CASTLEMAN, K. R. (1979). *Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall.
DEREWENDA, Z. S., DODSON, E. J., DODSON, G. G. & BIZOZOWSKI, A. M. (1981). *Acta Cryst.* A**37**, 407-413.
HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97-107.
PODJARNY, A. D., BHAT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Chem.* **16**, 351-373.
RANGO, C. DE, MAUGEN, Y., TSOUCARIS, G., DODSON, E. J., DODSON, G. G. & TAYLOR, D. J. (1985). *Acta Cryst.* A**41**, 3-17.
SAYRE, D. (1972). *Acta Cryst.* A**28**, 210-212.
SAYRE, D. (1974). *Acta Cryst.* A**30**, 180-184.
TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492-499.
WANG, B. C. (1985). *Methods Enzymol.* **115**, 90-112.

Acta Cryst. (1990). A**46**, 46-57

Extension of Molecular Replacement: a New Search Strategy based on Patterson Correlation Refinement

BY AXEL T. BRÜNGER

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, *USA*

(*Received* 28 *April* 1989; *accepted* 18 *August* 1989)

Abstract

A new search strategy is presented to obtain initial phases for single-crystal diffraction data by molecular replacement. It consists of carrying out 'Patterson refinements' of a large number of the highest peaks of a rotation function. The target function for Patterson refinement is proportional to the negative correlation coefficient between the squared amplitudes of the observed and the calculated normalized structure factors. If the root-mean-square difference between the search model and the crystal structure is within the radius of convergence of the minimization procedure employed, the correct orientation can be identified by having the lowest value of the target function after refinement. Similar to conventional crystallographic *R*-factor refinement, the target function for Patterson refinement may be combined with an empirical energy function describing geometric and non-bonded interactions. Patterson refinement of individual atomic coordinates or of rigid-group parameters may be carried out. Search models of crambin and of myoglobin with 1·6-2·0 Å backbone atomic r.m.s. differences from the target crystal structures show that the Patterson refinement strategy can solve crystal structures that cannot be solved by conventional molecular replacement or even by full six-dimensional searches.

Abbreviations

CPU, central processing unit; FFT, fast Fourier transformation; MR, molecular replacement; PC, standard linear correlation coefficient between $|E_{obs}|^2$ and $|E_{model}|^2$; RF, rotation function; r.m.s., root-mean-square; SA refinement, crystallographic refinement by simulated annealing with molecular dynamics; S/N, signal to noise; TF, translation function.

Introduction

In macromolecular crystallography, the initial determination of phases by 'molecular replacement' (MR) (Rossmann & Blow, 1962; Huber, 1965; Rossmann, 1972; Lattman, 1985) is often attempted if the structure of a similar or homologous macromolecule is known ('search model'). MR involves the placement (*i.e.* rotation and translation) of the search model in the unit cell of the target crystal in order to obtain the best agreement between calculated and observed diffraction data. The optimally placed search model is used to obtain initial phases for structure building and refinement. This approach may or may not succeed; many successful cases reported involve search models with a backbone atomic root-mean-square (r.m.s.) difference of less than 1 Å from the target structure (*e.g.* Wang, Bode & Huber, 1985; Schirmer, Huber, Schneider, Bode, Miller & Hackert, 1986).

Recent progress in obtaining approximate three-dimensional models of macromolecules from other information suggests an increased use of MR to solve crystal structures. For instance, the data base of known protein sequences and protein structures is growing rapidly. Techniques for aligning sequences such as consensus templates [see Taylor (1988) for a review] have been developed in order to recognize very distantly related proteins or protein domains and to carry out model building on the basis of the known protein structures. Another example is the determination of three-dimensional structures of small proteins and nucleic acids from nuclear magnetic resonance (NMR) NOESY experiments (Wüthrich, 1986). The

0108-7673/90/010046-12$03.00

search models obtained by these methods are often approximate. For example, the atomic r.m.s. differences in the protein core of homologous proteins range from 0·76 to 2·3 Å (Chothia & Lesk, 1986). Thus, MR could become very difficult, if not impossible. It is therefore desirable to extend the applicability of MR.

If there is one molecule in the crystallographic asymmetric unit, then three positional and three angular parameters fully describe the placement of the search model in the unit cell of the target crystal. Rossmann & Blow (1962) realized that this six-dimensional search can be reduced to a sequence of a three-dimensional angular search using a 'rotation function' (RF) followed by a three-dimensional positional search using a 'translation function' (TF). This procedure assumes that the highest peak of the rotation function yields the correct orientation. Examples are known where this is not true. Owing to advances in computer technology, multi-dimensional search strategies with more than three parameters are no longer beyond available computational resources [*e.g.* the program *BRUTE* (Fujinaga & Read, 1987)]. In fact, multi-dimensional searches in which symmetry or packing considerations reduce the number of parameters have already been successfully applied to solve, among others, a structure of haemoglobin (Baldwin, 1980), and various nucleic-acid structures (Shakked *et al.*, 1981; Rabinovich & Shakked, 1984).

Even six-dimensional searches may fail to solve the crystal structure. In this case we propose to vary the atomic coordinates of the search model in the neighborhood N of the initial positions. In the limiting case of $N = 0$, the method is identical to conventional MR, in the limiting case of $N \gg 0$ the method is equivalent to the crystallographic phase problem; in this report we focus on applications where N is small and falls within the radius of convergence of conjugate gradient minimization procedures. For instance, R-factor refinements could be carried out with the search model placed in the most likely orientations and positions as determined by a multi-dimensional search. However, this procedure would be computationlly intensive since the translation searches may not yield a unique solution and one would have to carry out refinements for several peaks for the translation search for each selected orientation.

We present a method to refine atomic coordinates of the search model *prior* to translation searches. The target function for the refinement consists of an effective Patterson energy term combined with an empirical energy function. The Patterson energy term is proportional to the negative correlation coefficient PC between the squared amplitudes of the observed and the calculated normalized structure factors. The normalized structure factors are computed with the search model placed in a triclinic unit cell identical in geometry to that of the crystal. The empirical energy function represents information about the geometry and non-bonded interactions of the macromolecule (Karplus & McCammon, 1983). PC refinement of individual atomic coordinates may be impractical for large molecules because of the large computational expense. In this case, generalized coordinates, such as the orientation and position of rigid groups, can be refined against a target function that simply consists of the Patterson energy term without an empirical energy function.

The following combined MR and PC refinement strategy is proposed. First, a conventional RF is carried out. All sampled orientations are sorted with respect to their RF value and then a large number of the highest peaks are selected for PC refinements. Finally, the PC-refined search models with the highest correlation coefficients are used for conventional translation searches.

This paper describes the results of computer studies that were aimed at evaluating the utility of PC refinement. A particular search model of crambin with a 2 Å backbone atomic r.m.s. difference from the crystal structure failed to provide the correct orientation when using a rotation function or a six-dimensional search. PC refinements of the search model in the most likely orientations resulted in the correct orientation having the lowest value of the target function. This enabled us to solve the crambin structure starting with the PC-refined search model. In an application to myoglobin we show that rigid-group PC refinement of the eight α helices has an approximate radius of convergence of 13° for orientational parameters. A search model of myglobin with the α helices artificially tilted by 13° failed to provide the correct orientation when using a rotation function or a six-dimensional search. PC refinements uniquely determined the correct overall orientation of the myoglobin search model by returning the α helices to their original placements.

