Acta Cryst. (1995). D**51**, 1036–1040

Direct-Space Methods in Phase Extension and Phase Determination. III. Phase Refinement using Sayre's Equation

By L. S. Refaat, C. Tate and M. M. Woolfson

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

(*Received* 2 *June* 1994; *accepted* 16 *May* 1995)

Abstract

An algorithm is described for refining a set of phases to agree with the Sayre equation. All operations are carried out using Fourier transforms with modest computer-store requirements even for very large systems. The procedure is tested with two moderate-sized proteins, one containing heavy atoms, and is found to give good refinement with data at more than atomic resolution (1.17 Å) and useful, if less good, refinement when the data resolution is lower (1.5 Å). It is concluded that at atomic resolution, or slightly below, the Sayre equation still has something to offer both for phase refinement and phase extension, especially if used cautiously with weighted multiple isomorphous replacement phases acting as a constraint on the phase changes. Even when the Sayre equation on its own refines phases badly, or not at all, it may still make an important contribution in conjunction with other real-space refinement procedures.

Introduction

Sayre (1952) showed that for an equal-atom structure a good set of phases, $\{\varphi(\mathbf{h})\}$, associated with observed magnitudes of sufficiently high resolution satisfies the equations,

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

where, $F(\mathbf{h}) = |F(\mathbf{h})| \exp\{i\varphi(\mathbf{h})\}$, $\theta(\mathbf{h}) = f(\mathbf{h})/g(\mathbf{h})$, $f(\mathbf{h})$ = atomic scattering factor, $g(\mathbf{h})$ = scattering factor for a squared atom and V = volume of the unit cell.

It was shown by Shiono & Woolfson (1991) that Sayre's equation is reasonably well satisfied even when the condition of equal atoms does not hold or when there is less than atomic resolution.

Sayre (1972) described a least-squares technique for phase refinement using (1) and tested it under somewhat idealized conditions. However, it was shown to be effective in extending the phases of rubredoxin from 2.5 to 1.5 Å resolution (Sayre, 1974). A computational difficulty in this work arises from the problem of storing the Jacobian matrix involved in the linear equations for a large number of reflections. Sayre (1975) proposed that very large computational savings could be made by reducing all matrix operations to Fourier transforms. The Sayre equation transformed into real space has been combined with several other real-space constraints in the phase extension and refinement (PER) procedure *SQUASH* (Zhang & Main, 1990*ab*; Cowtan & Main, 1993). The present study concentrates on refining phases using a large system of Sayre equations by modifying phases while retaining the observed magnitudes. The procedure is illustrated by application to a 96 amino-acid protein, RNAp1 (Bezborodova, Ermekbaeva, Shlyapnikov, Polyakov & Bezborodov, 1988), for which data is available at 1.17 Å resolution, and to 2-Zn Insulin (Baker *et al.*, 1988) for which multiple isomorphous replacement (MIR) phases are available at 1.9 Å resolution and native protein data to 1.5 Å resolution.

Optimizing phases for a system of Sayre equations

For a given set of phase estimates the residuals of the associated set of Sayre equations are,

$$r(\mathbf{h}) = F(\mathbf{h}) - [\theta(\mathbf{h})/V] \sum_{\mathbf{k}} F(\mathbf{k})F(\mathbf{h}-\mathbf{k}). \qquad (2)$$

If the structure factors are considered as the basic variables then the terms of the Jacobian are,

$$A(\mathbf{h},\mathbf{k}) = \delta(\mathbf{h},\mathbf{k}) - [2\theta(\mathbf{h})/V]F(\mathbf{h}-\mathbf{k}), \qquad (3)$$

where $\delta(\mathbf{h},\mathbf{k})$ is the Kronecker delta such that $\delta(\mathbf{h},\mathbf{h}) = 1$ and $\delta(\mathbf{h},\mathbf{k}) = 0$, $\mathbf{k} \neq \mathbf{h}$. For sufficiently small changes,

$$\delta r(\mathbf{h}) = \sum_{\mathbf{k}} A(\mathbf{h},\mathbf{k})\delta F(\mathbf{k}). \qquad (4)$$

