Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Ab initio structure determinations by direct-space methods: tests of low-density elimination

Naohiro Matsugaki† and Masaaki Shiono*

Faculty of Science, Kyushu University, 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan

† Present address: Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita-shi, Osaka 565-0871, Japan.

Correspondence e-mail: shio8scp@mbox.nc.kyushu-u.ac.jp

Received 4 July 2000
Accepted 10 October 2000

The low-density elimination method, which was developed for phase extension and refinement, has been investigated regarding its power to solve crystal structures starting from completely random phase sets. The method employs a multi-solution strategy. Low-symmetry structures are easily solvable where phase restrictions are only applied to a few reflections. Even with high-symmetry structures, a reasonable solution was obtained regarding centric reflections as general reflections. It is also shown that the structure of a small protein ribonuclease Ap1 is solvable if the positions of the five S atoms in the protein are known.

1. Introduction

The low-density elimination (LDE) method has been described in two successive papers [Shiono & Woolfson (1992) and Refaat & Woolfson (1993), referred to as paper I and paper II, respectively] and is used for removing negative peaks and sharpening peaks in the E map. In previous publications, we concentrated on phase extension and refinement for macromolecules and concluded that the LDE method is effective when high-resolution data are available. In this paper, we describe the effectiveness of the method for obtaining *ab initio* solutions.

2. Methods

We made slight modifications in our procedure as follows.

(i) In paper I, we used a simple modification function

$$\begin{cases} \rho'(\mathbf{r}) = \rho(\mathbf{r}) & \rho(\mathbf{r}) > 0.2\rho_c \\ \rho'(\mathbf{r}) = 0 & \rho(\mathbf{r}) \leq 0.2\rho_c, \end{cases} \tag{1}$$

where ρ_c is the expected average peak height of light atoms in the structure. This modification function was effective but rather slow in convergence because of the discontinuity in the modified density. To remove the discontinuity, we changed to a different function in paper II,

$$\begin{cases} \rho'(\mathbf{r}) = \rho(\mathbf{r})^{n+1}/[(0.2\rho_c)^n + \rho(\mathbf{r})^n] & \rho(\mathbf{r}) > 0 \\ \rho'(\mathbf{r}) = 0 & \rho(\mathbf{r}) \leq 0, \end{cases} \tag{2}$$

where n is an integer greater than unity. The best value for n was found to be 5. We further sought a better function and in the present work density is transformed by

$$\begin{cases} \rho'(\mathbf{r}) = \rho(\mathbf{r})\{1 - \exp[-\frac{1}{2}[\rho(\mathbf{r})/0.2\rho_c]^2]\} & \rho(\mathbf{r}) \geq 0 \\ \rho'(\mathbf{r}) = 0 & \rho(\mathbf{r}) < 0. \end{cases} \tag{3}$$

This function also removes the discontinuity in the modified map. Compared with function (2) with $n = 5$, the merit of function (3) is that it does not dramatically reduce peaks with

Table 1
Results of *ab initio* structure determinations for (*a*) MUCCAR, (*b*) CINOBUFAGIN, (*c*) AZET, (*d*) ALPHA-1, (*e*) aPP, (*f*) crambin, (*g*) rubredoxin and (*h*) cytochrome c_6.

MPE is the mean phase error for all reflections. WMPE is the mean phase error weighted with $|E\mathcal{F}|$ as defined in paper I (Shiono & Woolfson, 1992). LCFOM is defined in the text. The sets marked by an asterisk are the solutions. Observed structure factors of ALPHA-1, rubredoxin and cytochrome c_6 were obtained from the Protein Data Bank (Abola *et al.*, 1987; Bernstein *et al.*, 1977).

(*a*) MUCCAR.

Set number	Cycles	MPE	WMPE	LCFOM
1*	110	32.58	16.11	0.4304
2*	70	32.83	15.65	0.4211
3*	70	34.50	18.47	0.4293
4*	82	34.71	18.35	0.4016
5*	89	33.19	16.85	0.4265
6*	63	34.07	17.82	0.4241
7*	77	33.17	16.86	0.4249
8*	129	34.42	18.24	0.4001
9*	49	32.44	16.20	0.4223
10*	57	33.74	18.62	0.4327

(*b*) CINOBUFAGIN.

