Because of the completely general way in which the electron-density constraints can be applied, it is very easy to incorporate almost any other kind of constraint into the method. An obvious example of this is when a structure is partially known, *e.g.* the main chain of a macromolecule. The known density can be enforced on the map as part of the density modification and the remaining density determined along with the known density. An alternative approach would be to subtract the known density from the map and to determine the remaining density as a smaller structure. It is intended to try both of these developments of the method.

The most efficient way to propagate knowledge of phases throughout the reciprocal lattice is to use normalized structure factors, E's, rather than observed structure factors, F's. However, if E's are used, the map is subject to large series-termination errors and a considerable amount of negative density is produced. The background smoothing will get rid of this and, in so doing, extrapolate the E's to higher resolution. The extrapolated E's should be used in the calculation of the next map or the negative density will reappear. There are therefore two ways of proceeding with the calculations. One is to use F's, which produce no series-termination errors, and perform the calculations at the observed resolution. The other way is to use E's and perform the calculations at a suitably high resolution with extrapolated data, as is normally done in maximum-entropy calculations. This will increase the computing time considerably. The author has chosen to use F's to keep the computing time short, although the possibility of using E's needs to be investigated properly.

References

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65-67.

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A**41**, 286-290.

HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70-75.

HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* A**34**, 863-870.

KARLE, I. L. & KARLE, J. (1963). *Acta Cryst.* **16**, 969-975.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635-651.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180-184.

ZHANG, K. Y. J. & MAIN, P. (1990*a*). *Acta Cryst.* A**46**, 41-46.

ZHANG, K. Y. J. & MAIN, P. (1990*b*). *Acta Cryst.* A**46**, 377-381.

Acta Cryst. (1990). A**46**, 377-381

The Use of Sayre's Equation with Solvent Flattening and Histogram Matching for Phase Extension and Refinement of Protein Structures

BY KAM Y. J. ZHANG AND PETER MAIN

Department of Physics, University of York, Heslington, York YO1 5DD, England

(*Received* 21 *July* 1989; *accepted* 24 *October* 1989)

Abstract

A new method for phase refinement and extension, which combines Sayre's equation with solvent flattening and histogram matching, has been developed. Equations which express electron-density constraints are solved using discrete Fourier transforms to give a new approximation to the electron density. The formulation of the equations is in real space, which allows any set of constraints to be entered directly into the calculation. An application to the known structure of 2Zn insulin refined the 3 Å MIR phases from a mean phase error of 46 to 39° and extended the phases to 2 Å resolution with a mean overall phase error of 57°. Analysis of the phase errors shows that, for the strong reflexions, the new method determines phases with half the mean error of MIR phases.

Introduction

The dominant method in the determination of macromolecular structures is that of multiple isomorphous replacement (MIR). The phases obtained by MIR suffer from inaccuracies due to experimental error and lack of isomorphism and they are not always determined to the full resolution of the native data. All of this detracts from the quality of the electron-density map and may lead to difficulties in its interpretation. Thus, the ability to improve the quality of the MIR phases and to extend them to the full resolution of the native data would be a valuable contribution to protein crystallography.

The most successful technique of phase refinement and extension uses density modification. In its various forms it applies constraints to the electron density such as positivity, atomicity, boundedness, solvent flatness, connectivity and non-crystallographic symmetry. For a review, see Podjarny, Bhat & Zwick (1987). A recent addition to density modification is the histogram matching of Zhang & Main (1990) which imposes the correct electron-density histogram on the map. When combined with solvent flattening (Wang, 1985), it successfully refined the 1·9 Å MIR

0108-7673/90/050377-05$03.00

phases of 2Zn pig insulin (Baker, Blundell, Cutfield, Cutfield, Dodson, Dodson, Hodgkin, Hubbard, Isaacs, Reynolds, Sakabe & Vijayan, 1988) and extended them to 1·5 Å. At lower resolution, the method refined phases successfully, but it failed in phase extension.

