References

ANDERSON, D. M., NITSCHE, J. C. C., DAVIS, H. T. & SCRIVEN, L. E. (1990). *Adv. Chem. Phys.* **77**, 337.

COXETER, H. S. M. (1973). *Regular Polytopes.* New York: Dover.

DUBOIS-VIOLETTE, E. & PANSU, B. (1990). Editors. *J. Phys. (Paris) Colloq.* C7.

ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. & TRICOMI, F. G. (1953). *Higher Transcendental Functions.* Bateman Manuscript Project. New York: McGraw-Hill.

FISCHER, W. & KOCH, E. (1989). *Acta Cryst.* A**45**, 726-732.

HOPF, H. (1983). *Differential Geometry in the Large. Springer Lecture Notes in Mathematics.* Berlin: Springer.

HYDE, S. T. (1989). *Z. Kristallogr.* **187**, 165-185.

HYDE, S. T. & ANDERSSON, S. (1985). *Z. Kristallogr.* **170**, 225-239.

KARCHER, H. (1989). *Manuscr. Math.* **64**, 291-337.

KOCH, E. & FISCHER, W. (1990). *Acta Cryst.* A**46**, 33-40.

LIDIN, S. & HYDE, S. T. (1987). *J. Phys. (Paris)*, **48**, 1585-1590.

NEOVIUS, E. R. (1883). *Bestimmung Zweier Speciellen Periodische Minimalflaechen.* Helsinki: Frenckel.

NISSEN, H.-U. (1969). *Science*, **166**, 1150-1152.

NITSCHE, J. C. C. (1975). *Vorlesungen über Minimalflächen.* Berlin: Springer Verlag.

RIEMANN, B. (1953). *Über die Fläche vom kleinsten Inhalt bei gegebener Begrenzung. Gesammelte Mathematische Werke.* New York: Dover.

SCHOEN, A. H. (1970). *Infinite Periodic Minimal Surfaces Without Self-intersections.* NASA Technical Note No. D5541, Washington, DC, USA.

SCHWARZ, H. A. (1890). *Gesammelte Mathematische Abhandlungen.* Berlin: Springer.

SPIVAK, M. (1979). *A Comprehensive Introduction to Differential Geometry.* Berkeley: Publish or Perish.

SPRINGER, G. (1957). *Introduction to Riemann Surfaces.* Reading, MA: Addison-Wesley.

WILLIAMS, R. (1979). *The Geometrical Foundation of Natural Structure.* New York: Dover.

Acta Cryst. (1992). A**48**, 451-456

Direct-Space Methods in Phase Extension and Phase Determination. I. Low-Density Elimination

BY M. SHIONO* AND M. M. WOOLFSON

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

(*Received 9 August* 1991; *accepted* 10 *December* 1991)

Abstract

A density-modification procedure for phase extension and refinement is described which replaces all density less than one-fifth of the height of a light-atom peak by zero. Its effectiveness is demonstrated by applications to a small and a medium-size protein structure. With high-resolution data, for the small protein, it is possible to extend and refine from 3 to 1 Å with a mean phase error less than 30°. Successful phase extension from 4 Å is also possible. In general it is shown that phase extension to high resolution gives less error than extension to lower resolution. It has also been shown that for a small protein it is possible to obtain an *ab initio* solution of the structure by refining from a complete set of random phases for all reflexions.

The basis of direct methods

Most direct methods consist of mathematical procedures carried out in reciprocal space which are designed to produce sets of phases satisfying particular constraints. The first powerful and generally applicable direct methods were those based on the tangent formula, introduced by Karle & Hauptman (1956),

$$\tan \varphi(\mathbf{h}) = \left\{\sum_{\mathbf{k}} |E(\mathbf{k})E(\mathbf{h}-\mathbf{k})| \sin[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]\right\} \times \left\{\sum_{\mathbf{k}} |E(\mathbf{k})E(\mathbf{h}-\mathbf{k})| \cos[\varphi(\mathbf{k})+\varphi(\mathbf{h}-\mathbf{k})]\right\}^{-1}. \quad (1)$$

Although the tangent formula was derived by Karle & Hauptman from algebraic and statistical considerations, it can be given a real-space physical interpretation. The phase given by the tangent formula is just that which would be obtained if an E map, calculated with current phase estimates, was squared and the phase, $\varphi(\mathbf{h})$, of the Fourier coefficient of index $\mathbf{h}$ of the squared map was taken.

