Acta Cryst. (1998). D**54**, 81–85

A Comparison of Two Algorithms for Electron-Density Map Improvement by Introduction of Atomicity: Skeletonization, and Map Sorting Followed by Refinement

F. M. D. VELLIEUX

Institut de Biologie Structurale J.-P. Ebel CEA CNRS, 41 Avenue des Martyrs, 38027 Grenoble CEDEX 01, France. E-mail: vellieux@ibs.ibs.fr

(Received 11 April 1997; accepted 28 May 1997)

Abstract

A comparison has been made of two methods for electron-density map improvement by the introduction of atomicity, namely the iterative skeletonization procedure of the *CCP*4 program *DM* [Cowtan & Main (1993). *Acta Cryst.* D**49**, 148–157] and the pseudo-atom introduction followed by the refinement protocol in the program suite *DEMON/ANGEL* [Vellieux, Hunt, Roy & Read (1995). *J. Appl. Cryst.* **28**, 347–351]. Tests carried out using the 3.0 Å resolution electron density resulting from iterative 12-fold non-crystallographic symmetry averaging and solvent flattening for the *Pseudomonas aeruginosa* ornithine transcarbamoylase [Villeret, Tricot, Stalon & Dideberg (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10762–10766] indicate that pseudo-atom introduction followed by refinement performs much better than iterative skeletonization: with the former method, a phase improvement of 15.3° is obtained with respect to the initial density modification phases. With iterative skeletonization a phase degradation of 0.4° is obtained. Consequently, the electron-density maps obtained using pseudo-atom phases or pseudo-atom phases combined with density-modification phases are much easier to interpret. These tests also show that for ornithine transcarbamoylase, where 12-fold non-crystallographic symmetry is present in the *P*1 crystals, G-function coupling leads to the simultaneous decrease of the conventional *R* factor and of the free *R* factor, a phenomenon which is not observed when non-crystallographic symmetry is absent from the crystal. The method is far less effective in such a case, and the results obtained suggest that the map sorting followed by refinement stage should be by-passed to obtain interpretable electron-density distributions.

1. Introduction

Once an electron-density distribution has been obtained as the result of heavy-atom or density modification, several algorithms have been proposed to further improve the quality of this electron-density map by methods which introduce pseudo-atoms in the electron density.

Such methods currently being used in macromolecular crystallography are the iterative skeletonization procedure of Wilson & Agard (1993, *PRISM*) which simplifies the electron density as ridge points and ridge lines connecting these points, the automatic refinement procedure of Lamzin & Wilson (1993, *ARP*), the weighted automatic refinement procedure of Perrakis, Sixma, Wilson & Lamzin (1997, *wARP*), and the algorithm present in the program suite *DEMON/ANGEL* (Vellieux *et al.*, 1995), which consists of the introduction of pseudo-atoms in an electron-density map, followed by the refinement of these atomic coordinates and of the temperature factors of the pseudo-atoms. This idea is similar to that implemented in *ARP*, except that no additional pseudo-atom introduction, nor pseudo-atom deletion, are performed.

In this paper, a comparison is made between two of these algorithms, namely the iterative skeletonization procedure [as implemented in the *CCP*4 program *DM* (Cowtan & Main, 1993)], and the procedure found in the program suite *DEMON/ANGEL*. The tests have been carried out at a medium resolution of 3.0 Å in a case where non-crystallographic symmetry (NCS) is present in the crystal, where such algorithms are known not to give impressive results, and at 2.3 Å resolution when NCS is absent. The calculations were carried out using an initial phase and Sim weight set resulting from iterative 12-fold averaging and solvent flattening for the *Pseudomonas aeruginosa* catabolic ornithine transcarbamoylase (OTCase, Villeret *et al.*, 1995). For further analysis, the same tests were repeated for a structure devoid of non-crystallographic symmetry, chitinase from *Serratia marscescens* (Perrakis *et al.*, 1994) at 2.3 Å resolution, using initial phases at 2.5 Å resolution resulting from a procedure of iterative solvent flattening.

