Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 8 January 2008
Accepted 18 June 2008

The ($F_o - F_c$) Fourier synthesis: a probabilistic study

Rocco Caliandro, Benedetta Carrozzini, Giovanni Luca Cascarano, Liberato De Caro, Carmelo Giacovazzo* and Dritan Siliqi

Institute of Crystallography – CNR, Via G. Amendola, 122/O 70126 Bari, Italy. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

($F_o - F_c$) and ($2F_o - F_c$) Fourier syntheses are considered the most powerful tools for recovering the remainder of a structure and for correcting crystal structure models. A probabilistic approach has been applied to derive the formula for the variance for the expected value of the coefficient ($F_o - F_c$). This has allowed a better understanding of the features of the difference Fourier synthesis; in particular, a subset of well phased reflections has been separated from the subset of reflections best phased by the standard F_o Fourier synthesis. An iterative procedure, based on the electron-density modification of the difference Fourier map, has been devised which aims to improve phase and modulus estimates of the reflections with higher variance value, by using as lever arm the set of reflections with lower variance value. The new procedure (DEDM) has been implemented and verified on a wide set of test structures, the partial models of which were obtained by molecular replacement or by automatic model-building routines applied to experimental electron-density maps. Phase and modulus estimates of the difference Fourier syntheses improve in all the test cases; as a consequence, the quality of the difference Fourier maps also improves in the region where the target structure deviates from the partial model. A new procedure is suggested, combining DEDM with standard electron-density modification techniques, which leads to significant reduction of the phase errors. The procedure may be considered a starting point for further developments.

1. Notation

The following notation has been used in this article.

N: number of atoms in the unit cell of the true structure.

p: number of atoms in the structural model. Usually $p \leq N$, but it may also be that $p > N$.

$\mathbf{r}_j$, f_j, $j = 1, \ldots, N$: 'true' atomic positions and corresponding scattering factors (displacement parameters included).

$\mathbf{r}'_j = \mathbf{r}_j + \Delta\mathbf{r}_j$, g_j, $j = 1, \ldots, p$: positional vectors of the atoms belonging to the model structure and corresponding scattering factors (displacement parameter included).

B_j, $j = 1, \ldots, N$: isotropic displacement parameters of the N atoms in the unit cell of the true structure.

$B'_j = B_j + \Delta B_j$, $j = 1, \ldots, p$: isotropic displacement parameters of the atoms in the model structure.

$s = 2\sin\theta/\lambda$.

$F = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{h}\mathbf{r}_j)$: 'true' structure factor for the reflection $\mathbf{h}$.

$F_o = F + |\mu|\exp(i\theta)$: observed structure factor; $|\mu|$ is the measurement error.

$F_p = \sum_{j=1}^{p} g_j \exp(2\pi i \mathbf{h}\mathbf{r}'_j)$: structure factor of the model structure.

$E = A + iB = R\exp(i\varphi)$, $E_p = A_p + iB_p = R_p\exp(i\varphi_p)$: normalized structure factors of F and F_p, respectively.

ε: statistical Wilson coefficient (corrects for expected intensities in reciprocal-lattice zones).

$F_q = \sum_{j=p+1}^{N} f_j \exp(2\pi i \mathbf{h}\mathbf{r}_j)$: structure factor of the substructure constituted by the atoms that are part of the N-atom structure but not of the p-atom structure.

$\Sigma_N = \varepsilon \sum_{j=1}^{N} f_j^2$.

$\Sigma_p = \varepsilon \sum_{j=1}^{p} g_j^2$.

$D = \langle \exp(-s^2\Delta B/4)\cos(2\pi\mathbf{h}\Delta\mathbf{r})\rangle$: the average involves the p atoms and is performed per resolution shell.

$\sigma_A = D \sum_{j=1}^{p} f_j g_j / (\Sigma_N \Sigma_p)^{1/2}$.

$e = 1 + \langle\mu\rangle/\Sigma_N$.

$I_i(x)$: modified Bessel function of order i.

$m = \langle\cos(\varphi - \varphi_p)\rangle = I_1(X)/I_0(X)$: X is defined in the text.

2. Introduction

Many Fourier syntheses can be used to recover the remainder of a structure when a partial structure is known (Ramachandran & Raman, 1959; Srinivasan, 1961; Ramachandran & Srinivasan, 1970). Today the use is mostly restricted to the observed, to the difference and to the so-called ($2F_o - F_c$) Fourier syntheses.

The observed synthesis is the conventional one: noncentric reflections are usually weighted according to Sim (1959)

[$m|F|\exp(i\varphi_c)$ are the coefficients, where m is the Sim weight], centric reflections according to Woolfson (1956). The maps show missing atoms at about half the intensity, even less when more of the structure is missing (Luzzati, 1952).

In order to obtain peaks with equal intensity, irrespective of whether they belong to the known or to the remainder of the substructure, the $(2F_o - F_c)$ synthesis is used [it may be considered the sum of an F_o and an $(F_o - F_c)$ synthesis]. Main (1979) proposed the use of $(2m|F| - |F_c|)$ coefficients. Read (1986) observed that electron-density maps calculated from a p-atom substructure are biased towards the substructure. He used Main's algebraic approach to introduce errors into the cosine law and proposed to eliminate the model bias component by using $(2m|F| - D|F_c|)$ coefficients for noncentric reflections. The factor D compensates for errors in the atomic positions, scattering and B factors. More recently a paper by Caliandro *et al.* (2005) has generalized the expression of the factor D to take into account measurement errors also.

Cochran (1951) studied the properties of the $(F_o - F_c)$ Fourier synthesis. Its efficiency has been discussed by Henderson & Moffat (1971) and by Ursby & Bourgeois (1997). These last authors studied, *via* Bayesian statistics, the influence of measurement errors on the efficiency of the synthesis.

In the present paper, we present some unknown features and properties of the difference Fourier synthesis derived *via* a probabilistic approach (see §§3 and 4). The study has also suggested iterative procedures to improve both the quality of the difference Fourier synthesis (see §5) and the quality of the best observed electron-density map (see §6). The procedures have been implemented in the program *IL MILIONE* (Burla *et al.*, 2007). The first applications are quoted in §§6 and 7.

3. About the coefficients of the difference Fourier synthesis

Let ρ, ρ_p and ρ_q be the true electron density, a model density and the difference structure, respectively; F, F_p and F_q are their Fourier transforms. $\rho_q = \rho - \rho_p$ is the ideal difference Fourier synthesis. It is defined by the following property: summed to ρ_p it exactly provides ρ irrespective of the quality of the model structure. Its Fourier transform

$$F_q = F - F_p = |F|\exp(i\varphi) - |F_p|\exp(i\varphi_p) = |F_q|\exp(i\varphi_q) \tag{1}$$

provides the ideal difference structure factors which, added to F_p, enable the complete electron density ρ to be recovered from ρ_p. If ρ_p is part of ρ then ρ_q is positive everywhere; if ρ_p contains false or misplaced atoms then ρ_q is not a positive defined function.

