Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Phasing at resolution higher than the experimental resolution

Rocco Caliandro,[a] Benedetta Carrozzini,[a] Giovanni L. Cascarano,[a] Liberato De Caro,[a] Carmelo Giacovazzo[a,b]* and Dritan Siliqi[a]

[a]Istituto di Cristallografia, CNR, Via Amendola 122/o, 70126 Bari, Italy, and [b]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy

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

Received 16 December 2004
Accepted 4 February 2005

Limited experimental resolution is a unavoidable feature in macromolecular crystallography: it may hinder or make difficult the determination of the crystal structure. A novel procedure is presented which from an approximate electron-density map extrapolates the moduli and phases of non-measured reflections beyond and behind the experimental resolution limit. Applications to a set of test structures show that the extrapolation can be successfully accomplished. As a consequence, the phase estimates of the observed reflections are subsequently improved and the interpretability of the corresponding electron-density map increases. The use of the extrapolated values for the non-measured reflections provides additional information for the map, which shows a resolution higher than the experimental resolution.

1. Notation

EDM, electron-density modification.
NMRE, non-measured reflection extrapolation.
$\mathrm{RES_{obs}}$, resolution limit of the experimental data in angstroms.
$\mathrm{RES_{ext}}$, resolution limit of the reflections estimated by the NMRE procedure.
MPE, mean phase error of a given set of phases with respect to the refined (published) model. $\mathrm{MPE_{obs}}$ refers to the observed reflections and $\mathrm{MPE_{ext}}$ to the extrapolated reflections.
MIS, percentage of non-measured reflections up to $\mathrm{RES_{obs}}$.
GEN, the ratio number of extrapolated reflections/number of observed reflections.
Residues, number of residues in the protein molecule.
$|F_{\mathbf{h}}^{\mathrm{obs}}|$, observed structure-factor modulus; $|E_{\mathbf{h}}^{\mathrm{obs}}|$ is its normalized value.
$|F_{\mathbf{h}}^{\mathrm{ext}}|$, structure-factor modulus estimated by the NMRE procedure for extrapolated reflections; $|E_{\mathbf{h}}^{\mathrm{ext}}|$ is its normalized value.
$|F_{\mathbf{h}}^{\mathrm{true}}|$, structure-factor modulus of non-measured reflections as calculated from the published structure model.
$|E_{\mathbf{h}}^{\mathrm{calc}}|$, normalized structure-factor modulus estimated by the NMRE procedure for observed reflections.
$D_1(x) = I_1(x)/I_0(x)$, where $I_i(x)$ is the modified Bessel function of order i.

2. Introduction

The inability to collect diffraction data to sufficiently high resolution is a limit to the efficiency of all phasing procedures (irrespective of whether direct or Patterson techniques, SAD/MAD or SIR/MIR methods are used) and for the structure-refinement process (*e.g.* irrespective of whether this takes place *via* least-squares or *via* EDM techniques). Indeed, low data resolution implies a low quantity of information obtained

by the diffraction experiment and therefore objective difficulties in achieving interpretable electron-density maps.

A shortage of experimental data is usual and often critical in macromolecular crystallography, but may also occur for small molecules when the diffraction sample is of poor quality. Attempts have been made previously to extrapolate the experimental data beyond the observed range. Karle & Hauptman (1964) suggested the use of a sharpened origin-removed Patterson function on the basis that it is non-negative. Seeman *et al.* (1976) reported the use of the Karle and Hauptman technique to obtain improved estimates of the normalized structure factors within the set of measured data. Langs (1998) extended the method to macromolecular crystallography: he described some protocols to extend observed data from 1.0 to ∼0.5 Å resolution and applied them to two test structures (gramicidin A, with 36 residues, and scorpion toxin-II, with 64 residues).

A shortage of experimental data is more frequent in powder crystallography, where peak overlapping does not allow accurate estimation of the diffraction intensities. It has been shown that the accuracy of the experimental pattern decomposition (which is critical for the success of the phasing techniques) may benefit from using different sources of prior information; for example, the positivity of the integrated intensities (Jansen *et al.*, 1992; Sivia & David, 1994) or the Patterson positivity (Estermann & Gramlich, 1993; Altomare *et al.*, 1998). Improved estimates of the intensities of severely overlapping reflections can also be obtained *via* reciprocal-space relationships. Van der Putten *et al.* (1982) used a probabilistic expression for estimating $|E_{\mathbf{h}}|$ from the most reliable quartets in which **h** is a cross-term. David (1987) suggested two techniques: to maximize the entropy of the Patterson function under the constraint of the measured intensities and to square (Sayre, 1952) the Patterson map. His conclusive formula was

$$|F_{\mathbf{h}}|^2 = \sum_{\mathbf{k}} |F_{\mathbf{k}}|^2 |F_{\mathbf{h}-\mathbf{k}}|^2. \qquad (1)$$

Cascarano *et al.* (1991) suggested, *via* the application of joint probability distribution functions, replacing (1) by

$$|E_{\mathbf{h}}|^2 = 1 + N^{-1} \sum_{\mathbf{k}} (|E_{\mathbf{k}}|^2 - 1)(|E_{\mathbf{h}-\mathbf{k}}|^2 - 1). \qquad (2)$$

More recently, Xu & Hauptman (2000) suggested using the correlation coefficient

$$\frac{\langle |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \rangle - \langle |E_{\mathbf{k}}| \rangle \langle |E_{\mathbf{h}-\mathbf{k}}| \rangle}{(\langle |E_{\mathbf{k}}|^2 \rangle - \langle |E_{\mathbf{k}}| \rangle^2)^{1/2} (\langle |E_{\mathbf{h}-\mathbf{k}}|^2 \rangle - \langle |E_{\mathbf{h}-\mathbf{k}}| \rangle^2)^{1/2}} \qquad (3)$$

to estimate diffraction moduli. Its application to some macromolecular structures shows that (3) estimates moduli more accurately than (1).

All the techniques mentioned above aim at improving the estimates of the diffraction moduli beyond the resolution limit of the experimental data. The extrapolated moduli, mixed (with some supplementary care) with the observed ones, should increase the number of reliable structure invariants and seminvariants to be used in tangent procedures and should therefore make the crystal structure solution more straightforward. However, implementation of all the above techniques in well documented programs routinely devoted to phasing macromolecular structures has still not been accomplished.

Extrapolation of moduli and phases is possible when some phase information is available for the measured reflections. This may be accomplished in several ways.

(i) In reciprocal space. For instance, *via* the Sayre equation (Sayre, 1952)

$$F(\mathbf{h}) = [\theta(\mathbf{h})/V] \sum_{\mathbf{k}} F(\mathbf{k}) F(\mathbf{h}-\mathbf{k}),$$

where $\theta(\mathbf{h}) = f(\mathbf{h})/g(\mathbf{h})$, $f(\mathbf{h})$ is the atomic scattering factor for the generic atom and $g(\mathbf{h})$ is the scattering factor of the squared atom (it is supposed that the electron density consists solely of equal atoms). The free vector **k** runs over the set of observed (and phased) reflections and the index **h** refers to unobserved or non-phased reflections. The Sayre equation has been applied by Sayre (1974) to a small protein and by Main (1990) in order to refine phases. No attempt to extrapolate reflections beyond the data resolution was attempted.