Set number	Cycles	MPE	WMPE	LCFOM
1*	245	26.03	9.71	0.4321
2	343	84.13	80.75	0.0051
3*	263	26.23	9.73	0.4321
4	266	86.86	87.93	0.0135
5*	322	26.27	9.79	0.4312
6	457	87.74	87.67	0.0234
7*	183	26.30	9.73	0.4333
8*	524	26.34	9.75	0.4317
9*	397	26.34	9.76	0.4317
10*	300	26.34	9.75	0.4320

(*c*) AZET.

Set number	Cycles	MPE	WMPE	LCFOM
1*	282	29.70	12.59	0.2604
2*	116	29.54	12.26	0.2549
3*	460	29.23	11.81	0.2592
4*	279	29.46	12.19	0.2624
5*	124	29.69	12.35	0.2573
6*	274	29.22	11.92	0.2618
7*	661	29.39	12.07	0.2622
8*	356	29.33	11.94	0.2593
9*	482	29.74	12.58	0.2586
10*	258	29.92	12.72	0.2585
19	699	79.30	67.48	0.0416
66	604	74.30	57.05	0.0692
75	1191	84.43	82.09	0.0198

(*d*) ALPHA-1.

Set number	Cycles	MPE	WMPE	LCFOM
1	360	85.73	80.19	0.0329
2	308	87.19	82.77	0.0172
3	330	84.83	77.87	0.0245
4	371	86.47	81.50	0.0325
5	451	86.81	81.96	0.0323
6	534	86.82	82.57	0.0366
7	293	86.83	81.67	0.0201
8*	510	33.41	22.68	0.3633
9*	321	36.19	26.39	0.3634
10	417	83.77	76.27	0.0302

(*e*) aPP.

Set number	Cycles	MPE	WMPE	LCFOM
1*	53	34.03	21.73	0.2106
2*	154	32.06	19.10	0.2055
3*	145	34.61	22.25	0.2002
4*	120	30.26	17.06	0.2024
5*	273	32.63	19.95	0.2001
6*	163	33.60	20.82	0.2041
7*	87	33.94	21.33	0.2066
8*	142	35.09	22.87	0.2014
9*	87	34.82	22.20	0.2024
10*	181	31.50	18.65	0.2003

(*f*) Crambin.

Set number	Cycles	MPE	WMPE	LCFOM
1	941	89.25	89.66	0.0301
2	925	88.45	86.18	0.0181
3	881	88.85	88.59	0.0191
4	1031	88.81	86.93	0.0143
5	892	88.61	88.00	0.0198
13*	526	26.38	13.41	0.4748
14*	527	26.50	13.66	0.4740
17*	222	26.44	13.67	0.4758
18*	287	26.78	14.27	0.4773
19*	461	26.45	13.78	0.4784
30*	323	26.71	13.96	0.4738
73*	472	26.21	13.39	0.4748
90*	1161	26.67	13.81	0.4745

(*g*) Rubredoxin.

Set number	Cycles	MPE	WMPE	LCFOM
1	908	89.69	87.96	0.0895
2*	578	28.40	11.27	0.3817
3	1276	89.20	87.36	0.0770
4	1127	88.94	86.87	0.0183
5*	1259	28.43	11.32	0.3811
6	1198	89.54	89.47	0.0109
7	1713	89.49	89.39	0.0178
8	1276	89.66	88.26	0.0131
9	1422	89.31	89.81	0.0150
10	1011	88.61	86.22	0.0139

(*h*) Cytochrome c_6.

Set number	Cycles	MPE	WMPE	LCFOM
1*	238	49.22	32.84	0.1729
2*	259	46.18	24.73	0.1795
3*	320	48.59	32.48	0.1802
4*	287	56.14	46.06	0.1817
5*	488	53.90	42.10	0.1783
6*	261	54.10	43.59	0.1766
7*	445	54.72	41.90	0.1841
8*	517	54.46	41.50	0.1796
9*	315	51.07	40.71	0.1888
10*	380	56.09	44.88	0.1780

heights just below $0.2\rho_c$ and gives the opportunity of developing the small peaks. For direct determination of structures it is very important to allow unforced generation and extinction of peaks. Function (3) is very satisfactory for this purpose.

