Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

EDM–DEDM and protein crystal structure solution

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 5 February 2009
Accepted 9 March 2009

Electron-density modification (EDM) procedures are the classical tool for driving model phases closer to those of the target structure. They are often combined with automated model-building programs to provide a correct protein model. The task is not always performed, mostly because of the large initial phase error. A recently proposed procedure combined EDM with DEDM (difference electron-density modification); the method was applied to the refinement of phases obtained by molecular replacement, *ab initio* or SAD phasing [Caliandro, Carrozzini, Cascarano, Giacovazzo, Mazzone & Siliqi (2009), *Acta Cryst.* D**65**, 249–256] and was more effective in improving phases than EDM alone. In this paper, a novel fully automated protocol for protein structure refinement based on the iterative application of automated model-building programs combined with the additional power derived from the EDM–DEDM algorithm is presented. The cyclic procedure was successfully tested on challenging cases for which all other approaches had failed.

1. Notation

The following notation has been used in the article.
PDB: Protein Data Bank.
SIR/MIR: single/multiple isomorphous replacement.
SAD/MAD: single/multiple anomalous dispersion.
MR: molecular replacement.
AbI: *ab initio*.
AbIM: *ab initio* modelling.
AMB: automated model building.
EDM: electron-density modification.
DEDM: difference electron-density modification.
$N_{\rm asym}$: number of non-H atoms in the asymmetric unit.
RES: experimental data resolution.
ρ, ρ_p: electron density of the target and of the model structure, respectively.
$\rho_q = \rho - \rho_p$: the ideal difference Fourier synthesis. Summed to ρ_p, it exactly provides ρ, no matter what the quality of ρ_p is.
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.

2. Introduction

The solution of the phase problem for proteins is far from being routine. Four methods are traditionally used: *ab initio* phasing, SIR/MIR, SAD/MAD and MR. The first relies on the native diffraction data only, while the others require supple-

mentary experimental diffraction data (for SIR/MIR or SAD/MAD techniques) or some phasing information (for MR techniques).

RES and $N_{\rm asym}$ determine the limits for the *ab initio* approach. While success has been obtained up to a considerable size (*i.e.* $N_{\rm asym}$ = 7890), the limit for RES is 2 Å, which is only attained when some heavy atoms are present in the unit cell (Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo, Mazzone *et al.*, 2008).

SIR/MIR and SAD/MAD have wider limits of RES and $N_{\rm asym}$. Advances in the SIR/MIR procedures rely on new appealing techniques based on quick-soak methods, the use of noble gases, innovation in derivatization strategies and new mathematical approaches. SAD/MAD procedures have profited from the increasing power and tunability of synchrotron beamlines. Several well documented programs [*SnB* (Xu *et al.*, 2006), *SHELXD* (Sheldrick, 2008), *RANTAN* (Yao, 1981), *ACORN* (Yao, 2002), *IL MILIONE* (Burla *et al.*, 2007) and *SOLVE*/*RESOLVE* (Terwilliger, 2004)] are now available for *ab initio* and/or for SIR/MIR and SAD/MAD techniques.

MR is a key method for solution of the phase problem: it provides initial phase estimates for a given structure (the target) by using a previously known structure (the model). Recent advances in the method allow the use of inaccurate and/or incomplete model structures. Six-dimensional space search programs (*e.g.* Chang & Lewis, 1997; Kissinger *et al.*, 1999; Sheriff *et al.*, 1999; Glykos & Kokkinidis, 2000; Jamrog *et al.*, 2003) or two three-dimensional searches (Navaza, 1994; Vagin & Teplyakov, 1997; Read, 2001; Yao, 2002; Caliandro *et al.*, 2006; McCoy *et al.*, 2007) may be used to identify the target structure.

In the past few years, a new approach for protein crystal structure determination has been proposed (Bradley *et al.*, 2005; Qian *et al.*, 2007): *ab initio* structure prediction coupled with MR. According to this approach, fragments of known structures that are compatible with local sequences of the target protein are produced and clustered to represent the target structure; the most plausible models are then optimized *via* energy-based algorithms and are used in MR approaches to provide a starting set of crystallographic phases. This method will be referred to hereafter as *ab initio* modelling (AbiM).

No matter which of the above five methods is employed, the phases that they provide are usually not good enough to allow AMB programs [*e.g.* *ARP*/*wARP* (Perrakis *et al.*, 1999), *PHENIX* (Terwilliger *et al.*, 2008), *MAID* (Levitt, 2001), *MAIN* (Turk, 1992) and *Buccaneer* (Cowtan, 2006)] to provide complete and/or reliable models of the protein structure. EDM techniques are a necessary intermediate step (Cowtan, 1999; Abrahams, 1997; Abrahams & Leslie, 1996; Zhang *et al.*, 2001; Refaat & Woolfson, 1993; Giacovazzo & Siliqi, 1997): they modify electron-density maps to capture the desired features of the maps and therefore improve the phases. Often their application is not sufficient: the average phase error may remain large and the AMB programs have no chance of providing good models. Under these conditions an alternative consists of modifying the common EDM protocols or of attempting an iterated application of the AMB programs, either alone or in combination with EDM. As an ultimate possibility, one can attempt a manual interpretation of the electron-density map or manual adjustments of the intermediate models, with a great waste of time. Recently, He *et al.* (2007) have proposed a protocol for automatic MR model completion based on iterated use of the programs *OASIS* (Zhang *et al.*, 2007), *DM* (Cowtan, 1999) and *ARP*/*wARP*. The first program performs a reciprocal-space phase refinement based on direct-methods techniques, the second improves the phases using EDM algorithms and the third performs real-space model building and refinement. This dual-space approach to model completion has been tested on two protein structures by using several artificially constructed partial models and the results showed that the *OASIS* step greatly increases the efficiency of the whole protocol.