Methodology

A. Real-space rotation search

The real-space Patterson search method (Huber, 1965; Fehlhammer & Bode, 1975; Huber, 1985) was employed for the orientation of the search model. The search-model Patterson maps were computed by placing the search model into an orthorhombic P_1 box, evaluating the structure factors, and fast Fourier transforming (FFT) the squared structure factors. Patterson vectors were selected based on length and RF value. The selected vectors were rotated by application of a rotation matrix $\Omega(\theta_1, \theta_2, \theta_3)$ where $\theta_1, \theta_2, \theta_3$ are the Eulerian angles according to the convention of Rossmann & Blow (1962). Values of the observed Patterson map at the rotated vector

positions were obtained by a linear eight-point interpolation (Nordman, 1980). The orientations of the search models were sampled by using the pseudo-orthogonal Eulerian angles (Lattman, 1972)

$$\begin{aligned}\theta_+ &= \theta_1 + \theta_3 \\ \theta_- &= \theta_1 - \theta_3 \\ \theta_2 &= \theta_2.\end{aligned} \tag{1}$$

The interval Δ for θ_2 was constant. Following Lattman (1972) the interval for θ_+ is given by $\Delta/\cos(\theta_2/2)$ and the interval for θ_- is given by $\Delta/\sin(\theta_2/2)$. The rotation search was restricted to an asymmetric unit of the rotation function (Rao, Jih & Hartsuck, 1980). For each sampled orientation Ω the product function

$$\mathrm{RF}(\Omega) = \langle P_{\mathrm{obs}} P_{\mathrm{model}}(\Omega)\rangle \tag{2}$$

between the rotated vectors $P_{\mathrm{model}}(\Omega)$ and the interpolated values of the observed Patterson map P_{obs} was computed.

For two given rotation matrices Ω^1, Ω^2 a metric was defined as

$$m(\Omega^1, \Omega^2) = \min_{s=1,n_s}\left[\sum_{i=1,3}\sum_{j=1,3}(\Omega^1_{ij} - \mathcal{O}_s\Omega^2_{ij})^2\right]^{1/2} \tag{3}$$

where n_s is the number of symmetry operators of the space group of the crystal and $\mathcal{O}_s$ is the rotational part of the symmetry operator s. Two RF grid points are considered as being in the same cluster if the corresponding rotation matrices yield a small $m(\Omega^1, \Omega^2)$. The incorporation of crystallographic symmetry in (3) ensures that clusters of grid points at the boundaries of the asymmetric unit of the RF are treated properly.

All grid points of the RF were sorted with respect to their RF value. Starting with the highest grid point '1' all grid points i were found that satisfy

$$m(\Omega^1, \Omega^i) < 0{\cdot}2. \tag{4}$$

The highest grid point of this cluster was retained while the other grid points were discarded. The procedure was repeated for the next-highest grid point that did not yet belong to another cluster. This set of peaks was used for subsequent translation searches or PC refinements. We will simply refer to this set of peaks as the *selected* peaks of the RF. Peak indices are assigned in the order in which they occur within the selected set; the index 1 corresponds to the highest peak of the RF.

The rotation function RF was vectorized and implemented in the program *X-PLOR* (Brünger, 1988*a*). The implementation of the method in *X-PLOR* was tested by comparing results with the program *PROTEIN* (Steigemann, 1974). The vectorization resulted in a fourfold speed up on a Convex-C1 compared with scalar mode; *X-PLOR* performed 5·7 times faster on a Convex-C1 than *PROTEIN* on a VAX8700.

B. Translation search with correlation coefficients

Translation searches of the rotated search models were carried out with the program *X-PLOR* (Brünger, 1988*a*) by computing the standard linear correlation coefficient between the squares of the normalized observed structure factors (E_{obs}) and the normalized calculated structure factors (E_{calc}),

$$\begin{aligned}\mathrm{TF}(xyz, \Omega) = [&\langle |E_{\mathrm{obs}}|^2 |E_{\mathrm{calc}}(xyz, \Omega)|^2\rangle \\ &- \langle |E_{\mathrm{obs}}|^2\rangle\langle |E_{\mathrm{calc}}(xyz, \Omega)|^2\rangle] \\ &\times \{[\langle |E_{\mathrm{obs}}|^4\rangle - \langle |E_{\mathrm{obs}}|^2\rangle^2] \\ &\times [\langle |E_{\mathrm{calc}}(xyz, \Omega)|^4\rangle \\ &- \langle |E_{\mathrm{calc}}(xyz, \Omega)|^2\rangle^2]\}^{-1/2}.\end{aligned} \tag{5}$$

The symbols $\langle\ \rangle$ denote an averaging over the set of observed reflections. The search model was placed in the unit cell of the crystal with the position given by the coordinates x, y, z of the center of gravity and with the orientation given by the rotation matrix Ω. The correlation coefficient TF is independent of the scale of E_{obs} and E_{calc}, and it is normalized to yield values in the range $-1, \ldots, +1$. A value of one implies total agreement whereas a value of zero implies no agreement. It has been suggested that TF is a more efficient statistical property than the conventional crystallographic factor for translation searches (Fujinaga & Read, 1987). Maximizing the correlation coefficient TF is equivalent to minimizing the phase error (Hauptman, 1982).

The translation search was initiated by computing and storing the structure factors E_{calc} of the search model and those of its symmetry mates. The translation search was carried out by applying appropriate phase shift factors to the calculated structure factors of the search model and to those of its symmetry mates (Fujinaga & Read, 1987). In this way the computation is very fast and it is also highly vectorized. The normalization of the structure factors was carried out by individual normalization for each of 20 concentric shells of equal volume between the resolution limiting spheres in reciprocal space,

$$E(\mathbf{h}) = \frac{F(\mathbf{h})}{\langle |F(\mathbf{h})|^2\rangle^{-1/2}}. \tag{6}$$

A sorted list of the highest peaks of the translation search was generated and used for subsequent analysis. The vectorization of the algorithm produced a fivefold speed up on a Convex-C1 with respect to a scalar version of the routine.

C. Packing function

A packing function was defined for evaluating the likelihood of packing arrangements of the search

model and its symmetry mates in the crystal (Hendrickson & Ward, 1976). A finite grid that covers the unit cell of the crystal was generated. All grid points were marked that were within the van der Waals radii around any atom of the search model and its symmetry mates. The number of marked grid points was counted; this represents the union of the molecular spaces of the search model and its symmetry mates. Maximization of the union of molecular spaces is equivalent to minimization of the overlap. Thus, an optimally packed structure has a maximum of the packing function. This packing-function algorithm was vectorized. The edge length of the grid was set to 1 Å.

D. PC refinement

We define PC refinement of individual atomic coordinates $\mathbf{r}$ as minimization of the target function

$$E_{\rm tot}(\mathbf{r}) = E_{\rm PC}(\mathbf{r}) + E_i(\mathbf{r}). \tag{7}$$

E_i is an empirical energy function that describes the geometry and non-bonded interactions of the molecule. A slightly modified form of the *CHARMM* empirical energy function TOPH19, PARAM19 was used as discussed by Brünger, Karplus & Petsko (1989). $E_{\rm PC}$ is an effective energy term that is proportional to the standard linear correlation coefficient $\mathrm{PC}(\mathbf{r}, \Omega)$

$$E_{\rm PC}(\mathbf{r}) = W_{\rm PC}[1 - \mathrm{PC}(\mathbf{r}, \Omega)] \tag{8}$$

where

$$\begin{aligned}\mathrm{PC}(\mathbf{r}, \Omega) = {} & [\langle |E_{\rm obs}|^2 |E_m(\mathbf{r}, \Omega)|^2\rangle \\ & - \langle |E_{\rm obs}|^2\rangle \langle |E_m(\mathbf{r}, \Omega)|^2\rangle] \\ & \times \{[\langle |E_{\rm obs}|^4\rangle - \langle |E_{\rm obs}|^2\rangle^2]\,[\langle |E_m(\mathbf{r}, \Omega)|^4\rangle \\ & - \langle |E_m(\mathbf{r}, \Omega)|^2\rangle^2]\}^{-1/2}. \end{aligned} \tag{9}$$

The symbols $\langle\,\rangle$ denote an averaging over the set of observed reflections expanded to P_1. $E_{\rm obs}$ denote the normalized observed structure factors and $E_m(\mathbf{r}, \Omega)$ denote the normalized structure factors of the search model oriented according to Ω and placed in a triclinic unit cell identical in geometry to that of the crystal. The choice of $\mathrm{PC}(\Omega)$ is related to the product function $\langle |F_{\rm obs}|^2 |F_m(\Omega)|^2\rangle$ which is often used in conventional MR (Lattman, 1985). However, in conventional MR the structure factors F_m are computed in a large orthorhombic P_1 box.

