A modified *ACORN* to solve protein structures at resolutions of 1.7 Å or better

Yao Jia-xing,[a]* M. M. Woolfson,[b] K. S. Wilson[a] and E. J. Dodson[a]

[a]York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, England, and [b]Department of Physics, University of York, Heslington, York YO10 5YW, England

Correspondence e-mail: yao@ysbl.york.ac.uk

ACORN has previously been shown to provide an efficient density-modification procedure for the solution of protein structures using diffraction data to better than 1.3 Å. The initial phase set could be obtained from a variety of sources such as the position of a heavy atom, a set of scatterers such as S that had been positioned from anomalous dispersion measurements, a fragment or a very low homology model placed from a molecular-replacement search. Several structures solved using the early version of *ACORN* have been reported in the literature. Here, the effect of applying the original *ACORN* procedures at lower resolution is reported and new procedures that yield good-quality maps with data sets of resolution down to 1.7 Å are described. These new procedures involve the artificial extension of data to atomic resolution and new density-modification processes that develop density at atomic positions that was previously suppressed. The test calculations were aimed firstly towards a proof of principle using a small fragment of a known structure to demonstrate that the procedure could generate correct density and a derived model in initially empty regions of the cell. Further tests addressed the use of more realistic starting models.

1. Introduction

ACORN (Foadi *et al.*, 2000; Yao *et al.*, 2002) was developed as a procedure that can refine a poor starting phase set without intervention by the user when the resolution of the data is 1.3 Å or better (atomic resolution). With good-quality data, the procedure can reduce the phase error from an initial phase error of up to 80° to below 20° and thus provide a clear and complete image of the structure. It has been applied successfully to the solution of several new structures when the initial fragment is a poor or incomplete molecular-replacement solution, a fragment of known conformation, such as an α-helix (Yao, 2002), a set of heavier atoms positioned using anomalous or isomorphous differences (Gordon *et al.*, 2001) or, for metalloproteins, the heavy-atom structure alone (McAuley *et al.*, 2001).

The phase-improvement tool starts with the phases from the fragment and uses a density-modification procedure, DDM (dynamic density modification), to develop the complete structure. We give here a short description of the basic DDM process as the starting point for the improvements we have made.

Firstly, a map is calculated with appropriate Fourier co-efficients and phases derived from the previously modified map (or from the fragment in the first refinement cycle). The initial coefficients are $W(\mathbf{h})E_{\mathrm{o}}(\mathbf{h})$ for the stronger $E_{\mathrm{o}}(\mathbf{h})$. The

research papers

limit on $E_o(\mathbf{h})$ is chosen to give sufficient reflections for the procedure: about ten times the number of atoms in the asymmetric unit. The weight $W(\mathbf{h})$ is akin to a Sim weight (Sim, 1960) and is given by

$$W(\mathbf{h}) = \tanh(0.5X), \qquad (1)$$

where $X = |E_o(\mathbf{h})||E_c(\mathbf{h})|$ and $|E_c(\mathbf{h})|$ is the amplitude of the Fourier coefficient of the previous map scaled to the values of $|E_o(\mathbf{h})|$ in shells in reciprocal space.

Secondly, the standard deviation σ of the map density is calculated. The map density is then modified by

$$\begin{cases} \rho' = 0 & \text{if } \rho < 0 \\ \rho' = \rho \tanh[0.2(\rho/\sigma)^{3/2}] & \text{if } kn\sigma \geq \rho \geq 0 \ , \\ \rho' = kn\sigma & \text{if } \rho' > kn\sigma \end{cases}$$

where k is a constant given by the user (default value 3) and n is the minimum of the cycle number and 5.

The basis of DDM is that even when the mean phase error is high, say $75°$, the expected value of the map density at an atomic centre is several standard deviations σ above the average density, which will be close to zero (see supplementary material[1]). If the density at a point in the map is 1σ, then this is well within the bounds of random fluctuations and the probability that an atom is centred at that point is quite low. In contrast, if the density at a point is 4σ, then this represents a rare excursion from the mean by a random fluctuation so that the probability that an atom is situated there is quite high. The DDM process modifies density in such a way that lower density is proportionately reduced, thus increasing the probability that the residual density models true atomic positions. The high density that arises in the positions of fragment atoms is savagely truncated to reduce the initial phase bias.

To follow the process of refinement, a correlation coefficient, CC_s, is calculated between the magnitudes of the Fourier coefficients of the modified map and the magnitudes of the reflections not included in calculating the map. For a well behaved refinement with high-resolution data the value of CC_s rises steadily as the mean phase error reduces and saturation of the value of CC_s may be taken to indicate the end of the refinement process (Foadi et al., 2000).

It has been found beneficial to introduce one or two steps of what is called Sayre-equation refinement (SER) before commencing DDM and occasionally within the refinement procedure, especially if the refinement appears to stagnate. This procedure modifies phases in such a way that the Sayre equation is better satisfied for both large and small structure factors. It involves the calculation of six Fourier transforms per step (Foadi et al., 2000). Paradoxically, although SER usually makes both the value of CC_s smaller and the mean phase error larger, it conditions the phase set so that new peaks appear in the resultant map and the subsequent refinement by DDM progresses further than it would have done without the SER intervention.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: DZ5054). Services for accessing these data are described at the back of the journal.

2. Problems with low-resolution data

The analysis given in the supplementary material[1] indicates that the expectation density at an atomic centre in units of the standard deviation and assuming an equal atom distribution falls dramatically with a reduction in the resolution limit of the data. For data with resolution 1.7 Å and perfect phases the density at an atomic centre is about 5σ, which means that with perfect phases the atoms will show up well. The corresponding atomic centre density for perfect phases at 1.0 Å is about 11σ. Approximately, for a mean phase error of say $75°$, the expectation values at atomic centres must be multiplied by $\cos 75°$ (0.259), giving 2.8σ and 1.3σ for data resolutions of 1.0 and 1.7 Å, respectively. This gives a sufficiently good starting point for refinement at the higher resolution but not at the lower. A lower initial mean phase error, in the region 55–$60°$, might enable a structure to be refined at 1.7 Å, but such a starting point may be difficult to achieve.

However, even if a suitable starting point for refinement could be found, the difficulties with low-resolution data do not end there. The supplementary material[1] also shows that the expectation density at atomic centres favours heavier atoms disproportionately. In trials with known structures with data sets in the resolution range 1.5–1.7 Å, the application of DDM gives final mean phase errors of the order of 45–$50°$ and the final maps tend to show all S atoms and many O and N atoms but comparatively few C atoms, although C=O often appears as a single peak. For atoms bonded to three other atoms the density is often weak or missing because for 1.6 ± 0.1 Å data they lie in the diffraction ripple minimum of the neighbouring atoms. This category includes the C^α atoms, so the main chain appears broken and automated chain-tracing methods fail.

The atomic resolution required for a successful solution has severely limited the proportion of structures to which ACORN could be applied, so we have endeavoured to extend the resolution range of applicability, the results of which we now report.

