Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Advances in the EDM–DEDM procedure

Rocco Caliandro, Benedetta Carrozzini, Giovanni Luca Cascarano, Carmelo Giacovazzo,* Anna Maria Mazzone and Dritan Siliqi

Institute of Crystallography, CNR, Via G. Amendola 122/O, 70126 Bari, Italy

Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

Received 28 October 2008
Accepted 23 December 2008

The DEDM (difference electron-density modification) algorithm has been described in a recent paper [Caliandro *et al.* (2008), *Acta Cryst.* A**64**, 519–528]: it breaks down the collinearity between model structure phases and difference structure phase estimates. The new difference electron-density produced by DEDM, summed to the calculated Fourier maps, is expected to provide a representation of the full structure that is more accurate than that obtained by the observed Fourier synthesis. In the same paper, the DEDM algorithm was combined with the EDM (electron-density modification) approach to give the EDM–DEDM procedure which, when applied to practical molecular-replacement cases, was able to improve the model structures. In this paper, it is shown that EDM–DEDM suffers from some critical points that did not allow cyclical application of the procedure. These points are identified and modifications are made to allow iteration of the procedure. The applications indicate that EDM–DEDM may become a fundamental tool in protein crystallography.

1. Notation

The following notation has been used in this article.
ρ, ρ_p: true electron density and model density, respectively.
$\rho_q = \rho - \rho_p$: the ideal difference Fourier synthesis. Summed to ρ_p, it exactly provides ρ, no matter the quality of ρ_p.
F, F_p, F_q: the structure factors of ρ, ρ_p and ρ_q, respectively. φ, φ_p and φ_q are their phases.
E, E_p, E_q: the normalized structure factors of ρ, ρ_p and ρ_q, respectively.
N: number of atoms in the unit cell for the full structure.
p: number of atoms in the unit cell for the structural model. Usually, $p \leq N$, but it may also be that $p > N$.
ε: statistical Wilson coefficient (corrects for expected intensities in reciprocal-lattice zones).
$\Sigma_N = \varepsilon \sum_{j=1}^{N} f_j^2$.
$\Sigma_p = \varepsilon \sum_{j=1}^{p} g_j^2$.
$\Sigma_q = \varepsilon \sum_{j=p+1}^{N} f_j^2$ if $p < N$.
$D = \langle \cos(2\pi \mathbf{h} \Delta \mathbf{r}) \rangle$, where $\Delta\mathbf{r}$ is the error in the model structure atomic positions; the average involves the p atoms and it is performed per resolution shell (Luzzati, 1952).
$\sigma_A = D(\Sigma_p/\Sigma_N)^{1/2}$.
$\langle \mu \rangle$: expected measurement error on $|F|$.
$e = 1 + (\langle |\mu|^2 \rangle / \Sigma_N)$.
$I_i(x)$: modified Bessel function of order i.
$m = \langle \cos(\varphi - \varphi_p) \rangle = I_1(x)/I_0(x)$, where $x = 2\sigma_A |E||E_p|/(e - \sigma_A^2)$.

2. Introduction

Weighted observed Fourier syntheses [with coefficients $m|F|\exp(i\varphi_p)$ (Sim, 1959); referred to hereafter as the F_o-synthesis], weighted difference Fourier syntheses {with coefficients $\|m|F| - D|F_p\|\exp[i(\varphi_p + s\pi)]$, where $s = 1$ or 0 according to whether the sign of $m|F| - D|F_p|$ is positive or negative (Henderson & Moffat, 1971; Main, 1979; Read, 1986; Ursby & Bourgeois, 1997); referred to hereafter as the ΔF-synthesis} and their combination syntheses [with $(2m|F| - |F_p|)\exp(i\varphi_p)$ coefficients] are fundamental tools in structural crystallography. Their use substantially aims at recovering $\rho(\mathbf{r})$ from $\rho_p(\mathbf{r})$. All the three syntheses suffer from an unavoidable bias generated by the prior information. Indeed,

(i) the phases of the F_o- synthesis may only assume φ_p values: their efficiency relies on the supplemental information provided by the observed $|F|$ moduli;

(ii) in the ΔF- synthesis the phase values may be φ_p or $\varphi_p + \pi$. Hence, φ and φ_p are assumed to be collinear: in addition, $\|m|F| - D|F_p\|$ may be a rough approximation of $|F_q|$. Similar considerations hold for syntheses which are combinations of the previous two.

If $\rho_p(\mathbf{r})$ roughly approximates $\rho(\mathbf{r})$, the three syntheses are unable to offer a realistic representation of the corresponding ideal electron densities. For the F_o- synthesis a powerful technique is widely used known as the electron-density modification (EDM) procedure (Cowtan, 1994, 1999; Abrahams, 1997; Abrahams & Leslie, 1996; Zhang *et al.*, 2001, Refaat & Woolfson, 1993, Giacovazzo & Siliqi, 1997): it is a cyclic process based on sensible modifications of the F_o or $2F_o - F_c$ electron-density maps with the aim of improving the phase estimates by capturing the desired features of the map (*e.g.* its positivity, the molecular envelopes, its statistical description *via* density histograms *etc.*).

Recently, the ΔF- synthesis has been revisited (Caliandro *et al.*, 2008). The main achievements may be summarized as follows:

(i) the variance σ_q^2 of F_q was derived, *i.e.*

$$\sigma_q^2 = (1 - m^2)|F|^2 + \langle|\mu|^2\rangle, \tag{1}$$

where $\langle|\mu|^2\rangle$ is the expected measurement error on $|F|^2$. According to (1), the accuracy of the φ_q estimates, $\langle\varphi_q\rangle$, is expected to be inversely correlated with σ_q;

(ii) reflections with the same value of $(|m|F| - D|F_p|)$ have different values of σ_q (according to whether $|F|$ is large and $|F_p|$ is small or *vice versa*). $\langle\varphi_q\rangle$ estimates corresponding to strong negative values of $(|m|F| - D|F_p|)$ are the most accurate;

(iii) reflections with large (small) values of $|F_p|$ and small (large) values of $|F|$ may be quite useful for defining the ΔF-synthesis but of limited use for the F_o- synthesis (*i.e.* their phase φ_p is a poor approximation of φ).