Recently, a new algorithm based on modification of the difference electron density, called DEDM, has been described (Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo & Siliqi, 2008) and has been combined with classical EDM procedures to improve the phase sets obtained by MR, *ab initio* or SAD methods (Caliandro *et al.*, 2009). In some cases the EDM–DEDM approach succeeded where the application of EDM alone substantially failed. This algorithm is implemented in *IL MILIONE* (Burla *et al.*, 2007) and its potential has still to be fully explored. This paper deals with the application of EDM–DEDM to some challenging cases arising from the application of the five phasing methods described above: the new procedure combines EDM–DEDM with AMB programs and tries to lead to the solution of very imperfect model structures that were resistant to any other approach.

3. The procedure

The most popular coefficients of the ΔF- syntheses are $(m|F| - D|F_p|)\exp(i\varphi_p)$: in the absence of any supplementary information they are expected to be the most useful approximation to F_q. However, the above assumption involves a systematic bias: φ_q is expected to be collinear with φ_p. The recently proposed DEDM procedure (Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo & Siliqi, 2008) breaks down the collinearity between model structure phases and difference structure-phase estimates *via* the following steps.

(i) The ΔE- synthesis is modified by squaring the very positive and very negative parts of the map: the rest is set to zero.

(ii) Fourier inversion of the modified ΔE- synthesis generates phase shifts $\Delta\varphi_q$, which are used, *via* the Carnot theorem, to estimate the value $|E_q|$.

(iii) Estimates of the structure factors of the full structure are obtained by combining the model structure with the difference synthesis.

DEDM is expected to be complementary with respect to EDM, since reflections with large $|F_p|$ and $|F|$ values largely contribute to the F- synthesis but are of limited use for the ΔF- synthesis, while reflections with large $|F_p|$ and small $|F|$ values may not be useful for the F- synthesis but very infor-

mative for the ΔF- synthesis. For this reason, DEDM was combined with EDM in a cyclic procedure that proved to be more powerful than the two single techniques (Caliandro *et al.*, 2009). The EDM–DEDM flowchart may be obtained from Fig. 1 by eliminating the blue blocks and arrows. $\rho_{\rm mod}$ is the initial electron density corresponding to a partial model, ρ_q is the difference electron density provided by DEDM, $\rho_{\rm new}$ is the new electron density obtained by combining model and difference electron densities and $\rho'_{\rm new}$ is obtained from $\rho_{\rm new}$ by EDM techniques. A new partial structure $\rho_{\rm mod}$ is obtained by selecting a suitable percentage of $\rho'_{\rm new}$ and a new EDM–DEDM cycle is started until the procedure converges (a suitable figure of merit is checked) and the procedure stops.

In this paper, we propose a generalization of this procedure: the cyclic combination of DEDM, EDM and AMB programs (for shortness, we call it DEA) with the aim of profiting from the supplemental information provided by each of them. The flowchart of the new procedure coincides with that schematically drawn in Fig. 1 after elimination of the green block and arrow. An initial set of phases (or model, if MR or AbiM is used) is immediately submitted to an AMB program to obtain a starting model (from now on referred to as $\rm MOD_1$). In some cases this does not lead to an effective phase improvement, but it is still a useful step for subsequent calculations. The EDM–DEDM cycles are then executed and the final $\rho'_{\rm new}$ obtained after convergence is directly submitted to an AMB program to start a new DEA cycle. The size of the model and the percentage of the docked residues may be used as a figure of merit to stop the iterations.

The final purposes of the new procedure are the following.

(i) The automatic modification of a set of poor phases, no matter whether obtained *via* AbI, SIR/MIR, SAD/MAD or *via* a model from MR or AbIM, to obtain a new set of phases with sufficient quality for the derivation, *via* AMB programs, of substantially correct model structures.

(ii) Validation of the phase improvements obtained *via* EDM–DEDM. Indeed, the automatic application of a modern AMB program may be considered as an objective tool for assessing the quality of an electron-density map.

(iii) The making available of a new procedure of interest for high-throughput protein crystallography, in which automation is one of the necessary requisites. In all our applications the same standard procedure will be applied.

Since DEA has been designed to cope with difficult cases (for which AMB programs fail, even if iteratively applied), in our applications the size of the model as well as the percentage of docked residues are expected to be a small percentage of the total.

