Acta Cryst. (1997). D**53**, 262–268

Application of Known Triplet Phases in the Crystallographic Study of Bovine Pancreatic Trypsin Inhibitor. I: Studies at 1.55 and 1.75 Å Resolution

R. H. MATHIESEN AND F. MO

Institutt for Fysikk, Norges Teknisk-Naturvitenskapelige Universitet, N-7034 Trondheim, Norway. E-mail: fmo@phys.unit.no

(*Received 10 September 1996; accepted 27 November 1996*)

Abstract

The structure of bovine pancreatic trypsin inhibitor could be re-determined both with data at 1.55 and 1.75 Å resolution using direct methods strengthened by the application of about 90 and 130 triplet phases, respectively, assumed known with a mean error of ±20°. From the known triplets a similar number of single phases were derived and used as starting values, thereby reducing or eliminating the need for permuting reflections in the starting set. The known triplets provide better estimates for the mean direction parameters in the corresponding Cochran distributions. This improvement is crucial for obtaining a structure solution. Single phases derived from these triplets could be combined to yield a larger set of triplets which forms the basis for a new figure of merit (FOM). The new FOM appears superior to the conventional ones for initial selection of the best phase model, and can also be used for assessing the correctness of structure expansion if the number of triplet phase relationships is sufficient.

1. Introduction

In a recent study we have shown that structure solution of the small protein rubredoxin was feasible with data at 1.54 Å resolution when a small set of 45 triplet phases assumed known with a mean error of ±22.5° was used as input to direct methods (Mo, Mathiesen, Hauback & Adman, 1996). The presence of an FeS_4 cluster in this structure was employed as a criterion to identify the most probable phase models from a subset which was obtained after applying two other criteria to filter off the least probable solutions from the total set. It was of interest to see if a similar combination of triplet phases with estimators from direct methods could be successfully applied with a structure of comparable complexity but without heavy atoms. Bovine pancreatic trypsin inhibitor (BPTI) was chosen as a good candidate for this study. BPTI is a protein of 58 amino-acid residues containing approximately 500 non-H atoms, with no atoms heavier than sulfur. Its structure was originally solved at 2.5 Å by the multiple isomorphous replacement technique (Huber, Kukla, Rühlmann, Epp & Formanek, 1970). Later, a second form (form II) of this protein was solved at 1.0 Å resolution using joint refinement of X-ray and neutron data (Walter & Huber, 1983; Wlodawer, Walter, Huber & Sjölin, 1984). Form II crystallizes in $P2_12_12_1$ with $a = 74.1$, $b = 23.4$ and $c = 28.9$ Å. The X-ray data comprises 17 615 reflections with weights $w > 0$ in the range $1.0 < d < 7.2$ Å.

The primary aim for the present work was to improve strategies for expansion in real space (the structure model) and in reciprocal space (the phase set). Standard figures of merit (FOM's) of direct methods are known to be unreliable in work with macromolecules, in particular in the early critical stages of structure build-up (Woolfson & Yao, 1990; Mo *et al.*, 1996). Therefore, a further purpose was to develop a better FOM for monitoring progress. This part of the work in turn led to a reassessment of the question of how many triplet phases are needed to accomplish a reliable phase expansion. Finally, it was considered important to examine the possibilities for structure solution with data at a resolution lower than 1.5 Å.

2. Theoretical considerations

We assume here that a certain number of triplet phases acquired from three-beam diffraction experiments are available and can be used as a large starting set in direct methods. How can this additional phase information be utilized efficiently for the structure solution? Re-entering the mean direction parameter in the Cochran distribution (Cochran, 1955), one can derive the unweighted tangent formula (Karle & Hauptman, 1956) as a maximum-likelihood estimate, β, for the direction of the phase $\varphi_{\mathbf{H}}$,

$$\beta = \tan^{-1}\left[\frac{\sum_{\mathbf{L}} |E_{\mathbf{L}}||E_{\mathbf{H}-\mathbf{L}}|\sin(\varphi_{\mathbf{L}} + \varphi_{\mathbf{H}-\mathbf{L}} + \Phi_3)}{\sum_{\mathbf{L}} |E_{\mathbf{L}}||E_{\mathbf{H}-\mathbf{L}}|\cos(\varphi_{\mathbf{L}} + \varphi_{\mathbf{H}-\mathbf{L}} + \Phi_3)}\right], \quad (1)$$

where,

$$E_{\mathbf{L}} = |E_{\mathbf{L}}|\exp(i\varphi_{\mathbf{L}}),$$

$$\Phi_3 = \begin{cases} \Phi_3^{\text{est}} & \text{for measured triplet phases} \\ 0 & \text{for all other triplet phases} \end{cases}.$$

Only for those terms that contain an estimated triplet phase, Φ_3^{est}, from three-beam diffraction profiles, will this formula differ from the traditional estimator in direct methods. In the experimental part we have used the weighting scheme of Germain, Main & Woolfson (1971).