3. Artificially extending the data

A key strategy that has been found to improve the performance of ACORN at low resolution greatly is to extend the data to 1 Å artificially and to include these non-observed reflections in the phasing process. (Indeed, there is also considerable benefit in filling in any missing data in the observed resolution range, especially at low resolution.) If reasonable phase estimates can be made for these artificially introduced reflections then they add information to the resultant map, the peak shape becomes more 'atomic' and the DDM procedure more effective; it is well known that if phases are correct the resultant map will show the expected features, almost regardless of the magnitudes. The question then arises of what magnitudes to give these introduced reflections and how to weight their contributions to a map.

4. Weighting the extended data

The default weighting scheme used for the observed data when calculating maps is given in (1). For the extended reflections there is no problem in finding $|E_c(\mathbf{h})|$ since it can be calculated from the original fragment to any resolution and can also be obtained from successive maps. The problem is to find an estimate of $|E_o(\mathbf{h})|$ for the extended reflections, an estimate we represent by $|E_{ext}(\mathbf{h})|$. If precise estimates could be made for these reflections then the position would be equivalent to that of having high-resolution data and weighting scheme (1) could be used as it is. However, only crude estimates can be made of these normalized structure amplitudes, so that some down-weighting of their contribution to the map is desirable. This down-weighting should also take account of the number of extended reflections relative to the number of observed, otherwise these flawed reflections would overwhelm the effect of the observed reflections. An *ad hoc* weighting scheme that we have found to be effective for the extended reflections is

$$W_{ext}(\mathbf{h}) = \tanh(0.5 X_{ext}), \qquad (2)$$

where

$$X_{ext} = \left(\frac{n_{obs}}{n_{obs} + n_{ext}}\right)^{1/2} |E_{ext}(\mathbf{h})||E_c(\mathbf{h})|$$

and n_{obs} is the number of observed reflections, n_{ext} is the number of extended reflections and $|E_c(\mathbf{h})|$ is the structure amplitude from the fragment or map scaled in shells in reciprocal space to make $\langle|E_c|^2\rangle = 1$. This has the effect that the greater the percentage of extended reflections used, the smaller their weights. More sophisticated weighting schemes based on estimates of the phase error from correlation coefficients were also tried, but to date none of those led to consistent results for the current set of test structures.

5. Magnitudes for the extended data

It is necessary to provide some estimate of the normalized structure amplitudes for the artificially extended data. Various ideas were explored.

5.1. Extension (i): using the average value of $|E|^2$

By definition, the average value of $|E|^2$ is 1, so taking $|E_{ext}(\mathbf{h})| = 1$ for all the extended reflections provides a very simple estimate, considerably better than effectively setting these values to 0.0.

5.2. Extension (ii): an expectation value from $|E_{frag}(\mathbf{h})|$

Although the normalized contribution of the fragment itself, $E_{frag}(\mathbf{h})$, is found to be a poor indicator of the magnitude of the extended normalized structure amplitudes, they should still have some influence on the expectation value. Intuitively, it is evident that the greater the contribution from the fragment then the greater the expectation value of the normalized structure amplitude. For a structure with N equal atoms in the asymmetric unit and a fragment of m atoms, the root-mean-square expectation value of the normalized structure amplitude of an extended reflection is

$$\langle|E_{ext}(\mathbf{h})|\rangle = \left\{1 + \frac{m}{N}[|E_{frag}(\mathbf{h})|^2 - 1]\right\}^{1/2}. \qquad (3)$$

This gives some departure from the expectation value of unity as given by extension (i). For example, with $m/N = 0.05$ the expectation value is 0.975 for $|E_{frag}(\mathbf{h})| = 0$ and 1.183 for $|E_{frag}(\mathbf{h})| = 3$.

5.3. Extension (iii): using calculated $|E|$ values from the fragment

An important diagnostic tool in the correct location of a fragment is the correlation coefficient between the structure amplitudes calculated from the fragment and those of the observed reflections. For a small fragment this is weak (of the order 0.01–0.02, but positive), so we tested the idea of using the fragment contributions as estimates of the measured observations.

5.4. Extension (iv): squaring the map (Sayre's equation)

For an equal atom resolved structure, application of the Sayre equation is equivalent to taking the Fourier coefficients of the squared map, giving Fourier coefficients that double the resolution of the original map.

6. Testing the estimation methods

Tests were made with the structure of penicillopepsin (PDB code 1bxo; Ding *et al.*, 1998). The space group is $C2$, with observed data to 0.9 Å and 2977 independent non-H atoms in the asymmetric unit. For the tests the data were truncated to 1.5 Å so that comparisons with actual observations could be made in the 1.5–1.0 Å range. A starting fragment was selected containing 400 atoms (13.4% of the structure). Using data to 1.0 Å resolution, the structure can be developed by *ACORN* to a final mean phase error of 13.1°, yielding a final map that clearly shows individual atoms and that can be interpreted by automatic procedures (Foadi *et al.*, 2000). (A much smaller fragment can be used to give a successful solution with 1.0 Å data, but since 400 atoms are required with 1.5 Å data we use that size of fragment for comparison purposes.)

With the 400-atom fragment using 1.0 Å data the variation of the unweighted mean phase error for reflections with $|E| > E_{strong}$ (1.2 in this case) with cycles of DDM is shown in Table 1.

At the truncated 1.5 Å resolution it was necessary to keep all observations with $|E| > 0.8$ in order to have sufficient reflections for the strong set. The initial R factor and the correlation coefficient between calculated structure amplitudes from a 400-atom fragment and all observed structure amplitudes are more or less uniform over the whole resolution range and are equal to 55% and 0.15, respectively. Various weighted mean phase errors between phases calculated from the fragment and phases calculated from the refined structure were calculated in the form

Table 1
The variation of the overall mean unweighted phase error (MPE) for $E > E_{strong}$ with DDM cycles using 1.0 Å data and a starting fragment of 400 atoms for 1bxo.

Cycle No.	0	2	4	6	8	10	12
MPE (°)	52.2	33.6	22.3	16.3	14.0	13.3	13.1

Table 2
The unweighted mean phase error for reflections with $E_o > E_{strong}$ (= 0.8) with cycle of DDM using estimates (iii) and (i) for the extended reflections.

