Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

*ACORN*2: new developments of the *ACORN* concept

E. J. Dodson[a]* and M. M. Woolfson[b]

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

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

Received 12 March 2009
Accepted 2 May 2009

The density-modification procedures incorporated in *ACORN*, available in the *CCP*4 package, have proved to be very successful in solving and refining high-resolution crystal structures from very poor starting sets. These can be calculated from a correctly positioned initial fragment containing between 1 and 8% of the scattering power of the total structure. Improvements of *ACORN*, reported here and incorporated in the program *ACORN*2, have lowered the size of the fragment required and examples are given of structures solved with only 0.25% of the scattering power in the fragment, which may be a single atom. Applications of *ACORN*2 to structures with space group *P*1 have shown the remarkable property that when the starting point is a pair of equal atoms, or even a single atom placed at the origin, the refinement process breaks the centric nature of the initial phases and converges to phases corresponding to one of the two possible enantiomorphs. Examples are given of the application of *ACORN*2 to the solution and/or refinement of a number of known trial structures and to the refinement of structures when phases are available either from MAD or from a molecular-replacement model.

1. Basic *ACORN*

Many methods of density modification have been shown to be valuable for phase improvement. The density-modification procedure *ACORN* (Foadi *et al.*, 2000; Yao *et al.*, 2002, 2006; Jia-xing *et al.*, 2005), following the tradition of such methods, was developed for proteins for which diffraction data to atomic resolution (better than 1.2 Å) were available. At this resolution, individual atoms can be resolved in the electron density. The starting set of phases can be from any standard source: heavy atoms located by the Patterson function, moderately heavy atoms located by physical methods (*e.g.* from anomalous differences), a molecular-replacement (MR) model or a known fragment of the structure oriented and positioned by some MR technique. These phases are usually accompanied by weights, w_e, indicating their reliability and the starting density map can have Fourier coefficients of either $w_e|E_o|$, where $|E_o|$ is the observed normalized structure amplitude, or $|E_oF_o|E_{frag}$, where E_{frag} is the normalized structure factor for the fragment. The latter coefficients are those of a sum-function Patterson superposition map and experience shows that the mean phase errors obtained from the first *ACORN* step using this map are usually 1–2° lower than those obtained from the normal weighted map.

Nowadays, many structures have data sets in the resolution region 1.3–1.6 Å where *ACORN* refinement was previously either partial or impossible. A way of overcoming this restriction is to artificially extend the data to, say, 1 Å, giving each extended normalized structure amplitude an expectation value (Caliandro *et al.*, 2005; Yao *et al.*, 2006; Sheldrick, 2008). In the case of *ACORN* the value given is 1.0, the root-mean-square value for *E*s. If there are observational gaps within the observed resolution range then these gaps can also be filled in the same way and considered as part of the extended data. The data are now divided into three categories with limits chosen to give a reasonable spread of reflections in the following zones.

(i) $E_{\rm L}$, the large observed normalized structure factors with amplitudes greater than

$$E_{\rm lim} = 0.15 + \frac{1}{r}, \tag{1}$$

where r is the resolution of the data in angstroms.

(ii) $E_{\rm S}$, the smaller observed normalized structure factors with magnitudes between some lower limit, usually about 0.2, and less than $E_{\rm lim}$.

(iii) $E_{\rm ext}$, the structure factors for the extended reflections.

Starting with the initial density map calculated with estimated phases for the set $E_{\rm L}$, the steps in the refinement process are as follows.

(i) Density modification. The normal method of density modification, called DDM0, is

$$\rho' = \begin{cases} 0 & \text{if } \rho \leq 0 \\ \rho\tanh[0.2(\rho/\sigma)^{1.5}] & \text{if } \rho > 0 \\ K\sigma & \text{if } \rho' > K\sigma \end{cases}, \tag{2}$$

where σ is the standard deviation of the map density. The value of K is the greater of three times the iteration cycle number, with the latter having a maximum value of 15, or 0.8 times the maximum map density in σ units.

It is possible to introduce other density-modification procedures, designated DDM1 and DDM2, in which the density-map coefficients are some modification of $2F_{\rm o} - F_{\rm c}$. These are beneficial in later stages of the refinement. After each cycle of refinement a figure of merit, $CC_{\rm S}$, is calculated (see step iii). If the figure of merit ceases to increase, then a process called density enhancement is introduced that tends to bring up areas of weak but significant density. Another phase-refinement process, Sayre equation refinement (SER), can also be included if the figure of merit is changing but very slowly. Full details of DDM1, DDM2, density enhancement and SER are given in Yao *et al.* (2006).

(ii) The modified map is transformed to give new phase estimates and normalized map structure factors $E_{\rm c}$. The set of smaller normalized structure factors, $E_{\rm S}$, that has not been used for density calculations is now used to calculate a figure of merit that is in the form of a linear correlation coefficient,

$$CC_{\rm S} = \frac{\langle|E_{\rm c}E_{\rm S}|\rangle - \langle|E_{\rm c}|\rangle\langle|E_{\rm S}|\rangle}{\sigma_{\rm c}\sigma_{\rm S}}, \tag{3}$$

where the σs are the standard deviations of the normalized structure factors. When the value of $CC_{\rm S}$ reaches some designated limit (say 0.4), ceases to improve over a number of cycles or the mean phase shift between cycles is less than 0.5°, the refinement process is stopped.

(iii) A new weighted density map is calculated with both the large and extended normalized structure factors. The weights for the large structure factors are

$$w_{\rm L} = \tanh(0.5|E_{\rm c}E_{\rm o}|) \tag{4}$$

and those for the extended structure factors are

$$w_{\rm ext} = \tanh(0.5X_{\rm ext}), \tag{5a}$$

where

$$X_{\rm ext} = \left(\frac{n_{\rm obs}}{n_{\rm ext} + n_{\rm obs}}\right)^{1/2} |E_{\rm c}| \tag{5b}$$

and $n_{\rm obs}$ and $n_{\rm ext}$ are the numbers of observed and extended reflections, respectively. For the extended reflections the values of $|E_{\rm c}|$ are those obtained by scaling to a mean-square value of unity in shells.

The process then returns to step (i).

ACORN has been successful in refining poor initial phases for many protein structures. Initial phases have been based on fragments giving between 1 and 8% of the total scattering power. Continued development has led to improvements in the efficiency of *ACORN* and these will now be described.

2. *ACORN*2

The new version of *ACORN* is simultaneously somewhat simpler in application than the original version and also more effective. It uses a number of parameters that are based on characteristics of the structure, of the quantity of data and of the starting fragment. These characteristics are as follows.

N, the number of atoms in the unit cell.

B, the overall isotropic temperature factor in units of Å^2.

r, the resolution of the data in angstroms. This is correlated with B, but also depends on the efficiency of data collection.

$Z_{\rm max}$, the atomic number of the heaviest atom in the structure.

p, the proportion of the intensity arising from the fragment. This depends on the scattering angle, but the value used is that for zero scattering angle, *i.e.*

$$p = \frac{\sum\limits_{\rm fragment} Z^2}{\sum\limits_{\rm all\ atoms} Z^2}, \tag{6}$$

where the Zs are atomic numbers.

At this stage, it is worthwhile mentioning how the relationships between the refinement parameters and the various characteristics listed above were found for *ACORN*2. Essentially, it was an empirical process consisting of varying the parameters over ranges in a large number of experimental runs of *ACORN*2, determining the parameters that gave the best results and then finding a relationship that gave

approximately those parameters in terms of the characteristics that seemed to be relevant. Once these relationships had been found for all the parameters for a set of trial structures then they were tested with other structures to confirm that they gave reasonable results. The parameters that are found by the relationships are not necessarily the most efficient for any particular structure but they are the best that could be found to give phase refinement over a wide range of structures.

In place of the sum Patterson function used as a first step in *ACORN*, the function calculated in *ACORN*2 is an origin-removed sum Patterson function with coefficients $(|E_o|^2 - 1)E_c$. Experiments have shown that while it does not necessarily give a lower mean phase error than the *ACORN* function after the application of DDM0, it does tend to give a better final outcome.