The above results are the basis of the DEDM (difference electron-density modification) algorithm based on the modification of the difference electron density (Caliandro *et al.*, 2008). It works in a way that is substantially independent of EDM and hinges on the phases of reflections with small σ_q values (which are left practically unvaried during the process). In practice, a combined EDM–DEDM procedure was applied according to the scheme

$$\varphi_p \xrightarrow{\text{DEDM}} \varphi' \xrightarrow{\text{EDM}} \varphi_p' \xrightarrow{\text{DEDM}} \varphi'', \tag{2}$$

which may be described in four steps.

Step 1: A model structure [the model structure (almost) correctly placed by molecular replacement (MR) or built by *ARP*/*wARP* (Perrakis *et al.*, 1999)] is used as ρ_p: the corresponding phases are denoted in (2) by φ_p.

Step 2: The DEDM algorithm is applied, which is constituted of the following substeps.

(i) The ΔF- synthesis is calculated and modified: the very positive and very negative parts of the map are preserved or squared and the remainder is set to zero.

(ii) Fourier inversion of the modified ΔF- synthesis generates phase shifts $\Delta\varphi_q$: new estimates $\langle\Delta\varphi_q\rangle = (\varphi_q + s\pi) + \Delta\varphi_q$ are obtained.

(iii) For each reflection, the value $\langle|F_q|\rangle$ is calculated by applying the Carnot theorem to the triangle defined by mF, DF_p and $\Delta = |\varphi_p - \langle\varphi_q\rangle|$ according to

$$\langle|F_q|\rangle = -D|F_p|\cos(\Delta) \pm [m^2|F|^2 - D^2|F_p|^2\sin^2(\Delta)]^{1/2}. \tag{3}$$

Rules for resolving the sign ambiguity are given in the appendix of Caliandro *et al.* (2008).

(iv) New estimates for the full structure factors are obtained *via* the relation

$$|F'|\exp(i\varphi') = D|F_p|\exp(i\varphi_p) + w\langle F_q\rangle, \tag{4}$$

where w is a suitable weight proportional to $1/\sigma_q$. The new estimates are used to compute a new F_o- synthesis ($\langle\rho\rangle$), which is expected to be a more accurate representation of ρ.

Substeps (i), (ii) and (iii) are repeated several times during the procedure and constitute the DEDM cycles. Substeps (i) and (ii) are coupled in the phase-assignment cycles, while substep (iii) forms the modulus-assignment step, which is executed after a fixed number of phase-assignment cycles. The map modification in substep (i) varies as a function of the cycle number. Initially, 10% of the highest intensity pixels and 10% of the lowest intensity pixels are selected and their intensity is squared. In the phase-assignment cycles following the first modulus-assignment step, the squaring is no longer applied and only the top positive part of the map is selected, with a percentage of pixels ranging from 3.5% to 5%. In addition, the $\langle\varphi_q\rangle$ estimates obtained in substep (ii) are always combined with those obtained in the previous cycles.

Step 3: EDM is applied to $\langle\rho\rangle$, from which a new model structure ρ_p is chosen. Its corresponding phases are denoted in (2) by φ_p'.

Step 4: DEDM is again applied to improve the phases, starting from the new model structure.

Although it has successfully been applied to several cases, the proposed EDM–DEDM procedure suffers from several critical points, which are revised in the present paper. The applications suggest the great potential of this new crystallographic tool.

3. The new EDM–DEDM algorithm

Let us provide a more solid mathematical basis for the DEDM algorithm. By definition,

$$F_q = F - F_p = |F|\exp(i\varphi) - |F_p|\exp(i\varphi_p).$$

Denoting

$$\Delta\varphi = \varphi - \varphi_p \text{ and } \Delta F = |F| - |F_p|$$

we have

$$\begin{aligned} F_q &= |F|\exp[i(\varphi_p + \Delta\varphi)] - (|F| - \Delta F)\exp(i\varphi_p) \\ &= |F|\exp(i\varphi_p)\exp\left(i\frac{\Delta\varphi}{2}\right)\left[\exp\left(i\frac{\Delta\varphi}{2}\right) - \exp\left(-i\frac{\Delta\varphi}{2}\right)\right] \\ &\quad + \Delta F[\exp(i\varphi_p)] \\ &= \Delta F\exp[i(\varphi_p)] + |F|2\sin\left(\frac{\Delta\varphi}{2}\right)\exp\left[i\left(\varphi_p + \frac{\Delta\varphi}{2} + \frac{\pi}{2}\right)\right]. \end{aligned}$$

Then,

$$\begin{aligned} \rho_q(\mathbf{r}) &= \frac{1}{V}\sum \Delta F\exp(i\varphi_p)\exp(-2\pi i\mathbf{hr}) \\ &\quad + \frac{1}{V}\sum |F|2\sin\left(\frac{\Delta\varphi}{2}\right)\exp\left[i\left(\varphi_p + \frac{\Delta\varphi}{2} + \frac{\pi}{2}\right)\right]\exp(-2\pi i\mathbf{hr}). \end{aligned} \tag{5}$$

(5) suggests that $\rho_q(\mathbf{r})$ is the sum of two Fourier syntheses:

$$\rho_q(\mathbf{r}) = \Delta\rho(\mathbf{r}) + \rho_{\rm comp}. \tag{6}$$

The first term on the right-hand side of (6) coincides with the ΔF- synthesis. The complementary contribution, $\rho_{\rm comp}$, with coefficients $|F|2\sin(\Delta\varphi/2)$, shows phase values that are shifted by $(\Delta\varphi/2) + (\pi/2)$ with respect to the corresponding coefficients in the first synthesis. $\rho_{\rm comp}$ is usually non-negligible with respect to $\Delta\rho$, particularly when ρ_p is a poor approximation of ρ. In order to obtain ρ_q from $\Delta\rho$, the correct identification of the shifts $\Delta\varphi$ is necessary: this is the first aim of the DEDM procedure.

