ions are parameters that establish the almost fully ionic character of these compounds. The crystalline-field effect causes a contraction of the anions and a slight expansion of the cation. In the absence of recent and precise experimental data on structure factors, the values reported in this study may be used as a reference. They will be refined in the future by the exploration of the importance of polarization functions on the S^{2-} ion in order to improve the agreement factor for Li_2S and to calculate its deformation accurately by the rigorous calculation of scattering factors by Fourier transform of the electron density on a 'per energy band occupied' basis and by taking thermal agitation into account.

The authors wish to express their gratitude to Professor C. Pisani for his helpful suggestions. The calculations were carried out on the IBM-3090/600-VF of the Centre National Universitaire Sud de Calcul (CNUSC), which is gratefully acknowledged.

References

CAUSA', M., DOVESI, R., PISANI, C. & ROETTI, C. (1986*a*). *Phys. Rev. B*, **33**, 1308-1316.
CAUSA', M., DOVESI, R., PISANI, C. & ROETTI, C. (1986*b*). *Acta Cryst.* B**42**, 247-253.
CLEMENTI, E. (1963). *J. Chem. Phys.* **39**, 175-179.
DOVESI, R. (1985). *Solid State Commun.* **54**, 183-185.
DOVESI, R., PISANI, C., RICCA, F., ROETTI, C. & SAUNDERS, V. R. (1984). *Phys. Rev. B*, **30**, 972-979.
DOVESI, R., PISANI, C., ROETTI, C., CAUSA', M. & SAUNDERS, V. R. (1988). *CRYSTAL* 88. *QCPE*, Program no. 577. Indiana Univ., Bloomington, Indiana, USA.
DOVESI, R., ROETTI, C., FREYRIA-FAVA, C., PRENCIPE, M. & SAUNDERS, V. R. (1991). *Chem. Phys.* **156**, 11-19.
HOLBROOK, J. B., SABRY-GRANT, R., SMITH, B. C. & TANDEL, T. V. (1990). *J. Chem. Educ.* **67**, 304-307.
HUZINAGA, S. & HART-DAVIS, A. (1973). *Phys. Rev. A*, **8**, 1734-1738.
International Tables for Crystallography (1992). Vol. C, p. 488. Dordrecht: Kluwer Academic Publishers.
LICHANOT, A., GELIZE, M., LARRIEU, C. & PISANI, C. (1991). *J. Phys. Chem. Solids*, **52**, 1155-1164.
LUAÑA, V., RECIO, J. M. & PUEYO, L. (1990). *Phys. Rev. B*, **42**, 1791-1801.
O'KEEFFE, M. (1981). *Structure and Bonding in Crystals*, edited by M. O'KEEFFE & A. NAVROTSKY, Vol. I, p. 304. New York: Academic Press.
PANDEY, R. & VAIL, J. M. (1989). *J. Phys. Condens. Matter*, **1**, 2801-2820.
PANTELIDES, S. T., MICKISH, D. J. & KUNZ, A. B. (1974). *Phys. Rev. B*, **10**, 5203-5212.
PISANI, C., DOVESI, R. & ROETTI, C. (1988). *Hartree-Fock Ab Initio Treatment of Crystalline Systems*. Berlin: Springer-Verlag.
SANGER, P. L. (1969). *Acta Cryst.* A**25**, 694-702.
SCHMIDT, P. C., SEN, K. D. & WEISS, A. (1980). *Ber. Bunsenges. Phys. Chem.* **84**, 1240-1251.
SCHMIDT, P. C. & WEISS, A. (1979). *Z. Naturforsch. Teil A*, **34**, 1471-1481.
SCHWARZ, K. & SCHULZ, H. (1978). *Acta Cryst.* A**34**, 994-999.
SHANNON, R. D. (1981). *Structure and Bonding in Crystals*, edited by M. O'KEEFFE & A. NAVROTSKY, Vol II, pp. 67 & 55. New York: Academic Press.
SUZUKI, T. (1960). *Acta Cryst.* **13**, 279.
TOKONAMI, M. (1965). *Acta Cryst.* **19**, 486.
WATSON, R. E. (1958). *Phys. Rev.* **111**, 1108-1110.
YAMASHITA, J. & KOJIMA, M. (1952). *J. Phys. Soc. Jpn*, **7**, 261-263.

Acta Cryst. (1993). A**49**, 97-104

PRISM: Automated Crystallographic Phase Refinement by Iterative Skeletonization

BY CHARLES WILSON* AND DAVID A. AGARD

Howard Hughes Medical Institute, Department of Biochemistry and Biophysics, and Graduate Group in Biophysics, University of California at San Francisco, San Francisco, CA 94143-0448, *USA*

(*Received* 6 *May* 1991; *accepted* 22 *June* 1992)

Abstract

A phase-refinement procedure based on iterative skeletonization of electron density maps is presented. As with traditional solvent-flattening methods, refinement alternates between real-space and reciprocal-space representations of the scattering density. A pseudoatom list derived from the modified skeleton of an initial electron density map provides calculated structure-factor amplitudes and phases. Recombination with the observed F_{obs} values yields a new map that can serve as the starting point for another round of skeletonization. Tests using partial structures to provide starting crystallographic phases have shown this refinement procedure to have a significantly larger radius of convergence than solvent flattening.