2. Materials and methods

The iterative skeletonization procedure as implemented in the *CCP*4 program *DM* (Cowtan & Main, 1993) relies on the skeletonization algorithm of Swanson (1994). This is quite similar to the iterative skeletonization procedure of the program *PRISM* (Wilson & Agard, 1993), which relies on Greer's algorithm (Greer, 1985).

Since the procedure found in the program suite *DEMON/ANGEL* was not described in our previous

paper (Vellieux *et al.*, 1995), it will be described in detail below.

First, the electron-density map is sorted by decreasing density values. The result of this procedure is a sorted array containing the decreasing electron density values, together with the corresponding grid locations in the three-dimensional electron-density map array. This sorted array is read sequentially, a pseudo-atom corresponding to a water molecule O atom is placed at the corresponding cartesian Å coordinates after estimating the maximum in the three-dimensional map array by fitting a parabola along each axis. Following this, all electron-density values within a radius of 1.52 Å from this newly introduced pseudo-atom are set to zero in the three-dimensional map array (this value of 1.52 Å corresponding to *ca.* the closest distance of approach of two C atoms bound by a single bond). The process of pseudo-atom introduction is then repeated, each time checking that the electron-density values in the three-dimensional map array are not less than or equal to 0.0 (otherwise, this pseudo-atom position is skipped), until all requested pseudo-atoms have been introduced in the electron-density map. The output is a Protein Data Bank coordinate file containing atomic coordinates of water O atoms.

The positions of these pseudo-atoms, together with their artificial temperature factor of 15.0 $Å^2$, are then refined by least-squares (Konnert & Hendrickson, 1980) or maximum-likelihood (Pannu & Read, 1996) refinement with a new crystallographic refinement program (Adams, Pannu, Read & Brünger, 1997). The resulting phases can be used to compute a Sim- or *SIGMAA*-weighted electron-density map (Sim, 1959, 1960; Read, 1986; Vellieux, 1997), or they can be combined with the phases used in the initial map calculation to generate a phase-combined electron-density distribution. It can be noted that a similar technique was described 20 years ago (Agarwal & Isaacs, 1977), but has not found widespread use in macromolecular crystallography.

These two algorithms were applied separately to the 3.0 Å resolution electron-density distribution resulting from the procedure of 12-fold averaging and solvent flattening used in the structure determination of *Pseudomonas aeruginosa* catabolic ornithine transcarbamoylase (Villeret *et al.*, 1995).

Iterative skeletonization was performed for a total of 20 cycles using all data to 3.0 Å resolution, with the free-*R*-factor approach (Brünger, 1992). 5% of the reflections were selected from the data for the free *R*-factor calculation, and these were changed at each skeletonization cycle.

With the program *DARTH* of the *DEMON/ANGEL* program suite, a total of 3273 atoms were introduced in the region of the averaged electron density corresponding to a single OTCase monomer. This corresponds approximately to the expected number of atomic positions, including water O atoms, for an OTCase monomer. The electron-density map used for pseudo-atom introduction had been calculated using a grid of about 1/10th of the high-resolution limit (3.0 Å). The resolution limits for refinement were 6.0 and 3.0 Å, the same as those used in the refinement of the atomic model of OTCase (Villeret *et al.*, 1995). During refinement, 12-fold non-crystallographic symmetry constraints corresponding to the dodecamer found in the asymmetric unit, were applied both for positional refinement and for the refinement of the individual atomic temperature factors. The crystallographic refinement program's protein_rep.param parameter file was modified to give the pseudo-atoms a van der Waals radius of 1.5 Å. A total of 200 cycles of maximum-likelihood energy minimization were performed, followed by 17 cycles of individual atomic temperature-factor refinement.

