Acta Cryst. (1993). D**49**, 367–371

Direct-Space Methods in Phase Extension and Phase Determination. II. Developments of Low-Density Elimination

BY L. S. REFAAT AND M. M. WOOLFSON

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

(*Received* 24 *February* 1993; *accepted* 30 *March* 1993)

Abstract

The low-density elimination method for phase extension and refinement [Shiono & Woolfson (1992). *Acta Cryst.* A**48**, 451–456] has been improved by substituting a smoother density-modification procedure for the original sharp cut-off function. In addition, better criteria have been found for limiting the number of refinement cycles, which gives a better final result for much less work. The effectiveness of the process has been illustrated by phase refinement for a protein with high-resolution (1.17 Å) data containing 808 independent non-H atoms plus 83 water molecules in the asymmetric unit; the unweighted mean-phase error was reduced from 74 to 39.3°. Phase extension and refinement was also demonstrated for pig 2Zn insulin starting with multiple isomorphous replacement (MIR) phases at 1.9 Å and extending out to 1.5 Å. There was a significant improvement of phases and the final map had a correlation coefficient of 0.540.

Introduction

In the first paper of this series (Shiono & Woolfson, 1992, referred to as I) the low-density elimination (LDE) method was shown to be able not only to extend and refine the phases of a small protein but also to give an *ab initio* solution. The basic idea of the method is simply to transform the density by

$$\rho'(\mathbf{r}) = \rho(\mathbf{r}) \quad \rho(\mathbf{r}) > 0.2\rho_c$$

and

$$\rho'(\mathbf{r}) = 0 \qquad \rho(\mathbf{r}) \le 0.2\rho_c, \tag{1}$$

where ρ_c is the expected height of a light-atom peak in the protein region, to find which an objective procedure was described. The Fourier transform of the modified density then gave new phase estimates which were used to compute a new $\rho(\mathbf{r})$ for the next refinement cycle.

The example used in paper I was the structure of avian pancreatic peptide (aPP), a small globular protein with 36 amino acids which crystallizes with Zn^{2+} in space group $C2$ with unit-cell parameters $a = 34.18$, $b = 32.92$, $c = 28.44$ Å, $\beta = 105.30°$ and $Z = 4$ (Blundell, Pitts, Tickle, Wood & Wu, 1981). There are also 80 water molecules in each asymmetric unit. The LDE process was very effective but converged rather slowly and the initial goal of the work presented here was to try to accelerate the convergence. It occurred to us that the form of density modification given by (1) led to density with discontinuities, corresponding to a map with large amplitudes for high-resolution Fourier coefficients. The net effect was that in the succeeding map, diffraction ripples would appear in the region of the discontinuity thus introducing negativity which it was the purpose of the method to eliminate. For this reason we decided to use a smoother modification function and several forms were tried. The one which we eventually decided upon as the most satisfactory was

$$\rho'(\mathbf{r}) = \frac{\rho(\mathbf{r})^{n+1}}{\rho_c(\mathbf{r})^n + \rho(\mathbf{r})^n} \quad \rho(\mathbf{r}) > 0$$

and

$$\rho'(\mathbf{r}) = 0 \qquad \rho(\mathbf{r}) \le 0, \tag{2}$$

where n is some integer greater than unity. This modification makes $\rho'(\mathbf{r})$ both smooth and continuous; as n is increased so this modification approaches (1) and we explored values of n in the range 5–10.

A test with aPP

The original solution of aPP was obtained by a combination of single isomorphous replacement and one-wavelength anomalous scattering, both using a mercury derivative. This combination gave phases out to a resolution of about 3.0 Å with a mean-phase error (m.p.e.) slightly less than 30°. We chose as our starting point 614 reflections within the 3.0 Å resolution sphere with an m.p.e. of 30° imposed on the phases calculated from the refined structure. In Table 1 we show what happens when LDE phase extension and refinement is carried out by both the modification formulae (1) and (2). By trial and error we

found that $n = 5$ gave good results with (2) and it is the value used in all the work reported here. It will be seen that the use of (2) gave a useful reduction in the number of cycles required and also a small reduction in the final error. It is found that the LDE method gives phase errors negatively correlated with both the values of $|E|$ and also the magnitudes of the Fourier coefficients, $|\mathcal{F}|$ of the maps which are produced. For that reason we also show the m.p.e. with $|E.\mathcal{F}|$ weighting which is usually significantly lower in value than the unweighted one. Encouraged by this result for aPP we turned our attention to the application of LDE to larger structures.