Let us assume that triplets can be selected for measurements from the bottom region of the convergence map (Germain, Main & Woolfson, 1970). This assumption seems to be realistic if the resolution of the data is about 1.5 Å or less. For higher resolution, the relative number of weak structure-factor (s.f.) amplitudes will increase, giving fewer measurable triplets as compared with the total number of triplet phase relationships (TPR's) in consideration. This means that even if high-resolution data are available, it may well turn out preferable to start the solution of the phase problem from a subset at lower resolution. If the number of measurable triplets is sufficient, measurements can be made to set up a solvable set of linear equations giving single s.f. phase estimates, $\varphi_{\mathbf{H}}$, directly. If there are m measured triplet phases Φ_3, single phases, $\varphi_{\mathbf{H}}$, can be derived,

$$\boldsymbol{\Phi}_{m\times 1} = \mathbf{Z}_{m\times k}\cdot\boldsymbol{\varphi}_{k\times 1} + \mathbf{b}_{m\times 1}, \qquad (2)$$

$\mathbf{Z}$ is the design matrix containing elements ± 1 and 0, while the k elements in the φ matrix are single phases. The $\mathbf{b}$ matrix consists of terms that arise from using the space-group symmetry to transform the single reflections involved in a triplet such that they are represented with either $\pm\mathbf{H}$; $\mathbf{b}$ is, therefore, trivial to deduce. This system of equations differs from that of Debaerdemaeker & Woolfson (1975), since we have here the physical estimates of the elements in $\boldsymbol{\Phi}$.

The system (2) is generally solvable if $k \le m$, given $\mathbf{Z}$ has full rank. Since we are free to fix n phase values for origin definition, the set of equations can be solved for $k + n \le m$. The values obtained will be estimates because of the inaccuracies in the physically estimated Φ_3's. At present, a realistic mean deviation in Φ_3 is about 20° compared with the values calculated from the crystallographically refined phases. This is due in part to the experimental resolution forcing a triplet to take on values $i\cdot\pi/4$, $i\in\{0, \ldots, 7\}$, but also to inadequacies in the present dynamical theory for X-rays in the case of three interacting beams in a finite crystal. Thus, the physical estimates for the triplets can provide single-phase values as well as better values for the mean direction parameters in the Cochran distributions. Solvability of (2) reduces or eliminates the need for single s.f. phase permutations when applying (1) iteratively, and therefore reduces greatly the number of phase models to be developed and examined.

Physically estimated triplets also provide the basis for a new figure of merit. Let us assume that m Φ_3's have been measured and used as input to a direct-methods program. A set of linear equations like (2) can be solved to give k single-phase values. The k single phases, and their symmetry equivalents, can be combined to yield a total of t triplets, which include the set of m that has actually been measured. After a phase refinement, one should expect the refined values of the t triplets for a good phase model to be closer to their measured (m) and precalculated ($t-m$) values than for a poor model. A new FOM can be defined,

$$\text{PHIFOM} = 1/t\sum_{i=1}^{t}\min\Big[\Big|\text{mod}\Big(\varphi_{3,i}^{\text{est, calc}} - \varphi_{3,i}^{\text{ref}}, 2\pi\Big)\Big|, 2\pi - \Big|\text{mod}\Big(\varphi_{3,i}^{\text{est, calc}} - \varphi_{3,i}^{\text{ref}}, 2\pi\Big)\Big|\Big], \qquad (3)$$

where $\varphi_{3,i}^{\text{est, calc}}$ = a physically estimated or precalculated triplet phase and $\varphi_{3,i}^{\text{ref}}$ = a refined triplet phase.

A PHIFOM close to zero indicates a good solution, $\pi/2$ indicates a random solution.

3. Experimental

Intensity data comprising 17 615 unique reflections were kindly provided by Dr M. Schneider of Professor R. Huber's laboratory. The data were in the d range 1.0–7.2 Å. Within the shells 1.0–1.1 and 1.1–1.2 Å only 31.3 and 47.5%, respectively, of the total number of accessible reflections were observed. Therefore, this portion of the data is not well suited for normalization. The percentages also indicate that an *ab initio* direct-methods approach to the structure solution is unlikely to succeed (Sheldrick, 1990). Data with interplanar spacings $d > 7.2$ Å was not included in the transferred files, but since these low-index reflections were considered important for good triplet coupling, intensities from the shell $7.2 < d < 20$ Å were calculated from the published coordinates with an overall isotropic B. Within each subshell of thickness 0.025 Å^{-1} in $\sin\theta/\lambda$ unobserved reflections were given a value $0.5|F_o|_{\min}$ and then added to complete the data. Two program packages were employed for extension and refinement of the phases: *MULTAN*78 (Main *et al.*, 1978), which was modified locally to (*a*) accept up to 3000 E's, (*b*) develop and store up to 100 000 TPR's, and (*c*) accept triplet phases with user-defined weights as input in the *SIGMA*2 routine; and *MULTAN*88 *E* (Debaerdemaeker, Germain *et al.*, 1988), which employs the *SAYTAN* formalism (Debaerdemaeker, Tate & Woolfson, 1985, 1988). The particular version of this program that was used allows processing up to 100 000 TPR's, ranked according to weight. The program can also process quartet phase relationships, but quartets were not employed in this work.

