178	0·6178	0·0000	60·0	$B1$	G
	0·025	0·1977	0·4614	0·1640	60·0	1	–
	0·019	0·1339	0·5698	0·0942	60·0	$C1$	G
	0·014	0·7324	0·1485	0·1067	60·0	$C2$	–
	0·012	0·2391	0·5449	0·1021	60·0	1	–
	0·010	0·4289	0·0125	0·1354	60·0	1/2	–
	0·007	−0·0233	0·5826	0·1670	60·0	1	–
	0·007	0·1700	0·3826	0·1956	60·0	1/2	–
K$_2$HgI$_4$	0·072	0·1333	0·5542	0·0974	150·0	$C1$	G
	0·068	0·7195	0·1460	0·1232	150·0	$C2$	G
	0·059	0·1708	0·3707	0·1336	150·0	1/2	–
	0·055	0·3856	0·0077	0·3466	150·0	$B1$	G
	0·053	0·4790	0·0525	0·0179	150·0	1	–
	0·043	0·4593	0·0091	0·2918	150·0	1/2	–
	0·042	0·8920	0·3808	0·0886	150·0	1	–

The occupancies are on a relative scale; the Hg(Ac)$_2$ sites and the major sites for the other two derivatives are essentially fully occupied. x y z are given in fractional coordinates for unit cell $a = b = 166$, $c = 93$ Å. Letters in the site column denote recurrent sites where A is disulfide group, B is crystallographic twofold axis, C is on trimer threefold axes and the number denotes trimer 1 or 2. Sites found by *GROPAT* are marked G in the comments column.

Table 4. *SIR phasing and solvent flattening* (*Wang*) *statistics*

Phasing statistics are prior to any phase modification by solvent flattening.

Derivative	FM† (all data)	FHA/ETOT‡ (all data)	R_{crys}§	FM into Wang	R factor¶ Wang start	R factor end	FM end
K$_2$Pt(CN)$_4$	0·34	1·5	0·68	0·24	0·52	0·26	0·70
K$_2$HgI$_4$	0·34	1·8	0·56	0·19	0·56	0·25	0·68
Hg(Ac)$_2$	0·40	1·7	0·54	0·33	0·48	0·24	0·74
Hg(Ac)$_2$*	0·36	1·6	0·60	0·23	–	0·20	0·76

* Statistics on all data to 4 Å (previous rows refer to data from 20 to 6 Å).
† FM figure of merit.
‡ FHA heavy-atom structure factor; ETOT residual lack of closure.
§ $R_{crys} = \sum ||F_{ph} - F_p| - F_h|/\sum |F_{ph} - F_p|$ for the centric reflections.
¶ Wang R factor; the crystallographic R factor between observed structure-factor amplitudes and those calculated from the solvent-flattened map: $R = \sum |(|F_{obs}| - |F_{calc}|)|/\sum |F_{obs}|$.

derivative *GROPAT* was successful in detecting three of the strongest sites which it ranked 1, 18 and 78. Of these 18 was identical to the twofold site ranked 4 for the K$_2$Pt(CN)$_4$ derivative. 1 and 78 were the sites on the trimer threefold axes with 1 corresponding to site 117 in the K$_2$Pt(CN)$_4$ derivative. The poor quality of the K$_2$HgI$_4$ and K$_2$Pt(CN)$_4$ derivatives, combined with their failure to conform to the non-crystallographic symmetry and the binding at crystallographically special positions, led to us failing to make a successful analysis of the difference Patterson maps. Details of the heavy-atom sites are given in Table 3. It is noteworthy that *GROPAT* found greater difficulty in locating sites associated with one of the two trimers, trimer 2, and overall these sites showed lower occupancies.

The sites of the three derivatives may be examined in the light of the refined TNF structure. In the Hg(Ac)$_2$ derivative the mercury was indeed specifically inserted into the disulfide bridge for each of the six TNF subunits within the asymmetric unit.

The binding of the heavy-atom compounds for the other two derivatives arises from nonspecific electrostatic association of the Pt(CN)$_4^{2-}$ and HgI$_3^-$ species with basic side chains, *i.e.* Arg and Lys. For these derivatives the major sites lie at interfaces between subunits (either within a trimer on the non-crystallographic threefold axis or between trimers at regions of crystal contact) where they are stabilized by multiple interactions. Of the three other putative derivatives listed in Table 2, K$_2$PtCl$_4$ had very low-occupancy binding in the vicinity of the disulfide bridges, WAC had very low-occupancy binding at the major twofold site of the K$_2$Pt(CN)$_4$ derivative and KAu(CN)$_2$ showed a mass of non-specific binding to main-chain atoms in the β-sheets.

Solvent flattening

The solvent-flattening procedure of Wang (1985) (with modifications as described in the Appendix) was applied to SIR maps based on each of the three

heavy-atom derivatives. A schematic of the work carried out is presented in Fig. 4 with resultant statistics in Table 4. Several general principles were adhered to. After each outer cycle of envelope calculation sufficient cycles of solvent flattening were performed to achieve convergence as judged by the mean phase change for successive cycles. The outer cycles of envelope recalculation were continued until the envelopes showed no further change. Care was taken to ensure that the truncation of negative regions within the protein was limited to only 2 to 3% of the total protein volume since more significant modification had a deleterious effect on the overall quality of the map, in agreement with the experience

6 Å RESOLUTION PHASING

DERIVATIVE 1	DERIVATIVE 2	DERIVATIVE 3
ENVELOPE CALCULATION (60% solvent)	DERIVATIVE 1 ENVELOPE (60% solvent)	DERIVATIVE 1 ENVELOPE (60% solvent)
7 CYCLES	8 CYCLES	8 CYCLES
ENVELOPE CALCULATION	ENVELOPE CALCULATION	ENVELOPE CALCULATION
7 CYCLES	6 CYCLES	6 CYCLES
	Addition of unmatched native data	
8 CYCLES	8 CYCLES	8 CYCLES
ENVELOPE CALCULATION (65% solvent)	ENVELOPE CALCULATION (65% solvent)	ENVELOPE CALCULATION (65% solvent)
6 CYCLES	6 CYCLES	6 CYCLES

ADD MAPS

ENVELOPE CALCULATION (65% solvent)

6 Å phased derivative 1 data plus unphased derivative 1 6–4 Å data

ENVELOPE CALCULATION (64% solvent)

5 CYCLES at 6 Å

10 CYCLES at 5·5 Å

10 CYCLES at 5·0 Å

ENVELOPE CALCULATION (64% solvent)

2 CYCLES at 5·0 Å

6 CYCLES at 4·8 Å

6 CYCLES at 4·6 Å

6 CYCLES at 4·4 Å

ENVELOPE CALCULATION (64% solvent)

2 CYCLES at 4·4 Å

6 CYCLES at 4·3 Å

6 CYCLES at 4·2 Å

6 CYCLES at 4·1 Å

6 CYCLES at 4·0 Å

ENVELOPE CALCULATION (64% solvent)

10 CYCLES at 4·0 Å

PHASE EXTENSION

Fig. 4. Schematic of the strategy followed for solvent flattening. Derivative 1 = Hg(Ac)$_2$; derivative 2 = K$_2$HgI$_4$; derivative 3 = K$_2$Pt(CN)$_4$.

of Holden, Rypniewski, Law & Rayment (1987). This is contrary to the findings of Leslie (1989).