Cycle No.	0	5	10	15	20	25	30	35
$\|E_{ext}\| = \|E_{frag}\|$	59.5	52.4	50.3	49.2	48.4	47.8	47.4	47.1
$\|E_{ext}\| = 1$	59.5	49.5	46.6	45.2	44.2	43.5	43.0	42.7

$$\overline{|\Delta\varphi|} = \frac{\sum\limits_{\mathbf{h}} w_{\mathbf{h}} |\Delta\varphi_{\mathbf{h}}|}{\sum\limits_{\mathbf{h}} w_{\mathbf{h}}} \tag{4}$$

for $w_{\mathbf{h}} = 1$, $W(\mathbf{h})$, $|E_o(\mathbf{h})|$ and $W(\mathbf{h})|E_o(\mathbf{h})|$. These mean phase errors, calculated in resolution shells, are shown in Fig. 1. The lowest mean phase errors are for weights $W(\mathbf{h})|E_o(\mathbf{h})|$, which are the amplitudes of the coefficients used for the calculation of maps. Various criteria were used to estimate the effectiveness of the extension methods (i) to (iv) with data truncated to 1.5 Å. Fig. 2 illustrates the R factor between the different estimates of $|E_{ext}|$ and the true values in resolution shells for the resolution range 1.5–1.0 Å. It can be seen that the R factors for the pairs {(i), (ii)} and {(iii), (iv)} are similar and are much smaller for the first pair: approximately 0.41 compared with 0.55. From this it seems obvious that if we are looking for an effective way to extend magnitudes in the extended range we should just look at extension methods (i) and (ii); further testing confirms this conclusion.

Applying extension method (i) for the extended reflections gave the variations of mean phase error reported in Table 2 for reflections to 1.5 Å resolution with cycle number, which are compared with the results from the inferior magnitude-extension method (iii).

Applying DDM phase refinement with extension method (ii) was slightly worse than for method (i), but not significantly so. It can be concluded that as a default procedure the simple method (i) should be used since it is independent of the size and accuracy of the fragment.

7. Enhancements in density modification

An application of the original *ACORN* to the structure of 1bxo using only the truncated data at 1.5 Å gives a final mean phase error that differs very little from the starting error of ~60°: the introduction of the extension procedure has thus made a dramatic improvement. Nevertheless, for the reasons given in §2, the maps are still difficult to interpret, especially since many Cα atoms have no density. It is known that density refinement using several different procedures is often more effective than any one of the procedures alone (Abrahams &

Leslie, 1996; Cowtan & Main, 1996; Refaat *et al.*, 1996), so we have introduced some new density-modification processes to enhance weaker density and thus to reveal lighter atoms.

Some other types of map coefficients are used in these steps. The first maps used for DDM used coefficients $W(\mathbf{h})E_o(\mathbf{h})$. As the electron density shows more of the model, maps with other kinds of coefficients also become useful. We have investigated the use of coefficients $W(\mathbf{h})[|2E_o(\mathbf{h})| - |E_c(\mathbf{h})|]$ and $W_{ext}(\mathbf{h})[2|E_{ext}(\mathbf{h})| - |E_c(\mathbf{h})|]$ for the observed and extended data, respectively, or σ_A-weighted maps with map coefficients $2m(\mathbf{h})|E_o(\mathbf{h})| - \sigma_A|E_c(\mathbf{h})|$ and $2m(\mathbf{h})|E_{ext}(\mathbf{h})| - \sigma_A|E_c(\mathbf{h})|$. Density-modification procedures starting from these maps are designated as DDM1 and DDM2 in the following discussion and the original procedure designated as DDM0. The maps used for DDM1 and DDM2 have some similarity to those calculated with coefficient amplitudes $2|E_o| - |E_c|$ traditionally used to bring up the density for missing atoms.
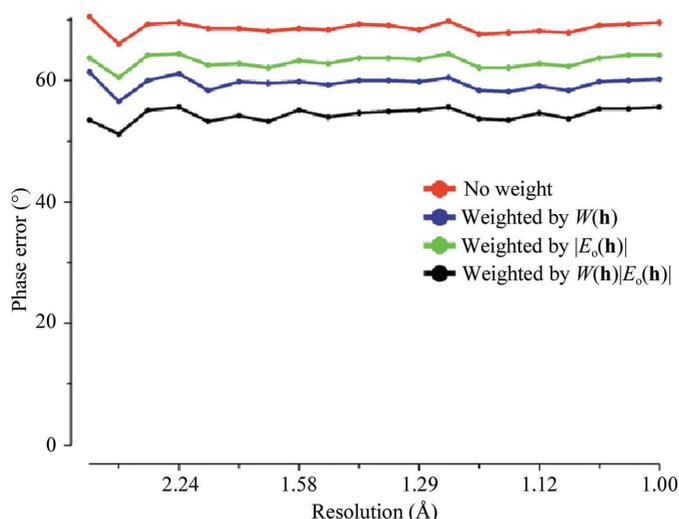
Figure 1
Initial phase errors from the 400-atom fragment of 1bxo in resolution shells with four weighting schemes for all reflections.
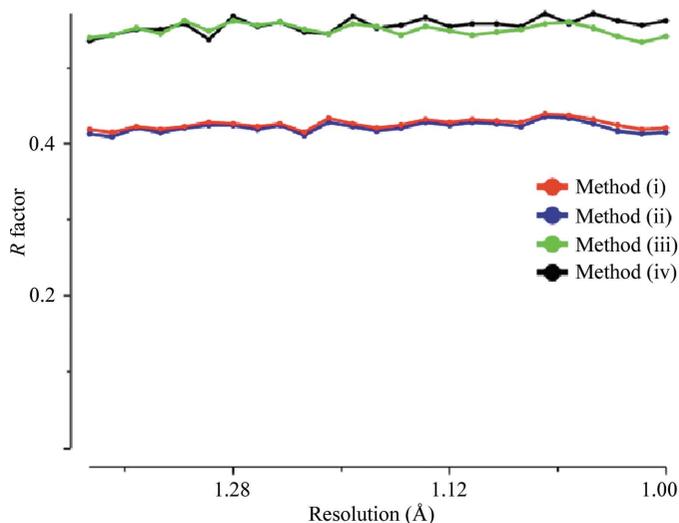
Figure 2
The R factor between observed normalized structure magnitudes and the magnitudes estimated in four different ways for extended reflections.

Table 3
Details of the test structures.

The phase errors are for strong E values with $|E_o| > 0.8$.

PDB code	1bxo	1bxo	1w2y	1w2y	1paz	1j3w	1uuq	1uuq
Space group	$C2$	$C2$	$P2_12_12_1$	$P2_12_12_1$	$P6_5$	$C222_1$	$P2_12_12$	$P2_12_12$
No. of residues	323	323	458	458	123	492	440	440
Resolution (Å)	10.0–1.0	10.0–1.5	48.59–1.65	48.6–1.65	8.97–1.55	29.82–1.5	33.98–1.5	33.98–1.5
Extended resolution (Å)	59.23–1.0	59.23–1.0	55.9–1.0	55.9–1.0	43.3–1.0	76.06–1.0	67.97–1.0	67.97–1.0
No. of reflections								
Theoretical total to 1.0 Å resolution	142670	142670	236768	236768	75026	348917	251645	251645
Measured	133476	42010	52878	52878	19659	104432	74865	74865
Extended	9194	100660	183890	183890	55367	244485	176780	176780
Extended/total (%)	6.4	70.6	77.7	77.7	73.8	70.1	70.2	70.2
First model/percentage of final model	400 atoms/ 13.5%	400 atoms/ 13.5%	MR and rigid-body refinement	400 atoms/ 9.8%	1 Cu/ 2%	150 atoms/ 3.2%	SAD + SHELXE 1.5, 2.1, 2.5 Å	16 Se/ 3%
Initial/final phase error (°)	60.7/13.9	60.7/34.5	52.5 /33.4	62.2/35.7	73.8/21.6	71.3/19.6	28.7/20.4, 41.6/24.9, 49.3/29.6†	74.1/21.0

† Tests where the input phase errors are for the observed reflections at the specified resolutions.