(ii) Various trials have suggested that all reflections are not necessary for *E*-map calculation. We calculated the *E* map with coefficients of normalized structure factors ($|E|$s) greater than unity and monitored the refinement by the linear

correlation coefficients between $|E|$s and Fourier coefficients of the modified map ($|\mathscr{F}|$s) as a figure of merit (LCFOM) defined by

$$\mathrm{LCFOM} = \frac{\left\langle \left(|E| - \langle|E|\rangle\right)\left(|\mathscr{F}| - \langle|\mathscr{F}|\rangle\right)\right\rangle}{\left[\left\langle \left(|E| - \langle|E|\rangle\right)^2\right\rangle \left\langle \left(|\mathscr{F}| - \langle|\mathscr{F}|\rangle\right)^2\right\rangle\right]^{1/2}}, \quad (4)$$

where averages are taken over all reflections which are not used for density modification.

(iii) In the course of refinement, centric reflections are treated as general reflections. All symmetry-related reflections are assigned independent random phases without phase restriction. This procedure is equivalent to treating the space group as $P1$. By doing so, a very high validity is achieved despite the increase in the number of independent atom positions to be determined. Since we start the refinement treating the structure as $P1$, the resultant maps have arbitrary crystallographic origins. It is thus necessary to fix the origin before introducing phase restrictions. The origin-shift vector $\Delta\mathbf{r}$ is evaluated by using centric reflections as follows. The phases of centric reflections are restricted to two values which differ by π. However, if the origin is not properly defined, the phase values become

$$\varphi(\mathbf{h}) = 2\pi\mathbf{h}\cdot\Delta\mathbf{r} + \varphi_c(\mathbf{h}) + (0 \text{ or } \pi), \quad (5)$$

where $\Delta\mathbf{r}$ is the position vector of the proper origin and $\varphi_c(\mathbf{h})$ is one of the two restricted phase values for a particular $\mathbf{h}$. We now subtract $\varphi_c(\mathbf{h})$ from $\varphi(\mathbf{h})$ and multiply the result by 2. Taking the 2π degeneracy of phase value into account, we find

$$2[\varphi(\mathbf{h}) - \varphi_c(\mathbf{h})] = 2\pi\mathbf{h}\cdot 2\Delta\mathbf{r}. \quad (6)$$

It is easily understood by considering (6) that if we perform Fourier transformation with the coefficients of unit magnitude and their phases, $2[\varphi(\mathbf{h}) - \varphi_c(\mathbf{h})]$, then we obtain a map which has a peak at the position $2\Delta\mathbf{r}$. We can therefore find the origin-shift vector $\Delta\mathbf{r}$ by halving the peak position. This procedure uniquely determines the proper origin for primitive lattices. If the lattice is non-primitive, for example face centred or body centred, then we may find more than one peak in the map caused by extinction rules. In such a case, there are ambiguities in determining the origin. In our latest program, we distinguish the correct origin by comparing LCFOMs after performing a cycle of refinement with phase restrictions for each origin. It must be noted that this procedure for origin fixing cannot apply to trigonal, hexagonal and rhombohedral crystal systems, where there is no centric reflection. For those space groups, phase differences between equivalents for general reflections are used for the origin determination in some effective manner, although the result is usually less accurate. Once the origin has been successfully fixed, phase restrictions are applied and the structure is refined with the appropriate space group for a few cycles. All phases of reflections generated by symmetry operations are abandoned and replaced with phases calculated from those of original reflections and the phases of centric reflections are set to their allowed values. If the origin shift fails, the phases are stored as $P1$.

3. Tests with small structures

All refinements in this section start from completely random phases. A multi-solution strategy was employed for the LDE method and we performed 100 trials in each test for the purpose of estimating its validity. In the first three cycles, density was squared after all negative density was set to zero in order to accelerate the refinement. Each trial is terminated when the average phase change becomes less than 0.5°. We chose four organic molecules and four small proteins for the tests. The data for MUCCAR, CINOBUFAGIN, AZET and aPP were supplied by Professor M. M. Woolfson, and the data for crambin were supplied by Dr A. Yamano of Rigaku Corporation. The other data were downloaded from the Protein Data Bank (Abola *et al.*, 1987; Bernstein *et al.*, 1977). The organic compounds are referred to by their code names for simplicity. The quality of the resultant phase sets are shown in Table 1 in terms of phase errors, which are calculated after properly determining the origin and enantiomorph.