The precursor of the tangent formula, the three-phase relationship

$$\varphi(\mathbf{h}) - \varphi(\mathbf{k}) - \varphi(\mathbf{h}-\mathbf{k}) \simeq 0 \text{ (modulo } 2\pi) \quad (2)$$

was derived by Cochran (1955) from the condition that a set of correct phases should give an electron density map for which $\int_V \rho^3 \, dV$ is a maximum. This condition, somewhat intuitive in origin, expresses the

---

[*] Present address: Department of Physics, Kyushu University, Hakozaki 6-chome, Higashi-ku, Fukuoka, Japan.

0108-7673/92/040451-06$06.00

idea that the electron density will be everywhere positive and also be concentrated around atomic centres.

To strengthen the tangent formula, higher-order phase relationships, notably quartets (Schenk, 1973; Hauptman, 1977), have sometimes been employed. In the simplest form, a quartet relationship involving the phases of four Es of large magnitude is

$$\varphi(\mathbf{h})-\varphi(\mathbf{k})-\varphi(\mathbf{l})-\varphi(\mathbf{h}-\mathbf{k}-\mathbf{l})\simeq 0 \text{ (modulo } 2\pi) \tag{3}$$

and may be thought of as expressing the condition that $\int_V \rho^4 \, \mathrm{d}V$ is a maximum. However, the even power of ρ in the integral does not discriminate against negative ρ, and quartets are best considered from a statistical point of view. In fact, Schenk (1973) and Hauptman (1977) showed that if the 'cross terms', $E(\mathbf{h}-\mathbf{k})$, $E(\mathbf{k}-\mathbf{l})$ and $E(\mathbf{l}-\mathbf{h})$, are all small in magnitude then

$$\varphi(\mathbf{h})-\varphi(\mathbf{k})-\varphi(\mathbf{l})-\varphi(\mathbf{h}-\mathbf{k}-\mathbf{l})\simeq \pi \text{ (modulo } 2\pi), \tag{4}$$

which is difficult to interpret in terms of E-map density.

A recent addition to direct-methods phasing formulae is the Sayre-equation tangent formula (Debaerdemaeker, Tate & Woolfson, 1985, 1988). This formula contains both three-phase and quartet derived terms but is chosen so that the derived phases tend to satisfy Sayre's equation

$$E(\mathbf{h})=\theta(\mathbf{h})\sum_{\mathbf{k}} E(\mathbf{k})E(\mathbf{h}-\mathbf{k}) \tag{5}$$

(Sayre, 1952), where $\theta(\mathbf{h})$ is known. This equation expresses the condition that if the E-map density shows equal resolved atoms then the squared density will show equal and resolved 'squared' atoms. In fact, it has been shown by Shiono & Woolfson (1991) that Sayre's equation holds remarkably well in many situations when the atoms are unresolved and unequal.

A final example of a very mathematical approach with a real-space condition attached to it is the use of Karle–Hauptman determinant inequalities (Karle & Hauptman, 1950). These are of the form

$$\begin{vmatrix} E(\mathbf{0}) & E(\mathbf{h}_1) & E(\mathbf{h}_2) & \cdots & E(\mathbf{h}_n) \\ E(\bar{\mathbf{h}}_1) & E(\mathbf{0}) & E(\mathbf{h}_2-\mathbf{h}_1) & \cdots & E(\mathbf{h}_n-\mathbf{h}_1) \\ E(\bar{\mathbf{h}}_2) & E(\mathbf{h}_1-\mathbf{h}_2) & E(\mathbf{0}) & \cdots & E(\mathbf{h}_n-\mathbf{h}_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E(\bar{\mathbf{h}}_n) & E(\mathbf{h}_1-\mathbf{h}_n) & E(\mathbf{h}_2-\mathbf{h}_n) & \cdots & E(\mathbf{0}) \end{vmatrix} \geq 0 \tag{6}$$

and they express the condition that electron density (or E-map density for an infinite data set) should be non-negative (Toeplitz, 1911).