A description of the quantity σ_A was first given by Srinivasan (1966) and developed for macromolecular applications by Read (1986). The value of σ_A reflects the current mean phase error [in *ACORN* estimated from the correlation coefficients between $|E_o|$ and $|E_c|$] and m is a figure of merit calculated for each reflection based on the agreement of $E_o(\mathbf{h})$ and $E_c(\mathbf{h})$ and the current value of σ_A. When the phase error is large, σ_A is small and DDM2 gives more or less the same results as DDM0.

There are four steps in an enhancement cycle.

(i) A map is calculated using the same coefficients as DDM1 (default) or DDM2 as specified by user.

(ii) All negative density is set to zero and the square root is taken of the remaining density to enhance weaker density.

(iii) A locally averaged map over a sphere of radius 2 Å is calculated for a map with coefficients $W(\mathbf{h})|F_o(\mathbf{h})|$ for the observed reflections to find an envelope (Wang, 1981, 1985). The Fourier coefficients of the locally averaged map are found by convoluting the original map with a uniform sphere of radius 2 Å. A cutoff for the averaged map density is found below which the volume equals that expected for the solvent. For all 'solvent regions' the density of the map obtained from process (ii) is set equal to zero.

(iv) To enhance the weaker density at each grid point in the 'protein region', the density is multiplied by the average in the 2 Å radius sphere centred on that point. Finally all densities higher than $n\sigma$ are set back to $n\sigma$ (the default value of n is 3.0).

The density-enhancement procedure is introduced for one cycle whenever the process of phase refinement slows down. After an enhancement step, further cycles (of DDM0, DDM1 or DDM2) will continue, as specified by user-given keywords. The default procedure is to use DDM0 only. The refinement is stopped when the values of the correlation coefficient CC$_s$ indicate that no further improvement is possible. We are investigating many more test cases to try to select the best protocol.

8. Testing the new *ACORN*

The new *ACORN* was tested on a series of known structures selected to establish its performance with data in the 1.5–1.7 Å range. Half of the tests described here were initiated from a fragment of the refined model. This is of course unrealistic, but allows us to analyse the final maps for features in regions outside the fragment. Details are summarized in Table 3.

The tested structures were the following.

(i) Penicillopepsin at 1.5 Å (PDB code 1bxo). For this structure the measured data really extend to 1.0 Å and so it was possible to evaluate the correctness of the 'extended' phases for those data between 1.5 and 1.0 Å resolution, where *ACORN* used estimated E values of 1.0. The phasing was started from a true 400-atom fragment.

(ii) *Campylobacter jejuni* dUTPase at 1.65 Å (PDB code 1w2y). This structure was used in two ways. Firstly, a fragment of 400 atoms from the known structure was used as a starting point for *ACORN* to provide a proof of principle for the new algorithms. Whilst this is clearly an unrealistic start for a real structure, it allowed us to evaluate the relative merits of the new options in the program, especially with regard to revealing atoms that were missing from the starting fragment. Secondly, the molecular-replacement structure, before and after rigid-body refinement, was used as a starting point for *ACORN*. This represents a realistic test on a real problem.

(iii) Pseudoazurin at 1.5 Å (PDB code 1paz). This was a structure solved by one of the authors (KSW) and colleagues some years ago. Pseudoazurin is a small metalloprotein with a single Cu atom. *ACORN* was used to locate the copper position and to extend the phases starting from the Cu atom alone, again providing a realistic case of using native data only. Alternatively anomalous scattering data, which is easily available, can be used to locate the copper position and then run *ACORN* with native data starting from the phases from the Cu atom.

(iv) Gliding protein at 1.5 Å (PDB code 1j3w). The phases for this test structure were generated from a very small percentage of the final model (3.2%).

(v) Exo-mannosidase (PDB code 1uuq). This is a known structure solved in York using data to 1.5 Å resolution. It has 440 residues with 16 methionines in the asymmetric unit (Dias *et al.*, 2004). Single-wavelength anomalous measurements were used to locate the selenium positions and estimate initial phases. Starting from these SAD phases at resolutions down to 2.5 Å, the new *ACORN* program successfully refined and extended them for the whole 1.0 Å set. This represents a novel and effective role for *ACORN*.

Taken together, these five examples provide an excellent validation of the usefulness of the new *ACORN*.

8.1. Penicillopepsin (PDB code 1bxo)

The data were restricted to 1.5 Å for the purposes of the test and extended to 1.0 Å by extension method (i). The starting
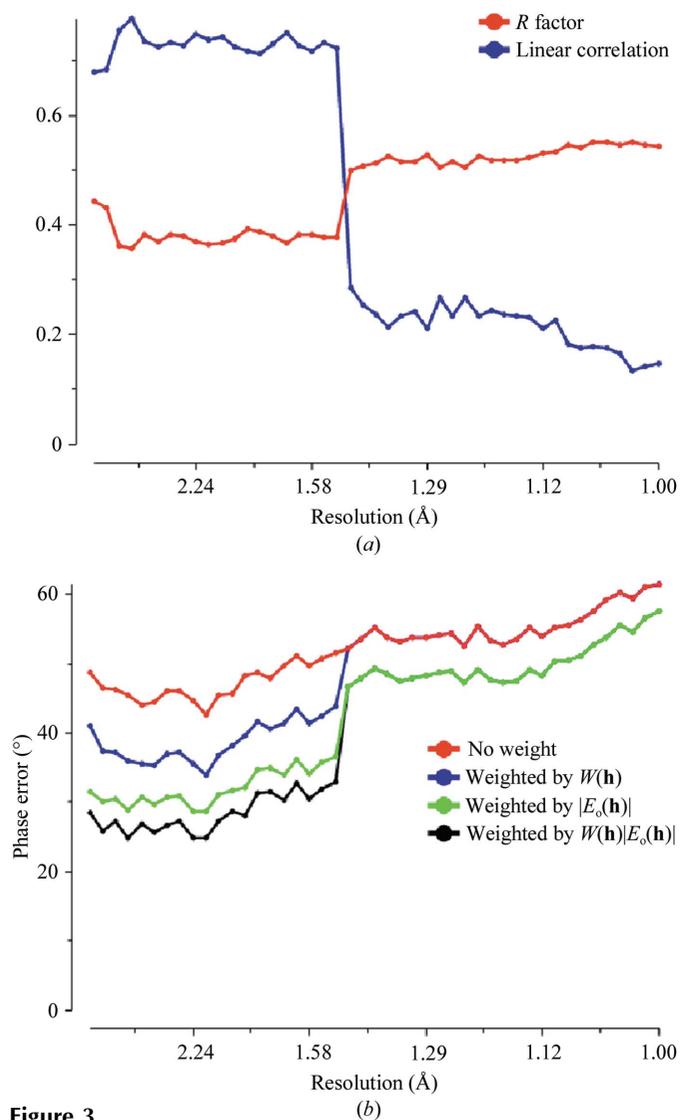
model of 400 atoms was taken from the final model and gave an unweighted mean phase error of 60.7° for observed reflections with $|E| > 0.8$ (a limit chosen to give sufficient reflections which is about ten times the number of atoms in the asymmetric unit). After 58 cycles of DDM0 refinement (without SER insertions) the value of CC_s was 0.084 and the mean phase error was reduced to 39.6°. The map correlation coefficient was 0.6860 for the whole structure, 0.7056 for the main chain and 0.6641 for side chains.

