Acta Cryst. (1996). D**52**, 813–825

The *Ab Initio* Crystal Structure Solution of Proteins by Direct Methods. VI. Complete Phasing up to Derivative Resolution

CARMELO GIACOVAZZO,[a] DRITAN SILIQI,[a] JAVIER GONZALEZ PLATAS,[b] HANS-JÜRGEN HECHT,[c] GIUSEPPE ZANOTTI[d] AND BRYANT YORK[e]

[a]*Dipartimento Geomineralogico, Universita' di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy,* [b]*Departamento de Fisica Fundamental y Experimental, Universidad de La Laguna, E-38203 La Laguna, Tenerife, Spain,* [c]*Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, D-38124 Braunschweig, Germany,* [d]*Dip. di Chimica Organica, Universita', Via Marzolo, 35100 Padova, Italy, and* [e]*Northeastern University, College of Computer Science, 161 Cullinane Hall, Boston, Massachusetts 02115, USA. E-mail: giacovazzo@area.ba.cnr.it*

(*Received 16 November 1995; accepted 20 February 1996*)

Abstract

The procedure described in the papers I–V of this series [Giacovazzo, Siliqi & Ralph (1994). *Acta Cryst.* A**50**, 503–505; Giacovazzo, Siliqi & Spagna (1994). *Acta Cryst.* A**50**, 609–621; Giacovazzo, Siliqi & Zanotti (1995). *Acta Cryst.* A**51**, 177–188; Giacovazzo & Gonzalez Platas (1995). *Acta Cryst.* A**51**, 398–404; Giacovazzo, Siliqi & Gonzalez Platas (1995). *Acta Cryst.* A**51**, 811–820], aiming at estimating protein phases *via* a single heavy-atom derivative, has been improved so as to extend phase determination to all the reflections up to derivative resolution. The quality of the resulting electron-density maps is checked for a number of test strutures. Some of the maps are immediately interpretable, and some can be interpreted after some cycles of solvent flattening and/or histogram matching. The correlation with classical SIR techniques is also discussed.

1. Symbols and abbreviations

Symbols and notations are basically the same as in the papers I–V of this series (Giacovazzo, Siliqi & Ralph, 1994; Giacovazzo, Siliqi & Spagna, 1994; Giacovazzo, Siliqi & Zanotti, 1995; Giacovazzo & Gonzalez Platas, 1995; Giacovazzo, Siliqi & Gonzalez Platas, 1995). For the reader's convenience they are quoted below.

$F_p = \lvert F_p \rvert \exp(i\varphi)$	Structure factor of the protein
$F_d = \lvert F_d \rvert \exp(i\psi)$	Structure factor of the isomorphous derivative
$F_H = F_d - F_p$	Structure factor of the heavy-atom structure (*i.e.* the atoms added to the native protein
$\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}$	
$E_p = R\exp(i\varphi)$	Normalized structure factor of the protein
$E_d = S\exp(i\psi)$	Normalized structure factor of the isomorphous derivative
N	Number of non-H atoms in the primitive cell for the native protein
$\sigma_i = \sum_{j=1}^{N} Z_j^i$	Z_j = atomic number of the jth atom
$N_{\rm eq} = \sigma_2^3/\sigma_3^2$	(Statistically equivalent) number of atoms in the primitive unit cell
$[\sigma_2^3/\sigma_3^2]_p$	Value of $N_{\rm eq}$ for the native protein
$[\sigma_2^3/\sigma_3^2]_H$	Value of $N_{\rm eq}$ for the heavy-atom structure
f_j	Atomic scattering factor of the jth atom
$\sum_p = \sum_p f_j^2$	The sum is extended to the native protein atoms
$\sum_H = \sum_H f_j^2$	The sum is extended to the heavy-atom structure
$\sum_d = \sum_d f_j^2$	The sum is extended to the derivative atoms
$D_i(x) = I_i(x)/I_0(x)$	I_i = modified Bessel function of order i
$E'_d = F_d/\sum_H^{1/2} = S'\exp(i\psi)$	Derivative pseudonormalized structure factor
$E'_p = F_p/\sum_H^{1/2} = R'\exp(i\varphi)$	Native pseudonormalized structure factor
$\Delta = S' - R'$	$\Delta' = S'T - R'$
$T = D_1(2R'S')$	
APP	Avian pancreatic polypeptide
CARP	Carp muscle calcium-binding protein
E2	Catalytic domain of *Azotobacter vinelandii* dihydrolipoyl transacetylase
M-FABP	Recombinant human muscle fatty-acid-binding protein
BPO	Bacterial haloperoxidase from *Streptomyces aureofaciens*

2. Introduction

A pioneering paper by Hauptman (1982) showed how direct methods may be integrated with isomorphous-replacement techniques. The triplet phase invariant of

the native protein $\Phi = \varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}}$ was estimated *via* a von Mises distribution of which the reliability coefficient A depends on an intricate interrelationship among the six moduli $R_{\mathbf{h}}$, $R_{\mathbf{k}}$, $R_{\mathbf{h}-\mathbf{k}}$, $S_{\mathbf{h}}$, $S_{\mathbf{k}}$, $S_{\mathbf{h}-\mathbf{k}}$. The Hauptman approach has been reconsidered by Giacovazzo, Cascarano & Zheng (1988): a simpler distribution,

$$P(\Phi|R_{\mathbf{h}}, R_{\mathbf{k}}, R_{\mathbf{h}-\mathbf{k}}, S_{\mathbf{h}}, S_{\mathbf{k}}, S_{\mathbf{h}-\mathbf{k}}) \simeq [2\pi I_0(A)]^{-1} \times \exp(A\cos\Phi), \quad (1)$$

was obtained for the case 'native heavy-atom derivative', where,

$$A = 2[\sigma_3/\sigma_2^{3/2}]_p R_{\mathbf{h}} R_{\mathbf{k}} R_{\mathbf{h}-\mathbf{k}} + 2[\sigma_3/\sigma_2^{3/2}]_H \Delta_{\mathbf{h}} \Delta_{\mathbf{k}} \Delta_{\mathbf{h}-\mathbf{k}}, \quad (2)$$

and $\Delta = (|F_d| - |F_p|)/\sum_H^{1/2}$ is the pseudo-normalized difference (with respect to the heavy-atom structure). Since $[\sigma_3/\sigma_2^{3/2}]_H \gg [\sigma_3/\sigma_2^{3/2}]_p$, the Cochran parameter is often negligible with respect to the term including pseudonormalized differences: this last may attain large values even for large proteins. Since $\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}}$ may be positive or negative, positive as well as negative triplets can be identified *via* (2).

Hauptman's and Giacovazzo, Cascarano & Zheng's formulae succeeded when applied to calculated data, but failed when applied to real experimental data. The common belief was that the experimental data were too inaccurate to be used in direct-methods applications: in particular the general feeling was that lack of isomorphism between the native and derivative structures combined with errors in the experimental data and/or in their mathematical treatment hinder any success when direct-phasing procedures are applied to experimental data even if the structure solution could be straightforwardly solved *via* ideal error-free data.

The entire situation has been reconsidered in paper I of this series, which focused the attention onto the case in which diffraction data of one isomorphous derivative are available. It was shown that in such a case the direct *ab initio* solution of protein structures is feasible in principle. In paper II a direct procedure successfully applying the formula by Giacovazzo, Cascarano & Zheng to real data was described. The approach was remarkably different from the typical procedures used for small-molecule crystal structure solution: the crucial innovations concerned the normalization process (of the derivative with respect to the native protein), the phasing procedure, and the figures of merit for finding the correct among the various solutions provided by the random-starting-phase approach. It was shown that a certain number of reflexions (roughly speaking, less than 0.15 of the total number of reflexions up to the derivative resolution) could be phased with a limited phase error.

The main limitation of the procedure described in paper II was the small number of phased reflexions (rather than the quality of the assigned phases). As a consequence, the corresponding electron-density maps suffered from severe series truncation effects. However, the assigned phases were of such a high quality that they could be used as a starting point for a reliable phase-extension process. This was just the aim of paper III: the probability distribution function $P(\Delta)$ was derived which suggested, as a rule of thumb, the phasing process could be extended to reflexions with $|\Delta| \geq 0.5$. The result was that about 40% of the measured reflections (up to derivative resolution) could be phased without paying too much in terms of the quality of the new phases. The phase-extension process was fast and could be run in a completely automatic way.

The following drawbacks were still limiting the usefulness of the procedure described in paper III.