In separate, preliminary studies, as many as 200 single phases with values calculated from the published structure were used as starting sets for direct methods to assign phases for the 500 largest E's. The resolution

limit ranged from 1.3 to 1.55 Å. All attempts based on single phases alone were unsuccessful. This result can be ascribed to the dominant uniform character of the Cochran distributions when the number of atoms in the unit cell becomes large. In the case of BPTI it turns out that of the 90 TPR's at the bottom of the convergence map, 41 triplets, or 45%, deviate by 90° or more in phase from the theoretical estimate 0°. In particular, triplets 1, 4 and 5 from the bottom are negative (180°).

3.1. *Work with data at 1.55 Å resolution*

A total of 7767 reflections, of which 7204 had $w > 0$, were normalized and sorted in order of descending amplitudes. The 500 largest amplitudes were selected to set up a convergence map from 8200 TPR's. The minimum κ value was 0.26; κ is the common estimate for the concentration parameter in the underlying distribution (Cochran, 1955),

$$\kappa_{\mathbf{H},\mathbf{L}} = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{H}}E_{\mathbf{L}}E_{\mathbf{H}-\mathbf{L}}|. \quad (4)$$

It was used here in this form even if the derivation of the prefactor $\sigma_3\sigma_2^{-3/2}$ rests on the assumption of fully resolved atoms (Cochran, 1955), which is generally not true for proteins. Clearly, a sufficient number t of known triplets is needed both for the calculation of a reliable PHIFOM, and for the use of (2) to avoid numerous solutions to the phase problem. This number may be larger than the minimum required to accomplish a breakthrough in the phase development which was a criterion in the work with rubredoxin (Mo *et al.*, 1996). In the case of BPTI, with no heavy-atom fragment in the structure, the role of a reliable FOM becomes even more important.

From the bottom of the convergence map 91 Φ_3's were selected according to experimental requirements on the absolute and relative strengths of the amplitudes of the $F_{\mathbf{H}}$, $F_{-\mathbf{L}}$ and $F_{\mathbf{L}-\mathbf{H}}$ s.f. that are involved in the TPR's. The set of Φ_3 should also provide a consistent basis for the calculation of single $\varphi_{\mathbf{H}}$. The Φ_3's were assigned values $i\cdot\pi/4$, with $i\in\{0, \ldots, 7\}$ according to values calculated from the refined structure. Physical estimation of 91 Φ_3's is experimentally quite feasible. Weckert (1997) has estimated about 700 triplet phases from crystals of tetragonal lysozyme.

Including reflections for the origin definition this process gave 94 single $\varphi_{\mathbf{H}}$ estimates, from which a total of t = 215 Φ_3's could be calculated for the use in PHIFOM. Two different reflections were permuted to yield four different sets of refined phase models. Table 1 shows several FOM's that have been calculated for these models. ABSFOM, PSI0, RESID and CFOM are FOM's defined in *MULTAN*78 (Main *et al.*, 1978) and in *MULTAN*88 *E* (Debaerdemaeker, Germain *et al.*, 1988). In addition PHIFOM and $\langle\Delta\Phi_3\rangle$ are given. The last quantity is the mean triplet phase error which can only be calculated *a posteriori*, when the phases from the crystallographic refinement are known. A contradiction in the traditional FOM's is apparent in Table 1. While ABSFOM and PSI0 give preference to models 1 and 2, RESID and CFOM (with default weighting) favour models 3 and 4. Table 1 also illustrates the inability of the traditional FOM's to identify the correct enantiomer. Starting phase values for models 1 and 2 (or 3 and 4) are equivalent except for an inversion of the sign of the phase for the enantiomer defining reflection. For the pair of models 1 and 2, RESID leads to the opposite conclusion of ABSFOM and PSI0, even though the difference in $\langle\Delta\Phi_3\rangle$ for the two models is significant and corresponds to a considerable amount of structure in favour of model 1. The enantiomer insensitivity of the FOM's becomes even more evident recalling that the measured TPR's input to the phase refinement contain information on the molecular chirality ($\Phi_3 = \pm 90°$). PHIFOM is able to discriminate between the related models, 1 and 2 or 3 and 4, and to identify the best model overall as well. Model 1 is clearly indicated as the best model by PHIFOM.