A test with RNApl

The structure of RNApl (Bezborodova, Ermekbaeva, Shlyapnikov, Polyakov & Bezborodov, 1988) has space group $P2_1$ 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 go out to high resolution by protein standards – 1.17 Å – and there are 23 853 independent observed reflections. Using LDE modification (1) a phase refinement was begun with an imposed m.p.e. of 70° for the complete data set. After 293 cycles of refinement the unweighted m.p.e. had reduced to 50.2° while that weighted with $|E.\mathcal{F}|$ was only 34.6°. The process was then repeated with modification formula (2) but this time the initial m.p.e. was 74°, greater than that used previously. There was a significant improvement in the result. After 243 cycles the unweighted m.p.e. had reduced to 39.3° and that weighted with $|E.\mathcal{F}|$ to 25.6°. This gave an excellent final map with a conventional map-correlation coefficient (MCC) of 0.697 compared with 0.222 for the original phases. These results are collected together in Table 2.

Having made successful tests with the modified LDE process on two proteins, both of which had data of high resolution by the standards of protein crystallography, we decided to see how well it could perform in a case less favourable as far as resolution is concerned. It is well known that direct methods are extremely dependent for their effectiveness on the resolution of the available data and it seemed possible that this would also be true for LDE.

Tests on pig 2Zn insulin

The structure pig 2Zn insulin (Baker, Blundell, Cutfield, Cutfield, Dodson, Dodson, Hodgkin, Hubbard, Isaacs, Reynolds, Sakabe, Sakabe & Vijayan, 1988) crystallizes in the rhombohedral space group $R3$ with $a = 49.0$ Å and $\alpha_R = 114.8°$. The whole cell contains six molecules of pig insulin each of composition

Table 1. *Comparison of phase extension from* 3 *to* 1 Å *for aPP*

Method (*a*) has a sharp cut-off at $0.2\rho_r$ and (*b*) has a smoothing-function density modification. The starting point in each case is 614 reflections within 3 Å resolution with a mean-phase error of 30°.

Method	No. of cycles required	Final No. of reflections	M.p.e. $\langle\Delta\varphi\rangle$(°)	Weighted m.p.e. $\lvert E.\mathcal{F}\rvert$(°)
(*a*)	77	16 524	27.6	15.5
(*b*)	58	16 524	27.3	15.3

Table 2. *Refinement of the complete data set of* 23 853 *reflections at* 1.17 Å *resolution for RNApl*

Phase-modification methods (*a*) and (*b*) are as given in Table 1. Phase errors are given both without weights and with $|E.\mathcal{F}|$ weighting. The figures in parentheses are the initial map-correlation coefficients (MCC) for case (*b*).

Method	No. of cycles required	Initial m.p.e. (°)	Unweighted final m.p.e. (°)	Weighted final m.p.e. (°)
(*a*)	293	70.0	50.2	34.6
(*b*)	243	74.0 (0.222)	39.3 (0.697)	25.6

$C_{256}N_{130}O_{152}S_{12}H_{776}$, *i.e.* two per asymmetric unit, and two Zn atoms which are therefore constrained to sit on threefold axes. There are also 283 H_2O molecules in each asymmetric unit. Native data were obtained to 1.5 Å resolution and, from a large number of derivatives, the multiple isomorphous replacement (MIR) technique gave phase estimates to 1.9 Å.

This structure has been the target of trials by many phase-extension and refinement methods (*e.g.* de Rango, Mauguen & Tsoucaris, 1975) and therefore seemed suitable as a test to explore what might be expected from the LDE method.

Refinement using 1.5 Å *data*

In some initial tests we started with a moderately high imposed m.p.e. (~60°) for the complete data set for the native protein and then refined with density-modification formula (2). Since the phases are known for the refined structure we were able to follow the refinement process in some detail. One aspect that we kept under investigation was the retention of enantiomorph discrimination since with heavy atoms present, particularly when situated on special positions, we knew from our direct-methods experience that phases could be found corresponding to the sum of enantiomorph structures. The way that we achieved this was simply to calculate the m.p.e. after each cycle by comparison with phases for both the target structure and its enantiomorph; if enantiomorph discrimination is lost then the two values of m.p.e. will be similar.