Assuming that the linear approximation is valid the residuals may be eliminated by values of $\delta F(\mathbf{k})$ coming from the linear equations,

$$\sum_{\mathbf{k}} A(\mathbf{h},\mathbf{k})\delta F(\mathbf{k}) = -r(\mathbf{h}). \qquad (5)$$

Premultiplying both sides by the Hermitian transpose of A gives,

$$\sum_{\mathbf{k}} H(\mathbf{l},\mathbf{k})\delta F(\mathbf{k}) = -\gamma(\mathbf{l}), \qquad (6)$$

where,

$$H(\mathbf{l},\mathbf{k}) = \sum_{\mathbf{h}} A(\mathbf{h},\mathbf{l}) * A(\mathbf{h},\mathbf{k}) \qquad (7)$$

©1995 International Union of Crystallography
Printed in Great Britain – all rights reserved

and,

$$\gamma(\mathbf{l}) = \sum_{\mathbf{h}} A(\mathbf{h}, \mathbf{l}) * r(\mathbf{h}). \tag{8}$$

Substituting (3) in (8) gives,

$$\gamma(\mathbf{l}) = r(\mathbf{l}) - Y(\mathbf{l}), \tag{9}$$

where,

$$Y(\mathbf{l}) = (1/V) \sum_{\mathbf{h}} \theta(\mathbf{h}) r(\mathbf{h}) F(\mathbf{l} - \mathbf{h}). \tag{10}$$

The factor of 2 in (3) does not appear in (9) because the summation in (10) is over a sphere of reciprocal space while $A(\mathbf{h}, \mathbf{k})$ includes only those reflections being refined, *i.e.*, those in a hemisphere. If the total residual is defined by,

$$R = (1/2) \sum_{\mathbf{h}} |r(\mathbf{h})|^2, \tag{11}$$

then $\gamma(\mathbf{l})$ and $H(\mathbf{l}, \mathbf{k})$ are its gradient and Hessian, respectively.

The changes in the structure factors found by solving the linear equations modify both their magnitudes and phases. The change of phase for a general reflection of index $\mathbf{k}$ is found by taking the component of $\delta F(\mathbf{k})$ perpendicular to $F(\mathbf{k})$. For a special reflection the component of $\delta F(\mathbf{k})$ along $F(\mathbf{k})$, $\Delta F(\mathbf{k})$, is considered. If $|F(\mathbf{k})| + \Delta F(\mathbf{k})$ is positive the phase is not changed but if it is negative then the phase is changed by π. This way of modifying phases works well in practice.

Two methods have been tried for minimizing R. The Hessian, $\mathbf{H}$, is strongly diagonal and the first approach was to use the diagonal approximation to (6). From (7) and (3) the diagonal elements of $\mathbf{H}$ are,

$$D(\mathbf{l}) = 1 - (4/V) \Big\{ \theta(\mathbf{l}) F(\mathbf{O}) - (1/V) \sum_{\mathbf{h}} \theta(\mathbf{h})^2 |F(\mathbf{l} - \mathbf{h})|^2 \Big\}, \tag{12}$$

and (6) reduces to,

$$\delta F(\mathbf{l}) = -\gamma(\mathbf{l})/D(\mathbf{l}). \tag{13}$$

Another approach is that of steepest descent for which,

$$\delta F(\mathbf{l}) = -\lambda \gamma(\mathbf{l}). \tag{14}$$

Substituting (14) into (5),

$$\lambda \sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k}) \gamma(\mathbf{k}) = r(\mathbf{h}). \tag{15}$$

If we now write,

$$Q(\mathbf{h}) = \sum_{\mathbf{k}} A(\mathbf{h}, \mathbf{k}) \gamma(\mathbf{k}), \tag{16}$$

then the optimum λ is that which minimizes $|\lambda Q(\mathbf{h}) - r(\mathbf{h})|^2$. This gives,

$$\lambda = \mathrm{Re}\Big\{ \sum_{\mathbf{h}} Q(\mathbf{h}) * r(\mathbf{h}) \Big\} \Big/ \sum_{\mathbf{h}} |Q(\mathbf{h})|^2,$$

and by the use of (8) it can be shown that,

$$\lambda = \sum_{\mathbf{h}} |\gamma(\mathbf{h})|^2 \Big/ \sum_{\mathbf{h}} |Q(\mathbf{h})|^2. \tag{17}$$