With the inclusion of the density-enhancement procedures, the value of CC_s became 0.119, the mean phase error was reduced to 34.5° and the map correlation coefficient was 0.7754 for the whole structure, 0.7959 for the main chain and 0.7488 for side chains. The improvement in the mean phase error and map correlation coefficient is substantial. The map now showed atoms that were previously missing. The R factor and CC found in resolution shells are illustrated in Fig. 3(*a*), in which the sharp changes beyond the imposed artificial resolution limit are clearly seen. The final mean phase errors, also in resolution shells, are given in Fig. 3(*b*) for the same weighting schemes used in Fig. 1. It is interesting that although no observed data are used beyond 1.5 Å resolution, the phases for the extended reflections are significantly better than random.

8.2. *C. jejuni* dUTPase (PDB code 1w2y)

For this structure, with space group $P2_12_12_1$, the experimental data reach 1.65 Å resolution and there are 430 residues in the asymmetric unit (Moroz *et al.*, 2004). A low-homology model, *Trypanosoma cruzi* dUTPase (PDB code 1ogk), was used to solve the structure by molecular replacement with data to 4 Å.
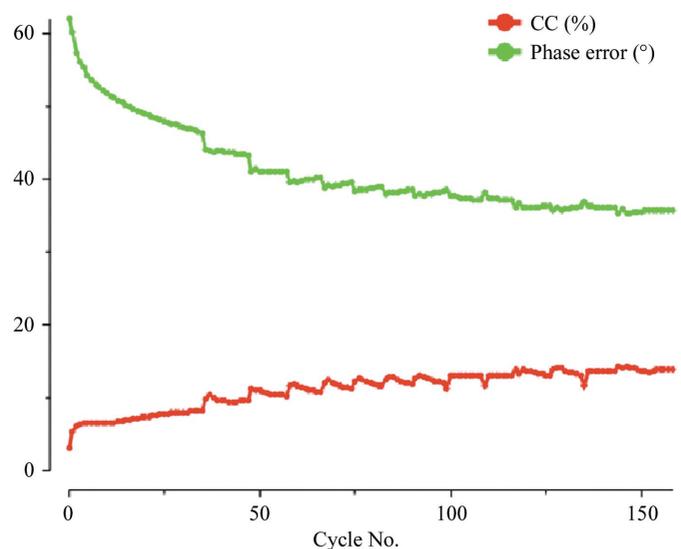
Figure 4
The variation of CC_s and phase error with cycle for 1w2y. DDM0 in cycles 1–35 and 37–47. ENH in cycles 36, 48, 58, 67, 75, 83, 91, 99, 109, 117, 127, 135 and 144. DDM1 in cycles 49–57, 59–66, 68–74, 76–82, 84–90, 91–98, 118–126, 128–134 and 145–152. DDM2 in cycles 100–116, 136–143 and 153–158.

Figure 3
The *ACORN* refinement of 1bxo using the observed amplitudes to 1.5 Å and $E = 1.0$ from 1.5 to 1.0 Å resolution. (*a*) The R factor and CC between the observed amplitudes and the calculated amplitudes by *ACORN* in resolution shells. (*b*) The weighted phase error in resolution shells.

As a proof of principle, *ACORN* was first applied with phases calculated from a starting fragment made up of 400 atoms (13% of the model) taken from the refined structure. The initial phase error was 62.2° and fell to 46.4° after 35 cycles of DDM0. The enhanced density-modification procedures were then applied. The sequence of application of the various components of density enhancement is given in Fig. 4, where ENH represents the enhancement cycle and the cycles using DDM1 and DDM2 are also indicated. The final mean phase error was reduced to 35.7° with a correlation coefficient of 0.137. Using these phases, the map correlation was 0.7521 for the total structure, 0.8142 for the main chain and 0.7238 for side chains. While this is clearly an unrealistic starting fragment, it demonstrated the ability of *ACORN* to introduce correctly placed atoms in regions missing from the starting fragment and allowed us to analyse maps for new features in detail.

This structure is used to illustrate the improvements described in this paper. Fig. 5(*a*) shows the final *ACORN* map at 1.65 Å without the use of data extension or density enhancement; in fact, by use of the original *ACORN*. Adding data extension to assist the application of DDM0 but still calculating the map using 1.65 Å data only gives Fig. 5(*b*), which shows the advantage from the data extension very clearly. Adding the extended terms with their associated weights gives the further improvement seen in Fig. 5(*c*). Using both data extension and density enhancement to improve phases, the 1.65 Å map is shown in Fig. 5(*d*) and it is seen that the density enhancement introduces density previously missing; as an example, in the left centre of Fig. 5(*d*) the main chain is indicated by an unresolved bar of density. Finally, with data extension and density enhancement a 1.0 Å map, shown in Fig. 5(*e*), breaks up the unresolved density in Fig. 5(*d*) into resolved features representing the individual atoms.

Once a final *ACORN* map has been produced it can be interpreted in terms of atomic positions, usually by an automatic procedure such as *ARP/wARP* (Perrakis *et al.*, 1999). Table 4 gives an analysis of the maps illustrated in Fig. 5 in terms of the peaks they contain and the proportion corresponding to true atomic positions. For this purpose a 'good' peak is defined as being within 0.4 Å of a refined atomic position.

The most important peaks for automatic interpretation of the map are those in the main chain and the number of these in the maps is given in Table 5, which also gives the number of heavier atoms (sulfur and magnesium) found. The amount of new information, corresponding to atoms outside the fragment, is significantly higher for the map illustrated in Fig. 5(*e*) than for any of the other maps.