Introduction

The procedure to find a solution to the crystallographic phase problem is generally the rate-limiting step in macromolecular structure determination by

* Current address: Department of Molecular Biology, Wellman 9, Massachusetts General Hospital, Boston, MA 02114, USA.

0108-7673/93/010097-08$06.00

X-ray diffraction (Blundell & Johnson, 1976). All crystallographic protein structures have relied upon a limited number of experimental or computational methods to obtain phase estimates before building a complete atomic model. These techniques, including isomorphous replacement, anomalous scattering and molecular replacement, often yield poor initial estimates of the phases, resulting in electron density maps that are difficult to interpret. The development of techniques capable of improving phases in the absence of a complete atomic model should dramatically increase the rate of structure determination by X-ray crystallography and minimize problems such as incorrect chain tracing that have plagued several recent structures (Branden & Jones, 1990).

The method of density modification (also known as solvent flattening) has been used in a number of cases to maximize the information available from a single derivative or from anomalous-scattering data (see, for example, Chen *et al.* 1991; Tulinsky, Park & Skrzypczak-Jankun, 1988; Messerschmidt *et al.* 1989). Density modification is an iterative procedure consisting of two distinct steps (Wang, 1985). Initially, an electron density map (calculated using the current phase estimates) is filtered such that the density of all regions lying outside a molecular boundary is set to some mean value and often a positivity constraint is also applied within the molecular boundary. Subsequent transformation of the filtered map yields a new set of crystallographic phases that can be recombined with the initial phases to produce an 'improved' electron density map for the next cycle of refinement. While often successful, density modification has a limited radius of convergence, especially when the solvent content of a crystal is low. In addition, since the procedure always converges to some solution, it is sometimes impossible to tell whether or not solvent flattening has improved the phases.

We have sought to develop new methods for phase refinement that can tolerate significantly larger errors in the starting phases. Our strategy has been to use the same cycle of real-space–reciprocal-space refinement, but to apply considerably stronger constraints on the real-space representation of the scattering. Our refinement applies these constraints by first skeletonizing the electron density map and then forcing this skeleton to adopt 'protein-like' characteristics. To evaluate our method, we have developed a number of test cases in which the starting phase information comes from partial atomic models. In the examples we have considered, our refinement procedure is able to converge to the correct phase solution, allowing the complete model to be built into the electron density map. In the same tests, solvent flattening produces a marginal improvement in the phases but leaves significant biases and errors in the final density map.

Refinement method

Fig. 1 outlines the key steps to our iterative phase-refinement protocol, *PRISM* (phase refinement by iterative skeleton modification). The procedure uses a set of experimental structure factors (F_{obs}) and corresponding initial phase estimates (α_0) to start the refinement. In the test cases we have considered, the α_0 values are calculated from a partial model. In principle, more conventional sources such as SIR, MIR, anomalous scattering or molecular replacement using a more complete model could provide the starting phases. From F_{obs} and α_0, one can calculate a starting electron density map corresponding to the asymmetric unit in the unit cell. Each cycle of refinement uses an electron density map as input and ends with the synthesis of another map based upon (hopefully) improved phases. By repeating several cycles of refinement, the phases gradually converge to final predicted values. The steps in each cycle of refinement are described below.

Step 1. An asymmetric unit of the electron density map (calculated on an ~1 Å rectangular grid) is skeletonized using the *MKSKEL* program, part of the *GRINCH* package (Williams, 1982; adapted for VMS by M. Carson, University of Alabama, USA). This program outputs a listing of nodes, corresponding to all local maxima along the x, y or z axes in the electron density map, and a listing of the connections between nodes. The nodes represent pseudo-atoms in the structure, with the node connections defining the bonds between them.

Step 2. The *CONNECT* program (available upon request from CW) automatically modifies the skeleton output by *MKSKEL* such that the scattering density resembles a single chain rather than a group of disconnected atoms or groups of atoms. In contrast to the *BONES* program of Alwyn Jones, University of Uppsala, Sweden, no user intervention is required to define the backbone, thus allowing the skeleton to be used in a completely automated fashion. *CONNECT* initially identifies connected graphs of nodes and immediately deletes the smallest graphs (those with fewer than three nodes) from the model. Nodes with only a single bond to all other nodes (termed end-points) are then identified and used to generate connections between the remaining isolated graphs.

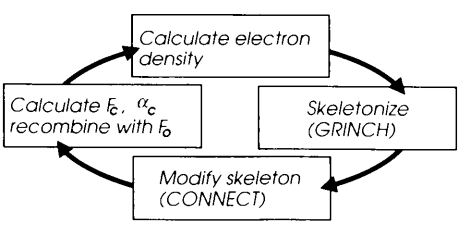

Fig. 1. Overview of the *PRISM* phase-refinement scheme. Details for each step are described in the text.