For the $Hg(Ac)_2$ derivative at 6 Å resolution the protein envelope was calculated for a crystal solvent content of 60% (two passes using a radius of averaging of 5·5 Å). This deliberately underestimated the true value by some 5% to allow room for error in the initial attempts at envelope determination. The first envelope was rather ragged but after a number of cycles of solvent flattening and three recalculations the protein envelope had become very stable with a smooth shape and excellent connectivity. During this procedure the crystallographic R factor between observed structure-factor amplitudes and those calculated from the solvent-flattened map fell from an initial value of 0·48 to a final value of 0·24 and the mean figure of merit (as defined in Table 4) rose from 0·33 to 0·74. The initial 6 Å SIR map had shown few features other than a generally raised electron density in the regions which were duly delineated as protein by the envelope. The crystallographic asymmetric unit of the final map contained two independently orientated conical shapes in which elements of internal threefold symmetry were readily apparent. The triangular clusters of mercury sites lay at the solvent/protein interface on the 'noses' of both these cones with the heavy-atom and protein non-crystallographic threefold axes approximately coincident. As described in the previous section phases from the final stage in the map-modification procedure led to the solution of two further derivatives from difference Fourier maps.

The Wang solvent-flattening procedure was now separately applied to the two new derivatives [K_2HgI_4 and $K_2Pt(CN)_4$] taking as the initial protein envelope the current $Hg(Ac)_2$ envelope. Only after exhaustive cycles of solvent flattening were the envelopes recalculated on the basis of the appropriate derivative data. However, the envelopes for the three derivatives remained essentially identical. At this stage an encouraging measure of the agreement between the three SIR maps was given by correlation coefficients (as defined earlier, taken between pairs of maps); these were consistently 72% when calculated on pixels greater than 2σ and 58 to 63% over all pixels. Unmatched native reflections (for which no isomorphous-replacement phase information was available) were now included and further solvent-flattening cycles carried out with 65% solvent assumed for all three derivatives. This resulted in markedly improved correlation coefficients between the maps of 78 to 79% on 2σ density and 64 to 68% over all map pixels. The maps were viewed on a PS300 Evans and Sutherland graphics system using *FRODO* (Jones, 1985). Two separately orientated protein regions exhibiting non-crystallographic threefold symmetry were apparent as was a generally slab-like structure indicative of β-sheets (Fig. 5). The extent of the adherence

(a)

(b)

(c)
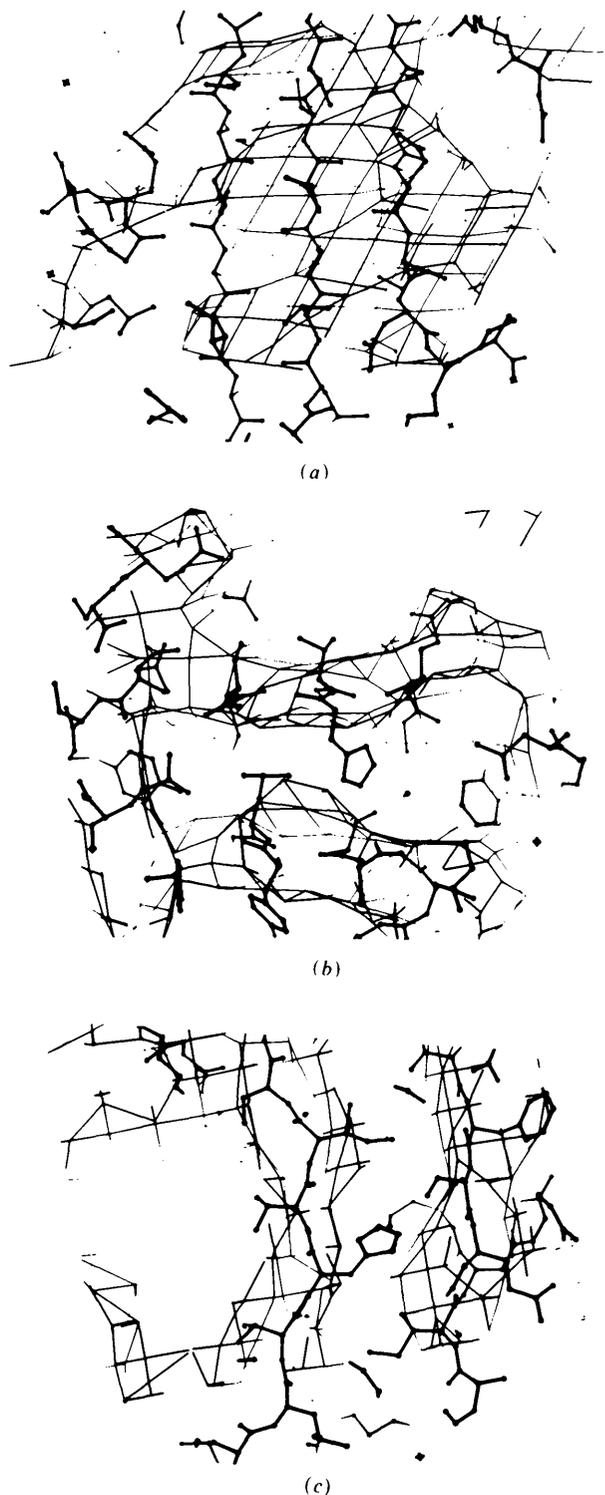
Fig. 5. Portion of the electron density map at 6 A resolution based on SIR phases from the $Hg(Ac)_2$ derivative after solvent flattening. The region shown is centred on histidine 78 in subunit 1 trimer 1; coordinates for the current refined model are superimposed on the density. Three orthogonal views are given (a) into the β-sheet; (b) down the β-strands of the β-sandwich; (c) edge on to the β-sheets. Note that the orientations and volumes shown are identical in Figs. 5, 6, 10 and 11. The contour level is 1·5σ.

(a)

(b)