The same calculations were also performed on a structure devoid of non-crystallographic symmetry, a chitinase from *Serratia marscescens* (Perrakis *et al.*, 1994). For skeletonization, the same procedure (20 skeletonization cycles with 5% of the reflections removed for the free-*R*-factor calculation) as for OTCase was used, starting from 2.5 Å solvent-flattened phases. For pseudo-atom refinement, all data between 8.0 and 2.3 Å were used, with the electron-density distribution used to generate the pseudo-atoms being a 2.5 Å resolution Sim-weighted electron-density synthesis resulting from iterative solvent flattening. A total of 4500 atoms were introduced in the protein region of the initial map.

The resulting phase sets were compared to the set of phases for the refined structure of OTCase, which has an *R* factor of 21.6% for all data between 6.0 and 3.0 Å (Villeret *et al.*, 1995). For further analysis, several electron-density maps were calculated, and the real-space correlation coefficient between these observed electron densities and a 'control map', without any resolution cutoff and computed from the refined atomic coordinates, was calculated with the graphics display program *O* (Jones, Zou, Cowan & Kjeldgaard, 1991). The same calculations were performed for chitinase, comparing the phase set resulting from these density-modification procedures to the phases from the refined molecular model (Perrakis *et al.*, 1994).

3. Results

The results are summarized in Tables 1, 2 and 3 and can be described as follows.

20 cycles of iterative skeletonization were performed for OTCase with the program *DM* (Cowtan & Main, 1993). The resulting combined phase set was further away from the correct, refined phases, by 0.4° (Table 1), which seems insignificant. However, the corresponding electron-density map, calculated with weighted observed structure-factor amplitudes and combined phases (Table 2) is less interpretable than the initial averaged electron-

Table 1. *Phase differences (°) between the different phase sets obtained and the phases from the refined model of OTCase*

The phase set density modification is that obtained by the iterative procedure of 12-fold averaging and solvent flattening used in the structure determination of OTCase (Villeret *et al.*, 1995). The skeletonization phase set is that obtained by combining the skeletonization phases with the density modification phases, as performed by the program *DM* (Cowtan & Main, 1993). The refinement phase set is that obtained by refinement of the coordinates and individual temperature factors of the pseudo-atoms resulting from the run of the program *DARTH*. The combined phase set is a phase set resulting from phase combination between the initial density-modification phases and the refinement phases [with the Hendrickson–Lattman coefficients corresponding to the density-modification phases modified to ensure that the resulting combined phases differ equally from each individual phase set (Vellieux *et al.*, 1995)].

Density modification	48.4
Skeletonization	48.8
Refinement	33.1
Combined phases	35.4

Table 2. *Real-space correlation coefficients between the artificial electron-density distribution obtained from the refined atomic coordinates [calculated with the program O (Jones et al., 1991)], and several electron-density distributions*

AVG: initial density-modification map (Villeret *et al.*, 1995); DM: electron-density map computed from combined phases resulting from 20 cycles of skeletonization; SIGMAA: electron-density distribution obtained with modified *SIGMAA* coefficients (Read, 1986; Vellieux & Dijkstra, 1997) and phases from the refined pseudo-atom coordinates; SIGMAA-phi comb: same as SIGMAA, except that the phases used result from phase combination between the initial density-modification phases and the pseudo-atom phases.

ED map	⟨CCF⟩
AVG	0.59
DM	0.49
SIGMAA	0.59
SIGMAA-phi comb	0.59

density distribution, since the real-space correlation coefficient (Jones *et al.*, 1991) decreases by 0.1.