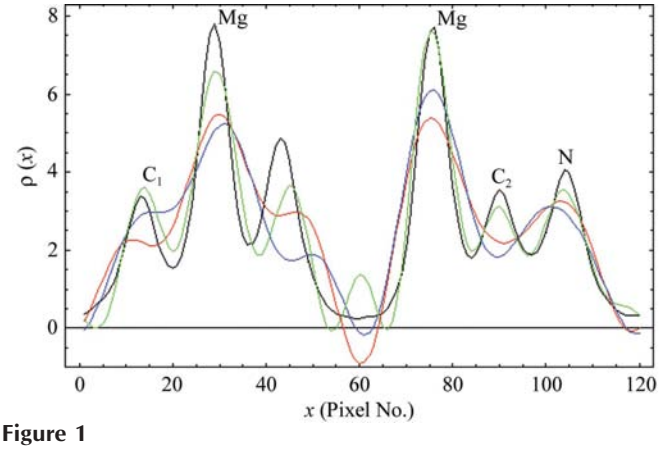

Figure 1
$\rho_{\rm true}$ (black line) and $\rho_{1.7}$ (red line) are sampled at 120 grid points. The electron density produced after 15 cycles of EDM is represented by the blue line and the electron density obtained after 15 cycles of EDM–NMRE is represented by the green line.

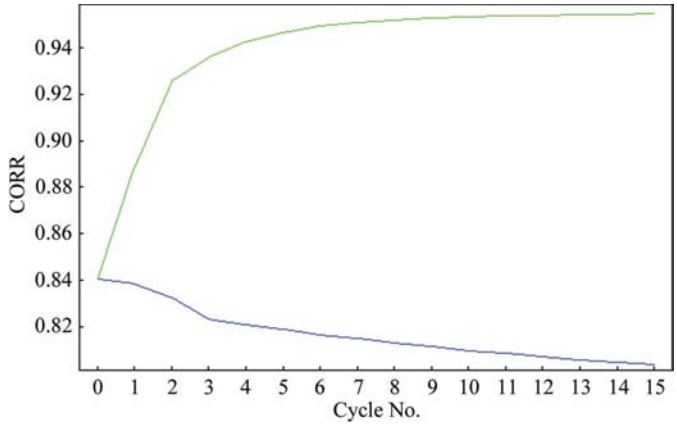

Figure 2
The value of CORR *versus* the cycle number when EDM (blue line) and EDM–NMRE (green line) procedures are applied.

In high-resolution electron microscopy, it is common practice to extract higher resolution structural information from low-resolution images *via* phase-extension procedures, which rely on an additional measured electron-diffraction pattern with higher resolution. Fan & Zheng (1975) proposed a method using the Sayre equation to extrapolate both the phases and magnitudes of structure factors and successfully applied the technique to a small molecule (Liu *et al.*, 1988).

(ii) By maximum-entropy techniques (Bricogne, 1984). These have been applied to macromolecular crystallography in several contexts [*e.g.* by Carter *et al.* (1990) to fix the envelope, by Schiltz *et al.* (1997) for maximum-likelihood refinement, by Roversi *et al.* (1998) for accurate studies of charge densities and by Voigt-Martin *et al.* (1995) in electron crystallography], but in practice have never been applied to extrapolate reflections beyond data resolution.

(iii) By EDM techniques. Any modification of the electron density is able to extrapolate moduli and phases: the accuracy of the estimates depends on the modification algorithm.

In this paper and in a subsequent one, we will show that it is possible to systematically extrapolate moduli and phases beyond and behind (*e.g.* for non-measured low or very low resolution reflections) the observed resolution, with a large advantage for the phasing process: this will therefore provide electron-density maps that are more interpretable and resolved than those obtainable *via* the measured reflections only. The papers will only concern macromolecular crystallography, where the following three typical situations could be encountered.

(i) *Ab initio* phasing: $\mathrm{RES_{obs}}$ in the interval 1.5–1.0 Å, an approximated electron density available, with $\mathrm{MPE_{obs}}$ in the range 25–60°.

(ii) SAD/MAD, SIR/MIR or SIRAS/MIRAS phases: $\mathrm{RES_{obs}}$ in the interval 2.8–1.5 Å, an approximated electron density available (*e.g.* after the application of EDM procedures), with $\mathrm{MPE_{obs}}$ in the range 40–65°.

(iii) *Ab initio* phasing, $\mathrm{RES_{obs}}$ in the interval 1.5–1.0 Å, no phase information available.

In all these cases, the ideal extrapolation procedure is expected to reduce the phase error of the measured reflections, to provide sensible estimates (of modulus and phase) for some additional reflections behind and beyond $\mathrm{RES_{obs}}$ and to increase the interpretability of the final electron-density map. In this paper, we will describe a general extrapolation technique suitable for the cases (i) and (ii); case (iii) will be treated in a subsequent contribution.

3. A preliminary example

The following simple preliminary example aims to introduce the concept that extrapolation of the experimental diffraction moduli and of the relative estimated phases can increase the quality of the electron-density map. It will also help us to make clear the mechanism responsible for the gain of information and to describe the general features of the new map.

We will combine a simple EDM technique with an NMRE procedure in order to demonstrate that this combination is able to improve the features of an electron-density map. We will study the simple case of a one-dimensional structure, with $a = 10$ Å, containing two Mg, one O, one N and two C atoms. The exact distribution of the electron density ($\rho_{\rm true}$) is represented in Fig. 1 by a black line sampled at 120 grid points. Suppose that the correct molecular model has been obtained *via* experimental data with resolution to 0.5 Å: let $\{|F_{\mathbf{h}}^{\rm obs}|\}_{0.5}$ and $\{\varphi_{\rm true}\}_{0.5}$ be the corresponding sets of measured moduli and calculated phases. Their Fourier transform will practically coincide with $\rho_{\rm true}$.

Let us now suppose that the available data resolution is 1.7 Å (let $\{|F_{\mathbf{h}}^{\rm obs}|\}_{1.7}$ be the set of measured reflections) and that the information provided by the experimental data, combined with some stereochemical prior information, leads us to the same molecular model obtained from the data at 0.5 Å resolution. The best electron-density distribution we can obtain by using data truncated at 1.7 Å ($\rho_{1.7}$), is that using $\{|F_{\mathbf{h}}^{\rm obs}|\}_{1.7}$ and $\{\varphi_{\rm true}\}_{1.7}$, which is shown in Fig. 1 by a red line. $\rho_{1.7}$ correctly locates the two Mg atoms and shows very faint peaks connected with the N, O and C_1 sites, but it has a minimum in the C_2 position and presents a region with negative electron density.

We now verify whether a simple EDM algorithm based on the atomicity and on the positivity of the electron density can improve the interpretability of $\rho_{1.7}$, even at the expense of the phase correctness. We use the following algorithm: at the jth cycle the electron density is modified according to

$$\rho_{\rm mod}(x) = \begin{cases} 0 & \text{if } \rho(x) < 0 \ \forall\, j \\ \rho(x)^{1.3} & \text{if } \rho(x) > 0 \text{ and } j \leq 3 \\ \rho(x) & \text{if } \rho(x) > 0 \text{ and } j > 3. \end{cases} \qquad (4)$$

The first of the conditions in (4) applies the positivity criterion and the second makes the atomic electron densities sharper in order to contrast the effect of the resolution limit. After 15 cycles of $\rho\rightarrow\varphi\rightarrow\rho$, the resulting electron density is represented by the blue curve in Fig. 1: it does not show the missed C_2 atom and is a rather distorted representation of $\rho_{\rm true}$. Increasing the number of cycles increases the overall distortion, which can be deduced from Fig. 2, where CORR is plotted (blue curve) against the cycle number. CORR is the correlation factor between $\rho_{\rm true}$ and the current electron density $\rho_{\rm cur}$, given by

$$\mathrm{CORR} = \frac{\langle \rho_{\rm true} \cdot \rho_{\rm cur} \rangle - \langle \rho_{\rm true} \rangle \cdot \langle \rho_{\rm cur} \rangle}{(\langle \rho_{\rm true}^2 \rangle - \langle \rho_{\rm true} \rangle^2)^{1/2} \cdot (\langle \rho_{\rm cur}^2 \rangle - \langle \rho_{\rm cur} \rangle^2)^{1/2}}. \qquad (5)$$