$W_{\rm PC}$ is a weight which relates $E_{\rm PC}$ to the empirical potential energy E_i. The procedure for determining the weight $W_{\rm PC}$ is similar to that developed for SA refinement (Brünger, 1988*b*; Brünger, Karplus & Petsko, 1989). It was chosen to make the gradients of $E_{\rm PC}$ and E_i of the same magnitude for a structure of the search model obtained after a short molecular dynamics simulation with $W_{\rm PC}$ set to zero.

We define PC refinement of generalized coordinates p, such as the orientation and position of rigid groups, as minimization against a target function that consists of $E_{\rm PC}$

$$E_{\rm tot}(p) = E_{\rm PC}(p). \tag{10}$$

Refinement of rigid groups is often used in conventional crystallographic R-factor refinement [program *CORELS* of Sussman, Holbrook, Church & Kim (1977)]. In principle, an empirical energy function could also be included .to describe interactions between the rigid groups. However, for the purposes of this study this appeared not to be necessary.

The effective energy $E_{\rm PC}$ and its first derivatives with respect to atomic positions were computed by using the subgrid fast Fourier transformation (SGFFT) method of Brünger (1989). The extension of the SGFFT method to the target $E_{\rm tot}$ [(7) or (10)] was straightforward. Individual atom and rigid-group minimization of $E_{\rm tot}$ was carried out by using a conjugate gradient algorithm (Powell, 1977).

E. The computer program

The algorithms described above were added to the program *X-PLOR* (Brünger, 1988*a*) which is running among others on VAX, Cray, Convex and Stellar computers. The procedures are highly automated and require minimal human intervention. The new features will be made available in a future version of *X-PLOR*. Requests for *X-PLOR* should be made to the author.

Results and discussion

A. Crambin

The diffraction data and the refined structure of crambin at 1·5 Å resolution were used [(Hendrickson & Teeter, 1981); referred to as $S_{\rm HT}$)]. The protein is composed of 46 amino acids. Previously, five approximate search models were obtained from an NMR structure determination that used simulated data (Brünger, Clore, Gronenborn & Karplus, 1986; Clore, Brünger, Karplus & Gronenborn, 1986). Of the five search models two were chosen for this work; they will be referred to as S_1 and S_3 [same notation as in Brünger *et al.* (1987); RDII′ and RDIC′ in Clore *et al.* (1986)].

All reflections were weighted equally. The atomic B factors of all atoms were set to 8·0 Å^2. Fig. 1 shows the r.m.s. differences averaged by residue of the NMR-derived search models S_1 and S_3 from the $S_{\rm HT}$ structure. The averaged r.m.s difference for backbone (C, C$^\alpha$, N) atoms of search model S_1 and S_3 from $S_{\rm HT}$ is 1·37 and 1·95 Å, respectively. The averaged r.m.s. difference for all non-hydrogen atoms of search model S_1 and S_3 from $S_{\rm HT}$ is 1·83 and 2·33 Å, respectively. It was shown by Brünger *et al.* (1987) that a

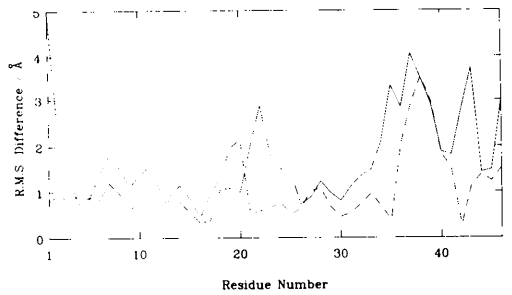
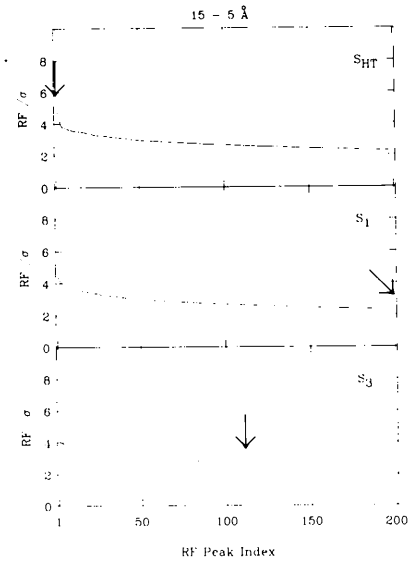
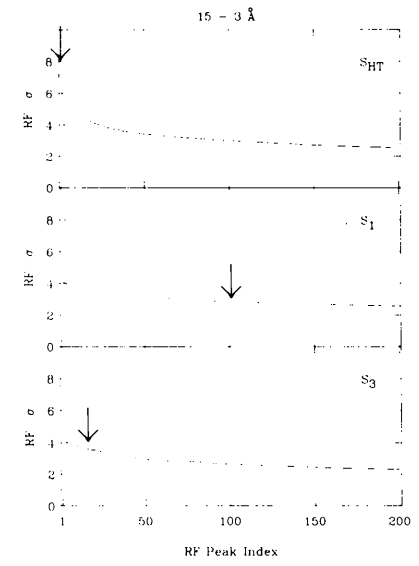


Fig. 1. Atomic r.m.s. differences in Å averaged over backbone (C^{α}, C, N) atoms *vs* residue number between crambin search model S_3 and structure S_{HT} (solid line), and between crambin search model S_1 and structure S_{HT} (dashed line).



(*a*)



(*b*)

reciprocal-space RF failed in the case of the search models S_1 and S_3. This situation changed little when the real-space RF described in the *Methodology* section was used (Fig. 2).

The RF of search model S_3 at 15-3 Å resolution exhibits a significant error peak which is about 1σ above all other peaks (Fig. 2*b*). Variation of the parameters of the RF (resolution range, vector range and grid size) changed the position of the correct peak but it never emerged as the highest peak. The 200 selected peaks of the RF for search models S_1 and S_3 at 15-3 Å resolution were used for further analysis.

Translation searches of S_1 and S_3 were carried out for each of the 200 selected RF peaks. The highest value (TF_{max}) of the translation function [(5)] and the value of the packing function of the position corresponding to TF_{max} is plotted *vs* RF peak index in Fig. 3. TF_{max} exhibits a significant global maximum and the packing function shows a local maximum for the correctly oriented search model S_1 (Fig. 3*a*). The translation search for this orientaton was successful with the highest peak being about 1σ above noise peaks (not shown). The *R* factor of search model S_1 in this orientation and position is 55% at 3 Å resolution and the structure refined readily to 36% at 2 Å resolution using SA refinement (Brünger *et al.*, 1989). Electron density maps were clearly interpretable. Thus, S_1 is an example of a search model that fails to provide the correct orientation when using a rotation function but is successful when using a six-dimensional search.