Substituting (3) into (16) we also find that

$$Q(\mathbf{h}) = \gamma(\mathbf{h}) - [2\theta(\mathbf{h})/V] \sum_{\mathbf{k}} \gamma(\mathbf{k}) F(\mathbf{h} - \mathbf{k}). \tag{18}$$

In this equation the factor of 2 is used if the $\mathbf{k}$ summation is over a hemisphere. It is not necessary if the summation is over the whole observed sphere in reciprocal space. A steepest descent step is the first part of the conjugate-gradient procedure and this we actually tried. However, it was found that it was better to use successive steps of steepest descent repeatedly.

In the applications we have made so far (13) and (14) give similar results because $D(\mathbf{l})$ lies approximately in the range 0.8–1.25 while λ also has a value close to unity (see Tables 1 and 2). However, while $D(\mathbf{l})$ is independent of phase, as can be seen from (12), λ calculated from (17) is phase dependent. It has been found that this phase dependence does give a clear advantage in some circumstances. For example, when Sayre-equation refinement is combined with some other techniques such as histogram matching or solvent flattening then values found for λ can differ appreciably from unity and the consequent phase improvement is better.

Practical procedures

The application of this approach is totally dependent on the use of FFT routines. The steps are as follows where FT represents either a Fourier transform or an inverse Fourier transform.

(i) Using current phases a density map, ρ, is calculated.

(ii) An FT of the square of this map gives $G(\mathbf{h}) = (1/V) \sum_{\mathbf{k}} F(\mathbf{k}) F(\mathbf{h} - \mathbf{k})$.

(iii) The values of $r(\mathbf{h})$ are then found from $F(\mathbf{h}) - \theta(\mathbf{h}) G(\mathbf{h})$, where the values of $\theta(\mathbf{h})$ are expressed analytically.

(iv) The FT of $\theta(\mathbf{h}) r(\mathbf{h})$ gives a real-space function, σ.

(v) The FT of $\sigma\rho$ then gives the coefficients $Y(\mathbf{l})$ in (10) which combine with $r(\mathbf{l})$ to give $\gamma(\mathbf{l})$ in (9).

(vi) The FT of $\gamma(\mathbf{k})$ gives a real-space function, ψ.

(vii) the FT of $\psi\rho$ gives the value of the summation in (18) from which the values of $Q(\mathbf{h})$ may be found.

Thus, a full treatment requires six FFT's per cycle. In fact two other FFT's could be used to give the summation in (12) but, since this is a phase-independent quantity, it needs to be carried out only once. In practice, we calculated this summation by assuming that

Table 1. *Refinement of* 23 853 *phases of RNAp*1 *at* 1.17 Å *resolution*

MPE1 = mean phase error; WMPE1 = E-weighted mean phase error; MPE2 = mean phase error for enantiomorph structure; MPEL = mean phase error for 1964 reflections with $|E| \geq 1.4$; MCC = map correlation coefficient; λ = value from equation (17). All phase errors are in °.

Cycle	MPE1	WMPE1	MPEL	MPE2	MCC	λ
0	59.7	59.4	57.5	87.2	0.419	0.808
1	52.7	50.0	38.9	85.9	0.571	0.687
2	50.2	47.1	35.2	87.0	0.610	0.685
3	46.2	43.0	30.1	86.7	0.661	1.088
4	42.0	38.0	21.8	86.9	0.727	1.037
5	39.9	35.7	19.4	86.9	0.755	1.044
6	39.2	35.0	18.3	86.7	0.762	1.027
7	38.6	34.4	18.0	86.8	0.770	1.025
8	39.2	35.0	18.1	86.8	0.763	0.984
9	38.8	34.6	18.0	86.8	0.768	1.029
10	39.5	35.3	18.3	86.8	0.760	0.998
11	39.1	34.9	18.2	86.7	0.764	1.029
12	39.7	35.4	18.4	86.8	0.758	1.000
13	39.4	35.1	18.4	86.7	0.762	1.028
14	40.0	35.7	18.7	86.8	0.755	1.000
15	39.6	35.4	18.6	86.7	0.759	1.028
16	40.3	36.0	18.8	86.8	0.752	0.998
17	39.9	35.7	18.8	86.7	0.755	1.028
18	40.5	36.2	18.9	86.8	0.749	0.999
19	40.2	35.9	19.0	86.7	0.753	1.028
20	40.7	36.4	18.9	86.8	0.747	1.000
21	40.4	36.2	19.1	86.8	0.750	1.028
22	41.0	36.7	19.1	86.8	0.744	1.000
23	40.7	36.4	19.3	86.8	0.747	1.028
24	41.3	36.9	19.2	86.8	0.741	1.001
25	40.9	36.7	19.4	86.8	0.744	1.027