We subsequently used a realistic starting model, namely the *T. cruzi* structure positioned by molecular replacement. Compared with the final structure, this model gave an r.m.s. deviation of 2.1 Å for the core 116 residues, an initial phase error of 78.3° for 26 859 $E > 0.8$ and $CC_s = 0.02206$. 841 cycles of *ACORN*, using DDM0 and DDM1 with enhancement procedures, reduced the phase error to 43.4° with $CC_s = 0.10683$. This was a challenging problem for *ACORN* as

(*a*)　　　　　　　　　　　(*b*)

(*c*)　　　　　　　　　　　(*d*)

(*e*)
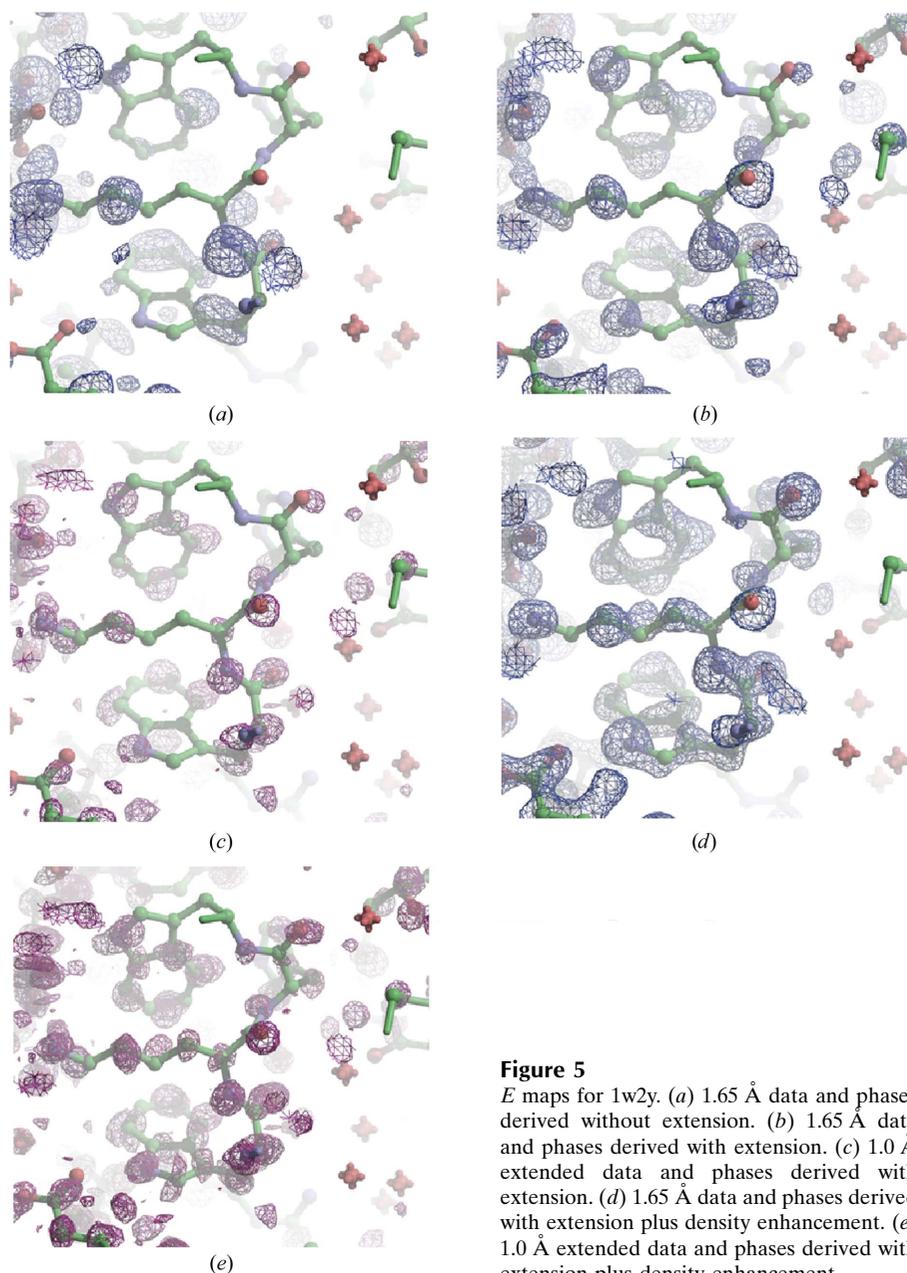
Figure 5
E maps for 1w2y. (*a*) 1.65 Å data and phases derived without extension. (*b*) 1.65 Å data and phases derived with extension. (*c*) 1.0 Å extended data and phases derived with extension. (*d*) 1.65 Å data and phases derived with extension plus density enhancement. (*e*) 1.0 Å extended data and phases derived with extension plus density enhancement.

Table 4
An analysis of the peaks in the maps shown in Fig. 5.

Map	Fig. 5(a)	Fig. 5(b)	Fig. 5(c)	Fig. 5(d)	Fig. 5(e)
Total peaks	2193	2153	2925	2006	2668
Total good peaks	542	844	1123	988	1297
Total good peaks outside fragment	429	652	811	845	1082
Total good peaks/ total peaks (%)	24.7	39.2	38.4	49.3	48.6
Total good peaks outside fragment/total peaks (%)	19.6	30.3	27.7	42.1	39.8

Table 5
Peaks corresponding to the main-chain and heavier atoms in the maps illustrated in Fig. 5.

The numbers in parentheses are peaks found outside the fragment.

Map	Fig. 5(a)	Fig. 5(b)	Fig. 5(c)	Fig. 5(d)	Fig. 5(e)
O	107 (83)	179 (141)	197 (155)	207 (171)	227 (190)
N	94 (69)	173 (134)	198 (154)	200 (169)	232 (191)
C	16 (14)	17 (16)	91 (50)	16 (15)	99 (73)
C^{α}	64 (51)	82 (57)	127 (83)	106 (87)	171 (136)
C^{β}	37 (31)	70 (51)	103 (70)	81 (66)	114 (95)
S, Mg	19 (15)	22 (17)	22 (17)	23 (18)	22 (18)

there are substantial changes in the relative orientations of the two domains between the *T. cruzi* and the *C. jejuni* enzymes. The final map correlation of 0.66 was nevertheless a considerable improvement on the starting model, giving a density map that was much improved compared with that from the MR model.

A third approach involved rigid-body refinement of the initial MR model before applying *ACORN*. The r.m.s. deviation was 1.38 Å for 165 residues and the initial phase error was 52.6° with a more even resolution distribution. *ACORN* refinement now reduced the phase error to 36.7° after 27 cycles of DDM0. Further refinement using the enhanced density-modification procedures reduced it to 33.4° after a total of 140 cycles. The resultant maps, with a correlation coefficient of 0.77, could be easily built using automated procedures.

8.3. Pseudoazurin (PDB code 1paz)

This is a known structure, space group $P6_5$, with data to 1.55 Å resolution containing 123 residues and one Cu atom in the asymmetric unit (Petratos *et al.*, 1988). The new *ACORN* program used a random search procedure with native data only to locate the Cu atom, a position that gave an initial mean phase error of 73.8° for reflections with $|E| > 0.8$. The reflections were extended to 1.0 Å by method (i), yielding 55 367 extended reflections compared with 19 625 observed.