4. Applications

For DEA applications, some challenging cases have been selected according to the following criterion: that the available starting set of phases (or models) obtained using one of the five phasing techniques mentioned in §2 do not provide correct structural models when processed using the most popular AMB programs.

The AMB programs used in our applications are *ARP/wARP* v.7.0.1 (Perrakis *et al.*, 1999) and the *PHENIX AutoBuild* wizard (Terwilliger *et al.*, 2008), which is part of the *PHENIX* v.1.3 project (Adams *et al.*, 2002). *ARP/wARP* is a package for automated model building and structure refinement. It is based on a unified approach to the structure-solution process by combining electron-density interpretation, pattern recognition in an electron-density map and maximum-likelihood model parameter refinement. The *PHENIX Autobuild* wizard is a highly automated tool for iterative model building, structure refinement and density modification using *RESOLVE* for statistical density modification and model building (Terwilliger, 2003*a*,*b*) and *phenix.refine* for structure refinement (to be published). As a rule of thumb, in our tests we first included *ARP/wARP* within the DEA cycles (because it is faster than the *PHENIX AutoBuild* wizard). If a reasonable

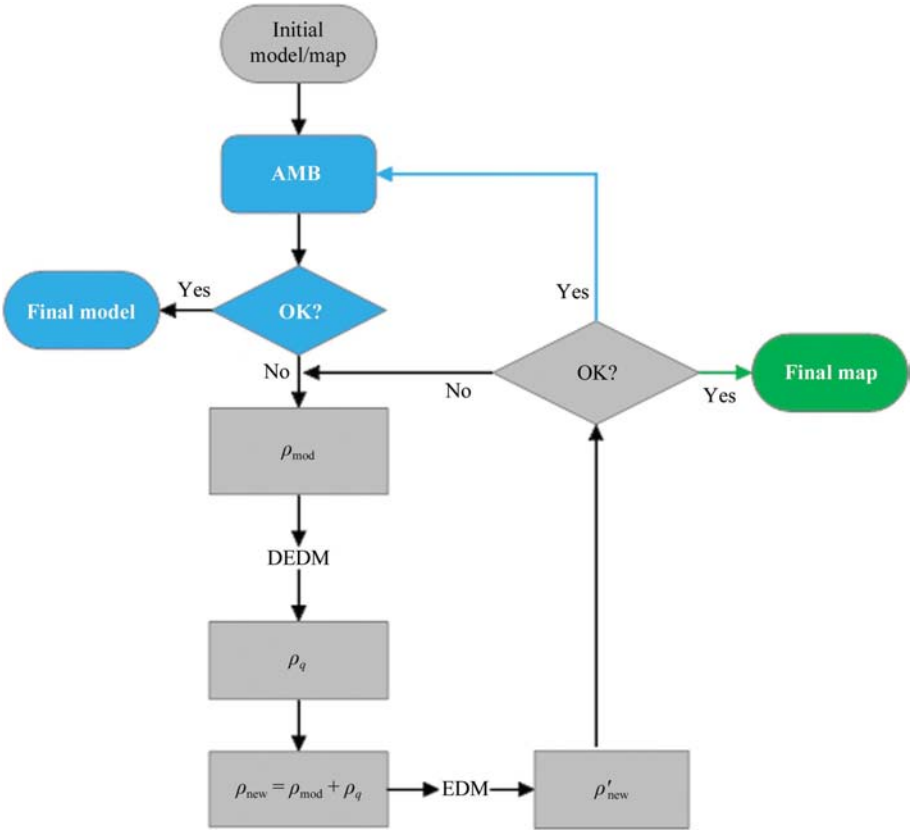

Figure 1
DEA iteration procedure flowchart.

Table 1
Test cases: see text for notation.

PDB code	NRES	RES (Å)	Phasing method	$SIZE_1/NDOCK_1$	$CORR_1$ ($\langle\lvert\Delta\varphi\rvert\rangle_1$)	AMB	Reference
1ujz	215	2.10	MR	74/39	0.49 (67)	A	Kortemme *et al.* (2004)
1yxa	742	2.10	MR	529/47	0.57 (67)	P	Horvath *et al.* (2005)
1cgn (Fe, 1)	127	2.15	MR	51/14	0.52 (73)	A	Dobbs *et al.* (1996)
2fzt (Se, 3)	156	2.05	AbIM	56/0	0.24 (78)	P	Uhrinova *et al.* (2001)
2f14 (Zn, 1; Hg, 1)	257	1.71	AbI	67/20	0.59 (64)	A	Alterio *et al.* (2006)
1arm (Cu, 1; Hg, 4)	307	1.76	AbI	87/28	0.59 (64)	P	Greenblatt *et al.* (1998)
1yfd (Hg, 13; Fe, 4)	681	1.90	AbI	193/21	0.58 (59)	P	Kolberg *et al.* (2005)
1svn (Ca, 3; Cl, 2; S, 2)	269	1.74	SAD	93/12	0.65 (58)	A	Betzel *et al.* (1988)
1lvo (Se, 60)	1798	1.95	MAD	900/310	0.68 (61)	A	Anand *et al.* (2002)

Table 2
Results obtained by DEA for the test cases: see main text for notation.