(*a*) A non-negligible number of reflections with $|\Delta| < 0.5$ but large R value remained unphased: phasing them could be a valuable contribution in making the electron-density maps interpretable.

(*b*) The solution with the highest figure of merit was not always the correct solution.

(*c*) The phasing procedure could not be applied when the derivative resolution was about 4 Å or lower.

(*d*) Pseudo-centrosymmetric phases were provided in specific space groups.

Points (*b*)–(*c*) were discussed in paper V: a more robust normalizing procedure was designed which made explicit use of the distribution $P(\Delta)$. Histogram-matching procedures were used to obtain an optimal fit of the observed Δ distribution with the expected one. The new Δ's proved statistically more meaningful and were able in most of the cases to overcome, for data up to 4 Å resolution, the disturbing consequences produced on the Wilson plot by strong Debye effects. A method was also suggested for discarding the false solutions: an *a posteriori* check of the heavy-atom positions allows the elimination of the trials corresponding to the so-called (in small-molecule direct-methods applications) 'uranium solutions'.

In this paper we will show that all the reflexions up to derivative resolution may be phased in principle, so improving the quality of the final electron-density map. The various sources of errors are briefly analysed, in order to provide greater insight into the limits and the advantages of the phasing method. Owing to such errors it is also suggested that up to about 20% of the total number of reflexions up to derivative resolution can be omitted from the phasing procedure without causing deterioration of the quality of the electron-density map. The method will be applied to a set of experimental data and its performance will be compared with those provided by the classical single isomorphous replacement (SIR) techniques.

Table 1. *Code name, space group and crystallochemical data for the test structures*

Structure code	Reference	Space group	Molecular formula	Z
APP	(1)	$C2$	$C_{190}N_{53}O_{58}Zn$	4
BPO	(2)	$P2_13$	$C_{2744}N_{712}O_{1073}$	12
E2	(3)	$F432$	$C_{1170}N_{310}O_{366}S_7$	96
M-FABP	(4)	$P2_12_12_1$	$C_{667}N_{170}O_{261}S_3$	4
NOX	(5)	$P4_12_12$	$C_{1034}O_{704}N_{299}S_2O_{1/8}$	8

References: (1) Glover *et al.* (1983); (2) Hecht *et al.* (1994); (3) Mattevi *et al.* (1992); (4) Zanotti *et al.* (1992); (5) Hecht *et al.* (1993).

Table 2. *Relevant parameters for diffraction data of the test structures*

Structure code	Native RES (Å)	Native NREFL	Heavy atom	Derivative $[\sigma_2]_H/[\sigma_2]_P$	Derivative RES (Å)	Derivative NREFL
APP	0.99	17058	Hg	0.23	2.0	2086
BPO	2.35	23956	Au	0.06	2.78	15741
			Pt	0.06	2.78	14786
E2	3.00	10388	Hg	0.08	3.0	9179
M-FABP	2.14	7595	Hg	0.06	3.0	7125
NOX	3.00	4619	Pt	0.08	3.0	4619

In Table 1 we show the code name, the space group and the crystallochemical data of our test structures, in Table 2 the relevant parameters concerning the diffraction data are given.

3. The scaling procedure and the sign inversion of Δ

According to (1) and (2) the expected value of the triplet phase Φ mostly depends on the value of $\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}}$. If $|\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}}|$ is sufficiently large and $\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}} > 0$ then Φ is expected close to zero; if $|\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}}|$ is large and $\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}-\mathbf{k}} < 0$ then Φ is expected close to π. Relationships (1) and (2) were obtained on assuming perfect isomorphism between native and derivative structures. Owing to lack of isomorphism and/or errors in the experimental data and/or in their mathematical treatment one Δ could invert its sign with respect to the sign corresponding to ideal error-free data: then the expected triplet phase should change by π.

High frequency of the sign inversion for Δ degrades the efficiency of our direct-phasing procedure. In Fig. 8 of paper II the percentage of the reflections that indergo sign inversion for Δ' was shown for APP, CARP, E2 and M-FABP. Such a figure discouraged any attempt at phasing reflections with small $|\Delta'|$ values because the frequency of sign inversion for them is too large; in particular it was larger than 0.5 for very small $|\Delta|$'s, that is, worse than for randomly distributed signs. In order to understand the reason of such a systematic error let us compare Fig. 8 of paper II (II.8) with Fig. 7 of paper II (II.7) where, for calculated error-free data, the percentage of Δ' which undergo sign inversion as a result of the mere normalization process is given. We observe that the frequency of the sign inversion at small $|\Delta|$ values in Fig. II.7 is extremely large, with a trend very close to the inversion frequency depicted in Fig. II.8. As a consequence, phase extension to reflections with small $|\Delta'|$ value is fruitful in practice only if a normalization procedure is available which avoids the systematic errors at small $|\Delta'|$ values shown in Fig. II.7.

Let us now apply the normalization process described in paper V to ideal error-free data in order to evaluate the percentage of the reflections which undergo sign inversion for Δ as a consequence of just the mathe-

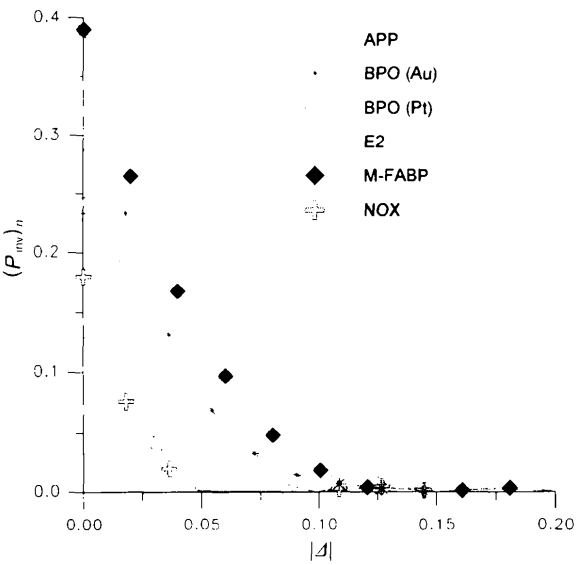


Fig. 1. Percentage of reflections which undergo sign inversion for Δ as a result of the normalization process described in paper V (calculated error-free data).

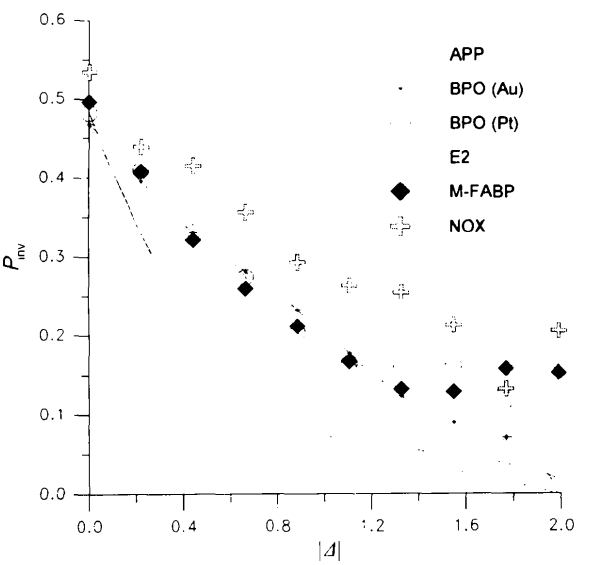


Fig. 2. Percentage of reflections that undergo sign inversion for Δ as a result of the combination of mathematical data treatment (*i.e.*, the normalization process), lack of isomorphism and errors in measurements (experimental data).

Table 3. *Cumulative average phase error (ERR) and differential average phase error (ERRD) when the phasing process is extended to reflexions with $|\Delta| > TR\Delta$ (experimental data)*

For each threshold $TR\Delta$ the value of the correlation factor (CORR) between our electron-density map and the 'correct' map obtained *via* model phases is given.