2.1. Derived parameters

2.1.1. The lower level cutoff for density. The primary density-modification procedure is similar to DDM0, as described in (2), but differs in the lower cutoff limit and the upper truncation level. The zero cutoff limit in DDM0 and the down-weighting of low positive density is based on the proposition that such density is as likely to correspond to noise as to true density. In *ACORN*2 this discrimination against low density is enhanced by making density lower than $L\sigma$ equal to zero, where L has an initial positive value L_1 for the first cycle but gradually reduces to zero in successive cycles. The maps used in *ACORN*2 do not include the $E(000)$ contribution since this would simply raise the density by a constant amount and modify the value of L.

The value of L_1 determined by this process was

$$L_1 = 1.05\left(\frac{B}{r^2} - 1\right)\Phi(Z_{\max}). \tag{7}$$

The function Φ is a cubic function going through the four points $(\Phi, Z) = (0.84, 16)$, $(0.96, 30)$, $(1.15, 34.5)$ and $(1.24, 48)$. For $Z_{\max} < 16$ the value for $Z = 16$ (corresponding to sulfur) is used and for $Z_{\max} > 48$ the value for $Z = 48$ (corresponding to cadmium) is used. In initial trials the factor Φ was not included in (7) and solutions were found. However, further trials showed that somewhat modified values of L_1 gave better results and the factor Φ moves the calculated values of L_1 towards those better values. Values of L_1 are constrained to be in the range 3–6; if (7) gives a value outside that range then it is taken as the nearest boundary value.

The value of L is reduced from L_1 down to zero in n cycles where

$$n = \text{nearest integer to } \left(\frac{0.5}{p}\right) \tag{8a}$$

and then

$$L = L_1 - L_1^{c/n}, \tag{8b}$$

where c is the cycle number. This changes L slowly at first and then more rapidly later on.

2.1.2. The upper truncation level. The upper truncation level is also different from that in the original DDM0. For *ACORN*2 the upper truncation level for the density is given, in σ units, as

$$T = T_1 + c + 0.5c^2 \tag{9}$$

with a maximum value of 100. The value of T_1 depends on the heaviest atom in the structure and the resolution of the data and is given by

$$T_1 = \left(\frac{M}{N}\right)^{1/2}\frac{Z_{\max}}{14} \text{ constrained by } 3 \leq T_1 \leq 15, \tag{10}$$

where M is the number of observable reflections within the resolution sphere and N is the number of atoms in the unit cell. The unconstrained value of T_1 is approximately one-half of the expected peak height of the heaviest atom in an E-map with perfect phases. For an equal-atom structure and perfect phases an atomic peak height is $(M/N)^{1/2}$ (Yao *et al.*, 2002) and this is also the average peak height for a protein structure where all but a few non-H atoms are carbon, nitrogen and oxygen. Taking an average peak height for nitrogen ($Z = 7$), the peak height for an atom of atomic number $Z_{\max}$ would be approximately twice the value given by (10).

(9) changes the upper truncation level slowly at first and then more rapidly and this has been found to be better than a linear change.

To summarize, in place of DDM0 we have DDM0′, for which

$$\rho' = \begin{cases} 0 & \text{if } \rho \leq L\sigma \\ \rho\tanh[0.2(\rho/\sigma)^{\eta}] & \text{if } \rho > L\sigma \\ T\sigma & \text{if } \rho' > T\sigma \end{cases}. \tag{11}$$

It will be noticed that the exponent within the argument of the tanh term is now given as η rather than 1.5 as in the original *ACORN*.

2.1.3. Exponent of the argument within the tanh term. *ACORN* has operated successfully with data resolutions as low as 1.65 Å, but for the lowest resolutions it has been found to be beneficial to use a higher value of the exponent η within the tanh term of DDM0′. To determine η we use

$$\eta = 17.24(r - 1)^5 + 1.5. \tag{12}$$

For higher resolutions this gives a value that is little different from 1.5, as used in *ACORN*. For the lowest resolution at which *ACORN*2 has been applied, 1.65 Å, it gives $\eta = 3.5$.

2.1.4. Density disturbance: POWDM. The original *ACORN* had four alternatives to DDM0 that could be deployed either to improve phase estimates or to disturb the density to a condition that would allow further refinement if the refinement process either slowed or stalled. These were DDM1, DDM2, density enhancement and SER. Our experience now is that the detailed form of the disturbance is not very critical, as long as it is reasonably sensible. In place of the four density-modification procedures in *ACORN*, *ACORN*2 has the single density-modification process POWDM, which is

$$\begin{cases} \rho' = \left(\dfrac{\rho}{\sigma}\right)^{Q}\sigma & \text{if } \rho > L\sigma \\ \rho' = 0 & \text{if } \rho \leq L\sigma \end{cases}. \qquad (13a)$$

The value of Q is given by

$$Q = Q_0 - 0.004c, \qquad (13b)$$

where

$$Q_0 = 1.4448 + 0.02222Z_{\max} \qquad (13c)$$

and c is the cycle number. When Q falls to $0.75Q_0$ it is thereafter kept constant. The rather specific coefficients on the right-hand side of (13*c*) come from a least-squares fit of a linear function of $Z_{\max}$ to the optimum values of Q_0 found empirically for a range of structures. POWDM is applied every tenth cycle until the value of CC_S rises to 0.1, after which DDM0′ (12) is applied in every cycle. The application of POWDM disturbs the smoothness of the refinement process and sometimes seems to be damaging in the sense that its application leads to deterioration in the value of CC_S and an increase in the mean phase error. This is also a characteristic of SER, which is much more expensive to apply as it requires the evaluation of six FFTs. Despite the apparent harm done by including POWDM, tests have shown that excluding it, especially in the early stages of refinement, always makes matters worse, leading either to slower refinement overall or to a failure to refine at all. This point is illustrated later.

2.1.5. Limit on *E* terms. When the initial phases are calculated from a fragment, the smaller p is, the larger the expected initial mean phase error will be. To produce peaks in the electron-density map significantly greater than of random height then requires $E_{\rm lim}$ to be smaller for smaller p. To allow for this, *ACORN*2 chooses

$$E_{\rm lim} = \left(0.15 + \frac{1}{r}\right)(0.85 + 10p), \qquad (14)$$

with the proviso that $E_{\rm lim}$ is restricted to be in the range 0.75–1.25. For initial phases derived from a physical method or MR there is no obvious value of p that could be used in (14) so the original *ACORN* equation (1) is used.

Another important difference in the application of *ACORN*2 is in the weighting scheme that modifies the Fourier coefficients of the calculated maps. The detailed procedure is as follows.

(i) The density-modified map is transformed to give Fourier coefficients F_c.

(ii) Defining

$$S = \frac{1}{m}\sum_{\rm obs}|F_c|^2, \qquad (15)$$

in which the summation is taken over only the observed reflections, scaled Fourier coefficients, which will have a mean square value of unity for the observed reflections, are found from

$$E_c = F_c/S. \qquad (16)$$

The values of E_c for both the observed and extended reflections are found using this scaling process.

(iii) The weights for the following map are

$$w_{\rm obs} = \tanh(0.5|E_cE_{\rm obs}|) \qquad (17a)$$

for the observed reflections and

$$w_{\rm ext} = \tanh(0.5|E_c|)\tanh(0.5X_{\rm est}) \qquad (17b)$$

for the extended reflections, where $X_{\rm est}$ is given by (5*b*). This process down-weights the extended reflections more than in *ACORN*, but tests show that this enhanced down-weighting gives superior results.

