Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 23 June 2000
Accepted 3 October 2000

The pair-functional method for direct solution of molecular structures. II. Small-molecule tests

A. D. McLachlan

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England. Correspondence e-mail: admcl@mrc-lmb.cam.ac.uk

The new pair-functional direct method has been implemented and tested. Like the Patterson function, the pairing force has valuable imaging properties at high resolution. Two simple iterative algorithms were designed to refine on the total pair potential and the normalized intensity correlation coefficient of an atomic model. The first algorithm is a peak-picking method which selects the best-paired high peaks from a density map and then uses the strong reflections to generate a new Fourier filtered map. The second algorithm, the pair-and-square method, uses a tangent formula step instead of the Fourier and is a little more efficient. Computational experiments on a point-atom grid model, with perfect data, reached exact *ab initio* solutions for up to 600 atoms. Point-atom models were also solved by searching for reduced structures that contained as few as one quarter of the atoms. Seeded searches, guided by a small known fragment, solved up to 30000 atoms on the grid. Realistic tests on actual molecules showed that Sheldrick's [*Acta Cryst.* (1990), A**46**, 467–473] test structures of 50–200 atoms can be solved under a variety of conditions.

1. Introduction

The theory of the pair-functional method, described in an earlier paper (McLachlan, 2001*a*; paper I), aims to solve molecular structures in a new way by searching for clusters of correctly placed atoms in a specially designed statistical ensemble. The practical value of the method depends on its power to solve real structures, otherwise the principle might only be an interesting mathematical curiosity. As a first step in the direction of practical use, we now describe simple tests of a new search algorithm. The aim here was to arrive at a working trial solution of the target structure that would be good enough to be refined to completion by standard methods.

2. Theoretical background

The basic theory shows how to set up a unique many-body ensemble of N equal atoms in the crystal cell, which corresponds with the experimental X-ray measurements on the actual molecule. The Patterson function of the cell ensemble is matched to the experimental Patterson derived from the normalized X-ray intensities of the crystal. This condition appears as a linear constraint on the pair-correlation function of the atoms and thus makes it possible to generate a unique maximum-entropy ensemble for the system (McLachlan & Harris, 1961). The atoms adopt an equilibrium Boltzmann distribution under strong fictitious pair-interaction forces, like the distribution of an atomic fluid under strong forces at a finite temperature (Hansen & McDonald, 1986). The result is a cloud or superposition of many atomic conformations whose average Patterson matches the data. Individual conformations have random deviations from the correct intensities.

The statistical pairing force that keeps the atoms in position is fully determined by the experimental intensity data. That is, the pairing force is specified implicitly by the outcome of the maximum-entropy calculation. This makes it possible to simulate the effect of the forces on individual groups of atoms and to pick out clusters that are candidates for a complete molecular structure. The process of solving the molecule is a little like a dynamical simulation of an atomic fluid in the cell, in which a random collection of atoms organizes itself and eventually condenses into a correct ordered structure. In practice, the Fourier components of the force can be estimated reasonably accurately from the direct correlation function of the ensemble (Ornstein & Zernicke, 1914). An account of the theory for deducing the forces from the intensities will be given in paper III of this series (McLachlan, 2001*b*).

3. Test procedures

The first series of trials, described below, are computational tests on an ideal system, a set of equal point atoms on a fine grid, in a crystal cell of symmetry $P1$. This system is useful because it is simple to compute and it tests the basic ideas while maintaining complete control over complications like limited resolution or inaccurate experimental data. All these grid trials either achieve a perfect solution or else fail completely.

The second series of trials uses real high-quality data at atomic resolution from a well known test set of small molecules (Sheldrick, 1990). The quality of these solutions now depends critically on the fidelity of the model and the resolution of the data (McLachlan, 1999).

The progress of the solution process is monitored using three figures of merit. The first two, which assume no detailed knowledge of the correct solution, are the *total pair potential*, Ψ, and the *residual intensity correlation coefficient*, R_I. The third figure is the best *atom–atom overlap*, Q_{MT}, between superposed maps of the current trial model and the known solution. The intensity correlation coefficient and its gradient are defined in Appendix *A*.

4. The total pair potential

The first step in setting up the calculation is to convert the observed intensities $I_{\rm obs}(\mathbf{H})$ of the measured reflections $\mathbf{H}$ into unphased normalized structure amplitudes $|E_{\rm obs}(\mathbf{H})|$ (Blessing *et al.*, 1996, 1998). These will be called the target amplitudes $T(\mathbf{H})$.

$$|T(\mathbf{H})|^2 = |E_{\rm obs}(\mathbf{H})|^2. \tag{1}$$

In paper I, we used a continuous particle density $p_N(\mathbf{r})$ for N atoms in a cell. When this is replaced by a set of localized point atoms with occupancies $p_j(\mathbf{r}_j)$ at fixed centres $\mathbf{r}_j$, the total pairing potential of any atomic model becomes

$$\Psi_N(\mathbf{r}^N) = \sum_{i<j} p_i(\mathbf{r}_i)\psi(\mathbf{r}_i - \mathbf{r}_j)p_j(\mathbf{r}_j). \tag{2}$$

Here we have introduced the notation $p_j(\mathbf{r}_j)$ as a reminder that atom j is now localized at the point $\mathbf{r}_j$. Also, $\psi(\mathbf{u})$ is the pairing force between two atoms at a separation $\mathbf{u}$. The force can also be expressed in terms of its scaled Fourier components $\psi_E(\mathbf{H})$ for the measured reflections.

$$\psi(\mathbf{u}) = (1/N)\sum_{\mathbf{H}\neq 0} \psi_E(\mathbf{H})\exp(-2\pi i\mathbf{H}\cdot\mathbf{u}). \tag{3}$$

We use an empirical form of potential adapted from the direct correlation function of the pair ensemble. This is

$$\psi_E(\mathbf{H}) = \frac{|T(\mathbf{H})|^2 - 1}{|T(\mathbf{H})|^2 + T_{\rm low}^2}, \tag{4}$$

where $T_{\rm low}$ has a small constant value, typically $T_{\rm low}^2 = 0.25$, chosen to soften the strong repulsive force for very weak reflections. We note that $\psi_E(\mathbf{H})$ represents a wavelike attractive force between atoms for all strong reflections with $|T(\mathbf{H})|^2 > 1$ and provides a repulsive force for all weak reflections. In this way, the natural intensity fluctuations of the atoms in the paired-atom ensemble are modulated to fit the experimental intensity measurements. It is usual to subtract the origin peak from $\psi(\mathbf{u})$ and construct an originless force $\psi^0(\mathbf{u})$ with corresponding Fourier components $\psi_E^0(\mathbf{H})$, as described in paper I. Under these conditions, the total pair potential of any atomic model with normalized structure factors $E(\mathbf{H})$ is

$$\begin{aligned}\Psi_{\rm trial} &= \sum_{\mathbf{H}>0} \psi_E^0(\mathbf{H})|E(\mathbf{H})|^2 \\ &= \sum_{\mathbf{H}>0} \psi_E(\mathbf{H})\{|E(\mathbf{H})|^2 - 1\}.\end{aligned} \tag{5}$$

The target value to be matched is evidently

$$\Psi_{\rm targ} = \sum_{\mathbf{H}>0} \psi_E^0(\mathbf{H})|T(\mathbf{H})|^2. \tag{6}$$

During the structure search, the atoms are moved in the gradient map of the *local pairing field*. This is defined as

$$V(\mathbf{x}) = \sum_{j=1}^{N} \psi^0(\mathbf{x} - \mathbf{r}_j)p_j(\mathbf{r}_j). \tag{7}$$

The atoms may also be treated as one single-particle probability distribution with a continuous density $\rho(\mathbf{r})$ in place of the discrete point occupancies $p_j(\mathbf{r}_j)$ at $\mathbf{r}_j$. In the grid model, we calculate the local pairing field and the structure factors with a simple Fourier transform over the grid points. In the trials on real molecules, the atoms are represented by single broadened Gaussian peaks that extend over a range of points on the map grid (Agarwal, 1978; Isaacs & Agarwal, 1978).