Figs. 3(*a*) and (*b*) justify the definition of the correlation coefficient PC [(9)]. PC is highly correlated with TF_{max}. The standard linear correlation

Fig. 2. Rotation searches for the crambin structure S_{HT}, search model S_1, and search model S_3 (*a*) at 15-5 Å resolution and (*b*) at 15-3 Å resolution. Shown are the values of the 200 highest selected peaks of the RF sorted by peak height. The correct orientation is indicated by an arrow (in the case of search model S_1 at 15-5 Å resolution the correct orientation is not among the first 200 peaks). The RF values are shown in units of standard deviations (σ) above the mean. The observed Patterson map was computed from the observed intensities by FFT with a grid size of 1·25 and 0·84 Å for (*a*) and (*b*), respectively. The model Patterson maps were computed by placing the search model into an orthorhombic box with 100 Å cell edges, evaluating the structure factors and FFT of the squared amplitudes with a grid size of 1·25 and 0·84 Å for (*a*) and (*b*), respectively. Model Patterson vectors were selected acording to length (between 30 and 4 Å), and according to peak height (2σ above the mean of the model Patterson map). The interval Δ for θ_2 was set to 5°. As the crambin crystal symmetry is monoclinic $P2_1$, *b* axis unique, the symmetry of the observed Patterson map is $P2/m$. The rotation search could therefore be restricted to the asymmetric unit $\theta_+ =$ 0-720°, $\theta_2 =$ 0-90°, $\theta_- =$ 0-360° (Rao, Jih & Hartsuck, 1980). RF peaks were selected by checking the cluster condition $m(\Omega^1, \Omega^i) < 0{\cdot}2$ [equation (3)]. The CPU time for each rotation search was 190 s and 650 s on a Cray-YMP for (*a*) and (*b*), respectively.

coefficient between the PC and TF_{max} graphs in Figs. 3(*a*) and (*b*) is 0·86 and 0·88, respectively. The correct orientation could have been predicted in Fig. 3(*a*) by inspecting the PC values for each selected orientation without having to carry out translation searches. In contrast to the rotation function, PC is not affected by possible interpolation errors and it properly treats all terms, strong or weak. In a different context PC was used as a criterion for minimal phase errors by Hauptman (1982).

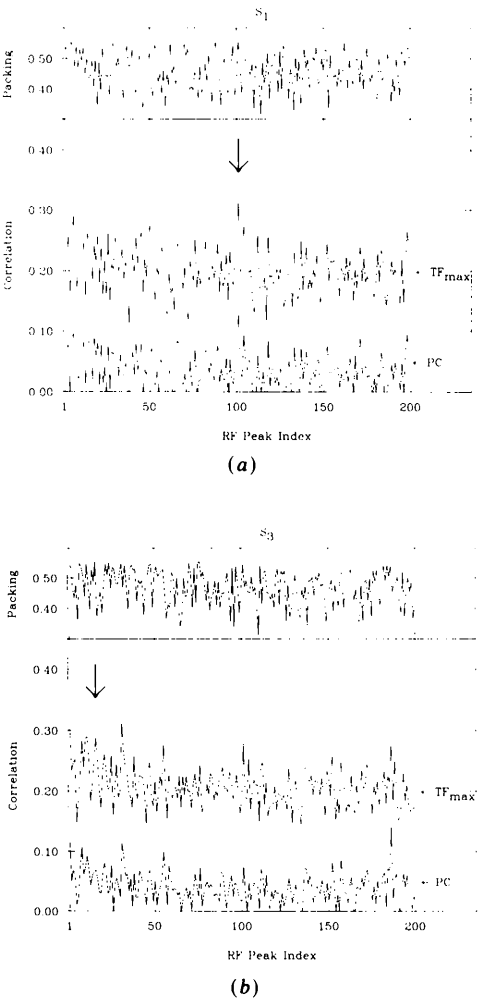


Fig. 3. Translation searches for each orientation of crambin search model (*a*) S_1 and (*b*) S_3 corresponding to the 200 highest selected peaks of the RF. Data were included between 15 and 3 Å resolution. Shown are the correlation coefficients (PC), the maximum value of the translation function TF_{max}, and the packing function value *vs* RF peak index. The correct orientation is marked by an arrow. The symmetry of the space group $P2_1$ of the crambin crystal implies that the *y* component of the center of gravity of the search model is arbitrary. Therefore, the search for a maximum of TF could be reduced to the asymmetric unit ($x = 0{\cdot}0$–$0{\cdot}5$, $y = 0{\cdot}0$, $z = 0{\cdot}0$–$0{\cdot}5$, in fractional units). The sampling interval was 0·01 in fractional units. The CPU time for (*a*) as well as (*b*) was 710 s on a Cray-YMP.

Neither the translation searches, nor the packing function, nor the correlation coefficient PC predict the correct orientation for search model S_3 (Fig. 3*b*). This implies that even a full six-dimensional search could not have solved the problem. In the following we address the question of whether PC refinement of the atomic coordinates against E_{tot} [(7)] could be used to locate the correct peak. We show in Table 1 that refinement against E_{tot} is meaningful if sufficient high-resolution data are used. If PC refinement of search model S_3 is started in the correct orientation the atoms move closer to the S_{HT} structure, whereas they move away from the S_{HT} structure when starting in a wrong orientation. This effect is most pronounced at 2 Å resolution. Furthermore, the total energy E_{tot} is higher, the empirical energy E_i is higher, and PC is lower when starting in the wrong orientation.

As a control we carried out a PC refinement starting with the correctly oriented S_{HT} structure (Table 1). As expected, this structure yields the highest PC value, the lowest empirical energy E_i, and the lowest total energy E_{tot} of all the structures listed in Table 1. However, the S_{HT} structure deviates by a small amount from the initial structure after PC refinement. This is due to the lower resolution [2·5 Å as opposed to 1·5 Å in Hendrickson & Teeter (1981)] and to missing information in PC about symmetry-related molecules.

It is sometimes important to carry out a fine-grid search around the highest peaks of the rotation function since the translation function critically depends on the accuracy of the orientation parameters (*e.g.* Wang, Bode & Huber, 1985). Rigid-body PC refinement is essentially equivalent to a fine-grid rotation-function search around the selected orientation. Rigid-body PC refinement is in fact more efficient than a fine-grid search since it only explores the shifts of the orientation parameters that improve the agreement between the calculated and observed diffraction data. However, as Fig. 4 shows, rigid-body PC refinements of search model S_3 followed by translation searches failed to determine the correct orientation.

PC refinements of individual atomic coordinates of search model S_3 at 3·0, 2·5 and 2·0 Å resolution yielded the lowest total energy E_{tot} for the correct orientation (Fig. 5). The correctly placed and PC-refined search model S_3 was used to solve the crambin crystal structure. Fig. 6 is a contour plot of the translation function and the packing function for search model S_3 after PC refinement at 2 Å resolution. The packing function is maximal along the 'ridge' between $x = 0{\cdot}2$ and $x = 0{\cdot}3$, and it falls off monotonically on both sides of the ridge. There are only a few maxima, including the correct one, of the translation function that are located on the ridge. The maximum of the translation function which corresponds to the correct position has a signal-to-noise ratio of 1·5 (Table 2).

Table 1. *Comparison of PC refinements for crambin*

PC refinements of the specified structures with the target E_{tot} consisting of rigid-body PC refinement (20 steps conjugate gradient minimization), followed by PC refinement of individual atomic coordinates (400 steps conjugate gradient minimization), at the specified resolution. The weight W_{PC} was set to 10 000 kJ mol^{-1}.

Target	Resolution range (Å)*	R.m.s. difference from S_{HT} All	(C^α, C, N)	5-30(C^α, C, N)	PC	E_{tot}	E_i (kJ mol^{-1})
Initial†	15-3	2·33	1·95	0·68	0·07		11 350·0
S_{HT} in correct orientation§							
E_{tot}	15-2·5	0·53	0·38	0·34	0·63	15 109·3	−453·0
S_3 in correct orientation§							
E_{tot}	15-5	2·43	1·96	1·06	0·93	3468·3	704·6
E_{tot}	15-4	2·15	1·59	1·11	0·85	7730·5	1373·7
E_{tot}	15-3	2·21	1·82	0·83	0·67	15 826·9	1938·5
E_{tot}	15-2·5	2·35	1·97	0·65	0·57	20 228·1	1622·4
E_{tot}	15-2	2·37	1·93	0·43	0·44	24 430·0	1121·2
S_3 in wrong orientation¶							
E_{tot}	15-5	2·42	1·92	1·13	0·94	3734·2	1189·5
E_{tot}	15-4	2·50	2·00	1·22	0·79	11 649·4	2636·8
E_{tot}	15-3	2·49	1·99	0·93	0·63	17 762·1	2336·2
E_{tot}	15-2·5	2·60	2·12	0·98	0·53	21 639·1	2023·5
E_{tot}	15-2	2·42	1·99	0·83	0·41	26 770·4	1934·9

* Resolution range for which the correlation coefficient PC(Ω) is computed.
† This is the initial structure S_3 without PC refinement.
§ Eulerian angles ($\theta_1 = 339{\cdot}2$, $\theta_2 = 75$, $\theta_3 = 240{\cdot}6°$).
¶ Eulerian angles ($\theta_1 = 152{\cdot}7$, $\theta_2 = 90$, $\theta_3 = 166{\cdot}2°$).