Let us now perform 15 $\rho\rightarrow\varphi\rightarrow\rho$ cycles by combining EDM with NMRE. In each half-cycle $\rho\rightarrow\varphi$ the electron density is modified according to (4) and, by Fourier inversion, moduli and phases are extrapolated up to 1.0 Å: the initial ρ map is $\rho_{1.7}$, calculated from measured moduli and true phases. In each $\varphi\rightarrow\rho$ half-cycle the electron density is calculated using the measured moduli and current phases for reflections up to $\mathrm{RES_{obs}} = 1.7$ and the calculated moduli and current phases for the extrapolated reflections. The final electron-density distribution is shown in Fig. 1 by the green curve. Its comparison with $\rho_{\rm true}$ suggests that the new procedure (i) produces a higher resolution map (peaks are more resolved than in $\rho_{1.7}$),

(ii) leads to an electron density that is much more interpretable in terms of atomic positions, (iii) shows maxima corresponding to the sites of all the atoms and, in particular, correctly locates the missed C_2 atom and (iv) produces false but low-intensity peaks.

The algorithm is stable (enlarging the number of cycles does not substantially improve or deteriorate the map), as may be seen from Fig. 2, where the trend in CORR (green curve) is shown *versus* the number of cycles. The higher interpretability of the map is obtained at the expense of the phase error (*i.e.* $\mathrm{MPE_{obs}} = 10^\circ$ and $\mathrm{MPE_{ext}} = 30^\circ$) and in terms of errors in the estimated moduli for the extrapolated reflections. For these, the crystallographic residual

$$R_{\mathrm{cryst}} = \frac{\sum_{\mathbf{h}} \left| |F_{\mathbf{h}}^{\mathrm{true}}| - |F_{\mathbf{h}}^{\mathrm{ext}}| \right|}{\sum_{\mathbf{h}} |F_{\mathbf{h}}^{\mathrm{true}}|} \qquad (6)$$

is 0.30.

The use of different EDM algorithms will produce different results; however, the above example proves the potential usefulness of the extrapolated reflections in the phasing process. Additional tests, not shown for brevity, suggest that (i) extrapolation may be more difficult when the data resolution is lower (for example, we do not succeed in locating the missed C when the experimental data resolution is 2.0 Å and/or reflections are extrapolated to 0.5 Å) and (ii) enlarging the extrapolation interval too much makes the extrapolated moduli and phases at higher resolution affected by large errors.

The two points (i) and (ii) may be explained as follows. High data resolution corresponds to a large amount of information gained by the diffraction experiment. In this case, if the phasing process is able to provide sufficiently good phases, the missed structural information is a small percentage of that available at the end of the phasing process and may be (at least partially) recovered by suitable EDM techniques which are able to produce new electron densities closer to that corresponding at zero resolution. If the data resolution is low (*e.g.* 4 Å), the quantity of information gained by the experiment is low: the missed structural information at the end of the phasing process is relatively high and extrapolation by EDM is expected to be less efficient, particularly when the extrapolation interval is high.

4. The algorithm

The example described in §3 does not completely fit the cases usually encountered in crystallography. Indeed, in §3 we supposed that $\mathrm{MPE_{obs}}$ is zero, while it is frequently in the range 25–60° for *ab initio* phasing or in the range 40–65° for SAD/MAD, SIR/MIR or SIRAS/MIRAS phasing. In addition, in real cases the completeness of the data set is seldom reached and missing reflections are usually present within the experimental resolution limit. Extrapolation will therefore be a more difficult process, requiring efficient EDM and NMRE algorithms. They should be applied to the electron density available at the end of the phasing process (the best attained using current techniques) based on measured reflections only. The application of the EDM algorithm allows the extrapolation of reflections beyond and behind $\mathrm{RES_{obs}}$, including non-measured reflections lying below $\mathrm{RES_{obs}}$. Part of the extrapolated reflections, together with the observed ones, are used to calculate a new electron-density map, which is again submitted to the EDM algorithm.

The procedure is performed in two steps, each one including a number of cycles which may be represented by $\rho \rightarrow \varphi \rightarrow \rho$. In the first step, consisting of 50 cycles, the NMRE process plays a major role, with extrapolated reflections progressively added to the measured ones; the EDM criteria governing the EDM process are kept fixed. In the second step, 20 further EDM cycles are performed; the number of considered reflections is kept constant, but the EDM criteria are continuously varied. At the end of the procedure, the resolution limit of the observed and extrapolated reflections is $\mathrm{RES_{ext}} < \mathrm{RES_{obs}}$.

Let us consider the first step of the procedure, as designed for *ab initio* phasing ($\mathrm{RES_{obs}}$ in the interval 1.5–1.0 Å, an approximate electron density available). In accordance with Langs (1998), in the half-cycle $\rho \rightarrow \varphi$ we found it advantageous to extrapolate all non-measured reflections in one step from $\mathrm{RES_{obs}}$ to $\mathrm{RES_{ext}}$, rather than to increase the extrapolation resolution gradually. However, not all the extrapolated reflections are used in the half-cycle $\varphi \rightarrow \rho$, but only a percentage of them, which increases with the cycle number (it ranges from 10 to 75% of the number of measured reflections). The selection of the extrapolated reflections is performed on the basis of their moduli $|E_{\mathbf{h}}^{\mathrm{ext}}|$ estimated by map inversion. In fact, the largest modulus reflections strongly influence the quality of the electron-density map, can be phased with larger accuracy and are able to pilot the subsequent extrapolation. On the other hand, an excessive number of actively used extrapolated reflections could corrupt the initial observed reflections phase set, so that it will not exceed 75% of the number of observed reflections.

Other features of the procedure are as follows.

(i) In the half-cycle $\varphi \rightarrow \rho$ the Fourier coefficients are $|E_{\mathbf{h}}^{\mathrm{obs}}|$ for observed reflections and the estimated moduli for the extrapolated reflections. The latter are truncated to their expected average value $\langle |E_{\mathbf{h}}^{\mathrm{ext}}| \rangle = 0.886$, since we have verified that phases are more reliably estimated than moduli.

(ii) In the half-cycle $\rho \rightarrow \varphi$ only a fraction of ρ corresponding to 10% of the volume occupied by the protein is used in each map inversion.

(iii) The $|E_{\mathbf{h}}^{\mathrm{ext}}|$ values obtained after each map inversion are rescaled according to the distribution of normalized structure factors expected for a random-atom structure.

(iv) A Sim-like weight is associated with each reflection: $w_{\mathbf{h}} = D_1(k|E_{\mathbf{h}}^{\mathrm{obs}}||E_{\mathbf{h}}^{\mathrm{calc}}|)$ for an observed reflection and $w_{\mathbf{h}} = D_1(k|E_{\mathbf{h}}^{\mathrm{ext}}|^2)$ for an extrapolated reflection. k is an empirical constant set to 0.5. The above scheme does not coincide with the classical Sim weighting and is based on our past experience of *ab initio* phasing.