The existing practice is that phase extension and refinement, particularly for proteins, are carried out either by real-space methods – for example, Wang's solvent flattening method (Wang, 1985) – or by reciprocal-space methods – for example, by applying the maximum determinant rule (Tsoucaris, 1970). On the other hand, phase determination by direct methods is done almost exclusively in reciprocal space by applying relationships between phases.

Although many direct-methods procedures could be translated into real-space processes, reciprocal-space operations have the advantage that they are more economical in computer time as long as one is interested in phasing only a subset of Es of largest magnitude. For example, in a typical direct-methods solution of a small structure, it is adequate to find phases for, say, 300 reflexions and these may be linked by 5000 three-phase relationships. However, experience indicates that if large structures, *e.g.* proteins, are to be tackled using direct methods then a large proportion of the data, ideally all the data, must be deployed. The number of three-phase relationships increases approximately as the square of the number of reflexions but the real-space operation of calculating a Fourier transform, using the standard fast Fourier transform (FFT) algorithm, depends, to first order, directly on the number of coefficients. Since we are interested in developing methods for solving macromolecular structures by directly determining phases, we have recently been turning our attention to real-space procedures and we report here our first experiences in this direction.

The low-density modification procedure

The basic program we set up to examine direct-space methods consisted of three components:

(1) calculate electron density from structure factors (including phase) with a FFT routine;

(2) modify density;

(3) calculate structure factors from the modified density using the FFT routine.

It was our expectation that we would be looking at some quite complex density modifications in which we would be forcing the density towards some idealized configuration. In the event, the first very simple density modification we tried was so effective that we have concentrated on exploring its properties and limitations. We shall be presenting the procedures we have found by trial and error to be most useful; we cannot completely justify them or the parameters we use by any formal theoretical reasoning.

The density modification we have used is low-density elimination (LDE). This consists of replacing both negative and positive density that is less than a certain magnitude by zero. The practical details of this procedure are as follows.

(i) Calculate the E map in sections, *e.g.* at intervals in y of approximately $1/(4k_{\max})$, which ensures that the sampling effectively and efficiently follows the variation of density. Similar grid intervals are used in the other principal directions. All reflexions of known phase are used in the E-map calculation, including $E(\mathbf{0})$.

(ii) Find the average peak height, ρ_c, of the light atoms. This is estimated by finding the average of the maximum density in each section. Thus if there are M sections and the maximum density in the jth section is $\rho_{\max,j}$ then

$$\rho_c = \left(\sum_{j=1}^{M} \rho_{\max,j}\right) \Big/ M. \qquad (7)$$

If there are heavy atoms in the structure then their influence can be removed by eliminating any sectional maximum density appreciably higher than the average.

(iii) Modify the E-map density by

$$\rho_{\text{new}} = \rho, \qquad \rho \geq 0.2\rho_c;$$

$$\rho_{\text{new}} = 0, \qquad \rho < 0.2\rho_c.$$

(iv) Fourier transform the modified density to give Fourier coefficients

$$\mathfrak{F}(\mathbf{h}) = |\mathfrak{F}(\mathbf{h})| \exp\left[i\varphi_c(\mathbf{h})\right].$$

(v) Calculate a weighted E map with coefficients

$$w(\mathbf{h})|E(\mathbf{h})| \exp\left[i\varphi_c(\mathbf{h})\right],$$

where

$$w(\mathbf{h}) = \tanh\left[K|\mathfrak{F}(\mathbf{h})E(\mathbf{h})|/2\right] \qquad (8)$$

and K, a normalizing constant for the Q values of $\mathfrak{F}(\mathbf{h})$ is calculated from

$$K = \left(Q \Big/ \sum_{\mathbf{h}} |\mathfrak{F}(\mathbf{h})|^2\right)^{1/2}. \qquad (9)$$

If the LDE method is being used for phase extension from a small base of known phases then these are not changed until the final cycle of the procedure. The weighting scheme, which is similar to that derived by Sim (1960), was based on a function whose value goes from zero to one as the argument goes from zero to infinity with the values $\mathfrak{F}(\mathbf{h})$ and $E(\mathbf{h})$ involved in a sensible way. The constant parts of the argument were chosen by trial and error but the behaviour of the phase extension and refinement procedure does not depend upon them too sensitively.

(vi) Go to (ii) to begin the next cycle.