Crystallographic SA refinement (Brünger, 1988*b*) readily refined the correctly oriented PC-refined and positioned search model S_3 to an R factor of 34% at 2 Å resolution with good geometry (Table 3). Electron density maps were clearly interpretable and indicated regions where changes were necessary to refine the structure fully. We considered the solution of crambin through PC refinements of search model S_3 in selected orientations as completed for the purposes of this report.

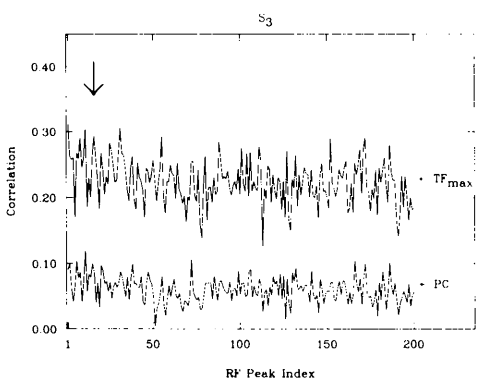


Fig. 4. Combined rigid-body PC refinements and translation searches for each orientation of crambin search model S_3 corresponding to the 200 highest selected peaks of the RF. Shown are the correlation coefficients (PC) after rigid-body PC refinement and the maximum value of the translation function TF_{max} *vs* RF peak index. The correct orientation is marked by an arrow. The PC refinement consisted of 20 steps rigid-body conjugate gradient minimization of E_{tot} at 15-2·5 Å resolution. The grid size for the FFT evaluation of the structure factors was set to one third of the high-resolution limit. The translation searches were carried out with the rigid-body refined search models at 15-3·0 Å resolution. The parameters for the translation searches are described in Fig. 3. The CPU time for these calculations was 1670 s on a Cray-YMP.

Table 2. *Signal-to-noise ratio of the translation function for crambin*

Translation searches of the specified search models of crambin in the correct orientation ($\theta_1 = 339{\cdot}2$, $\theta_2 = 75$, $\theta_3 = 240{\cdot}6°$) were carried out at 15-3 Å resolution using a grid with a spacing of 0·01 in fractional coordinates. The peak heights are expressed as the number of standard deviations above the mean. The S/N is calculated as the peak-height ratio of the correct peak to the highest error peak.

Structure	Correct peak (σ)	Error peak (σ)	S/N
S_{HT}	7·3	3·6	2·0
S_3	4·1	3·1	1·3
S_3, PC refined at 5·0 Å	3·7	2·8	1·3
S_3, PC refined at 4·0 Å	2·5	3·3	0·8
S_3, PC refined at 3·0 Å	2·9	2·2	1·3
S_3, PC refined at 2·5 Å	2·9	2·6	1·1
S_3, PC refined at 2 Å	3·8	2·5	1·5

B. Myoglobin

Metmyglobin room-temperature diffraction data at 1·5 Å resolution and a refined structure (referred to as M_{FPT}) were used (Frauenfelder, Petsko & Tsernoglou 1979). Myoglobin is composed of 153 amino acids and a heme group. It contains eight α helices (*A*: 3-18, *B*: 20-35, *C*: 36-42, *D*: 51-57, *E*: 58-77, *F*: 86-94, *G*: 100-118, *H*: 125-148) connected by various β turns and loops. Myoglobin has been studied previously by Nordman (1972) to investigate whether individual α helices can be located by rotation and translation searches.

For the purposes of this study a myoglobin search model was generated consisting of all atoms of the eight α helices without the β turns, loops, and the heme group. This comprised about 75% of the M_{FPT} crystal structure. The search model was made inaccurate by rotating the eight α helices of the search model

Table 3. *SA refinement of structure* S_3

The starting structure is the correctly oriented and positioned search model S_3 after PC refinement at 2 Å resolution ($\theta_1 = 339{\cdot}2$, $\theta_2 = 75$, $\theta_3 = 240{\cdot}6°$, $x = 0{\cdot}275$, $z = 0{\cdot}195$). SA refinement was carried out using a slow-cooling protocol: 60 steps C^α-restrained conjugate gradient minimization, followed by molecular dynamics starting at 4000 K and ending at 300 K over a period of 3·7 ps, followed by 120 steps of conjugate gradient minimization.

	Resolution range (Å)	*R* factor	Δ_{bonds} (Å)	Δ_{angles} (°)	R.m.s. difference from S_{HT} All	(C^α, C, N)	5-30(C^α, C, N)
Initial	10-2·5	0·49	0·033	6·56	2·37	1·93	0·43
SA refined	10-2	0·34	0·021	4·24	1·86	1·46	0·18

by angles between 0 and 16° around certain rotation axes. The rotation axes were placed through the centers of gravity of each α helix. The rotation axis slopes were chosen arbitrarily [the slopes are (3, 4, 1) for α helix *A*, (1, 2, 1) for *B*, (−1, 4, −2) for *C*, (3, −4, −5) for *D*, (3, −4, 1) for *E*, (3, −4, −1) for *F*, (−1, 4, 1) for *G* and (3, 4, −1) for *H*]. All reflections were weighted equally. The atomic *B* factors of all atoms were set to 15·0 Å^2.

Fig. 7 and Table 4 show the differences of the myoglobin search models M_4, M_{10}, M_{13} and M_{15} from the myoglobin crystal structure M_{FPT}. The myoglobin search models have only a few close contacts that are introduced by the α-helix rotations. All these contacts could easily be relieved by moving a few side-chain atoms. Therefore, each myoglobin search model has a chemically reasonable conformation.

Table 4 illustrates the radius of convergence for rigid-group PC refinement at 15-3 Å resolution. It consisted of conjugate gradient minimization of the three rotational and translational parameters of the eight α helices against the target function E_{tot} [(10)]. It appears that, up to an α-helix tilt angle of 13°, PC

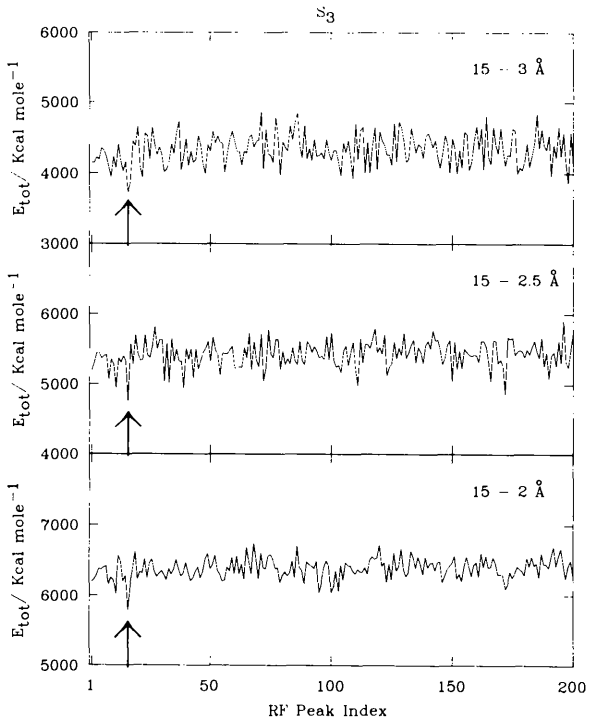


Fig. 5. Individual-atom PC refinements for each orientation of crambin search model S_3 corresponding to the 200 highest selected peaks of the RF. Shown is the total energy E_{tot} after PC refinement *vs* RF peak index. The correct orientation is marked by an arrow. The individual-atom PC refinement consisted of 20 steps rigid-body conjugate gradient minimization of E_{tot} followed by 400 steps individual-atom conjugate gradient minimization of E_{tot} at 15-3, 15-2·5 and 15-2 Å resolution. The grid size for the FFT evaluation of the structure factors was set to one third of the high-resolution limit. The weight W_{PC} in equation (8) was set to 41 868 kJ mol^{-1}. The CPU time for the individual-atom PC refinements was 7, 9·5 and 13 h on a Cray-YMP at 15-3, 15-2·5 and 15-2 Å resolution, respectively.