APP					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	2107	62.4 (57.9)	—	—	0.494
0.1	1864	60.1 (56.6)	243	75.5 (79.6)	0.494
0.2	1623	56.9 (54.5)	241	81.7 (81.3)	0.496
0.3	1396	54.1 (52.8)	227	73.9 (72.7)	0.496
0.4	1173	50.4 (49.9)	223	73.7 (73.5)	0.494
0.5	980	48.2 (48.1)	193	61.6 (61.1)	0.487
BPO (Au)					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	15731	62.3 (56.3)	—	—	0.461
0.1	13929	59.2 (54.5)	1802	86.6 (85.6)	0.464
0.2	12159	56.2 (52.6)	1770	75.1 (74.4)	0.461
0.3	10452	53.1 (50.4)	1707	68.6 (67.3)	0.459
0.4	8865	50.3 (48.3)	1587	68.2 (67.3)	0.448
0.5	7389	47.7 (46.2)	1476	62.4 (61.6)	0.429
BPO (Pt)					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	14777	62.9 (57.0)	—	—	0.468
0.1	13089	59.8 (55.2)	1688	86.5 (86.5)	0.469
0.2	11427	56.6 (53.1)	1662	81.8 (81.7)	0.468
0.3	9832	53.8 (51.1)	1595	74.1 (73.4)	0.462
0.4	8329	50.9 (48.9)	1503	69.8 (68.6)	0.450
0.5	6971	48.4 (46.8)	1358	63.6 (62.3)	0.437
E2					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	7756	59.8 (53.4)	—	—	0.539
0.1	6868	56.2 (51.6)	888	87.6 (87.6)	0.539
0.2	5991	52.3 (49.2)	877	82.6 (82.3)	0.538
0.3	5143	48.8 (46.7)	848	73.5 (74.5)	0.532
0.4	4369	45.5 (44.3)	774	67.3 (67.3)	0.527
0.5	3644	42.6 (42.0)	725	60.3 (60.4)	0.507
M-FABP					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	7122	69.4 (64.0)	—	—	0.401
0.1	6301	66.4 (62.3)	821	92.6 (92.4)	0.401
0.2	5499	64.2 (60.8)	802	81.5 (81.0)	0.398
0.3	4727	61.3 (58.9)	772	81.5 (79.7)	0.393
0.4	3998	58.9 (56.9)	729	74.8 (73.2)	0.379
0.5	3337	56.1 (54.7)	661	73.0 (70.9)	0.367
NOX					
TRΔ	NREFL	ERR(W-ERR)	NREFLD	ERRD(W-ERRD)	CORR
0.0	4613	77.7 (72.6)	—	—	0.281
0.1	4084	75.7 (71.4)	529	93.1 (92.8)	0.281
0.2	3559	74.7 (70.5)	525	82.3 (81.6)	0.280
0.3	3065	72.9 (69.1)	494	86.2 (84.2)	0.280
0.4	2595	71.6 (67.8)	470	80.3 (79.4)	0.277
0.5	2163	69.5 (65.7)	432	82.1 (80.0)	0.269

matical data treatment. The results are shown in Fig. 1: the inversion frequency [called here $(P_{\text{inv}})_n$] is practically negligible up to $|\Delta| = 0.08$, and is never larger than 0.4, so confirming the higher efficiency of the new normalization process. In order to have a counter check that such a normalization process allows a fruitful phase extension to small $|\Delta|$ values we show in Fig. 2 the inversion frequency for the test structures in Table 1 as obtained by applying to the experimental data the normalization procedure described in paper V. The inversion frequency in Fig. 2 is dramatically smaller at low $|\Delta|$ than in Fig. II.8; this confirms the higher quality of the normalization procedure proposed in paper V, and suggests that even reflections with $|\Delta|$ smaller than 0.5 could be conveniently phased.

The problems that now exist are as follows. Is it worthwhile phasing all the reflections up to derivative resolution, or can some reflexions be excluded from the phasing process without detriment to the quality of the electron-density map? How can we select the reflections to which it is worthwhile extending the phasing process? Can some general criteria be fixed? We will show that a minimum threshold for $|\Delta|$ and the standard deviation associated with reflection

intensity measurements are sensible criteria to be applied.

4. The average phase error as a function of $|\Delta|$

Fig. 2 shows that for most experimental data $P_{inv} \leq 0.50$. The maximum value is attained close to $|\Delta| \simeq 0$, where the combined effect of the mathematical treatment of the data (see Fig. 1) and of the lack of isomorphism make the experimental sign of Δ completely unreliable. A reasonable criterion has to be found of balancing the advantage of employing more data against the disadvantage that the extra data may have a higher phase error.

The phasing process described in the papers III and V extends step by step the phase determination to batches of reflexions with progressively smaller values of $|\Delta|$. The trend of the phase error as a function of the threshold TRΔ for $|\Delta|$ may be deduced from Table 3, where, for each batch, we give the number of reflections (NREFL), the cumulative average phase error (ERR) and the weighted cumulative phase error (W-ERR). A better insight can be obtained by calculating the differential average phase error (ERRD). While ERR refers to all the reflections with $|\Delta| > \text{TR}\Delta$, ERRD refers only to the NREFLD reflections with $|\Delta|$ between the current TRΔ value and the preceding one.

Both ERR and ERRD increase by decreasing TRΔ values, but ERRD in a dramatic way. However, extending phases to a larger number of reflections improves the quality of our electron-density map ρ (Lunin & Woolfson, 1993). This may be monitored by calculating the correlation factor (CORR) between ρ and the 'correct' map ρ_{mod} (obtained *via* model phases, all reflections up to derivative resolution included),

$$\text{CORR} = (\langle\rho\rho_{mod}\rangle - \langle\rho\rangle\langle\rho_{mod}\rangle)/ \times [(\langle\rho^2\rangle - \langle\rho\rangle^2)^{1/2}(\langle\rho_{mod}^2\rangle - \langle\rho_{mod}\rangle^2)^{1/2}].$$

In Table 3 CORR is given for the various maps calculated by using reflections with $|\Delta| > \text{TR}\Delta$. CORR increases with decreasing values of TRΔ except for very small TRΔ values. It may be argued from Table 3 that in most cases reflections with $|\Delta| \leq 0.1$ do not provide valuable additional information to the electron-density map, so that they could be removed from the phasing process.

5. The role of the standard deviation of the intensity measurements

While the effects of the lack of isomorphism are *a priori* unpredictable, the effects of the errors in measurements may be partially controlled by exploiting the standard deviation $\sigma(|F|)$ usually associated with the structure-factor modulus $|F|$. The value of $\sigma|(F|)$ is probably an underestimate of the total error since it takes into account only errors coming out from the statistical fluctuations in the X-ray intensity, and neglects systematic variations like inaccuracy in absorption corrections, misalignment of the crystal, *etc.* However, $\sigma|(F|)$ provides a reasonable estimate of the relative reliability of different measurements and therefore can be used for estimating the percentage of sign inversions for Δ caused by counting errors in intensity measurements.

Let us suppose that errors in measurements are distributed according to the normal distribution then,

$$\sigma(|\Delta|) \simeq [\sigma^2(|F_p|) + \sigma^2(|F_d|)]^{1/2} \sum_H^{-1/2}.$$

On the first approximation the variable $y = |\Delta|/\sigma(|\Delta|)$ may be considered to be distributed, in the absence of systematic errors, according to the Gaussian distribution $N[y; y_m, 1]$ where $y_m = [|\Delta|/\sigma(|\Delta|)]_m$ is the values of y obtained from measurements. For positive Δ values (but the final results will hold also for negative values of Δ) the probability of the sign inversion is equal to the integral,

$$\int_{-\infty}^{0} N[y; y_m, 1]\,dy. \qquad (3)$$

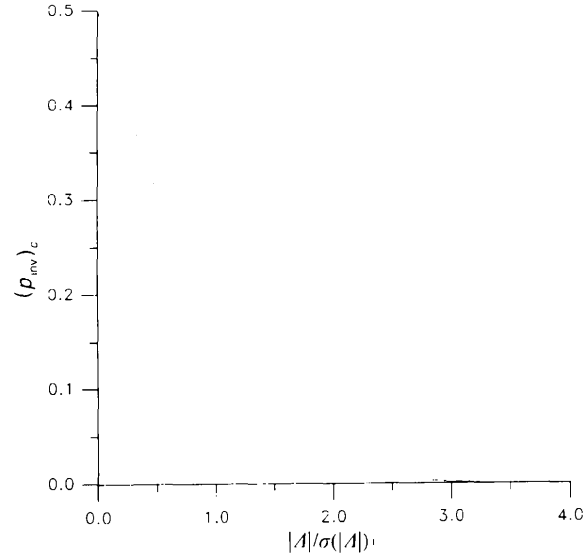


Fig. 3. The sign-inversion frequency for Δ as a result of the counting errors in intensity measurements (experimental data).

Table 4. *NREF is the number of reflections with* $|\Delta|/\sigma(|\Delta|) < 0.8$*; ERR is the average phase error for such reflections, (W-ERR) is the weighted average phase error*

Structure code	NREF	ERR (W-ERR)
APP*	—	—
BPO (Au)	3196	82.6 (81.4)
BPO (Pt)	2455	85.2 (84.4)
E2	1352	90.8 (90.8)
M-FABP	1376	89.7 (89.5)
NOX	440	95.4 (95.4)

* $\sigma(|F|)$ are not available for this structure.