Table 1. *Calculated FOM's and mean triplet phase error $\langle\Delta\Phi_3\rangle$ for four models based on 500 E*

Starting phase values for the pair of models 1 and 2 or 3 and 4, are equivalent except for one reflection defining opposite enantiomers.

Model	ABSFOM	PSI0	RESID	CFOM	PHIFOM	$\langle\Delta\Phi_3\rangle$
1	1.0312	0.954	45.37	1.000	31.05	72.43
2	1.0875	0.957	43.45	1.834	40.64	77.76
3	1.2591	1.397	42.28	2.026	37.51	78.72
4	1.2769	1.446	42.34	1.980	43.16	79.34

A density map was calculated from the 500 refined E's of model 1. From an automatic peak search (APS) the 20 strongest peaks were selected and treated as possible atomic positions. Two maxima, separated by about 2.15 Å, were four times more dense than any of the other 18 peaks in the map. They were tentatively assigned as S atoms while the other maxima were labled C atoms. The use of a random peak search was motivated by experience from the rubredoxin work (Mo *et al.*, 1996) where it was found that E maps, even at an early stage, contain much correct information.

From this point on the iterative phase estimation with the tangent formula (1) based on the convergence map and the known triplet phases was abandoned, as the program *MULTAN*88 *E* had not been modified to incorporate such measurements. Instead, the *SAYTAN* formalism (Debaerdemaeker, Tate *et al.*, 1985, 1988) was employed with additional information consisting of: (i) the 500 $\varphi_{\mathbf{H}}$ values from the initial refinement cycle used as starting values for the refinement and extension of phases, and (ii) the molecular fragment present at that stage incorporated to modify the calculation of the concentration parameters, $\kappa_{\mathbf{H},\mathbf{L}}$, and to calculate new theoretical values, generally different from 0, for the mean direction parameters of the corresponding von Mises distributions (Main, 1976).

Table 2. *Study at 1.55Å resolution*

Development of FOM's and mean triplet phase error $\langle\Delta\Phi_3\rangle$ after this cycle, number of E's phased by direct methods and number of atoms included from map prior to this cycle.

Cycle	ABSFOM	PSI0	RESID	PHIFOM	$\langle\Delta\Phi_3\rangle$	Model + APS atoms	$(N-M)$ E's
0	1.031	0.954	45.37	31.05	72.43	0+0	500
1	3.355	2.201	81.43	58.60	83.55	2+18	1200
2	3.064	1.979	75.22	58.13	81.94	4+16	1200
3	2.640	1.804	60.59	59.07	79.86	17+15	1200
4	1.934	1.618	35.67	53.68	76.65	33+15	1200
5	1.685	1.432	32.41	53.53	77.49	53+15	1800
6	1.248	1.308	30.32	44.14	71.45	71+15	1800
9	1.065	1.319	28.41	44.57	66.93	71+32	1800
12	0.833	1.195	26.49	36.45	56.04	79+68	2000
15	0.811	1.155	26.65	35.03	48.85	131+79	2200
18	0.797	1.153	27.27	28.70	42.10	162+75	2600
21	0.743	1.132	28.13	29.31	39.64	205+85	2600
24	0.713	1.116	28.75	26.79	37.80	234+74	2200

Phase refinement and expansion was carried out in a recycling process.

(*a*) Calculate new $\kappa_{\mathbf{H},\mathrm{L}}$ and Φ_3^{calc} for all TPR's among the set of N largest E's (see Table 2) from current molecular fragment. Accept all TPR's with $\kappa_{\mathbf{H},\mathrm{L}} > \kappa_{\min}$ for the phase refinement.

(*b*) Use $(N-M)$ φ from initial cycle to retain the origin and enantiomorph and to avoid multiple solutions. Develop and refine phases among the $(N-M)$ best determined E's using *SAYTAN*. Include the starting set phases in the refinement for six additional cycles after convergence is achieved.

(*c*) Calculate a new E map from $(N-M)$ E's and search for larger fragment. From cycle 19 include for the calculation the following 5000 – $(N-M)$ E's with phases calculated directly from the coordinates of current molecular fragment.

(*d*) Return to (*a*).