(5) provides some further information that allows optimization of the algorithm. Indeed,

(i) reflections with small values of $|F|$ and $|F_p|$ may be neglected;

(ii) reflections with small $|F|$ and large $|F_p|$ values do not contribute to $\rho_{\rm comp}$: accordingly, DEDM does not try to identify the corresponding $\Delta\varphi$ values (we assume that $\Delta\varphi = 0$ for them). It is also useful to underline that such reflections strongly contribute to $\Delta\varphi(\mathbf{r})$ (they have the smallest variance);

(iii) reflections with large values of $|F|$ are most important, particularly those for which $\Delta\varphi$ is large.

The conclusions outlined above may easily be extended to ΔF- syntheses using Read coefficients [one only needs to replace $|F|$ by $m|F|$ in (5)]. In this case the calculations require, for each reflection, the previous estimate of the coefficients m and D. Such information, which is also necessary for the application of (3), is only available if the scattering power of the model structure is perfectly known. Indeed, only in this case may the value of $\sigma_{\rm A}$ available from suitable statistical calculations be decomposed into $(\Sigma_p/\Sigma_N)^{1/2}$ and D. When the starting point is an electron-density map rather than a molecular model, the decomposition of $\sigma_{\rm A}$ is ambiguous. The algorithm falls into this situation at Step 4 and this is the main reason that hindered Caliandro *et al.* (2008) in making it cyclic.

To overcome this problem, we decided to perform all calculations by using normalized structure-factor moduli in all the Fourier syntheses. Since

$$mF - DF_p = (mE - \sigma_{\rm A}E_p)\Sigma_N^{1/2}, \tag{7}$$

using E rather than F has the advantage that $\sigma_{\rm A}$ is directly computable from the relation (Caliandro *et al.*, 2005)

$$\sigma_{\rm A} = \langle |E|^2|E_p|^2\rangle - e.$$

Moreover, since $mF - DF_p = \langle F_q\rangle$, from (7) we obtain

$$(mE - \sigma_{\rm A}E_p) = \langle E_q\rangle\left(\frac{\Sigma_q}{\Sigma_N}\right)^{1/2}, \tag{8}$$

and (3) may be replaced by

$$\begin{aligned} \langle |E_q|\rangle = \left(\frac{\Sigma_N}{\Sigma_q}\right)^{1/2}\{&-\sigma_{\rm A}|E_p|\cos(\Delta) \\ &\pm [m^2|E|^2 - \sigma_{\rm A}^2|E_p|^2\sin^2(\Delta)]^{1/2}\}. \end{aligned} \tag{9}$$

However, the coefficient $(\Sigma_N/\Sigma_q)^{1/2}$ on the right-hand side of (8) and (9) is unknown. Luckily, such information is not necessary: indeed, when (4) is expressed in terms of E values, it is the product $\langle |E_q|\rangle(\Sigma_q/\Sigma_N)^{1/2}$ that is used,

$$\begin{aligned} |E'|\exp(i\varphi') &= D\left(\frac{\Sigma_p}{\Sigma_N}\right)^{1/2}|E_p|\exp(i\varphi_p) \\ &\quad + w\langle |E_q|\rangle\exp(i\varphi'_q)\left(\frac{\Sigma_q}{\Sigma_N}\right)^{1/2} \\ &= \sigma_{\rm A}E_p + w\langle E_q\rangle\left(\frac{\Sigma_q}{\Sigma_N}\right)^{1/2}. \end{aligned} \tag{10}$$

The above reformulation allowed us to replace the old EDM–DEDM scheme with the following one:

$$(\varphi_p)_i \overset{\rm DEDM}{\longrightarrow} (\varphi')_i \overset{\rm EDM}{\longrightarrow} (\varphi'_p)_i \quad \text{for } i = 1, \ldots, N.$$

The DEDM and EDM cycles are now part of a block which is iterated N times. At the end of the ith iteration (called the supercycle), the resulting phase set $(\varphi'_p)_i$ is expected to represent an improved partial model and is taken as input for a further iteration of the procedure, *i.e.* $(\varphi_p)_{i+1} = (\varphi'_p)_i$.

Further improvements with respect to the procedure described by Caliandro *et al.* (2008) are as follows.

(i) The EDM cycles have been revised according to the package *IL MILIONE* (Burla *et al.*, 2007). They are performed in two different ways depending on the data resolution limit. If this is high, the EDM procedure described by Giacovazzo & Siliqi (1997) is used: only a small fraction of the electron density is inverted to obtain new phases and a mask from molecular-envelope calculations is applied. Otherwise, EDM cycles are performed by using strategies such as solvent flattening (Leslie, 1987), histogram matching (Zhang & Main, 1990) and γ correction (Cowtan, 1999).

(ii) The molecular-envelope mask is also applied within the DEDM cycles. It is calculated in each supercycle from the map obtained by the phase set $(\varphi_p)_i$ and is applied to the Fourier difference map during the first DEDM cycles in order to constrain the $\langle\varphi_q\rangle$ phase assignments. In the final DEDM cycles this constraint is relaxed, since it would prevent the appearance of new features in the map.