Table 5. *For each test structure the number of reflections with* $|\Delta|/\sigma(|\Delta|) > 0.8$ *and* $|\Delta| > 0.1$ *are given, together with the average phase error and the correlation coefficient CORR of our map with the model electron-density map*

	NREFL	ERR(W-ERR)	CORR
APP*	1864	60.1 (56.6)	0.494
BPO (Au)	12449	56.9 (52.9)	0.463
BPO (Pt)	11736	57.2 (53.4)	0.470
E2	6240	53.2 (49.5)	0.540
M-FABP	5664	64.1 (60.5)	0.402
NOX	4039	75.4 (71.2)	0.285

* Only the criterion $|\Delta| > 0.1$ is used.

If the normalized variable $z = y - y_m$ is substituted for y, the probability of the sign inversion due to counting errors is equal to,

$$(P_{\text{inv}})_c = \int_{-\infty}^{-y_m} N[z; 0, l]\,dz$$
$$= 1 - \left\{ 1/(2\pi)^{1/2} \int_{-\infty}^{y_m} \exp(-z^2/2)\,dz \right\}$$
$$= 1 - \Phi(y_m), \qquad (4)$$

which is tablulated in standard books.

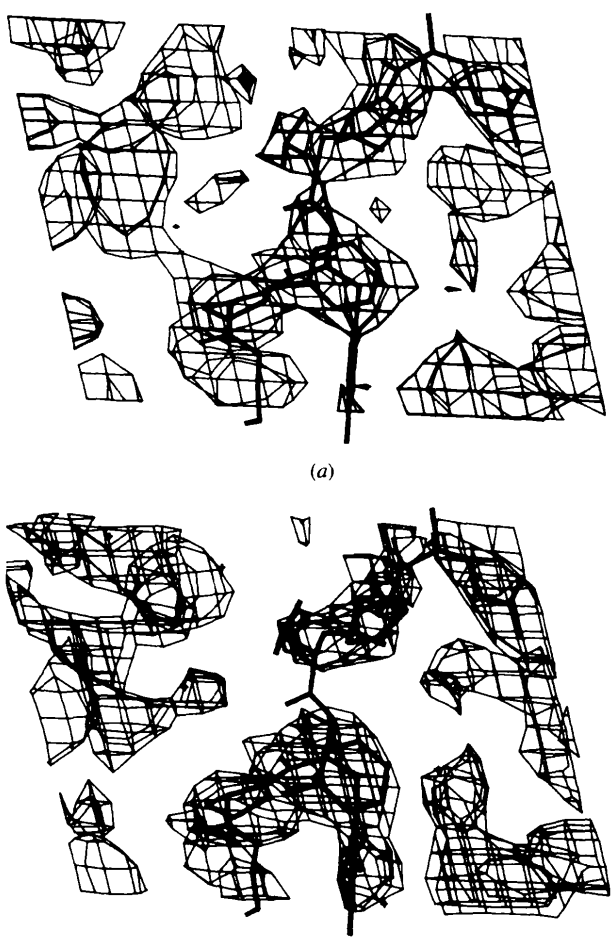
$\Phi(y_m)$ is the well known error functions: to guide the reader we show $(P_{\text{inv}})_c$ in Fig. 3. The figure suggests that: (*a*) $(P_{\text{inv}})_c$ is about 0.15 for the reflections with $|\Delta|/\sigma(|\Delta|) = 1$. Since the robustness of the phasing method can tolerate larger sign-inversion frequency, even reflections with $|\Delta|/\sigma(|\Delta|) \leq 1$ can be fruitfully involved in the phasing process. A more reasonable criterion may be to include reflections with $|\Delta|/\sigma(|\Delta|) \geq 0.8$: this corresponds to $(P_{\text{inv}})_c \simeq 0.20$. Such a criterion does not characterize a fixed percentage of the reflections, since this number depends on the quality of the data.

The efficiency of the criterion may be judged from Table 4, where, for the various test structures, the average phase error (as calculated by our phasing procedure) is given for the reflections with $|\Delta|/\sigma(|\Delta|) \leq 0.80$. It is seen that such reflections, once phased, should not add relevant information to the electron-density map, and could, therefore, be excluded from the phasing process.

When reflections with $|\Delta|/\sigma(|\Delta|) \leq 0.80$ or $|\Delta| \leq 0.1$ are excluded from the procedure the CORR values shown in Table 5 are obtained. The table shows that the average phase of the phased reflections is markedly smaller than in Table 3 at $\text{TR}\Delta = 0$, and that no structural information should be substantially lost by applying the above omit criteria.

6. About the quality of the electron-density maps

For some of the test structures the mean-phase error is sufficiently low to suggest that some of the electron-density maps could be straightforwardly interpreted. We discuss here the quality of the various maps calculated at the end of our procedure (DM maps from now on) and we compare these with the maps obtained with standard single ismorphous techniques.



(*a*)

(*b*)

Fig. 4. Portion of the electron-density map of APP, showing the region around residues from 4 to 8. Phases from direct methods (*a*) and SIR (*b*).

6.1. *APP*

Avian pancreatic polypeptide is a 36-amino-acid hormone. It crystallizes in space group $C2$, with cell dimensions $a = 34.18$, $b = 32.92$, $c = 28.45$ Å, $\beta = 105.26°$ and crystals diffract to 0.98 Å resolution (Glover *et al.*, 1983). The single heavy atom in the asymmetric unit simulates, in space group $C2$, a centric heavy-atom substructure: in this case, as already stressed in paper II, the enantiomorph may get lost during the phasing process, and the electron-density map may be pseudo-centrosymmetric. The electron density for the main chain, calculated with direct-method phases, is reasonably continuous in some regions (see Fig. 4*a*), but contains breaks and is discontinuous in other areas. Moreover, the level of noise is quite high, possibly owing to the pseudo-centrosymmetric features of the DM phases in this space group. Anyhow, since the molecule is quite small, at least for a protein, the map could be interpreted, at least partially. Maps calculated with SIR phases are definitively worse and not interpretable (see Fig. 4*b*). This result should be rather unexpected if one remembers that such a small protein can also be solved directly from the native data (Woolfson & Yao, 1990; Sheldrick, Dauter, Wilson, Hope & Sieker, 1993). However, Δ area available only up to 2 Å resolution, while *ab initio* solution methods exploit all the R_p's up to atomic resolution.

6.2. *BPO*

The bacterial haloperoxidase from *Streptomyces aureofaciens* consists of 277 residues and belongs structurally to the class of the α/β hydrolases (Hecht, Sobek, Haag, Pfeifer & Van Pee, 1994). It crystallizes in the cubic space group $P2_13$ with two molecules per asymmetric unit and a cell dimension $a = 126.5$ Å. Its structure has been solved by multiple isomorphous replacement (MIR) using two derivatives and has been refined to 2.05 Å resolution with an R value of 18.4%.

SIR maps of the platinum and gold derivative, phased with *MLPHARE* (Collaborative Computational Project, Number 4, 1984), show, when compared with the refined structure, numerous breaks and wrong connections in the electron density and would have been difficult to interpret in an attempted structure solution. Solvent flattening, histogram matching and phase extension from 3.0 to 2.8 Å with the package *CCP*4 (Cowtan, 1994) reduces the $R_{\rm free}$ value, calculated for 10% of the reflections, from 52 to 36% and yields much improved electron-density maps, which still show some discontinuities and incorrect connections but would have been interpretable. The electron-density map obtained after phasing with *MLPHARE* using both derivatives shows, after solvent flattening, histogram matching and phase extension from 3.0 to 2.8 Å, very easily interpretable density.