(c)
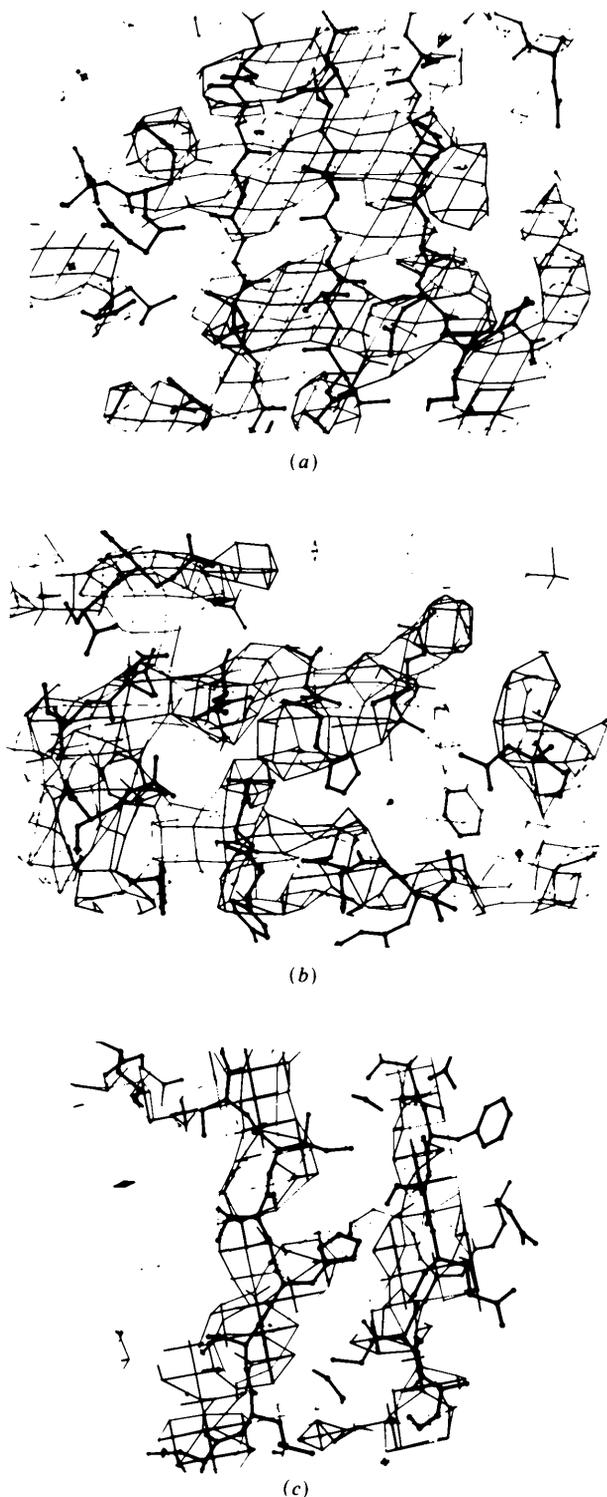
Fig. 6. Portion of the electron density map at 4 Å resolution based on SIR phases from the Hg(Ac)$_2$ derivative after solvent flattening. The region shown is centred on histidine 78 in subunit 1 trimer 1; coordinates for the current refined model are superimposed on the density. Three orthogonal views are given (a) into the β-sheet; (b) down the β-strands of the β-sandwich; (c) edge on to the β-sheets. The contour level is 1·5σ.

to threefold symmetry provided an additional criterion on which to judge the goodness of the map. The quality of the Hg(Ac)$_2$ derivative map was clearly better than that of the other two. Thus a composite map was calculated by addition of the three maps with double weight attached to that for the Hg(Ac)$_2$ derivative. This map-addition procedure is not exactly equivalent to any of the more usual reciprocal-space phase-combination methods but worked well and avoided the difficulty of weighting the different derivatives in reciprocal space by providing a simple objective real-space alternative. The protein envelope recalculated on the basis of this composite map was applied to modify the Hg(Ac)$_2$ phases. Using this as a starting point the resolution of the native data phased on the basis of the Hg(Ac)$_2$ derivative was gradually extended to 4 Å. The K$_2$HgI$_4$ and K$_2$Pt(CN)$_4$ derivatives yielded minimal phasing power beyond 6 Å (nominally still above 1 but with figures of merit below 0·3 and essentially random values for the

$$R_{\mathrm{crys}} = \sum \| F_{ph} - F_p | - F_h | / \sum | F_{ph} - F_p |$$

for centric reflections in the resolution range 6 to 5 Å) and thus were abandoned at this point.

Particular caution was exercised in the extension to 4 Å resolution because of the relatively poor quality of even the Hg(Ac)$_2$ derivative, the best of the available derivatives. A minor modification in the program suite allowed this gradual extension of the resolution to be carried out as a single submitted job (CPU time of 11·5 h on a MicroVAX3000). For simplicity a map sectioning of 1·5 Å, sufficient for the final 4 Å resolution, was used throughout. Calculated phases were combined with the original SIR phase probabilities using the weighting scheme

$$\exp - [| | F_{\mathrm{calc}} | - | F_{\mathrm{obs}} | | / | F_{\mathrm{obs}} |]$$

(Rayment, 1983) and the solvent volume was taken as 64%. The final 4 Å map showed that the slabs of density seen at 6 Å were now beginning to be resolved into separate β-strands (Fig. 6). However, it was apparent that no further phasing power was available from the isomorphous heavy-atom derivatives; efforts at refining the best [Hg(Ac)$_2$] derivative indicated a fall to 1 in the value of (mean heavy-atom amplitude)/(error of closure) by 4·0 Å resolution.

Real-space symmetry averaging

The procedure adopted for symmetry averaging was based on the programs of Bricogne (1976). The use of non-crystallographic symmetry restraints to extend the phasing beyond an available set of low-resolution phases was originally suggested by Rossmann & Blow (1962) and first reported at high resolution (from 3·5 to 3·0 Å) by Gaykema, Hol, Vereijken, Soeter, Bak

Table 5. *Orientation of non-crystallographic threefold axes*

	Trimer 1				Trimer 2			
	ω	φ	KAPPA*	TRANS†	ω	φ	KAPPA	TRANS
Unrefined	61·8	89·6	120·0	0·0	17·1	78·5	120·0	0·0
Smeared‡	60·9	90·8	120·2	0·3	18·4	82·2	121·6	0·9
Final§	60·9	90·8	120·0	0·2	18·7	81·1	120·4	0·7
Coords¶	61·0	90·7	119·9	0·0	18·4	80·7	120·1	0·0

* The orientation of the threefold axis (KAPPA) is defined using polar coordinates where ω is the angle from the pole and φ is the angle around the equator.
† TRANS is the magnitude of the error in the translation along the threefold axis.
‡ Smeared refers to the results of the refinement on the original 4 Å Wang map smeared with an effective B factor of 500 Å^2.
§ Final refers to the values used in the final stages of the averaging at 2·9 Å; these were obtained by further refinement using $2f_{obs} - f_{calc}$ maps obtained during the course of the phase extension.
¶ Coords refers to the values calculated in retrospect from the refined model coordinates.

& Beintema (1984). The method has been mainly applied to, and achieved notable successes in, the field of virus crystallography (Rossmann, Arnold, Erickson, Frankenberger, Griffith, Hecht, Johnson, Kamer, Luo, Mosser, Rueckert, Sherry & Vriend, 1985; Luo, Vriend, Kamer, Minor, Arnold, Rossmann, Boege, Scraba, Duke & Palmenberg 1987; Hogle, Maeda & Harrison, 1986; Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989). In order to proceed with real-space averaging and phase extension two pieces of information must already be available. Firstly, an accurate orientation and position must be obtained for the symmetry axis about which the averaging is to take place. Secondly, an envelope must be drawn to delineate the region exhibiting that non-crystallographic symmetry.