Since the values of φ are unknown, the ideal difference structure factors [equation (1)] remain unknown; common practice is to calculate the difference Fourier synthesis by using the difference structure factors, estimated *via* the expression

$$\langle F_q\rangle = \langle F\rangle - \langle F_p\rangle. \tag{2}$$

The averages in equation (2) may be calculated by using the conditional distributions of the joint probability function (see Luzzati, 1952; Read, 1986; Caliandro *et al.*, 2005),

$$P(R, R_p, \varphi, \varphi_p) = RR_p\pi^{-2}(e - \sigma_A^2)^{-1}\exp\{-(e - \sigma_A^2)^{-1} \times [R^2 + eR_p^2 - 2\sigma_A RR_p\cos(\varphi - \varphi_p)]\}, \tag{3}$$

from which the following conditional phase probability distribution may be derived:

$$P(\varphi|R, R_p, \varphi_p) = [2\pi I_0(X)]^{-1}\exp[X\cos(\varphi - \varphi_p)], \tag{4}$$

where $X = 2\sigma_A RR_p/(e - \sigma_A^2)$. The value of σ_A may be obtained *via* the relation $\langle R^2R_p^2\rangle = (e + \sigma_A^2)$.

According to equation (3),

$$\langle|F|\exp(i\varphi)\rangle = m|F|\exp(i\varphi_p). \tag{5}$$

The values provided by equation (5) are the coefficients of the best observed Fourier synthesis as defined by Blow & Crick (1959): it is 'that Fourier transform which is expected to have the minimum mean-square difference from the Fourier transform of the true F's when averaged over the whole unit cell' and it is 'obtained by using the centroid of the probability distribution for F.'

The concept of best synthesis may be extended to the difference Fourier synthesis provided we are able to estimate the mean value of $|F_p|\exp(i\varphi_p)$. The result is (Read, 1986)

$$\langle|F_p|\exp(i\varphi_p)\rangle = D|F_p|\exp(i\varphi_p), \tag{6}$$

where D may be estimated as $\sigma_A/b^{1/2}$, b being the intercept of the line that fits the distribution of σ_A *versus* s. Accordingly, the coefficients of the best difference Fourier synthesis may be assumed to be the difference between two expected values:

$$\begin{aligned}\langle F_q\rangle &= (m|F| - D|F_p|)\exp(i\varphi_p)\\ &= |m|F| - D|F_p||\exp[i(\varphi_p + s\pi)],\end{aligned} \tag{7}$$

where $s = 1$ or 0 according to whether the sign of $m|F| - D|F_p|$ is positive or negative. The main problem arising from equation (7) is the following: while for the estimation of the observed Fourier synthesis we use the estimated phases and the observed moduli $|F|$, for the calculation of ρ_q both moduli (that is $|m|F| - D|F_p||$) and phases are estimates.

A relevant point is to establish how reliable such estimates are. To answer the question it may be useful to calculate the variance (say σ_q) of F_q. The usefulness of considering the variances has already been presented by Ursby & Bourgeois (1997), who introduced suitable weights to associate with the difference Fourier coefficients, in order to take into account measurement errors. To estimate σ_q we first calculate

$$\begin{aligned}\langle|F_q|^2\rangle &= \langle|F|^2 + |F_p|^2 - 2|FF_p|\cos(\varphi - \varphi_p)\rangle\\ &= ||F|^2 + D^2\langle|F_p|^2\rangle - 2mD|F||F_p|| + \langle|\mu|^2\rangle\end{aligned} \tag{8}$$

and then we derive

$$\langle|F_q|\rangle^2 = m^2|F|^2 + D^2|F_p|^2 - 2mD|FF_p|, \tag{9}$$

Table 1
Test structures used in the analysis.

PDB is the PDB code of the protein structure (apart from the structure enhexa, which is not yet deposited in the PDB), Space is its space group, Res is the data resolution in Å and NresT is the number of residues; Model indicates either the PDB code of the model structure used in the molecular replacement procedure or that the structure was built by *ARP/wARP* starting from a map obtained by *ab initio* phasing, NresM is the number of residues of the model structure, and MPE is the mean phase error (in degrees) of the phases φ_p, calculated from the model structure, with respect to the published phases φ.

PDB	Space	Res	NresT	Model	NresM	MPE
enhexa†	$P4_32_12$	1.2	130	1paz	130	22
1kf3	$P2_1$	1.0	124	7rsa	124	26
6rhn	$P4_32_12$	2.2	115	4rhn	104	32
1zs0	$P2_12_12_1$	1.6	163	1i76	163	42
3ebx	$P2_12_12_1$	1.4	62	by *ARP/wARP*	40 (+10 Gly)	45
1na7	$P2_1$	2.4	329	1m2r	327	48
1a6m	$P2_1$	1.0	151	1mbc	153	48
1e3u	$P2_1$	1.7	970	by *ARP/wARP*	520 (+492 Gly)	51
2sar	$P2_12_12_1$	1.8	192	1ucl, chain *A*	96	53
1pm2	$P2_12_12_1$	1.8	678	by *ARP/wARP*	363 (+322 Gly)	56
1i9a	$P4_12_12$	2.4	364	by *ARP/wARP*	165 (+78 Gly)	58
1kqw	$I4$	1.8	134	1opa	133	60
1lys	$P2_1$	1.7	258	2ihl	129	64
6ebx	$P2_12_12_1$	1.7	124	3ebx	62	72
2iff	$P2_1$	2.6	556	2hem	129	72
1cgn	$P6_522$	2.2	127	2ccy	122	73
1bxo	$C2$	1.0	323	1er8	338	74

† Honnappa (2008).

from which the variance

$$\sigma_q^2 = \langle |F_q|^2 \rangle - \langle |F_q| \rangle^2 = (1 - m^2)|F|^2 + \langle |\mu|^2 \rangle \qquad (10)$$

is obtained. The role of σ_q may be elucidated by three extreme examples.

Case 1: Both m and D are large, and both $|F|$ and $|F_p|$ are large and have comparable moduli (see Fig. 1*a*). In this situation the relation $\varphi \simeq \varphi_p$ is highly reliable, but it is impossible to decide if F lies (in the Gauss plane) on the right or on the left side of F_p. In Fig. 1(*a*) the two possibilities are shown simultaneously; the expected value of φ_q approximately changes by π when passing from one choice to the other. The high phase uncertainty is indicated in equation (10) by a large variance value.

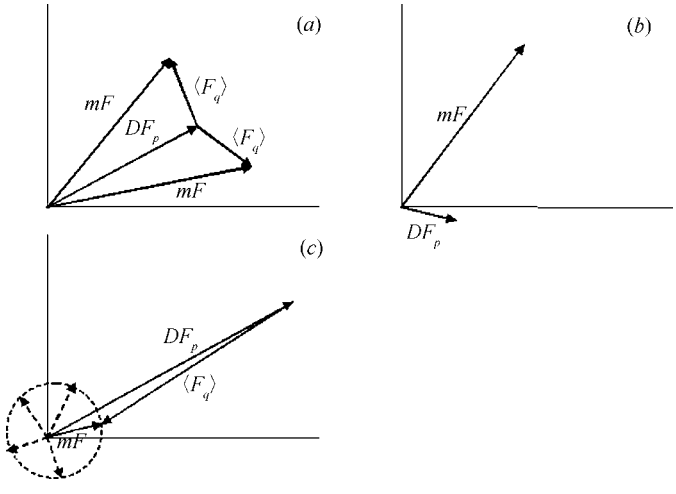

Figure 1
Schematic view of three situations that can occur in the difference Fourier synthesis: (*a*) both m and D are large, and both $|F|$ and $|F_p|$ are large and have comparable moduli; (*b*) $|F|$ is quite large, $|F_p|$ is small, and m and D are close to 0.5; (*c*) $|F|$ is small while $|F_p|$ is quite large.