(i) MUCCAR (Bianchi *et al.*, 1978; $C_{13}H_{11}N$; $P1$; a = 8.310, b = 7.026, c = 9.508 Å, α = 100.89, β = 97.82, γ = 113.48°, Z = 2). There are 1940 observed reflections; 681 $|E|$s greater than unity were used for refinement. All 100 resultant maps showed the complete structure. Final phase errors and LCFOMs for the first ten sets are given in Table 1(*a*).

(ii) CINOBUFAGIN (Declercq *et al.*, 1977; $C_{26}H_{27}O_6$; $P2_12_12_1$; a = 7.663, b = 15.900, c = 19.291 Å, Z = 4). There are 2231 independent reflections; 803 $|E|$s were used for refinement. We first tried to solve this structure with phase restrictions; that is, to treat the structure as space group $P2_12_12_1$; only four complete solutions were found within 100 trials. Next, we removed the phase restrictions and found 90 complete solutions. Final phase errors and LCFOMs for the first ten sets are given in Table 1(*b*). There were three failures with low LCFOM values in the first ten sets.

(iii) AZET (Colens *et al.*, 1974; $C_{21}H_{16}ClNO$; $Pca2_1$; a = 36.042, b = 8.730, c = 11.084 Å, Z = 8). There are 1910 independent reflections; 664 $|E|$s were used for refinement. With phase restrictions, we found 37 complete solutions within 100 trials. This result is tolerable and since the first set eventually led to a complete solution, we actually solved the structure within 1 min. Next, we removed the phase restrictions and found 97 complete solutions. The result for the first ten sets and all three failures are given in Table 1(*c*). Here we can again see that the correct solutions clearly indicate high values of LCFOM.

(iv) ALPHA-1 (Patterson *et al.*, 1999; $P1$; a = 20.846, b = 20.909, c = 27.057 Å, α = 102.40, β = 95.33, γ = 119.62°, Z = 4). This is a designed peptide with 12 amino-acid residues. The unit cell contains 408 non-H atoms belonging to the peptides, 30 water molecules and 41 other non-H atoms including a chloride ion. There are no S atoms in the peptide. The resolution of the data is 0.90 Å, with 21 831 independent reflections. We found 19 solutions out of 100 trials. The results for the first ten sets are given in Table 1(*e*).

(v) Avian pancreatic polypeptide (aPP; Glover *et al.*, 1983; $C2$; a = 34.18, b = 32.92, c = 28.45 Å, β = 105.26°, Z = 4). This is

a small protein with 36 amino-acid residues. The asymmetric unit contains 301 non-H atoms belonging to the protein, a Zn atom and 80 water molecules. We used 6425 $|E|$s for the refinement out of 17 454 independent reflections to 0.98 Å resolution. With phase restrictions, we found 11 solutions. Without phase restrictions, all 100 trials reached the correct solution. The results for the first ten sets are given in Table 1(*d*). The reasonable values of the phase errors suggest that all maps should reveal most of the structure.

(vi) Crambin (Hendrickson & Teeter, 1981; $P2_1$; a = 40.76, b = 18.49, c = 22.33 Å, β = 90.61°, Z = 2). This is also a small protein, with 46 amino-acid residues. The PDB file records 393 non-H atoms belonging to the protein molecule. The resolution of the data is 0.89 Å, with 25 951 independent reflections. The program failed to solve this structure with phase restrictions. Without phase restrictions, eight solutions were found. The results for the first five sets and all eight solutions are given in Table 1(*f*).

(vii) Rubredoxin (Bau *et al.*, 1998; $P2_12_12_1$; a = 34.123, b = 34.874, c = 43.683 Å, Z = 4). This protein consists of 53 amino-acid residues. The asymmetric unit contains 413 non-H atoms belonging to the protein molecule, 137 water molecules and an iron ion. The resolution of the data is 0.95 Å, with 32 303 independent reflections. The trials were only performed without phase restrictions and ten solutions were found. The results for the first ten sets are given in Table 1(*g*).