Table 2. *Refinement of* 11 414 *phases of RNAp*1 *at* 1.5 Å *resolution*

The tabulated quantities are described in Table 1. The number of reflections for which $|E| \geq 1.4$ is 1139. Only the first ten cycles are given. The pattern of gradual phase degeneration thereafter is similar to what is seen in Table 1. The MCC relates to an ideal map at 1.17 Å resolution. All phase errors are in °.

Cycle	MPE1	WMPE1	MPEL	MPE2	MCC	λ
0	59.4	59.0	56.5	86.1	0.292	0.672
1	57.8	54.9	46.4	86.3	0.345	0.881
2	56.4	52.8	42.2	86.3	0.370	0.861
3	56.1	51.9	39.4	86.0	0.382	1.004
4	56.2	51.7	38.4	85.9	0.386	0.991
5	56.1	51.4	36.9	85.9	0.390	1.005
6	56.3	51.5	36.7	85.9	0.390	0.996
7	56.5	51.5	35.5	85.9	0.391	1.006
8	56.9	51.9	35.6	86.0	0.388	1.005
9	57.1	51.9	35.2	86.0	0.388	1.007
10	57.6	52.4	35.5	86.1	0.385	1.006

$|F(\mathbf{l}-\mathbf{h})|$ is independent of $\theta(\mathbf{h})$ and deriving $|\theta(\mathbf{h})|^2$ by analytical integration.

There is an option in the program to use either sharpened F's or E's, the normalized structure factors. For all cases examined so far results have been better with E's. If sharpened F's are used then two Fourier transforms are required to give the values of $\theta(s)$, a function of position in reciprocal space. One of these is to calculate model density for a single atom and the other to transform back the square of the model density.

Results

The first trial structure we used was RNAp1 (Bezborodova *et al.*, 1988). This has space group $P2$ with $a = 32.01$, $b = 49.76$, $c = 30.67$ Å and $\beta = 115.83°$ with $Z = 2$. The asymmetric unit contains 808 non-H atoms in the protein, including five S atoms and, in addition 83 water molecules. The data extend to a resolution of 1.17 Å and there are 23 853 independent reflections. In all our tests we started the refinement with a random mean phase error (MPE) imposed on the calculated phases. Our method of applying random MPE's is to produce a weighted mixture of calculated phases, φ_c, with completely random phases, φ_r, by,

$$\varphi = \arctan\{[w \sin\varphi_c + (1-w)\sin\varphi_r] \\ \times [w\cos\varphi_c + (1-w)\cos\varphi_r]\}. \qquad (19)$$

For $w = 1$ the calculated phases are obtained while for $w = 0$ the final phases are completely random. It is difficult to choose a value of w to achieve a precise value of MPE but with experience one can get close to a target value.