We found that for the first few cycles of refinement there was a reduction in the m.p.e. and an improvement in the MCC while enantiomorph discrimination was maintained. However, the phases then began to deteriorate, together with the MCC, and

Table 3. *Results of refinement experiments with the* 1.5 Å *data for* 2Zn *insulin*

The various methods used are: (*a*) Smoothing function only. (*b*) Change of ρ_c cycle by cycle. (*c*) Truncation of high density. (*d*) Change of ρ_c cycle by cycle plus termination by the criterion LCEF. The initial and final mean-phase errors are given for the target structure and, in parentheses, for its enantiomorph. For method (*d*) the initial and final value of the LCEF is given in parentheses below the m.p.e. values.

Method	No. of cycles required	Initial m.p.e. (°)	Unweighted final m.p.e. (°)	Weighted final m.p.e. (°)	MCC
(*a*)	92	61.2 (81.1)	59.0 (60.7)	56.9	0.403
(*a*)	127	78.8 (86.9)	60.5 (61.5)	61.6	0.371
(*b*)	95	61.2 (81.1)	56.6 (76.2)	46.4	0.533
(*b*)	133	67.6 (83.1)	60.4 (77.7)	50.4	0.485
(*c*)	197	61.2 (81.1)	77.4 (79.6)	75.6	0.169
(*c*)	279	78.8 (86.9)	82.4 (82.7)	81.6	0.105
(*d*)	10	61.2 (81.1)	47.3 (74.9)	34.8	0.682
(*d*)	15	67.6 (83.1) (0.040)	54.0 (75.1) (0.261)	42.1	0.599

enantiomorph discrimination became poor. Nevertheless, we ran the LDE process to completion in the normal way, which is when the mean-phase change in a cycle falls below some predetermined value, which was 1° in this case, and the result of doing so is shown in the top row of Table 3. The initial phases had an m.p.e. of 61.2 and 81.1° for the target and enantiomorph structure, respectively. The MCC for these phases was 0.402. After 92 refinement cycles the values of m.p.e. for the target structure and its enantiomorph were 59.0 and 60.7°, respectively, with an MCC of 0.410. In addition the $|E.\mathcal{F}|$-weighted m.p.e. was not not much less than the unweighted value. While the m.p.e. had marginally been reduced, the loss of enantiomorph discrimination would almost certainly have made the final phases less useful than the initial ones. The second row of the table shows a similar effect when starting with a higher m.p.e.

It seemed that the damage to the process might be as a result of the Zn atoms on special positions. Although ρ_c was no longer a sharp cut-off level with modification (2), nevertheless the result of the density modification was to increase greatly the relative contribution of the heavy atom in the modified map as compared to the original map. This effect became greater as the refinement proceeded since the proportion of the volume of the cell above the level ρ_c tended to decline. It was decided to try to correct this effect by changing ρ_c cycle-by-cycle in such a way that a fixed fraction of the volume of the cell corresponded to density higher than ρ_c and we selected, after some trials, what seemed a suitable fraction of 0.015. The result of doing this is seen in the rows marked (*b*) of Table 3. After 95 cycles the m.p.e.'s for the target structure and its enantiomorph were 56.6 and 76.2° respectively and the $|E.\mathcal{F}|$-weighted m.p.e. was 46.4°. The map-correlation coefficient was now 0.533 and it is clear that there was a great improvement from using the revised ρ_c-selection procedure.

Finally we tried reducing further the influence of the Zn atoms by truncating the density such that if $\rho(\mathbf{r}) > 2\rho_c$ then $\rho(\mathbf{r}) = 2\rho_c$, but the effect of this, as given in rows (*c*) of Table 3, can be seen to have been quite catastrophic.

Our diagnostic interrogation of the refinement process, step by step, showed that we were going much too far in the refinement and that after a few cycles of improvement the remainder of the process, corresponding to the bulk of the work done, was actually making matters worse. At this point a different criterion for stopping the refinement was suggested to us by Dr. V. Y. Lunin and this we now describe.