As a final process that differs from that in *ACORN*, in the final stages of *ACORN*2, when the mean change of phase in a cycle is less than 0.01 radians the change in phase from the previous cycle to the new one is doubled by using

$$\varphi'_{\rm new} = 2\varphi_{\rm new} - \varphi_{\rm old}, \qquad (18)$$

in which $\varphi_{\rm new}$ is the phase derived from the transformed map and $\varphi_{\rm old}$ is the phase in the previous cycle. This has the effect

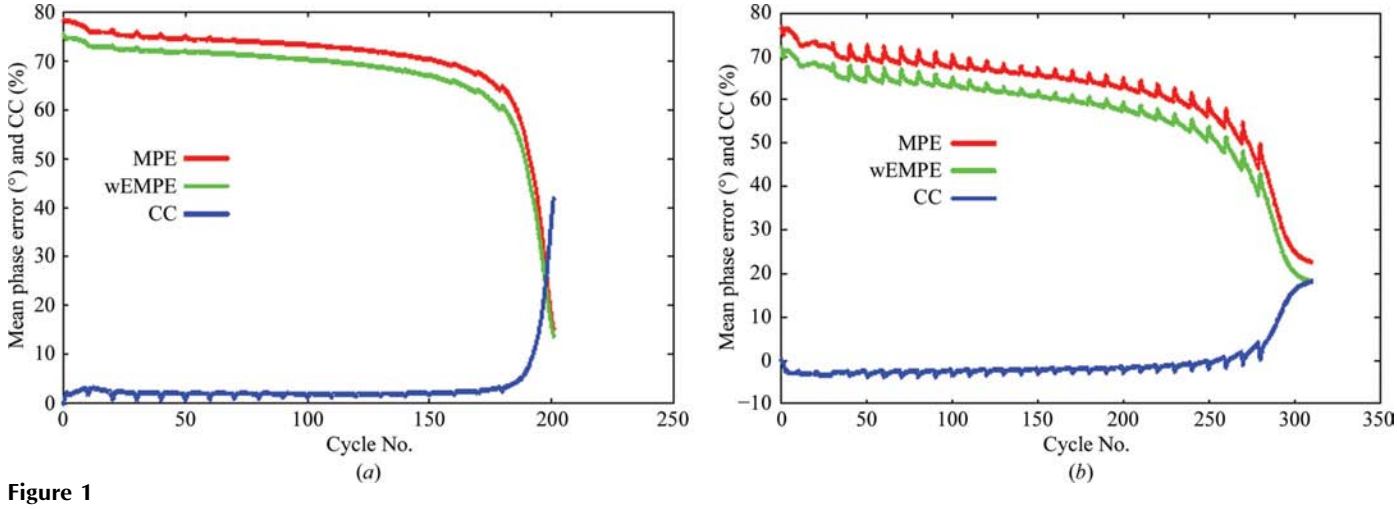


Figure 1
*ACORN*2 refinement of phases starting with small fragments. The quantity in parentheses is the value of p as a percentage. (*a*) 1itx (0.82%), (*b*) 1mgo (1.31%).

of accelerating the refinement process and sometimes reinvigorating it.

3. Trials with many-atom fragments

Most of the initial trials were carried out for structures with space group $P1$ for which the origin is arbitrary so that positioning a fragment was not necessary. However, this had the accompanying disadvantage that there were no symmetry elements to assist the phasing process. Some interesting and unexpected results were found, described in §§4 and 5, that are relevant to all space groups where the starting fragment is centrosymmetric. In all cases, the phases found from *ACORN*2 were compared with phases calculated from the complete deposited models without bulk-solvent correction.

3.1. Catalytic domain of chitinase A1 from *Bacillus circulans* WL-12

This structure belonged to space group $P1$, with unit-cell parameters a = 43.01, b = 46.84, c = 55.70 Å, α = 109.29, β = 95.45, γ = 116.68° (PDB code 1itx; Matsumoto *et al.*, 1999). The unit-cell content was 3516 non-H atoms including seven S atoms. The data resolution was 1.10 Å.

*ACORN*2 was run with a fragment of five S atoms, corresponding to p = 0.0082, which is less than 1% of the scattering power. The data were artificially extended to 1.0 Å. The progress of the refinement is illustrated in Fig. 1(*a*), in which the unweighted mean phase error (MPE), the mean phase error weighted by the product $w|E_o|$ (wEMPE) and 100 CC_S are plotted against the cycle number. The initial mean phase error for all reflections with $|E| \geq 0.987$ (the weighted value is given in parentheses) was 78.4° (75.2°) and the final mean phase error was 15.2° (13.8°).

3.2. Horse liver alcohol dehydrogenase Phe93Ala mutant

This structure belonged to space group $P1$, with unit-cell parameters a = 44.14, b = 51.31, c = 93.48 Å, α = 91.98, β = 103.01, γ = 109.81° (PDB code 1mgo; Rubach & Plapp, 2002). The unit-cell content was 6153 non-H atoms including four Zn atoms. The data resolution was 1.20 Å.

*ACORN*2 was run with a fragment of four Zn atoms, corresponding to p = 0.0131. The data were artificially extended to 1.0 Å. The progress of the refinement, shown in Fig. 1(*b*), is slower than that seen in Fig. 1(*a*) despite the higher value of p, which probably reflects the lower resolution of the data. The initial mean phase error for all reflections with $|E| \geq 0.965$ (the weighted value is given in parentheses) was 76.6° (72.1°) and the final mean phase error was 22.7° (18.2°).

4. Trials with two equal-atom fragments for *P*1

For structures containing several heavy atoms it is often possible from the Patterson function to find the vectors corresponding to one pair of heavy atoms, usually those with the lowest temperature factors. If this is performed for space group $P1$ by choosing the origin plus the position of the highest Patterson peak, there is the problem that when the atoms are equal the fragment is centrosymmetric. However, it turns out that from this starting point *ACORN*2 somehow develops a set of phases corresponding to one or other of the possible enantiomorphs, as the following examples will illus-