M was 0 in the intial cycle, then 50 for the next four cycles, and thereafter 100 to the end. In the recycling process $\kappa_{\min}$ increased from 0.35 to its final value of 3.4. In cycle 1 the data set was increased to include the 1200 largest E's. This strong expansion of the phase set caused a large increase in all FOM's (Table 2). A more gradual expansion would have allowed a higher value to be retained for the threshold $\kappa_{\min}$. The effect on the FOM's can be understood in terms of this threshold as low κ values will reduce the inner consistency of TPR's used to estimate the same $\varphi_{\mathbf{H}}$ (Woolfson & Yao, 1990; Mo *et al.*, 1996). The increase in PHIFOM, on the other hand, is not dominated by this effect alone. It is also important for this figure that known Φ_3's are no longer part of the input to control the phase refinement (*i.e.* all TPR's are assigned Φ_3 calculated from the current molecular fragment only), and the phase refinement was extended for six cycles after convergence which may cause a general drift from the optimized starting values of the single phases that are used for the calculation of PHIFOM. The total number of Φ_3's, t, that are involved in PHIFOM should be sufficient to monitor the expansion process. Here t = 215 which was 2.6% of the TPR's in the initial refinement cycle. Later experience suggests that this fraction should be larger, maybe closer to 10% to produce a reliable FOM. The undulations in PHIFOM in Table 2 should be studied with this in mind.

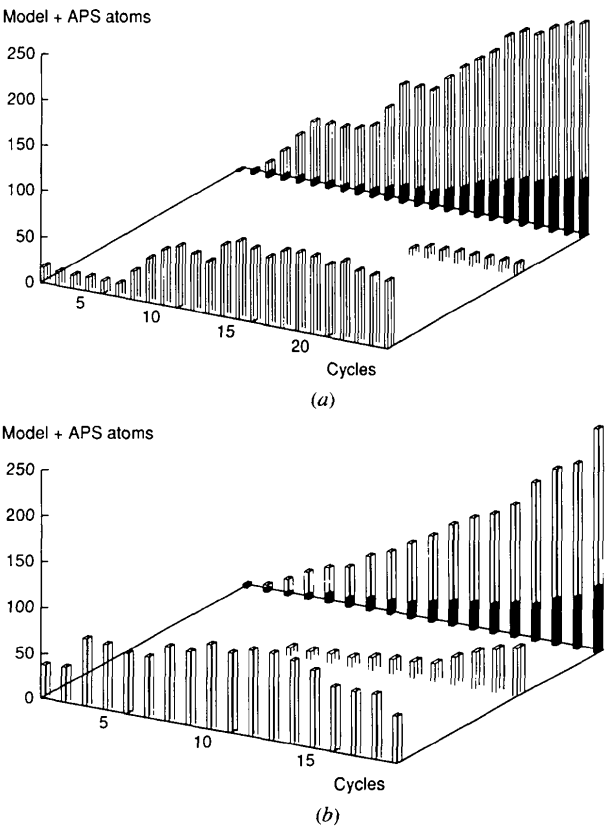

Fig. 1. Progress in the structure modelling of BPTI. Rear histogram shows total number of atomic positions assigned to the connected molecular fragment, shaded fraction is number of atomic positions obtained from APS maxima. The middle histogram gives number of APS maxima excluded during modelling. The front histogram gives number of APS maxima remaining unconnected at each stage. (*a*) The study at 1.55 Å resolution. (*b*) The study at 1.75 Å resolution.

Table 3. *Study at 1.75 Å resolution*

Development of FOM's and mean triplet phase error $\langle\Delta\Phi_3\rangle$ after this cycle, number of E's phased by direct methods and number of atoms included from map prior to this cycle.

Cycle	ABSFOM	PSI0	RESID	PHIFOM	$\langle\Delta\Phi_3\rangle$	Model + APS atoms	$(N-M)$ E's
0	0.785	0.773	63.18	22.28	70.0	0+0	400
1	1.780	1.310	27.28	41.97	78.97	4+38	500
2	1.914	1.270	32.30	45.44	82.58	8+40	700
3	1.325	1.271	28.36	46.14	82.40	18+75	900
4	1.185	1.272	27.95	38.91	81.58	30+73	1100
5	1.176	1.248	28.12	42.53	79.93	39+68	1100
6	1.124	1.175	28.05	35.89	79.43	44+68	1100
7	1.145	1.139	29.46	34.30	77.59	60+83	1500
8	1.078	1.120	29.17	34.17	75.34	69+82	1700
9	0.995	1.104	28.88	32.75	72.72	83+94	1900
12	0.839	1.092	27.32	33.06	64.45	95+89	2200
15	0.776	1.126	27.82	25.79	52.98	131+93	2200
18	0.730	1.103	28.48	18.91	40.43	245+52	2200

The remaining FOM's are gradually reduced towards normality from cycle 1 onwards. However, they cannot be used to monitor progress unless κ_{min} is adjusted to keep the mean number of TPR's nearly constant for each reflection involved in the phase refinement.