5. Imaging properties of the local field

We now illustrate the key properties of the pairing force $\psi^0(\mathbf{u})$ and the local field $V(\mathbf{r})$ with contour plots for a simple ideal example structure of 20 equal atoms. For display purposes only, it is much clearer to use a two-dimensional object in place of a realistic three-dimensional one. So these 20 atoms have been placed randomly in a square cell of side 10 Å with a minimum collision diameter of 1 Å. The space group is *P*1. At a resolution of 1 Å, the ideal data set contains 162 reflections and all the peaks are well separated (Fig. 1). Here the density

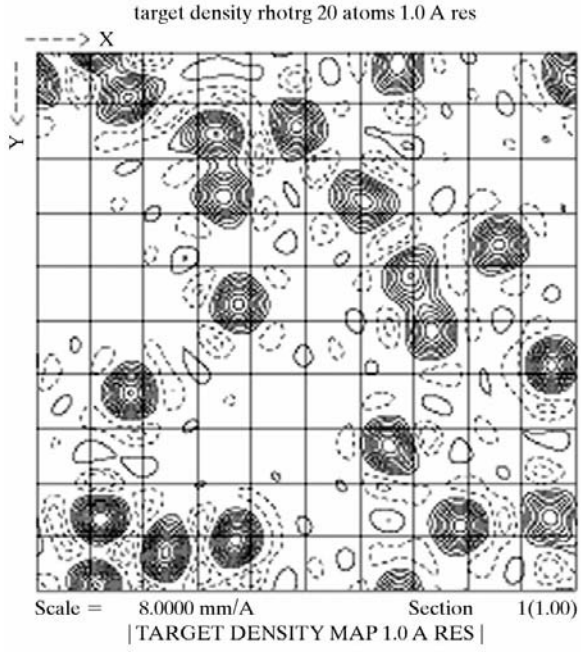

Figure 1
Density contour plot for ideal target molecule of 20 atoms in a 10 × 10 Å cell. Data truncated at 1.0 Å resolution. Negative contours are dashed.

has been synthesized from the exact normalized E structure factors, truncated at 1 Å resolution. The originless pairing force (Fig. 2) was calculated from the exact intensities, taking the same cutoff intensity $T^2_{\rm low} = 0.25$ as we normally use in real applications. $\psi^0(\mathbf{u})$ is a centrosymmetric function with a mean value of zero, which displays a complex pattern of positive and negative peaks. For comparison, the levelled originless Patterson (Fig. 3) has many similar peaks in almost the same positions. However, the pairing force peaks tend to be rounder and do not run together in such large islands. The general similarity between the appearance of the pairing force and Patterson maps holds under a wide range of conditions. For example, at a higher resolution of 0.8 Å, using 257 measured reflections, the two kinds of plot each contain many extra peaks, but look even more alike. Also, in a series of maps calculated with different values of $T^2_{\rm low}$, ranging from 10.0 to 0.025, there was a continuous gradation of type, gradually shifting from the Patterson function ($T^2_{\rm low} = \infty$) towards the ideal direct correlation function ($T^2_{\rm low} = 0$). The next plots illustrate the imaging properties. The local field $V(\mathbf{r})$ for any trial atomic model is the convolution of the model density $p_N(\mathbf{r})$ with $\psi^0(\mathbf{u})$ and tends to 'echo' the peak positions of correctly placed atoms. In particular, if the target structure itself is echoed, with its correct density $p_T(\mathbf{r})$, we obtain the echo field

$$V^T_{\rm echo}(\mathbf{x}) = \int p_T(\mathbf{x}-\mathbf{u})\psi^0(\mathbf{u})\,{\rm d}\mathbf{u}. \qquad (8)$$

This may be compared with the well known Patterson image map of the target

$$V^T_{\rm Patt}(\mathbf{x}) = \int p_T(\mathbf{x}-\mathbf{u})\Delta P^{(2)}_N(\mathbf{u})\,{\rm d}\mathbf{u} \qquad (9)$$

(Buerger, 1959; Simpson *et al.*, 1965). The pairing field echo of the target at 1 Å resolution (Fig. 4) correctly picks out almost every atom, with only a few spurious peaks (one atom has almost vanished and there is at least one strong spurious peak). The echo of the Patterson (Fig. 5) has peaks in similar positions but many of them are elongated. If the same maps are recalculated at 0.8 Å resolution, both sets of echoed peaks become very much clearer (not shown) but the pairing field echo is still a little sharper than the Patterson echo. These findings are typical of many varied trials. Lastly we show the echo of a subfragment of the true structure, containing ten randomly chosen atoms in their correct positions. The echo is rather noisy at 1 Å resolution, and so the maps are now plotted at 0.8 Å resolution for both the density (Fig. 6) and its echo (Fig. 7). As expected, the pairing field echo brings back all the ten fragment atoms, but it also points up all the other ten missing atoms. The echo is not perfect, as about five spurious peaks are as strong as the correct ones. At the lower resolution of 1 Å, nine fragment atoms and seven missing atoms still appear clearly but they are accompanied by a large number of spurious peaks. The exact numbers of well marked peaks are sensitive to the resolution and the contour levels of the maps.

To sum up, the pairing force has very good imaging characteristics at high resolution so that the pair field maps echo the correct structure and its subfragments rather faithfully. The noise in these maps is often less than in a Patterson echo. As a systematic structure search proceeds, the spurious peaks become weaker and disappear.

6. The correlation gradient map

The pair potential alone is not enough to guide the atoms into their correct positions. Model structures with a high potential can still have some intensities $|E(\mathbf{H})|^2$ that deviate strongly

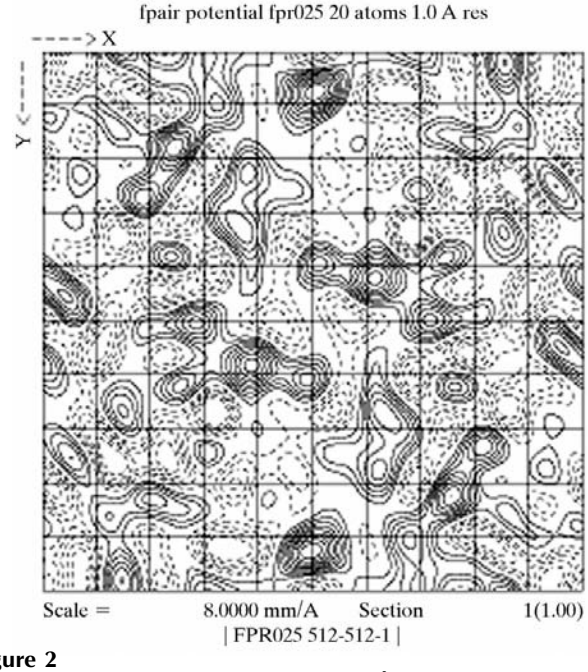

Figure 2
Pairing force $\psi^0(\mathbf{u})$ with $T^2_{\rm low} = 0.25$, at 1.0 Å.