All the trials we have made so far have been on known structures but using real observed data. To judge the effectiveness of the procedure, we have used mean phase errors, either unweighted or weighted in three different ways. These are:

$$\langle\Delta\varphi\rangle = \sum_{\mathbf{h}} |\varphi_t(\mathbf{h}) - \varphi_c(\mathbf{h})|/Q; \qquad (10)$$

$$\langle\Delta\varphi\rangle_E = \left[\sum_{\mathbf{h}} |E(\mathbf{h})||\varphi_t(\mathbf{h}) - \varphi_c(\mathbf{h})|\right] \Big/ \sum_{\mathbf{h}} |E(\mathbf{h})|; \qquad (11)$$

$$\langle\Delta\varphi\rangle_F = \left[\sum_{\mathbf{h}} |\mathfrak{F}(\mathbf{h})||\varphi_t(\mathbf{h}) - \varphi_c(\mathbf{h})|\right] \Big/ \sum_{\mathbf{h}} |\mathfrak{F}(\mathbf{h})|; \qquad (12)$$

$$\langle\Delta\varphi\rangle_{EF} = \left[\sum_{\mathbf{h}} |E(\mathbf{h})\mathfrak{F}(\mathbf{h})||\varphi_t(\mathbf{h}) - \varphi_c(\mathbf{h})|\right] \Big/ \sum_{\mathbf{h}} |E(\mathbf{h})\mathfrak{F}(\mathbf{h})|; \qquad (13)$$

$\varphi_t(\mathbf{h})$ is the 'true' phase calculated from the coordinates of the refined structure.

Phase extension for a small protein

We applied LDE to a small protein, avian pancreatic polypeptide (App) (Glover *et al.*, 1983). The space group is $C2$ with dimensions $a = 34.18$, $b = 32.92$, $c = 28.45$ Å and $\beta = 105.26°$. The asymmetric unit contains 301 non-hydrogen atoms belonging to the protein, a zinc atom and 80 H_2O molecules. The data are of good quality by protein crystallography standards and extend to a resolution of 0.98 Å.

Our first experiments were in extending phases from 3 Å to various resolutions. In the original solution of this structure, phases were found out to 3 Å resolution by a combination of one-wavelength anomalous scattering and single isomorphous replacement using data from a mercury derivative. The mean phase error for these low-resolution reflexions was less than 30°. In our experiments, we introduced the phase errors artificially by the formula

$$\tan\varphi(\mathbf{h}) = [p \sin\varphi_t(\mathbf{h}) + (1-p)\sin\varphi_r(\mathbf{h})] \times [p\cos\varphi_t(\mathbf{h}) + (1-p)\cos\varphi_r(\mathbf{h})]^{-1}, \qquad (14)$$

where $\varphi_t(\mathbf{h})$ is the true phase and $\varphi_r(\mathbf{h})$ is a random phase chosen from a distribution with uniform probability in the range from 0 to 2π. If $p = 0$ then the phases $\varphi(\mathbf{h})$ are completely random; on the other hand if $p = 1$ then $\varphi(\mathbf{h})$ are the true phases. We chose a value of p such that the initial low-resolution phases had a mean phase error of approximately 30°.

The advantage of using random phase errors calculated in this way is that we could systematically explore the way in which the effectiveness of the method depended on the errors. In practice, real phase errors as derived from, say, multiple isomorphous replacement are not truly random; there will be a tendency for the errors to be greater for the smaller structure factors and for those of higher resolution (Cannillo, Oberti & Ungaretti, 1983). For a given mean phase error of a set of reflexions, a MIR set will be better than the set we use because we would have greater error in the stronger reflexions - although the resolution effect will be in the other direction.

Our process was divided into two distinct stages - phase extension followed by phase refinement. The initial decisions taken were that n_1 new phases would be estimated in the first cycle of phase extension and that the phase extension process would consist of L

Table 1. *Phase extension and refinement from* 3 *and* 4 Å *to the stated resolution for App*

The values in parentheses in the column 'Initial number of reflexions' are the unweighted and weighted mean phase errors in degrees for the low-resolution starting reflexions after refinement. The initial errors for these phases are approximately 30°.