Initial estimates for the directions and positions of the non-crystallographic threefold axes of the two TNF trimers were obtained from the refined heavy-atom sites of the Hg(Ac)$_2$ derivative supplemented by additional points taken from inspection of the Wang map on the graphics. These points were input to a simple program (*TOSS-BITS*) based upon the superposition algorithm of Kabsch (1976); this yielded an orientation for the threefold axes. Programs (*EXTRACT, ROTATE, MATCH*) to refine the orientation of an axis of non-crystallographic symmetry by density correlation using the initial electron density map have been written by Bricogne and these were implemented. A program *CIRCLES* was written to allow the creation of simple envelopes comprising any number of spheres of given centre and radius. The threefold orientations for the two trimers were refined separately. For each trimer the envelope produced to select the electron density for the correlation was made up of two small spheres (radii 20 Å) centred between 20 to 30 Å apart along the threefold axis of the trimer. This gave a conservative estimate of the area obeying the local threefold symmetry to avoid introducing excess noise into the process. At first the procedure failed; tests showed an apparent lack of convergence for any errors in the initial threefold-axis orientation in excess of 1°. Although some modifications were made to the refinement program

MATCH success was only achieved by smearing the 4 Å Wang map; this was carried out in reciprocal space by applying an effective B factor of 500 Å^2. Tests now showed very effective convergence to the correct orientation from a wide range of errors in the initial rough orientation (convergence from an error of 30° was observed). The results of the refinement for the threefold operators of both TNF trimers are presented in Table 5. The success of the refinement procedure is illustrated by the agreement between these values and the threefold orientations calculated in retrospect from the refined model coordinates.

A rough envelope approximated by a sphere large enough to contain a trimer was produced using *CIRCLES*. The position of the sphere centre and a suitable radius (50 Å) was estimated from the existing Wang map. The sphere was centred in turn on the two trimers and a single run of real-space averaging (GENERATE, SORT, INTERPOLATE, SORT and MODIFY) was applied to each using the appropriate refined threefold operators. The aim of this procedure was to preserve the density for the selected trimer while attenuating any closely associated density from neighbouring trimers which would not obey the same threefold symmetry operators. Thus the method should produce a clarified image of the selected trimer in isolation to facilitate the tracing of an accurate envelope. The strategy for the production of the envelopes is given in Fig. 7. The maps for both trimers were separately plotted in sections and envelopes for each were determined by inspection. Each section was then displayed on a pericom MG200 terminal emulating a Tektronix 4014. The chosen solvent/protein boundary was entered by hand using a program *TRACE* developed for the purpose. Thus a hand-traced envelope was obtained section by section independently for trimer 1 and trimer 2. An automatic reciprocal-space envelope determination (program by David Stuart, Elizabeth Fry and Ravi Acharya) was also carried out for the entire $P3_121$ cell assuming 60% crystal solvent content and applying a smearing B factor of 1000 Å^2. The final tight envelopes about the two trimers were obtained as a pixel-by-pixel logical .AND. of the automatic and hand-drawn

envelopes. For trimer 1 the resultant envelope, for single copy of the trimer in isolation, constituted 3·5% of the total volume of the unit cell and for trimer 2 the rather lower value of 2·9%.

The initial cycles of real-space averaging and phase refinement were carried out at 4 Å. The protocol employed for this iterative process, at a fixed resolution and ultimately for the phase extension to 2·9 Å, is illustrated in Fig. 8. The two trimers were separately threefold averaged thus allowing a comparison between the two to act as an independent check on the success of the procedure. A system of coarsely and finely sectioned maps was employed; the coarse maps sectioned at half the resolution of the data were used for the modified maps and the stages following, the fine maps sectioned at a quarter of the resolution, *i.e.* oversampled, were used to increase the accuracy of the linear interpolation as recommended by Bricogne (1976). The actual real-space averaging was carried out in a $P1$ cell which was rapidly rebuilt to a $P3_121$ cell by combining the relevent Fs in reciprocal space. This novel method has considerable advantages in terms of speed and simplification of the treatment of symmetry and edge effects during the map modifications. However, particular care was required in the use of the envelopes. In the program *TRUNC4* (see Fig. 6) the automatically determined envelope for the $P3_121$ cell was applied. This was used to set the mean density in the solvent region to zero. The $P1$ maps could then be obtained by isolating in turn the two unique trimers, cut out using their individual envelopes, in otherwise empty maps which retained the dimensions appropriate for the $P3_121$ cell. The isolation of a single copy for each of the two unique trimers prior to interpolation ensured that any deviation from threefold symmetry in the envelopes could not result in spurious density being picked up from outside the envelope. Similarly, *MODIFY* (Bricogne, 1976) was altered to preclude any density being put down outside the original envelope. After real-space averaging the two $P1$ maps were added; since the mean density in the solvent regions had been set to zero (in *TRUNC4* above), the addition of these maps did not distort the background level. The $P3_121$ cell was then rebuilt from this $P1$ cell using the reciprocal-space procedure. The scaling between F_{obs} and F_{calc} was performed in resolution shells with Rayment weighting applied. Initial test runs demonstrated that satisfactory behaviour over repeated cycles of real-space averaging could only be obtained if this iterative procedure was performed using maps calculated with amplitudes of $2F_{obs}-F_{calc}$. The initial heavy-atom/Wang phases were discarded after the first cycle. Apart from in the very low-resolution region (below 10 Å) the native data set to 4 Å was 99% complete. During averaging all reflections of resolution less than 15 Å were set to F_{calc} and unmeasured reflections were predicted to a resolution slightly in excess of the resolution limit on F_{obs} to alleviate series truncation errors (Rayment, 1983).

The results of the averaging at 4 Å are included in Fig. 9. For the first cycle the overall averaging R factor (the crystallographic R factor between observed structure-factor amplitudes and structure-factor amplitudes calculated from the averaged and solvent flattened map) was 39·3%. After 30 cycles of averaging (including three envelope redetermina-
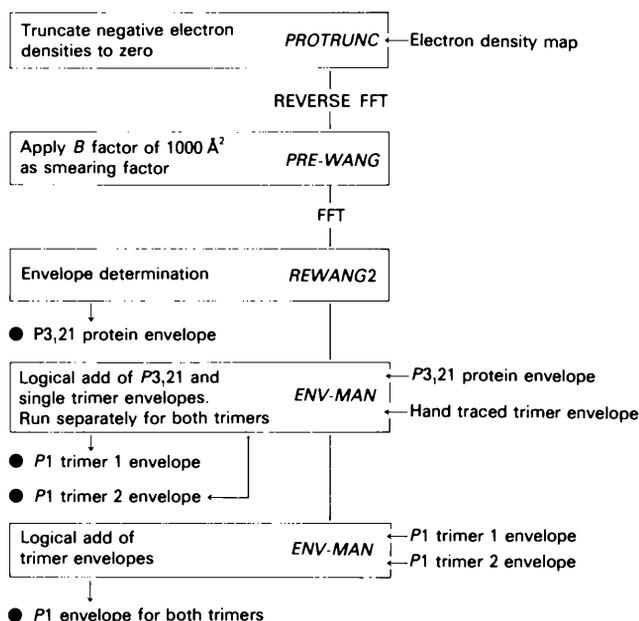
Fig. 7. Schematic of the strategy followed for the production of the envelopes required for averaging.