PDB code	NCycle	$SIZE_f/NDOCK_f$	$CORR_f$ ($\langle\lvert\Delta\varphi\rvert\rangle_f$)
1ujz	3	200/200	0.97 (27)
1yxa	9	693/640	0.90 (32)
1cgn	3	123/123	0.98 (19)
2fzt	4	150/145	0.91 (31)
1arm	10	263/263	0.93 (30)
1yfd	13	588/588	0.91 (31)
2f14	11	227/222	0.94 (27)
1svn	3	250/250	0.96 (22)
1lvo	18	1309/1003	0.83 (42)

model was not reached after a reasonable number of DEA cycles, then the *PHENIX Autobuild* wizard was used.

In Table 1 for each test structure we give the following.

(i) The Protein Data Bank code (PDB code). Heavy and/or anomalous scattering atoms that are helpful in obtaining starting phase estimates are quoted in parentheses, together with their numbers.

(ii) The number of residues in the asymmetric unit (NRES).

(iii) The experimental data resolution of the target structure (RES).

(iv) The phasing method used to obtain the initial set of phases.

(v) The parameters characterizing the model obtained after the first AMB application (MOD_1), *i.e.* the size of the model in terms of residues ($SIZE_1$) and the number of docked residues ($NDOCK_1$).

(vi) The correlation factor between the electron-density map obtained using observed moduli and phases calculated from MOD_1 and the map calculated *via* observed moduli and published phases ($CORR_1$).

(vii) The corresponding mean phase error with respect to the published phases ($\langle\lvert\Delta\varphi\rvert\rangle_1$).

(viii) The letters A or P, which denote the AMB program used (A stands for *ARP*/*wARP* and P for *PHENIX*).

In Table 2 we show the results obtained *via* DEA. For each test structure we give the PDB code, the number of DEA cycles (NCycle), the number of built and docked residues of the final model ($SIZE_f$ and $NDOCK_f$, respectively). For the map and phases calculated from the final model the correlation value $CORR_f$ and the average phase error $\langle\lvert\Delta\varphi\rvert\rangle_f$ are also given.

Comparing Table 2 with Table 1 it may be seen that a good solution ($\langle\lvert\Delta\varphi\rvert\rangle_f < 30^\circ$, an almost completely docked model) is also reached in extreme cases for which the initial phase set has $\langle\lvert\Delta\varphi\rvert\rangle_1 > 70^\circ$ and MOD_1 is very incomplete.

In the following the default results of DEA are described in greater detail for each phasing technique. In order to correctly appreciate them, we note that the automation of modern ABM programs is very high and that we have used them in a default way. Therefore, when we state that the mere application of EDM, *ARP*/*wARP*, *PHENIX* and of their combinations to the selected test cases do not lead to satisfactory models we only refer to their default use: we do not claim that manual inspection of the electron-density maps combined with nondefault applications of such programs or manual adjustment of the models could not lead to good final models.

4.1. MR cases

The test case 1ujz has been used by He *et al.* (2007) to test their dual-space model-completion procedure in a simulated difficult situation. 1ujz is a 215-residue protein constituted of two molecules: molecule *A* of 87 residues and molecule *B* of 128 residues. Molecule *A* of a similar complex, PDB entry 1bxi, shows 60% sequence identity and a root-mean-square deviation of 1.38 Å with respect to 1ujz. Molecule *A* of 1bxi was pruned using the program *CHAINSAW* (Schwarzenbacher *et al.*, 2004) to produce a search model for the MR program *Phaser* (Read, 2001). The MR solution was submitted to *ARP*/*wARP*, which built a 46-residue model (only 13 of which had side chains) amounting to about 20% of the 1ujz structure, which was completely destroyed by iterating the *DM*–*ARP*/*wARP* cycle. Success was obtained using the combination *OASIS*–*DM*–*ARP*/*wARP*, which was able to produce a model of 201 residues all docked in the sequence after seven cycles.

The same 46-residue model as used by He and coworkers was used in our tests: for the model $\langle\lvert\Delta\varphi\rvert\rangle_1 = 67^\circ$ and $CORR_1 = 0.49$. Firstly, the EDM–*ARP*/*wARP* iteration procedure was applied to $\rho_{\rm mod}$: it did not provide any useful model in spite of the large number of iterations, which is in agreement with the results of He and coworkers. In Fig. 2(*a*) the dashed line shows the average phase error $\langle\lvert\Delta\varphi\rvert\rangle$ *versus* the EDM–*ARP*/*wARP* iteration number (up to 7): the open circles refer to the phases obtained after the application of EDM and the filled squares to those obtained after the application of AMB. In contrast, in three cycles DEA produced (unbroken line in Fig. 2*a*) a model of 200 residues,

all docked in the sequence, for which $\langle|\Delta\varphi|\rangle_f = 27^\circ$ and $\mathrm{CORR}_f = 0.97$. Thus, the proposed protocol was able to decrease the phase error by 40° in only three iterations.