Resolution from (Å)	Resolution to (Å)	Number of refinement cycles	Initial number of reflexions	Final number of reflexions	$\langle\Delta\varphi\rangle$ (°)	$\langle\Delta\varphi\rangle_{EF}$ (°)
3.0	1.0	77	614 (33.2; 18.7)	16 524	27.6	15.5
	1.25	178	(38.5; 22.8)	8 502	40.4	23.4
	1.50	191	(49.6; 31.3)	4 932	50.9	32.6
	1.75	90	(53.6; 34.5)	3 118	65.9	49.0
	2.00	63	(58.0; 37.9)	9 095	72.2	56.5
	2.25	47	(54.0; 35.8)	1 475	69.1	54.8
	2.50	23	(60.5; 35.5)	1 081	68.2	49.3
4.0	1.0	144	253 (39.6; 21.8)	16 524	30.8	18.0

cycles. Then the number of estimated phases in cycle j is

$$n_j = n_1\gamma^{j-1} \tag{15}$$

where γ is found from the total number of phases to be estimated by extension, n_{ext}, by

$$n_{\text{ext}} = \sum_{j=1}^{L} n_j = n_1(\gamma^L - 1)/(\gamma - 1). \tag{16}$$

For example, with $n_1 = 30$, $n_{\text{ext}} = 15\,910$ and $L = 12$, the value of γ is found to be 1.62162 and the numbers of phases in the extension process in successive cycles are 30, 49, 79, 128, 207, 336, 546, 885, 1435, 2326, 3772, 6117.

In extending phases cycle by cycle, estimates are accepted in order of the magnitude of $w(\mathbf{h})|E(\mathbf{h})|$ where $w(\mathbf{h})$ is given in (8). During the process of phase extension, phases are unchanged once they have been estimated and the magnitude of the corresponding Fourier coefficient is also unchanged in the subsequent maps; changes to the initially estimated phases are only made when the phase-refinement process is begun.

Once an estimated phase is available for every reflexion in the extension set, cycles of phase refinement are set in train following the steps (ii) to (vi) previously described. The process is continued until the average change of phase in each of several successive cycles is less than some predetermined limit (say 0.1°); one more cycle is then gone through in which the original low-resolution phases are allowed to change. It is necessary to have stability for several cycles before terminating the refinement as the process has the characteristic that it sometimes pauses momentarily and then takes off again.

The results of phase extension from 3 Å for App are shown in Table 1. Two interesting conclusions can be drawn from this. The first is that extending to 1 Å, the limit of the data, gives a very low mean phase error; a map with these phases shows the structure clearly. The second point is that extension to low resolution is much less accurate than extension to high resolution; instinctively it might be thought that a modest extension from 3 to 2.5 Å should give very accurate phases since the extrapolation range is small, but instinct is wrong in this case.

Finally, in Table 1 we show the result of phase extension from 4 to 1 Å, a much greater range than is usually attempted. Once again, the final phase error is low and it will be clear that this process of phase extension is very robust.

The influence of heavy atoms

From previous experience in applying the Sayre-equation tangent formula to App for phase extension (Yao & Woolfson, 1988) and for *ab initio* solution (Woolfson & Yao, 1990), it was suspected that the Zn atom was having an important influence on the LDE phase extension. To test this we modified the normalized structure factors to correspond to those for a zinc-free (ZF) structure. This was done by the formula

$$E_{\text{ZF}}(\mathbf{h}) = E(\mathbf{h}) - [f_{\text{Zn}}(\mathbf{h})^2/\sigma_2]^{1/2}E_{\text{Zn}}(\mathbf{h}), \tag{17}$$

where $\sigma_2 = \sum_{j=1}^{N} f_j(\mathbf{h})^2$, $f_j(\mathbf{h})$ is the scattering factor for the jth atom, $f_{\text{Zn}}(\mathbf{h})$ is the scattering factor for zinc and $E_{\text{Zn}}(\mathbf{h})$ is the normalized structure factor for zinc calculated from the known coordinates. By doing this calculation, we were able to eliminate the contribution of zinc while retaining the experimental error in the structure-factor magnitudes. The results of extending phases from 3 Å to 1.0, 1.25 and 1.5 Å are shown in Table 2(a). The results are much inferior to those for the zinc-containing structure and the results are probably not useful for phase extension to less than 1.25 Å resolution.