Case 2: $|F|$ is quite large, $|F_p|$ is small, and m and D are close to 0.5 (see Fig. 1*b*). In this situation, the phase value to associate with $|m|F| - D|F_p||$ is φ_p, but this assignment is rather unreliable; indeed, owing to the small $|F_p|$ value, the relation $\varphi_p \simeq \varphi$ is not reliable. Again the large σ_q value recognizes that the estimate is highly uncertain.

Case 3: $|F|$ is small while $|F_p|$ is quite large (see Fig. 1*c*). In this situation the phase value to associate with $|m|F| - D|F_p||$ is $\varphi_p + \pi$. The small value of the variance suggests that the phase indication is highly reliable; indeed, whatever the value of φ, the value of φ_q must be close to $\varphi_p + \pi$ to satisfy the constraint $|F| = |F_p| + |F_q| \simeq 0$. The reliability of the phase indication does not depend on the quality of the model density.

The above three examples suggest that σ_q, according to circumstances, may assume quite different values. That opens new perspectives for an improved estimation of ρ_q.

4. An experimental check of the variance

A step frequently encountered in protein crystallography is the refinement of a model structure. In a high percentage of cases (about 70%) the model is obtained by molecular replacement techniques; in the remaining cases it is attained by applying single/multiple-wavelength anomalous diffraction (SAD–MAD), single/multiple isomorphous replacement (SIR–MIR) or *ab initio* methods. Electron-density modification techniques are applied to the best electron-density maps obtained by one of the above methods, then automatic model building routines are often applied.

In the case of molecular replacement the well known problem of the model bias has to be overcome: *i.e.* the model structure (almost) correctly placed in the unit cell has to be modified until the target structure is reached. To check the correctness of the theoretical results obtained in §3 we used, as targets, the test structures quoted in Table 1. For each structure we report its Protein Data Bank (PDB) code, space group (Space), data resolution (Res) and number of residues (NresT). The model structures were located by *REMO* (Caliandro *et al.*, 2006), a molecular replacement program included in the package *IL MILIONE*. Their PDB codes (Model), the corresponding number of residues (NresM) and the average phase error (*i.e.* $\langle |\varphi - \varphi_p| \rangle$) at the end of the molecular replacement procedure (MPE) are also listed in Table 1.

For three test cases, *i.e.* 2sar, 6ebx and 1lys, the target proteins are constituted by two monomers related by noncrystallographic symmetry, while the corresponding models contain only one monomer (in fact in Table 1 the number of residues of the model is half that of the target structure). The molecular replacement run has been performed by searching for only one copy of the model, and the first *REMO* solution, representing one monomer correctly

placed, has been retained as model structure. In the following these cases will be considered as preferential tests for Fourier difference applications; for these cases the region where the target structure is nearly reproduced by the model is well separated from the region where only the target structure is present.

In Table 1, four additional test cases were included; their electron-density maps were obtained *ab initio* by *IL MILIONE* and the starting models were built by *ARP/wARP* (Perrakis *et al.*, 1999). In these cases the models consist of docked residues and of nondocked residues considered as glycines. Their numbers are reported in the column NresM.

In equation (4) $|R|$ and $|R_p|$ play a symmetrical role in establishing the reliability of the relation $\varphi \simeq \varphi_p$; indeed the reliability parameter depends, *via* suitable parameters, on the product RR_p. Conversely, the results obtained in §3 suggest that $|F|$ and $|F_p|$ play different roles in establishing the reliability of the phase φ_q. Indeed, owing to the corresponding different variances, the case in which $|F|$ is large and $|F_p|$ is small is quite different from the case in which $|F|$ is small and $|F_p|$ is large, even if the value of $||m|F| - D|F_p||$ is the same.

To check this property, we have divided the reflections into batches, each batch corresponding to a given value of $(m|F| - D|F_p|)$ and containing an equal number of reflections. For each batch we calculated CrRES_q and MPE_q. The first is the classical crystallographic residual between $|F_q|$ [the modulus of the ideal difference structure factor, as given by equation (1)] and its estimate $\langle|F_q|\rangle = |m|F| - D|F_p||$ (*i.e.* the modulus of the best difference Fourier synthesis). The second is the average difference between φ_q [the phase of the ideal difference structure factor, as given by equation (1)] and its expected value $(\varphi_p + s\pi)$. CrRES_q and MPE_q estimate, respectively, the modulus and the phase error resulting from the replacement of the ideal difference Fourier coefficient F_q by the best difference Fourier coefficient $\langle F_q\rangle$. The values of MPE_q *versus* $m|F| - D|F_p|$ for the three test structures are superimposed in Fig. 2. We observe that

(i) MPE_q shows a maximum when $|m|F| - D|F_p||$ is close to zero, as expected. We have verified (but not shown for brevity) that this is also true for CrRES_q.

(ii) According to equation (10), of the two batches having the same value of $|m|F| - D|F_p||$, that with a negative value of $m|F| - D|F_p|$ is expected to have smaller values of MPE_q. Fig. 2 shows that this trend is fully satisfied. We verified that the distribution of CrRES_q is instead symmetric with respect to $m|F| - D|F_p|$.

The above experimental results agree with equation (10) and indicate that the ratio $\langle|F_q|\rangle/\sigma_q$ may be used as a criterion to select the reflections for which φ_q and $|F_q|$ are more accurately estimated. To this end we calculated the standard deviation σ_q according to equation (10), and we introduced a weight w proportional to $1/\sigma_q$. In Fig. 3 we plot MPE_q *versus* $w\langle|F_q|\rangle$ for the test structure 2sar; for reference, the plot of MPE_q *versus* $\langle|F_q|\rangle$ is superimposed. In accordance with the theoretical expectations, $w\langle|F_q|\rangle$ ranks the phase error (and also CrRES_q) much better than $\langle|F_q|\rangle$.