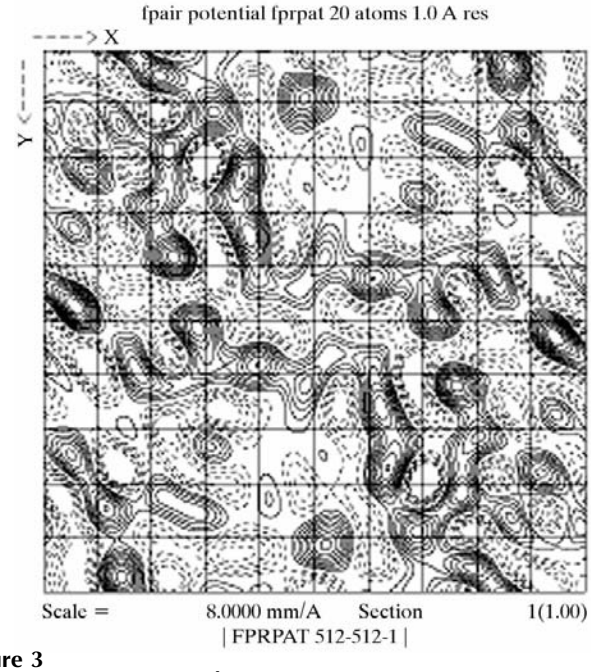

Figure 3
Originless Patterson at 1.0 Å.

from the target values $|T(\mathbf{H})|^2$ and which never converge. This may be because $\psi(\mathbf{u})$ is not estimated well or for other reasons. Therefore, we use the intensity correlation residual $R_I = (1 - C_I)$ as an additional active term in the refinement (Fujinaga & Read, 1987). Here C_I is the intensity correlation coefficient of the model and its spatial gradient $C(\mathbf{r})$ is a difference map that indicates how the electron density should change in order to improve the fit of $|E(\mathbf{H})|^2$ to $|T(\mathbf{H})|^2$. R_I is one of the most reliable figures of merit for the early and middle stages of refinement, being scale-independent, and is often used in rotation and translation searches (Harada *et al.*, 1981). The trial calculations described below use a weighted mixed difference map of $V(\mathbf{r})$ and $C(\mathbf{r})$ in combination as a safeguard to eliminate peaks that are clearly inconsistent with the observed target intensities $|T(\mathbf{H})|^2$.

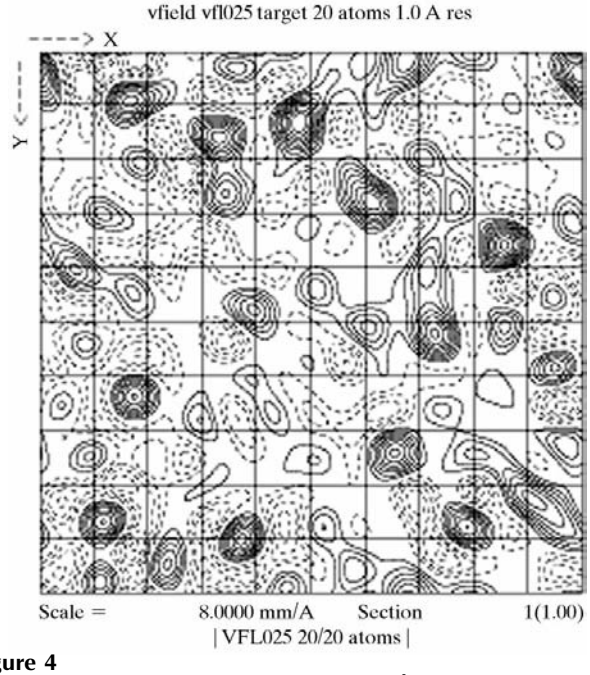

Figure 4
Local field echo of the target structure at 1.0 Å.

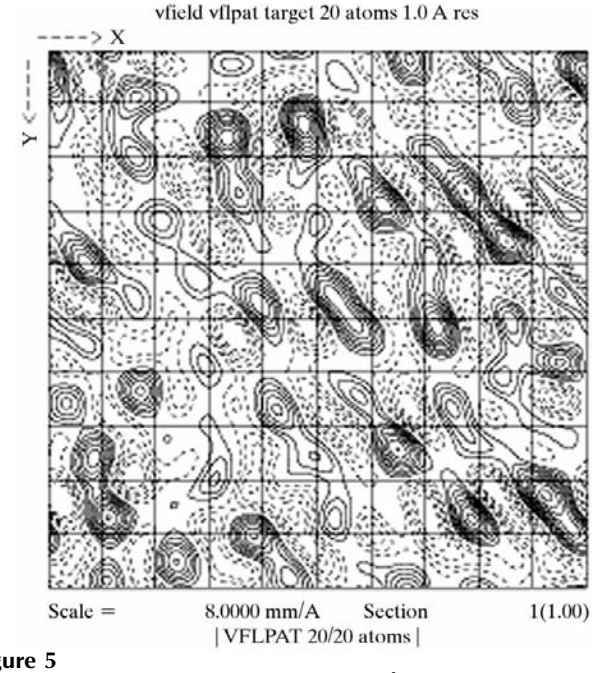

Figure 5
Patterson echo of the target structure at 1.0 Å.

7. A grid-based filter and pair algorithm

7.1. A grid point-atom model

A typical computer experiment starts by generating a random structure of N equal atoms, selecting N random points on a fine rectangular grid that contains a large number of points, L. This point density of N peaks is the target for solution. The normalized structure factors of the target are written

$$T(\mathbf{h}) = E_{\rm targ}(\mathbf{h}) \tag{10}$$

and the object of the experiment is to solve the target structure from the given values of $|T(\mathbf{H})|^2$ for the observed reflections $\mathbf{H}$. The grid model gives rise to approximately $\frac{1}{2}L$ independent pairs of reflections $(\mathbf{h}, -\mathbf{h})$ from the half-sphere of reciprocal space ($h \geq 0$). All of the data were used in our grid calculations, so that the set of measured reflections $\mathbf{H}$ is the entire half-sphere out to the Brillouin zone boundary.

7.2. A three-step algorithm

The solution method outlined below is designed to make maximum use of the pair potential in a straightforward way. No statistical phasing relationships are used, in order to make clear that the pair potential is able to pick out the correct atoms without such help.

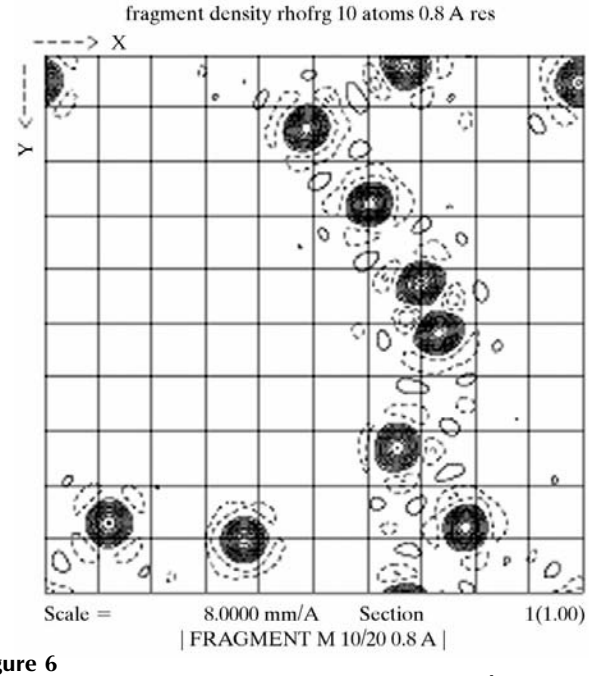

Figure 6
Fragment density for ten correctly placed atoms at 0.8 Å.

The algorithm is a three-step iterative cycle carried out on an electron-density map. The main steps are Fourier power filtering, peak picking and paired peak selection.

(i) The Fourier filter is carried out on the current model density $\rho_1(\mathbf{r})$.

(*a*) Calculate the model Fourier $E_N(\mathbf{h})$ for N atoms with the resulting phases $\varphi(\mathbf{h})$.

(*b*) Apply these phases to the target amplitudes $|T(\mathbf{H})|$.

(*c*) Filter out and remove the weak reflections of the target, where $|T(\mathbf{H})|^2 < T^2_{\rm filt}$, with a typical threshold $T^2_{\rm filt} = 1.75$. This normally selects the top 17% of strongest reflections and the selection list need be made only once.

(*d*) Synthesize the filtered density $\rho_2(\mathbf{r})$, which has both positive and negative peaks.

(ii) Pick peaks from $\rho_2(\mathbf{r})$ to generate a new trial set of peaks with many extra atoms.