It is worthwhile noticing that each DEA cycle includes several EDM–DEDM microcycles, as shown in Fig. 2(*b*), in which the second DEA cycle is enlarged. While in Fig. 2(*a*) we only report the phase errors after each cycle, in Fig. 2(*b*) we show the phase error after each microcycle; the filled circles correspond to the phase error after application of DEDM. The synergy between the EDM and DEDM algorithms may be schematized as follows: DEDM cycles add new features to the partial model which may be independent (indeed, DEDM relies on real-space modifications of the difference electron-density map) of the electron-density modifications carried out by EDM.

The remaining two MR test cases correspond to structures originally solved by MR *via* quite imperfect models. We will use the results provided by the program *REMO* (Caliandro *et al.*, 2006), included in *IL MILIONE* (Burla *et al.*, 2007), by exploiting the same model structures originally employed to solve the target structures.

The structure 1cgn (NRES = 124) has a model (2ccy) consisting of 127 residues. The root-mean-square distance between the C^α atoms of the model and target structures is 1.73 Å, while the sequence identity is 31%. For the model structure provided by *REMO* $\langle|\Delta\varphi|\rangle_1 = 73^\circ$ and $\mathrm{CORR}_1 = 0.52$. Direct application of *ARP*/*wARP* to the MR solution produced a model with 51 residues (only 14 of which had side chains). Application of the EDM–AMB procedure provided a 121-residue model, all of which were docked in the sequence, after seven cycles. The DEA procedure provided the same result (122 residues all docked in the sequence) in only three cycles.

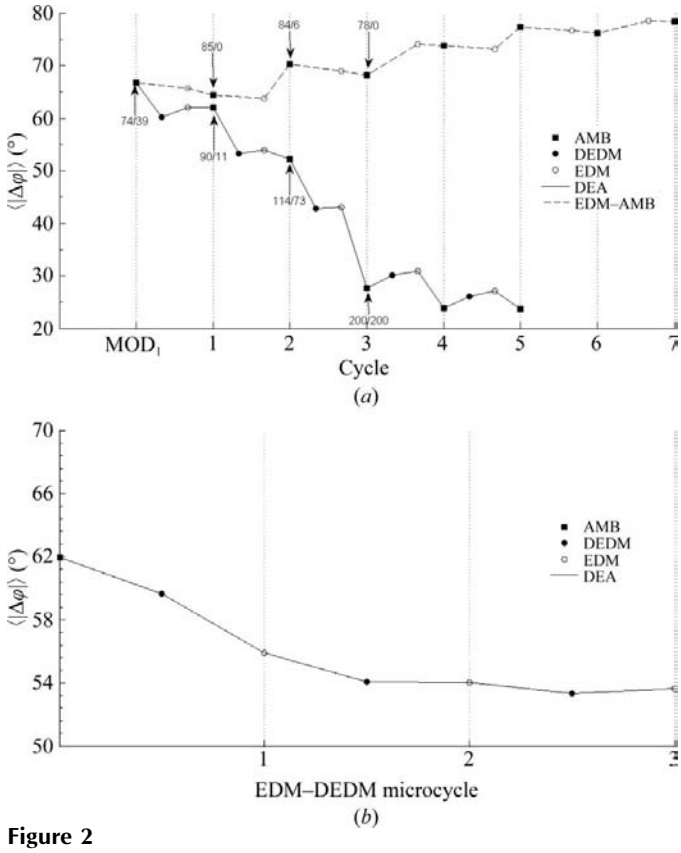

Figure 2
(*a*) 1ujz with EDM–AMB and DEA procedures. (*b*) 1ujz: enlargement of DEA cycle 2. Three EDM–DEDM microcycles are performed in cycle 2.

For the structure 1yxa (NRES = 742), the model (1qlp) consists of 372 residues and has a homodimeric structure. The root-mean-square distance between the C^α atoms of the model and target structure is 1.68 Å, while the sequence identity is 46%. In this case, the best results were obtained by using the *PHENIX AutoBuild* wizard as an AMB program: it produced a starting model (MOD_1) with 529 residues (but only 47 of these had side chains). Application of the EDM–AMB procedure did not provide any improved model, even after several cycles. DEA provided a model with 693 residues (640 of which had side chains) after nine cycles.

4.2. AbIM cases

We have applied the DEA approach to some challenging cases recently described by Rigden *et al.* (2008), who searched the PDB (Berman *et al.*, 2007) for proteins of fewer than 100 residues, deposited not earlier than 2006, with RES < 2.2 Å and with a sequence identity to previously determined structures of not greater than 30%. For each of the selected cases at least 3000 models were produced and clustered by *ROSETTA*, a program for predictive modelling of proteins (Simons *et al.*, 1997, 1999; Shortle *et al.*, 1998). Secondary-structure predictions were provided by *PSIRED* (Jones, 1999), side chains were added to the polyalanine models using *SCWRL* (Dunbrack & Cohen, 1997) and energy minimization was performed using *MODELLER* (Sali & Blundell, 1993). Only in five of the 16 selected cases (PDB codes 2pmr, 2nn4, 2fzt, 2o3l and 2duy) were correct MR solutions obtained by *Phaser* (McCoy *et al.*, 2007) and for only two of them (2pmr and 2nn4) was *ARP*/*wARP* able to automatically build a well docked model (74 of 76 and 174 of 186 residues were docked, respectively). We then used 2fzt, 2o3l and 2duy as test structures for DEA by using both *ARP*/*wARP* and the *PHENIX AutoBuild* wizard as the AMB program.