In giving the results of our tests for RNAp1 in Tables 1 and 2 the column MPE1 gives the MPE for all the reflections and WMPE the E-weighted MPE. The column MPE2 gives the MPE for the enantiomorph structure. In investigations of methods the MPE alone can sometimes be misleading if the phases give a map with poor enantiomorph discrimination – by imposing a pseudo centre-of-symmetry, for example. This is revealed by similar MPE's for both enantiomorphs and we find both values as a matter of course in all our investigations. Finally, we give the mean phase error, MPEL, for the subset of large magnitude reflections for which $|E| \geq 1.4$. The value of λ is given for each cycle and also MCC, the standard map correlation coefficient corresponding to the current phases. It should be noted that the MCC is always given relative to an ideal map with calculated phases corresponding to all the data available – even when the data set being refined is at a lesser resolution. Other workers prefer to compare their maps with ideal maps at the same resolution as the data which has been phased. The value of the MCC is calculated from the phase errors using the formulae given by Lunin & Woolfson (1993).

Test 1

In this test all the data to 1.17 Å resolution were refined (see Table 1). The initial value of MPE1 was 59.7° and after 25 cycles this had reduced to 40.9°. However, WMPE reduced from 59.4 to 37.5° so it is clear that the reflections with higher E values are most improved. This is also confirmed by the reduction in the value of MPEL, the MPE for the 1964 reflections with $E > 1.4$, from 57.5 to 19.4°. Although 25 cycles of refinement are shown it is clear that the total improvement came in the first seven cycles as is seen both in the MPE and in the value of MCC, 0.770, which corresponds to a map of good quality for interpretation at 1.17 Å resolution. After cycle 7 there is a slow deterioration in the results and one of the problems we have not yet tackled is to find out when to stop. However, there are figures of merit that might be effective, for example that proposed by Mishnev & Woolfson (1994). The variation in the value of λ confirms what was stated above, that it has a value close to unity.

Test 2

This is similar to Test 1 except that it uses only the 11 414 data to 1.5 Å resolution (see Table 2). It is clear that refinement is much less effective and again reaches the best position in seven cycles as judged by the value of the MCC, 0.391. While there has been a significant improvement of the MCC from its original value of 0.292 if the MCC had ben related to the ideal map at 1.5 Å then the MCC would have improved from 0.42 to 0.57 but we think that the smaller numbers give a more realistic impression of the information content of the phase set.

In an original version of this paper, when the algorithm we were using was somewhat less effective, there was very little improvement in the phases for RNAp1 at 1.5 Å resolution and we stated that the Sayre equation seemed not to be effective at this resolution. A referee pointed out that we ourselves had stated that Sayre (1974) had refined rubredoxin at 1.5 Å resolution and the referee also reminded us that Agarwal & Isaacs (1977) had successfully extended MIR phases at 1.9 Å for insulin to 1.5 Å using a tangent-formula approach. This persuaded us that we should look again at our procedure both with a view to improving it and also to applying it to insulin.

There were three changes we found which improved the results obtained.

(i) We originally calculated normalized structure factors by the K-curve method which constrains $\langle|E|^2\rangle$ to be equal to unity within all spherical shells in reciprocal space. Now the Wilson plot is used to find temperature and scale factors for converting F's to E's individually. For the $|E|$'s found in this way the value of $\langle|E|\rangle^2$ depends on the scattering angle. The values are closer to calculated values and more appropriate for inclusion in the Sayre equation.

(ii) The complex quantity of $\delta F(\mathbf{h})$ is now projected at the end of each cycle as appropriate for special or general reflections. Previously we projected the gradient $\gamma(\mathbf{h})$ before calculating λ.

(iii) We had originally, in error, included a factor 2 in (9).

The structure of 2-Zn insulin (Baker *et al.*, 1988) has space group $R3$ with $a = 49.0$ Å and $\alpha_R = 114.8°$. The asymmetric unit contains 806 non-H atoms, excluding solvent but including two Zn atoms. There are 6447 reflections to 1.9 Å resolution, for which MIR phases are available, and 13 289 reflections for the native protein to 1.5 Å resolution. For these results we give at each cycle MPE's for the MIR set, MPEM; for the extended set, MPEE; for the total set, MPET; E-weighted for the total set, WMPET, and for the subset with $|E| \geq 1.4$, MPEL.