In the refinement procedure, DDM0 was carried out for 25 cycles, at which stage CC_s began to fall again and the mean phase error had fallen slightly to 71.4°. Two cycles of SER were then carried out, which had the effect of making the CC_s lower and the mean phase error higher, although the latter information would not be available for an unknown structure. After the SER cycles DDM0 was recommenced until changes in CC_s were sluggish, when SER cycles were again introduced. By cycle 192 there had been ten interventions of SER and the mean phase error had fallen to 26.4°. The same general procedure was then continued, but with DDM1 replacing DDM0, until cycle 256, at which stage the mean phase error had reduced to 25.5°. The refinement process was then continued with the addition of density-enhancement with DDM1 from cycle 257 to 295, reducing the phase error further to 22.0°. Finally, from cycle 296 to 301 DDM2 was used to give a final mean phase error of 21.6°. The final weighted phase errors in resolution shells are illustrated in Fig. 6. The whole process took less than 2 h on a laptop (Toshiba Tecra M1).

DDM0 was replaced by DDM1 or DDM2 when the $CC_s > 0.15$, but further tests will be needed to select the best protocol.

8.4. Gliding protein (PDB code 1j3w)

ACORN was applied to the gliding protein-MGLB from *Thermus thermophilus* (Lokanath *et al.*, 2004). The structure belongs to space group $C222_1$ with four molecules in the asymmetric unit and the observed data reach 1.5 Å. It has 535 residues with 4665 atoms.

The first 150 main-chain atoms in the PDB file (3.2% of the structure) were taken as a starting fragment, giving an initial mean phase error of 71.3°. The number of extended reflections to 1.0 Å was 244 485, giving a ratio $n_{obs}/(n_{obs} + n_{ext})$ of 0.30. After 85 cycles of DDM0 the phase error was reduced to 29.8° and the enhanced density-modification steps were then carried out and followed by DDM0. After 136 cycles, with five interventions of enhanced density modification, the value of CC_s was 0.211 and the mean phase error was 24.0°. Density-enhancement procedures followed by DDM1 were then carried out to cycle 162, at which stage the mean phase error was 19.7°. Finally, DDM2 was used from cycles 163 to 168, reducing the phase error to 19.6°. The *ACORN* map correla-
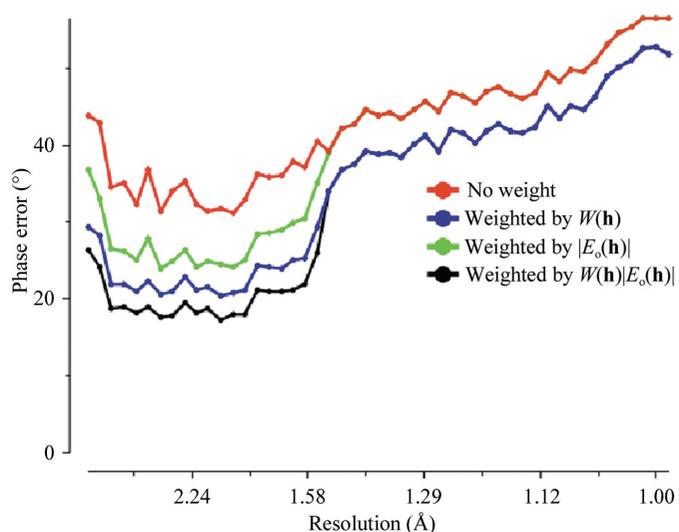
Figure 6
The final weighted phase errors from *ACORN* for 1paz in resolution shells.

tion coefficient was 0.8838 for the complete structure, 0.9243 for the main chain and 0.8556 for side chains. As for 1bxo, the phases for extended reflections were by no means random and the quality of maps can be improved by adding the extended reflections with estimated magnitudes, with appropriate weights and with phases coming from the refinement. Fig. 7 shows constant density surfaces for an F map with observed data to 1.5 Å and an E map which includes the extended reflections to 1.0 Å. The F map was easily interpreted or traced automatically by computer and the E map showed most individual atoms clearly.

8.5. Exo-mannosidase (PDB code 1uuq)

This is a known structure, space group $P2_12_12$, solved in York using data to 1.5 Å resolution. It has 440 residues with 16 methionines (three in two conformations) in the asymmetric unit (Dias *et al.*, 2004). All data were collected from a single crystal. Measurements at the Se peak wavelength were available to 1.6 Å and for the remote wavelength to 1.5 Å resolution. The peak data were used to locate the selenium positions and estimate initial experimental phases at a range of resolutions using *SHELXD* and *SHELXE*. Starting from these phases, the new *ACORN* program was applied in two different modes.

The first used the experimental phases determined to 2.5, 2.1 and 1.6 Å by *SHELXE* (Sheldrick, 2002) as a starting point with phase errors of 49.3° (16 105 reflections), 41.6° (27 284 reflections) and 28.7° (74 907 reflections), respectively. (For comparison, using *SHELXE* to extend the phases to 1.5 Å using the native and the 1.6 Å anomalous set the mean phase error was 29° for all 74 907 reflections). The reflections were extended to 1.0 Å by method (i), yielding 176 780 extended reflections compared with 74 865 observed. The standard

procedure rapidly improved and extended the phase set in all three cases, requiring only a few cycles of DDM0. The initial and final phase errors for all reflections are illustrated in Fig. 8. The phase error for the strong reflections $|E| > 0.8$ (38 372 reflections to 1.5 Å resolution) reduced to 29.6° after 38 cycles of DDM0, 24.9° after 18 cycles of DDM0 and 20.4° after 11 cycles of DDM0. The resultant map was of excellent quality. This is an extremely cost-effective use of *ACORN*: the SAD phasing set could be carefully measured to a limited resolution, conventional density-modification procedures used to refine the phases and excellent high-resolution phases generated very quickly from *ACORN*.

The second approach used the Se positions alone as a starting fragment. These gave an initial mean phase error of 74.1° for the strong reflections $|E| > 0.8$. The refinement procedure used cycles of DDM0, with single passes of SER carried out whenever CC_s stopped increasing and slowly reduced the phase error. The improvement was extremely sluggish, but these refinements were carried out in a fully automatic mode. The procedure continued until cycle 2712, when the CC_s was 0.17 and the mean phase error was 30°. The same general scheme was continued, but with DDM1 replacing DDM0 from cycle 2712 to 2756 until a final CC_s of 0.25 and mean phase error of 21.0° were obtained. The whole process took about 9 h on a laptop (Toshiba Tecra M1) but with no human intervention. An F map with observed data to 1.5 Å and an E map with the extended reflections to 1.0 Å are shown in Fig. 9.