A new E map was calculated from the refined phases from cycle 1, and a new peak search confirmed the 20 maxima from the first map. Two of the maxima that had been ascribed to C atoms now appeared with densities comparable to the pair defined as S atoms. As their interdistance was 2.20 Å, the two peaks were interpreted tentatively as a second disulfide bridge. After another cycle, modelling based on connectivity and geometric restraints could be started from the assumed S atomic positions in a new E map. Essential for the modelling was the presence of one of the cysteine N atoms in a peak identified in the first APS. This particular cysteine residue was completed during the first attempts at modelling and enabled us to start tracing out the protein backbone. Progress in the structure modelling and a history of the unassigned APS positions extracted from each new E map can be studied in Fig. 1(*a*).

After cycle 4 the four cysteines were complete, as well as some connected peptide groups. In each cycle typically a fragment of 1–2 atoms could be localized in each 'tail' of the previously modelled chain, leading to a total increase of 10–20 new atoms. The number of reflections was increased gradually to ensure a more reliable development of phases. Following cycle 6, an additional fragment comprising 17 maxima from an APS was included as C atoms. New maxima were accepted with (*a*) electron density > 0.9 e Å^{-3},* (*b*) a minimum separation of 1.0 Å from other maxima. Table 2 shows that the 17 unconnected APS maxima lead to a change in the FOM's not much different from that of a connected fragment of similar size.

From this stage, structure modelling and APS were performed regularly towards the end. From cycle 19 a total of 5000 reflections were used to calculate maps. The $(N-M)$ E values indicated in Table 2 were phased by direct methods whereas the remaining 5000 – $(N-M)$ terms were phased from the coordinates of the connected molecular fragment at that stage. This was carried out to ensure a sufficient number of E_H values to calculate more detailed maps. After 25 cycles, a 234-atom fragment had been fitted geometrically to match the protein backbone and cysteine residues connected in disulfide bridges. 74 APS maxima were still unconnected but were retained as atoms belonging to various side groups along the chain. The r.m.s. deviation of the connected molecular fragment from that of the refined structure was 0.22 Å.

Calculations were made on microVax and on Sun/Unix machines. The model fitting was carried out using a combination of printed density maps and contour maps produced by the *Xtal* program packages (Hall & Stewart, 1990; Hall, Flack & Stewart, 1992).

3.2. *Work with data at 1.75 Å resolution*

5470 reflections, of which 5121 had $w > 0$, were normalized. The 400 largest E's were used to construct a convergence map from 5689 TPR's; κ_{min} for these triplets was only 0.19, as all possible relationships among the 400 E's were used. From 130 experimentally accessible triplet phases selected from the bottom of the convergence map 134 single phases could be assigned a starting value *via* (2), and a total of t = 580 Φ_3's became available for the calculation of PHIFOM; t was about 10% of the total number of TPR's. After intial was less straightforward. Since only one S—S bridge was identified, just four sites were available for modelling of the connected backbone. Assuming that the sequence of amino-acid residues was known, linking the backbone between the two cysteines would allow identification of the residues in this loop. After cycle 9 this sequence was completed identifying the cysteines as residues Cys30 and Cys51. A third cysteine (Cys38) belonged to this part of the sequence, and the modelling

* Scaling of the coefficients to put the densities on an approximate e Å^{-3} level is described by Hall & Stewart (1990).

of its S atom revealed a second S–S bridge, giving access to a new region of the molecule. Both the latter S atomic sites were among the 15 densest APS maxima.

From this point progress was steady. Work was discontinued when 245 atoms had been connected. Most of the atoms were backbone and cysteines linked together by S—S bridges. Some side-group atoms were also fitted geometrically when their presence was obvious in the maps, but no refinement was carried out. After cycle 18 there were 52 unexplained maxima, of which 12 had no corresponding atomic position closer than 0.875 Å (intergrid distance in the maps) in the published coordinate file. The r.m.s. deviation for the 245 connected atoms was 0.320 Å, and became 0.346 Å with the 40 unassigned peaks included. From cycle 12, 5000 reflections were phased and used for each map, as was done in the 1.55 Å study, with $(N-M)$ E's phased by *MULTAN*88 E and the rest phased from the molecular fragment. Fig. 1(*b*) shows the history of the structure modelling at 1.75 Å resolution.