For 2fzt (NRES = 156 and RES = 2.05 Å) we used one of the nine polyalanine models used by Rigden *et al.* (2008), which has a total of 112 residues. Similar to Rigden and coworkers, we did not succeed in building an improved model using *ARP*/*wARP* or the *PHENIX AutoBuild* wizard: 69 and 112 residues were built, respectively, but both were unable to dock any significant portion of the sequence. The same result was obtained by application of the EDM–AMB procedure. In contrast, DEA (with the *PHENIX AutoBuild* wizard as the AMB program) automatically generated after four cycles a 150-residue model with 145 residues docked in the sequence, corresponding to $\langle|\Delta\varphi|\rangle_f = 31^\circ$ and $\mathrm{CORR}_f = 0.91$. We did not succeed with the other two test structures 2o3l and 2duy: the best values for $\langle|\Delta\varphi|\rangle_1$ were 85° and 88°, respectively, which were too large for the present DEA procedure.

4.3. AbI cases

The starting sets of phases for 2f14, 1arm and 1yfd were provided by *ab initio* phasing *via* Patterson methods followed

by EDM techniques according to the program *SIR*2008 (see Caliandro, Carrozzini, Cascarano, De Caro, Giacovazzo, Mazzone *et al.*, 2008), which is now included in *IL MILIONE*. It is worthwhile noticing that these test structures were originally solved by different approaches: MR for 2f14 and 1yfd, and SIR for 1arm.

For the phase sets obtained by AbI phasing the average phase errors were 51, 52 and 52°, respectively, and the corresponding electron-density maps showed correlation values with the published maps of 0.69, 0.66 and 0.66, respectively. Such good parameters should not deceive: indeed, a large part of the correlation arises from the correct location of the heavy atoms, while the rest of the structure is weakly correlated with the true structure. As an example, we show in Fig. 3(*a*) for 1yfd the electron-density map at 1.5σ near two of the 13 mercury ions contained in the asymmetric unit (the published model is represented by backbones). This lack of correlation agrees well with the poor figures characterizing MOD_1 in Table 1 (see $SIZE_1$ and $NDOCK_1$): it is thus not surprising that for all three test structures any attempt to build a significant portion of the molecule *via* the EDM–AMB procedure failed, despite a large number of efforts.

For 2f14 (NRES = 257, RES = 1.71 Å) DEA provided a very good model of 227 residues (222 of which were docked in the sequence) after 11 cycles using *ARP*/*wARP* as the AMB program.

For 1arm and 1yfd, DEA was unable to provide significant models when *ARP*/*wARP* was used as the AMB program. We then employed the *PHENIX AutoBuild* wizard and only obtained satisfactory results by using the option (input_ha_file) for the active use of the heavy-atom substructure in the EDM step: the program truncates the electron density near the heavy-atom positions (provided by *IL MILIONE*) at a maximum of 2.5σ. Using this approach for 1arm (NRES = 307 residues, RES = 1.76 Å) we obtained a 263-residue model all docked in the sequence after ten cycles and for 1yfd (NRES = 681 residues, RES = 1.90 Å) a 588-residue model all docked in the sequence was obtained after 13 cycles.

Figs. 3(*b*), 3(*c*) and 3(*d*) concern the same 1yfd unit-cell region as shown in Fig. 3(*a*). The electron-density maps are all