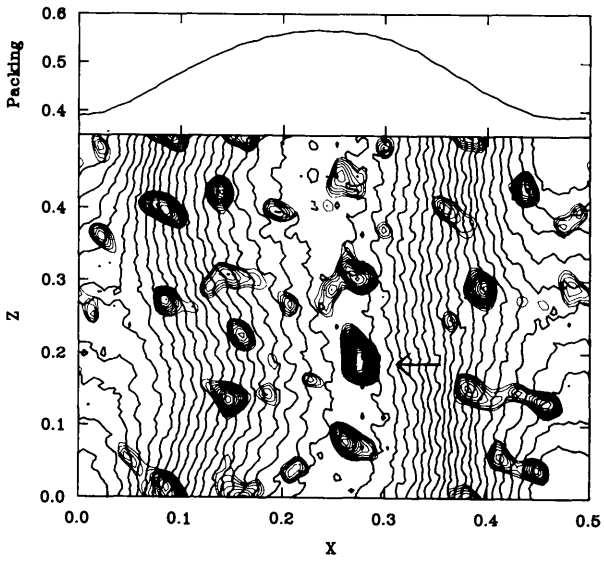


Fig. 6. Contour plot of the translation function TF (thin lines) and the packing function (thick lines) for the correctly oriented crambin search model S_3 after individual-atom PC refinement at 15-2 Å resolution. The translation search itself used data at 15-3 Å resolution. An asymmetric unit in translation space was computed. The sampling interval was 0·005 in fractional units. The mean of the translation function is 0·28 and σ is 0·034. The maximum (0·41) is marked by an arrow and corresponds to the correct position. The lowest contour of the translation function is drawn at 1σ above the mean and higher contours are drawn at intervals of 0·005. The packing function was evaluated on a 1 Å grid. The maximum of the packing function is 0·58 and the minimum is 0·39. The lowest contour of the packing function is drawn at the minimum and higher contours are drawn at intervals of 0·01. Also shown is the $Z = 0{\cdot}0$ section of the packing function. The CPU time for the translation and packing calculations was 326 s on a Cray-YMP.

Table 4. *Comparison of PC refinements for myoglobin*

PC refinements with the target E_{PC} consisting of rigid-body PC refinement (20 steps conjugate gradient minimization) followed by rigid-group PC refinement of the placement (*i.e.* rotation and translation) of the eight α helices (400 steps conjugate gradient minimization) at 15-3 Å resolution.

Structure	α-helix tilt angle (°)	R.m.s. difference (Å)¶		PC		
		Initial*	Final§	Initial*	Rigid†	Final§
	0	0	0·26	0·30	0·30	0·32
	2	0·24	0·26	0·28	0·29	0·32
M_4	4	0·48	0·27	0·23	0·24	0·32
	6	0·72	0·27	0·18	0·19	0·32
	8	0·96	0·26	0·13	0·14	0·32
M_{10}	10	1·19	0·26	0·09	0·10	0·32
	12	1·43	0·29	0·06	0·07	0·32
M_{13}	13	1·55	0·31	0·05	0·06	0·31
	14	1·67	1·64	0·04	0·05	0·18
M_{15}	15	1·80	1·68	0·04	0·05	0·18
	16	1·91	2·14	0·03	0·03	0·18

* 'Initial' refers to the myoglobin structure with the eight α helices tilted by the specified angle.
† 'Rigid' refers to the structure after rigid-body PC refinement.
§ 'Final' refers to the structure after rigid-body PC refinement followed by rigid-group PC refinement of the placement of the eight α helices.
¶ R.m.s. difference from the myoglobin crystal structure.

Table 5. *Signal-to-noise ratio of the translation function for myoglobin*

Translation searches of the PC-refined search models of myoglobin were carried out at 15-3 Å resolution using a grid with a spacing of 0·01 in fractional coordinates. The peak heights are expressed as the number of standard deviations above the mean. The signal-to-noise ratio (S/N) is calculated as the peak height ratio of the correct peak to the highest error peak.

Structure	Initial*			Rigid†			Final§		
	Correct	Error	S/N	Correct	Error	S/N	Correct	Error	S/N
M_4	8·04	2·00	4·02	8·36	1·86	4·50	8·19	2·31	3·50
M_{10}	5·00	2·50	2·00	5·84	2·00	2·92	7·94	2·31	3·40
M_{13}	3·50	3·13	1·12	4·06	2·39	1·70	7·71	2·36	3·27
M_{15}	2·81	3·00	0·94	2·59	2·65	0·98	3·39	2·65	1·28

* 'Initial' refers to the myoglobin structure with the eight α helices tilted by the specified angle.
† 'Rigid' refers to the structure after rigid-body PC refinement.
§ 'Final' refers to the structure after rigid-body PC refinement followed by rigid-group PC refinement of the placement of the eight α helices.

refinement returns the α helices to their original positions. The resulting structures are very close to that obtained by PC refinement of the crystal structure M_{FPT}. The maximum deviation of the refined structures from M_{FPT} for C^{α} atoms is 0·6 Å. For tilt angles between 14 and 15°, PC refinement reduces the initial r.m.s. difference somewhat but it fails to return the α helices completely to their original positions. This is corroborated by the reduced value of the PC correlation coefficient (0·18) compared with 0·32 for the converged structures. This suggests that the applicability of PC refinement is limited by the radius of convergence of the minimization procedure employed. The ability of PC refinement to return the α helices to their original positions is also reflected in the improvement of the signal-to-noise (S/N) ratio of translation searches (Table 5); for example, the S/N ratio for search model M_{13} improves by 2·15 after rigid-group PC refinement. Even in the case of M_{15} the S/N ratio is improved (initially 0·98, then 1·28 after rigid-group PC refinement).

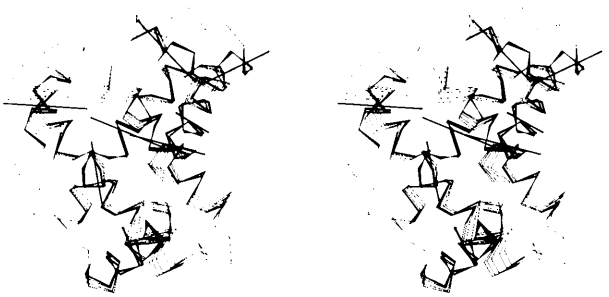
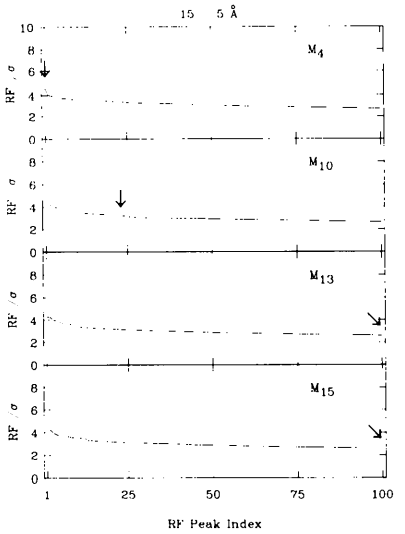
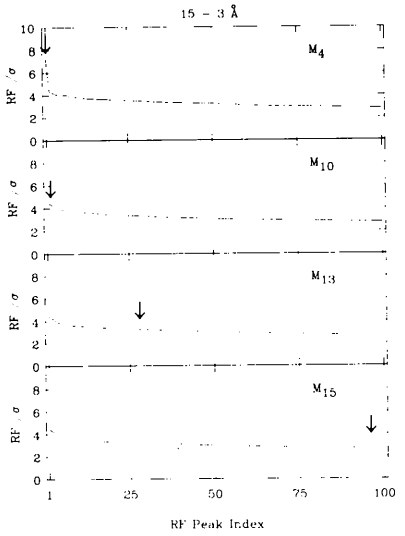


Fig. 7. C^{α}-backbone stereo plot of the myoglobin X-ray structure M_{FPT} (thin lines), the four myoglobin search models M_4, M_{10}, M_{13}, M_{15} (thin lines), and the rotation axis (thick lines) superimposed on it.