(iii) In the EDM cycles, the current phases $(\varphi'_p)_i$ are combined with the phases $(\varphi')_i$ obtained at the end of the DEDM step. Two criteria are used. According to the first one, a larger weight is given to the phases $(\varphi'_p)_i$ when $m|E| - \sigma_A|E_p| \simeq 0$ (these phases are badly estimated by DEDM); the opposite occurs for reflections with $m|E| - \sigma_A|E_p| << 0$, which are well estimated by DEDM. The second criterion concerns the amount of the structural model that is missing: if the initial value of Σ_p/Σ_N is much lower than unity (a very incomplete initial model), the new phases $(\varphi'_p)_i$ are given a larger weight with respect to the old phases $(\varphi')_i$ in order to increase the chance of finding the remainder of the structure.

(iv) The quantities $\langle|E_q|\rangle(\Sigma_q/\Sigma_N)^{1/2}$ are submitted to a histogram-matching procedure in order to fit their experimental distribution to the Wilson distribution for acentric reflections (the fitting preserves the original experimental mean value).

Two critical points still exist in the new EDM–DEDM scheme. The first involves the connection between the EDM and DEDM sections, in which the new phases $(\varphi')_i$ are calculated using (10). Indeed, some ambiguity arises because the scale between $\langle|E_q|\rangle$ and $|E_p|$ is unknown and because the weight w is used. For this reason, we decided to combine the phases $(\varphi')_i$ determined using (10) with the phases calculated from direct space,

$$\rho = \rho_p + K\rho_q, \quad K = \begin{cases} 1.2 & \text{if } \rho_q > 0 \\ 0 & \text{if } \rho_q \leq 0 \end{cases}. \qquad (11)$$

The coefficient K allows enhancement of the Fourier difference contribution in regions which are not anticorrelated with ρ_p. The second critical point deals with the reformulation of the new initial model. At the end of the the ith supercycle an observed electron-density map is calculated with phases $(\varphi_p)_{i+1} = (\varphi'_p)_i$. This is quite different from what is performed in the first supercycle, in which a molecular model is used from MR or from model-building calculations. Starting a supercycle from an electron-density map introduces a certain amount of ambiguity that is inherent to the part of map to be inverted and to the type of Fourier coefficients (normalized or not) used to calculate the map. We will see in the applications that this hinders optimal recursivity of the procedure.

3.1. Figures of merit for the EDM–DEDM procedure

Let us now calculate the overall variance of ρ_{comp}: it is given by

$$\langle\rho^2_{comp}\rangle = \frac{1}{V^2}\sum_{\mathbf{h}}|F|^2\sin^2\Delta\varphi. \qquad (12)$$

Even if the $\Delta\varphi$ values are never perfectly known, approximate estimates of them are available *via* the DEDM algorithm. Therefore, for practical applications we can substitute $\Delta\varphi$ with $\Delta\varphi = \varphi' - \varphi_p$ in (12) and use $\langle\rho^2_{comp}\rangle$ as a figure of merit for the DEDM cycles. It is expected that the value of $\langle\rho^2_{comp}\rangle$ should increase from one cycle to the next as the current phases φ' deviate more and more from their initial estimates φ_p. At the same time, its value is expected to diminish with increasing DEDM supercycle number. In fact, if the procedure converges then $(\varphi_p)_{i+1}$ should approximate φ better than $(\varphi_p)_i$ and therefore the corresponding phase shifts $(\Delta\varphi')_{i+1}$ should be smaller than $(\Delta\varphi')_i$.

Another figure of merit that is sensitive to the supercycle order is CORR($|E|, |E'|$), the correlation between the observed and calculated structure-factor moduli. It is expected to increase with the supercycle order.

The following strategy is followed.

(i) The number N of supercycles is not fixed: the procedure ends if the relative difference between the current supercycle and the previous one is below 1% for CORR($|E|, |E'|$) or above 1% for $\langle\rho^2_{comp}\rangle$.

(ii) The number of DEDM cycles is variable: it increments until the current cycle has a relative difference of $\langle\rho^2_{comp}\rangle$ from the previous cycle of below 1%.

4. Applications

The test cases used to validate the new EDM–DEDM procedure are listed in Table 1. For each structure, the Protein Data Bank (PDB) code, data resolution limit (Res, in Å), data completeness (Complet, a percentage) and number of residues (NresT) are reported. The PDB data files of the structures were processed by *IL MILIONE* using different phasing strategies. Most of the structures were phased by MR using the program *REMO* (Caliandro *et al.*, 2006): in this case the columns Model, NresM and Id of Table 1 contain the PDB code and number of residues of the model structure and the sequence identity between the target and model structures, respectively. For the remaining structures, *ab initio* or single-wavelength anomalous diffraction (SAD) phasing was applied and the model structure was built using *ARP*/*wARP*: in this case the column NresM contains the number of docked residues; the number of nondocked residues, considered as glycines, is given in parentheses. In two difficult cases (1xyg and 1ujz), *ARP*/*wARP* was also used after MR phasing in order to improve the model structure located by MR. The test cases have been sorted on the basis of the initial value of the parameter CORR, *i.e.* the correlation factor between the electron-density map calculated using observed moduli and initial phases φ_p and the map calculated using observed moduli and published phases φ. CORR is reported in the last column of Table 1.

The results are summarized in Table 2, in which CORR and MPE are reported per supercycle: MPE is the mean phase error (in degrees) of the current phases (φ_p, φ' or φ'_p) with respect to the published phases φ. The values of both parameters follow the scheme $(\varphi_p)_1 \rightarrow (\varphi')_1/(\varphi')_2/\ldots/(\varphi')_N \rightarrow (\varphi'_p)_N$

Table 1
Test structures.

PDB is the PDB code of the protein structure (apart from the structure enhexa, which has not yet been deposited in the PDB), Res is the data resolution limit in Å, Complet is the data completeness (%), NresT is the number of residues of the target structure, Model is the PDB code of the model structure used in the molecular-replacement procedure or indicates that the structure was built by *ARP/wARP* starting from a map obtained by *ab initio* phasing (or by SAD phasing for 1i9a), NresM is the number of residues of the model structure, Id is the sequence identity (%) between the target and model structures (only calculated for molecular-replacement cases), CORR is the correlation factor between the electron-density map calculated using observed moduli and phases φ_p and the map calculated using observed moduli and published phases φ.