Table 6. *BPO: statistics characterizing the electron-density maps calculated with CCP4 fft (resolution range 10.0–2.80 Å)*

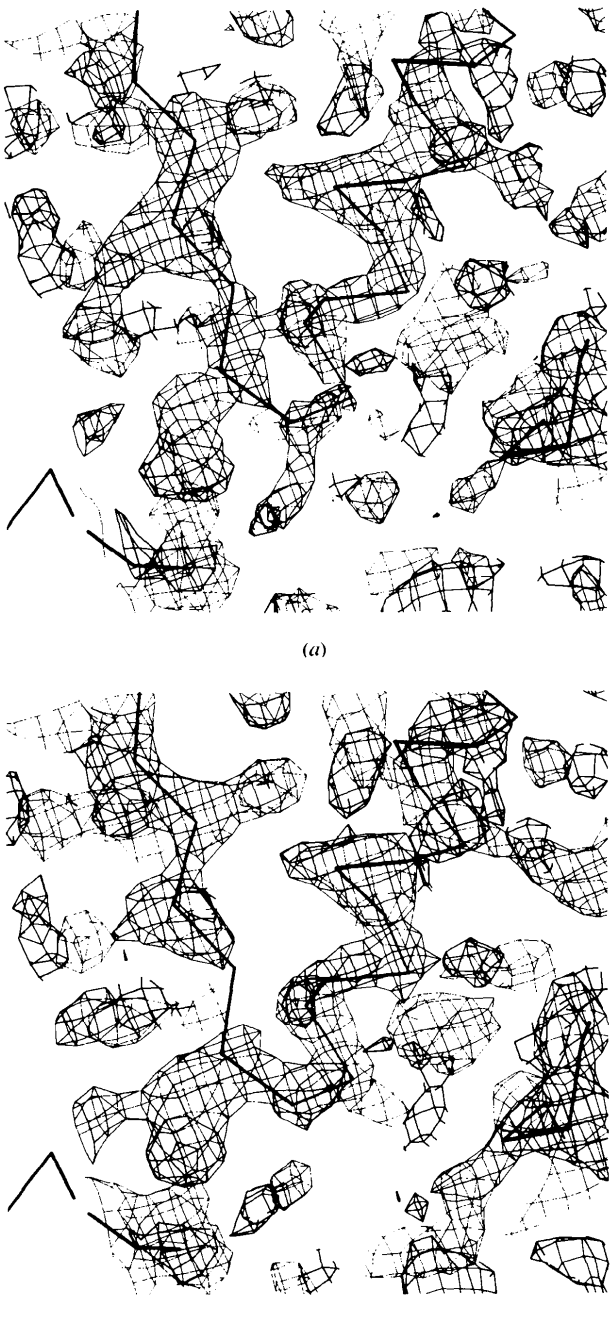
aui, pti, 'incomplete' data sets: reflections with $|\Delta|/\sigma(|\Delta|) \leq 0.80$ or $|\Delta| < 0.1$ are excluded; suffix '_dm', after density modification with *CCP*4 dm; suffix 'DM', obtained by direct methods (the present procedure); r.m.s., r.m.s. value from *CCP*4 *fft*; max, maximum density from *CCP*4 *fft*; fragm > 100, No. of fragments > 100 in skeletonization with *mapman* star value 1.3 r.m.s., increment 1.0 r.m.s.; max fragm, maximum fragment length in skeletonization with *mapmar*; skewness, calculated with *mapman*; skew/sig, skewness/sigma(skewness) calculated with *mapman*.

	max/r.m.s.	fragm > 100	max fragm	skewness	skew/sig
mir	5.40	8	543	0.30	104.3
mir_dm	6.27	6	859	0.76	268.1
au-sir	5.03	4	188	0.04	14.5
au_sir_dm	6.19	2	1224	0.73	255.9
pt_sir	5.24	3	437	0.15	53.5
pt_sir_dm	6.62	2	1047	0.79	275.3
pt_DM	12.53	5	220	0.18	63.7
pt_DM_dm	8.97	2	1562	0.62	216.0
au_DM	12.76	5	163	0.17	58.7
au_DM_dm	8.91	3	1767	0.57	199.8
aui_DM	12.94	5	131	0.17	60.1
aui_DM_dm	9.26	2	1755	0.60	209.7
pti_DM	12.59	3	217	0.19	66.2
pti_DM_dm	9.25	3	1412	0.64	223.4
$2f_o - f_c$	6.16	2	2481	0.83	290.7

For comparison of MIR and SIR results, electron-density maps before and after density modification were calculated from four phase sets obtained by direct methods. These data sets also were subjected to the same density-modification procedure as the MIR and SIR data sets. The statistics characterizing the maps are compiled in Table 6. The data sets indicated 'pt_DM' and 'au_DM' contained phases with weight larger than zero for 89.7 and 95.5% of the reflections in the resolution range 10.0–2.8 Å, data sets 'pti_DM' and 'aui_DM' were 71.2 and 75.6% complete in the same resolution range [we omitted reflections with $|\Delta|/\sigma(|\Delta|) \leq 0.80$ or $|\Delta| < 0.1$].

For a visual evaluation of the maps the program *O* (Jones, Zou, Cowan & Kjeldgaard, 1991) was used. All electron-density maps calculated with phases obtained by direct methods show very high correlation to the corresponding SIR maps (Table 7), while the correlation between the platinum- and the gold-derived maps is low, similar to the low correlation between the platinum and gold SIR maps. When compared to the MIR maps before and after density modification these correlation factors are slightly lower for the directly phased maps than for the SIR-phased maps. This may be caused in part by high residual electron density at the position of the derivative atom in maps, which is only partly eliminated by density modification. Compared with the SIR maps these directly phased maps show at least in some regions improved connectivity (Figs. 5*a* and 5*b*). Therefore, the map derived from the platinum derivative would have been interpretable, while the map derived from the gold derivative shows poorer

connectivity and would have been more difficult to interpret. Density modification reduces the $R_{\rm free}$ from 54 to 41% for the platinum, and the gold data set again yields much improved and interpretable electron density (Fig. 6). The less complete data sets 'pti_DM' and 'aui_DM' show map-correlation factors very similar to the more complete data sets, but somewhat worse connectivity, and, therefore, interpretability, of the density, as expected from the lower completeness.



(*a*)

(*b*)

Fig. 5. BPO. (*a*) Electron density of the platinum map phased by direct methods showing the nucleophile ellbow at the active site Ser, and C^{α} trace. (*b*) Electron density of the platinum SIR map showing the nucleophile ellbow at the active site Ser. The C^{α} trace of the refined structure is indicated with trick lines. The maps are countered at 1σ and were drawn with *O* (Jones *et al.*, 1991).

6.3. *E2*

The crystal structure of the catalytic domain of *Azotobacter vinelandii* dihydrolipoyl transacetylase was determined at 2.6 Å resolution by Mattevi *et al.* (1992). One subunit is present in the asymmetric unit of the cubic cell (space group *F*432) with $a = b = c = 224.8$ Å. The structure was solved using MIR techniques, with three heavy-atom derivatives, one of which (Hg), of quite good quality has been used for own tests. The structure has been refined to a crystallographic *R* factor of 0.188 at 2.6 Å resolution. We have calculated four electron-density maps: map *A* was calculated *via* direct methods, map *B* *via* SIR phases, map *C* *via* DM phases improved by solvent-flattening procedure (Wang, 1995), map *D* *via* SIR phases improved by solvent flattening. A portion of the unit cell according to maps *A*, *B*, *C* and *D* are shown in Fig. 7. On the map in Fig. 7(*b*) residuals corresponding to heavy-atom position can be noted. Maps *C* and *D* are in fact the better, since solvent regions have been cut away and the tracing of the protein chain appears straightforward. Anyhow, the quality of the other two is also reasonably good. On closer inspection all

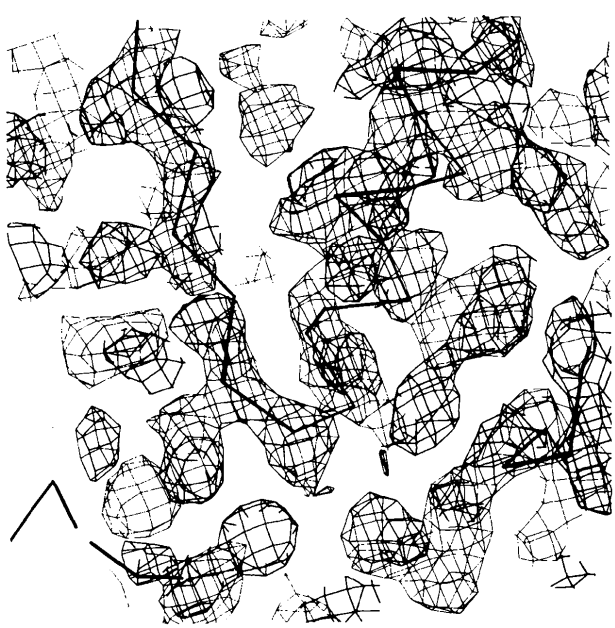


Fig. 6. BPO. Electron-density map after density modification of the gold directly phased map showing the nucleophile elbow at the active site Ser, and C^{α} trace.