Since many proteins contain S-S bridges, we also explored the effect of replacing Zn with S atoms. The results, shown in Table 2(b), are very little different from those without a heavy atom at all and sulfur seems to give no assistance in phase refinement.

In fact, it turns out that if the positions of the Zn atoms are known and high-resolution data are avail-

Table 2. *Phase extension and refinement for App*

(*a*) App with zinc contribution removed from 3 Å (614 reflexions) to the stated resolutions. The initial phase errors for the initial phases are approximately 30°. (*b*) As for (*a*) except that the App structure has sulfur atoms substituted for zinc.

	Resolution from (Å)	Resolution to (Å)	Number of refinement cycles	Final errors in initial reflexions	Final number of reflexions	$\langle\Delta\varphi\rangle$ (°)	$\langle\Delta\varphi\rangle_{EF}$ (°)
(*a*)							
	3.0	1.0	92	(33.4; 18.7)	16 524	42.4	26.3
		1.25	235	(46.1; 25.3)	8 502	54.2	33.5
		1.50	160	(54.2; 36.4)	4 932	72.3	57.0
(*b*)							
	3.0	1.25	200	(46.2; 26.1)	8 502	54.8	34.3
		1.50	148	(52.4; 35.7)	4 932	73.6	58.3

Table 3. *Results of refinement using initial phases from the zinc position*

Resolution (Å)	Number of reflexions	Initial $\langle\Delta\varphi\rangle$(°)	Number of cycles	$\langle\Delta\varphi\rangle$ (°)	$\langle\Delta\varphi\rangle_{EF}$ (°)
1.00	16 524	66.9	39	32.9	20.2
1.25	8 502	64.2	80	62.0	45.1
1.50	4 932	65.9	88	73.2	61.1

able then the App structure can be solved from this information alone with the LDE technique. Starting phases are assigned from the zinc contributions to all reflexions and are then refined in the usual way. The results are shown in Table 3. With 1 Å data, the final mean phase error is very low and it is clear that the structure has been solved by this process. For resolutions 1.25 and 1.5 Å, the results are very much inferior, although a weighted 1.25 Å map could provide a starting point for a detailed solution. The phases after refinement for 1.5 Å are actually worse than those obtained from the zinc position alone.

Application to a larger protein

We next tried the LDE procedure on a larger test structure, ribonuclease Apl (RNApl) (Polyakov, 1991). This has space group $P2_1$ with $a = 32.01$, $b = 49.76$, $c = 30.67$ Å and $\beta = 115.83°$. The asymmetric unit contains 808 non-H atoms, including five S atoms, and 83 water molecules. There are 23 853 independent observed reflexions out to 1.17 Å resolution.

Our goal here was to see how well initially poor phases could be refined. By the use of (14) we assigned phases to all reflexions with a 74° mean phase error. Experiences with the use of one-wavelength anomalous-scattering data (Fan, Hao & Woolfson, 1990) indicate that for a structure the size of RNApl, or even larger, a mean phase error of 70°, or even less, is attainable - perhaps even with comparatively unfavourable anomalous scattering from sulfur. After 293 cycles of refinement, the mean phase errors, unweighted and weighted, were: $\langle\Delta\varphi\rangle = 50.2$, $\langle\Delta\varphi\rangle_F = 45.1$, $\langle\Delta\varphi\rangle_E = 39.7$ and $\langle\Delta\varphi\rangle_{EF} = 34.6°$. There is a considerable improvement in phases and also in the quality of the resultant maps. The other interesting point is that the weighted errors show that the reliabilities of the phase estimates are well indicated by the weighting schemes and this is reflected in the improved quality of maps when calculated with the normal weights as given by (8), as compared with maps using unweighted coefficients.

During the investigation of refining RNApl we discovered a device for speeding up the refinement. Early experiments in which density was just squared showed that it usually gives markedly improved phases for a few cycles but after that it leads to a rapid deterioration of the phases. In the RNApl refinement, for the first three cycles, negative density was made equal to zero and then the modified density was squared. This accelerated the refinement without affecting its final outcome.