In contrast, refinement of the pseudo-atoms with the algorithm of the *DEMON/ANGEL* program suite (Vellieux *et al.*, 1995) using maximum likelihood (Pannu & Read, 1996) both for positional and individual atomic temperature-factor refinement proceeded well, since the free R factor (Brünger, 1992) decreased from 38.0 to 22.6%, and the R factor decreased from 38.4 to 21.2%, for all data between 6.0 and 3.0 Å. This decrease in the free R factor, computed with a subset of 5% of the available reflections, clearly indicates that the phases are improving at this medium resolution. Consequently, the resulting phase set shows an improvement of 15.3° with respect to the initial density-modification phases (Table 1). The resulting *SIGMAA* electron-density distribution (Read, 1986), computed with coefficients of the form $3m|F_o| - 2D|F_c|$ for acentric reflections and

Table 3. *Phase differences (°) between the different phase sets obtained and the phases from the refined model of chitinase*

The phase set solvent flattening is that obtained by the iterative procedure of solvent flattening used in the structure determination of this chitinase (Perrakis *et al.*, 1994). The skeletonization phase set is that obtained by combining the skeletonization phases with the solvent flattening phases. The phase set 'before refinement' corresponds to the phases obtained from the pseudo-atom coordinates before refinement. The refinement phase set is that obtained by refinement of the coordinates of the pseudo-atoms resulting from the run of the program *DARTH*.

Solvent flattening	51.9
Skeletonization	63.5
Before refinement	64.1
Refinement	58.0

$2m|F_o| - D|F_c|$ for centric reflections (Vellieux & Dijkstra, 1997) gives a real-space correlation coefficient equal to that of the initial density-modification map. Examination of this *SIGMAA* map clearly indicates that it is superior to the density-modification map (Fig. 1*c*).

Phase combination of these pseudo-atom phases with the initial density-modification phases was also performed, ensuring that the resulting combined phases differed equally from each individual phase set used for phase combination (this was done because the pseudo-atom phases are dominant in the phase combination process, Vellieux *et al.*, 1995). The resulting phase set shows an improvement of 13.0° with respect to the initial phases (Table 1). The resulting *SIGMAA* electron-density distribution also has the same real-space correlation coefficient as the initial density-modification map, but an examination indicates that it is superior (Fig. 1*d*).

In the case of chitinase, where non-crystallographic symmetry is absent from the crystals (Perrakis *et al.*, 1994), both the R factor and free R factor resulting from refinement of the pseudo-atoms are less convincing than in the case of OTCase, where non-crystallographic symmetry is present: the starting free R factor and conventional R factor were 41.2 and 41.4%, respectively. After pseudo-atom refinement, these figures were 46.1 and 27.2%. Such an increase in the free R factor suggests a phase degradation. This is confirmed by a comparison of the phase differences with the true phases (Table 3): the initial solvent-flattening phases are 51.9° away from the true phases. Pseudo-atom introduction followed by refinement leads to a rise in this figure to 58.0°, whereas iterative skeletonization degrades the phases further, since the phase difference to the true phases obtained after this procedure is 63.5°. It must be noted that in this case where no non-crystallographic redundancy is present, omitting the map sorting followed by refinement stage leads to better results. These observations clearly indicate that, in cases where non-crystallographic symmetry is present, there is a strong relationship in reciprocal space between structure factors (G-function coupling, Rossmann & Blow, 1962; Vellieux & Read,

1997), which leads to the simultaneous decrease of the conventional R factor and of the free R factor. This confirms that non-crystallographic symmetry leads to an effect on the free R factor (Kleywegt & Brünger, 1996). The free R factor then becomes a less sensitive indicator of phase correctness since the reflections which are removed for the free R-factor calculations are coupled to the reflections used for the conventional R factor by this G-function coupling effect. The results also indicate that the pseudo-atom refinement method will be far more effective when non-crystallographic symmetry is present in the crystals under investigation.

Examination of the electron-density maps computed with pseudo-atom phases or with combined phases suggest the presence of several minor errors, such as the misplacement of side chains, in the 3.0 Å resolution refined structure of *Pseudomonas aeruginosa* ornithine transcarbamoylase (Villeret *et al.*, 1995).