The above results seem to suggest that the weighted difference Fourier synthesis with coefficients $w(m|F| - D|F_p|)$ may be more useful than the simple $m|F| - D|F_p|$ Fourier synthesis. To check the correctness of the expectation we calculated for all the test structures the two corresponding electron densities ($\langle\rho_{qw}\rangle$ and $\langle\rho_q\rangle$, respectively) and we evaluated their correlation with the ideal difference structure ρ_q calculated *via* the coefficients $F - F_p$. The results were disappointing: in spite of the larger correlation values, the weighted Fourier difference maps contain a negligible amount of new information with respect to the model electron density. We identified two main reasons: (*a*) quite often $\langle\rho_{wq}\rangle$ is anti-correlated with ρ_p, owing to the enhanced weight given to the reflections with negative value of $\langle F_q\rangle$; (*b*) the total correlation with the true electron density ρ does not improve, since the

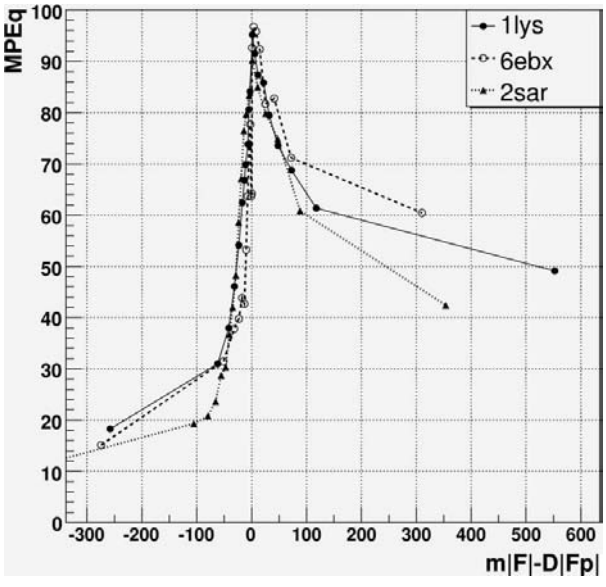

Figure 2
MPE_q *versus* $m|F| - D|F_p|$ for the three test structures used. The bins are chosen so as to contain an equal number of reflections.

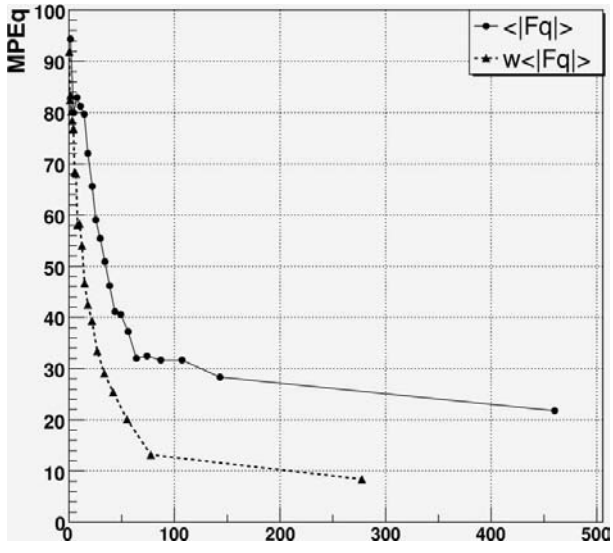

Figure 3
MPE_q *versus* $\langle|F_q|\rangle$ (circles) and $w\langle|F_q|\rangle$ (triangles) for the test structure 2sar. The bins are chosen so as to contain an equal number of reflections.

reflections with largest weight have vanishing $|F|$ values and therefore do not contribute to ρ. This last observation suggests that better electron-density maps may be obtained if a procedure is devised that is able to improve the phases of the reflections with large $|F|$ modulus (see §5).

5. The difference electron-density modification procedure

The theoretical results described in §3 and tested in §4 indicate that

(1) there are reflections for which the φ values may be accurately estimated *via* equation (4) [*i.e.* when F and F_p are sufficiently large] while the corresponding φ_q values cannot be reliably estimated;

(2) there are reflections for which the φ values cannot be accurately estimated *via* equation (4) [*i.e.* when F and/or F_p are small] while their φ_q values are reliably estimated [*i.e.* when $\langle F_q \rangle$ is sufficiently large and negative].

The reflections of point (1) play a central role in the electron-density modification (EDM) procedures that aim to recover a more accurate observed electron density starting from a map obtained *via ab initio*, SAD–MAD, SIR–MIR or molecular replacement methods. Any EDM algorithm modifies an F_o or a $2F_o - F_c$ electron-density map to fit the positivity of the map, also *via* histograms in one or more dimensions, in direct and/or reciprocal space, and tries to capture the stereochemical information and improve the phase estimates.

It may be argued that the reflections of point (2) may play a crucial role in a difference electron-density modification (DEDM) procedure, designed to improve difference electron-density maps. DEDM and EDM should work in substantially independent ways. Indeed,

(*a*) while the reflections playing a crucial role in EDM are those defined in point (1), those more useful for DEDM are the reflections defined in point (2);

(*b*) the restraints on the difference Fourier map are not based on the positivity of the electron density (indeed $\langle \rho_q \rangle$ shows both positive and negative regions).

In this section we plan to improve, *via* a DEDM procedure, the pairs $(\langle \varphi_q \rangle, \langle |F_q| \rangle)$ by exploiting the good phase information available for the reflections with a large negative value of $(m|F| - D|F_p|)$. To this purpose, we divide the set of reflections into three subsets:

Subset A: reflections with $m|F| - D|F_p| \simeq 0$.

Subset B: reflections with $m|F| - D|F_p| \gg 0$.

Subset C: reflections with $m|F| - D|F_p| \ll 0$.

The three subsets are sketched in Fig. 4, where the initial values of MPE_q are plotted *versus* $m|F| - D|F_p|$ for the structure 2sar. Their limits are initially set so that subset B includes 70% of the reflections with $m|F| - D|F_p| > 0$ and subset C includes 70% of the reflections with $m|F| - D|F_p| < 0$ (the limits are slightly varied during the DEDM cycles to avoid border effects).

Each subset will play a different role in the DEDM procedure. The phases and the moduli of the reflections in subset C are well determined. They are considered as prior information in the DEDM procedure. The starting phases of the reflections belonging to subset B are strongly correlated with the model phases: thus, bringing them closer to φ_q will improve the difference Fourier map. The phases of the reflections in subset A are weakly correlated with the model phases and may influence the quality of the difference map if the DEDM procedure is able to improve their values, provided the vanishing $\langle |F_q| \rangle$ moduli are replaced by more correct and larger values.

The DEDM procedure we propose may be described in terms of four main steps:

Step 1: Initialization. The observed amplitudes are converted to an absolute scale and normalized (Giacovazzo *et al.*, 2002), the structure factors F_p are calculated from the model structure and the starting difference Fourier synthesis is calculated according to equation (7). In accordance with the previous observations, the weight associated with σ_q does not multiply the coefficients of the difference Fourier synthesis; it is only used throughout the DEDM procedure to select reflections (see Step 2).

Step 2: Phase assignment. Several DEDM cycles $\rho_q \rightarrow \langle \varphi_q \rangle \rightarrow \rho_q$ are performed, in which $\langle |F_q| \rangle$ remain equal to their initial values, while some of the $\langle \varphi_q \rangle$ values are allowed to vary; the cycles aim to move the phase values away from their initial values $(\varphi_p + s\pi)$ by, hopefully, guessing the correct sense of variation. In each half cycle $\rho_q \rightarrow \langle \varphi_q \rangle$ the electron-density map is modified as follows: 80% of the positive map (*i.e.* the smallest intensity pixels) and 90% of the negative map (*i.e.* the pixels with intensity close to zero) are set to zero; the rest is squared. The new $\langle \varphi_q \rangle$ values are immediately associated with the reflections of subset A (they have the worst initial phases). For low-weight reflections belonging to subset B the $\langle \varphi_q \rangle$ values obtained at the end of