tions) the overall averaging R factor was 26·0% and the quality of the 4 Å electron density map was greatly improved (Fig. 10). However, considerable concern was experienced over the behaviour of trimer 2. The real-space R factor (the R factor between non-crystallographically related electron density) for this trimer was consistently worse (42·7%) than the equivalent value for trimer 1 (30·0%). The electron density seemed more diffuse for trimer 2, and as a result it made a smaller contribution to the Wang envelope. The threefold operators were re-refined after a couple of cycles of averaging using the detailed envelopes and a $2F_{obs} - F_{calc}$ map but the change in the trimer orientations was very slight and gave no significant

improvement when averaging was resumed. Difference maps were calculated to check for areas omitted or overlapped in the envelopes; however, none were apparent. Trimer 1 density was cut out and pasted into the trimer 2 position but a few cycles of averaging rapidly restored the previous situation and the crystallographic R factor for the initial cycle with trimer 1 density was high. Similarly the trimer 1 envelope flipped into position to act as the trimer 2 envelope gave no improvement in the behaviour. Thus the second trimer appeared to obey non-crystallographic threefold symmetry rather weakly. It had been the intention to turn on additional twofold averaging between the trimers to give overall sixfold

Adjust map zero level		Wang envelope for $P3_121$
Flatten solvent to zero	TRUNC4	Hand drawn + Wang envelope for 2 trimers in $P1$
Convert to 2 trimers in $P1$		Electron density map

SHRINK

	INTERPOLATE	INTERPOLATE	Sorted indices from GENERATE
Average trimers separately in $P1$	SORT	SORT	
	MODIFY	MODIFY	Hand drawn + Wang envelope for single trimer

Combine averaged maps for trimer 1 and trimer 2 in $P1$	ADD MAPS

Produce F_{calc} and φ_{calc}	REVERSE FFT

Build $P3_121$ cell from $P1$ cell in reciprocal space	CADLCF BUILD-F

Scale to master native set and weight phases	FSMELT SHELL-SCALE $\longleftarrow F_{obs}$

Monitor phase changes	MONPHASE $\longleftarrow \varphi_{calc}$ from previous cycle

Create new map	FORWARD FFT EXTEND

Fig. 8. Schematic of the protocol employed for real-space averaging and phase extension.

averaging during phase extension but because of the behaviour of trimer 2 it was decided to continue with independent threefold averaging of each molecule.

Phase extension was carried out in a continuous series of $1/450\ \text{Å}^{-1}$ steps in reciprocal space (i.e. steps of about a third of a reciprocal-lattice unit). The strategies employed were as detailed above for the real-space averaging at 4 Å. Rayment weights were applied to the data with F_{calc}s (down weighted by 0·7) used in place of unobserved reflections; the most recently added data were scaled separately with F_{calc}s for the unobserved data down weighted by 0·5. The behaviour of the system remained stable over the 67 cycles performed to reach 2·8 Å resolution. 13 additional iterations were performed at 2·8 Å to ensure convergence. The resultant map in the region of trimer 1 was of very good quality, the trimer 2 electron density, although of improved resolution, still appeared poor by comparison. Similarly the values of the real-space R factor continued to show a marked disparity, with trimer 1 giving 26·3% and trimer 2 giving 42·2%; the overall averaging R factor was 27·8% and the correlation coefficient 0·77 (Fig. 9).

(a)

(b)

(c)
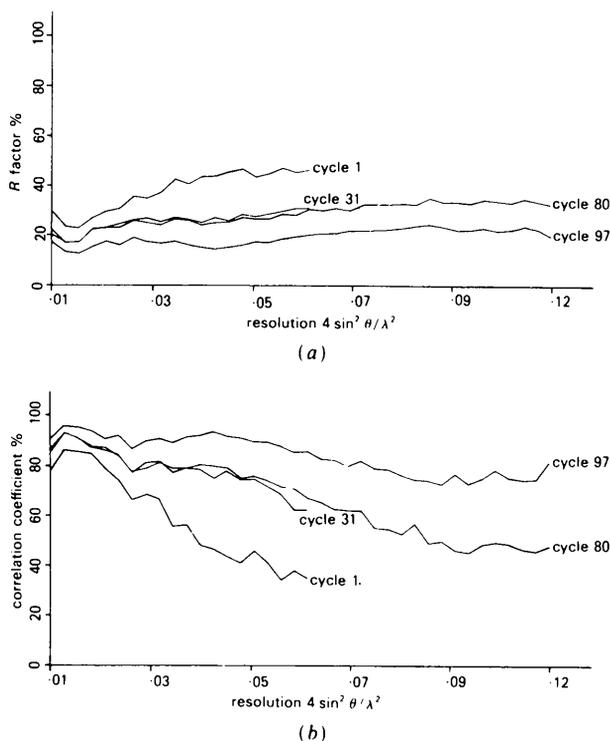
Fig. 9. Statistics for TNF phase extension. Cycles 1 and 31 are the first and last cycles of averaging at 4 Å, cycle 80 is the last cycle of averaging on both trimers at 2·9 Å and cycle 97 is the final cycle at 2·9 Å averaging on only trimer 1. Statistics plotted against resolution are given for (a) R factor

$$R = \sum|(|F_{obs}| - |F_{calc}|)|/\sum|F_{obs}|;$$

and (b) correlation coefficient:

$$C = \frac{\sum_h(\langle F_{obs}\rangle - |F_{h,obs}|)(\langle F_{calc}\rangle - |F_{h,calc}|)}{[\sum_h(\langle F_{obs}\rangle - |F_{h,obs}|)^2\sum_h(\langle F_{calc}\rangle - |F_{h,calc}|)^2]^{1/2}}.$$
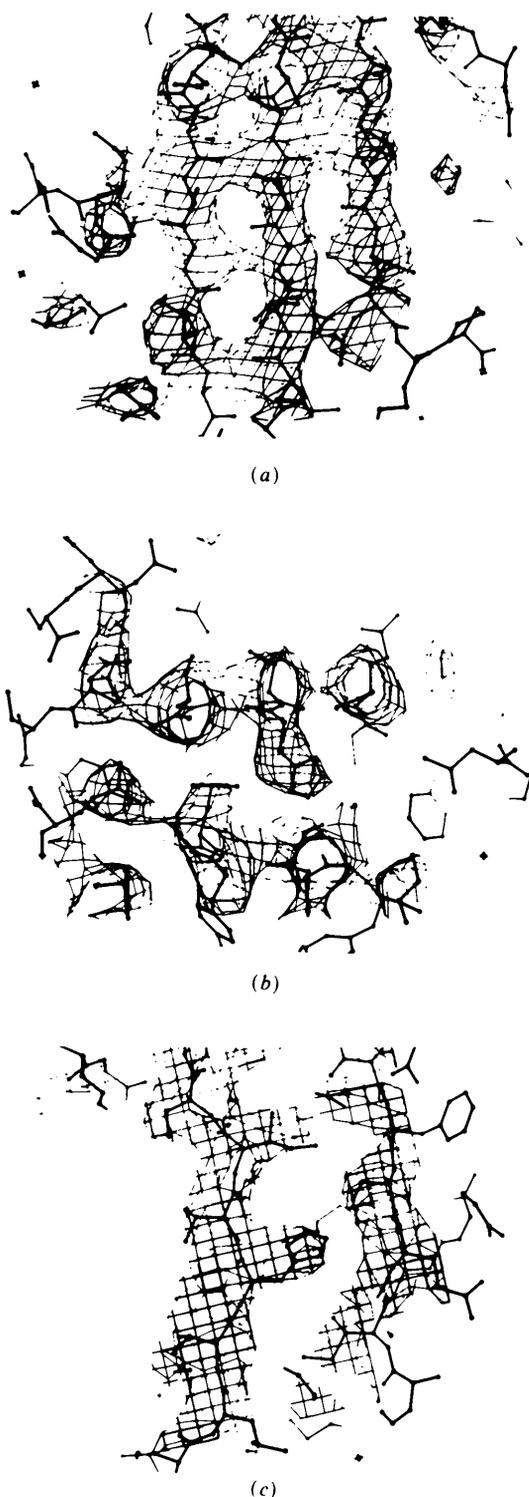
Fig. 10. Portion of the electron density map at 4 Å resolution after threefold averaging and solvent flattening. The region shown is centred on histidine 78 in subunit 1 trimer 1; coordinates for the current refined model are superimposed on the density. Three orthogonal views are given (a) into the β-sheet; (b) down the β-strands of the β-sandwich; (c) edge on to the β-sheets. The contour level is 1·5σ.