(*a*) The number of peaks generated will be $N_A = A_{\rm peak}N$, where $A_{\rm peak}$ is an amplification factor. Typically, $A_{\rm peak} = 2.75$.

(*b*) Select the N_A highest peaks from the filtered density $\rho_2(\mathbf{r})$.

(*c*) Generate a new positive density $\rho_3(\mathbf{r})$ with equal peaks at these marked N_A points.

(iii) Peak pairing. Choose the best N out of the N_A peaks to maximize a weighted average of the pair potential and the intensity correlation coefficient.

(*a*) Calculate the pair potential of the density $\rho_3(\mathbf{r})$, using all the N_A peaks, and their potential gradient in real space as a density map $V(\mathbf{r})$, using all the known reflections, both strong and weak, even those rejected in step (i)(*c*).

(*b*) Calculate the intensity correlation coefficient, using all the reflections, and its gradient $C(\mathbf{r})$, as described in Appendix *A*. Use $C(\mathbf{r})$ as a difference map, with the weight factor described below.

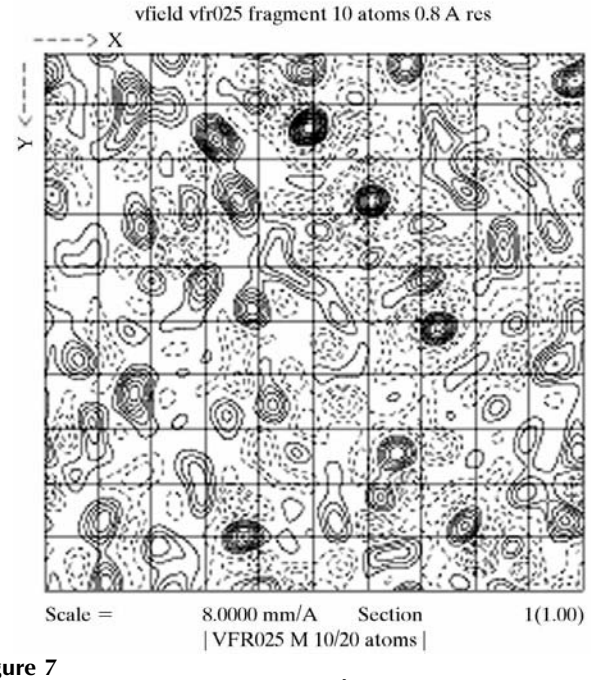

Figure 7
Local field echo of the fragment at 0.8 Å.

(*c*) Apply scaling factors to each of the maps $V(\mathbf{r})$ and $C(\mathbf{r})$ so that the mean absolute value of each map over all space is normalized to 1. The scaled maps are now $V_s(\mathbf{r})$ and $C_s(\mathbf{r})$.

(*d*) Construct a weighted scaled map $V_{sw}(\mathbf{r}) = [V_s(\mathbf{r}) + w_c C_s(\mathbf{r})]$ with a weight w_c for the intensity correlation, typically $w_c = 0.2$. Evaluate $V_{sw}(\mathbf{r})$ at each of the N_A assigned trial peaks and select the N peaks with the highest values. Do not select any other new unmarked peaks.

(*e*) Generate a new density $\rho_4(\mathbf{r})$ with these N points as equal peaks. $\rho_4(\mathbf{r})$ will become the starting density $\rho_1(\mathbf{r})$ for the next cycle.

Each cycle requires at least five Fourier transform operations and two peak-picking operations.

The role of the power filter in step (i)(*c*) is to generate variety in the trial sets of peaks, in such a way as to match the strongest reflection amplitudes of the target. The intensity correlation coefficient gradient $C(\mathbf{r})$ in step (iii)(*b*) is used as a necessary safeguard to eliminate peaks that are clearly inconsistent with the observed target intensities. The replacement of the selected peaks by equal peaks in steps (ii)(*c*) and (iii)(*e*) implements the assumption that there are N equal atoms in the target.

An experimental trial starts from a random starting model of N atoms and is set to run for a certain number of cycles, say 500. Progress is monitored by recording the total pair potential $\Psi_{\rm trial}$ and the intensity correlation residual R_I. In the first few cycles $\Psi_{\rm trial}$ and C_I increase rapidly but they soon level off at a plateau, with small fluctuations about a slowly rising or constant background. A successful solution appears as a sudden unheralded rapid rise in $\Psi_{\rm trial}$ and C_I, leading to completion in a few tens of cycles. Otherwise the test may run out of cycles, or more rarely it may become stuck in a repeated limit cycle. Here the same incorrect density $\rho_1(\mathbf{r})$ is regenerated repeatedly on the grid in every cycle. This fault commonly occurs if there are not enough points on the grid.

To monitor the rate of change of the density map from one cycle to the next, we use the quantity S_{11}, defined as the fractional overlap of the current density ρ_1 with ρ_1 from the previous cycle, using a fixed origin. Thus, $S_{11} = 1.0$ indicates a limit cycle, while values of S_{11} in the range 0.80–0.95 indicate a typical reasonably stable refinement process. It may seem surprising that a great number of trial peaks is needed, with a large amplification factor $A_{\rm peak} = 2.75$, but in practice most of the same peaks are reselected in each iteration so that a large number are required to encourage change. The positions of the freshly created peaks are fixed at step (ii) of each cycle and do not move until the next cycle.

8. Grid model *ab initio* solutions

8.1. Pick-and-pair experiments

Table 1 shows the results of some typical successful solutions of random target structures on grids of various sizes. The part of the table headed 'Pick-and-pair algorithm' used the standard algorithm above, on three different grid sizes. Each experiment consisted of about ten repeated trials on the same

Table 1
Ab initio solutions on a rectangular grid.

Grid points	Atoms	Success/trials	Cycles used				
			Mean	Median	Max.	Min.	Limit
Pick-and pair algorithm							
(64,64,64)							
262144	400	10/11	596	545	894	416	1500
	500	0/10					1500
(64,128,128)							
1048576	400	10/10	171	195	271	46	500
	450	10/10	286	204	470	45	500
(128,128,128)							
2097152	500	10/10	206	190	331	83	500
Pair-and-square algorithm							
(128,128,128)							
2097152	500	10/10	85	66	179	20	500

target but with different starting models. The results provide a rough estimate of the typical success rate.

With the coarsest grid dimensioned (64, 64, 64), the solution is trivial for up to 300 atoms and is completed very quickly. For 400 atoms, a solution is almost always successful but takes longer. A measure of the average cost of a solution is the median number of cycles used. For example, the median number with a 400-atom target is 545 and a series of trials with a limit of 545 would achieve a 50% success rate. With increasingly large targets, the number of cycles needed for a solution increases very steeply and 500 atoms represents a practical upper limit on this size of grid.

Finer grids increase the chance of success and reduce the number of cycles required. The (128, 128, 128) grid only needs about 200 cycles to solve 500 atoms. But the computational cost of each iteration cycle becomes high, with lengthy Fourier transforms, so that experiments with more than 500 atoms are again expensive.

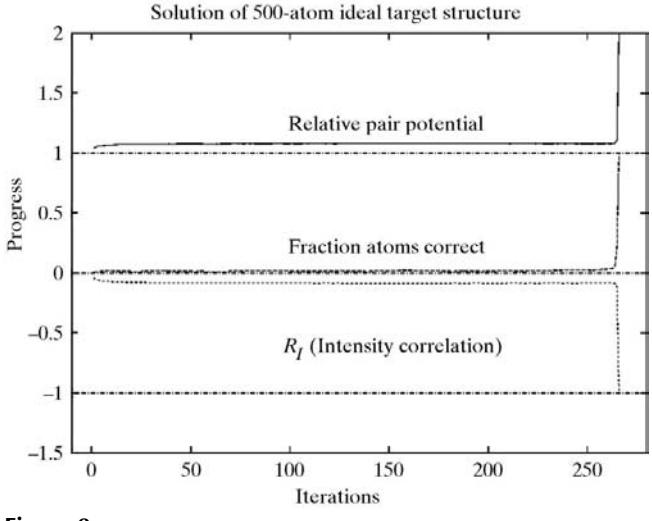

Figure 8
Progress curves for complete all-atom *ab initio* solution of a random target structure of 500 atoms. Upper curve: relative pair potential Ψ_{rel}. Middle curve: fraction of atoms correct. Lower curve: intensity correlation residual R_I. Note that although R_I may range between 0 and +2 the values observed here did not exceed +1.