The connections formed between graphs are defined by a path of grid points in the electron density map. To force connections between graphs, endpoints are allowed to diffuse on the lattice of grid points in the unit cell subject to the cost function

$$C(i, k) = -\sum_j d_{ij}^{-1} + w_i \rho(k),$$

where $C(i, k)$ is evaluated for the ith endpoint at the kth grid point in the density map; d_{ij} is the minimum distance between the ith and jth endpoints (taking adjacent symmetry mates into account); w_i is an adjustable weighting parameter; and $\rho(k)$ is the electron density of the kth grid point. The summation over j endpoints includes all those that do not belong to the same graph as the ith endpoint. By minimizing the cost function, endpoints are forced to form connections (decreasing d_{ij}) while remaining in regions of high electron density [keeping $-\rho(k)$ small].

Minimization is done in a limited number of cycles with each endpoint moving a maximum of one grid point per cycle (thereby limiting the length of the connections that can be generated). For each endpoint in a given cycle, the cost function at all adjacent grid points is evaluated and the endpoint is moved to the neighboring position with the lowest cost. w_i is initially set such that the starting gradient is as close to zero as possible, thus balancing the density and the distance terms. If the current position has the lowest cost, w_i is dropped, removing the bias toward maintaining good electron density and favoring connection formation. If a pair of endpoints diffuse such that they lie on adjacent grid points, the two isolated graphs are grouped into a single larger graph and the pair of endpoints are removed from the minimization. Small graphs that remain unconnected to others after the maximum number of cycles of endpoint diffusion (those with fewer than six nodes) are deleted from the skeleton. The grid points defining the connections between graphs are added to the skeleton, which is then output as a list of atoms.

Step 3. The pseudoatom list output by *CONNECT* is used to calculate structure factors (F_c) and phases (α_1) for the next cycle (with the assumption of a carbon scattering curve for the pseudoatoms). The F_c are scaled to the observed structure factors (using zonal scaling).

Step 4. A new electron density map is calculated using the updated phases, α_1 values and modified Fourier coefficients, $F_w = 2wF_o - F_c$ (w = Sim's weight; Sim, 1960). Maps calculated using the F_w coefficients are ideal for this procedure since they include the true scattering vector at full scale while minimizing the systematic noise directed along α_1 (see Appendix). Alternative arguments for using these coefficients have been provided, and appropriate correction factors to account for experimental errors have been derived by Read (1986).

The ability of the refinement scheme to converge to the correct electron density map was estimated by several criteria. Because the refinement is done using calculated 'observed' data, the phase errors can be calculated exactly using the known structure. To estimate the accuracy of the phases, we have calculated the weighted phase error,

$$\Delta\Phi = \sum F_{obs} |\alpha_{obs} - \alpha_c| / \sum F_{obs},$$

where the summation is done over all reflections. The crystallographic R factor between the scaled calculated F values and the experimental F values,

$$R = \sum |F_{obs} - F_c| / \sum F_{obs},$$

provides a measure of the convergence and can also provide a crude measure of the error in the electron density map in cases in which the observed phases are not known.

Control experiments

The refinement procedure is based on the assumption that an electron density map can be accurately represented by the nodes in a skeleton. To confirm this assumption, the following test was performed. An electron density map (~1 Å grid spacing) was calculated for the first fifty residues of myoglobin using coordinates obtained from the Protein Data Bank entry 1MBD (Phillips, 1978). The space group and unit cell of apolipoprotein-E, another all-helical protein, was used for this and all other calculations ($P2_12_12_1$, $a = 41.26$, $b = 54.51$, $c = 87.09$ Å, $\alpha = \beta = \gamma = 90°$; see Wilson, Wardell, Weisgraber, Mahley & Agard, 1991). The electron density map for the myoglobin fragment was subsequently skeletonized using the *MKSKEL* program (Williams, 1982). The nodes in the skeleton were treated as pseudoatoms and used to obtain new structure factors and phases. Fig. 2 shows the zonal R factor between the correct

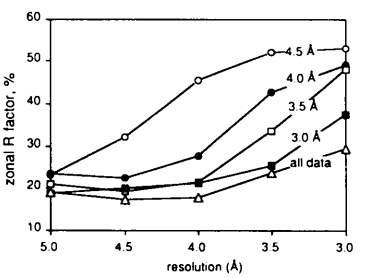

Fig. 2. Accuracy of the skeletonization procedure. Structure factors calculated from the atomic model for the first 50 residues of myoglobin were used to calculate a series of electron density maps with decreasing resolution. The maps were subsequently skeletonized using the *GRINCH MKSKEL* program. The unfiltered output (a list of local maxima in the density map) was used to calculate new structure factors (each local maximum was assumed to represent a dummy C atom). The zonal R factor is shown as a function of zonal resolution for each of the maps.