In summary, these results indicate that at medium resolution (3.0 Å), where the comparison was performed, pseudo-atom introduction into an electron-density map followed by refinement performs much better than iterative map skeletonization (at least in the case of OTCase, where non-crystallographic symmetry is present), with a phase improvement of 15.3° over the initial phases. These results probably apply as well to other procedures which rely on the introduction of pseudo-atoms in an electron-density map, followed by refinement of these atoms (namely the *ARP* and *wARP* procedures). It has also not escaped my attention that this procedure of

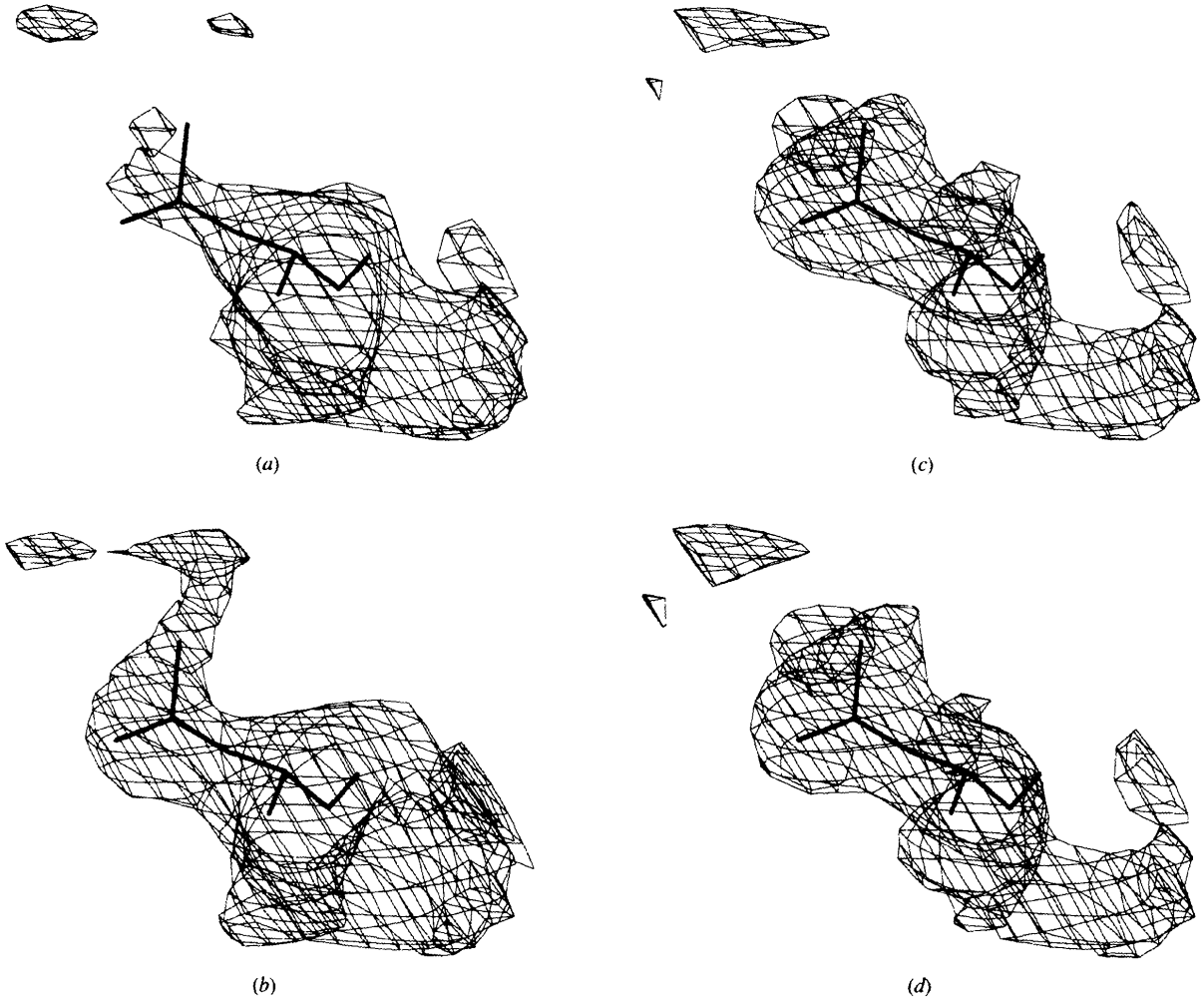

Fig. 1. 3.0 Å resolution electron densities corresponding to residue Leu24 of *Pseudomonas aeruginosa* ornithine trans-carbamoylase. (*a*) Electron density resulting from the procedure of iterative 12-fold averaging and solvent flattening (Villeret *et al.*, 1995). The map was calculated using $m\mathrm{Sim}|F_o|\exp(i\varphi_{calc})$; (*b*) electron density resulting from 20 cycles of iterative skeletonization with *DM* (Cowtan & Main, 1993). This map was calculated using coefficients of the form $w\mathrm{Sim}|F_o|\exp(i\varphi_{comb})$; (*c*) *SIGMAA* electron density computed with phases resulting from the refinement of pseudo-atoms. This map was calculated with coefficients of the form $3m|F_o| - 2D|F_c|\exp(i\varphi_{calc})$ for acentric reflections and $2m|F_o| - D|F_c|\exp(i\varphi_{calc})$ for centric reflections (Read, 1986; Vellieux & Dijkstra, 1997); (*d*) *SIGMAA* electron density computed with combined pseudo-atom refinement and density-modification phases, as described in §3. This map was calculated with coefficients of the form $3m|F_o| - 2D|F_c|\exp(i\varphi_{comb})$ for acentric reflections and $2m|F_o| - D|F_c|\exp(i\varphi_{comb})$ for centric reflections. All maps are contoured at a 1.0σ level, and were displayed with the program *O* (Jones *et al.*, 1991).

pseudo-atom introduction may be modified to automatically introduce and delete water O atoms in difference Fourier maps.