In order to improve the method, we have sought to combine it with ideas from direct methods of phase determination. With the success of direct methods on small molecules (Karle, 1986; Hauptman, 1986; Woolfson, 1987), more attention is being paid to their use in the determination of macromolecular structures. Serious attempts have been made to use direct methods to improve MIR phases such as the use of Sayre's equation (Sayre, 1974), the maximum determinant rule (de Rango, Mauguen, Tsoucaris, Dodson, Dodson & Taylor, 1985) and the tangent formula (Blundell, Pitts, Tickle, Wood & Wu, 1981). We report in this paper a technique of phase refinement and extension which combines Sayre's equation with the density modification of Zhang & Main (1990) referred to earlier. Since Sayre's equation, solvent flattening and histogram matching are used simultaneously, the technique has attracted the acronym SQUASH. It has been tested on the known structure of 2Zn pig insulin and achieved phase refinement at 3 Å with extension to 2 Å. This is a significant improvement over our previous results (Zhang & Main, 1990).

Method

The method is described in detail by Main (1990). It consists of expressing the electron-density constraints due to solvent flattening and histogram matching as a system of equations in terms of the unknown electron density. These equations are solved simultaneously with Sayre's equation (Sayre, 1952) to obtain a least-squares estimate of $\rho(x)$. Since the equations are non-linear, this forms one cycle of an iterative procedure which starts with an approximate electron-density map and then alters it to become consistent with all the criteria built into the equations.

Sayre's equation is normally expressed as

$$F(\mathbf{h}) = [\theta(h)/V] \sum_{\mathbf{k}} F(\mathbf{k})F(\mathbf{h}-\mathbf{k}) \qquad (1)$$

where the scale factor $\theta(h)$ is the ratio of the scattering factors of the real and squared atoms. Equation (1) may also be expressed in real space as

$$\rho(\mathbf{n}) = (V/N) \sum_{\mathbf{m}} \rho^2(\mathbf{m})\psi(\mathbf{n}-\mathbf{m}) \qquad (2)$$

where the electron density, $\rho(\mathbf{n})$, is expressed as a discrete function evaluated at N grid points and $\psi(\mathbf{n})$ is the Fourier transform of $\theta(h)$. The convolution of $\psi(\mathbf{n})$ with $\rho^2(\mathbf{n})$ has the effect of changing the peak shapes of the squared density back to the normal shape. Because of the method used in the solution of the equations, it does not matter whether (1) is solved for the unknown phases or (2) is solved for the unknown electron density. In practice, we have used (2).

The scale factor $\theta(h)$ in Sayre's equation is sensitive to resolution. Its shape may be predicted from the atomic scattering factors and then a linear scale factor is calculated to minimize the residual of the equations. This works best at atomic resolution. A more satisfactory method for the low-resolution work described here was to set up the equations for a similar but known structure at the same resolution and then determine $\theta(h)$ as a function of $(\sin \vartheta)/\lambda$ by spherical averaging.

The constraints on the electron density due to solvent flattening and histogram matching may be expressed in the equations

$$\rho(\mathbf{n}) = H(\mathbf{n}) \qquad (3)$$

where $H(\mathbf{n})$ is the electron-density map modified as described by Zhang & Main (1990). The process of density modification makes (3) non-linear.

Equations (2) and (3) represent a system of non-linear simultaneous equations with as many unknowns, $\rho(\mathbf{n})$, as grid points in the asymmetric unit of the map and twice as many equations as unknowns. The functions $H(\mathbf{n})$ and $\psi(\mathbf{n})$ are both known. The least-squares solution, using either the full-matrix or the diagonal approximation, is obtained using the Newton–Raphson technique as described by Main (1990).

This determination of the electron density is part of an iterative procedure of map improvement very similar to that given by Zhang & Main (1990). Starting from an approximate map calculated from MIR phases, one proceeds as follows.