F values (obtained from the Fourier transform of the true electron density map) and those calculated from the corresponding skeleton. The errors in the skeleton model are a strong function of the resolution of the initial map used to generate the skeleton. If the starting map is calculated directly from the atomic coordinates (*i.e.* with infinite resolution), the zonal *R* factor remains below 25% for all data in the range 10–3 Å, but rises sharply at higher resolutions. A similar result is found using a 3 Å resolution starting map. If lower-resolution starting maps are used to calculate the skeleton, the zonal *R* factor remains at about 25% for data below the resolution of the map but rises to that expected for a random structure at higher resolutions (Fig. 2). If the correct map is used as the input to the *PRISM* procedure and refined for several cycles, the overall *R* factor (10–3.0 Å data) rises to 30% while the phase error rises to 27°, providing a best-case estimate for the final expected amplitude and phase errors.

Refinement tests with defined partial models

The *PRISM* method was initially tested using the 50-residue myoglobin fragment described above as a model unknown structure. Several structures were constructed to test different possible biases in the starting models. The first model contained only C_α atoms for all 50 residues of the myoglobin fragment. While this model has a small fraction of the total scattering density (50 atoms of the 417 making up the full structure), the electron density is well distributed and there are no regions that are systematically under-represented. A second model included all backbone and C_β atoms for the first 25 residues of the myoglobin fragment. While this half-backbone test case includes a higher fraction of the scattering density (122 atoms), the distribution is highly asymmetric with the entire C-terminal half missing from the initial map. To test the effect of measurement errors and coordinate errors on refinement, both the observed structure factors (*i.e.* those calculated for the full 50-residue model) and the starting myoglobin model coordinates were perturbed by the addition of a Gaussian distribution of random shifts.

The first test of the refinement procedure used the C_α-atom-only model of the myoglobin fragment as a starting partial model. An initial map calculated at 3 Å resolution using the observed structure factors and C_α-based phases is shown in Fig. 3(*a*). This map was skeletonized using the *MKSKEL* program (Williams, 1982; Fig. 3*b*), the resulting skeleton was then modified by the *CONNECT* program and a new map was calculated with the modified skeleton providing new phases and modified structure factors (Fig. 3*c*). This procedure was iterated ten times to produce the final refined map shown in Fig. 3(*d*). Fig. 4 shows the change in overall *R* factor and phase error during refinement. Whereas the initial phase error for the C_α model is 58°, refinement lowers this to 33.5°. Direct comparison of the true electron density map with the final map indicates that refinement restores the majority of the missing side-chain density (allowing the side chains to be easily modeled), as well as removing the extraneous density lying outside the protein region. The *R* factor and phase error in the final cycle closely approaches that obtained when refining from the true electron density map. Whereas the direction-

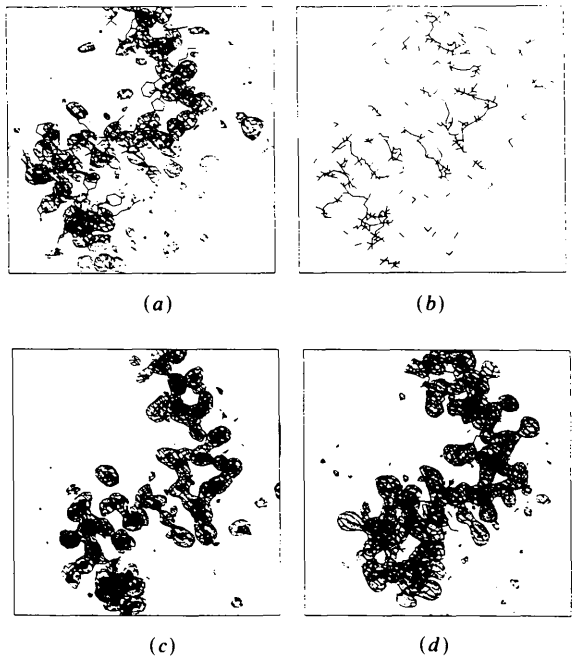

Fig. 3. Steps in the refinement of a C_α-atom-only starting model. (*a*) Starting electron density map calculated using phases derived from the C_α-atom model. (*b*) Skeleton produced from the map in (*a*) using the *GRINCH MKSKEL* program. (*c*) Skeleton following modification by the *CONNECT* program. Electron density is calculated using phases derived from the modified skeleton. (*d*) Electron density calculated after ten cycles of *PRISM* refinement.

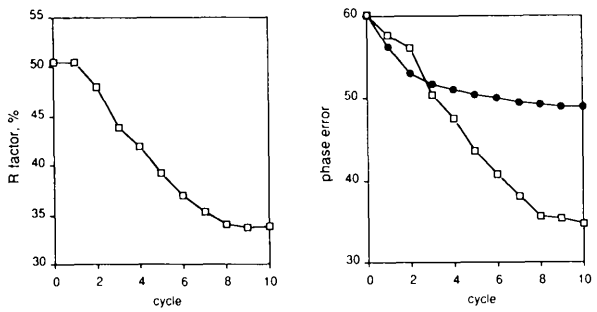

Fig. 4. Results for C_α-atom-only starting model. Overall *R* factor and phase error calculated as a function of refinement cycle, using the C_α-atom model to start refinement of the myoglobin fragment. Phase error is shown for both the *PRISM* refinement (open squares) and solvent-flattening refinement (filled circles).

ality of the starting model is ambiguous, density from the peptide carbonyl O atoms becomes readily apparent in the refined map, permitting the full atom backbone to be modeled.