(v) In order to prevent extrapolated reflections with large modulus predominating in the phasing process, stalling the

Table 1
Code names and crystallochemical data for protein test structures.

SEQ is a sequential number for the structures sorted with respect to RES_{obs}, PDB is the file code in the Protein Data Bank, Residues is the number of residues and MIS is the percentage of missing reflections inside the experimental resolution limit. MPE_{obs} is the mean phase error for observed reflections (the starting set of phases to which the EDM–NMRE will be applied).

SEQ	Structure code	PDB	Residues	RES_{obs} (Å)	MIS (%)	MPE_{obs} (°)	Reference
1	Cyto549	1e29	135	1.21	4.1	34	Frazao *et al.* (2001)
2	Bpti9	9pti	58	1.22	11.5	24	C. Eigenbrot, M. Randall & A. A. Kossiakoff, unpublished work
3	Dna145d	145d	24	1.27	0.8	60	Schroth *et al.* (1993)
4	Cupre	1aac	105	1.30	2.3	30	Durley *et al.* (1993)
5	Hewl133	193l	129	1.33	9.0	31	Vaney *et al.* (1996)
6	Wild	1fs3	124	1.35	1.6	25	Chatani *et al.* (2002)
7	Buto140	3ebx	62	1.40	0.7	46	Smith *et al.* (1997)
8	Hewl140	194l	129	1.40	9.2	35	Vaney *et al.* (1996)
9	Vitad3	1ie9	259	1.40	23.8	57	Berkowitz *et al.* (2002)
10	Cole	1lri	98	1.45	1.2	35	Lascombe *et al.* (2002)
11	Adeny	1fx2	235	1.46	4.7	58	Bieger & Essen (2001)
12	Dnajes	1jes	24	1.49	2.7	57	Atwell *et al.* (2001)
13	Pazur	1paz	123	1.55	3.3	34	Petratos *et al.* (1988)
14	Kpr	1ks9	291	1.70	4.3	58	Silinski *et al.* (2001)
15	Sav3	1svn	269	1.74	0.8	51	Betzel *et al.* (1988)
16	Haptbr	1fj2	464	1.80	0.3	54	Devedjiev *et al.* (2000)
17	Mdd	1fi4	832	2.28	0.2	61	Bonanno *et al.* (2001)
18	Idi	1i9a	364	2.40	4.4	54	Bonanno *et al.* (2001)

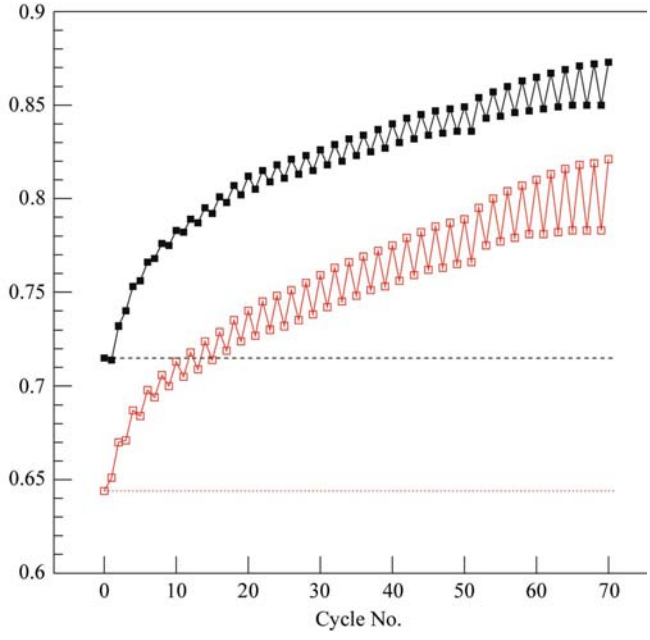

Figure 3
Buto140: $CORR_{obs}$ (full squares) and $CORR_{tot}$ (open squares) *versus* the cycle number when EDM–NMRE procedures are applied.

EDM–NMRE procedure, the distribution of weights was dynamically modified during the procedure. Specifically, the observed reflection weights were raised to the power $\langle|E_{\mathbf{h}}^{ext}|\rangle/\langle|E_{\mathbf{h}}^{calc}|\rangle$: this ratio tends to decrease with the cycle number as long as extrapolated reflections with lower moduli are fed in the procedure and it is mostly lower than one. This operation, which also allows the reduction of the impact of the newcomer extrapolated reflections onto those already phased, is performed every two cycles.

(vi) A substantial gain in efficiency is obtained by calculating the molecular envelope in the half-cycle $\varphi\rightarrow\rho$ and by using it as a mask in the following half-cycle (Wang, 1985; Leslie, 1987). The calculation includes all the reflections phased in the current cycle (hence also the extrapolated reflections, particularly those at very low resolution) and is performed using a sphere of varying radius, which decreases with the cycle number from the value RES_{obs} + 6 to RES_{ext}.

In the second step of the procedure the fraction of ρ used in each map inversion varies from 10 to 30% of the protein volume, depending on the cycle number. The following procedures were performed.

(i) To reduce the impact of the background, the pixel intensity is halved if it is below one standard deviation of the whole electron-density map.

(ii) To limit the overvaluing of large-moduli reflections, every two cycles the map is truncated to a threshold value which ranges from five to ten times the standard deviation, depending on the cycle number.

(iii) The molecular envelope is not applied and the exponent used for the modification of weights is decreased from its last value in the first step to 0.5 in order to enhance the contribution of lower weight reflections.

Let us now consider how the procedure has been modified to handle the case in which experimental (SAD/MAD, SIR/MIR or SIRAS/MIRAS) phases are available. Then, owing to the lower experimental data resolution (RES_{obs} supposed to be in the interval 2.8–1.5 Å), the risk that extrapolated reflections can corrupt the starting phase set is high. To overcome this tendency, current phases of the observed reflections are combined with their 'experimental' values, using a relative weight which is progressively in favour of the current phases. Since the combination can slow down the convergence of the NMRE procedure, it is performed every two cycles. Additional features are as follows.

(i) The number of cycles in the first step has been reduced to ten (a larger number of cycles does not increase the quality of the electron-density map).

(ii) Weights for observed reflections are not recalculated during the procedure, since their initial distribution is always more suitable than that provided by the NMRE process (*i.e.* too much biased towards large-moduli reflections). For structures with very low resolution (say RES_{obs} > 2.0), however, this bias has the effect of enhancing the role of the better phased reflections: hence, for this category of structures

the weights are again recalculated during the procedure, as for the *ab initio* case.

(iii) The molecular envelope is used also in the second step of the procedure for structures with $RES_{obs} > 2.0$ as an additional constraint for the phasing process.

5. Applications

The EDM–NMRE procedure has been implemented in the *SIR*2004 program (Burla *et al.*, 2005) and applied to the diffraction data of the 18 biological macromolecules listed in Table 1: they are sorted with respect to their experimental resolution limit (SEQ is the corresponding sequential number). In the same table we show their code names, the PDB codes and the corresponding Residues, RES_{obs} and MIS values. The structures listed in the first 13 rows have RES_{obs} lying in the interval 1.55–1.20 Å and were submitted to *ab initio* phasing using the program *SIR*2004. For each structure MPE_{obs} is the minimum phase error attainable by *SIR*2004. The correct solution was not attained by *SIR*2004 for Adeny, Vitad3 and Dnajes: we randomly perturbed the true phase values to obtain an mean phase error of around 60°.

For the five structures listed in rows 14–18 SAD/MAD data are available. We applied the procedures described in Burla *et al.* (2004) to locate the anomalous substructures and in Giacovazzo & Siliqi (2004) for phase assignment and refinement. For each structure MPE_{obs} is the minimum phase error attained.