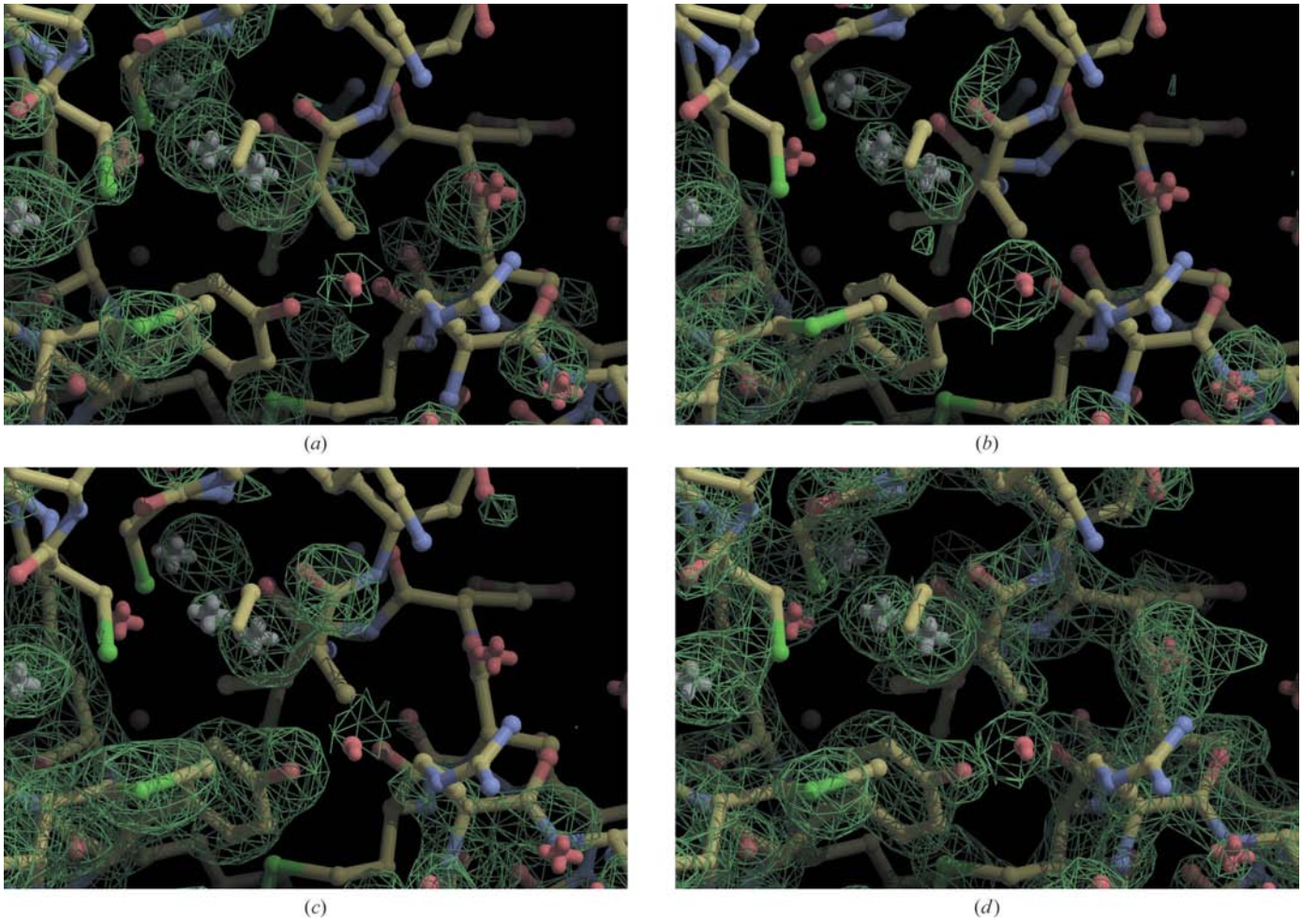

Figure 3
1yfd: unit-cell region near Hg612, Hg611 and residues Ala265 and Asp285 (both on chain *B*). The published protein structure from the PDB is shown as backbones and the current electron-density map is drawn at 1.5σ using *Coot* (Emsley & Cowtan, 2004): (*a*) after *ab initio* phasing, (*b*) after the *PHENIX AutoBuild* wizard from the *ab initio* map (ρ_{mod}), (*c*) after five DEA cycles and (*d*) at the end of DEA.

drawn at 1.5σ: in Fig. 3(*b*) after the first application of the *PHENIX AutoBuild* wizard (to obtain MOD_1), in Fig. 3(*c*) after five DEA cycles (model consisting of 314 residues, with 142 residues docked in the sequence) and in Fig. 3(*d*) at the end of the DEA procedure (model consisting of 588 residues, all docked in the sequence).

4.4. SAD/MAD cases

The starting sets of phases for 1svn and 1lvo were provided by the SAD/MAD phasing routines included in *IL MILIONE* and improved *via* EDM cycles: the average phase error was 53° for 1svn and 65° for 1lvo and the correlation values of the corresponding electron-density maps with the published maps were 0.63 and 0.71, respectively. The figures characterizing MOD_1 as obtained by *ARP*/*wARP* were discouraging for the 1svn case (12 residues docked in a model of 93 residues) and encouraging for 1lvo (310 residues docked in a model of 900 residues). Any attempt to improve the models *via* iterated use of *ARP*/*wARP* failed.

For 1svn, EDM–*ARP*/*wARP* provided a very good model after ten cycles (250 docked residues in a model of 250 residues): the same result was obtained by DEA in only three cycles.

For 1lvo, EDM–*ARP*/*wARP* did not improve the initial model, even after several cycles. With DEA, 1003 residues of a 1309-residue model were docked in the sequence after 18 cycles, corresponding to $\langle|\Delta\varphi|\rangle_f = 42°$ and $CORR_f = 0.83$.

5. Conclusions

Modern crystallography may use different techniques for phasing proteins (*e.g. ab initio* phasing, SAD/MAD, SIR/MIR, MR or *ab initio* modelling), several tools for extending and improving phases (*i.e.* EDM, DEDM and hybrid direct methods) and automatic procedures to build and refine structural models. The final aim of the methodologies is to reduce the manual effort as much as possible in all steps of the phasing process. This paper indicates how EDM, DEDM and AMB programs may be combined into a single procedure (DEA) to automate phase-refinement and model-building steps: it has been shown that DEA succeeds when different combinations of the single programs fail. In particular, it is shown that the iterated use of particular combinations of available tools may greatly increase the efficiency of the structure-solution process. From the point of view of execution time, DEA shows a supplementary practical advantage. Most of the computing time is spent on the iterated application of the AMB programs: the use of the EDM–DEDM cycles reduces the number of AMB applications and thus dramatically reduces the total computing time.