To allow a direct comparison between *PRISM* refinement and conventional density modification, we have used the Wang (1985) solvent-flattening programs with the backbone-atom model providing a starting-phase probability distribution for each reflection. Ten cycles of iterative map filtering and phase recombination were carried out (recalculating the molecular boundary every cycle). A range of solvent fraction parameters were tested and optimal results were obtained with the input value for the solvent content set to 85%. The final recombined phases calculated by solvent flattening have an average phase error of 49.0° (Fig. 4), only marginally better than the starting-phase distribution (phase error = 58.0°). The results are marginally worse if the phases calculated in each cycle are used without being recombined with the partial starting-model phases (final phase error = 54.5°) or if $2wF_o - F_c$ weights are used to produce the recombined map in each cycle (final phase error = 52.3°). In this test case, at least, *PRISM* refinement seems considerably more effective than the method of solvent flattening.

The ability of the *PRISM* method to improve the phase estimates may in some way depend upon the high helical content of the myoglobin structure used for the test models. To demonstrate its more general utility, *PRISM* refinement was carried out with the first 50 residues of the protein alpha-lytic protease (Protein Data Bank entry 2ALP) as the test structure. In contrast to the largely helical myoglobin fragment, the alpha-lytic protease test model contains mostly extended β strands. A partial starting model was generated from the C_α coordinates for the 50 residues and refinement was carried out as described above for the myoglobin fragment. After fifteen cycles of refinement the phase error had dropped significantly, from 51.9 to 31.8°. This result suggests that successful use of *PRISM* refinement is not limited to highly ordered helical structures.

A more pertinent test of the refinement procedure was constructed using a model containing only backbone atoms for the first 25 residues of the 50-residue myoglobin fragment. Because this model lacks the majority of the side-chain atoms and the entire C-terminal half of the molecule, it more accurately simulates a potential starting map that one might obtain from molecular replacement with a fragment-search model. Fig. 5 shows the starting and final maps for this test case. Similar improvements in the overall *R* factor and the phase error are seen with this half-backbone model (Fig. 6) as with the C_α-only model. The improvement is especially dramatic in the C-terminal region (Figs. 5*c*, *d*), with the disordered, unconnected density becoming continuous and easily interpretable in the final map. The final phase error in this test (27.2°) is essentially identical to that obtained after several cycles of refinement with the complete model as a starting structure (26.9°). In contrast, the solution obtained following extensive solvent flattening had an average phase error of 52.6° (Fig. 6). Traditional density modification thus only slightly improves the phases relative to their initial estimates (phase error = 56.6°) and yields a final map that is significantly noisier than the *PRISM*-refined map, especially in the C-terminal region.

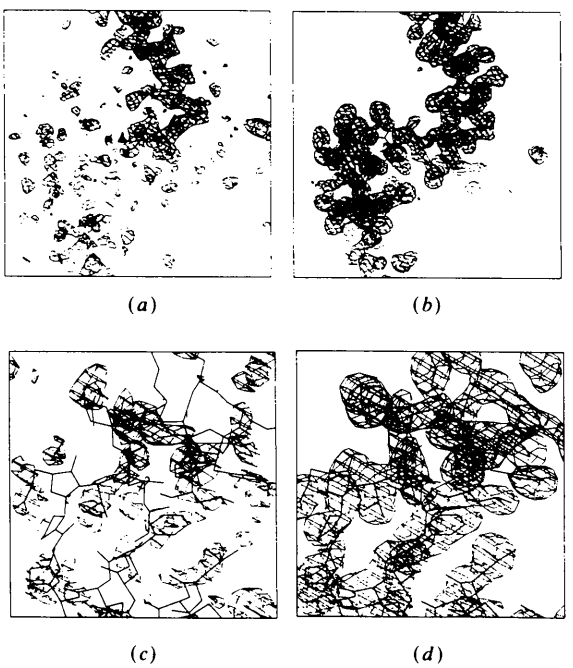

Fig. 5. Starting and final maps using a half-backbone-atom starting model. (*a*) Starting electron density map calculated using phases derived from the half-backbone-atom model. (*b*) Final electron density map after ten cycles of *PRISM* refinement. (*c*), (*d*) C-terminal region of the electron density (*c*) before and (*d*) after *PRISM* refinement.