(*a*) Determine the molecular envelope.

(*b*) Set the density within the solvent region to a constant.

(*c*) Modify the density within the molecular envelope to match the expected histogram.

(*d*) Solve (2) and (3) for the electron density $\rho(\mathbf{n})$. [The map resulting from the modifications in (*b*) and (*c*) is the function $H(\mathbf{n})$ in (3).]

(*e*) Calculate structure factors and their Sim weights from $\rho(\mathbf{n})$.

(*f*) Combine the new phases with the original MIR phases, taking their weights into account. Extended phases and weights are accepted at their calculated values.

(*g*) Calculate a new map and repeat from (*a*) until the process has converged.

Results

The known structure of 2Zn pig insulin was chosen as a test for the method. The space group is $R3$ with two Zn atoms in the unit cell and 806 non-H atoms belonging to the protein in the asymmetric unit. The

Table 1. *Phase refinement and extension for* 2Zn *insulin using the full-matrix calculation*

Resolution (Å)	Number of reflexions	Mean phase error (°) Original phases	Extended phases	All phases
3·0	1677	46·2	—	46·2
3·0	1677	43·0	—	43·0
2·8	2141	40·9	69·4	47·1
2·6	2632	39·9	67·4	49·9
2·4	3140	39·3	66·4	51·9
2·2	4294	39·2	65·6	55·3
2·0	5551	39·3	64·8	57·1
2·0	5551	38·7	64·5	56·7

Table 2. *Phase refinement and extension for* 2Zn *insulin using the diagonal-approximation calculation*

Resolution (Å)	Number of reflexions	Mean phase error (°) Original phases	Extended phases	All phases
3·0	1677	46·2	—	46·2
3·0	1677	42·2	—	42·2
2·8	2141	40·7	65·8	46·1
2·6	2632	39·9	65·1	49·0
2·4	3140	39·5	66·5	52·1
2·2	4294	39·4	65·8	55·5
2·0	5551	39·7	65·4	57·6
2·0	5551	39·6	65·4	57·6

magnitudes used in the calculations were observed F's which had been sharpened by removing the overall temperature factor.

The 3·0 Å MIR phases were refined using the procedure outlined in the previous section and convergence was reached after five iterations. The phases were then extended to 2·0 Å in five stages, increasing the resolution by 0·2 Å at each stage. Separate calculations were carried out using the full-matrix and the diagonal approximation for comparison. It was found that satisfactory convergence for each stage of phase extension was reached after about five iterations, regardless of the kind of calculation used. The full-matrix results are shown in Table 1 and these may be compared with the diagonal-approximation results in Table 2. The measure of quality of the phases which appears in Tables 1 and 2 is an unweighted mean phase error, *i.e.* all phases are given the same weight. The last entry at 2·0 Å resolution in each table is the result of additional phase refinement.

It was of interest to see if the processes of density modification and Sayre's equation both added to the quality of the final result. The phase refinement at 3·0 Å and extension to 2·0 Å was repeated using Sayre's equation alone, *i.e.* without density modification. This gave the results shown in Table 3, where it is seen that the original MIR phases at 3·0 Å are not improved in quality and the phase extension to 2·0 Å is poor. The overall phase error of 65° is significantly worse than the 57° of the SQUASH phases. A further calculation was carried out using density modification alone, *i.e.* solvent flattening and histogram matching. The results of this are also in Table 3, where they are seen to be only a small improvement on Sayre's equation. The fact that the SQUASH phases are better than either shows that each constraint applied to the electron density adds useful and different information.