Table 7. *BPO: map-correlation factors (calculated with CCP4 overlapmap)*

	mir	mir_dm	au_sir	au_sir_dm	pt_sir	pt_sir_dm	pt_DM	pt_DM_dm	au_DM	au_DM_dm	aui_DM	aui_DM_dm	pti_DM	pt_DM_dm
$2f_o - f_c$	0.66	0.83	0.49	0.73	0.53	0.74	0.46	0.65	0.45	0.65	0.46	0.65	0.46	0.65
mir	—	0.81	0.77	0.70	0.78	0.69	0.64	0.64	0.66	0.65	0.67	0.66	0.65	0.64
mir_dm	—	—	0.61	0.82	0.63	0.81	0.55	0.72	0.55	0.72	0.56	0.73	0.55	0.72
au_sir	—	—	—	0.69	0.32	0.41	0.27	0.36	0.85	0.68	0.86	0.68	0.27	0.36
au_sir_dm	—	—	—	—	0.43	0.66	0.37	0.57	0.61	0.81	0.61	0.82	0.37	0.57
pt_sir	—	—	—	—	—	0.69	0.81	0.67	0.28	0.38	0.28	0.39	0.82	0.66
pt_sir_dm	—	—	—	—	—	—	0.60	0.79	0.36	0.56	0.37	0.58	0.60	0.80
pt_DM	—	—	—	—	—	—	—	0.75	0.24	0.33	0.24	0.33	0.98	0.73
pt_DM_dm	—	—	—	—	—	—	—	—	0.33	0.49	0.33	0.51	0.74	0.95
au_DM	—	—	—	—	—	—	—	—	—	0.76	0.98	0.74	0.24	0.33
au_DM_dm	—	—	—	—	—	—	—	—	—	—	0.75	0.94	0.33	0.50
aui_DM	—	—	—	—	—	—	—	—	—	—	—	0.75	0.24	0.33
aui_DM_dm	—	—	—	—	—	—	—	—	—	—	—	—	0.34	0.51
pti_DM	—	—	—	—	—	—	—	—	—	—	—	—	—	0.74

the maps appear to be easily interpretable, the main chain is continuous and the shape of side chains is often reasonably clear.

6.4. *M-FABP*

Human muscle fatty-acid-binding protein is a single-chain protein of 132 amino acids. The protein contains ten antiparallel β-strands, arranged into approximately β-sheets, and two α-helices. Crystals are orthorhombic, $P2_12_12_1$, with $a = 35.4$, $b = 56.7$, $c = 72.7$ Å. The crystal structure was solved by using two heavy-atom derivatives and molecular replacement (Zanotti, Scapin, Spadon, Veerkamp & Sacchettini, 1992) and refined to an R factor of 0.195 at 2 Å resolution.

The electron-density maps calculated with direct methods show encouraging correlation coefficient ($\simeq 0.401$) with the map obtained with calculated phases. The density in the map is quite continuous and corresponds reasonably to the model, but several breaks along the main chain would make difficult the interpretation. For comparison, an SIR map was also calculated. The latter, despite being not too different from the previous one, is definitively not interpretable. A portion of the two maps, showing the region around residues from 121 to 127, is shown in Fig. 8. Finally, direct methods and SIR phases were improved (Figs. 8*c* and 8*d*) by a density-modification procedure (Wang, 1985). Contours of molecules in the cell are now clearly visible in Fig. 8(*c*), and the continuity of the main chain has been improved.

6.5. *NOX*

The structure of NADH oxidase from *Thermus thermophiilus* (Hecht *et al.*, to be published) was solved by MIR techniques (four derivatives). The P6 *cis*-Pt derivative is used in our calculations. In this case the derivative data were of very poor quality and showed significant lack of isomorphism. Consequently the SIR-phased maps were uninterpretable. The DM-phased map showed, as for BPO, high electron density at the location of the heavy atom, and was not interpretable. Density modification by solvent flattening and histogram matching with *CCP4* programs lowered the free R value from 62.2 to 54.6%, but failed to produce an interpretable electron-density map. The calculations were performed in the resolution range 10.0–3.0 Å and the free R value was calculated for 10% of these reflections. The completeness of the data sets was 93% for the derivative data set and the directly phased data set was 84% complete.

7. The relation between SIR techniques and direct methods

Our direct-methods procedure estimates triplet phases Φ in absence of any information on the heavy-atom structure, while the SIR method estimates single phases provided the heavy-atom structure is known. In spite of such formal differences the numerical results shown in the preceding paragraph show that DM and SIR electron-density maps are strongly correlated with each other. This suggests that the two techniques themselves should be highly correlated. Two questions arise: what is the nature of the correlation? Are DM and SIR estimates consistent? In order to answer the above questions we will analyse a few practical cases.

Let us suppose that Δ_1, Δ_2, Δ_3 are all positive and large. Then DM suggest $\Phi \approx 0$, where '$\approx$' represents 'probably equal to'. Once the heavy-atoms have been located SIR techniques suggest,

$$\text{if } \Delta_1 > 0, \varphi_{p_1} \approx \varphi_{H_1} \tag{5a}$$

$$\text{if } \Delta_2 > 0, \varphi_{p_2} \approx \varphi_{H_2} \tag{5b}$$

$$\text{if } \Delta_3 > 0, \varphi_{p_3} \approx \varphi_{H_3} \tag{5c}$$

summing (5*a*)–(5*c*) gives,

$$\Phi \approx \Phi_H, \tag{6}$$

where $H_1 + H_2 + H_3 = 0$ and $\Phi_H = \varphi_{H_1} + \varphi_{H_2} + \varphi_{H_3}$. If $|\Delta_1|$, $|\Delta_2|$, $|\Delta_3|$ are sufficiently large $|E_{H_1}|$, $|E_{H_2}|$, $|E_{H_3}|$ will also be large (by definition $|\Delta_i| \leq |E_{H_i}|$). If the number of heavy atoms in the unit cell is small (as

usual) then Φ_H is expected close to zero, so that DM and SIR triplet estimates agree. When one of the Δ_i's is relatively small DM and SIR techniques may diverge since Φ_H is expected far from zero.

Suppose now that $\Delta_1 > 0$, $\Delta_2 > 0$ and $\Delta_3 < 0$. Then DM suggest $\Phi \approx \pi$ and SIR techniques provide, once the heavy atoms have been located, the following indications,