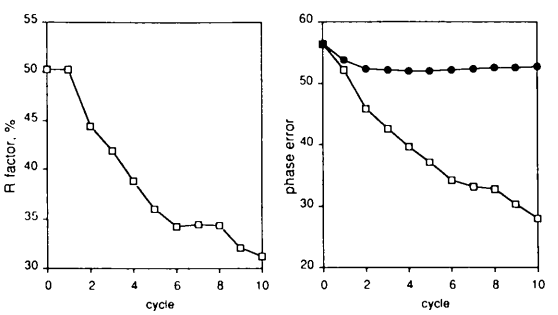

Fig. 6. Results for half-backbone-atom starting model. Overall *R* factor and phase error as a function of refinement cycle, using the half-backbone-atom model to refine the myoglobin fragment. Phase error is shown for both the *PRISM* refinement (open squares) and the solvent-flattening refinement (filled circles).

Understanding the requirement for phase convergence

Several additional tests were performed to simulate potential problems that might be encountered in a true structure determination. To approximate data-collection errors, the 'observed' structure factors were altered by the addition of a Gaussian distribution of random shifts such that the R factor between the true data and the modified data was $\simeq$7%. This level of noise, significantly more than one might expect with typical diffraction data, appeared to have only minor effects on the refinement procedure (slowing it slightly) when using the half-backbone starting model (Table 1). Similar effects were observed when using a half-backbone starting model that had been perturbed by the application of 0.5 Å random shifts to the atomic coordinates (Table 1). While systematic errors are likely to cause artifacts in the maps, the refinement procedure does not appear to require data free of random errors or a perfectly accurate partial starting model.

To understand which aspects of the procedure are required for the phases to converge to the true values, we repeated the half-backbone test several times altering steps in the refinement protocol. If conventional $2|F_o|-|F_c|$ Fourier coefficients are used in place of the optimally weighted coefficients ($2w|F_o|-|F_c|$), the phases improve only marginally after several cycles (Table 1). Modification of the skeleton also appears to play a key role; if the skeleton is not pruned or if connections in the density are not generated, the phases improve somewhat but the drop in the phase error is not as significant (Fig. 7). The resolution of the starting electron density map is important for proper refinement; a map with less than 3.5 Å resolution does not converge whereas one calculated at higher than 3.5 Å resolution converges easily (Table 1).

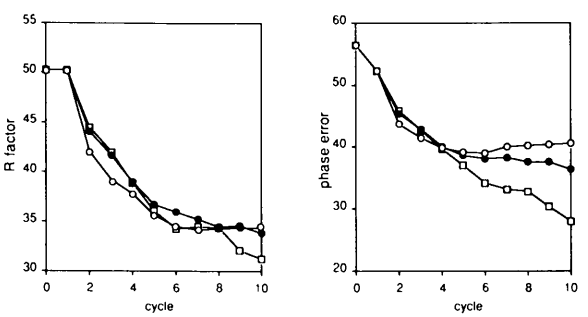

Fig. 7. Effects of skeleton modification on refinement. The half-backbone-atom model was used to start refinement of the myoglobin fragment. In addition to the standard protocol (open squares), the refinement was also carried out in tests in which connections between isolated subgraphs were not generated (open circles) or small graphs were not pruned out of the skeleton (filled circles).

Table 1. *Requirements for phase refinement*

The half-backbone-atom test (using backbone and C_β atoms for residues 1 to 25 to model the first 50 residues of myoglobin) was carried out using the standard refinement scheme described in the test. In addition, the refinement was repeated several times with slight changes in the refinement protocol as described above.

Test conditions	Starting R factor (%)	Starting phase error (°)	Final R factor (%)	Final phase error (°)
Standard refinement	50.5	60.1	30.5	27.2
Using $2F_o - F_c$ recombination	50.5	60.1	37.0	42.4
Not pruning small groups of atoms	50.5	60.1	33.9	36.3
Not generating connections between groups	50.5	60.1	34.5	40.4
Addition of 7% noise to the data	52.4*	60.1	34.3*	33.1
0.5 Å random shifts of starting-model atoms	56.35	61.9	33.2	33.4
Low-resolution (4 Å) starting map	50.5†	61.6†	39.7	53.4

* R factor calculated with F_o being error-added data, not error-free data.
† Starting parameters calculated after the first cycle of skeletonization.

Discussion

This work describes the development of an automated phase-refinement procedure to be used with phases provided by a partial model of the scattering density. As with standard density-modification (solvent-flattening) techniques, our method alternates between real-space and reciprocal-space representations of the scattering density. Whereas solvent flattening uses the real-space constraint that all density lie within a molecular boundary, our constraint is significantly more restrictive. By skeletonizing the map and using the nodes in the skeleton as dummy atoms, we immediately enforce positivity and atomicity on the electron density. The scattering density in a protein crystal can generally be represented as a single connected chain (Greer, 1985). By modifying the skeleton such that distinct groups of atoms are connected to one another, we force this constraint to be satisfied at every cycle in the refinement procedure. Because both the skeletonization and skeleton-modification steps are implemented as computer programs, these relatively strong constraints on the electron density can be applied with no undue human bias.