For these data, starting from the experimental SAD phases is much faster than from phases calculated from the positions of the Se atoms as an initial fragment. This contrasts with previous successful applications of *ACORN* with measured data to atomic resolution, for example positioning the S atoms using 2.5 Å resolution sulfur-SAD data, and using these as a starting fragment with the 1.15 Å resolution data from the same crystal (Gordon *et al.*, 2001).
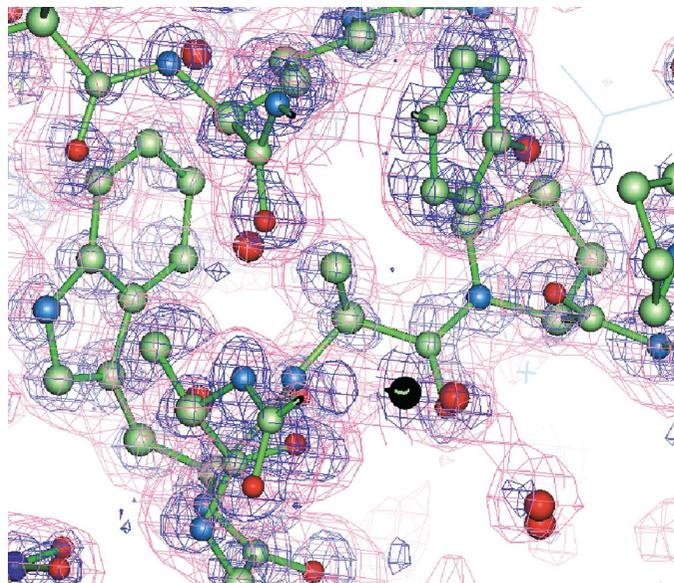
Figure 7
A final F map with observed data to 1.5 Å (red) for 1j3w and E map with data artificially extended to 1.0 Å (blue).
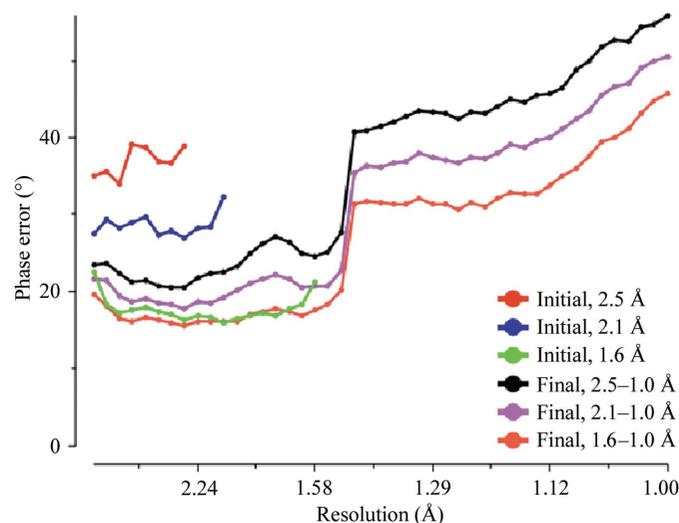
Figure 8
The initial and final phase errors weighted by WtE_x in resolution shells. $E_x = E_{obs}$ for observed reflections and $E_x = 1.0$ for extended reflections.

9. Discussion

The effectiveness of *ACORN* has previously been established for structures with atomic resolution data and several previously unknown structures have been solved by its use. However, we realise that comparatively few protein data sets are collected to atomic resolution, but that if the resolution limit for *ACORN* could be raised to 1.5 Å or thereabouts, the application would extend to a much greater proportion of structures. For this reason, we set ourselves the objective of lowering the resolution limit of application. What has been demonstrated here is that if a suitable fragment can be found consisting typically of 10% of the structure, then *ACORN* can still give considerable phase improvement with lower resolution data. Rajakannan *et al.* (2004) tested *ACORN* with 1.9 Å resolution data in solving a 39.5 kDa structure; starting with a fragment consisting of 14% of the structure, they obtained a map with a correlation coefficient around 0.5 that was of too low a quality to interpret directly. However, it gave a successful starting point for *ARP/wARP*.

Our new improved *ACORN* with the enhanced density-modification techniques offers a viable approach using data of resolution 1.7 Å or better. In particular, it should be very effective for structures containing heavier atoms whose positions can usually be readily determined. If heavier atoms are not present in the structure, then the barrier to using the improved *ACORN* is that of finding a suitable starting fragment: this will be the basis of future work.

During the preparation of this paper, but before its submission, a contribution by Caliandro *et al.* (2005) appeared which dealt with phasing beyond the resolution limit, a component of the present work. This approach was different

to our own and confirmed the effectiveness of the procedure. They estimate the extended *E* amplitudes based on values calculated from the current map which depend on the phase error. This corresponds to procedure (iii) discussed in §5. In contrast, we found that the best approach was to set the extended amplitudes to their mean expectation value of 1, which is independent of the phase error. In the Caliandro paper, the number of extended amplitudes is not allowed to exceed 75% of the measured number, whereas for some of our applications, *e.g.* the case of 1w2y, the extended/measured ratio is as high as 3.5. However, the major difference in our studies lies in the quality of the starting model. Caliandro and coworkers typically started from phases with errors around 30–60° and gain a modest improvement, whilst we are able to start from a much poorer phase set from an earlier stage of the solution process.

A key feature of *ACORN* remains the significance of the CC estimate as a correct indicator of a successful application. It is straightforward to see when *ACORN* has produced a meaningful increase in CC and perhaps more importantly to see when it has failed to do so.

Figure 9
A final *F* map with observed data to 1.5 Å (red) and *E* map with data artificially extended to 1.0 Å (blue) for 1uuq.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005). *Acta Cryst.* D**61**, 556–565.
Cowtan, K. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Dias, F. M. V., Vincent, F., Pell, G., Prates, J. A. M., Centeno, M. S. J., Tailford, L. E., Ferreira, L. M. A., Fontes, C. M. G. A., Davies, G. J. & Gilbert, H. J. (2004). *J. Biol. Chem.* **279**, 25517–25526.
Ding, J., Frasere, M. E., Meyer, J. H. & Bartlett, P. A. (1998). *J. Am. Chem. Soc.* **120**, 4610–4621.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J.-X. & Zheng, C.-D. (2000). *Acta Cryst.* D**56**, 1137–1147.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Lokanath, N. K., Shiromizu, I., Ohshima, N., Nodake, Y., Sugahara, M., Yokoyama, S., Kuramitsu, S., Miyano, M. & Kunishima, N. (2004). *Acta Cryst.* D**60**, 1816–1823.
McAuley, K. E., Yao, J.-X., Dodson, E. J., Lehmbeck, J., Ostergaard, P. R. & Wilson, K. S. (2001). *Acta Cryst.* D**57**, 1571–1578.
Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., González-Pacanowska, D. & Wilson, K. S. (2004). *J. Mol. Biol.* **342**, 1583–1597.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Petratos, K., Dauter, Z. & Wilson, K. S. (1988). *Acta Cryst.* B**44**, 628–636.
Rajakannan, V., Selvanayagam, S., Yamane, T., Shirai, T., Kobayashi, T., Ito, S. & Velmurugan, D. (2004). *J. Synchrotron Rad.* **11**, 358–362.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Refaat, L. S., Tate, C. & Woolfson, M. M. (1996). *Acta Cryst.* D**52**, 1119–1124.

Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.

Sim, G. A. (1960). *Acta Cryst.* **13**, 511–512.

Srinivasan, R. (1966). *Acta Cryst.* **20**, 143–144.

Wang, B.-C. (1981). *Acta Cryst.* A**37**, C11.

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

Yao, J.-X. (2002). *Acta Cryst.* D**58**, 1941–1947.

Yao, J.-X., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2002). *Z. Kristallogr.* **217**, 636–643.