(viii) Cytochrome c_6 (Frazao *et al.*, 1995; $R3$; a = 40.430 Å, Z = 3). This is a haem-containing protein with 89 amino-acid residues. There are 724 non-H atoms belonging to the protein molecule and 151 waters and 43 haem atoms including an iron ion. The resolution of the data is 1.10 Å, with 32 653 independent reflections. The trials were only performed without phase restrictions; the final phase errors for all 100 trials showed reasonably low values. The results for the first ten sets are given in Table 1(*h*).

The above results are very satisfactory and show that a structure containing about 800 atoms is tractable using the LDE method. However, we are not saying that solving structures as large as cytochrome c_6 is an easy matter. Solving cytochrome c_6 by LDE is easy despite its complex structure owing to the influence of the heavy atom(s). Presumably, the computation process is to find the Fe peaks first and after several cycles of refinement to gradually determine other atom positions. In the case of crambin and ALPHA-1, the S atoms and the chloride ion, respectively, play a role as heavy atoms as described in the next section.

4. Phase refinement initiated by partial structure

As discussed above, the LDE method is very useful when the target structure contains heavy atoms. Now the question arises: how heavy should they be? Since many proteins contain S atoms, we tested the effectiveness of S atoms for solving the small proteins ribonuclease Ap1 and crambin. While carrying out this test, for the first 15 cycles we used $w|E|$ as the Fourier coefficient for map calculation as well as that used in paper I, where

$$w = \tanh(K|E\mathscr{F}|/2). \tag{7}$$

K is the scale factor for $|\mathscr{F}|$. After cycle 16 we replaced the Fourier coefficient by

$$\begin{cases} w(2|E| - |\mathscr{F}|) & 2|E| - |\mathscr{F}| > 0 \\ 0 & 2|E| - |\mathscr{F}| < 0. \end{cases} \tag{8}$$

Physical interpretation of this coefficient is the combination of normal weighted Fourier and the difference Fourier which is commonly used in protein crystallography. Our view on the effect of using this coefficient is that this coefficient allows the peak heights to increase quickly and decreases the ripple peaks around large peaks; it is expected to be effective in refining the phases calculated from the partially determined structures. Phase restrictions are applied from the first cycle.

The first test structure is ribonuclease Ap1 (RNAp1; Bezborodova *et al.*, 1988), which is a protein crystallized in space group $P2_1$, with unit-cell parameters a = 32.01, b = 49.76, c = 30.67 Å, β = 115.83°, Z = 2. The protein molecule contains 808 non-H atoms including five S atoms; there are 83 ordered water molecules in the asymmetric unit. The observed data have a resolution of 1.17 Å with 23 853 independent reflections. The data were supplied by Professor M. M. Woolfson.

We tried to refine a phase set calculated from all five sulfur coordinates. We used 8623 $|E|$s greater than unity for the map calculation. B factors of 0.0 Å^2 are used in the calculations of the initial phases. The initial mean phase error was 74.76° for all reflections. After 468 cycles of refinement, the phase error decreased to 51.98° and the mean phase error weighted with $|E\mathscr{F}|$ fell to a small value of 35.55°. The value of final LCFOM was −0.0664. Since the value of LCFOM was large and negative at the beginning of the refinement (*e.g.* −0.1847 at cycle 20), in spite of its negative value it was still useful in judging if the refinement is successful. For comparison, we also tried refinement using the coefficient $w|E|$ throughout, but no significant improvement was obtained. Results for both procedures are shown in Table 2(*a*).

Next, we tried to refine the phase sets calculated with four sulfur coordinates. There are five combinations in the choice of four sulfurs and we tried all five cases. The results are shown in Table 2(*b*), where only one trial, with the combination S1, S2, S3 and S5, was successful. We then tried to refine phases with three sulfur combinations out of four sulfurs. The results are shown in Table 2(*c*), where we found one successful refinement with S1, S2 and S5 among the four possible combinations. It must be mentioned that we also tried with two sulfur contributions but none of the ten possible combinations refined successfully. Here, we used a stopping criterion very similar to that described in paper II. When the mean phase change becomes less than 0.5°, we calculated LCFOM in each cycle. The values of LCFOM for the three most recent cycles are stored and when the current LCFOM becomes the smallest of these then the refinement is stopped. Usually the value of LCFOM smoothly increases cycle by cycle if the refinement is successful and this is the reason why successful trials need more cycles of computation than unsuccessful ones. On seeing the results, it seems that the contributions of some

Table 2
Results of the refinement for RNAp1 from sulfur contributions Refinements from (*a*) five sulfur contributions, (*b*) four sulfur contributions, (*c*) three sulfur contributions, (*d*) four sulfur contributions and (*e*) one sulfur contribution.