Table 3. *Comparison of mean phase errors obtained in various calculations*

	Mean phase error (°) ∞–3·0 Å	3·0–2·0 Å	∞–2·0 Å
MIR phases	46·2	63·0	57·9
SQUASH phases (full matrix)	38·7	64·5	56·7
SQUASH phases (diagonal approximation)	39·6	65·4	57·6
Sayre's equation alone	46·5	73·5	65·3
Density modification alone	43·2	72·7	63·8

Table 4. *Correlation coefficients with the map of the refined structure at* 2·0 Å

3·0 Å MIR map	2·0 Å MIR map	2·0 Å SQUASH map Full matrix	2·0 Å SQUASH map Diagonal approximation
0·625	0·676	0·739	0·728

Table 5. *Mean phase errors of the strongest F's at* 2·0 Å

Number of strongest reflexions	Mean phase error (°) MIR phases	SQUASH phases Full matrix	SQUASH phases Diag. approx.
250	30·6	16·8	18·3
500	32·6	20·5	21·5
1000	37·0	27·6	28·3
2000	46·0	38·9	39·5

The computing time depended heavily upon resolution, but at 2·0 Å each iteration required about 70 min for the full-matrix calculation and 8 min for the diagonal approximation on a VAX 8650. Later program improvements have reduced these times significantly.

Discussion

A comparison of the SQUASH phases with the MIR phases in the same resolution range is shown in Table 3. It is satisfying to note that the original 3·0 Å phases have been improved and the extended phases appear to be just as good as those determined by MIR. However, the unweighted mean phase error hides the distribution of errors. The fact that the SQUASH phases are better is seen in Table 4, which shows the correlation coefficients between various maps and that calculated from the refined structure. Correlation coefficients are calculated as described by Zhang & Main (1990). Clearly the SQUASH maps are of better quality than the MIR map at the same resolution.

This suggests that the phases of the strong reflexions are much better determined by SQUASH than by MIR. Tables 5 and 6 confirm this. The errors in the

Table 6. *Mean phase errors of strongest E's at* 2·0 Å

		Mean phase error (°)		
E value	Number of reflexions	MIR phases	SQUASH phases Full matrix	SQUASH phases Diag. approx.
>1·5	526	55·0	38·8	38·4
>1·0	2047	50·4	45·4	45·8
>0·1	5545	57·9	56·6	57·5

SQUASH phases for the strongest reflexions are considerably smaller than those obtained by MIR.

There are many places where the improvement in the quality of the map is obvious, but one example is given in Figs. 1 to 4. Fig. 1 shows the molecular structure for residues 62B (Val), 63B (Asn) and 64B (Gln). The glutamine side chain is disorded and Fig. 1 shows the atoms in both positions (the occupancy for each is about the same). Figs. 2 to 4 show the associated densities from the 3·0 Å MIR map, the 2·0 Å SQUASH map and the 2·0 Å map calculated from the refined structure. The main chain between CA62B and N63B is seen to be broken in the MIR map in Fig. 2, but is continuous in the SQUASH map in Fig. 3. In addition, the side-chain density from CA63B to ND63B is broken in Fig. 2 but continuous in Fig. 3 and new density is present around the disordered atoms in the glutamine side chain in the SQUASH map which was not present at all in the original MIR map. Continuity of density is very important for the interpretation of the map and it is clear that this is significantly improved by our process.

The tables of results show that the full-matrix phases are consistently better than those obtained by the diagonal approximation. However, the difference is not large. The full-matrix calculations take nearly ten times longer than the diagonal approximation, so it is doubtful whether the extra computing is worth the small gain in accuracy.

The addition of Sayre's equation to our method of histogram matching has applied an extra and important constraint to the electron density. That is, the atomic shape. Because the histogram does not contain any positional information, the peaks in the map without Sayre's equation can be any shape consistent with the observed magnitudes. With Sayre's equation, the peak shape is fixed by the function $\psi(\mathbf{n})$ in (2) and this has added considerably to the power of the