The results will be examined by first considering one structure in detail; the results will then be summarized for the overall set of test structures.

5.1. Results for Buto140

Buto140 is a protein with RES_{obs} = 1.40 Å and an initial value of MPE_{obs} of 46°; the correlation factor calculated according to (5) between its initial density map and that obtained by the published model is equal to 0.72. Since both the maps were calculated using observed reflections only, in the following we will refer to this correlation value as $CORR_{obs}$. The EDM–NMRE procedure tries to extend the phasing process to RES_{ext} = 1.0 Å: the number of generated reflections is GEN = 1.7 times the number of observed reflections. In Fig. 3, $CORR_{obs}$ is plotted by full squares as a

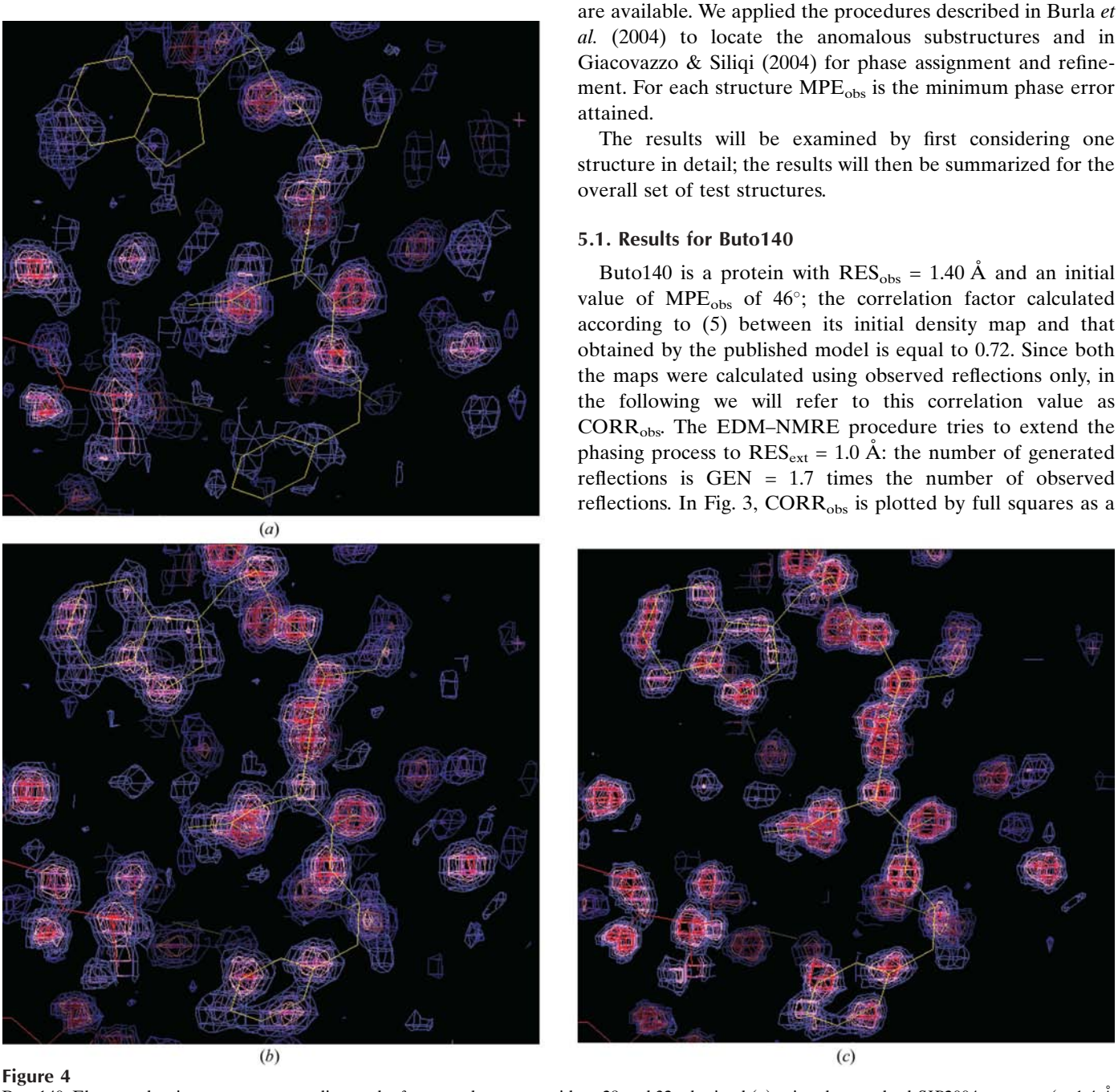

Figure 4
Buto140. Electron-density maps corresponding to the fragment between residues 29 and 32, obtained (*a*) using the standard *SIR*2004 program (at 1.4 Å resolution), (*b*) at the end of the EDM–NMRE procedure using only the observed reflections and (*c*) at the end of the EDM–NMRE procedure using observed and extrapolated reflections.

function of the cycle number: it increases from the starting value of 0.72 (highlighted in Fig. 3 by a dashed line) to a final value of 0.87. Most of this enhancement is yielded during the first step of the procedure (50 cycles) where the NMRE algorithm takes place. In the last 20 cycles, where only the EDM process is active, the further increase of $CORR_{obs}$ essentially arises from the variation of the parameter ruling the modification of the distribution of weights. Such a modification is performed every two cycles and is responsible for the alternation of lower and higher correlation values for subsequent cycles (see Fig. 3). Although this operation seems to dampen the NMRE procedure, we have verified that it allows the attainment of higher final values of $CORR_{obs}$.

A relevant feature of the NMRE procedure is that the increase in correlation is supported by a reduction in the phase error. Indeed, the MPE_{obs} relative to the final phase set is 37°, 9° lower than its starting value.

Let us now consider the extrapolated reflections. The suitability of their estimates is assessed by their MPE_{ext} value and by the correlation (say $CORR_{tot}$) between the final electron-density map made at RES_{ext} resolution and the reference map at the same resolution obtained from the published model. The final value of MPE_{ext} is 48°, which is comparable with the initial MPE_{obs}. The final R_{cryst} value for the extrapolated reflections, calculated using the normalized structure-factor moduli, is 0.45. This points to the moduli playing a minor role with respect to phases and demonstrates that the procedure can converge even in the presence of roughly estimated moduli.

The $CORR_{tot}$ values as a function of the cycle number are shown in open squares in Fig. 3. The starting $CORR_{tot}$ value (equal to 0.64), highlighted by a dotted line, has been calculated by assuming $|E_{\mathbf{h}}^{ext}| = 0$ for all the extrapolated reflections. The increase in $CORR_{tot}$ during the procedure follows that of $CORR_{obs}$, thus indicating the interplay existing between the more accurate phase determination for the observed reflections and the better phase and modulus estimates for the extrapolated reflections. The final value of $CORR_{tot}$ is 0.82, which is well above its starting value and also above the starting $CORR_{obs}$ value. Thus, the procedure is expected to provide a final map with higher resolution and better quality with respect to the initial map. To verify this expectation, we report in Fig. 4 the electron densities corresponding to residues 29–32 (in yellow) and to the symmetrical equivalent of residue 53 (in red) (the corresponding PDB molecular fragment is superimposed). Fig. 4(*a*) shows the electron density provided by *SIR*2004 at 1.4 Å resolution, Fig. 4(*b*) shows the corresponding map obtained at the end of the EDM–NMRE procedure using only the observed reflections and Fig. 4(*c*) shows the final electron-density map calculated at the end of the procedure using observed and extrapolated reflections. Comparison of the figures clearly indicates that Fig. 4(*b*) is more informative than Fig. 4(*a*) (the phase improvement for observed reflections leads to a more interpretable map), with almost all the molecular model enclosed in high electron-density regions. In addition, Fig. 4(*c*) adds further information owing to the increase in resolution from 1.4 to 1.0 Å: the atoms are well located and the map appears sharper, despite it including the contribution of extrapolated reflections with worse phase estimates and roughly estimated moduli. In particular, the ring conformations of Trp29 and Phe32 are fully indicated by the peaks in Fig. 4(*c*) but are almost completely missed in Fig. 4(*a*).