Figs. 8-11 show the results of rotation searches, translation searches, combined rigid-body PC refinements and translation searches, and rigid-group PC refinements of the selected RF peaks using the myoglobin search models M_4, M_{10}, M_{13}, and M_{15}. The results can be summarized as follows. The RFs at 15-5 and at 15-3 Å resolution failed to determine



(a)



(b)

Fig. 8. Rotation searches for the myoglobin search models M_4, M_{10}, M_{13} and M_{15} at (*a*) 15-5 Å resolution and (*b*) 15-3 Å resolution. Shown are the values of the 100 highest selected peaks of the RF sorted by peak height. The correct orientation is indicated by an arrow (in the case of search models M_{13} and M_{15} at 15-5 Å resolution the correct orientation is not among the first 100 peaks). The RF values are shown in units of standard deviations (σ) above the mean. The observed Patterson map was computed from the observed intensities by FFT with a grid size of 1·25 and 0·84 Å for (*a*) and (*b*), respectively. The model Patterson maps were computed by placing the search model into an orthorhombic box with 100 Å cell edges, evaluating the structure factors and FFT of the squared amplitudes with a grid size of 1·25 and 0·84 Å for (*a*) and (*b*), respectively. Model Patterson vectors were selected according to length (between 30 and 4 Å), and according to peak height (2σ above the mean of the model Patterson map). The interval Δ for θ_2 was set to 5°. The rotation search was carried out in the asymmetric unit $\theta_+ = 0$-$720°$, $\theta_2 = 0$-$90°$, $\theta = 0$-$360°$. RF peaks were selected by checking the cluster condition $m(\Omega^1, \Omega') < 0{\cdot}2$ [equation (3)]. The CPU time for each rotation search was 223 and 696 s on a Cray-YMP for (*a*) and (*b*), respectively.

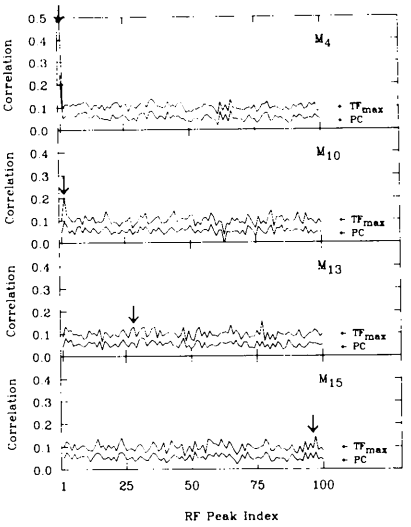


Fig. 9. Translation searches for each orientation of the myoglobin search models M_4, M_{10}, M_{13} and M_{15} corresponding to the 100 highest selected peaks of the RF. Data were included between 15 and 3 Å. Shown are the correlation coefficients (PC) and the maximum value of the translation function TF_{max}. The correct orientation is marked by an arrow. The search for a maximum of the TF was carried out in the asymmetric unit ($x = 0{\cdot}0$-$0{\cdot}5$, $y = 0{\cdot}0$, $z = 0{\cdot}0$-$0{\cdot}5$, in fractional units). The sampling interval was 0·02 in fractional units. The CPU time for the set of translation searches was 900 s on a Cray-YMP.

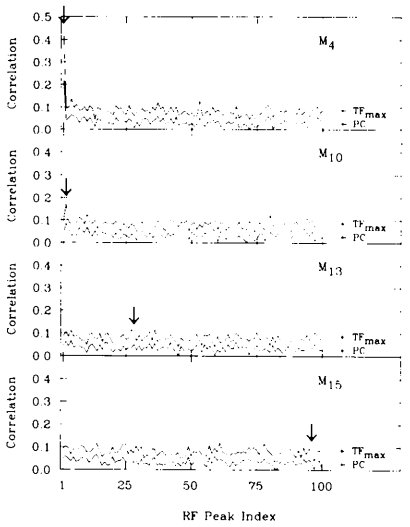


Fig. 10. Combined rigid-body PC refinements and translation searches for each orientation of the myoglobin search models M_4, M_{10}, M_{13} and M_{15} corresponding to the 100 highest selected peaks of the RF. Shown are the correlation coefficients (PC) after rigid-body PC refinement and the maximum value of the translation function TF_{max} *vs* RF peak index. The correct orientation is marked by an arrow. The PC refinement consisted of 20 steps rigid-body conjugate gradient minimization of E_{tot} at 15-3 Å resolution. The grid size for the FFT evaluation of the structure factors was set to one third of the high-resolution limit. The translation searches were carried out with the rigid-body refined search models at 15-3·0 Å resolution. The parameters for the translation searches are described in Fig. 3. The total CPU time for each set of PC refinements and translation searches was 1 h on a Cray-YMP.

the correct orientation for tilt angles greater than or equal to 10° (Fig. 8). The 100 selected peaks of the RF at 15-3 Å were used for further analysis. The translation searches for the selected RF peaks determined the correct orientation up to a tilt angle of 10° but they failed for larger tilt angles (Fig. 9). Rigid-body PC refinement of the orientation and position of the myoglobin search models combined with translation searches did not improve this result (Fig. 10). As in the crambin case there is a large correlation between the PC and TF_{max} graphs (the standard linear correlation coefficient is 0·95, 0·84, 0·82 and 0·84 for search models M_4, M_{10}, M_{13} and M_{15}, respectively). In particular, PC alone without subsequent translation searches could have been used to predict the correct orientation in Figs. 9 and 10.

Rigid-group PC refinement of the α-helical positions and orientations determines the correct orientation up to a tilt angle of 13° (Fig. 11) which is the radius of convergence of the conjugate gradient minimization (see above). This proves, at least in this case, that PC refinement is a superior method compared with six-dimensional searches to locate the correct peak of a noisy rotation function. The computations for the PC refinements in Fig. 11 require 15 times more CPU time than the translation searches in Fig. 9. However, it should be noted that the translation searches were only two-dimensional because of the low symmetry of the space group $P2_1$. Therefore, the comparison of the CPU requirements would be more favorable for PC refinement in the case of higher-symmetry space groups where three-dimensional translation searches would be required.

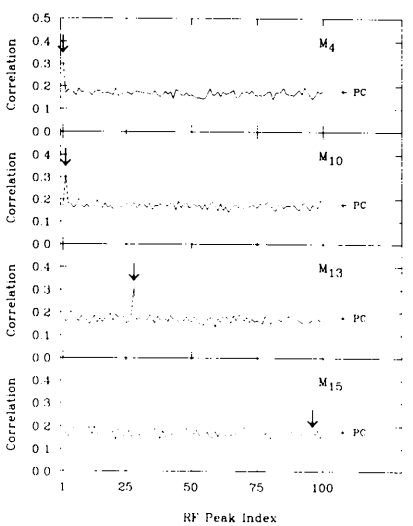


Fig. 11. Rigid-group PC refinements for each orientation of the myoglobin search models M_4, M_{10}, M_{13} and M_{15} corresponding to the 100 highest selected peaks of the RF. Shown is the total energy E_{tot} after PC refinement *vs* RF peak index. The correct orientation is marked by an arrow. The individual-atom PC refinement consisted of 20 steps rigid-body conjugate gradient minimization of E_{tot} followed by 120 steps rigid-group conjugate gradient minimization of E_{tot} at 15-3 Å resolution. The grid size for the FFT evaluation of the structure factors was set to one third of the high-resolution limit. The CPU time for each PC refinement was 3·5 h on a Cray-YMP.