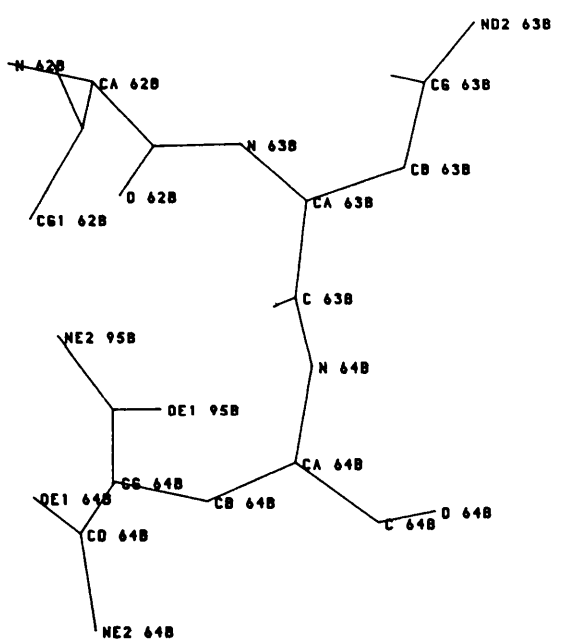

Fig. 1. Molecular structure for residues 62B to 64B of 2Zn insulin.

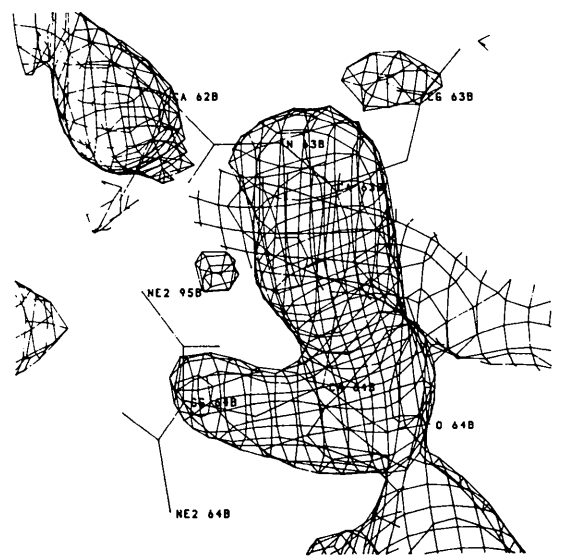

Fig. 2. Electron density for residues 62B to 64B of 2Zn insulin obtained from the 3 Å MIR phases.

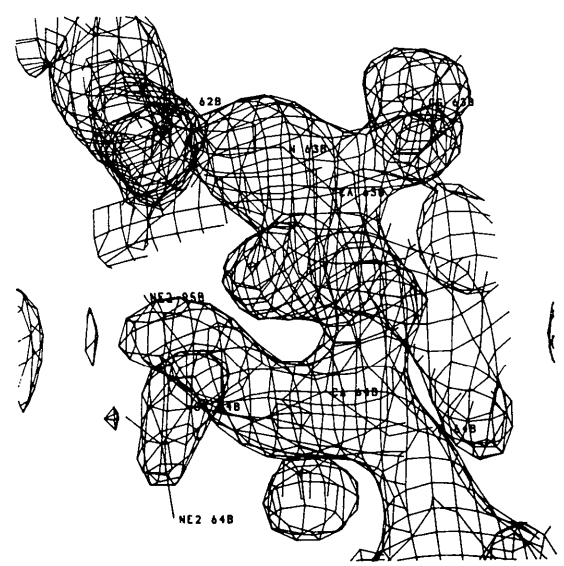

Fig. 3. Same electron density as Fig. 2 but obtained from 2 Å SQUASH phases.

method. Its power is necessarily limited, however, because we cannot see atoms at low resolution. The electron-density peaks will no longer be spherically symmetric because of overlap and their shapes cannot be predicted without knowing the structure. This means the atomic shape function $\psi(\mathbf{n})$ becomes structure dependent. Fortunately, this does not mean that atomic resolution is an absolute necessity and we have shown that good results can be obtained starting from 3·0 Å resolution.

