Discrimination of solvent from protein regions in native Fouriers as a means of evaluating heavy-atom solutions in the MIR and MAD methods

Thomas C. Terwilliger[a]* and Joel Berendzen[b]

[a]Structural Biology Group, Mail Stop M888, Los Alamos National Laboratory, Los Alamos, NM 87545, USA, and [b]Biophysics Group, Mail Stop D454, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Correspondence e-mail: terwilliger@lanl.gov

An automated examination of the native Fourier is tested as a means of evaluation of a heavy-atom solution in MAD and MIR methods for macromolecular crystallography. It is found that the presence of distinct regions of high and low density variation in electron-density maps is a good indicator of the correctness of a heavy-atom solution in the MIR and MAD methods. The method can be used to evaluate heavy-atom solutions during MAD and MIR structure solutions and to determine the handedness of the structure if anomalous data have been measured.

1. Introduction

In the multiple isomorphous replacement (MIR) and multi-wavelength anomalous dispersion (MAD) approaches to determining macromolecular structures, a key step is the identification of the heavy-atom sites in the crystal lattice. There are two general approaches in current use for identifying these heavy-atom sites. These are Patterson-based searches, often carried out manually or by semi-automated procedures (Terwilliger *et al.*, 1987) or genetic algorithm-based methods (Chang & Lewis, 1994), and direct methods (Sheldrick, 1990; Miller *et al.*, 1994). Patterson-based and direct methods both begin by extracting differences between amplitudes of structure factors at different wavelengths or for derivative and native structures. The differences are then used to estimate structure factors corresponding to the heavy atoms that differ between the native and derivative structures or that scatter differently from X-ray wavelength to wavelength, and subsequently to deduce the partial structure of the heavy atoms. In extracting these differences, information on the structure as a whole and its handedness is discarded. Evaluating the quality of potential heavy-atom solutions is often difficult, particularly for Patterson-based methods, because many solutions often appear to agree to similar extents with a relatively noisy Patterson function. The purpose of this work is to point out that even a very simple but automatic evaluation of the features of a native electron-density map resulting from a heavy-atom model can be of enormous use in discriminating between correct and incorrect models. This information is complementary to the information contained in the differences used for Patterson-based or direct-methods identification of heavy-atom sites. Comparison of native Fourier maps based on different heavy-atom solutions can potentially discriminate between correct and incorrect heavy-atom solutions that otherwise appear of equal quality. If anomalous data have been measured, native Fourier maps can potentially distinguish the correct hand of the structure.

There are many features of an electron-density map that could be readily examined automatically and used to evaluate

whether the map is likely to represent a macromolecule in a crystal. Some of these are exactly the features that are examined and modified in current density-modification procedures and include the flatness of solvent regions (Wang, 1985; Podjarny *et al.*, 1987), differentiation of solvent and protein regions based on local r.m.s. density (Abrahams *et al.*, 1994) and the histograms of electron densities in a map (Zhang & Main, 1990). Other features that could potentially be used might include more detailed features of a map, such as connectivity of electron-dense regions and the shapes of these regions (Baker *et al.*, 1993).

We have chosen to make use of one of the simplest features of macromolecular crystals, the presence of distinct regions of solvent and macromolecule, to examine and evaluate the quality of an electron-density map in an automated fashion. Our approach is essentially to take the idea of solvent flattening to the level of a diagnostic. A typical electron-density map of a macromolecule consists of well defined regions that are relatively flat (solvent) and other regions that have a larger amount of variation (the macromolecule). In contrast, a map with random phases has a relatively uniform amount of

variation throughout. The measure of the non-random nature of the native electron-density map we use is the standard deviation, over the whole unit cell, of the local r.m.s. density (where the F_{000} term is not included in the calculation of the map). This standard deviation reflects how much the local r.m.s. electron density varies from position to position in the map. For an electron-density map with clearly defined solvent and macromolecule, the standard deviation in local r.m.s. density will be large (*i.e.* the r.m.s. density will vary from solvent region to macromolecule in the unit cell), while for a random map the standard deviation of r.m.s. density will be small (*i.e.* the r.m.s. density will be constant over the cell). Recent solvent-flattening approaches have used the variation in r.m.s. density as a means of identification of solvent regions in an electron density (*e.g.* Abrahams *et al.*, 1994). The approach taken here is similar to evaluating whether or not solvent flattening could be advantageously applied to a particular electron-density map.

We show here that an automatic examination of electron-density maps based on the variation of local r.m.s. density can be a useful indicator of the correctness of the heavy-atom solutions used to construct the maps and can be used to obtain the handedness of a heavy-atom solution.