Fig. 8 is a progress curve that illustrates the course of a successful *ab initio* solution on a random target structure of 500 atoms, which succeeded after 284 cycles. The relative pair potential in the upper curve is $\Psi_{rel} = \Psi_{trial}/\Psi_{targ}$, where Ψ_{targ} is the known total pair potential of the target structure. At first Ψ_{rel} increases rapidly, then it reaches a fluctuating equilibrium about a slowly rising baseline. The rapid increase that signals a solution occurs suddenly and without warning in the last 10–15 cycles, when Ψ_{rel} increases to unity. The fraction of atoms correct, as measured by the cross-correlation of the model density map with the known target, behaves similarly and shows little change until the slowly rising value passes 5% after about 250 cycles. Lastly, the intensity-correlation residual curve R_I maintains a high value throughout, until the last 5–10 cycles.

This behaviour is typical of a dynamical phase transition in a highly cooperative system. The original disordered fluid state of mispaired atoms organizes itself rapidly into a perfectly ordered solid structure once a sufficient number of correct subclusters of atoms have formed. The long-range pairing forces between atoms provide the driving force for constructing these clusters, and once the cluster size reaches a critical value condensation takes place about this nucleus. The solution, if reached, is completely exact.

8.2. The pair-and-square method

A variant of the standard algorithm is to replace the power filter step (1) above by a conventional tangent formula step that uses the same strong reflections. The tangent step works with the square of the filtered density $\rho_2(\mathbf{r})$ (Sayre, 1952). This 'pair-and-square' iteration solves 500-atom targets consistently faster on the finer of the two grids above in about 100 cycles, but it still cannot solve the difficult larger targets.

These *ab initio* tests demonstrate that the pair potential, coupled with an elementary peak selection and filtering procedure, can easily solve ideal structures of a few hundred atoms.

9. *Ab initio* subcluster searches

During these grid experiments, it became clear that a correct solution almost always emerges in a few cycles whenever the current model contains at least one good subcluster of correct atoms, amounting to 10–15% of the whole target. For example, a target of 500 atoms is solved rapidly once the trial model is good enough to match 50–75 atoms. The pair forces generated by this small group of correct atoms therefore contain sufficient information to guide the other 425–450 atoms into position.

Table 2
Ab initio subcluster solutions on a grid.

Grid points	Cluster/atoms	Success/trials	Cycles used				
			Mean	Median	Max.	Min.	Limit
(64, 64, 64)							
262144	75/300	10/19 9 repeats	40	41	53	9	100
(128, 128, 128)							
2097152	200/500	10/10	103	95	253	11	500
	150/600	5/9 3 repeats	206	195	287	121	500

This cluster-building property of the pair potential means that a target structure can be solved *ab initio* in two stages: first find a sufficiently large correct subcluster, leaving out the rest of the atoms, and then use this to nucleate the remainder in a later full-atom refinement. Table 2 illustrates *ab initio* solutions using a search for subclusters. For instance, a partial search for a quarter of the atoms in a target of 300 on the (64, 64, 64) grid usually found a perfect 75/300 subcluster in only 40 cycles. However, half of these trials ended unproductively in a limit cycle. With the finer (128, 128, 128) grid, the searches for subclusters of 200/500 and 150/600 atoms were efficient and successful. These subcluster solutions using a reduced set of atoms can always be completed by adding in further randomly placed atoms for a second whole-molecule refinement.

10. Seeding with a known fragment

Two further sets of experiments investigated the use of 'seeds' or small known correct structural fragments of the target to produce a partially assisted solution. As a first illustration of seeding into a complete structure on a (128, 128, 128) grid, we took a 5% seed of 500 guide atoms for a target of 10000 atoms. This was solved almost instantly (see Table 3). A more ambitious experiment successfully solved 30000 atoms from a seed of 1500 guide atoms using 400 cycles. A seeded full structure seems to require as few as 3–5% of guide atoms in order to ensure success. Under these conditions, the search almost always converges uniformly and without difficulty to a correct solution.

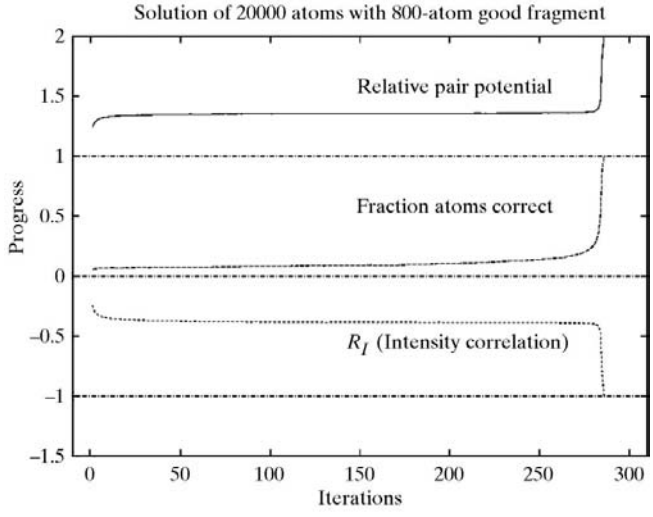

Figure 9
Progress curves for a seeded solution of a random target structure of 20000 atoms, building on a known fixed fragment of 800 atoms. Upper curve: relative pair potential $\Psi_{\rm rel}$. Middle curve: fraction of atoms correct. Lower curve: intensity correlation coefficient residual R_I.

The progress curve of a typical seeded solution, in Fig. 9, has some characteristic differences from the unguided *ab initio* trials. The middle portion of the refinement shows a slow but steady rise in the pair potential as the correct subcluster of atoms nucleated by the original seed fragment gradually increases in size over a long period. The onset of the final phase transition is less sudden and takes about 50 cycles to complete.

Seeding can also be used to assist a solution with reduced subclusters. Table 3 shows the results of two series of trials, one with a target of 1800 atoms on the (64, 64, 64) grid and the other with very large targets on the (128, 128, 128) grid. The conclusion is clear. Under these ideal conditions, it is possible to obtain fast and complete solutions of very large random structures (up to 30000 atoms) using subclusters of fewer than 10% of the total number of atoms and aided by seeds of as few as 2% of the total.

These experiments, taken together, show that the pair potential, when used efficiently with seeds, is capable of solving very large grid models. The current practical limit of 500–600 atoms for *ab initio* grid solutions could probably be raised significantly by developing more refined methods which are better at finding starter subclusters of correct atoms.

11. The effective resolution of a grid model

The data generated by the point atoms in the ideal *P*1 grid model correspond to a very high effective resolution. Thus if the numbers of grid divisions along the three cell axes (x, y, z) are (L_1, L_2, L_3), then the limits on the magnitudes of the reciprocal Bragg spacing $1/d_{hkl}$ in the first Brillouin zone are $(L_1/2a, L_2/2b, L_3/2c)$ and the number of measured reflections per atom is $(L_1L_2L_3/2N)$. Hence, in an ideal trial on 500 atoms placed on a (128, 128, 128) grid, there are about 2000 reflections per atom. This should be compared with a typical well packed protein unit cell with a little solvent (Matthews, 1968) and 500 atoms, each occupying a mean volume $v_A = (V_{\rm cell}/N) = 25\text{Å}^3$, in a real cell of volume 12500 Å^3. If the cell is taken to be a cube of side 23 Å, the grid spacing will be 0.18 Å, and becomes suitable for modelling a structure to about 0.60 Å resolution. The number of independent reflections out to a Bragg spacing of $d_{\rm res}$ is proportional to the volume of the hemisphere in reciprocal space, with

$$n_R = \frac{2\pi}{3}\frac{V_{\rm cell}}{d_{\rm res}^3} \qquad (11)$$

Table 3
Solutions seeded with a known fragment.