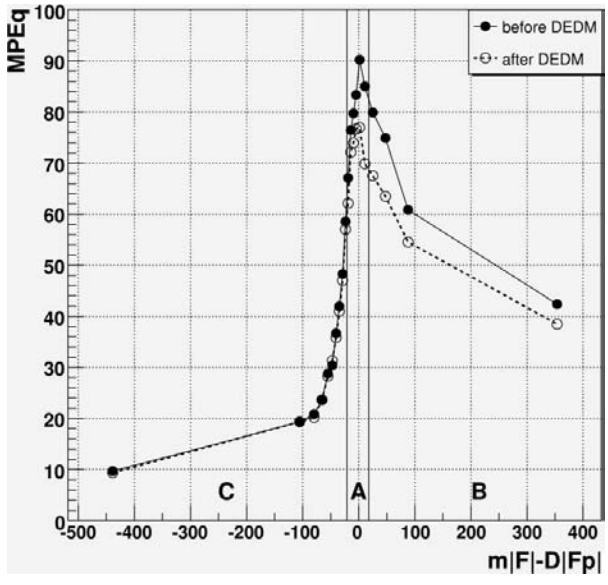

Figure 4
MPE_q *versus* $m|F| - D|F_p|$ for the test structure 2sar, before and after the application of DEDM. Two vertical lines define the subsets A, B and C described in the text.

Table 2
Results of the DEDM procedure.

$CrRES_q$ is the crystallographic residue of the estimated moduli $\langle|F_q|\rangle$ with respect to the ideal ones $|F_q|$; MPE_q is the mean phase error (in degrees) of the estimated phases $\langle\varphi_q\rangle$ with respect to the ideal ones φ_q; $CORR_q$ is the correlation factor between the difference Fourier map and the ideal one. For each parameter the values before and after the application of the DEDM procedure are reported.

PDB	$CrRES_q$	MPE_q	$CORR_q$
enhexa	0.8 → 0.8	52 → 45	0.67 → 0.74
1kf3	0.7 → 0.7	45 → 40	0.72 → 0.76
6rhn	0.9 → 0.9	52 → 46	0.73 → 0.77
1zs0	1.2 → 1.1	55 → 44	0.59 → 0.69
3ebx	1.5 → 1.3	64 → 50	0.50 → 0.64
1na7	1.7 → 1.6	56 → 54	0.60 → 0.62
1a6m	1.7 → 1.7	58 → 45	0.42 → 0.56
1e3u	1.6 → 1.4	55 → 41	0.57 → 0.69
2sar	1.7 → 1.6	54 → 49	0.54 → 0.58
1pm2	1.8 → 1.6	57 → 47	0.49 → 0.60
1i9a	1.9 → 1.8	59 → 52	0.56 → 0.64
1kqw	2.1 → 2.1	56 → 51	0.52 → 0.56
1lys	2.1 → 2.0	66 → 63	0.40 → 0.43
6ebx	3.6 → 3.4	64 → 60	0.29 → 0.34
2iff	4.7 → 4.5	75 → 75	0.21 → 0.23
1cgn	3.2 → 3.0	59 → 54	0.37 → 0.45
1bxo	4.3 → 4.1	47 → 43	0.37 → 0.42

the cycle $(n + 1)$ are combined with the phases obtained at the cycle n; the phases of the high-weight reflections belonging to subsets B and C are not changed (they are expected to have good initial phases). At the end of this step, the boundary condition (see Appendix *A*) is applied to the reflections for which $m|F| < D|F_p|$ and the following final phase values are assigned:

$$\langle\varphi_q\rangle = (\varphi_p + \pi) + \text{sign} \times \text{rand} \times |\langle\varphi_q\rangle - (\varphi_p + \pi)|_{\text{MAX}}. \quad (11)$$

sign is the direction of variation determined during the preceding cycles, rand is a random number generated between 0 and 1, and $|\langle\varphi_q\rangle - (\varphi_p + \pi)|_{\text{MAX}}$ is the largest deviation compatible with the boundary condition [see equation (17)]. The rationale is to force the phase values of the considered reflections to move away from their initial values, trusting in the direction defined by the DEDM cycles. The symmetry-restricted phases are excluded from the above procedure.

Step 3: Modulus assignment. As soon as new phase values are assigned to $\langle F_q\rangle$, new moduli are calculated according to Appendix *A*. The new moduli are combined with the previous ones with a coefficient of 0.8 for the old values.

Step 4: Phase and modulus assignment. Further cycles of DEDM are carried out, where the moduli are fixed to their new values and all the phases are allowed to vary with respect to the phase-assignment step. At the end the modulus values are updated to restore the geometrical correspondence between phases and moduli. As in the phase-assignment step, the difference Fourier map is modified by selecting the very positive and negative parts, and the new phase values are combined with the previous ones. If $m|F| < D|F_p|$, the phase values that do not satisfy condition (14) are corrected after each cycle as described in Appendix *A*.

6. Applications of the DEDM procedure

The DEDM procedure has been applied to the test cases quoted in Table 1; the outcome is summarized in Table 2, where, for each of the parameters $CrRES_q$ and MPE_q, two values are reported, calculated before and after the DEDM application, respectively. The last column of Table 2 contains two values of the correlation factor ($CORR_q$), calculated between the ideal difference Fourier map and the difference maps available before and after the application of DEDM, respectively.

As a general trend, we notice that, after the application of DEDM, both the parameters $CrRES_q$ and MPE_q diminish and, correspondingly, the $CORR_q$ values increase. In particular we note the following:

(i) The $CrRES_q$ values, both before and after DEDM, are anti-correlated with MPE and their improvement is marginal (*i.e.* it seems easier to improve phases than moduli).

(ii) Before DEDM, MPE_q is weakly correlated with MPE; their correlation increases after DEDM (owing to the fact that phase improvement is larger for better starting models).

(iii) In various cases $CORR_q$ significantly improves after DEDM. As expected, the most resistant cases are those for which the model structure is quite poor (see the last lines in Table 2).

To check which phases benefit more by DEDM, for the structure 2sar we plot in Fig. 4, *versus* $m|F| - D|F_p|$, the final MPE_q values, before and after DEDM. As a result of DEDM, the $\langle\varphi_q\rangle$ values of the reflections lying in regions A and B have been significantly improved (*i.e.* by levering on the phased reflections of the set C), thus reducing the asymmetry of the distribution. Similar plots, not shown for brevity, have been obtained for the other test structures.

The crucial point is now whether the difference Fourier maps obtained at the end of DEDM actually contain features of the protein structure that were not present in the conventional difference Fourier maps. For this purpose we selected some test structures (see Table 3) for which such analysis leads to non-ambiguous results; we used the *CCP4* package (Collaborative Computational Project, Number 4, 1994) to calculate the correlation coefficient, residue by residue, between the various difference Fourier maps $\langle\rho_q\rangle$ and the map built from the published coordinates ρ. The selected test structures were the following:

(*a*) 2sar, 1lys and 6ebx. They contain two symmetry-independent monomers, one of which is superimposed on the molecular replacement solution (denoted monomer *A*) while the second is to be located (monomer *B*). In Table 3 the average values of the correlation coefficients calculated for the main-chain atoms of the two monomers are reported; the columns Initial, DEDM and Ideal refer to the difference Fourier maps obtained, respectively, before and after the application of DEDM, and to the ideal difference Fourier map ρ_q. In all cases, the correlation value for monomer *B* increases when DEDM is applied. Although far from the Ideal value, the larger DEDM values imply that new features are present in the final $\langle\rho_q\rangle$ in a region far away from the model structure

Table 3
Average correlation coefficients, calculated residue-by-residue, between the electron-density map calculated from the published coordinates and various types of difference Fourier maps.