A *post mortem* analysis of the phase and modulus estimates for extrapolated reflections is reported in Figs. 5 and 6. The mean values of $\big||E_{\mathbf{h}}^{ext}| - |E_{\mathbf{h}}^{true}|\big|/|E_{\mathbf{h}}^{true}|$ (left) and phase error (right) are calculated for each resolution (Fig. 5) and for each estimated modulus (Fig. 6) bin: the corresponding errors are shown by vertical lines. It can be seen that both estimates exhibit a loose dependence on the resolution, but improve for higher values of $|E_{\mathbf{h}}^{ext}|$. This justifies our choice of selecting the extrapolated reflections (to be added to the procedure)

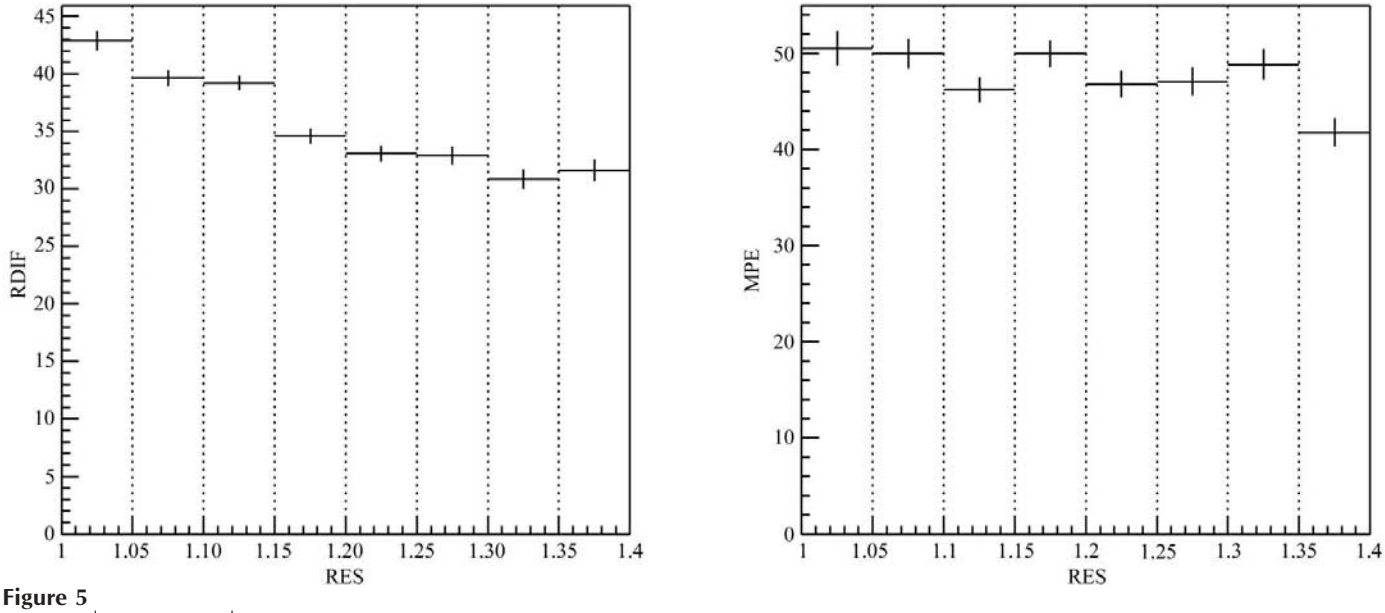

Figure 5
Buto140: $\big||E_{\mathbf{h}}^{ext}| - |E_{\mathbf{h}}^{true}|\big|/|E_{\mathbf{h}}^{true}|$ (left) and MPE (°) (right) *versus* the resolution for the extrapolated reflections.

Table 2
NMRE procedure: values of the main parameters and results of its application.

Structure code	RES_{ext} (Å)	GEN	MPE_{obs} (°)	$CORR_{obs}$	MPE_{ext} (°)	$CORR_{tot}$	R_{cryst} (%)
Cyto549	1.0	0.8	34→27	0.75→0.86	47	0.65→0.77	61
Bpti9	1.0	1.0	24→20	0.90→0.93	32	0.83→0.90	41
Dna145d	1.0	1.0	60→41	0.67→0.85	52	0.61→0.82	50
Cupre	1.0	1.3	30→24	0.87→0.91	29	0.80→0.88	47
Hewl133	1.0	1.7	31→26	0.85→0.91	31	0.75→0.87	47
Wild	1.0	1.3	25→23	0.89→0.92	32	0.77→0.85	47
Buto140	1.0	1.7	46→37	0.72→0.87	48	0.64→0.82	45
Hewl140	1.0	2.0	35→27	0.82→0.91	32	0.72→0.87	44
Vitad3	1.0	2.6	57→42	0.53→0.75	46	0.41→0.70	49
Cole	1.0	2.0	35→25	0.80→0.91	32	0.67→0.87	42
Adeny	1.0	2.2	58→43	0.57→0.76	48	0.48→0.70	46
Dnajes	1.0	2.4	57→57	0.60→0.65	60	0.51→0.63	49
Pazur	1.0	2.5	34→27	0.84→0.90	33	0.72→0.84	48
Kpr	1.2	1.9	58→50	0.79→0.82	49	0.69→0.77	47
Sav3	1.2	2.0	51→51	0.68→0.71	47	0.59→0.66	48
Haptbr	1.2	2.4	54→51	0.72→0.75	51	0.65→0.71	45
Mdd	1.8	1.0	61→59	0.69→0.70	70	0.55→0.63	51
Idi	1.8	1.4	54→52	0.81→0.81	65	0.62→0.73	47

according to their estimated modulus instead than to their resolution.

5.2. Results on all the test structures

The main parameter values adopted for the NMRE procedure and the overall results obtained for each test structure are briefly given in Table 2, where we give the values of RES_{ext} and of GEN, the initial and final values of MPE_{obs}, the initial and final values of $CORR_{obs}$, the final value of MPE_{ext}, the initial and final values of $CORR_{tot}$ and the value of R_{cryst} calculated using normalized structure-factor moduli. The test structures having $RES_{obs} < 1.6$ Å have been extended to 1.0 Å and those phased by SAD/MAD procedures have been extended to 1.2 Å if $RES_{obs} < 2.0$ Å or to 1.8 Å if $RES_{obs} \geq 2.0$ Å. As a result, the GEN values are smoothly dependent on the RES_{obs} and RES_{ext} values. We note the following.

(i) The initial value of MPE_{obs} is usually larger than its final value (with two exceptions, Dnajes and Sav3, for which MPE_{obs} remains invariant). In some cases the phase improvement is quite remarkable.

(ii) Conversely, the initial value of $CORR_{obs}$ is always smaller than its final value. In some cases the difference is quite remarkable.