Ab initio solution by the LDE method

The default procedure on many presently used direct methods is to start a process of phase refinement from initially random phases. This has been done using the LDE process with the 1 Å data for App. Forty trials were made starting with different sets of random phases for all 16 524 reflexions. After 106 cycles of refinement, the sixth trial gave a set of phases with the unweighted and weighted mean errors: $\langle\Delta\varphi\rangle = 33.9$, $\langle\Delta\varphi\rangle_F = 28.0$, $\langle\Delta\varphi\rangle_E = 25.2$ and $\langle\Delta\varphi\rangle_{EF} = 20.7°$.

The potential of the LDE method as a means of solving large, perhaps macromolecular, structures will be the subject of further study.

Discussion

It was a surprise to us that a density-modification process as simple as LDE could be so powerful as a phase-refining procedure. Part of the process, that which eliminates the negativity of the map, can clearly be related to the physical condition that electron density cannot be negative and this is the basis of some reciprocal-space direct-methods procedures. A referee has also reminded us of a paper by Barrett & Zwick (1971) in which phase extension and refinement was carried out by a process of first removing the negativity of a map and then squaring the modified map. On rereading this paper we find in it a reference to a private communication from G. Kartha, dated 1969, in which a method similar to that we describe here is mentioned as having been applied to some small-molecule structures; no further reference to this approach can be found.

The LDE procedure which replaces not only negative density but also some positive density by zero cannot so easily be rationalized - although it might be related to the solvent flattening idea used by Wang (1985). It should be said that we investigated the level for elimination in the form $c\rho_c$ for various values of c. While small variation about $c = 0.2$ makes little difference, the values $c = 0$ and $c = 0.4$ are markedly inferior in performance to $c = 0.2$. The selection of a threshold for replacement by zero was also done by Cannillo, Oberti & Ungaretti (1983), who replaced everything above the threshold by a constant value. Their method worked reasonably well for ideal data with both magnitudes and phases calculated from refined structure coordinates but less well with real data.

Another characteristic of the LDE method, which it has in common with more conventional direct methods, is that its effectiveness depends greatly on the resolution of the data being deployed; this is clear from the results displayed in Tables 1 and 2. We suspect that for low-resolution data it may be better to use Fs rather than Es, or at least Es modified to correspond to low-resolution atoms with little negative diffraction ripple. Our experience also leads us to question the usual philosophy in phase extension which is to do it by degrees - first from 3 to 2.5 then to 2.2 Å *etc.* The message that we obtain from our work is that all available data should be used from the very beginning. Data of the resolution that we have for App and RNApl are the exception rather than the rule in protein crystallography and it must be emphasized that the good results here were highly dependent on the data resolution - as is evident from all the tables. One of our future aims is to find ways of improving the performance of density-modification methods at lower resolution.

A final point is that the LDE method, which requires two Fourier transforms per cycle, is very expensive in computer time. At present, with supercomputer capability becoming available at desk-top level, we feel that this is a matter of little importance; within reason we can expect computers to catch up with requirements within a very short timespan.

We thank C. Tate, L. S. Refaat, Yao Jia-xing and Hao Quan for helpful discussions. We are also grateful to the Science and Engineering Research Council, the Wellcome Trust and the Wolfson Foundation for their support of direct-methods activity in York.

References

BARRETT, A. H. & ZWICK, M. (1971). *Acta Cryst.* **A27**, 6-11.

CANNILLO, E., OBERTI, R. & UNGARETTI, L. (1983). *Acta Cryst.* **A39**, 68-74.

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473-478.

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286-290.

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353-357.

FAN, H., HAO, Q. & WOOLFSON, M. M. (1990). *Acta Cryst.* **A46**, 656-659, 659-664.

GLOVER, I., HANEEF, I., PITTS, J., WOOD, S., MOSS, T., TICKLE, I. & BLUNDELL, T. (1983). *Biopolymers*, **22**, 293-304.

HAUPTMAN, H. (1977). *Acta Cryst.* **A33**, 553-555.

KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181-187.

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

POLYAKOV, K. M. (1991). Private communication.

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

SCHENK, H. (1973). *Acta Cryst.* **A29**, 77-82.

SHIONO, M. & WOOLFSON, M. M. (1991). *Acta Cryst.* **A47**, 526-533.

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

TOEPLITZ, O. (1911). *Rend. Circ. Mat. Palermo*, **32**, 191-192.

TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492-499.

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

WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* **A46**, 409-413.

YAO, J. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 410-413.