This present work may be regarded as a development of Sayre's own work (Sayre, 1974) on the phase extension and refinement for rubredoxin. The improvements we have brought to it are an enormous reduction of computing time, the addition of density modification and the least-squares solution of the equations. The latter enables us to make full use of very weak and accidentally absent reflexions, which contain useful information about the distribution of atoms. Sayre had to remove these from his equations because the phases of these reflexions (the variables with which he was working) have no meaning.

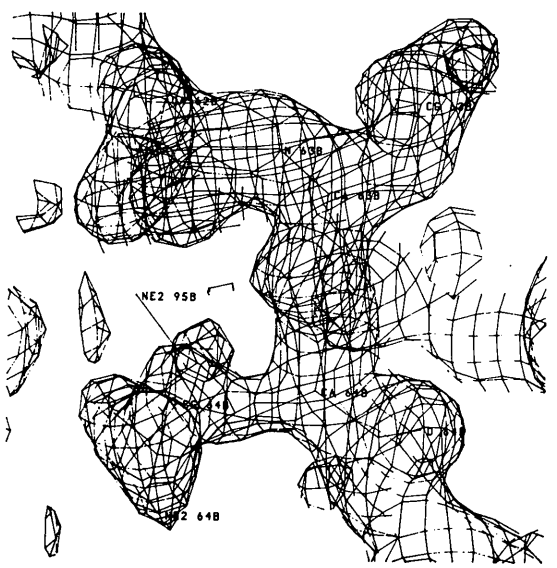

Fig. 4. Same electron density as Fig. 2 but obtained at 2 Å from the refined structure.

We wish to thank Professor G. Dodson for kindly supplying the 2Zn insulin data and atomic coordinates. We are also indebted to Mrs E. Dodson for the use of computer programs and helpful discussions. One of us (KYJZ) is grateful to the Rigaku Corporation of Japan for a research studentship, to the Dodsons for the use of their laboratory facilities and also to Professor M. M. Woolfson for encouragement and support in this project.

References

BAKER, E. N., BLUNDELL, T. L., CUTFIELD, J. F., CUTFIELD, S. M., DODSON, E. J., DODSON, G. G., HODGKIN, D. C., HUBBARD, R. E., ISAACS, N. W., REYNOLDS, C. D., SAKABE, N. & VIJAYAN, M. (1988). *Philos. Trans. R. Soc. London Ser. B*, **319**, 369-456.

BLUNDELL, T. L., PITTS, J. E., TICKLE, I. J., WOOD, S. P. & WU, C.-W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.

HAUPTMAN, H. (1986). *Science*, **233**, 178-183.

KARLE, J. (1986). *Science*, **232**, 837-843.

MAIN, P. (1990). *Acta Cryst.* A**46**, 372-377.

PODJARNY, A. D., BHAT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Chem.* **16**, 351-373.

RANGO, C. DE, MAUGUEN, Y., TSOUCARIS, G., DODSON, E. J., DODSON, G. G. & TAYLOR, D. J. (1985). *Acta Cryst.* A**41**, 3-17.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180-184.

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

WOOLFSON, M. M. (1987). *Acta Cryst.* A**43**, 593-612.

ZHANG, K. Y. J. & MAIN, P. (1990). *Acta Cryst.* A**46**, 41-46.

Acta Cryst. (1990). A**46**, 381-387

An Electron-Density Study of Germanium: Evaluation of the Available Experimental Data

BY A. S. BROWN AND M. A. SPACKMAN

Department of Chemistry, University of New England, Armidale, 2351 *NSW, Australia*

(*Received* 30 *May* 1989; *accepted* 5 *December* 1989)

Abstract

Deformation and valence electron densities in germanium are derived *via* Fourier summation and multipole refinement of a selectively merged set of X-ray structure factors. The deformation density for germanium appears to be qualitatively different from that in silicon and diamond. The available experimental data are evaluated in the light of problems encountered in the electron-density analysis. In

0108-7673/90/050381-07$03.00