Grid points	Seed/cluster	Atoms	Cycles used
Seeded all-atom solutions			
(128,128,128)	500/10000	10000	25
2097152	300/10000	10000	232
	1500/30000	30000	400
Seeded subcluster solutions			
(64,64,64)	75/450	1800	9
262144	50/450	1800	58
	35/450	1800	495
	35/300	1800	87
	30/300	1800	480
	30/180	1800	232
	30/150	1800	387
	25/150	1800	410
	30/120	1800	34
(128,128,128)	1000/5000	30000	60
2097152	750/5000	30000	368
	750/3750	30000	218
	750/3000	30000	217
	750/2000	30000	216
	675/2000	30000	290
	600/2000	30000	913

and the number of reflections per atom will be

$$\frac{n_R}{N} = \frac{2\pi}{3}\frac{v_A}{d_{\rm res}^3}. \qquad (12)$$

Thus, at 0.60 Å resolution, $n_R/N = 242$ and is very much less than the figure of 2000 quoted above. The grid model described here is therefore a test of structure solutions with extremely highly redundant data.

12. Solution of real structures

12.1. Test molecules

In the first part of this paper, we tested the power of the pair-functional method to solve structures by running computational experiments on random sets of point atoms scattered over a grid. The grid algorithm used a simple three-stage cycle: (i) Fourier filtering of strong reflections; (ii) picking a large excess of the highest peaks from a density map; (iii) pruning the peaks to keep those with the highest pair potential. The algorithm has been developed further to deal with real molecules.

We now describe tests on moderately large small molecules, already solved at atomic resolution. Compared with a grid target model, the new complications are: space-group symmetry, limited resolution, data errors and the non-random pattern of the actual atomic positions. There is also the need to impose the condition that each atom has a certain hard-sphere radius.

The chosen molecules were six samples from a set used by Sheldrick (1990) to test phase-annealing methods in the *SHELX90* program. Each molecule has 50–200 atoms in a unit cell. The data are of high quality and the small errors are not a significant obstacle to solution.

12.2. Effective *P*1 symmetry

In practice, there are two ways to deal with space-group symmetry. One is to symmetrize the trial atomic structure in every iteration of the search procedure. The electron density may then be stored as a single asymmetric unit and the pair potentials should be calculated with the correct statistical weights for each class of reflection, using correctly scaled $E(\mathbf{H})$ values (Stewart & Karle, 1976; Giacovazzo, 1980). The other way is to treat the symmetry as effectively $P1$ and let the calculation 'discover' the correct space group as it proceeds. We have used the second way because it often works better. A possible reason for this effect is that the tentative symmetry elements float freely in the cell until they finally become established by the features of the emerging atomic model. In this treatment, it is, however, still essential to include the centring number correction for the statistical weights of the data from centred cells, even if the minor classes of special reflections are not treated with the correct weights.

12.3. Limited resolution

The simplest form of pair-functional theory treats all light atoms (C, N, O) as point objects of equal weight. So the test calculations worked from a representation in terms of point-atom objects and with maps generated from undamped E structure factors. Atoms were given a hard-sphere diameter of about 0.80 Å and a repulsive collision potential then prevented overlapping peaks from being selected. The peaks were chosen from high-density regions of the map without testing for valid bond distances or angles. A fine grid was used to allow a spread-out discrete representation of each point atom as a broadened Gaussian and structure factors were calculated by Agarwal's method (Agarwal, 1978; Isaacs & Agarwal, 1978) using a fast Fourier transform. The high-resolution terms of the transform are then amplified to correct for the Gaussian atomic density. A suitable grid mesh for a molecule at 0.80 Å resolution is 0.16–0.20 Å, coupled with a typical Gaussian radius of 0.26 Å. This Gaussian representation on the grid is fine enough to calculate correct pair potentials and intensity correlation coefficients to an accuracy of one part in 10^4 (See Appendix *B* for a discussion).

12.4. Strong reflections

The Fourier filter in our search algorithm generates each map of trial peaks from the strong reflections only. For the filter to work correctly, there must be enough strong reflections to generate all the peaks at precisely the correct positions. This requires 10–15 strong reflections per atom and, as only 15–20% of the reflections are classed as strong, the method normally needs about 100 measured reflections per atom. These different conflicting criteria for the grid and the reflections have to be reconciled if an efficient structure

Table 4
Small-molecule pick-and-pair solutions.

	Test molecule					
	LOG	SUOA	NEWQB	BHAT	HOPS	MBH2
Space group	$P2_12_12_1$	$P2_12_12_1$	$P\bar{1}$	Pc	$R3$ (hex.)	$P1$
Atoms/cell	108	188	124	84	234	54
Cell edges (Å)	8.187	18.350	10.376	7.888	23.978	8.701
	14.277	21.441	13.276	6.779	23.978	10.924
	15.693	8.350	14.655	21.595	8.918	12.574
Cell angles (°)	90.0	90.0	98.71	90.0	90.0	70.30
	90.0	90.0	95.23	108.21	90.0	80.56
	90.0	90.0	93.20	90.0	120.0	82.43
Reflections						
Expanded to $P1$	9093	11447	3948	3673	5377	3764
Resolution (Å)	0.75	0.83	1.04	0.83	0.84	0.85
Grid size	64,128,128	108,128,48	72,96,108	64,64,192	144,144,54	48,64,64
Points	1048576	663552	746496	786432	1119744	196608
Target potential	5961	6983	4593	2308	9897	2559
Final potential	6267	7558	4920	2340	10577	2716
Intensity residual	0.127	0.141	0.022	0.222	0.190	0.034
Density overlap	0.972	0.926	0.950	0.909	0.917	0.918
E^2 cutoff	1.75	1.50	1.75	1.75	1.75	1.75
Strong reflections	1622	2256	769	601	863	613
Intensity mix	0.15	0.15	0.15	0.15	0.15	0.15
Cycles used	30	500	84	30	95	10

solution is to be possible. For this reason, the cell and grid dimensions for the test molecules are listed in Table 4.

The basic algorithm, with Agarwal sharpening, as described above, has been extended by adding various additional options that sometimes improve the performance. One is to relax the strict requirement of equal atomic peaks in the intermediate maps. Another is to use a two-stage pruning procedure in each cycle to select the best-paired atoms more consistently.

12.5. Results

Each molecule was given a few trial runs to explore the best conditions. Once these were found, a solution was found in almost every test (see Table 5). The most difficult molecule to solve was SUOA, which needed about 500 iterations and 5–10 trial runs for each success. This molecule has a rather low ratio of strong reflections to atoms.

The quality of the solution has been measured against several standards. The first measure, which assesses the extent to which the atoms are paired correctly in the pair-functional force field, is to compare the pair potential of the known target structure (deduced from the atomic coordinates of the correct solution) with the final potential of the model. The numbers compare well for each of the six molecules. They are calculated from all the measured reflections. The second measure is R_I, the intensity correlation residual. Values of R_I range from 0.222 for the very high resolution structure of BHAT to an extremely low value of 0.022 for NEWQB, where the resolution is 1.04 Å.

The third quality measure, which is only possible with a previously known atomic solution, is the optimum overlap, Q_{MT}, of the final model point density with the point density of the correct solution (Appendix A). This is calculated with a small B factor of 7.5 to accommodate small positioning errors in the densities. All the density overlaps are better than 0.90. The quality values could probably be improved by further refinement, using standard methods.