The modified stopping criterion

This criterion is of a kind that is used by protein crystallographers to test the self-consistency of their current structural model when, for example, they calculate structure factors excluding the density in one small part of the unit cell and then see how well that part is restored when a new map is calculated. Here what is done is to exclude some small percentage of the reflections from the refinement process, in the range 1–10% of them depending on the total number being used, and then at each stage to compare the magnitude of the Fourier coefficient of the modified map with the normalized structure magnitude. This comparison is performed by means of a linear-correlation coefficient (LCEF) so that scaling problems do not occur. Tests showed that there was usually a good correlation between the LCEF and the m.p.e. Sometimes the behaviour of the LCEF is irregular at the beginning of the refinement process, *i.e.* it might fluctuate even though the m.p.e. is reducing, so for that reason we took as our criterion for stopping that the LCEF should decrease for four successive cycles. We then took the phases from four steps backwards where the LCEF was at its maximum.

The result of doing this for refining the 1.5 Å phases of 2Zn insulin is shown in rows (*d*) of Table 3. There is a considerable improvement in the final outcome in terms of the final m.p.e. and the MCC and a significant economy in the computing effort.

The original MIR phases for the 1.9 Å resolution reflections had an m.p.e. of 61°. We now did some experiments to see how well the LDE method would work at this resolution starting with the experimental MIR phases. Our initial results were extremely disappointing; the m.p.e. actually increased right from the beginning. In the LDE method at each stage of refinement a new map is calculated with Fourier coefficients weighted by a factor

$$W = \tanh(|E.\mathcal{F}|/2).$$

We decided to experiment with the weighting scheme and we tried the following:

(i) $W = \tanh(|E.F|/2)$, the original scheme used in paper I.

(ii) No weight, $W = 1$, *i.e.* the magnitude of the Fourier coefficient was $|E|$.

(iii) W = minimum of $\{1, |.F|/|E|\}$, which means that the magnitude of the Fourier coefficient is the lesser of $|E|$ and $|.F|$.

(iv) W = minimum of $\{1, |F/E|^{1/2}\}$, which means that the magnitude of the Fourier coefficient is the lesser of $|E|$ and $|E.F|^{1/2}$.

We noticed that initially, for all the weighting schemes, the LCEF increases whether the m.p.e. decreases or declines then, for a few cycles, the LCEF continues to increase while the m.p.e. slightly increases and then finally the LCEF decreases as the phases get worse. It was at this stage that we found that restricting the number of refinement cycles to somewhere between five and ten was just as effective as relying on the value of the LCEF and in much of the following work that is what we did. In Table 4 the results of refinement are shown after five and ten cycles for all four weighting schemes. Scheme (iii) seems to be slightly better than (ii) and (iv) while the original scheme (i) seems worst of all. However, it is important to notice that while the original m.p.e. of 61.0° had no correlation with $|E|$ after refinement there is a strong correlation with $|E.F|$, for all the schemes, and this means that the information content of the phases will be higher – even for scheme (i) where the unweighted m.p.e. actually increases.

We continued to experiment with other modifications of the refinement procedure. We tried one scheme where the newly accepted phase was not that of the new map but a weighted tangent-formula combination of that from the new map and that from the old. One such combination was of the form

$$\varphi_{\text{next}} = \text{phase of } [W_{\text{new}} \exp(i\varphi_{\text{new}}) + W_{\text{old}} \exp(i\varphi_{\text{old}})], \quad (3)$$

where $W_{\text{new}} = |E.F^2_{\text{new}}|^{2/3}$, W_{old} is similarly defined, and old and new refer to the previous and new maps respectively. This gave results similar to, and marginally better, than those in Table 4 but it was not considered worthwhile to incorporate this process routinely.

Another scheme we tried was to use simulated F's instead of E's throughout the refinement process. This was achieved by

$$F = E \exp(-\alpha s^2) \quad (4)$$

where s is the usual scattering vector of magnitude $(2\sin\theta)/\lambda$. Values of α in the range 5–10 were explored and the results of doing this are shown in Table 5 where option (iii) is used for the weighting of the map Fourier coefficients. Two values of α, 6 and 9, are used; the only features of interest are the low $|E.F|$-weighted m.p.e. values but this just indicates that lower phase errors are found for lower resolution data where the average values of $|F|$ are higher.