PDB	Res	Complet	NresT	Model	NresM	Id	CORR
1kf3	1.0	96	124	7rsa	124	97	0.95
enhexa†	1.2	100	130	1pa7	130	100	0.92
6rhn	2.2	90	115	4rhn	104	100	0.91
1zs0	1.6	95	163	1i76	163	100	0.86
3ebx	1.4	98	62	*ARP/wARP*	40 (+10 Gly)	—	0.84
1a6m	1.0	93	151	1mbc	153	100	0.83
1na7	2.4	99	329	1m2r	327	76	0.78
1e3u	1.7	99	970	*ARP/wARP*	520 (+492 Gly)	—	0.77
1s31	2.7	100	273	1c8z	265	96	0.76
2p0g	2.3	99	318	2oka	336	64	0.76
2sar	1.8	98	192	1ucl, chain *A*	96	98	0.72
1pm2	1.8	98	678	*ARP/wARP*	363 (+322 Gly)	—	0.69
1kqw	1.8	100	134	1opa	133	74	0.69
1i9a	2.4	96	364	*ARP/wARP*	165 (+78 Gly)	—	0.68
1lys	1.7	85	258	2ihl	129	96	0.65
1xyg	2.2	93	1380	*ARP/wARP*	117 (+388 Gly)	—	0.64
1bxo	1.0	94	323	1er8	338	55	0.58
1cgn	2.2	90	127	2ccy, chain *A*	122	31	0.51
6ebx	1.7	87	124	3ebx	62	100	0.50
2pby	2.1	97	1155	1mki	1248	46	0.45
1yxa	2.1	95	740	1qlp	744	46	0.45
1ujz	2.1	100	215	*ARP/wARP*	13 (+33 Gly)	—	0.43
2iff	2.6	65	556	2hem	129	98	0.39

† Honnappa (2008).

so that intermediate results may also be appreciated. The final values of CORR and MPE may be compared with $\mathrm{CORR_{EDM}}$ and $\mathrm{MPE_{EDM}}$, respectively, which are the corresponding values obtained by applying EDM alone to the phase set $(\varphi_p)_1$. The results in Table 2 suggest the following considerations.

(i) The number of intermediate CORR and MPE values varies from one structure to the next, since the number of supercycles is automatically determined by the figures of merit (see §3.1). As a general trend, a steady reduction in the average phase error with the supercycle number can be noted: in the last supercycles the improvement decreases, thus justifying the use of the figures of merit.

(ii) Structures with higher data resolution show quite high values of the final CORR value when EDM–DEDM is applied (their final CORR is above 0.90).

(iii) The final values of CORR and MPE are always better and are often remarkably better than the corresponding $\mathrm{CORR_{EDM}}$ and $\mathrm{MPE_{EDM}}$ values (the exceptions are 6rhn and 1na7, for which the values are substantially equivalent). This means that EDM–DEDM is more efficient than EDM alone.

(iv) EDM–DEDM reduces the phase error more efficiently than EDM by itself. For 1kf3, 1zs0, 1a6m, 1e3u, 1s31, 2sar, 1pm2, 1i9a, 1xyg, 1bxo, 1cgn and 6ebx the difference in the respective performances is more than 5°; it is more than 10° for 1a6m, 1e3u, 1s31, 2sar, 1pm2, 1i9a, 1bxo, 1cgn and more than 20° for 1e3u and 1bxo. This happens irrespective of the data resolution limit (which lies in the range 1.0–2.7 Å), of the quality of the initial phase set (at least for MPE < 78°) and of the complexity of the target structure (NresT ranges from 62 to 1380). It is worthwhile noting that similar results for the application of EDM only were obtained using the program *DM* (Cowtan, 1999) from the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994).

(v) Greater efficiency of the EDM–DEDM procedure also occurs for structures (*i.e.* 6ebx, 1ujz and 2sar) which have very incomplete starting models (see the NresM/NresT ratio). Indeed, while the target structures of 6ebx and 2sar contained two monomers related by noncrystallographic symmetry, only one monomer is found in their model structures. Also, the target structure of 1ujz is constituted of two different monomers: the model structure very partially covers one monomer (NresM/NresT = 0.2).

(vi) For the structures 1lys, enhexa, 6rhn and 1na7, EDM–DEDM (as well as EDM) is not able to substantially improve the initial model. While enhexa, 6rhn and 1na7 have good starting models (and therefore limited room for improvement), an effective loss of efficiency of both procedures occurs for 1lys. One possible reason is that 1lys contains two monomers translated by the vector $(\frac{1}{2}, 0, \frac{1}{2})$ while the starting model consists of a single monomer placed by MR: in this case pseudo-translational symmetry effects may limit the efficiency of the procedures. In addition, 1lys has a data completeness of 85%, with missing reflections concentrated at resolutions below 10 Å (100% missing) and above 1.8 Å (73% missing).

(vii) Both EDM–DEDM and EDM failed to improve the initial phase set of 2iff. This structure represents a borderline case: it has a modest data resolution limit (Res = 2.6 Å) and a very incomplete model structure (NresM/NresT = 0.23); indeed, its target structure consists of five monomers, the smallest of which has been approximated by the MR model. In addition, in contrast to the 1ujz case, which has a similar NresM/NresT ratio but 100% completeness, the data completeness for 2iff is only 65% and falls to 30% at resolutions higher than 3.0 Å. The initial value of CORR for 2iff (0.39) may be taken as limiting parameter for an electron-density map to have any chance of being improved by EDM–DEDM.