The letters *A* and *B* in the first column refer to monomer *A* (the one on which the model structure superimposes after the molecular replacement step) and monomer *B*, respectively. For the structure 3ebx, holes refers to residues not matched with any residue of the model and Gly refers to residues matched with non-docked residues of the model. The headings Initial, DEDM and Ideal refer to the difference Fourier map before and after the application of DEDM, and to the ideal difference Fourier map, respectively.

PDB	Initial	DEDM	Ideal
2sar *A*	−0.10	−0.18	−0.16
2sar *B*	0.56	0.58	0.89
1lys *A*	0.44	0.43	0.37
1lys *B*	0.48	0.49	0.87
6ebx *A*	0.29	0.24	0.42
6ebx *B*	0.36	0.37	0.86
1a6m	0.79	0.88	0.96
1kf3	0.52	0.65	0.89
enhexa	0.66	0.70	0.90
3ebx (holes)	0.72	0.81	0.89
3ebx (Gly)	0.70	0.76	0.90

which were not present in the initial $\langle\rho_q\rangle$. As counterpart, the correlation values for monomer *A* are decreased when DEDM is applied; for 2sar and 1lys the corresponding Ideal values are approached. This means that DEDM is able to reduce the strong bias constituted by the model used in the molecular replacement step. For 6ebx, which has the worst initial model of the three (see its MPE in Table 1), the method is not able to recover the differences between model and target structures in the region of monomer *A*.

The residue-by-residue correlation calculated for the side chain (not reported) is less influenced by DEDM, although the corresponding DEDM values, for the three test cases, are still increased with respect to the Initial values.

Fig. 5 is a visual counterpart of the above conclusions, obtained using the program *COOT* (Emsley & Cowtan, 2004); for the structural fragment of 2sar comprising residues 31–33 of monomer *B* we report the Initial and the DEDM $\langle\rho_q\rangle$ maps, respectively, in blue and red. It can be noted that the DEDM map describes the atomic positions of residue 32 (Gln) and its peptide bond with residue 33 (Asp) much better than the initial map.

(*b*) 1a6m and 1kf3. The first structure includes a protoporphyrin IX containing iron ($C_{34}H_{32}FeN_4O_4$), an oxygen molecule and two sulfate ions (SO_4) in the asymmetric unit. These ligands were not present in the model used for molecular replacement; therefore a study of the correlation coefficients between $\langle\rho_q\rangle$ and ρ calculated for such atoms will indicate the quality of the maps. The increase of the average value of the correlation coefficient, reported in Table 3 (from 0.79 to 0.88), indicates that the DEDM difference Fourier map contains much more detail with respect to the Initial difference Fourier map and approaches the Ideal one. Similar results, calculated for a sulfate ion present as ligand, are obtained for 1kf3 (from 0.52 to 0.65).

(*c*) enhexa. The target protein shows a six-residue loop which does not fit the model positioned *via* molecular replacement. The average values of the residue-by-residue correlation coefficients calculated along this loop, reported in Table 3, indicate an improvement in the DEDM difference Fourier map with respect to the initial one.

(*d*) 3ebx. The initial model was built by *ARP*/*wARP*: it contains holes (for a total of 12 residues) and non-docked residues (ten residues, interpreted as glycines). To assess the improvements in the difference Fourier map, we calculated the main-chain residue-by-residue correlation coefficient in the regions not covered by the model and the side-chain residue-by-residue correlation coefficient in the regions of non-docked residues. Their average values are reported in Table 3, denoted, respectively, by 'holes' and 'Gly'; in the first case the average correlation increases from 0.72 to 0.81, and in the second case from 0.70 to 0.76. In Fig. 6 the improvements in the difference Fourier map may be appreciated. From right to left, the first four residues of the sequence (Arg, Ile, Cys and Phe) are shown. This terminal region was not covered by the model and cannot be identified by the initial difference Fourier map (in blue). The difference Fourier map after DEDM (in red), instead, is much more continuous and fits well the side chains of the four residues.

7. The EDM–DEDM procedure

The above results suggest a straightforward strategy for the use of the DEDM approach: just replacing the standard difference Fourier synthesis in automatic model building or refinement procedures. An alternative strategy, based on the

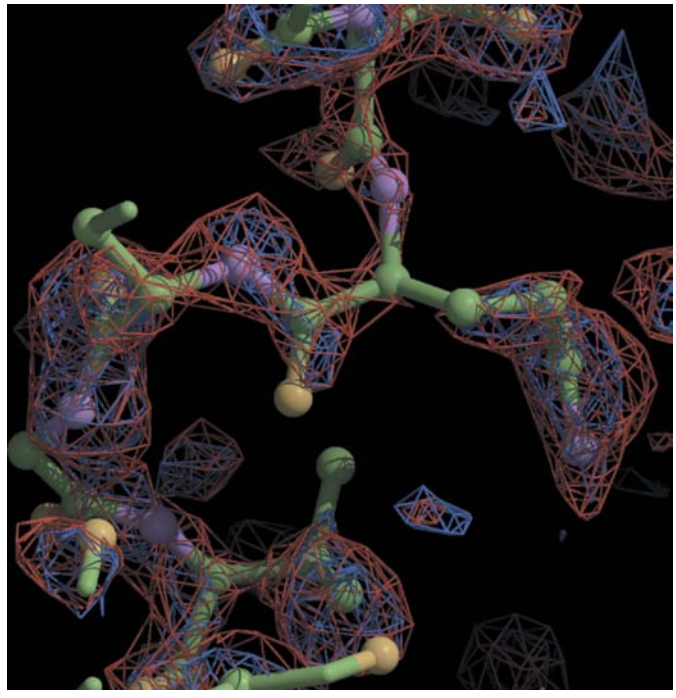

Figure 5
2sar: difference electron-density maps calculated before (blue) and after (red) the application of the DEDM procedure, superimposed on the structural fragment corresponding to residues 31–33 (Ser, Gln and Asp) of monomer *B*.

expected complementarity between EDM and DEDM, may be applied. Let us suppose the following:

(i) The standard EDM procedure ends with the electron-density map ρ_p (the best map available *via* such procedure): $|F_p|\exp(i\varphi_p)$ is the generic corresponding structure factor.

(ii) The application of DEDM to ρ_p leads to $\langle\rho_q\rangle$, and to the corresponding structure factors $\langle F_q\rangle = \langle|F_q|\exp(i\varphi_q)\rangle$, with weight w.

(iii) The overall structure factors

$$|F'|\exp(i\varphi') = D|F_p|\exp(i\varphi_p) + w\langle F_q\rangle \qquad (12)$$

are calculated for all the observed reflections.