The radius of convergence of any refinement procedure is likely to be a function of the ratio between the number of variables used to represent the scattering density and the number of structure-factor observations. As this ratio increases, it becomes easier to fall into incorrect minima in the refinement cost function since errors in the electron density can be accurately modeled by the excess parameters. The skeleton representation requires relatively few parameters (the grid coordinates for each of the pseudo-atoms, approximately equal to the number of true

atoms in the structure) but models the scattering density with reasonable accuracy (as shown in Fig. 2). In contrast, traditional density-modification techniques treat each electron density map pixel within the molecular boundary as a continuous variable during refinement. With a 1 Å grid (appropriate for 3 Å data), a single atom contributes significant electron density to about 30 surrounding pixels. The skeletonization procedure thus reduces the effective number of parameters by approximately one order of magnitude. This effect alone may help explain the larger apparent radius of convergence of the *PRISM* method. Further improvements in the skeletonization procedure, which are now under way, should result in an even greater radius of convergence.

In addition to differences in the real-space constraints, application of the reciprocal-space constraints also differs between the solvent-flattening and *PRISM* procedures. In Wang's (1985) implementation of solvent flattening, the filtered-map phase probability distribution is multiplied at each cycle by the starting (experimentally determined) probability distribution to arrive at a new estimate. In contrast, our procedure ignores all previous estimates and uses only the current electron density map to obtain phases. Reciprocal-space recombination is achieved using modified Fourier coefficients $(2w|F_o|-|F_c|)$, yielding a new map that has the full component of the missing density yet the minimal amount of systematic noise (*i.e.* errors directed along the α_{calc} vector). If experimental phase information exists, it should be possible to implement a reciprocal-space recombination that uses this in each refinement cycle. The current procedure is appropriate, however, when the starting phase estimates are based on an incomplete atomic model that contains significant errors.

Future work using this methodology will incorporate additional constraints in the refinement procedure. For instance, *PRISM* refinement could be applied to only the poorly defined regions in a map derived from a relatively complete molecular-replacement model. At each cycle, new phases would be obtained using the combined real atoms from the molecular-replacement model and the nonoverlapping pseudoatoms from the *CONNECT* output. Alternatively, if a heavy-atom derivative has been found, its phase information could be used in a standard recombination scheme to direct the refinement. The addition of constraints on the phases should extend the radius of convergence of the method and allow the use of even more abbreviated partial models to start refinement.

We wish to thank Dr Lynn Ten Eyck for a conversation that inspired us to take this approach. Funding was provided by the Howard Hughes Medical Institute. CW was supported by a Fannie and John Hertz Foundation fellowship in applied physics.

APPENDIX
Optimal weighting scheme for modified Fourier coefficients

Given phases calculated for a partial structure, our goal is to find optimal Fourier coefficients that properly weight the contribution of the scattering from the missing atoms. A standard difference map (using $|F_o|-|F_c|$ coefficients and α_c phases) can be decomposed into three components (see Fig. 8), corresponding to the true missing signal, to systematic noise (directed along α_c) and to random noise (Blundell & Johnson, 1976):

$$
\begin{aligned}
&(F_0-F_c)\exp i\alpha_c \\
&\quad=[F_c/(F_0+F_c)]F_m \exp i\alpha_m \quad \text{(scaled true signal)} \\
&\qquad+[F_m^2/(F_o+F_c)]\exp i\alpha_c \quad \text{(systematic noise)} \\
&\qquad+[F_cF_m/(F_o+F_c)]\exp i(-\alpha_m+2\alpha_c) \\
&\qquad\qquad\qquad\qquad \text{(random noise).}
\end{aligned}
$$

An optimal map can be obtained by subtracting the best estimate for the systematic noise $\{[\langle F_m^2\rangle/(F_o+F_c)]\}$ and by scaling the (F_o-F_c) map such that the true scattering is given its full weight $\{[F_c/(F_o+F_c)]^{-1}\}$. To estimate the average systematic noise we may use (Sim, 1960)

$$
\begin{aligned}
\frac{\langle F_m^2\rangle}{(F_o+F_c)} &= \frac{F_o^2+F_c^2-2F_oF_c\int p(x)\cos x\,dx}{(F_o+F_c)} \\
&= \frac{F_o^2+F_c^2-2F_oF_cw}{(F_o+F_c)},
\end{aligned}
$$

where $w=I_1(X)/I_0(X)$; $X=2F_oF_c/\langle\Delta F^2\rangle$. After the average noise is subtracted, the true signal is scaled and the model scattering is added (to give the complete structure rather than the difference signal), we have

$$
\begin{aligned}
&\frac{\{(F_o-F_c)\exp i\alpha_c-[\langle F_m^2\rangle/(F_o+F_c)]\exp i\alpha_c}{[F_c/(F_o+F_c)]} \\
&\quad+F_c\exp i\alpha_c \\
&\qquad=\frac{(F_o-F_c)(F_o+F_c)-F_o^2-F_c^2+2F_oF_cw+F_c^2}{F_c\exp i\alpha_c} \\
&\qquad=[2wF_o-F_c]\exp i\alpha_c.
\end{aligned}
$$

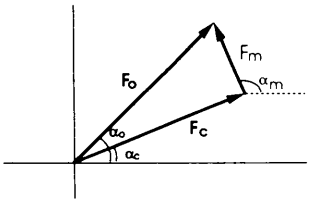

Fig. 8. Decomposition of standard difference map into three components. Model scattering = $F_c\exp i\alpha_c$. Missing scattering = $F_m\exp i\alpha_m$. True scattering = $F_o\exp i\alpha_o = F_c\exp i\alpha_c + F_m\exp i\alpha_m$.