13. Conclusions

These preliminary tests of the pair-functional method give some insight into the use of the pair potential to solve structures at atomic resolution. The trials with a simple three-stage iterative algorithm show that the pairing force plays a primary part in guiding the atoms into position and is essential for success. The intensity correlation coefficient acts as a secondary quality control. In the early stages of the calculation, it helps to eliminate models that have seriously wrong X-ray intensities. At the end, it helps to place the atoms in exact final positions. The Fourier filter step operates by generating fresh sets of trial peaks that follow the phases of the most recent atomic model. In the pair-and-square algorithm, the tangent-formula step acts as an efficient accelerator in the final stages of a solution but seems to have little overall effect on the success rate in long runs. As the pair-functional method works with atomic positions rather than phase variables, it is not surprising that the tangent-formula step makes only a marginal improvement. When working with real molecules and data at limited resolution, the correct choice of the grid dimensions, following the principles in Appendix B, is a critical factor.

The total pairing potential $\Psi_{\rm trial}$ is an estimate of the logarithm of the likelihood for any atomic conformation and it uses all the $N(N-1)/2$ separation vectors in a highly cooperative way. But the simple originless Patterson function is not at all useful as a potential function. The pair-functional ensemble, and the pair potential deri2ved from it, embody much of the same cooperative many-particle statistical information as the triplet and quartet phase probability distributions that are used in classical direct methods (Hauptman & Karle, 1953; Hauptman, 1975, 1991).

The experimental tests in this paper suggest that a successful solution normally proceeds by constructing many trial clusters of atoms that grow and redissolve randomly until a sufficiently large correct nucleus is formed. The pairing forces guide additional atoms into position around existing nuclei and have a larger range of convergence than the intensity-correlation difference maps. These latter maps behave rather like difference Patterson densities.

The point-atom grid trials are unrealistic because of their over-determined data and their very high effective resolution. They do, however, indicate some useful properties of the

Table 5
Small-molecule pair-and-square solutions.

	Test molecule					
	LOG	SUOA	NEWQB	BHAT	HOPS	MBH2
Grid size	64,128,128	144,162,64	72,96,108	64,64,192	144,144,54	48,64,64
Points	1048576	1492992	746496	786432	1119744	196608
Target potential	5961	7129	4593	2308	9897	2559
Final potential	6267	7123	4922	2354	10556	2725
Intensity residual	0.128	0.190	0.024	0.225	0.196	0.039
Density overlap	0.970	0.901	0.963	0.908	0.916	0.929
E^2 cutoff	1.50	1.25	1.75	1.25	1.75	1.75
Strong reflections	2013	3198	769	968	863	613
Intensity mix	0.15	0.15	0.20	0.20	0.15	0.15
Cycles used	40	383	84	25	30	20

pairing force. In particular, the successes of *ab initio* solutions which use only a small subset of atoms and of the seeded solutions which start from a very small known fragment are both encouraging. It may be possible to solve structures from high-resolution data by combining the pair potential with only a little extra phase information.

The pair-functional method has a long way to go before it can match the best existing classical *ab initio* methods (Sheldrick, 1990; DeTitta *et al.*, 1994; Weeks *et al.*, 1994). An improved algorithm is being developed that is an optimization program rather than a simple iterative procedure. This is developed from the temperature-dependent self-consistent-field method in statistical mechanics and seeks the minimum of an effective free-energy function built from the pair potential and the single-particle entropy.

APPENDIX *A*
Measures of quality

A1. Normalized intensity correlation coefficient

The intensity correlation coefficient between the model and the target Fouriers is calculated for the measured reflections. It is convenient to introduce weights, $w(\mathbf{h})$, for the reflections, scaled so that $\sum_{\mathbf{h}} w(\mathbf{h}) = 1$. This makes it possible to allow for symmetry weights and temperature factors. Let the normalized intensities of the model and the target be $|E(\mathbf{h})|^2$ and $|T(\mathbf{h})|^2$, with respective weighted mean values E_m^2 and T_m^2. Define the average cross correlation of E^2 with T^2 as

$$k_{ET} = \sum_{\mathbf{h}} w(\mathbf{h})\{|E(\mathbf{h})|^2 - E_m^2\}\{|T(\mathbf{h})|^2 - T_m^2\}, \qquad (13)$$

with analogous expressions k_{EE} and k_{TT} for the correlations of E^2 with E^2 and T^2 with T^2. Then the cross-correlation coefficient (Fujinaga & Read, 1987) is

$$C_I = k_{ET}/(k_{EE}k_{TT})^{1/2} \qquad (14)$$

and has a value between -1 and $+1$. The cross-correlation residual is $R_I = 1 - C_I$. The gradient of C_I for variations of the model structure factors is

$$\partial C_I/\partial E(\mathbf{h}) = 2\gamma_1 w(\mathbf{h})E(-\mathbf{h})(\{|T(\mathbf{h})|^2 - T_m^2\} - \gamma_2\{|E(\mathbf{h})|^2 - E_m^2\}), \qquad (15)$$

where γ_1 and γ_2 are scale factors.

$$\gamma_1 = (k_{EE}k_{TT})^{-1/2}, \quad \gamma_2 = k_{ET}/k_{EE}. \qquad (16)$$

Here γ_2 allows for differently scaled average intensity fluctuations in the model and the target, measured from their respective mean values. This scale independence means that one can compare Fourier intensity data sets with very different levels of modulation but similar relative intensity patterns. Each Fourier component of $\partial C_I/\partial E(\mathbf{h})$ is seen to be the product of $E(-\mathbf{h})$ with a term that resembles a scaled originless difference Patterson Fourier having coefficients

$$[|T(\mathbf{h})|^2 - T_m^2] - \gamma_2[|E(\mathbf{h})|^2 - E_m^2].$$

When this is transformed back into real space, the resulting gradient map $C(\mathbf{r})$ comes out as the convolution of the model density with this effective Patterson function. It provides a difference map that can be used to refine the current model.

A2. Density overlap between model and target maps

Suppose that the single-particle probability density of the target structure, with N_T atoms, is known to be $p_T(\mathbf{r})$, while the model, with N_M atoms, has a density $p_M(\mathbf{r})$. This model density may match the target well if it is superimposed on it with a suitable origin shift $\mathbf{u}$ and the quality of match for any given choice of $\mathbf{u}$ is measured by the overlap quality function or convolution

$$Q_{MT}^{(+)}(\mathbf{u}) = \int p_M(\mathbf{r})p_T(\mathbf{r} - \mathbf{u})\,\mathrm{d}\mathbf{r} = A_{MT}\sum_{\mathbf{h}} E_M(\mathbf{h})E_T(-\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{u}), \qquad (17)$$

where $A_{MT} = (N_M N_T)^{1/2}$. The model may also have the opposite hand to the target and this alternative fit can be tested by the inverted overlap function

$$Q_{MT}^{(-)}(\mathbf{u}) = \int p_M(\mathbf{r})p_T(\mathbf{u} - \mathbf{r})\,\mathrm{d}\mathbf{r} = A_{MT}\sum_{\mathbf{h}} E_M(\mathbf{h})E_T(\mathbf{h})\exp(-2\pi i\mathbf{h}\cdot\mathbf{u}). \qquad (18)$$

The functions $Q_{MT}^{(+)}(\mathbf{u})$ and $Q_{MT}^{(-)}(\mathbf{u})$ are calculated as maps in real space. The highest peaks with heights $Q_{MT}^{(+)}$(max.) at $\mathbf{u} = \mathbf{u}_m^{(+)}$ and $Q_{MT}^{(-)}$(max.) at $\mathbf{u} = \mathbf{u}_m^{(-)}$ identify the best matches for the two hands, and their corresponding origin shifts.