DEA has been included in the package *IL MILIONE*, which may be integrated by external AMB programs *via* suitable scripts. As an outlook, we foresee that embedding DEDM cycles within AMB procedures will greatly increase their efficiency.

We would like to thank the CASPUR Consortium for technical and computational support, Dr Hai-fu Fan for providing the model used to test the *OASIS* procedure and Dr Daniel John Rigden for providing the structural models generated by *ROSETTA*.

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.
Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Alterio, V., Vitale, R. M., Monti, S. M., Pedone, C., Scozzafava, A., Cecchi, A., De Simone, G. & Supuran, C. T. (2006). *J. Am. Chem. Soc.* **128**, 8329–8335.
Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J. & Hilgenfeld, R. (2002). *EMBO J.* **21**, 3213–3224.
Berman, H., Henrick, K., Nakamura, H. & Markley, J. L. (2007). *Nucleic Acids Res.* **35**, D301–D303.
Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
Bradley, P., Misura, K. M. & Baker, D. (2005). *Science*, **309**, 1868–1871.
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., Mazzone, A. & Siliqi, D. (2008). *J. Appl. Cryst.* **41**, 548–553.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2008). *Acta Cryst.* A**64**, 519–528.
Caliandro, R., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Mazzone, A. M. & Siliqi, D. (2009). *Acta Cryst.* D**65**, 249–256.
Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
Cowtan, K. (2006). *Acta Cryst.* D**62**, 1002–1011.
Dobbs, A. J., Anderson, B. F., Faber, H. R. & Baker, E. N. (1996). *Acta Cryst.* D**52**, 356–368.
Dunbrack, R. L. Jr & Cohen, F. E. (1997). *Protein Sci.* **6**, 1661–1681.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Giacovazzo, C. & Siliqi, D. (1997). *Acta Cryst.* A**53**, 789–798.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.
Greenblatt, H. M., Feinberg, H., Tucker, P. A. & Shoham, G. (1998). *Acta Cryst.* D**54**, 289–305.
He, Y., Yao, D.-Q., Gu, Y.-X., Lin, Z.-J., Zheng, C.-D. & Fan, H.-F. (2007). *Acta Cryst.* D**63**, 793–799.
Horvath, A. J., Irving, J. A., Rossjohn, J., Law, R. H., Bottomley, S. P., Quinsey, N. S., Pike, R. N., Coughlin, P. B. & Whisstock, J. C. (2005). *J. Biol. Chem.* **280**, 43168–43178.
Jamrog, D. C., Zhang, Y. & Phillips, G. N. (2003). *Acta Cryst.* D**59**, 304–314.
Jones, D. T. (1999). *J. Mol. Biol.* **292**,195–202.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kolberg, M., Logan, D. T., Bleifuss, G., Potsch, S., Sjoberg, B. M., Graslund, A., Lubitz, W., Lassmann, G. & Lendzian, F. (2005). *J. Biol. Chem.* **280**, 11233–11246.
Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). *Nature Struct. Mol. Biol.* **11**, 371–379.
Levitt, D. G. (2001). *Acta Cryst.* D**57**, 1013–1019.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007). *Nature (London)*, **450**, 259–264.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Refaat, L. S. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 367–371.
Rigden, D. J., Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 1288–1291.
Sali, A. & Blundell, T. L. (1993). *J. Mol. Biol.* **234**,779–815.
Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.
Shortle, D., Simons, K. T. & Baker, D. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 11158–11162.
Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. (1997). *J. Mol. Biol.* **268**, 209–225.
Simons, K. T., Ruczinski, I., Kooperberg, C., Fox, B. A., Bystroff, C. & Baker, D. (1999). *Proteins*, **34**, 82–95.
Terwilliger, T. C. (2003*a*). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. (2003*b*). *Acta Cryst.* D**59**, 1174–1182.
Terwilliger, T. C. (2004). *J. Synchrotron Rad.* **11**, 49–52.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Turk, D. (1992). PhD thesis. Technische Universitat München, Germany.
Uhrinova, S., Uhrin, D., Nairn, J., Price, N. C., Fothergill-Gilmore, L. A. & Barlow, P. N. (2001). *J. Mol. Biol.* **306**, 275–290.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Xu, H. & Hauptman, H. A. (2006). *Acta Cryst.* D**62**, 897–900.
Yao, J.-X. (1981). *Acta Cryst.* A**37**, 642–644.
Yao, J.-X. (2002). *Acta Cryst.* D**58**, 1941–1947.
Zhang, K. Y. J., Cowtan, K. D. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by E. Arnold & M. G. Rossmann, pp. 311–331. Dordrecht: Kluwer Academic Publishers.
Zhang, T., He, Y., Gu, Y.-X., Zheng, C.-D., Hao, Q., Wang, J.-W. & Fan, H.-F. (2007). *OASIS06: A Direct-Method Program for SAD/SIR Phasing and Reciprocal-Space Fragment Extension.* Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China.