(viii) The phase improvement obtained using this procedure shows a relevant dependence on three parameters: the initial value of MPE, the data resolution and the sequence identity, at least in the ranges covered by the test structures [Res between 2.8 and 1.0 Å, Id between 100% and 30%]. These parameters are not independent of each other, so that it is their combination which influences the degree of success of the method. For example, in very favourable circumstances (a small value of the initial MPE, a small value of Res and a large value of Id) the starting model is already a good approximation of the target structure, so that the room for improvement is limited (this is the case for 1kf3, enhexa, 6rhn and 1zs0).

The procedure is very likely to succeed when the initial MPE is large but Res is small (the method can exploit a large

Table 2
Results of the EDM–DEDM procedure.

CORR is the correlation factor between the electron-density map calculated using observed moduli and a given phase set and the map calculated using observed moduli and published phases φ; MPE is the mean phase error (in degrees) between a given phase set and the published phases. The considered phase sets are the starting phase set and the final phase set; intermediate results after each DEDM step are shown between slashes.

PDB code	CORR	$CORR_{EDM}$	MPE	MPE_{EDM}
1kf3	0.95→0.96/0.97/0.97→0.97	0.96	26→23/19/18→18	25
enhexa	0.92→0.94/0.95/0.95→0.95	0.93	22→20/19/20→21	23
6rhn	0.91→0.92/0.94/0.93→0.94	0.93	32→30/27/28→29	27
1zs0	0.86→0.89/0.92/0.93→0.93	0.88	42→39/32/31→31	38
3ebx	0.84→0.88/0.90/0.91→0.91	0.87	45→42/37/36→39	43
1a6m	0.83→0.88/0.93/0.94/0.94→0.94	0.86	50→45/32/27/25→30	43
1na7	0.78→0.79/0.81/0.81→0.81	0.81	47→46/45/47→48	45
1e3u	0.77→0.83/0.89/0.90→0.91	0.76	51→45/35/33→33	50
1s31	0.76→0.80/0.88/0.90→0.89	0.83	52→49/40/37→36	46
2p0g	0.76→0.78/0.83/0.84/0.84→0.84	0.81	57→56/50/49/49→49	53
2sar	0.72→0.77/0.83/0.85/0.85→0.85	0.70	53→50/43/41/40→40	56
1pm2	0.69→0.76/0.82/0.83→0.85	0.67	56→51/43/42→42	56
1kqw	0.69→0.73/0.79/0.80/0.80→0.81	0.76	60→57/49/47/46→47	51
1i9a	0.68→0.71/0.81/0.82→0.84	0.66	58→56/49/47→47	57
1lys	0.65→0.66/0.69/0.68→0.69	0.66	64→63/59/60→61	62
1xyg	0.64→0.65/0.69→0.70	0.63	64→64/60→59	64
1bxo	0.58→0.66/0.80/0.85/0.91/0.94→0.94	0.68	78→71/52/41/31/25→25	65
1cgn	0.51→0.54/0.66/0.73/0.75→0.77	0.60	74→72/63/58/55→52	66
6ebx	0.50→0.53/0.56/0.58/0.59/0.60→0.60	0.48	72→70/66/64/63/62→63	72
2pby	0.45→0.45/0.53/0.55/0.55→0.55	0.50	77→76/73/71/72→70	72
1yxa	0.45→0.45/0.52/0.52/0.51→0.51	0.48	77→77/73/72/72→72	74
1ujz	0.43→0.46/0.52/0.57/0.58→0.59	0.48	70→68/65/62/62→62	66
2iff	0.39→0.38/0.37/0.34→0.29	0.29	72→72/74/76→77	79

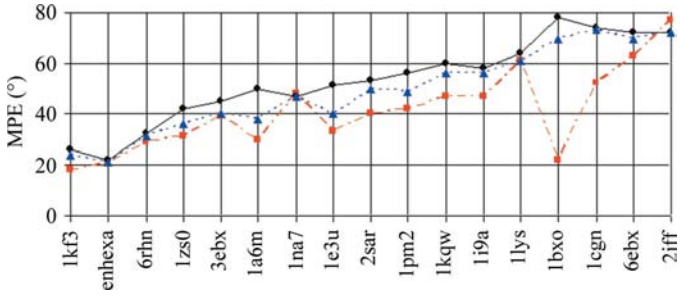

Figure 1
Initial (circles, solid line) and final (triangles, dashed line) CORR values for the previous version of the EDM–DEDM procedure, taken from Table 4 of Caliandro *et al.* (2008), together with the final CORR values for the present version of the procedure (squares, dashed/dotted line).

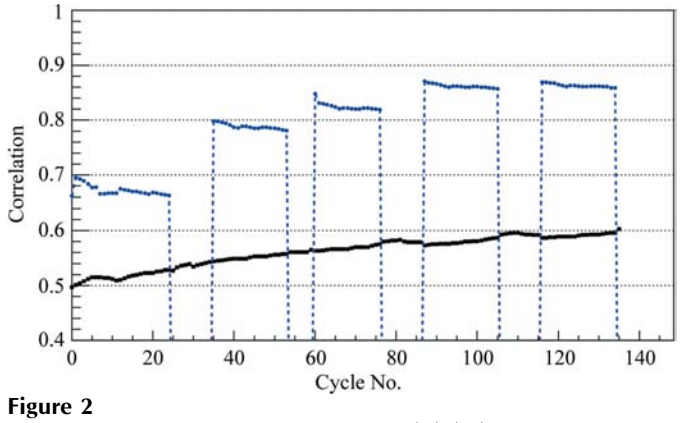

Figure 2
Values of CORR (solid line) and CORR($|E|$, $|E'|$) (dashed line) as a function of the EDM–DEDM cycle number for the test structure 6ebx. The vertical dashed lines separate DEDM from EDM cycles.

amount of experimental information). This is the case for 1bxo, for which Res = 1.0 Å, Id = 55% and the starting MPE is 78°: the final value of MPE is 25°, while the application of EDM alone ends with a value of 65°.