3.3. *Measurability of triplet phases*

In this work triplet phases were assumed known, but were not actually measured. It is of interest to discuss briefly some factors that affect triplet phase acquisition from three-beam experiments. The feasibility of such experiments with protein crystals has been demonstrated most convincingly by Hümmer, Weckert and collaborators: myoglobin (Hümmer, Schwegle & Weckert, 1991), tetragonal lysozyme and katalase oxidoreductase (Weckert, Schwegle & Hümmer, 1993). A critical parameter in physical phase estimation (PPE) is the crystal mosaicity as obtained from rocking-curve experiments. In the myoglobin study a crystal consisting of three distinct blocks was used, FHWM for the maxima from the best individual was 0.023°. We have made successful PPE on crystals of a large organic complex with rocking-curve FWHM values about 0.03°. The limiting mosaicity will be larger, in particular if the required precision ±20° in the phase assignments is relaxed, perhaps about 0.05°. A second important parameter is the ratio $R_F = |F_{-\mathbf{L}}||F_{\mathbf{L}-\mathbf{H}}|/|F_{\mathbf{H}}|$, where $F_{-\mathbf{L}}$, $F_{\mathbf{L}-\mathbf{H}}$ and $F_{\mathbf{H}}$ are the secondary, the coupling and the primary s.f., respectively [*cf.* equation (6), Mo *et al.* (1996)]. Favourable conditions for PPE are when $F_{-\mathbf{L}}$ and $F_{\mathbf{L}-\mathbf{H}}$ are both strong, and R_F is in the range $3|F_{\mathbf{H}}|$–$10|F_{\mathbf{H}}|$. These conditions are frequently satisfied for TPR's near the bottom of the convergence map with data at moderate refinement of the 400 largest E's a map was calculated. The 40 most dense maxima were selected from an APS, all with similar densities > 0.9 e Å^{-3}. Several interdistances in the range 2.2–2.75 Å indicated possible disulfide bridges. In an attempt to identify S–S pairs PHIFOM was calculated for 14 different combinations of 38 C + 2 S positions after recycling by *SAYTAN*. Interestingly, PHIFOM did identify the same disulfide bridge as was found in the first cycle with the 1.55 Å data. This particular S–S pair was the only combination giving a lower PHIFOM than obtained after recycling all 40 positions as C atoms. The E map also provided the Cβ atoms of the two linked cysteines.

A careful expansion from 400 to 500 reflections was carried out prior to the first recycling. The new E map provided only three new APS maxima with density > 0.9 e Å^{-3}. Modelling in the second map gave four new atoms, among them a cysteine N atom in a maximum from the initial APS. Structure modelling and phase expansion was done much in the same way as in the 1.55 Å work, but with some improved features: (*a*) Calculations were performed on a Silicon Graphics Indigo 2 workstation, using the same software as in the 1.55 Å study. The *FRODO/TOM/Marseille* software was used for structure modelling. (*b*) Expansion of the phase set was accomplished in a more gradual manner to maintain $\kappa_{\min}$ as high as possible. This led to more normal FOM values compared with the 1.55 Å work (Table 3). Recycling was performed continuously, so the cycles in this study represent particular stages selected for illustration. (*c*) 400 φ's from the initial phase model were used as starting values. In the first seven cycles refinement involving all phases excluding the starting set was discontinued when convergence had been reached. After cycle 7 refinement was extended for six final rounds to include the starting set phases as in the 1.55 Å study. (*d*) Structure build-up and the inclusion of APS maxima were accomplished in a more continuous process. After cycle 5, APS maxima decreasing in density below 1.1 e Å^{-3} were filtered out. The value of M was varied as in the 1.55 Å study, and $\kappa_{\min}$ increased during recycling from 0.3 to 3.15.

In cycle 2, the set of E's was expanded to 700. The input fragment before this run consisted of eight atomic sites fitted to the model and, in addition, 40 APS maxima. In the new map ten atoms were modelled and 35 new unconnected APS maxima with densities above the 0.9 e Å^{-3} threshold were accepted. In the next few cycles the phase set was expanded stepwise by 200 E's, and about ten atoms could be fitted to the structure in each cycle.

In the 1.55 Å work four S atoms were identified at an early stage, allowing structure modelling to proceed from eight different sites. With the 1.75 Å data progress resolution. In our experience PPE was feasible with R_F in the range $0.2|F_{\mathbf{H}}|$–$240|F_{\mathbf{H}}|$, the majority of those with $R_F < 1.5|F_{\mathbf{H}}|$ being *Aufhellung* cases. Triplets with all s.f. of medium strength could also be measured. Synchrotron radiation with its small angular divergence, narrow spectral bandwidth and high flux is necessary in PPE with protein crystals. The spectral distribution furthermore in general allows the wavelength to be selected so that there are no strong three-beam cases adjacent to the one under study (Hümmer, Weckert & Bondza, 1990). Pure ψ scans, such as can be made with a six-circle diffractometer (Hümmer, Weckert & Bondza,

1989) are superior to scans with a four-circle instrument where a rotation in ψ is accomplished by compounding rotations in the three axes ω, χ and Φ (Mo, Hauback, Mathiesen, Kvick & Weckert, 1997). For crystals with limited tolerance to X-radiation it is imperative to carry out accurate ψ scans as fast as possible.