These values compare unfavourably with, for instance, the corresponding ones for foot and mouth disease virus [final averaging R factor 14·9% and correlation coefficient 0·92 (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989)]. It was therefore decided to try a final series of cycles at 2·8 Å with the threefold averaging on trimer 2 turned off. To our surprise after 17 cycles this had resulted in some improvement throughout the map although the effective averaging was now only 1·5-fold. Thus, given the additional phase redundancy provided by constraining 65% of the cell to solvent regions, less than two-fold averaging was actually sufficient to improve the phases. The final overall averaging R factor was 18·9% (Fig. 9).

The quality of the final map (Fig. 11) allowed a chain trace to be rapidly established and the sequence fitted for one of the subunits in trimer 1. The full model was then generated by application of the non-crystallographic symmetry operators to this subunit and the coordinates submitted to molecular dynamics refinement in *XPLOR* (initial R factor 48·5%). This resulted in a crystallographic R factor of 25·5% (on all data between 6 and 2·9 Å resolution with deviation from ideal bond lengths of 0·015 Å) without any further manual intervention. After several rounds of minor rebuilding and refinement the current crystallographic R factor is 20·4% (on all data 6 to 2·9 Å with good stereochemistry and no model for the solvent). Maps based on $2F_{obs} - F_{calc}$ amplitudes show the electron density for trimer 2 is now of comparable quality to that of trimer 1. As expected from the Wilson plot for the native diffraction data the B factors are high but agree well between the six TNF subunits. Refinement with the six subunits restrained to obey non-crystallographic symmetry resulted in the distortion of regions of the model away from their correct positions (clearly demonstrated by omit maps), thus the refinement has continued without any non-crystallographic restraints. The mean positional deviations between main-chain atoms for the subunits in trimer 1 and trimer 2 are, in both cases, 0·6 Å. Neither trimer obeys true threefold symmetry; the TNF trimer is apparently rather malleable and thus is easily distorted by crystal contacts. In trimer 1 the deviations from 120° in superimposing subunit 2 on 1, subunit 3 on 2 and subunit 1 on 3 are respectively −3.8, 1·9 and 2·1° (with translational displacements of −0·29, 0·39 and −0·14 Å). For trimer 2 the corresponding deviations are 4·4, −4·2 and −0·4° (with translational displacements of 0·68, −0·43 and −0·24 Å). This marked deviation from the assumed threefold symmetry offers an explanation for the particularly poor behaviour of the TNF trimer during the averaging. The overall statistics for the averaging, even of just trimer 1, were markedly worse than those obtained concurrently by Acharya *et al.* (1989) for foot and mouth disease virus (a final averaging R factor of

(a)

(b)

(c)
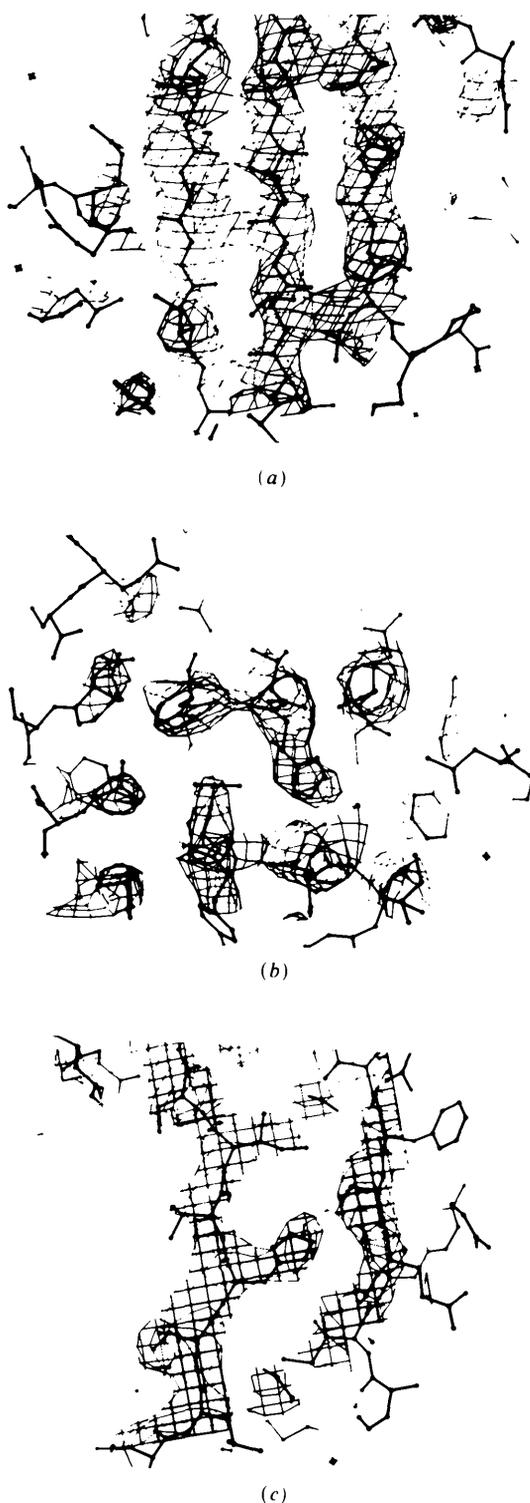
Fig. 11. Portion of the electron density map at 2·9 Å resolution after phase extension using threefold averaging and solvent flattening. The regions shown are centred on histidine 78 in subunit 1 trimer 1; coordinates for the current refined model are superimposed on the density. Three orthogonal views are given (a) into the β-sheet; (b) down the β-strands of the β-sandwich; (c) edge on to the β-sheets. The contour level is 1·5σ.

14·9% on all data from infinity to 2·8 Å); however, this discrepancy now seems reasonable given that the virus adheres to exact non-crystallographic symmetry.