Calculated structure factors were used for refinements (*d*) and (*e*). *B* factors (Å^2) of the five S atoms are 7.48 (S1), 8.87 (S2), 19.17 (S3), 22.64 (S4) and 26.29 (S5).

(*a*) Refinements from five sulfur contributions. Results of the refinement using $w|E|$ throughout (scheme 1) and introducing $w(2|E| - |\mathcal{F}|)$ after cycle 16 (scheme 2) are both tabulated.

	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
Scheme 1	204	74.76	75.84	65.60	−0.1932
Scheme 2	468	74.76	51.98	35.55	−0.0664

(*b*) Refinements from four sulfur contributions.

Combination	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
S1 + S2 + S3 + S4	174	75.98	80.09	72.10	−0.1809
S1 + S2 + S3 + S5	422	76.17	51.71	35.27	−0.0634
S1 + S2 + S4 + S5	207	76.33	77.80	69.56	−0.1945
S1 + S3 + S4 + S5	224	76.98	84.12	78.90	−0.1674
S2 + S3 + S4 + S5	163	77.03	81.90	76.99	−0.1970

(*c*) Refinements from three sulfur contributions.

Combination	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
S1 + S2 + S3	140	77.44	77.37	67.92	−0.1907
S1 + S2 + S5	538	77.38	51.71	35.22	−0.0626
S1 + S3 + S5	191	78.38	81.34	74.22	−0.1771
S2 + S3 + S5	242	78.33	83.68	78.44	−0.1773

(*d*) Refinements from four sulfur contributions using calculated structure factors.

Combination	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
S1 + S2 + S3 + S4	86	75.98	47.71	29.97	0.2375
S1 + S2 + S3 + S5	78	76.17	47.79	30.14	0.2379
S1 + S2 + S4 + S5	90	76.33	47.76	30.04	0.2375
S1 + S3 + S4 + S5	280	76.98	81.20	74.73	−0.0035

(*e*) Refinements from one sulfur contribution using calculated structure factors.

Sulfur	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
S1	261	82.43	48.49	30.59	0.2238
S2	195	82.64	84.39	79.19	0.0087
S3	165	83.36	85.88	84.16	−0.0029
S4	186	83.83	86.79	85.55	0.0069
S5	158	83.82	86.78	86.39	0.0001

sulfurs are more important than those of others in the refinements. Interestingly, we succeeded in refining phases from S1 + S2 + S5 contributions, but adding the S4 contribution caused the result to worsen.

Finally, we used calculated structure factors. The results from four sulfur contributions are shown in Table 2(*d*). We found three solutions out of five trials. The results indicate that for successful refinement it is necessary to involve contributions of S1 and S2, which have small *B* factors compared with the others. It is worth noting that the final LCFOMs of the successful refinements are large and positive as we expect; they were negative when we used the observed data. This shows that there are large errors in the observed intensities, especially for weak reflections. We went on further and tried to solve the structure with one sulfur contribution. The phases of one atom in space group $P2_1$ are centrosymmetric and this fact seriously disturbs the refinements. To overcome this difficulty, we introduced random errors in the initial phases using the formula

$$\tan[\varphi(\mathbf{h})] = \frac{0.9\sin[\varphi_S(\mathbf{h})] + 0.1\sin[\varphi_R(\mathbf{h})]}{0.9\cos[\varphi_S(\mathbf{h})] + 0.1\cos[\varphi_R(\mathbf{h})]}, \qquad (9)$$

where $\varphi_S(\mathbf{h})$ is the phase from an S atom and $\varphi_R(\mathbf{h})$ is a random phase. We can easily escape the centrosymmetric position with this procedure. The results are shown in Table 2(*e*). It turned out that this structure is solvable if the position of S1 is known.