Concluding remarks

Advances in computer technology have made it possible to sample orientations and positions of a search model in a much more complete manner than previously possible. These MR search procedures are intrinsically well suited for vectorization and parallelization on multiprocessor computer architectures. We propose a new strategy here which consists of incorporating internal flexibility of the search model into MR. All sampled orientations are sorted with respect to their RF value and then a large number of the highest peaks are selected for PC refinements followed by translation searches.

The use of the correlation coefficient PC makes it possible to filter the peaks of a rotation search. We provided evidence that PC suppresses noise peaks in the original rotation function and enhances the correct peak. Furthermore, we showed in two examples that individual atomic coordinates or rigid-group coordinates can be refined against PC. The target function E_{tot} of the PC-refined search model as a function of orientation assumed a minimum for the correct orientation. Furthermore, in one example, the S/N ratio of the translation function of the PC-refined model was improved. PC refinement is a superior method to six-dimensional searches in locating the correct peak of a noisy rotation function.

We expect that PC refinement of generalized coordinates such as bending or deformation parameters of secondary structural elements (Barlow & Thornton, 1988) will further extend the applicability of the method (work in progress). At present, PC refinement is limited by the radius of convergence of the minimization procedure employed. We expect that the use of simulated annealing or other modern non-linear optimization procedures could increase the radius of convergence and extend the applicability of MR even further.

The author thanks R. Huber, G. A. Petsko and L. Howell for stimulating discussions, M. M. Teeter for providing the X-ray data of crambin, G. A. Petsko for providing the X-ray data of metmyoglobin, and W. Steigemann for providing the *PROTEIN* program, and he acknowledges support by the Pittsburgh Supercomputer Center of the National Science Foundation (grant no. DMB890008P).

References

BALDWIN, J. M. (1980). *J. Mol. Biol.* **136**, 103-128.
BARLOW, D. J. & THORNTON, J. M. (1988). *J. Mol. Biol.* **201**, 601-619.

BRÜNGER, A. T. (1988*a*). *X-PLOR Manual, Version* 1.5. Yale Univ., New Haven, USA.
BRÜNGER, A. T. (1988*b*). *J. Mol. Biol.* **203**, 803-816.
BRÜNGER, A. T. (1989). *Acta Cryst.* A**45**, 42-50.
BRÜNGER, A. T., CAMPBELL, R. L., CLORE, G. M., GRONENBORN, A. M., KARPLUS, M., PETSKO, G. A. & TEETER, M. M. (1987). *Science*, **235**, 1049-1053.
BRÜNGER, A. T., CLORE, G. M., GRONENBORN, A. M. & KARPLUS, M. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 3801-3805.
BRÜNGER, A. T., KARPLUS, M. & PETSKO, G A. (1989). *Acta Cryst.* A**45**, 50-61.
CHOTHIA, C. & LESK, A. M. (1986). *EMBO J.* **5**, 823-826.
CLORE, G. M., BRÜNGER, A. T., KARPLUS, M. & GRONENBORN, A. M. (1986). *J. Mol. Biol.* **191**, 523-551.
FEHLHAMMER, H. & BODE, W. (1975). *J. Mol. Biol.* **98**, 683-692.
FRAUENFELDER, H., PETSKO, G. A. & TSERNOGLOU, D. (1979). *Nature (London)*, **280**, 558-563.
FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.
HAUPTMAN, H. (1982). *Acta Cryst.* A**38**, 289-294.
HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107-112.
HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* A**32**, 778-780.
HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.
HUBER, R. (1985). In *Molecular Replacement.* Proc. of the Daresbury Study Weekend, Daresbury, February 1985, pp. 58-61. SERC Daresbury Laboratory, Daresbury, Warrington, England.
KARPLUS, M. & MCCAMMON, J. A. (1983). *Annu. Rev. Biochem.* **52**, 263-300.
LATTMAN, E. E. (1972). *Acta Cryst.* B**28**, 1065-1068.
LATTMAN E. E. (1985). *Methods Enzymol.* **115**, 55-77.
NORDMAN, C. E. (1972). *Acta Cryst.* A**28**, 134-143.
NORDMAN, C. E. (1980). *Acta Cryst.* A**36**, 747-754.
POWELL, M. J. D. (1977). *Math. Programming*, **12**, 241-254.
RABINOVICH, D. & SHAKKED, Z. (1984). *Acta Cryst.* A**40**, 195-200.
RAO, S. N., JIH, J.-H. & HARTSUCK, J. A. (1980). *Acta Cryst.* A**36**, 878-884.
ROSSMANN, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
SCHIRMER, T., HUBER, R., SCHNEIDER, M., BODE, W., MILLER, M. & HACKERT, M. L. (1986). *J. Mol. Biol.* **188**, 651-676.
SHAKKED, Z., RABINOVICH, D., CRUSE, W. B. T., EGERT, E., KENNARD, O., SALA, G., SALISBURY, S. A. & VISWAMITRA, M. A. (1981). *Proc. R. Soc. London Ser. B*, **213**, 479-487.
STEIGEMANN, W. (1974). PhD Thesis. Technische Univ. München, Federal Republic of Germany.
SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800-804.
TAYLOR, W. R. (1988). *Protein Eng.* **2**, 77-86.
WANG, D., BODE, W. & HUBER, R. (1985). *J. Mol. Biol.* **185**, 595-624.
WÜTHRICH, K. (1986). *NMR of Proteins and Nucleic Acids.* New York: Wiley.

Acta Cryst. (1990). A**46**, 57-68

Direct Phase Determination for the Molecular Envelope of Tryptophanyl-tRNA Synthetase from *Bacillus stearothermophilus* by X-ray Contrast Variation

BY CHARLES W. CARTER JR, KATHERINE V. CRUMLEY, DAVID E. COLEMAN AND FRANK HAGE

Department of Biochemistry, CB No. 7260, *University of North Carolina at Chapel Hill, Chapel Hill, NC* 27599-7260, *USA*

AND GERARD BRICOGNE

LURE, Bâtiment 209*D*, 91405 *Orsay, France*

(*Received* 7 *April* 1989; *accepted* 21 *August* 1989)

Abstract

Monoclinic crystals of *Bacillus stearothermophilus* tryptophanyl-tRNA synthetase grown in the presence of substrate trytophan (space group $P2_1$) display evidence of a low-resolution trigonal space group ($P321$). The origin and averaging transformations for the local 32 point group of this unusually clear sixfold non-crystallographic symmetry may be inferred without prior estimation of the electron density. This local symmetry was exploited in conjunction with solvent density contrast variation to determine the shape of the molecular envelope. X-ray intensities measured from crystals equilibrated in mother liquors of three different electron densities were used to estimate three parameters for each reflection: the modulus of the envelope transform, $|G_\mathbf{h}|$; and components, $X_\mathbf{h}$ and $Y_\mathbf{h}$, relative to $G_\mathbf{h}$, of the structure-factor vector for the transform of intramolecular density fluctuations. The moduli $\{|G_\mathbf{h}|\}$ behave somewhat like structure-factor amplitudes from small-molecule crystals, and estimation of their unknown phases was successfully carried out by statistical direct methods. Reflections to 18 Å resolution, which obey rather well the symmetry of space group $P321$, were merged to produce an asymmetric unit in that space group. $|G_\mathbf{h}|$ values for the 34 strongest of these were phased using the small-molecule direct-methods package *MITHRIL* [Gilmore (1984). *J. Appl Cryst.* **17**, 42-46]. The best phase set was expanded back to the $P2_1$ lattice and negative density was truncated to generate initial phases for all reflections to 18 Å

0108-7673/90/010057-12$03.00