Our conclusion is that, on the basis of the experiments we have performed, the best scheme is to use option (iii) to weight map coefficients and, for low-resolution data, to refine for no more than five to ten cycles.

Table 4. *Refinement of the* 1.9 Å *data set for* 2Zn *insulin using four different weighting schemes for the Fourier coefficients of the calculated maps*

The weighting schemes, given as (i), (ii), (iii) and (iv) are described in the text. The initial mean-phase error is 61.0° in each case. The m.p.e. for the enantiomorph structure is shown in parentheses.

Weighting scheme	No. of cycles	Unweighted final m.p.e. (°)	Weighted final m.p.e. (°)
(i)	5	61.8 (86.7)	45.6
	10	63.5 (87.2)	46.5
(ii)	5	58.3 (86.3)	46.0
	10	59.6 (86.0)	47.2
(iii)	5	56.8 (86.3)	42.6
	10	58.5 (86.4)	44.8
(iv)	5	57.4 (86.4)	44.6
	10	58.6 (86.6)	44.9

Table 5. *Refinement of the* 1.9 Å *data set for* 2Zn *insulin using simulated F's* [$E\exp(-\alpha s^2)$] *instead of E's*

Weighting scheme (iii) is used for the Fourier coefficients. The initial error is 61.0° in each case. The mean-phase error for the enantiomorph structure is shown in parentheses.

α	No. of cycles	Unweighted final m.p.e. (°)	Weighted final m.p.e. (°)
6	5	59.9 (87.2)	37.2
6	10	60.9 (88.0)	38.1
9	5	65.4 (87.8)	38.3
9	10	68.3 (88.7)	40.3

Table 6. *Result of phase refinement and extension starting with* 1.9 Å *data for* 2Zn *insulin and extending to* 1.5 Å *data*

The initial mean-phase error for the 1.9 Å data is 61.0° and it is refined for five cycles. Extension is carried out in shells as described in the text. Weighting of map Fourier coefficients is done according to scheme (iii). MCC is the final map-correlation coefficient.

Resolution (Å)	No. of cycles	Unweighted m.p.e. (°)	Weighted m.p.e. (°)	MCC
1.9	0	61.0		—
1.9	5	56.8 (86.5)	42.4	—
1.9	69	54.8	41.3	
1.5	69	57.3 (82.8)	43.9	0.540

Phase extenstion for 2Zn insulin

We refined the 1.9 Å MIR-phased set of reflections for five cycles as indicated above and then gradually introduced the higher resolution data. This we did in shells: 1.9–1.8, 1.8–1.7, 1.7–1.6, 1.6–1.55 and finally

1.55–1.5 Å. Each new shell was refined for ten cycles, keeping the phases as well as the weighted Fourier coefficients the same for all the previous shells. After the final shell had been so treated then the whole system was relaxed and everything was refined to completion which, in this case, was taken as when the phases changed by an average of less than 3.5°.

The results, shown in Table 6, indicate that the phase extension had been quite effective with a map at 1.5 Å which could probably be interpreted in terms of a structural model.

Concluding remarks

We believe that we have demonstrated that the LDE process, using the best parameters and weighting schemes, is an economical and effective procedure for phase extension and refinement, at least for moderate-sized proteins. However, it is unlikely that it could be as effective as the *SQUASH* procedure described by Zhang & Main (1990*a*,*b*), which brings much more information to bear on the problem. On the other hand its requirements on computer power and time are much more modest and it may therefore find a limited use within the community of protein crystallographers.

We are grateful to the Science and Engineering Research Council who have supported this project. We also wish to thank Dr M. Shiono, Dr C. Tate and Dr V. Y. Lunin for their most valuable advice and contributions.

References

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**, 469–456.

BEZBORODOVA, S. I., ERMEKBAEVA, L. A., SHLYAPNIKOV, S. V., POLYAKOV, K. M. & BEZBORODOV, A. M. (1988). *Biokhimiya*, **53**, 965–973.

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

RANGO, C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* A**31**, 227–233.

SHIONO, M. & WOOLFSON, M. M. (1992). *Acta Cryst.* A**48**, 451–456.

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.