For model structures made of point atoms, it is more useful to calculate a spatially smoothed overlap map using an appropriate quality temperature factor B_Q or vibrational amplitude v_Q. The Fourier coefficients of $Q_{MT}^{(+)}$ then become

$$Q_{MT}^{(+)}(\mathbf{h}) = A_{MT}E_M(\mathbf{h})E_T(-\mathbf{h})\exp(-\tfrac{1}{4}B_Q|\mathbf{r}^*|^2), \quad B_Q = 8\pi^2 v_Q^2, \qquad (19)$$

where $\mathbf{r}^*$ is the reciprocal-lattice vector. The overlap quality functions $Q_{MT}^{(+)}(\mathbf{u})$ and $Q_{MT}^{(-)}(\mathbf{u})$ are measured in terms of the

number of atoms. They may also be converted into scaled overlap correlation coefficients by dividing by the self-overlaps of the model and the target maps

$$C^{(+)}_{MT}(\mathbf{u}) = Q^{(+)}_{MT}(\mathbf{u})/[Q^{(+)}_{MM}(0)Q^{(+)}_{TT}(0)]^{1/2} \quad (20)$$

$$C^{(-)}_{MT}(\mathbf{u}) = Q^{(-)}_{MT}(\mathbf{u})/[Q^{(-)}_{MM}(0)Q^{(-)}_{TT}(0)]^{1/2}. \quad (21)$$

In this scaled form, the coefficients measure the fraction of correctly placed atoms in the model.

APPENDIX *B*
Grid representation of point atoms

In Agarwal's method, point atoms are described *via* an intermediate representation, as broadened Gaussian peaks which are mapped onto the vertices of a fine grid. For simplicity, we consider here a one-dimensional Gaussian, with vibrational amplitude u, where the cell edge is divided into L_1 parts along x, with a spacing $s = a/L_1$. The Gaussian, in Cartesian coordinates, is

$$g(X) = [1/(2\pi u^2)^{1/2}]\exp(-X^2/2u^2). \quad (22)$$

In free space, this would have the Fourier transform

$$G_{\rm free}(h) = \exp[-2\pi^2 w^2(h^2/L_1^2)], \quad (23)$$

where the Fourier wavelength is a/h and w is the width ratio

$$w = u/s. \quad (24)$$

Fourier transforms on the discrete grid are periodic functions in h, with a period of L_1, and the first Brillouin zone bounded by $h = \pm\frac{1}{2}L_1$ marks the limits of the unique region (Ziman, 1969). The transform of a Gaussian, referred to the first zone, now includes satellites, or aliases (Bricogne, 1993), from the tails of transforms in adjacent zones, so that

$$G_{\rm grid}(h) = \sum_{\nu} G_{\rm free}(h - \nu L_1), \quad \nu = 0, \pm1, \pm2, \ldots, \quad (25)$$

where ν is the number of the displaced zone. Suppose that the measured reflections extend to a limited resolution, with Bragg spacing $d_{\rm res} = R$, and a corresponding limit $h_{\rm res}$ on h, where $h_{\rm res} = L_1(s/R)$. The calculated structure factors of these outer reflections are attenuated by an exponential factor

$$M_{\rm att} = G_{\rm free}(h_{\rm res}) = \exp[-2\pi^2(u^2/R^2)], \quad (26)$$

which must be compensated afterwards by dividing back by $M_{\rm att}$. The structure factors are also subject to aliasing corrections in both amplitudes and phases, which give relative errors of order

$$M_{\rm alias} = \sum_{\nu} G_{\rm free}(h - \nu L_1)/G_{\rm free}(h). \quad (27)$$

Normally the only important correction is from the nearest zone ($\nu = 1$), so that

$$M_{\rm alias} = G_{\rm free}(L_1 - h_{\rm res})/G_{\rm free}(h_{\rm res}) = \exp[-2\pi^2 w^2(1 - 2s/R)]. \quad (28)$$

The conditions necessary for an accurate calculation are:

(i) The attenuation factor $M_{\rm att}$ must not be too small, to avoid amplifying errors in the calculated structure factors. Typically, $(u/R) = 1/2$ leads to $M_{\rm att} = 0.007$, and this represents a useful upper limit for (u/R).

(ii) The aliasing error $M_{\rm alias}$ must be small. This requires a moderately large w and a small (s/R). For example, with $u/R = 1/2$ and $s/R = 1/3$, we obtain $w = 3/2$ and $M_{\rm alias} = 4 \times 10^{-7}$.

(iii) The Gaussian in real space decays rapidly so that its value at a position n grid spacings from the centre is proportional to $M_{\rm Gauss} = \exp(-\frac{1}{2}n^2/w^2)$. The bounding limit of n should be chosen so that $M_{\rm Gauss}$ is negligible for all the neglected grid points outside the boundary. This condition requires a moderately large w. Thus, when $w = 4/3$ and $M_{\rm Gauss} \leq 10^{-4}$, we need a range of $n = 6$.

These three requirements, taken together, place narrow limits on the acceptable ranges of w and s/R. The size of the Fourier transform grid is $L_1L_2L_3$ and thus proportional to $1/s^3$, while the cost of evaluating three-dimensional Gaussians on a grid in real space is proportional to n^3 and thus depends on w^3.

We also need to calculate other local point-atom properties, such as the pair force field $V(\mathbf{x}_A)$, which acts on an atom at a general continuous cell position $\mathbf{x}_A$. If the atom is represented by the broadened grid Gaussian

$$g(\mathbf{r}_j - \mathbf{x}_A), \quad (29)$$

which covers a cluster of grid points j in the neighbourhood of $\mathbf{x_A}$, and if $V(\mathbf{x})$ has Fourier components $V(\mathbf{h})$ then we construct a sharpened field $V_{\rm sharp}(\mathbf{r}_j)$ on the grid with Fourier components $V(\mathbf{h})/G_{\rm free}(\mathbf{h})$. We evaluate $V(\mathbf{x}_A)$ from the grid sum

$$V(\mathbf{x}_A) = \sum_j g(\mathbf{r}_j - \mathbf{x}_A)V_{\rm sharp}(\mathbf{r}_j). \quad (30)$$

This operation is also accurate provided that $M_{\rm att}$ is not too small.

I thank George Sheldrick for providing the data for his small test molecules and for practical advice on the trial solutions, David Eisenberg and Dan Anderson for discussions, and Tony Crowther for helpful comments on the draft manuscript.

References

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
Blessing, R. H., Guo, D. Y. & Langs, D. A. (1996). *Acta Cryst.* D**52**, 257–266.
Blessing, R. H., Guo, D. Y. & Langs, D. A (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 47–71. Dordrecht: Kluwer.
Bricogne, G. (1993). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 23–91. Dordrecht: Kluwer.
Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: John Wiley.
De Titta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.
Fujinaga, M. & Read, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
Giacovazzo, C. (1980). *Direct Methods in Crystallography*. London: Academic Press.
Hansen, J. P. & McDonald, I. R. (1986). *Theory of Simple Liquids*, 2nd ed. London: Academic Press.

Harada, Y., Lifchitz, A., Berthou, J. & Jolles, P. (1981). *Acta Cryst.* A**37**, 398–406.
Hauptman, H. (1975). *Acta Cryst.* A**31**, 671–679.
Hauptman, H. (1991). *Crystallographic Computing 5. From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 324–332. Oxford University Press/IUCr.
Hauptman, H. & Karle, J. (1953). Editors. *The Solution of the Phase Problem: I. The Centrosymmetric Crystal. American Crystallographic Association Monograph* No 3. Pittsburgh, PA: Polycrystal Book Service.
Isaacs, N. W. & Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 782–791.
McLachlan, A. D. (1999). *Acta Cryst.* A**55** Supplement, Abstract P12.BB.007.
McLachlan, A. D. (2001*a*). *Acta Cryst.* A**57**, 125–139.
McLachlan, A. D. (2001*b*). *Acta Cryst.* A**57**, 152–162.
McLachlan, A. D. & Harris, R. A. (1961). *J. Chem. Phys.* **34**, 1451–1452.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Ornstein, L. S. & Zernicke, F. (1914). *Proc. Akad. Sci. (Amsterdam)*, **17**, 793–806.
Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.
Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. (1965). *Acta Cryst.* **18**, 169–179.
Stewart, J. M. & Karle, J. (1976). *Acta Cryst.* A**32**, 1005–1007.
Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A**50**, 210–220.
Ziman, J. M. (1969). *Principles of the Theory of Solids.* Cambridge University Press.