Table 3. *Phase extension and refinement for 2-Zn insulin with extended phases originally given random values.*

Results are shown for even cycle numbers only. The MIR-phased set to 1.9 Å resolution contains 6447 reflections and there were 6842 extra reflections to 1.5 Å resolution. All phase errors are in °. Quantities tabulated are: MPEM = mean phase error for MIR-phased set; MPEE = mean phase error for extension set between 1.9 and 1.5 Å resolution; MPET = mean phase error for all 13 289 reflections; WMPET = weighted mean phase error for all 13 289 reflections; MPEL = mean phase error for 1055 reflections with $|E| \geq 1.4$; MCC = map correlation coefficient. Values of MPE2 (see Table 1) are all $\geq 86.1°$.

Cycle	MPEM	MPEE	MPET	WMPET	MPEL	MCC
0	61.5	89.8	76.1	74.8	70.8	0.211
2	58.7	80.3	69.8	67.8	62.4	0.323
4	57.8	76.3	67.3	64.7	57.5	0.372
6	57.3	74.0	65.9	62.9	54.5	0.401
8	57.0	72.6	65.0	61.8	52.7	0.418
10	56.6	71.7	64.4	61.1	51.8	0.428
12	56.4	71.1	64.0	60.6	50.9	0.435
14	56.3	70.7	63.7	60.3	50.0	0.441
16	56.2	70.4	63.5	60.0	49.4	0.445
18	56.2	70.2	63.4	59.9	48.7	0.448
20	56.3	70.0	63.4	59.7	48.3	0.451
22	55.9	69.8	63.0	59.3	47.3	0.459
24	55.4	69.4	62.6	58.7	46.0	0.468
26	54.9	69.1	62.2	58.2	44.6	0.477
28	54.8	68.8	62.0	57.7	43.1	0.486
30	55.1	68.5	62.0	57.4	41.4	0.494
32	55.6	68.3	62.1	57.3	40.0	0.499
34	56.2	68.2	62.2	57.3	39.4	0.500
36	57.2	68.1	62.8	57.7	38.4	0.498
38	58.5	68.1	63.4	58.2	38.3	0.494
40	59.9	68.1	64.1	58.8	38.4	0.487

Test 3

The MIR phases which were available to us for the 1.9 Å data had a MPEM of 61.5° rather than the much lower value of 52° which was reported by Agarwal & Isaacs (1977). The first approach we tried was the simple one of starting with MIR phases for the 1.9 Å data and random phases for the 6842 reflections between 1.9 and 1.5 Å. We adopted the usual scheme of combining new phase estimates, $\varphi_{\rm NEW}$, with MIR phases, $\varphi_{\rm MIR}$, so that the phase taken forward to the next cycle of refinement is,

$$\varphi = \text{phase of } [w_{\rm MIR} \exp(i\varphi_{\rm MIR}) + (1 - w_{\rm MIR}) \exp(i\varphi_{\rm NEW})], \tag{20}$$

where $w_{\rm MIR}$ is the weight from the MIR phasing procedure. However, after 20 cycles of refinement we relaxed the MIR weighting according to the cycle number, n, so that the weight given to the MIR phases was,

$$w'_{\rm MIR} = w_{\rm MIR} \times [(40 - n)/20], \tag{21}$$

Table 4. *Phase extension and refinement for* 2-Zn *insulin with extended phases given values derived from the square of the MIR-phased map*

All quantities given are as in Table 3.

Cycle	MPEM	MPEE	MPET	WMPET	MPEL	MCC
0	61.5	70.8	66.3	63.4	52.6	0.395
2	57.1	69.9	63.7	60.1	47.5	0.447
4	56.2	68.5	62.6	58.8	45.7	0.467
6	56.0	67.8	62.1	58.2	45.2	0.474
8	55.8	67.3	61.7	57.9	44.9	0.478
10	55.8	67.1	61.6	57.7	44.6	0.481
12	55.8	67.0	61.6	57.7	44.5	0.483
14	55.7	66.8	61.4	57.5	44.4	0.485
16	55.7	66.7	61.4	57.4	44.1	0.487
18	55.8	66.5	61.3	57.4	43.9	0.487
20	55.8	66.5	61.3	57.4	43.8	0.488
22	55.2	66.4	61.0	57.0	42.9	0.495
24	54.6	66.3	60.7	56.5	41.9	0.502
26	54.2	66.2	60.4	56.1	40.9	0.509
28	54.1	66.1	60.3	55.8	39.9	0.515
30	54.4	66.0	60.4	55.6	39.0	0.520
32	54.8	65.9	60.5	55.6	38.3	0.523
34	55.3	65.9	60.7	55.7	37.9	0.523
36	55.9	66.1	61.2	56.1	37.6	0.520
38	57.2	67.0	62.2	57.0	37.8	0.509
40	58.6	67.0	62.9	57.7	38.0	0.502