4. Conclusions

We have shown that a small protein structure like BPTI could be solved with data at resolution in the range 1.55–1.75 Å using direct methods with the additional information obtained from approximately 100 triplet phases assumed known with a mean error of about ±20°. The amount of structure information in the first Fourier maps is remarkably high. In the 1.55 Å study, 18 of the 20 primary APS maxima eventually could be fitted to the model. In the study with 1.75 Å data, 33 out of 40 primary APS maxima were fitted; 72 or about 30% of the 245 atomic sites in the final model at 1.75 Å originated from APS maxima adjusted to fit the geometrical restraints.

A new FOM, PHIFOM, appears superior to traditional FOM's for initial selection of the best phase model, but can also be used for evaluating structure building, provided the number of TPR's is sufficient. The attainable accuracy in measured TPR's allows calculation of single-phase starting values with good accuracy as well as better estimates for the directional parameters in the probabilistic formulae of direct methods. In particular for proteins containing no heavy elements, physically estimated TPR's could be of great value for solving the structure. Measurement of phases (MAD) or combinations of phases (TPR's) appears necessary when data of extremely high resolution is unavailable. At present, a practical limit in resolution to obtain an *ab initio* solution of structures of comparable size seems to be about 1.2 Å (Weeks *et al.*, 1995; Sheldrick, Dauter, Wilson, Hope & Sieker, 1993). Not many proteins diffract that well, and, therefore, techniques that can aid in the structure solution at more modest resolution are important.

In a continuation of the present study it appears that structure solution of BPTI is also feasible at 2.0 Å resolution starting from a small set of known triplet phases. In this case, a more extensive use of the known TPR's is required.

We are grateful to Professor R. Huber and Dr M. Schneider for providing the intensity data, and to Professor M. M. Woolfson and Dr C. Tate for making the program *MULTAN*88 *E* available. Support from Norges Forskningsråd under Grant 101166/432 is gratefully acknowledged.

References

Cochran, W. (1955). *Acta Cryst.* **8**, 473–478.
Debaerdemaeker, T., Germain, G., Main, P., Refaat, L. S., Tate, C. & Woolfson, M. M. (1988). *MULTAN*88. *Computer programs for the automatic solution of crystal structures from X-ray diffraction data*. University of York, York, England.
Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *Acta Cryst.* A**41**, 286–290.
Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1988). *Acta Cryst.* A**44**, 353–357.
Debaerdemaeker, T. & Woolfson, M. M. (1975). *Acta Cryst.* A**31**, 401–404.
Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274–285.
Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368–376.
Hall, S. R., Flack, H. D. & Stewart, J. M. (1992). Editors. *Xtal*3.2 *Reference Manual*. Universities of Western Australia, Australia, Geneva, Switzerland, and Maryland, USA.
Hall, S. R. & Stewart, J. M. (1990). Editors. *Xtal*3.0 *Reference Manual*. Universities of Western Australia, Australia, and Maryland, USA.
Huber, R., Kukla, D., Rühlmann, A., Epp, O. & Formanek, H. (1970). *Naturwissenschaften*, **57**, 389–392.
Hümmer, K., Schwegle, W. & Weckert, E. (1991). *Acta Cryst.* A**47**, 60–62.
Hümmer, K., Weckert, E. & Bondza, W. (1989). *Acta Cryst.* A**45**, 182–187.
Hümmer, K., Weckert, E. & Bondza, W. (1990). *Acta Cryst.* A**46**, 393–402.
Karle, J. & Hauptmann, H. (1956). *Acta Cryst.* **9**, 635–651.
Main, P. (1976). *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 97–105. Copenhagen: Munksgaard.
Main, P., Hull, S. E., Lessinger, L., Germain, G., Declercq, J.-P. & Woolfson, M. M. (1978). *MULTAN*78. *A system of computer programmes for the automatic solution of crystal structures from X-ray diffraction data*. University of York, York, England.
Mo, F., Hauback, B. C., Mathiesen, R. H., Kvick, Å. & Weckert, E. (1997). *J. Synchr. Rad.* Submitted.
Mo, F., Mathiesen, R. H., Hauback, B. C. & Adman, E. T. (1996). *Acta Cryst.* D**52**, 893–900.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). *Acta Cryst.* D**49**, 18–23.
Walter, J. & Huber, R. (1983). *J. Mol. Biol.* **167**, 911–917.
Weckert, E. (1997). *Acta Cryst.* D**53**. Submitted.
Weckert, E., Schwegle, W. & Hümmer, K. (1993). *Proc. R. Soc. London Ser. A*, **442**, 33–46.
Weeks, C. M., Hauptman, H. A., Smith, G. D., Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.
Wlodawer, A., Walter, J., Huber, R. & Sjölin, L. (1984). *J. Mol. Biol.* **180**, 301–329.
Woolfson, M. M. & Yao, J.-X. (1990). *Acta Cryst.* A**46**, 409–413.