Thus, a map calculated with the weighted coefficients $F_w = 2w|F_o| - |F_c|$ will provide the full scattering signal with the correct phase, plus additional random and systematic noise that has been minimized if the true phase is unknown,

$$(2wF_o - F_c)\exp i\alpha_c = F_m \exp i\alpha_m + F_c \exp i\alpha_c$$
(true signal at full strength)
$$+ (F_m^2 - \langle F_m^2 \rangle)/F_c \exp i\alpha_c$$
(minimal systematic noise)
$$+ F_m \exp i(-\alpha_m + 2\alpha_c)$$
(random noise).

References

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.

BRANDEN, C.-I. & JONES, T. A. (1990). *Nature (London)*, **343**, 687-689.

CHEN, L. Q., ROSE, J. P., BRESLOV, E., YANG, D., CHANG, W. R., FUREY, W. F., SAX, M. & WANG, B. C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240-4244.

GREER, J. (1985). *Methods Enzymol.* **115**, 206-224.

MESSERSCHMIDT, A., ROSSI, A., LADENSTEIN, R., HUBER, R., BOLOGNESI, M., GATTI, G., MARCHESINI, A., PETRUZZELLI, R. & FINAZZI-AGRO, A. (1989). *J. Mol. Biol.* **206**, 513-529.

PHILLIPS, S. E. V. (1978). *Nature (London)*, **273**, 247-248.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140-149.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511-512.

TULINSKY, A., PARK, C. H. & SKRZYPCZAK-JANKUN, E. (1988). *J. Mol. Biol.* **202**, 885-901.

WANG, B.-C. (1985). *Methods Enzymol.* **115**, 90-112.

WILLIAMS, T. (1982). PhD thesis, Univ. of North Carolina at Chapel Hill, USA.

WILSON, C., WARDELL, M., WEISGRABER, K. H., MAHLEY, R. W. & AGARD, D. A. (1991). *Science*, **252**, 495-506.

Acta Cryst. (1993). A**49**, 104-106

The Decomposition Scheme for Direct Methods

BY VÁCLAV KŘÍŽ

Fyzikální ústav ČSAV, Cukrovarnická 10, 162 00 *Praha* 6, *Czechoslovakia*

(*Received* 18 *November* 1990; *accepted* 19 *June* 1992)

Abstract

A new method for the decomposition of a set of start phases in two subsets is described. The decomposition enables the derivation of the subsets that have good values of some figures of merit to different ones with nearly the same value. By this operation a new set of start phases is obtained for the next refinement process. The method presented can be used as a simple but useful extension of the advanced program systems for the solution of the phase problem by direct methods.

1. Introduction

Each direct-method routine consists in forming some real function $G(\varphi)$ of the phase set φ and in generating a limited number of sets $\varphi_1, \ldots, \varphi_r$ of phases for which the values $G(\varphi_i)$ are close to the value expected for the correct solution estimated by the statistical theory.

The function $G(\varphi)$ can be, for example, some combination of good figures of merit. We denote by V the range of the function G. V is a subset of an n-dimensional hypercube,

$$V = I_1 \times \ldots \times I_n,$$

where n is the number of unknown phases, I_j is an interval $(-\pi, \pi)$ in the case of general phases or two values (that differ by π) for special phases.

Let the function G be chosen such that, for the correct phase set φ^*, the value $G(\varphi^*)$ can be expected to be very small, *i.e.* the routine has to generate 'the best' minima of G.

The most important component of a direct-method routine is a transformation $P: V \to V$. For some starting set $\varphi \in V$ this transformation derives a set of phases $P(\varphi) \in V$ for which a small value of $G[P(\varphi)]$ can be expected. As an example of P we can consider the traditional tangent procedure. Let $D \subset V$ be a subset of all the sets of starting phases used as the arguments of function P. D may be, for example, a result of the magic-integer procedure (White & Woolfson, 1975) or it may be identical to V [see, for example, the 'random approach' of Yao (1981)]. If P is based on developing phase values directed by a 'convergence map' (Germain, Main & Woolfson, 1970) from a small subset of fixed (origin definition) and trial phases then D is constructed by setting arbitrary acceptable values to the remaining phases. The direct-methods routine is completely described by the triad $\{D, P, G\}$; the pair $\{D, P\}$ determines which phase sets may be produced by means of this routine.

0108-7673/93/010104-03$06.00