Under more hostile conditions (resolution between 2.0 and 3.0 Å and/or sequence identity between 40% and 70%) the EDM–DEDM procedure may significantly improve a poor starting model. For example, for 1cgn (Res = 2.2 Å, Id = 31%) MPE drops from 74° to 52° when EDM–DEDM is applied: the application of EDM alone leads to an MPE of 66°.

The worst situation occurs when Res and Id are both small and the initial MPE value is large; in this case the quantity of experimental information may not be sufficient for a significant improvement. This is the case for 2pby (initial MPE = 77°, Res = 2.1 Å, Id = 46%) and 1yxa (initial MPE = 77°, Res = 2.1 Å, Id = 46%). In the first case EDM–DEDM improves the phase error from 77° to 70° and in the second case from 77° to 72°. The application of EDM alone reduces the MPE to 72° and 74°, respectively.

Most of the present test cases were also used to validate the first EDM–DEDM algorithm. A comparison of the above results with those reported in Table 4 of Caliandro *et al.* (2008) is shown in Fig. 1, in which for each test structure tested by the two algorithms we give the initial (circles) and the final phase error according to Caliandro *et al.* (2008) (triangles) and according to the present paper (squares). The increase in efficiency is evident for all the considered structures and is dramatic for some of them.

To figure out the evolution of the initial phase set during the procedure, in Fig. 2 the values of CORR (solid line) and CORR($|E|$, $|E'|$) (dashed line) are plotted as a function of the cycle number for 6ebx. In this case five supercycles are executed, each consisting of a variable number of DEDM cycles [in Fig. 2 they correspond to the points for which CORR($|E|$, $|E'|$) is calculated] and EDM cycles. We observe that (i) CORR exhibits a steady increase throughout the procedure, with EDM and DEDM complementing each other, and tends to become constant for higher supercycle number and (ii) the CORR($|E|$, $|E'|$) values fluctuate during the DEDM cycles, but follow the increase of CORR from one supercycle to the other rather faithfully, thus allowing its use as a figure of merit.

5. Analysis of local improvements

In §4, we monitored the global features of the EDM–DEDM algorithm *via* the parameters CORR and MPE. Since the

Table 3
Correlation coefficients for missing monomers.

Average values of the correlation factor, calculated residue by residue, between the electron-density map calculated from the published coordinates and that obtained before (Initial) and after application of the EDM–DEDM procedure (EDM–DEDM) and after application of the EDM procedure (EDM). The residues considered are those of the monomer that were not included in the model structure used for MR.

PDB code	Initial	EDM–DEDM	EDM
2sar	0.58	0.78	0.54
1lys	0.64	0.66	0.65
6ebx	0.31	0.43	0.26
1ujz	0.37	0.52	0.38

Fourier difference is particularly skilled at showing up structural details that are not present in the starting electron-density map, special attention may be dedicated to the capacity of the algorithm to improve the electron density in those parts of the target which are not present in the model structure. Relevant structural features to check, for example, may be the residues of a monomer or the atoms of a ligand that are not contained in the model but are present in the target structure. In Table 3 the average values of the correlation coefficient calculated residue by residue for the main-chain atoms of the missing monomer of 2sar, 1lys, 6ebx and 1ujz are reported. The calculations were carried out using the *CCP*4 package. The columns Initial, EDM–DEDM and EDM refer to the initial electron-density map, to that obtained after the EDM–DEDM procedure and to that obtained by EDM alone. For 2sar, 6ebx and 1ujz the electron-density map in the region of the missing monomer is improved by EDM–DEDM remarkably more than by EDM alone. A smaller increase in correlation occurred for 1lys, for which the two procedures are almost equivalent.

Capability to determine the ligand structure is of fundamental importance in structure-aided drug design. To check the performance of EDM–DEDM, we considered test cases that contained a ligand with more than five non-H atoms in the target structure. The cases selected are reported in Table 4, together with the identification name of the ligand and its chemical formula. The average values of the atom-by-atom correlation coefficient calculated for the ligand atoms are reported in the last three columns of Table 4. It can be seen that the atom-by-atom correlation (again calculated using *CCP*4) increases for all the ligands after application of the EDM–DEDM procedure. The best results were attained for 1a6m and 1bxo, owing to the high data resolution, but relevant increments were also obtained for more difficult cases such as 1cgn, 1s31 and 2sar. A more intuitive view of the improvement may be achieved by a graphical inspection of the electron-density map in the surroundings of the protein–ligand contact. With this aim, in Fig. 3 we report the starting electron-density map (in blue) and that obtained after the EDM–DEDM procedure (in green) for the region around the PGE ligand of 1s31. The two maps, plotted at 1.5 standard deviations using the program *Coot* (Emsley & Cowtan, 2004), reveal a

Table 4
Correlation coefficients for ligands.

PDB is the PDB code of the protein structure; ID is the identifier name of the ligand included in its structural model; Formula is the ligand chemical formula; Initial, EDM–DEDM and EDM are the average values of the correlation factor, calculated atom by atom, between the electron-density map calculated from the published coordinates and that obtained before and after application of the EDM–DEDM procedure and after application of the EDM procedure, respectively. The considered atoms are those of the corresponding ligand, which was not included in the model structure used for MR.