up to $n = 40$. As will be seen from Table 3 this accelerated the improvement in phases after cycle 20 and the best phases, according to the MCC value (0.500) were obtained in cycle 34. Because of the very different starting points it is not straightforward to compare the result of this refinement with that of Agarwal & Isaacs (1977). Their final MPE for the whole data set was 3° greater than for the original MIR phases while in Table 3 it is only 0.7° higher. As far as can be assessed the two approaches give a comparable performance.

Test 4

This was similar to test 3 except that the starting phases for the reflections between 1.9 and 1.5 Å resolution were obtained by Fourier transformation of the squared 1.9 Å map. This is equivalent to finding initial phase estimates for the extended reflection data from the tangent formula using only MIR phases on the right-hand side. It will be seen from Table 4 that not only is the starting point of the refinement a better one but that a better final MCC (0.523) is obtained, again in cycle 34. For this approach the total MPE is 0.8° less than the MPE of the original MIR phases.

Concluding remarks

From our results we conclude that the Sayre equation on its own is an effective tool for phase refinement if used with a complete data set at about atomic resolution (1.5 Å) or better. Trials at lower resolutions show that not much benefit is available. This does not fly in the face of the conclusion by Shiono & Woolfson (1991) that Sayre's equation holds reasonably well for low resolution and with unequal atoms; one must distinguish the approximate validity of the equation from its ability to constrain phases towards their correct values. For a complete data set at atomic resolution or better the equations are well conditioned and can be used as a useful phasing tool on their own but the conditioning quickly degenerates as the resolution is reduced.

Where Sayre's equation does come into its own, even at well below atomic resolution, is when it is used as a constraint on phases in conjunction with other phase refining processes, as used by Zhang & Main (1990*b*) in their *SQUASH* procedure. We have found that inclusion of a Sayre-equation refinement stage may even make the phases worse in terms of MPE but better in terms of the ability of another process to take over the refinement. There is some element of conditioning the set of phases, which we do not completely understand and which we do not know how to describe, which comes about from applying Sayre's equation and makes it a valuable refinement tool even well outside the range of its stand-alone validity.

This paper owes much to the helpful comments of a referee of the original version and we take pleasure in acknowledging that assistance. We also wish to express our gratitude to the Science and Engineering Research Council for the support of this and related work.

References

Agarwal, R. C. & Isaacs, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.

Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M. C., Hubbard, R. E., Isaacs, N. W., Reynolds, C. D., Sakabe, K., Sakabe, N. & Vijayan, N. M. (1988). *Philos. Trans. R. Soc. London Ser. B*, **319**, 456–469.

Bezborodova, S. I., Ermekbaeva, L. A., Shlyapnikov, S. V., Polyakov, K. M. & Bezborodov, A. M. (1988). *Biokhimiya*, **53**, 965–973.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* **D49**, 148–157.

Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* **D49**, 530–533.

Mishnev, A. F. & Woolfson, M. M. (1994). *Acta Cryst.* **D50**, 842–846.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Sayre, D. (1972). *Acta Cryst.* **A28**, 210–212.

Sayre, D. (1974). *Acta Cryst.* **A30**, 180–184.

Sayre, D. (1975). In *Crystallographic Computing Techniques*, edited by F. R. Ahmed, pp. 322–327. Copenhagen: Munksgaard.

Shino, M. & Woolfson, M. M. (1991). *Acta Cryst.* **A47**, 526–533.

Zhang, K. Y. J. & Main, P. (1990*a*). *Acta Cryst.* **A46**, 41–46.

Zhang, K. Y. J. & Main, P. (1990*b*). *Acta Cryst.* **A46**, 377-381.