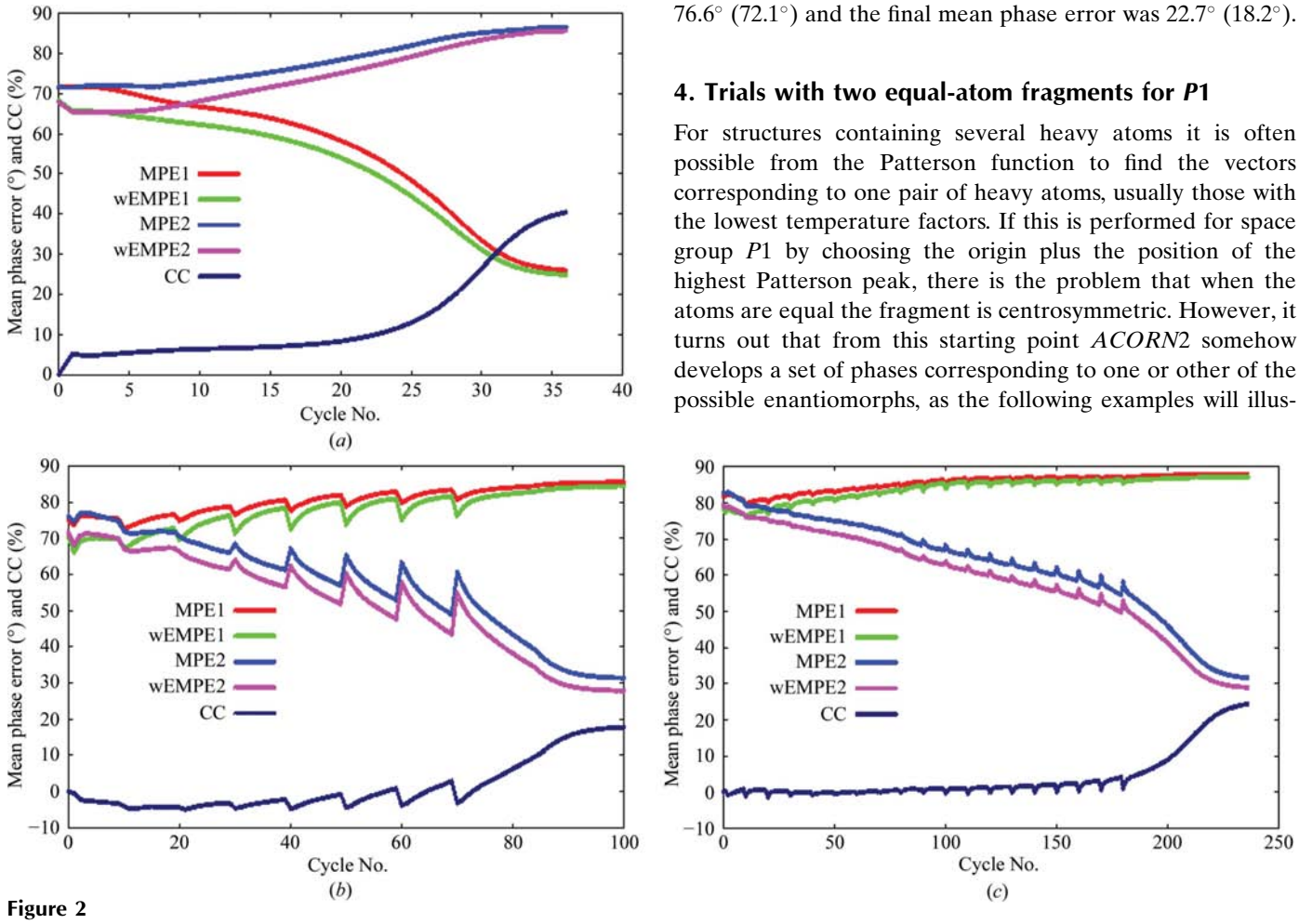


Figure 2
Refinement for structures with space group $P1$ and a starting fragment consisting of two equal atoms. The unweighted and wE-weighted mean phase errors are shown for both enantiomorphs. The structures are (*a*) 3ltz (1.11%), (*b*) 1o8b (1.62%), (*c*) 1f1g (0.50%).

trate. For all the examples that follow, with the exception of lysozyme, data were artificially extended to 1.0 Å.

4.1. *P*1 lysozyme

This structure belonged to space group *P*1, with unit-cell parameters $a = 26.65$, $b = 30.80$, $c = 33.63$ Å, $\alpha = 88.30$, $\beta = 107.40$, $\gamma = 112.20^\circ$ (PDB code 3ltz; Walsh *et al.*, 1998). The unit-cell content was 1001 non-H atoms including ten S atoms. The data resolution was 0.93 Å.

*ACORN*2 was run with a fragment of two S atoms, corresponding to $p = 0.0111$, with one atom at the origin. The result is shown in Fig. 2(*a*), in which the mean phase errors are now displayed for both enantiomorphs. It may be wondered how the phases develop into a noncentric form starting with what are essentially centric phases. The answer might seem to be that since all Fourier transforms are calculated on a grid and one of the fragment atoms is not on a grid point, the MPEs are initially not precisely the same for both choices of enantiomorph after the sum Patterson function has been calculated, modified and transformed. It will be seen that as the refinement progresses the phases evolve towards those of one enantiomorph, but not necessarily the one that is true for the crystal nor necessarily the one initially favoured by having a slightly smaller MPE. In this case, the enantiomorph with the slightly larger initial MPE is the one that is eventually selected. The initial mean phase error for all reflections with $|E| \geq 1.177$ (the weighted value is given in parentheses) was 71.7° (68.2°) and the final mean phase error was 25.9° (24.8°).

4.2. *Escherichia coli* ribose-5-phosphate isomerase

This structure belonged to space group *P*1, with unit-cell parameters $a = 42.05$, $b = 42.40$, $c = 60.19$ Å, $\alpha = 90.23$, $\beta = 100.98$, $\gamma = 98.98^\circ$ (PDB code 1o8b; Zhang *et al.*, 2003). The unit-cell content was 2960 non-H atoms including 11 Se atoms. The data resolution was 1.25 Å.

The input fragment was two Se atoms, corresponding to $p = 0.0162$. One atom was at the origin and the other was at the location of the largest Patterson peak. The progress of the *ACORN*2 refinement is shown in Fig. 2(*b*). The initial mean phase error for all reflections with $|E| \geq 0.961$ (the weighted value is given in parentheses) was 76.1° (71.7°) and the final mean phase error was 31.3° (27.8°). The high final MPE reflects the lower resolution of the observed data.

4.3. Oxidoreductase

This structure belonged to space group *P*1, with unit-cell parameters $a = 72.50$, $b = 72.48$, $c = 72.47$ Å, $\alpha = 109.20$, $\beta = 109.55$, $\gamma = 109.21^\circ$ (PDB code 1f1g; Haynes *et al.*, 1994). The unit-cell content was 7884 non-H atoms including six Zn atoms and six Cu atoms. The data resolution was 1.35 Å.

Again, a fragment of two atoms was taken with positions at the origin and the location of the highest Patterson peak. From the coordinates of the known structure these were identified as two Zn positions. The corresponding value of p is 0.0050, indicating that just 0.5% of the scattering power is in the initial fragment.

The refinement, shown in Fig. 2(*c*), takes more cycles of refinement than in the case of 1o8b, although the time taken is substantially greater since the cell is larger and thus requires more grid points in the FFTs. The initial mean phase error for all reflections with $|E| \geq 0.802$ (the weighted value is given in

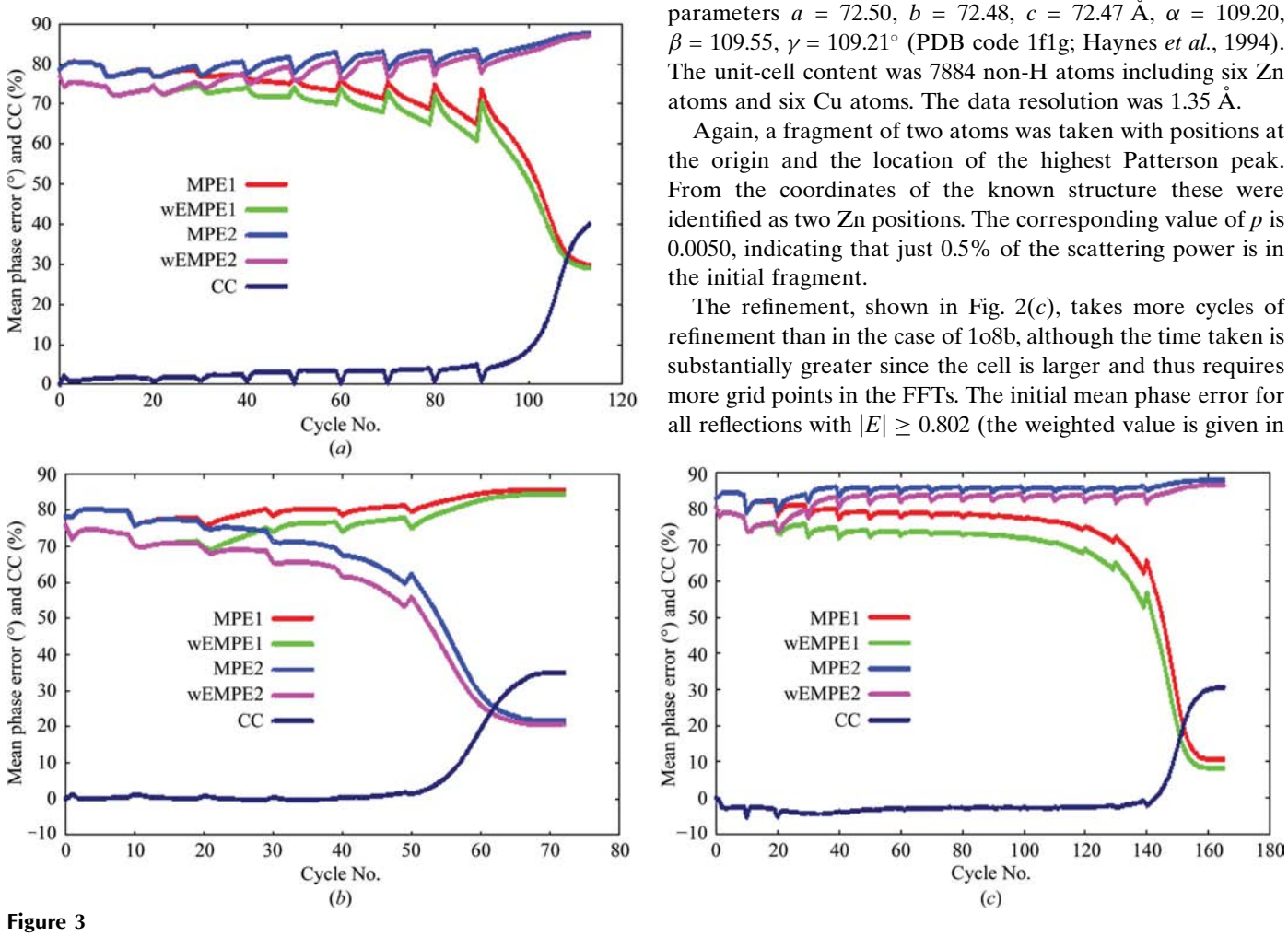


Figure 3
Refinement for structures with space group *P*1 and a starting fragment consisting of one atom situated at the origin. The unweighted and *wE*-weighted mean phase errors are shown for both enantiomorphs. The structures are (*a*) 1pwl (0.84%), (*b*) 1heu (0.65%), (*c*) 1het (0.255%).