Discussion

In Figs. 5, 6, 10 and 11 the same portion of electron density map is viewed at different points in the structure determination. Fig. 12 shows the mean phase differences between the phases corresponding to these stages and those of the current refined model (i.e. nominally the 'correct' phases). This represents our best estimate of the 'phase error' at any stage, however, it is likely to be over-pessimistic. Thus, if the error in the calculated phases is 30° an error of 30° in the comparison phase would produce a mean difference of 43° (apparent error) and an apparent error of 60° would correspond to a true error of 56°. For the resolution shell 5 to 4·7 Å the mean phase 'error' after Wang solvent flattening at 4 Å essentially based on SIR phases from the Hg(Ac)$_2$ derivative is 67° (equivalent to a figure of merit of 0·39), the magnitude of the error then increases for the higher-resolution shells up to 4 Å. The available derivatives were all of relatively poor quality (Table 4), but sufficient information was available for the solvent boundary to be correctly defined. Thus careful application of the solvent-flattening procedures did yield considerable phase improvement. From Table 4 the figure of merit for the Hg(Ac)$_2$ derivative SIR phasing to 4·5 Å was 0·38, the current calculations indicate a value of 0·51 after solvent flattening. It should be noted that the value quoted by the solvent-flattening programs (the Wang figure of merit), 0·70, is artificially high; this underestimation of the errors in the phase angles is an inherent weakness of the usual Wang procedure. The 4 Å solvent-flattened map (Fig. 6) shows some resolution of β-strand position emerging from the generally featureless slabs apparent at 6 Å (Fig. 5), thus an important shell of reciprocal space had been correctly phased. The phasing of the low-resolution reflections which carry information on the general positioning of the secondary

units is a crucial and often problematic step given weak phase information (Acharya, Fry, Stuart, Fox, Rowlands & Brown, 1989; Stuart, unpublished observations). However, the inclusion of threefold averaging obviously led to a major improvement in the quality of the phases at 4 Å as is evident from Fig. 12 and the dramatic improvement in the quality of the electron density (Fig. 10). The mean phase error in the resolution shell 5·0 to 4·7 Å has dropped to 53° (figure of merit 0·61) and the electron density shows clear β-strands and indications of side-chain size. One immediate result of the phase extension to 2·9 Å is the correction of series-termination errors still apparent in the higher-resolution shells at 4 Å. Thus the mean phase error for the 5·0 to 4·7 Å shell further drops to 49·5° (figure of merit 0·65). The phase extension does succeed in providing meaningful phases for the 4·0 to 2·9 Å region of reciprocal space (Fig. 12) with the estimated overall phase error for the 4·0 to 2·9 Å range in resolution of 71° (figure of merit 0·32). However, the improvement in the phases obtained by one- and-a-half-fold averaging using only trimer 1 is particularly noteworthy. Errors in the relatively low-resolution phases are corrected as evidenced by the improvement in the 5·0 to 4·7 Å shell phases (43·7° mean error corresponding to a figure of merit of 0·72) and the resultant electron density is improved (Fig. 12). The final 2·9 Å map was based on phases with an overall error when compared to the current refined phases of 61° which would correspond to an estimated figure of merit of 0·48. At this stage the side chains were now well defined and there were indications of some main-chain carbonyl bulges; thus the map was of sufficient quality to allow an unambiguous model to be built.

In successfully determining the structure the identification of putative derivatives and the solution of the corresponding Patterson difference maps was obviously crucial. At least one of the derivatives had to be of sufficient quality to allow the protein/solvent boundary to be correctly determined. This then allowed considerable phase improvement by solvent flattening applied to the relatively large solvent regions (crystal solvent content 65%). Once sufficient information was available to accurately define the non-crystallographic threefold axes and the individual trimer boundaries threefold averaging combined with solvent flattening provided adequate power to phase extend successfully to the 2·9 Å diffraction limit of the native data. This was despite the significant deviation of the TNF trimers from exact threefold symmetry.
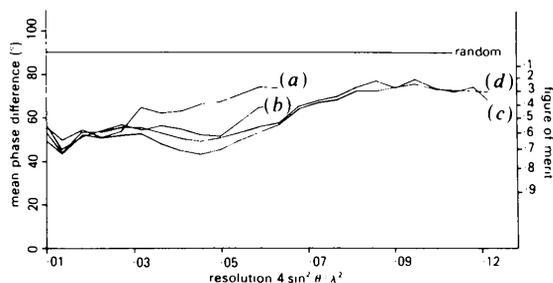
Fig. 12. Phase differences from final refined model. Curve (a) indicates the error in the phases after Wang solvent flattening at 4 Å, (b) indicates the error after averaging at 4 Å, (c) indicates the error after averaging on both trimers at 2·9 Å and (d) indicates the error after averaging only on trimer 1 at 2·9 Å.

APPENDIX

A modification of the Wang algorithm for the automatic determination of a molecular envelope. By D. I. STUART and E. Y. JONES

The knowledge that the solvent regions in a crystal should have constant electron density introduces an effective oversampling in the diffraction data which may be used to restrain the phases in a similar manner to the redundancy introduced by non-crystallographic symmetry. Thus, for crystals with reasonably high solvent content, solvent flattening may serve as a powerful method of breaking the phase ambiguity associated with SIR phases and further refining these phases (or indeed improving phases from any source). However, a vital prerequisite to the successful application of the method is the definition of an accurate protein/solvent boundary. Thus the regions of protein must be correctly distinguished from the regions of solvent in an electron density map calculated using the initial phases.

The solvent-flattening procedure developed by Wang (1985) is now widely and routinely employed for phase improvement. The protein/solvent boundary is delineated in the Wang methodology by making use of the assumption that the mean density in the protein regions is higher than that of the background solvent. A combination of truncation of negative densities and convolution of density over surrounding pixels is used to heighten the contrast between protein and solvent regions, the molecular boundary is then defined by setting a threshold level for protein density which yields the correct ratio of protein-to-solvent volume. This method has proved to be strikingly successful in many cases. However, the correct identification of the solvent/protein boundary, often made on the basis of extremely poor-quality electron density maps, is obviously crucial, thus, we here suggest additional criteria on which to base the envelope determination.

If the fluctuation in electron density is compared for a region of protein and a region of solvent the variations may be expected to be more extreme in the protein (reflecting the underlying structure of the protein). In the Wang algorithm a measure of the convoluted density at a pixel is taken as proportional to the weighted sum of the densities within a sphere centred on that pixel. By direct analogy, a measure of the density fluctuation at a pixel may be taken as proportional to the weighted sum of the absolute values of the deviations in the input map from this convoluted density within a sphere centred on that pixel. The probability that a pixel is in a protein region may then be based on a sum of the usual density measure and this value for the local average magnitude of the density fluctuation, typically giving equal weight to both criteria (we choose not to multiply these measures since the map has been truncated prior to analysis so that the measures are no longer truly independent). The modification required to incorporate this assessment of density fluctuation into the existing code by B. C. Wang for automatic envelope determination (ENVELP) is relatively minor.