$$\text{if } \Delta_1 > 0, \varphi_{p_1} \approx \varphi_{H_1} \tag{7a}$$

$$\text{if } \Delta_2 > 0, \varphi_{p_2} \approx \varphi_{H_2} \tag{7b}$$

$$\text{if } \Delta_3 < 0, \varphi_{p_3} \approx \varphi_{H_3} + \pi, \tag{7c}$$

summing (7*a*)–(7*c*) gives,

$$\Phi \approx \Phi_H + \pi,$$

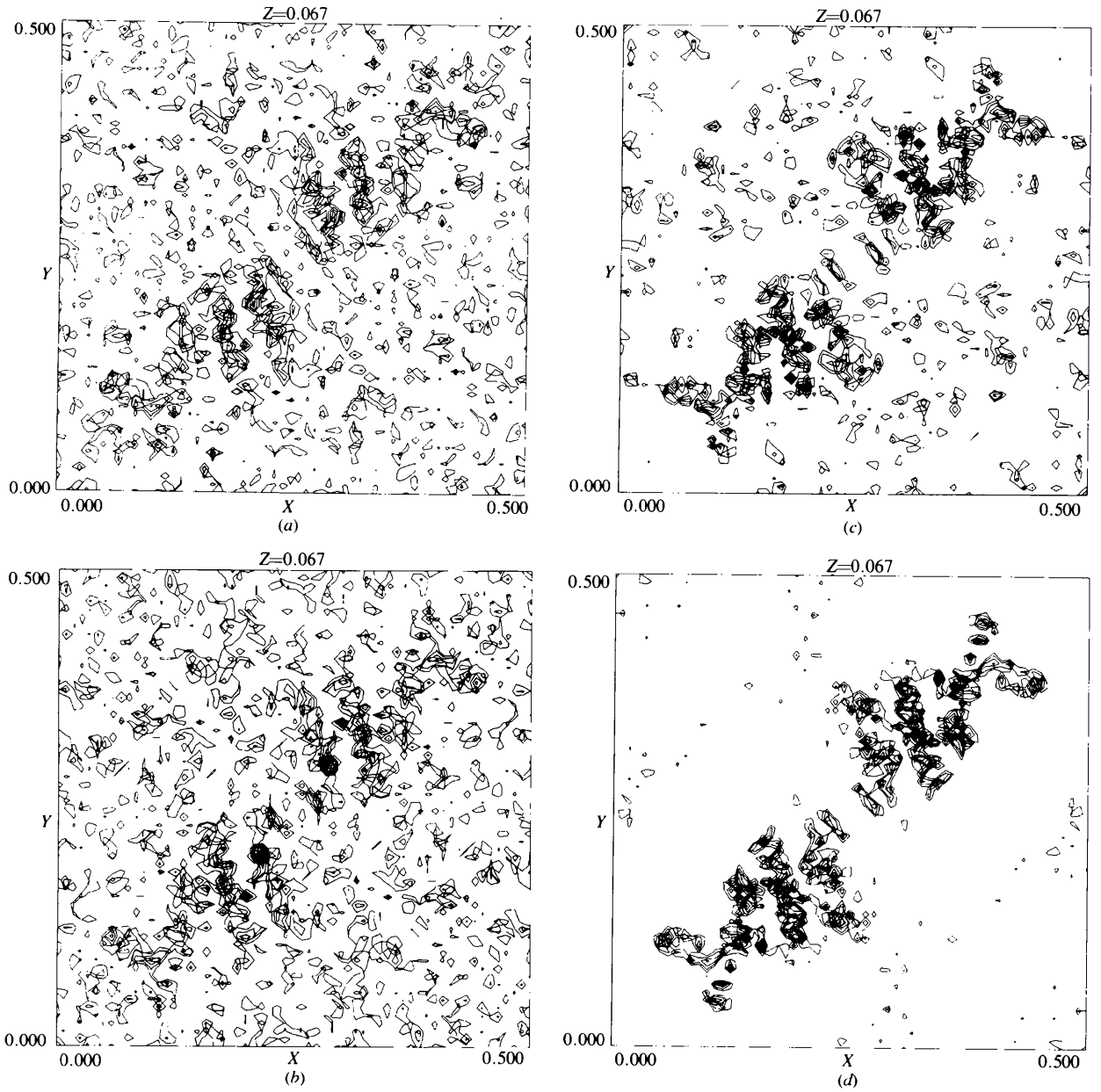


Fig. 7. E2. Electron-density maps of a portion of the unit cell. x ranges from 0 to 0.5, y from 0 to 0.5 in the vertical axis, and z from 0.012 to 0.067, perpendicularly to the plane of the paper. Contour levels are drawn with 1σ increment, starting from 1σ. Phases used are: (*a*) DM phases; (*b*) SIR phases; (*c*) DM phases improved by solvent-flattening procedure; (*d*) SIR phases improved by solvent-flattening procedure.

Table 8. *E2: the values* $\langle|\Phi_H|\rangle$ *for selected ranges of* $G_H = 2[\sigma_3/\sigma_2^{3/2}]_H|E_{H_1}E_{H_2}E_{H_3}|$

NTRIP is the number of triplet invariants for which $\langle|\Phi_H|\rangle$ is calculated (TRΔ = 0.84).

G_H	NTRIP	$\langle\|\Phi_H\|\rangle$
0.0–0.2	15292	85.9
0.2–0.4	25129	79.2
0.4–0.8	66819	68.0
0.8–1.2	51528	53.7
1.2–1.6	30043	41.8
1.6–2.0	17011	33.6
2.0–15.0	18163	25.9

Table 9. *E2: the values* $\langle|\Phi_H|\rangle$ *for selected ranges of* $G_H = 2[\sigma_3/\sigma_2^{3/2}]_H|E_{H_1}E_{H_2}E_{H_3}|$

NTRIP is the number of triplet invariants for which $\langle|\Phi_H|\rangle$ is calculated (TRΔ = 0.0).

G_H	NTRIP	$\langle\|\Phi_H\|\rangle$
0.0–0.2	176086	85.6
0.2–0.4	224541	78.8
0.4–0.8	446161	67.4
0.8–1.2	295015	53.1
1.2–1.6	165372	41.5
1.6–2.0	90630	33.7
2.0–15.0	105796	25.3

which again agrees with DM estimates only if $\Phi_H \approx 0$. Two questions now arise. (1) Is the condition $\Phi_H \approx 0$ fulfilled in the practice for most of the cases? (2) If DM and SIR techniques diverge which ones should be considered more reliable?

As for the first question we show in Table 8 for E2 the average value of $\langle|\Phi_H|\rangle$ calculated for all the reflections with $|\Delta| > 0.84$. It is easy seen that $\langle|\Phi_H|\rangle$ is far from being close to zero in most of the cases. Since the number of heavy atoms in the unit cell is small the high value of $\langle|\Phi_H|\rangle$ is rather surprising. However, all becomes clear when one considers that, because of the errors above discussed, large experimental values of $|\Delta|$ are often associated to smaller values of R_H. In the practice a non-negligible percentage of reflexions with $|\Delta| > 0.84$ show R_H values markedly smaller than 0.84. The statistics does not qualitatively change (see Table 9) if calculations are made for all the triplets used in the phasing process (*i.e.* for TRΔ = 0).

We have so far proved that DM and SIR techniques are not equivalent methods: therefore, high but not unitary correlations are to be expected between SIR and DM maps, as confirmed by our practical applications. A second question now arises: if SIR and DM estimates for a single triplet diverge, which one should be considered more efficient?

Let us first examine the case of triplets with phase value symmetry restricted to $(0, \pi)$, and having $\Phi_H = \pi$. In Fig. 9(*a*) we show the case in which $\Delta_1\Delta_2\Delta_3 > 0$: according to DM Φ is expected to be zero while its true value is π. In Fig. 9(*b*) the

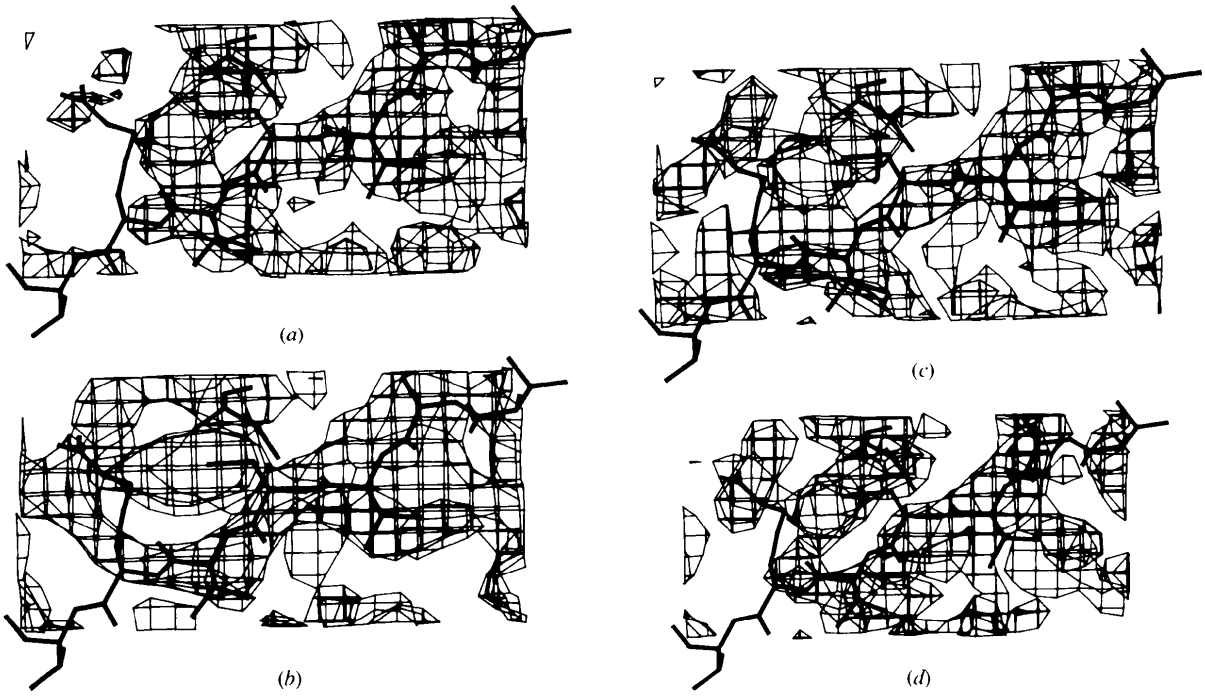


Fig. 8. Electron-density map of MFABP. A detail of the region around residues from 121 to 127 is shown. Maps were calculated (*a*) with DM; (*b*) with SIR; (*c*) with DM phases improved by density modification and (*d*) with SIR phases improved by density modification.

case $\Delta_1\Delta_2\Delta_3 < 0$ is depicted: according to (1) Φ is expected to be π while the true value is 0. In both the cases the SIR method successfully phases the three reflections while DM fail.