PDB	ID	Formula	Initial	EDM–DEDM	EDM
1zs0	EIN	$C_{17}H_{22}NO_6PS$	0.73	0.80	0.75
1zs0	MES	$C_6H_{13}N$	0.81	0.86	0.80
1a6m	HEM	$C_{34}H_{32}FeN_4O_4$	0.73	0.91	0.82
1s31	PGE	$C_6H_{14}O_4$	0.52	0.79	0.59
2sar	3GP	$C_{10}H_{14}N_5O_8P$	0.61	0.77	0.64
1kqw	RTL	$C_{20}H_{30}O$	0.49	0.54	0.49
1bxo	GOL	$C_3H_8O_3$	0.11	0.66	0.19
1bxo	MAN	$C_6H_{12}O_6$	0.38	0.78	0.35
1bxo	PP7	$C_{30}H_{47}N_4O_9P$	0.56	0.92	0.69
1cgn	HEM	$C_{34}H_{32}FeN_4O_4$	0.77	0.78	0.74

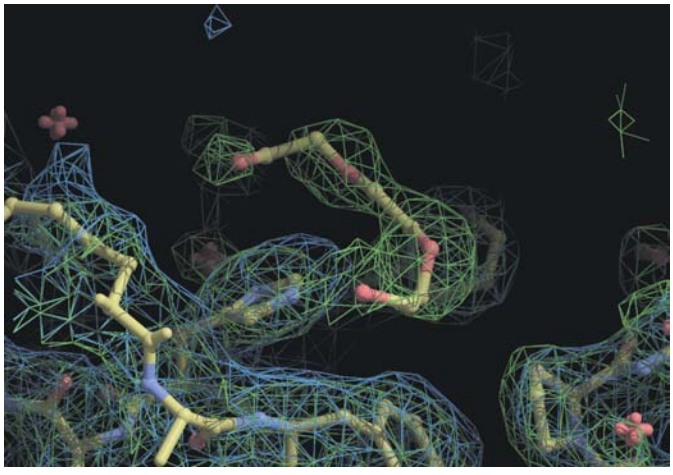

Figure 3
1s31: electron-density map calculated before (blue) and after (green) application of the EDM–DEDM procedure, superimposed on the structure of the ligand PGE. The two densities were plotted at 1.5 standard deviations using the program *Coot* (Emsley & Cowtan, 2004).

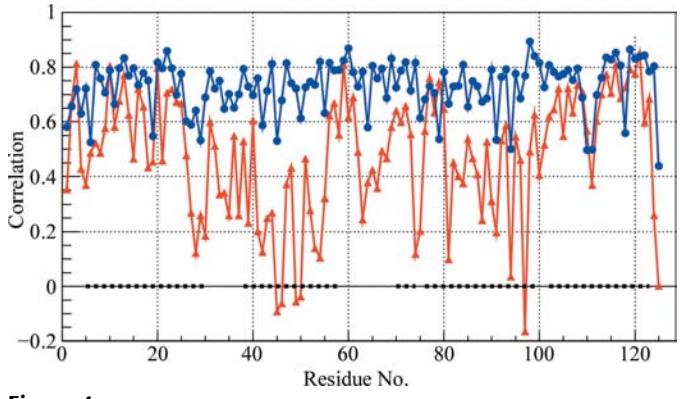

Figure 4
1cgn: main-chain residue-by-residue correlation coefficients between the electron-density map calculated from the published coordinates and that calculated before (triangles) and after (circles) application of the EDM–DEDM procedure. Dotted segments along the zero-level horizontal line indicate residues belonging to helix domains in the published model.

completely different description of the ligand, which may only be identified from the EDM–DEDM map. The blue density only starts to include the ligand if contoured at 0.8 standard deviations: only a few PGE atoms are included, together with much false detail.

One of the structures that most benefitted from EDM–DEDM was 1cgn. A detailed analysis of local improvements is given in Fig. 4, in which the correlation coefficient calculated for the main-chain atoms is plotted against residue number. It can be noted that the triangles (the initial electron-density map) are systematically below the circles (the map after the application of EDM–DEDM). The local improvements are correlated with the differences between the MR model and the target structure. The fold of 1cgn is a four-helical up-and-down bundle: in Fig. 4, dotted segments along the zero-level horizontal line indicate residues belonging to helix domains identified in the published structure. It may be noted that modest improvements occur for residues belonging to the first and last helices, where the MR model reproduces the target structure well. The largest improvements occur for residues belonging to loops, where the MR model substantially deviates from the target structure, and for the remaining helices. In this latter case the helix of the MR model is shifted along the axis of the corresponding helix of the target structure, so that the C^{α} atoms deviate up to 4 Å from their true positions.

6. Conclusions

The recently proposed DEDM algorithm has been optimized. Its expected complementarity with the well established EDM procedure has been carefully exploited by the development of a cyclic procedure in which DEDM and EDM steps were combined. The proposed EDM–DEDM procedure operates both in direct space *via* map modification and inversion and in reciprocal space by using more suitable Fourier coefficients for the difference Fourier synthesis. The operation of adding the difference to the calculated Fourier synthesis to obtain improved phase sets is also performed in both direct and reciprocal space. From a general point of view, adding the DEDM step to standard EDM procedures may be interpreted as an attempt to perform a special density modification which uses different parameters in regions covered by the model and in regions in which atoms should be placed.

Substantial improvements of the initial electron-density map were achieved even in difficult cases, *i.e.* with very incomplete initial models (containing up to 20% of the residues of the target structure) and at data resolution up to 2.7 Å. As a rule of thumb, the procedure is able to improve electron-density maps that have a correlation coefficient with the true map of larger than 0.4, while it is penalized by incomplete data sets, at least when data completeness falls below 85%.

An open question, which still represents the major obstacle for iterative application of the procedure, is represented by the different efficiency of the DEDM step when triggered by a structural model or by an electron-density map.

Practical applications of the EDM–DEDM procedure may range from structural model completion to drug design through ligand structure determination as well as elimination of model bias in MR. We foresee its use both within phase-refinement procedures and in connection with model-building programs.

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
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.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2008). *Acta Cryst.* A**64**, 519–528.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. D. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Henderson, R. & Moffat, J. K. (1971). *Acta Cryst.* B**27**, 1414–1420.
Honnappa, S. (2008). Personal communication.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
Main, P. (1979). *Acta Cryst.* A**35**, 779–785.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Ursby, T. & Bourgeois, D. (1997). *Acta Cryst.* A**53**, 564–575.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 311–331. Dordrecht: Kluwer Academic Publishers.
Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.