parentheses) was 82.8° (79.4°) and the final mean phase error was 31.6° (28.9°). Once again, the relatively high final MPE reflects the lower resolution of the data.

5. Trials with single-atom fragments for *P*1

It was suggested that starting with a two-atom fragment for *P*1, the positions of one or both of the atoms from grid points began to differentiate between the two possible solutions. The next stage is to consider the situation when a single atom is chosen as the starting fragment and this is situated at the origin; there can be no possible initial bias towards one enantiomorph in this case.

5.1. Human aldose reductase complexed with NADP and minalrestat

This structure belonged to space group *P*1, with unit-cell parameters a = 40.019, b = 47.128, c = 47.367 Å, α = 75.71, β = 67.48, γ = 76.75° (PDB code 1pwl; El-Kabbani *et al.*, 2004). The unit-cell content was 3209 non-H atoms including one Br atom, 14 S atoms and three P atoms. The data resolution was 1.10 Å.

The starting fragment was the Br atom situated at the origin, corresponding to p = 0.0084. As might be expected, for 15 cycles the MPE with respect to the two enantiomorphs remained the same (Fig. 3*a*). Thereafter, there was convergence towards one enantiomorph. This indicates that there is a numerical instability in the refinement process. The first suspicion was that it might have been a consequence of round-off errors when the FFT routine is used and logically this can be the only explanation. Interestingly, when double-precision real numbers were used in the refinement the result was substantially the same. It seems that once the phases depart from a centric form, no matter by how little, they quickly embark on a passage towards one or other of the possible solutions. The initial mean phase error for all reflections with $|E| \geq 0.989$ (the weighted value is given in parentheses) was 78.8° (77.0°) and the final mean phase error was 29.9° (29.0°).

To confirm that this behaviour pattern was not atypical, two other structures were tested where the fragment was a single heavy atom placed at the origin.

5.2. Liver alcohol dehydrogenase containing cadmium and a hydroxide adduct to NADH

This structure belonged to space group *P*1, with unit-cell parameters a = 51.18, b = 44.57, c = 94.10 Å, α = 104.30, β = 101.07, γ = 70.67° (PDB code 1heu; Meijers *et al.*, 2001). The unit-cell content was 7618 non-H atoms including four Cd atoms and 58 S atoms. The data resolution was 1.15 Å.

The starting point was a single Cd atom at the origin, corresponding to p = 0.0065. Again, after about 15 cycles of refinement in which there was no bias towards either enantiomorph there was convergence towards one of them (Fig. 3*b*). The solution was obtained quite quickly in this case. The initial mean phase error for all reflections with $|E| \geq 0.933$ (the weighted value is given in parentheses) was 78.3° (76.1°) and the final mean phase error was 21.7° (20.6°). The final phases gave a map that showed the structure quite clearly. Fig. 4(*a*) shows the density calculated with a weighted *E*-map for the region around the origin (the position of the placed Cd atom) for the initial unrefined phases. The map is centrosymmetric with a large peak of height 405σ at the origin, around which there are significant diffraction ripples. There are also peaks in the other three Cd positions of heights 7.4σ, 6.2σ and 5.7σ, but of course in pairs about the origin so that the enantiomorph is undefined. Fig. 4(*b*) shows the original map in the region of two tryptophan residues and Fig. 4(*c*) the same region of the map at the end of the refinement. The complete final map shows individual atoms quite clearly and is a starting point for a straightforward interpretation and further refinement of the structure.

5.3. Liver alcohol dehydrogenase containing a hydroxide adduct to NADH

This structure belonged to space group *P*1, with unit-cell parameters a = 51.10, b = 44.40, c = 94.00 Å, α = 104.60,

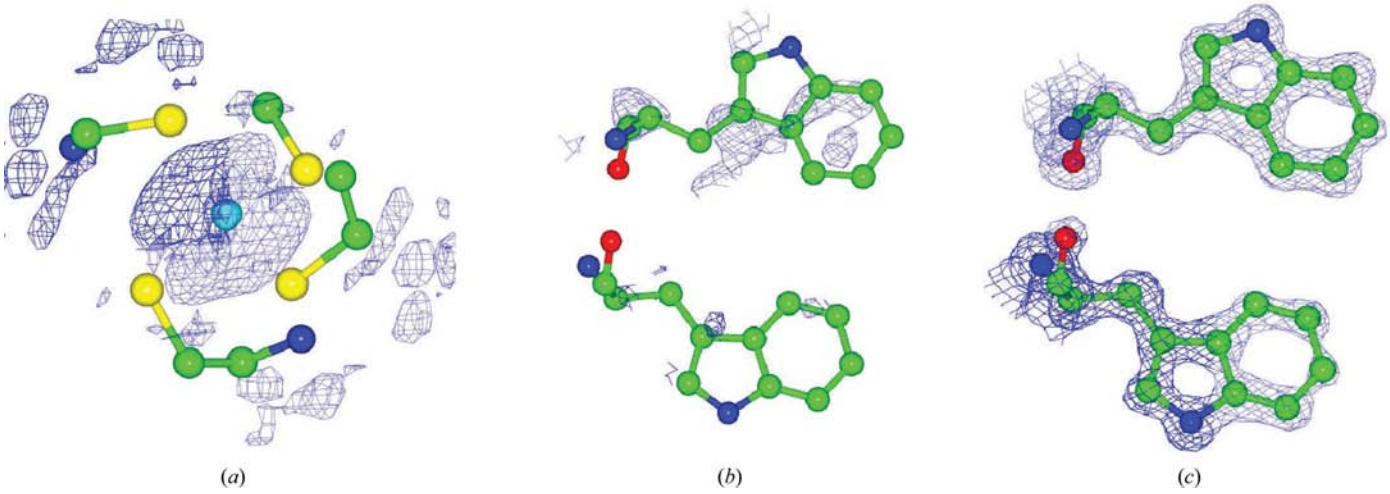


Figure 4
(*a*) Initial map in the region of the origin. (*b*) Initial map in the region of two tryptophan residues. (*c*) The same region as in (*b*) in the final map.

β = 101.50, γ = 70.50° (PDB code 1het: Meijers *et al.*, 2001). The unit-cell content was 7696 non-H atoms including four Zn atoms and 52 S atoms. The data resolution was 1.15 Å.

This structure is similar to 1heu but with Zn instead of Cd. The refinement began with a single Zn atom at the origin, corresponding to p = 0.00255, slightly over 0.25% of the scattering power in the initial single-atom fragment. The break point towards one enantiomorph came more quickly for this structure, after about 13 cycles (Fig. 3*c*). Despite the very small value of p and the correspondingly large initial value of the initial mean phase error, 83.0° (80.7°) for $|E| \geq 0.893$, the final MPE, 10.6° (8.2°), was remarkably low.

6. Trials in space group $P2_12_12_1$

Trials using *ACORN*2 in space group *P*1 have a special interest in that they bring out some rather unexpected behaviour with respect to determining a noncentrosymmetric structure from a centrosymmetric fragment. However, the majority of protein structures have more general space groups. Trials for three structures with the common space group $P2_12_12_1$ will now be described.