Similar tests were performed using the observed data of crambin. Since the protein includes six S atoms with small *B* factors as shown in Table 3, this structure is a good example for investigating the effect of S atoms. We tried to solve this structure using phases calculated from the contributions of the S atoms. The LDE procedure easily solved this structure from any one-sulfur contribution, as shown in Table 3. Furthermore, we tried to solve the structure using the phases of one sulfur in a unit cell regarding the structure as *P*1, which means that all starting phases are nearly zero. The structural information in the initial map reveals that at least one heavy atom is present in the unit cell. The program successfully produced the solution from this very small amount of information. We also tried to solve the other structures discussed previously starting from near-zero phases and the results are given in Table 4. Trials were successful except for CINOBUFAGIN, which does not contain heavy atoms.

These results are collected to indicate that LDE method is capable of solving small proteins if the data have high resolution and good quality. It is also required that the structure includes heavy atoms at least as heavy as sulfur.

Table 3
Results of refinement for crambin from one sulfur contribution.

B factors (Å^2) of the six S atoms are 4.70 (S1), 3.15 (S2), 3.42 (S3), 3.12 (S4), 3.14 (S5) and 5.11 (S6). All cases gave solutions.

Sulfur	Cycles	Initial MPE	Final MPE	Final WMPE	LCFOM
S1	163	81.33	27.57	14.18	0.4146
S2	62	78.04	27.75	14.44	0.4147
S3	53	78.96	26.91	13.29	0.4155
S4	57	78.59	27.67	14.50	0.4181
S5	78	78.33	29.29	16.60	0.4147
S6	147	82.09	28.63	15.82	0.4149

Table 4
Results of refinement from all-zero phases.

All cases except one reached solution.

	Cycles	Final MPE	Final WMPE	LCFOM
MUCCAR	110	32.58	16.11	0.4304
CINOBUFAGIN	299	87.96	86.54	0.0129
AZET	177	28.72	11.89	0.2594
ALPHA-1	396	37.77	27.52	0.4351
aPP	152	33.25	20.53	0.2175
Crambin	579	26.94	13.43	0.4136
Rubredoxin	212	28.44	11.29	0.3815
Cytochrome c_6	292	47.61	30.78	0.1882

5. Discussion

Since the LDE technique easily produces *ab initio* solutions for small structures, it might be a useful replacement for conventional direct methods. Especially if heavy atoms are present, it is worth trying to solve the structure starting from all-zero phases. The method described here, however, is time-consuming and therefore the authors expect that the LDE method may be used for structures which cannot be solved with conventional methods. Furthermore, this method has the potential to solve small proteins when high-resolution data are available.

It was shown in the last section that the LDE method can lead to complete solutions with very little structural information.

Our next aim is to develop the technique in order to solve structures as large as RNAp1 or even larger. For this purpose, a combination of Patterson methods and the LDE method is under investigation.

The authors are very grateful to Professor M. M. Woolfson for his advice and useful discussions. We thank Dr A. Yamano of Rigaku Corporation for supplying the observed data of crambin.

References

Abola, F. C., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107–132. Data Commission of the International Union of Crystallography: Bonn/Cambridge/Chester.

Bau, R., Rees, D. C., Kurtz, D. M., Scott, R. A., Huang, H., Adams, M. W. W. & Eidsness, M. K. (1998). *J. Biol. Inorg. Chem.* **3**, 484–493.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. E. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

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

Bianchi, R., Pilati, T. & Simonetta, M. (1978). *Acta Cryst.* B**34**, 2157–2162.

Colens, A., Declercq, J. P., Germain, G., Putzeys, J. P. & Van Meerssche, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.

Declercq, J. P., Germain, G. & King, G. S. D. (1977). Fourth Eur. Crystallogr. Meet. Abstr. A, pp. 279–280.

Frazao, C., Soares, C. M., Carrondo, M. A., Pohl, E., Dauter, Z., Wilson, K. S., Hervas, M., Navarro, J. A., De la Rosa, M. A. & Sheldrick, G. M. (1995). *Structure*, **3**, 1159–1169.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, T., Tickle, I. & Blundell, T. (1983). *Biopolymers*, **22**, 293–304.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–133.

Patterson, W. R., Anderson, D. H., DeGrado, W. F., Cascio, D. & Eisenberg, D. (1999). *Protein Sci.* **8**, 1410–1422.

Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.

Shiono, M. & Woolfson, M. M. (1992). *Acta Cryst.* A**48**, 451–456.