(iii) The final value of MPE_{ext} is comparable with the initial value of MPE_{obs}; for Kpr, Sav3 and Haptbr we have $MPE_{ext} < MPE_{obs}$.

(iv) The value of R_{cryst} for the extrapolated reflections usually lies in the interval 42–50%.

The efficiency of the NMRE procedure depends on the experimental data resolution. In Figs. 7 and 8 we plot for each test structure the values of $(MPE_{obs})_{initial} - (MPE_{obs})_{final}$ and $(CORR_{obs})_{final} - (CORR_{obs})_{initial}$, respectively, against RES_{obs}. The numbers near the symbols coincide with the sequential number of the test structure, as reported in the first column of Table 1. The general trend is a decrease in the efficiency with the experimental resolution. At higher resolutions some structures may exhibit substantial improvement of the mean phase error and correlation: $(MPE_{obs})_{initial} - (MPE_{obs})_{final}$ attains 20° for Dna145d and $(CORR_{obs})_{final} - (CORR_{obs})_{initial}$ reaches 0.23 for Vitad3. The procedure does not seem to strongly depend on the quality of the initial map nor on the completeness of the experimental data if high experimental resolution data are available. In this respect, it is interesting to note that Vitad3, which has the largest GEN ratio, essentially arising from the large number of missing reflections (MIS = 23.8), exhibits the largest improvement of the correlation, despite starting from $(CORR_{obs})_{initial} = 0.53$.

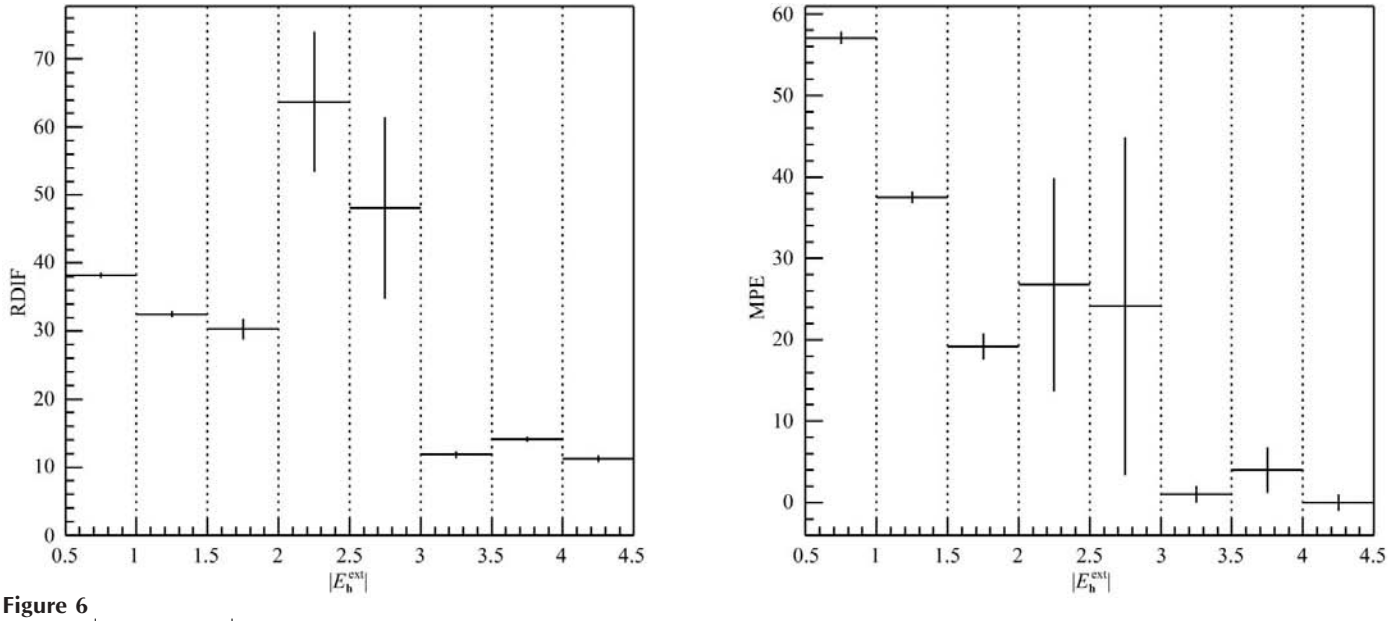

Figure 6
Buto140: $\big||E_{\mathbf{h}}^{ext}| - |E_{\mathbf{h}}^{true}|\big|/|E_{\mathbf{h}}^{true}|$ (left) and MPE (°) (right) *versus* $|E_{\mathbf{h}}^{ext}|$ for the extrapolated reflections.

To assess the improvement in the interpretability of the electron-density map obtained by the EDM–NMRE procedure, we calculated the average values of the residue-by-residue correlation for the main chains and the side chains for all the test structures. To perform this, we used the programs *SFALL* and *OVERLAPMAP* (Collaborative Computational Project, Number 4, 1994) in their default configuration: the former builds an atom map from the published molecule coordinates and generates a mask indicating the area of map corresponding to each residue, while the latter correlates the atom map with that produced by the EDM–NMRE procedure over regions around each residue of the protein. In Table 3 we show the correlation values: the left and right values in both columns refer to electron-density maps obtained before and after the application of the EDM–NMRE procedure, respectively. In all cases the combined effect of phase improvement and resolution extension leads to an increase in the correlation values of both main and side chains, even for the extreme cases of proteins with $RES_{obs} > 2.0$ Å.

The CPU time spent by the EDM–NMRE procedure is limited with respect to that required to obtain the initial electron-density map using the *SIR*2004 program or by the three-step SAD/MAD procedure: it requires at most about half an hour for Cyto549, Vitad3 and Kpr and a mean value of 17 min considering all the test structures (tests were performed using a Xeon-1.7 GHz processor, Linux operating system).

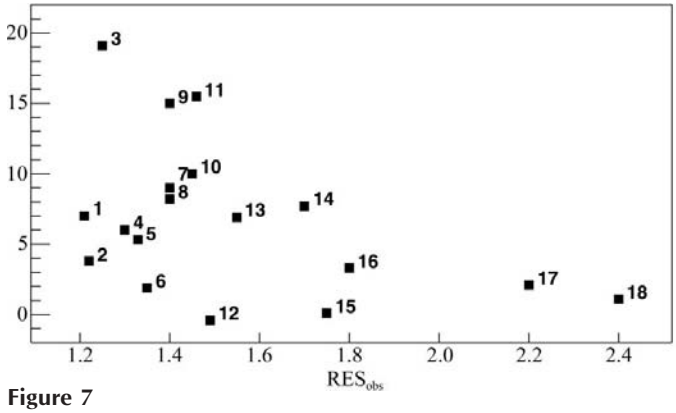

Figure 7
$(MPE_{obs})_{initial} - (MPE_{obs})_{final}$ *versus* RES_{obs} for all the test structures. The number near to the symbols coincide with the sequential number reported in the first column of Table 1.

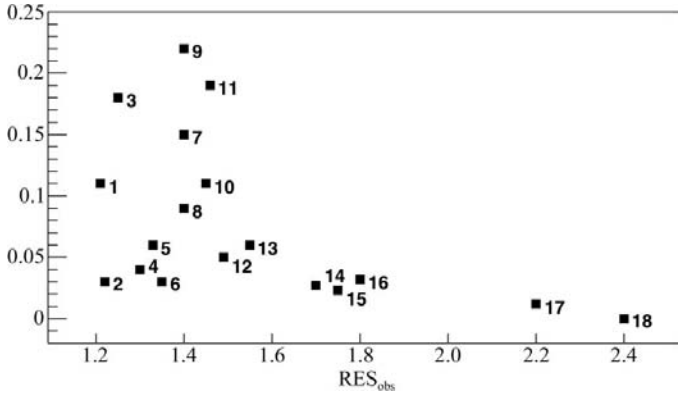

Figure 8
The value of $(CORR_{obs})_{final} - (CORR_{obs})_{initial}$ *versus* RES_{obs} for all the test structures. The number nearby the symbols coincide with the sequential number reported in the first column of Table 1.