Since the phases $\langle\varphi_q\rangle$, obtained at the end of step (ii), are usually better than those calculated by the standard difference electron density, it may be supposed that the new phases φ' will be more accurate than those available at the end of step (i) (*i.e.* the φ_p values). This expectation led to the suggestion to iterate the EDM–DEDM cycles, according to the following scheme:

$$\varphi_p \overset{\text{DEDM}}{\longrightarrow} \varphi' \overset{\text{EDM}}{\longrightarrow} \varphi_p \overset{\text{DEDM}}{\longrightarrow} \varphi'.$$

During the EDM cycles the phases of reflections having high values of $m|F|$ and $D|F_p|$ are allowed to vary (they are badly estimated by the DEDM procedure, since they have $|F_q| \simeq 0$), while the phases of reflections having high values of $D|F_p|$ and low values of $m|F|$ are kept fixed (they are well estimated by the DEDM procedure, since they have $|F_q| \ll 0$). The mean phase errors $\langle|\varphi' - \varphi|\rangle$ obtained after the first and the second application of the DEDM cycles, together with the correlation factors of the corresponding Fourier maps, are quoted in Table 4. For most of the test structures the procedure significantly improves the quality of the phase set; the most resistant are the structures for which a quite poor model was available. In some cases, such as 1zso, 1a6m, 1e3u and 1pm2, the improvement is relevant (42 → 36°, 48 → 38°, 51 → 40° and 56 → 49°, respectively). We verified that such improvements cannot be obtained by using EDM and DEDM separately, and that a further iteration of the procedure leads to better results only for a few test structures.

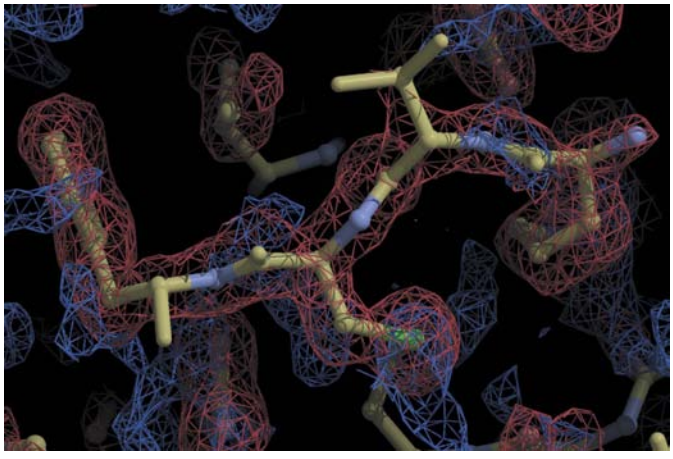

Figure 6
3ebx: difference electron-density maps calculated before (blue) and after (red) the application of the DEDM procedure, superimposed on the structural fragment corresponding to residues 1–4 (Arg, Ile, Cys and Phe).

Table 4
Results of the EDM–DEDM procedure.

The first MPE value corresponds to the best electron-density map available after the application of the standard EDM procedure [*i.e.* it is the mean phase error (in degrees) of the phases φ_p with respect to the published values φ]. The second and the third MPE values correspond to the mean phase error of the phases φ' with respect to the published φ values, after the first and the second application of the DEDM cycles, respectively. CORR are the correlation factors between the electron-density map calculated by using measured moduli and phases φ_p (for the first value) and φ' (for the second and third values), and the map calculated *via* measured moduli and published phases φ.

PDB	MPE	CORR
enhexa	22 → 21 → 21	0.92 → 0.93 → 0.94
1kf3	26 → 25 → 24	0.96 → 0.96 → 0.96
6rhn	32 → 31 → 31	0.91 → 0.92 → 0.92
1zs0	42 → 40 → 36	0.86 → 0.89 → 0.90
3ebx	45 → 42 → 40	0.84 → 0.88 → 0.89
1na7	48 → 47 → 47	0.78 → 0.78 → 0.79
1a6m	48 → 44 → 38	0.82 → 0.86 → 0.89
1e3u	51 → 47 → 40	0.77 → 0.81 → 0.85
2sar	53 → 52 → 50	0.72 → 0.73 → 0.75
1pm2	56 → 53 → 49	0.70 → 0.73 → 0.76
1i9a	58 → 57 → 56	0.68 → 0.71 → 0.72
1kqw	60 → 59 → 56	0.69 → 0.70 → 0.73
1lys	64 → 63 → 61	0.65 → 0.66 → 0.67
6ebx	72 → 71 → 70	0.49 → 0.51 → 0.49
2iff	72 → 72 → 72	0.38 → 0.38 → 0.36
1cgn	73 → 73 → 73	0.51 → 0.53 → 0.53
1bxo	74 → 74 → 70	0.58 → 0.59 → 0.63

The 1e3u electron-density maps, the first corresponding to the initial *ARP*/*wARP* model and the second to the final map available at the end of the EDM–DEDM cycles (observed moduli and phases φ'), are superimposed on the published model in Fig. 7, respectively, in blue and red. Moving from right to left, it can be noted that the residues 205 (Lys) and 206 (Thr) of chain *A* may be identified only by the EDM–DEDM electron-density map. The improved quality of this map has also been automatically recognized by *ARP*/*wARP*; its outcome was processed by the SSM function (Krissinel & Henrick, 2004) within *COOT* to compare the *ARP*/*wARP* model with the published one. The corresponding routines align the two structures by matching graphs built on the protein's secondary-structure elements, followed by an interactive three-dimensional alignment of protein backbone C atoms. The results may be summarized as follows: in the initial model 353 residues were aligned, with an r.m.s. deviation of 0.38 Å; 430 residues were aligned in the model obtained from the EDM–DEDM map, with an r.m.s. deviation of 0.32 Å.

A similar result is obtained for 1pm2: 207 residues were aligned in the initial model, with an r.m.s. deviation of 0.46 Å, and 299 residues in the model obtained from the EDM–DEDM map, with an r.m.s. deviation of 0.25 Å.

8. Conclusions

The statistical approach applied to the $(F_o - F_c)$ Fourier synthesis has revealed new interesting features, related to the

variance σ_q of the expected value of F_q. In particular, it has been established that the case with $|F|$ large and $|F_p|$ small is quite different from the case with $|F|$ small and $|F_p|$ large, even if the value of $|F_q| = ||m|F| - D|F_p||$ is the same. In addition, it has been shown that the phases φ_q of the reflections with large σ_q values cannot be accurately estimated, while their φ_p values are reliable estimates of the phase φ. *Vice versa*, reflections with small σ_q values provide reliable estimates of the phase φ_q, while their φ_p phases are rough approximations of the φ values. The above results suggested to us a procedure (DEDM) aimed at improving the φ_q estimates provided by the conventional $(F_o - F_c)$ Fourier synthesis by recursive cycles based on a difference electron-density modification technique. The method levers on the phases of the reflections with small σ_q values to improve the φ_q estimates with larger σ_q values.

The DEDM procedure has to face additional difficulties with respect to those tackled by EDM techniques; indeed, the coefficients of the F_o Fourier syntheses are observed quantities, while a DEDM procedure is obliged to use estimated values. Nevertheless, the proposed DEDM procedure succeeded in yielding better modulus and phase estimates when applied to cases routinely encountered in crystallography, in particular, in the refinement of molecular replacement models or of structural fragments built by automatic procedures from experimental electron-density maps.