Let us now consider the so called 'cross-over' case: despite the condition $\Delta > 0$ the sign of F_p is reversal with respect to F_H, and $\cos(\varphi_d - \varphi_p) < 0$. Triplets with one or more reflections showing cross-over cannot be correctly estimated by (1) even if $\Phi_H = 0$. As an example, let us consider (see Fig. 10) the case in which: (*a*) the three reflections forming the triplet have symmetry-restricted phase values to $(0, \pi)$; (*b*) the Δ_i, $i = 1, 2, 3$ are all positive; (*c*) the third reflection shows a cross-over. Then (1) will always estimate $\Phi \approx 0$ while, correctly, SIR method would provide,

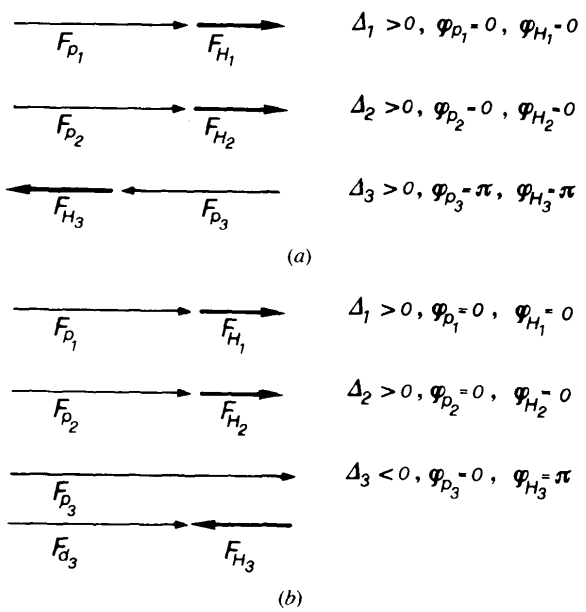


Fig. 9. Triplets with phase value symmetry restricted to $(0, \pi)$, having $\Phi_H = \varphi_{H_1} + \varphi_{H_2} + \varphi_{H_3} = \pi$. (*a*) $\Delta_1\Delta_2\Delta_3 > 0$: according to (1) the triplet phase is expected to be 2π but the true value is π. (*b*) $\Delta_1\Delta_2\Delta_3 < 0$: according to (1) the triplet phase is expected to be π but the true value is 2π.

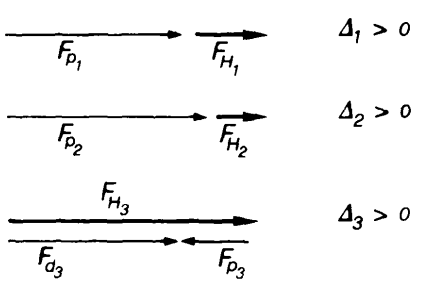


Fig. 10. Triplet with phase values symmetry restricted to $(0, \pi)$ having $\Phi_H = 0$. The third reflection shows a 'cross-over'.

$$\varphi_{p_1} \approx \varphi_{H_1} \tag{8a}$$

$$\varphi_{p_2} \approx \varphi_{H_2} \tag{8b}$$

$$\varphi_{p_3} \approx \varphi_{H_3} + \pi, \tag{8c}$$

from which by summation,

$$\Phi \approx \Phi_H + \pi = \pi.$$

What can we conclude about the relative reliability of DM and SIR techniques from the above examples? Unlike DM, the SIR method requires the prior information on the heavy-atom structure. Extracting it from the experimental data requires preliminary work, but, once available, the knowledge of F_H constitutes a valuable supplementary information. Thus, if we focus our attention on a single triplet there is no doubt that the DM estimate of Φ, derived in absence of any information on the heavy-atom structure, is equivalent or inferior to the SIR estimate. However, DM have some basic advantages.

(*a*) The SIR method is restricted to assign phase values equal to φ_H or $\varphi_H + \pi$. On the contrary, reflections are phased by DM by combining the indications provided by hundreds and often by thousands of triplets. This cooperative action can drive phases far away from φ_H or $\varphi_H + \pi$, so compensating the lack of information on the heavy-atom substructure.

(*b*) Once protein phase estimates are available, the heavy-atom positions can be easily identified. This supplementary information can be exploited by modified probabilistic procedures (Fortier, Moore & Fraser, 1985; Klop, Krabbedan & Kroon, 1987).

(*c*) DM potentially is not exhausted by triplet estimates. Quartet invariants can also be estimated, providing (negative quartet in particular) supplementary information to the phasing process (Giacovazzo & Siliqi, 1996*a*,*b*).

(*d*) The theory of representation (Giacovazzo, 1977, 1980) can be applied to the invariant estimation process. More accurate estimates will lead to better electron-density maps.

(*e*) Direct-methods algorithms, so efficient for the crystal structure solution of, small molecules, can be used to automatize the entire phasing process.

8. Conclusions

This paper is the conclusion of a series which, starting from the theoretical results obtained by Hauptman (1982) and by Giacovazzo *et al.* (1988), has been devoted to make those achievements practicable. The phasing procedure designed throughout this series can be described as follows: without any information on the heavy-atom positions the phasing process is able to provide in favourable cases electron-density maps which may be directly interpreted. The process is able in principle to phase all the reflections up to the

derivative resolution and may be accomplished in a fully automatic way, so adding appeal to the method. Bad isomorphism between the native and the derivative hinder the success: the maps are then not straightforwardly interpretable but still show an interesting correlation with the correct map.

The method described is highly correlated with traditional SIR techniques, but observations have been made suggesting that direct methods still have a reserve of power. The guidelines for future improvements of the method are described.

The authors are grateful to Consejeria de Educacion y Ciencia, del Gobierno de Canarias, for the partial financial support for the stay for one of the authors. Thanks are also due to Drs A. Mattevi and W. G. J. Hol for making the E2 data available.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–767.

Cowtan, K. (1994). *Jnt CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38.

Fortier, S., Moore, N. J. & Fraser, M. E. (1985). *Acta Cryst.* A**41**, 571–577.

Giacovazzo, C. (1977). *Acta Cryst.* A**33**, 933–944.

Giacovazzo, C. (1980). *Acta Cryst.* A**36**, 362–372.

Giacovazzo, C., Cascarano, G. & Zheng, C.-D. (1988). *Acta Cryst.* A**44**, 45–51.

Giacovazzo, C. & Gonzalez Platas, J. (1995). *Acta Cryst.* A**51**, 398–404.

Giacovazzo, C. & Siliqi, D. (1996*a*). *Acta Cryst.* A**52**, 133–142.

Giacovazzo, C. & Siliqi, D. (1996*b*). *Acta Cryst.* A**52**, 143–151.

Giacovazzo, C., Siliqi, D. & Gonzalez Platas, J. (1995). *Acta Cryst.* A**51**, 811–820.

Giacovazzo, C., Siliqi, D. & Ralph, A. (1994). *Acta Cryst.* A**50**, 503–510.

Giacovazzo, C., Siliqi, D. & Spagna, R. (1994). *Acta Cryst.* A**50**, 609–621.

Giacovazzo, C., Siliqi, D. & Zanotti, G. (1995). *Acta Cryst.* A**51**, 177–188.

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

Hauptman, H. (1982). *Acta Cryst.* A**38**, 289–294.

Hecht, H., Erdmann, H., Park, H., Sprinzl, M., Schmid, R. D. & Schomburg, D. (1993). *Acta Cryst.* A**49**, Suppl. 86.

Hecht, H., Sobek, H., Haag, T., Pfeifer, O. & Van Pee, K. H. (1994). *Nature Struct. Biol.* **1**, 532–537.

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Klop, E. A., Krabbendam, H. & Kroon, J. (1987). *Acta Cryst.* A**43**, 810–820.

Lunin, V. Y. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K. H., Westphal, A. H., De Kok, A. & Hol, W. G. J. (1992). *Science*, **255**, 1544–1550.

Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. (1993). *Acta Cryst.* D**49**, 18–23.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Woolfson, M. M. & Yao, J.-X. (1990). *Acta Cryst.* A**46**, 409–413.

Zanotti, G., Scapin, G., Spadon, P., Veerkamp, J. H. & Sacchettini, J. C. (1992). *J. Biol. Chem.* **267**, 18541–18550.