6.1. Zinc protease

This structure belonged to space group $P2_12_12_1$, with unit-cell parameters a = 55.10, b = 55.20, c = 37.24 Å (PDB code 1c7k; Kurisu *et al.*, 2000). The content of the asymmetric unit was 1134 non-H atoms including one Zn atom, one Ca atom and three S atoms. The data resolution was 1.0 Å.

This structure was quickly solved from the coordinates of either the Zn or Ca atoms. To explore the power of *ACORN*2, the position of a single S atom was taken, corresponding to p = 0.0049. The result of the refinement is shown in Fig. 5(*a*). Despite the high initial MPE, 83.8° (81.9°) for $|E| \geq 1.034$, the refinement is fairly rapid and gives a final MPE of 17.1° (13.5°). The structure also refined from a starting fragment of two O atoms (p = 0.00248) in 170 cycles to give a final unweighted MPE of 16°.

6.2. Endoglucanase Cel5a in complex with unhydrolysed and covalently linked 2,4-dinitrophenyl-2-deoxy-2-fluoro-cellobioside

This structure belonged to space group $P2_12_12_1$, with unit-cell parameters a = 54.35, b = 69.57, c = 77.09 Å (PDB code 1h2j; Varrot & Davies, 2003). The content of the asymmetric unit was 2903 non-H atoms including ten S atoms, some with partial occupancy. The data resolution was 1.15 Å.

The input fragment consisted of eight S atoms, three of which have three partial occupancies: 0.65, 0.38 and 0.38. The initial mean phase error for all reflections with $|E| \geq 1.000$ (the weighted value is given in parentheses) was 76.2° (72.3°) and the final mean phase error was 15.7° (13.5°). The value of p = 0.0131 is fairly high by the standards of most of the other structures reported here. However, there was no successful refinement with a smaller fragment, so this example can be regarded as only a partial success for *ACORN*2. It will be seen from Fig. 5(*b*) that the final MPE would have decreased if the refinement had been continued.

6.3. Transaldolase (YP_208650.1) from *Neisseria gonorrhoeae* FA 1090

This structure belonged to space group $P2_12_12_1$, with unit-cell parameters a = 42.14, b = 83.03, c = 89.79 Å (PDB code 3clm; Chen & Gotschlich, 2001). The content of the asymmetric unit was 3295 non-H atoms including eight Se atoms, all with partial occupancy, and one Cl atom. The data resolution was 1.14 Å.

The starting fragment for this structure was the Cl atom plus the Se atoms, which had partial occupancies of 0.75, 0.60, 0.42, 0.33, 0.27, 0.24, 0.23 and 0.15. This gave p = 0.01312. Such a starting point may have been achievable by the use of anomalous difference data. The refinement was very similar to that shown in Fig. 5(*b*). The initial unweighted MPE was 75.9° (72.1°) for $|E| \geq 1.008$ and fell to 13.8° (12.4°) after 438 cycles.

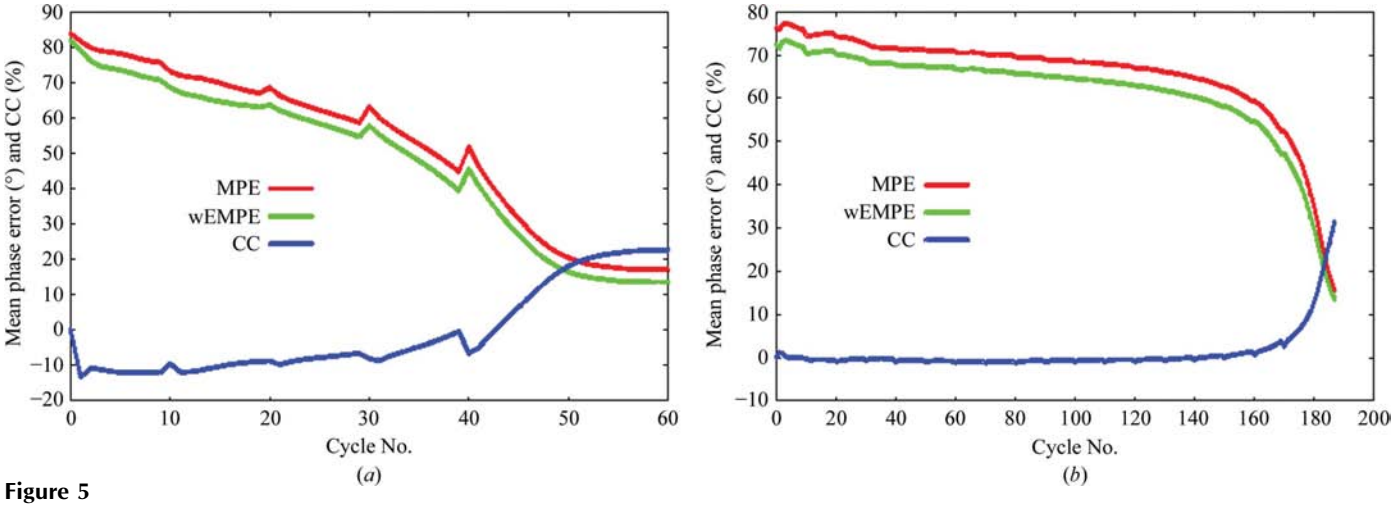


Figure 5
Refinement of structures with space group $P2_12_12_1$. (*a*) 1c7k starting with the position of a single S atom (0.49%). (*b*) 1h2j starting with the positions of eight S atoms (1.31%).

7. Trials with physically derived initial phases

The availability of tuneable synchrotron radiation has seen a spectacular rise in the use of anomalous measurements to derive experimental phases, using MAD or SAD as a solution technique followed by density modification using programs such as *DM* (Cowtan, 1999). The characteristic of this method and also of MIR is that initial phases typically have MPEs in the 50–60° range for the low-resolution data, but these increase to 80–90° for the highest resolution range. For these cases *ACORN*2 is used without the use of the POWDM step every tenth cycle, using only DDM0′ in all cycles. Also, as previously mentioned, since there is no determinable value of p we use (1) to determine $E_{\rm lim}$ and we take $L_1 = 0$, the latter condition usually prevailing in the final stages of refinement when starting with a fragment. This is nearly equivalent to using the original *ACORN* except for the variation of the upper truncation level and there being no refinement process other than DDM0′.

Here, we give examples involving two structures illustrating refinement from MAD-derived initial phase estimates.

7.1. MazG nucleotide pyrophosphohydrolase (13816655) from *Sulfolobus solfataricus*

This structure belonged to space group $I4_122$, with unit-cell parameters a = 79.749, b = 79.749, c = 95.909 Å (PDB code 1vmg; Javid-Majd *et al.*, 2008). The content of the asymmetric unit was 827 non-H atoms including three Se atoms. The data resolution was 1.46 Å.

Using the initial phase estimates from MAD and *DM*, with *ACORN*2 the MPE fell from 57.9° (42.6°) to 16.2° (12.2°) in six cycles for all $|E| \geq 0.835$ (Fig. 6*a*).

7.2. Unknown protein from 2D-PAGE (spot PRS1; b2898) from *E. coli* K12

This structure belonged to space group $P1$, with unit-cell parameters a = 44.81, b = 50.08, c = 52.01 Å, α = 94.68, β = 115.29, γ = 116.46° (PDB code 1vly; Joint Center for Structral Genomics, unpublished work). The unit-cell content was 3021 non-H atoms including nine Se atoms, one Ca atom, one Cl atom and two S atoms. The data resolution was 1.30 Å.

Using the initial phase estimates from MAD and *DM*, with *ACORN*2 the MPE reduced from 54.1° (38.6°) to 20.6° (17.3°) in five cycles for all $|E| \geq 0.919$ (Fig. 6*b*).