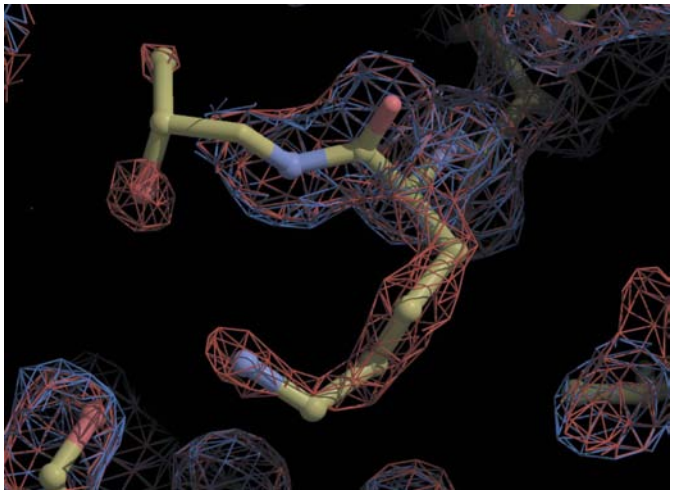

Figure 7
1e3u: electron-density maps calculated for the partial structure (blue) and after the application of the EDM–DEDM procedure (red), superimposed on the structural fragment corresponding to residues 205 (Lys) and 206 (Thr) of chain *A*.

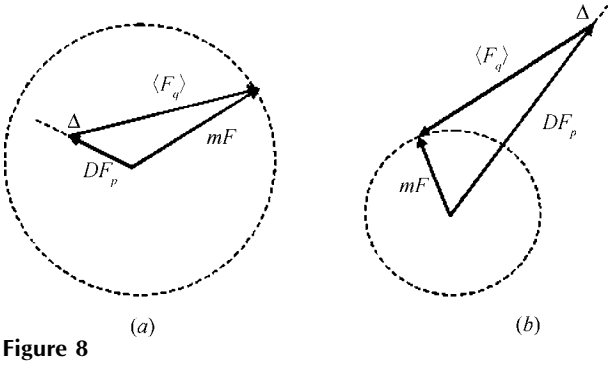

Figure 8
Schematic view of the vector equation $mF = DF_p + \langle F_q \rangle$ when (*a*) $m|F|$ is larger than $D|F_p|$ and (*b*) $m|F|$ is smaller than $D|F_p|$.

The expected complementarity between EDM and DEDM procedures suggested to us a next goal: the integration of the DEDM procedure with additional cycles of EDM. The results were very encouraging.

The procedure was applied to various cases, including proteins constituted by two monomers, one of which is correctly placed by molecular replacement. The quality of the final electron-density maps significantly improved, so allowing us to recover parts of the target structure not covered by the partial structure. The procedure seems to have a reserve of power and may be considered a first step for further developments.

The application of DEDM to macromolecular crystallography opens the question of how to handle the free *R* reflections. While they may be used during the EDM cycles that precede or follow DEDM, their application is doubtful within the DEDM cycles; indeed, no modulus constraint is used for calculating the difference Fourier map (contrary to what happens for the observed electron-density maps). As a consequence, in our tests the free *R* criterion is not well correlated with the variables MPE or CORR. Its use within DEDM cycles requires substantial modifications and a supplementary study.

APPENDIX *A*
Modulus estimation and phase estimation constraints in the DEDM procedure

Let us assume that $m|F|$, $D|F_p|$ and φ_p are known and that an estimate $\langle \varphi_q \rangle$ of φ_q is available after application of some DEDM cycles. Then $\langle |F_q| \rangle$ may be calculated *via* the Carnot theorem (see the triangles sketched in Fig. 8):

$$\langle |F_q| \rangle = -D|F_p| \cos(\Delta) \pm \left[m^2|F|^2 - D^2|F_p|^2 \sin^2(\Delta) \right]^{1/2}, \tag{13}$$

where $\Delta = |\varphi_p - \langle \varphi_q \rangle|$, provided

$$\sin^2(\Delta) \leq m^2|F|^2 / D^2|F_p|^2. \tag{14}$$

If $m|F| \geq D|F_p|$, condition (14) is always verified and one of the two roots in equation (13) is always negative; then the unique acceptable solution is

$$\langle |F_q| \rangle = -D|F_p| \cos(\Delta) + \left[m^2|F|^2 - D^2|F_p|^2 \sin^2(\Delta) \right]^{1/2}, \tag{15}$$

with no limitation on Δ. The geometrical interpretation of this case is sketched in Fig. 8(*a*). A circumference of radius $m|F|$ is drawn; since the orientation of F is not known, the triangle can be closed for any value of Δ.

Conversely, if $m|F| < D|F_p|$, condition (14) is verified for a limited range of Δ values [*i.e.* in Fig. 8(*b*) the triangle may be closed only if the vector $\langle F_q \rangle$ intersects the circle]. In this case

both the roots in equation (13) are positive; choosing the lower root is a way to minimize the error on $\langle|F_q|\rangle$.

Suppose now that $m|F| < D|F_p|$, but condition (14) is not satisfied; then $\langle\varphi_q\rangle$ is probably wrong. This constraint is used in step 4 to correct the phase values obtained after each cycle. Intuitively, one would assign the new $\langle\varphi_q\rangle$ value so that the tangent condition

$$\sin(\Delta) = m|F|/D|F_p| \qquad (16)$$

is verified, but in practice our tests lead to better results if the initial value $\langle\varphi_q\rangle = \varphi_p + \pi$ is restored. This indicates that in this case the side on which $\langle F_q\rangle$ is located with respect to F_p, suggested by the cycles, is also wrong.

In step 2 condition (16) is used to bound the phase estimation. If $m|F| > D|F_p|$, (16) may be rewritten as

$$|\langle\varphi_q\rangle - (\varphi_p + \pi)|_{\rm MAX} = \arcsin\big(m|F|/D|F_p|\big), \qquad (17)$$

which gives the limiting values of $\langle\varphi_q\rangle$ given its initial value $(\varphi_p + \pi)$.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G., Siliqi, D. & Spagna, R. (2007). *J. Appl. Cryst.* **40**, 609–613.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2006). *J. Appl. Cryst.* **39**, 185–193.

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Moustiakimov, M. & Siliqi, D. (2005). *Acta Cryst.* A**61**, 343–349.

Cochran, W. (1951). *Acta Cryst.* **4**, 408–411.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Giacovazzo, C., Monaco, H. L., Artioli, G., Viterbo, D., Ferrarsi, G., Gilli, G., Canotti, G. & Catti, M. (2002). *Fundamentals of Crystallography*, 2nd ed., edited by C. Giacovazzo, pp. 696–698. Oxford University Press.

Henderson, R. & Moffat, J. K. (1971). *Acta Cryst.* B**27**, 1414–1420.

Honnappa, S. (2008). Personal communication.

Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

Main, P. (1979). *Acta Cryst.* A**35**, 779–785.

Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Ramachandran, G. N. & Raman, S. (1959). *Acta Cryst.* **12**, 957–964.

Ramachandran, G. N. & Srinivasan, R. (1970). *Fourier Methods in Crystallography*. London: Wiley-Interscience.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Srinivasan, R. (1961). *Acta Cryst.* **14**, 607–611.

Ursby, T. & Bourgeois, D. (1997). *Acta Cryst.* A**53**, 564–575.

Woolfson, M. M. (1956). *Acta Cryst.* **9**, 804–810.