Table 3
Average values of the residue-by-residue correlation calculated for the main chains and the side chains between the map calculated from the published model and that obtained before (left) and after (right) the application of the EDM–NMRE procedure.

Structure code	Main chains	Side chains
Cyto549	0.71→0.81	0.64→0.78
Bpti9	0.85→0.92	0.81→0.87
Dna145d	0.13→0.45	0.64→0.83
Cupre	0.82→0.89	0.79→0.85
Hewl133	0.77→0.90	0.76→0.85
Wild	0.82→0.89	0.81→0.86
Buto140	0.68→0.88	0.68→0.85
Hewl140	0.76→0.90	0.74→0.85
Vitad3	0.49→0.78	0.44→0.68
Cole	0.71→0.89	0.69→0.84
Adeny	0.53→0.77	0.49→0.68
Dnajes	0.20→0.31	0.58→0.66
Pazur	0.79→0.90	0.73→0.83
Kpr	0.74→0.83	0.71→0.76
Sav3	0.64→0.69	0.52→0.58
Haptbr	0.68→0.71	0.65→0.66
Mdd	0.64→0.72	0.57→0.60
Idi	0.74→0.80	0.70→0.74

6. Conclusions

We have presented a novel procedure called NMRE which, when combined with classical electron-density modification techniques, is able to extrapolate the moduli and phases of non-measured reflections with resolution lower or higher than the experimental resolution. The procedure has been applied to protein maps obtained by *ab initio* or by SAD/MAD phasing. Modulus and phase estimates were so reliable that they could provide electron-density maps that were more resolved and more informative than those available at experimental resolution.

The results show that the quality and resolution of the final electron-density map is increased with respect to the initial map and that this does not occur at the expense of the mean phase error of the set of measured reflections, which is instead lowered. The result can substantially relieve the crucial task of interpreting the map, making the process of side-chain location and structure refinement easier.

References

Altomare, A., Foadi, J., Giacovazzo, C., Moliterni, A. G. G., Burla, M. C. & Polidori, G. (1998). *J. Appl. Cryst.* **31**, 74–77.
Atwell, S., Meggers, E., Spraggon, G. & Schultz, P. G. (2001). *J. Am. Chem. Soc.* **123**, 12364–12367.
Berkowitz, B., Huang, D. B., Chen-Park, F. E., Sigler, P. B. & Ghosh, G. (2002). *J. Biol. Chem.* **277**, 24694–246700.
Betzel, C., Dauter, Z., Dauter, M., Ingelman, M., Papendorf, G., Wilson, K. S. & Branner, S. (1988). *J. Mol. Biol.* **204**, 803–804.
Bieger, B. & Essen, L. O. (2001). *EMBO J.* **20**, 433–445.

Bonanno, J. B., Edo, C., Eswar, N., Pieper, U., Romanowski, M. J., Ilyin, V., Gerchman, S. E., Kycia, H., Studier, F. M., Sali, A. & Burley, S. K. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 12896–12901.
Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. (2005). *J. Appl. Cryst.* **38**, 381–388.
Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Moustiakimov, M., Polidori, G. & Siliqi, D. (2004). *Acta Cryst.* D**60**, 1683–1686.
Carter, C. W. Jr, Crumley, K. V., Coleman, D. E., Hage, F. & Bricogne, G. (1990). *Acta Cryst.* A**46**, 57–68.
Cascarano, G., Giacovazzo, C., Guagliardi, A. & Steadman, N. (1991). *Acta Cryst.* A**47**, 480–484.
Chatani, E., Hayashi, R., Moriyama, H. & Ueki, T. (2002). *Protein Sci.* **11**, 72–81.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
David, W. I. F. (1987). *J. Appl. Cryst.* **20**, 316–319.
Devedjiev, Y., Dauter, Z., Kuznetsov, S. R., Jones, T. L. & Derewenda, Z. S. (2000). *Structure Fold. Des.* **8**, 1137–1146.
Durley, R., Chen, L., Lim, L. W., Mathews, F. S. & Davidson, V. L. (1993). *Protein Sci.* **2**, 739–752.
Estermann, M. A. & Gramlich, V. (1993). *J. Appl. Cryst.* **26**, 396–404.
Fan, H.-F. & Zheng, Q.-T. (1975). *Acta Phys. Sin.* **24**, 97–104.
Frazao, C., Enguita, F. J., Coelho, R., Sheldrick, G. M., Navarro, J. A., Hervas, M., De La Rosa, M. A. & Carrondo, M. A. (2001). *J. Biol. Inorg. Chem.* **6**, 324–332.
Jansen, I., Peschar, R. & Schenk, H. (1992). *J. Appl. Cryst.* **25**, 231–236.
Giacovazzo, C. & Siliqi, D. (2004). *Acta Cryst.* D**60**, 73–82.
Karle, J. & Hauptman, H. (1964). *Acta Cryst.* **17**, 392–396.
Langs, D. A. (1998). *Acta Cryst.* A**54**, 44–48.
Lascombe, M. B., Ponchet, M., Venare, P., Milat, M. L., Blein, J. P. & Prangé, T. (2002). *Acta Cryst.* D**58**, 1442–1447.
Leslie, A. G. W. (1987). *Acta Cryst.* A**43**, 134–136.
Liu, Y., Fan, H.-F. & Zheng, C.-D. (1988). *Acta Cryst.* A**44**, 61–63.
Main, P. (1990). *Acta Cryst.* A**46**, 372–377.
Petratos, K., Dauter, Z. & Wilson, K. S. (1988). *Acta Cryst.* B**44**, 628–636.
Roversi, P., Irwin, J. J. & Bricogne, G. (1998). *Acta Cryst.* A**54**, 971–996.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Sayre, D. (1974). *Acta Cryst.* A**30**, 180–184.
Schiltz, M., Shepard, W., Fourme, R., Prangé, T., de La Fortelle, E. & Bricogne, G. (1997). *Acta Cryst.* D**53**, 78–92.
Schroth, G. P., Kagawa, T. F. & Ho, P. S. (1993). *Biochemistry*, **32**, 13381–13392.
Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. (1976). *J. Mol. Biol.* **104**, 109–144.
Silinski, P., Allingham, M. J. & Fitzgerald, M. C. (2001). *Biochemistry*, **40**, 14493–14502.
Sivia, D. S. & David, W. I. F. (1994). *Acta Cryst.* A**50**, 703–714.
Smith, G. D., Blessing, R. H., Ealick, S. E., Fontecilla-Camps, J. C., Hauptman, H. A., Housset, D., Langs, D. A. & Miller, R. (1997). *Acta Cryst.* D**53**, 551–557.
Van der Putten, N., Schenk, H. & Tsoucaris, G. (1982). *Acta Cryst.* A**38**, 98–102.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Voigt-Martin, I., Yan, D., Yakimansky, A., Schollmeyer, D., Gilmore, C. J. & Bricogne, G. (1995). *Acta Cryst.* A**51**, 849–868.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Xu, H. & Hauptman, H. A. (2000). *Acta Cryst.* A**56**, 284–287.