8. Trials with phases from molecular replacement

Although starting phases derived from a molecular-replacement (MR) model can be very variable in quality, they have the same property as physically derived phases with much lower MPEs at low resolution than at high. Experience has shown that using the same parameters in *ACORN*2 gives good results. The phases derived from a fragment show the fragment very well in the initial electron-density maps but may show very little interpretable density elsewhere, whereas MR-derived phases give initial maps that tend to spread the error in electron density over the whole volume of the cell. The *ACORN*2 refinement enhances the correct features of the solution and eliminates the incorrect. We illustrate *ACORN*2 refinements starting with MR phases for three structures, for the first two of which the fragment-based solutions have already been described.

8.1. Oxidoreductase (PDB code 1f1g; Haynes *et al.*, 1994)

The model structure was superoxide dismutase (PDB code 1sos, chain *A*). The initial mean phase error was 75.3° (71.6°) and after 132 cycles of refinement this fell to 20.7° (17.2°) (Fig. 7*a*). It is interesting to note that the final mean phase error is appreciably less than that obtained by starting with fragment-derived phases.

8.2. Zinc protease (PDB code 1c7k; Kurisu *et al.*, 2000)

The model structure was a member of the peptidase M10 family (PDB code 1hv5, chain *A*). The initial mean phase error was 82.5° (81.6°) and after 125 cycles of refinement this fell to 18.3° (14.8°) (Fig. 7*b*).

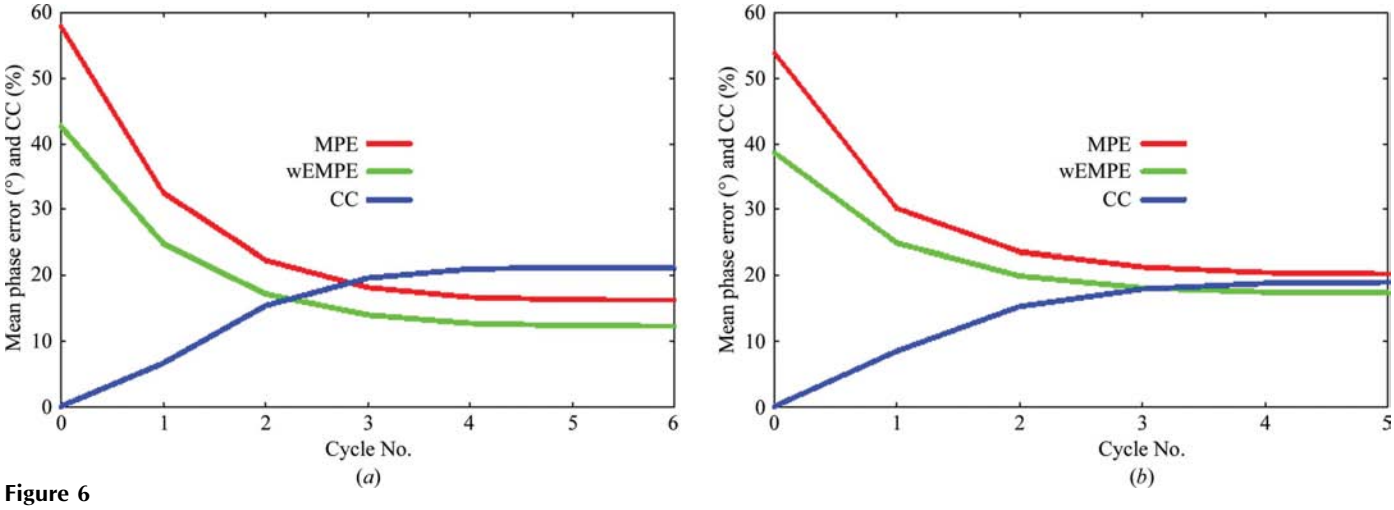


Figure 6
*ACORN*2 refinements starting with MAD-derived phases. (*a*) 1vmg, (*b*) 1vly.

8.3. *Campylobacter jejuni* dUTPase

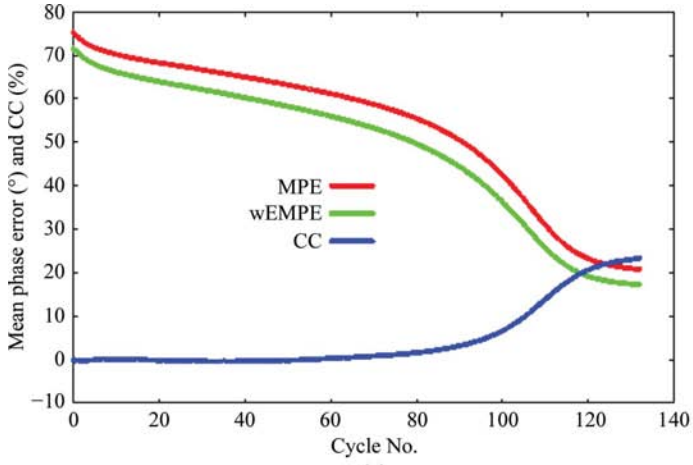
This structure belonged to space group $P2_12_12_1$, with unit-cell parameters a = 66.96, b = 70.63, c = 92.85 Å (PDB code 1w2y; Moroz *et al.*, 2004). The content of the asymmetric unit content was 4069 non-H atoms including five Mg atoms. The data resolution was 1.65 Å. The model structure was *Trypanosoma cruzi* dUTPase in complex with DUDP (PDB code 1ogk). The initial mean phase error for $|E| \geq 0.756$ was 44.6° (37.7°) and after seven cycles of refinement this fell to 32.6° (25.8°) (Fig. 7*c*). The resolution of this structure was at the lower end of what *ACORN*2 can be expected effectively to deal with.

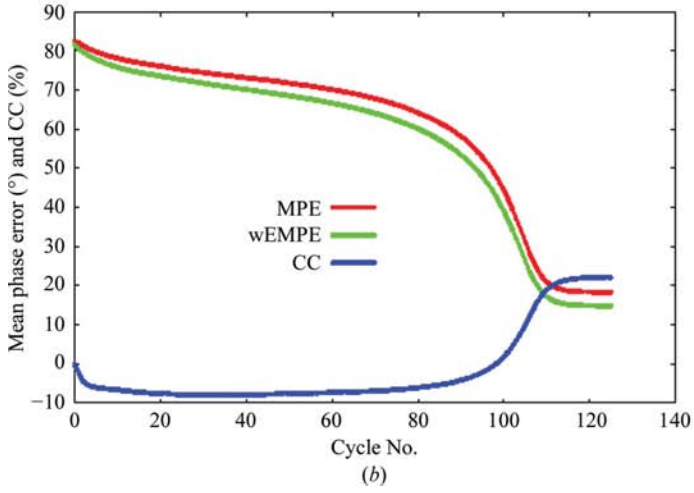
9. Concluding comments

The consistency with which *ACORN*2 changes a density map that seems to be almost random outside the fragment region to one that is readily interpretable by the application of a few automatically applied standard processes is quite remarkable. Much has been learned by trial and error and not everything in the *ACORN*2 process can be explained in an entirely rational way. It is a fact that the application of DDM0′ on its own may fail. This density-modification procedure can be justified theoretically as enhancing density that is more likely to be structurally significant relative to that more likely to be just random (Yao *et al.*, 2006). Adding the POWDM step may then turn a failing attempt at a solution into a successful one, but just how and why POWDM achieves this is difficult to explain. Fig. 8 shows the effect of using different values of Q_0 (13*c*) in the solution of 1heu. The solution marked Q = 1.0 actually corresponds to leaving out the POWDM step completely. It looks as though it might eventually reach a solution but only after a very large number of refinement cycles. The results for Q_0 = 2.0, 3.0 and 4.0 are all similar and the result from the value calculated by and used by *ACORN*2, 2.51, cannot be much improved.