I thank Professor A. T. Brünger (Yale University) for having provided a new crystallographic program system which includes an implementation of Pannu & Read's maximum-likelihood refinement. I also thank Dr O. Dideberg (IBS/LCM) and Dr V. Villeret (University of Gent) for having provided the data of OTCase, and Dr A. Perrakis (NKI, Amsterdam) for the chitinase data. Part of this work was carried out during a short working visit to the Weizmann Institute of Science, Rehovot, Israel, in 1993 (with thanks to Professor J. L. Sussman). Financial support of the Commissariat á l'Energie Atomique and of the Centre National de la Recherche Scientifique is gratefully acknowledged.

References

Adams, P., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Agarwal, R. C. & Isaacs, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.
Greer, J. (1985). *Methods Enzymol.* **115**, 206–226.
Jones, T. A., Zou, Y.-J., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Brünger, A. T. (1996). *Structure*, **4**, 897–904.
Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* A**36**, 344–350.
Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* D**49**, 129–147.
Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.
Perrakis, A., Sixma, T. K., Wilson, K. S. & Lamzin, V. S. (1997). *Acta Cryst.* D**53**, 448–455.
Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S. & Vorgias, C. E. (1994). *Structure*, **2**, 1169–1180.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
Sim, G. A. (1960). *Acta Cryst.* **13**, 511–512.
Swanson, S. (1994). *Acta Cryst.* D**50**, 695–708.
Vellieux, F. M. D. (1997). *J. Appl. Cryst.* **30**, 400–401.
Vellieux, F. M. D. & Dijkstra, B. W. (1997). *J. Appl. Cryst.* **30**, 396–399.
Vellieux, F. M. D, Hunt, J. F., Roy, S. & Read, R. J. (1995). *J. Appl. Cryst.* **28**, 347–351.
Vellieux, F. M. D. & Read, R. J. (1997). *Methods Enzymol.* **277**, 18–53.
Villeret, V., Tricot, C., Stalon, V. & Dideberg, O. (1995). *Proc. Natl Acad. Sci. USA*, **92**, 10762–10766.
Wilson, C. & Agard, D. A. (1993). *Acta Cryst.* A**49**, 97–104.