A comparison of the performance of the original and modified algorithm for envelope determination has been carried out using SIR data to 4 Å for the tumour necrosis factor (TNF) Hg(Ac)$_2$ derivative. The results are presented in Fig. 13. Initial phases were calculated on the basis of the original heavy-atom positions from heavy-atom refinement. A total of seven cycles of envelope determination were run in parallel using the two algorithms. The envelopes obtained at each stage were compared with the final envelope at 4 Å obtained in the TNF structure determination. This 'ideal' envelope was based on information from three derivatives at 6 Å and very gradual extension in resolution to 4 Å for phases from the Hg(Ac)$_2$ derivative. Although both algorithms perform well given an initial electron density map calculated using rather poor SIR phases (a figure of merit on input of 0·23) the modified method appears superior. The suggested modification may prove vital in defining the correct envelope for a few very marginal cases and, as demonstrated here, should routinely improve both the initial accuracy of the envelope and its convergence to the best possible boundary.

Fig. 13. Rate of convergence of protein/solvent envelopes to 'true' envelope. Results for the original Wang algorithm lie on the upper curve and those for the modified algorithm correspond to the lower curve. Disagreement is measured in terms of the number of map pixels wrongly allocated to protein or solvent regions. Each cycle consisted of an envelope determination (assuming 64% solvent with two passes of convolution radius 5·6 Å) and six iterations of solvent flattening and phase modification (Rayment weighting). Details of the data used are given in the text.

References

ACHARYA, R., FRY, E., STUART, D., FOX, G., ROWLANDS, D. & BROWN, F. (1989). Nature (London), 337, 709–716.
BEUTLER, B. & CERAMI, A. (1988). Ann. Rev. Biochem. 57, 505–518.
BRICOGNE, G. (1976). Acta Cryst. A32, 832–847.
BRUNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). Science, 235, 458–460.
CROWTHER, R. A. (1972). In The Molecular Replacement Method, edited by M. G. ROSSMANN. Int. Sci. Rev. Ser. No. 13, p. 10. New York: Gordon and Breach.

DIAMOND, R. (1971). *Acta Cryst.* A27, 436-452.

DODSON, E. J. (1985). In *Molecular Replacement*, edited by P. A. MACHIN, pp. 33-45. Warrington: SERC Daresbury Laboratory.

ELY, K. R., GIRLING, R. L., SCHIFFER, M., CUNNINGHAM, D. E. & EDMUNDSON, A. B. (1973). *Biochemistry*, 12, 4233.

FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* 21, 273-278.

GAYKEMA, W. P. J., HOL, W. G. J., VEREIJKEN, J. M., SOETER, N. M., BAK, H. J. & BEINTEMA, J. J. (1984). *Nature (London)*, 309, 23-29.

HOGLE, J. M., MAEDA, A. & HARRISON, S. C. (1986). *J. Mol. Biol.* 191, 625-638.

HOLDEN, H. M., RYPNIEWSKI, W. R., LAW, J. H. & RAYMENT, I. (1987). *EMBO J.* 6, 1565-1570.

JONES, A. (1985). *Methods Enzymol.* 115, 157-171.

JONES, E. Y., STUART, D. I. & WALKER, N. P. C. (1989). *Nature (London)*, 338, 225-228.

KABSCH, W. (1976). *Acta Cryst.* A32, 922-923.

LADENSTEIN, R., BACHER, A. & HUBER, R. (1987). *J. Mol. Biol.* 195, 751-753.

LESLIE, A. G. W. (1989). In *Improving Protein Phases*, edited by S. BAILEY, E. DODSON & S. PHILLIPS, pp. 13-24. Warrington: SERC Daresbury Laboratory.

LUO, M., VRIEND, G., KAMER, G., MINOR, I., ARNOLD, E., ROSSMANN, M. G., BOEGE, U., SCRABA, D. G., DUKE, G. M. & PALMENBERG, A. C. (1987). *Science*, 235, 182-191.

MUIRHEAD, H., COX, J. M., MAZZARELLA, L. & PERUTZ, M. F. (1967). *J. Mol. Biol.* 28, 117-150.

O'HALLERAN, T., LIPPARD, S. J., RICHMOND, T. J. & KLUG, A. (1987). *J. Mol. Biol.* 194, 705-712.

OLD, L. J. (1987). *Nature (London)*, 326, 330-331.

PETSKO, G. A. (1985). *Methods Enzymol.* 114, 147-156.

RAYMENT, I. (1983). *Acta Cryst.* A39, 102-116.

ROSSMANN, M. G., ARNOLD, E., ERICKSON, J. W., FRANKENBERGER, E. A., GRIFFITH, J. P., HECHT, H.-J., JOHNSON, J. E., KAMER, G., LUO, M., MOSSER, A. G., RUECKERT, R. R., SHERRY, B. & VRIEND, G. (1985). *Nature (London)*, 317, 145-153.

ROSSMANN, M. G. & BLOW, D. (1962). *Acta Cryst.* 15, 24-31.

STUART, D. I., LEVINE, M., MUIRHEAD, H. & STAMMERS, D. K. (1979). *J. Mol. Biol.* 134, 109-142.

TANAKA, N. (1977). *Acta Cryst.* A33, 191-193.

WALKER, N., MARCINOWSKI, S., HILLEN, H., MACHTLE, W., JONES, E. Y. & STUART, D. (1990). *J. Cryst. Growth*, 100, 168-170.

WANG, B. C. (1985). *Methods Enzymol.* 115, 90-111.

Extinction Effects in Polarized Neutron Diffraction from Magnetic Crystals. I. Highly Perfect MnP and YIG Samples

BY JIŘÍ KULDA*

Nuclear Physics Institute, ČSAV, 250 68 Řež near Prague, Czechoslovakia

JOSÉ BARUCHEL

Institut Laue-Langevin, 156X, 38042 Grenoble CEDEX, France, and Laboratoire Louis Néel, CNRS/UJF, 166X, 38042 Grenoble CEDEX, France

AND JEAN-PIERRE GUIGAY AND MICHEL SCHLENKER

Laboratoire Louis Néel, CNRS/UJF, 166X, 38042 Grenoble CEDEX, France

Abstract

Wavelength-dependent flipping-ratio data obtained with polarized neutrons are analysed with the help of two extinction models: Kato's [*Acta Cryst.* (1980), A36, 763-769, 770-776] statistical dynamical theory (SDT) and the random elastic deformation (RED) approach [Kulda (1987). *Acta Cryst.* A43, 167-173; (1991). *Acta Cryst.* A47, 775-779]. Because the crystals used were nearly perfect, *Pendellösung* oscillations are present in all cases and the extinction theories are thus tested under extreme conditions. Both treatments give fair agreement with the experimental data in the short-wavelength range and there is no significant difference between them in terms of the goodness of fit, but the parameters involved in the RED approach have a more straightforward physical meaning. There are, however, inconsistencies in the values of the SDT parameters obtained, as well as probable deviations at large wavelength and crystal thickness, which could be related to a new validity condition for SDT.

1. Introduction

The problem of adequate correction of observed Bragg reflection intensities for the effect of extinction, *i.e.* reduction with respect to the prediction of the kinematical theory, has been subject to continuous efforts of both theorists and experimentalists for many years. In the last decade the availability of a statistical diffraction theory (SDT) capable, in principle, of dealing with the whole range of crystal perfection (Kato, 1976, 1980) has spurred a few experimental