Looking at the effect of POWDM in the 70th step of solving 1o8b (Fig. 2*b*), which increases the mean phase error by about 12°, one might conclude that the phase-refinement process was about to be ruined, but in fact it quickly recovers. Leaving out the POWDM step at this point gives a similar solution one cycle faster than leaving it in. With prior knowledge of a structure and hence the ability to track the mean phase error, there is no doubt that the *ACORN*2 process could be tailored for each structure to obtain a solution more quickly. However, the real situation of solving an unknown structure has to be carried out with only CC_S as a guide, which is not very informative in the early stages of solution. Hence, one must accept that when a recipe for the greatest overall success rate over a wide range of structures is applied then it may not always be the optimum procedure for all or indeed any of them. The aberrations, or apparent aberrations, seen in the solutions we present here are the price one has to pay for having a technique that is widely applicable.

A feature of *ACORN*2 that was not present to the same extent in the original *ACORN* is the use of knowledge of the heaviest atom in the structure to set the various parameters that control the refinement. The value of Z_{max} is used for setting L_1, T_1 and Q_0 and getting these values more or less right is important in order to obtain a solution efficiently or even to obtain a solution at all. For example, many of the





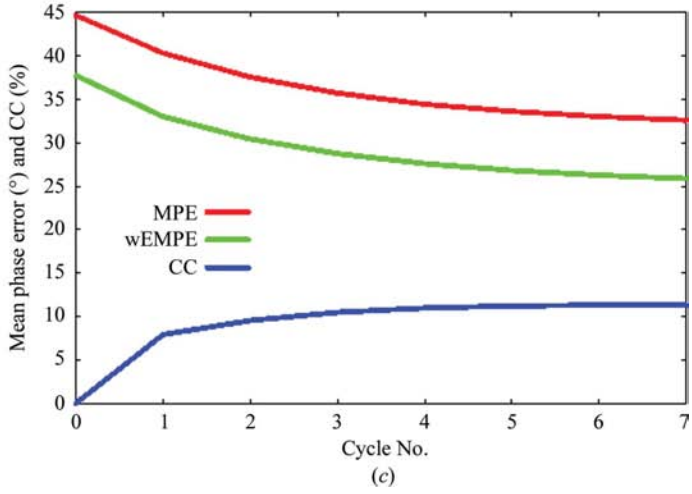


Figure 7
Refinement starting with MR-derived phases. (*a*) 1f1g, (*b*) 1c7k, (*c*) 1w2y

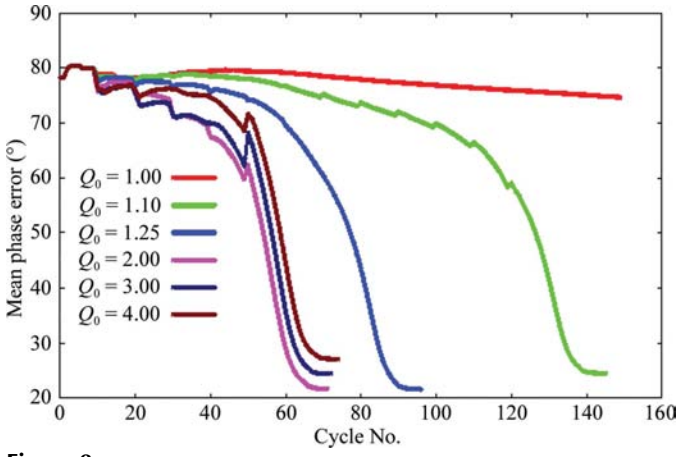


Figure 8
The effect of changing the value of Q_0 in the solution of 1heu.

structures solved here will give no solution with $L_1 = 0$, the value used by the original *ACORN*.

*ACORN*2 is a very flexible program that can be used in a variety of ways. We have seen from the *P*1 examples that in favourable cases where there are moderately heavy atoms in the structure one can obtain a virtually *ab initio* solution by taking a single atom at the origin as the starting point. This is equivalent to taking all starting phases equal to zero with the atomic number of the fragment atom just affecting the weights. In the case of the structure 1het, which contains 7696 non-H atoms (two similar chains each with 374 residues), the total scattering power in the 'fragment' amounted to just 0.25% of the whole. In less favourable cases taking just two fragment atoms, one at the origin and the second at the position of the highest Patterson peak, can also lead to a solution, which if not *ab initio* is nearly so. Where the starting point is either a single atom or a pair of equal atoms for space group *P*1, *ACORN*2 is able to reach a possible solution despite the initial centric phases. For some higher space groups just locating the position of a single atom in one asymmetric unit may also give a centrosymmetric fragment, *e.g.* for space group $P2_1$ just locating the coordinates x and z of a single atom from a Harker section. There is no reason to believe that *ACORN*2 would not yield a solution in such cases, as long as the single atom contains sufficient scattering power to give a suitable starting point.

An important feature of *ACORN*2 is that it is both automatic and quick in its application. The *ACORN*2 applications for all the examples given here were performed using a 3.4 GHz laptop and took between 2 min for the MR examples to somewhat over 3 h for 3clm, which required 438 refinement cycles. As previously mentioned, it is clear from the progression of CC_S during the refinements that during the early and middle stages it gives very little reliable indication of the progress of the refinement, so it may not be clear for some time whether or not a solution will eventually be found. The recommendation that we make is that whenever the situation looks as though it might be favourable for an *ACORN*2 application then it should be tried, with a maximum of around 1000 cycles that, depending on the size of the structure, might require up to 6 h of computing time on a good PC. The starting point could be a fragment with as little as 0.25% of the total scattering power. If the attempt is successful then it may greatly shorten the solution or refinement process; if it fails then there will have been very little waste of time and effort.

ACORN is being replaced by *ACORN*2 in the *CCP*4 system.

References

Caliandro, R., Carrozzini, B., Cascarano, G. L., De Caro, L., Giacovazzo, C. & Siliqi, D. (2005). *Acta Cryst.* D**61**, 556–565.
Chen, I. & Gotschlich, E. C. (2001). *J. Bacteriol.* **183**, 3160–3168.
Cowtan, K. (1999). *Acta Cryst.* D**55**, 1555–1567.
El-Kabbani, O., Darmanin, C., Schneider, T. R., Hazemann, I., Ruiz, F., Ok, M., Joachimiak, A., Schulze-Briese, C., Tomizaki, T., Mitschler, A. & Podjarny, A. (2004). *Proteins*, **55**, 805–813.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
Haynes, M. R., Stura, E. A., Hilvert, D. & Wilson, I. A. (1994). *Science*, **263**, 646–652.
Javid-Majd, F., Yang, D., Ioerger, T. R. & Sacchettini, J. C. (2008). *Acta Cryst.* D**64**, 627–635.
Jia-xing, Y., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2005). *Acta Cryst.* D**61**, 1465–1475.
Kurisu, G., Kai, Y. & Harada, S. (2000). *J. Inorg. Biochem.* **82**, 225–228.
Matsumoto, T., Nonata, T., Hashimoto, M., Watanabe, T. & Mitsui, Y. (1999). *Proc. Jpn Acad. Ser. B*, **75**, 269–274.
Meijers, R., Morris, R. J., Adolph, H. W., Merli, A., Lamzin, V. S. & Cedergren-Zeppezauer, E. S. (2001). *J. Biol. Chem.* **276**, 9316–9321.
Moroz, O. V., Harkiolaki, M., Galperin, M. Y., Vagin, A. A., González-Pacanowska, D. & Wilson, K. S. (2004). *J. Mol. Biol.* **342**, 1583–1597.
Rubach, J. K. & Plapp, B. V. (2002). *Biochemistry*, **41**, 15770–15779.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Varrot, A. & Davies, G. J. (2003). *Acta Cryst.* D**59**, 447–452.
Walsh, M. A., Schneider, T. R., Sieker, L. C., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1998). *Acta Cryst.* D**54**, 522–546.
Yao, J. X., Dodson, E. J., Wilson, K. S. & Woolfson, M. M. (2006). *Acta Cryst.* D**62**, 901–908.
Yao, J.-X., Woolfson, M. M., Wilson, K. S. & Dodson, E. J. (2002). *Z. Kristallogr.* **21**, 636–643.
Zhang, R., Andersson, C. E., Savchenko, A., Evdokimova, E., Beasley, S., Arrowsmith, C. H., Edwards, A. M., Joachimiak, A. & Mowbray, S. L. (2003). *Structure*, **11**, 31–42.