2. Methods: calculation of the standard deviation of r.m.s. electron density in the unit cell

A set of heavy-atom sites is tested by using it to calculate phases and an electron-density map for the native structure, not including the F_{000} term in the map calculation. The electron-density map is calculated on a grid with a spacing of approximately one-third of the resolution of the data. To calculate the standard deviation of the local r.m.s. density, the asymmetric unit of the map is divided into cubes five grid units on an edge. Partial cubes with less than half the volume of a full cubes are ignored. The r.m.s. electron density in each cube is calculated using the grid points in the cube that are contained within the asymmetric unit of the crystal. Then the standard deviation of this set of r.m.s. values over the entire asymmetric unit is determined. Overlapping sets of cubes offset by one grid unit are used to cover the entire asymmetric unit. It is possible that inaccuracies in heavy-atom parameters can lead to large peaks or valleys in the native electron-density map at the positions of the heavy atoms. In order to reduce any systematic errors introduced in this way, grid points within three grid
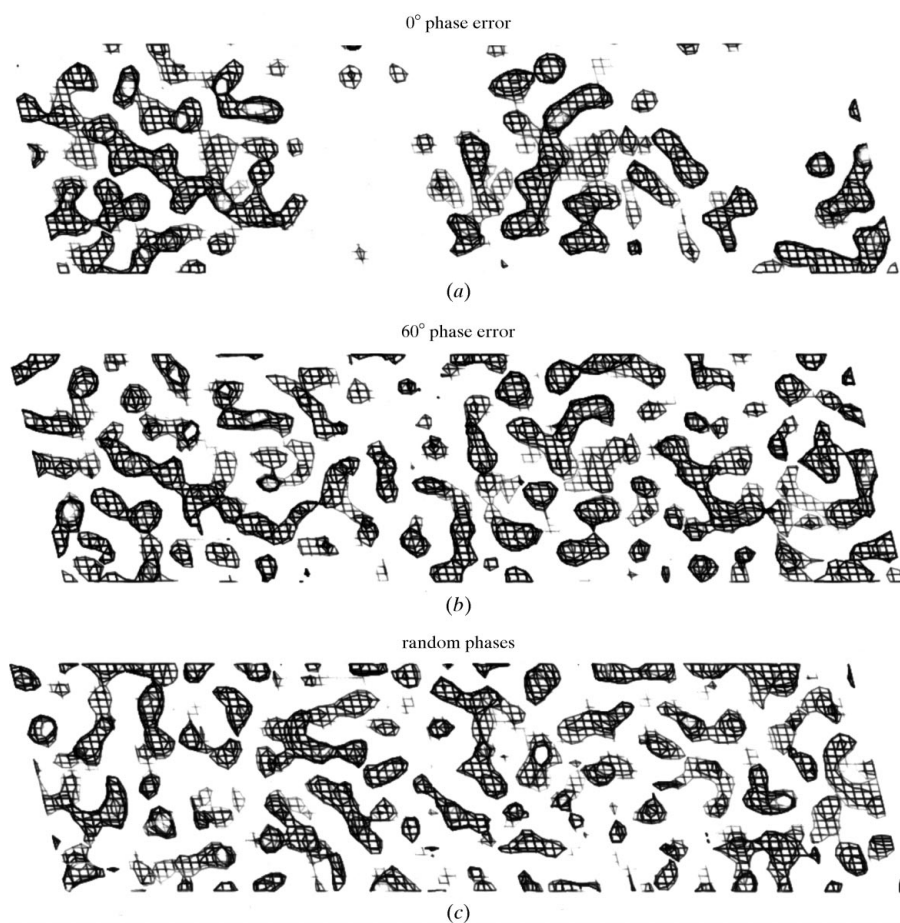

Figure 1
Sections through a model map, a map with a mean phase error of 60° and a map with random phases. Each map is calculated at a resolution of 2.5 Å. Amplitudes and phases of structure factors were calculated based on the gene V protein structure (PDB entry 1BGH) in space group *C*2 with unit-cell parameters $a = 76.08$, $b = 27.97$, $c = 42.36$ Å, $\beta = 103.2°$. Electron-density maps were calculated from these amplitudes and phases directly (*a*), after adding random errors to the phases to yield a mean phase error of 60° (*b*) and with random phases (*c*). Sections through each electron-density map are shown.

units of the highest and lowest N peaks in the map are excluded from the calculation. The number of peaks excluded (N) is chosen to be twice the number of expected heavy-atom sites.

3. Results and discussion

3.1. Standard deviation of r.m.s. density as a measure of distinction between solvent and macromolecule

To assess whether the standard deviation of r.m.s. density would be a useful measure of the quality of an electron-density map, we calculated model electron-density maps based on a known protein structure but with varying amounts of phase error. Fig. 1 shows sections through three model electron-density maps and Fig. 2 shows the distribution of r.m.s. electron density in local $5 \times 5 \times 5$ cubes within these maps. Each of these electron-density maps was calculated using the gene V protein structure in space group $C2$ (Skinner *et al.*, 1994) at a resolution of 2.5 Å. About half the unit cell is protein and half is solvent in this case. The section shown in Fig. 1(*a*) is from a map calculated from the gene V protein model structure with no added phase error. The map shows clear regions of solvent (which are flat) and of protein (where there is a high degree of variation). As expected (Fig. 2) curve *A* shows that many of the $5 \times 5 \times 5$ cubes sampled had r.m.s. variations near zero (the solvent region) and the remainder had a range of r.m.s. variations (the protein region). The overall standard deviation of the r.m.s. variation was 0.48 in units of normalized density (electron density/r.m.s. of the entire map, ρ/σ). In contrast, a map calculated using random phases results in an r.m.s. variation that varied very little for all the cubes sampled (Fig. 1*c*; Fig. 2, curve *C*). This map had a standard deviation of the r.m.s. variation of 0.17 units. A map calculated using phases offset from the model phases by about 60°, leading to an effective figure of merit of about 0.59, results in a distribution of r.m.s. variation that is close to the one observed for a random set of phases, but that has a slightly
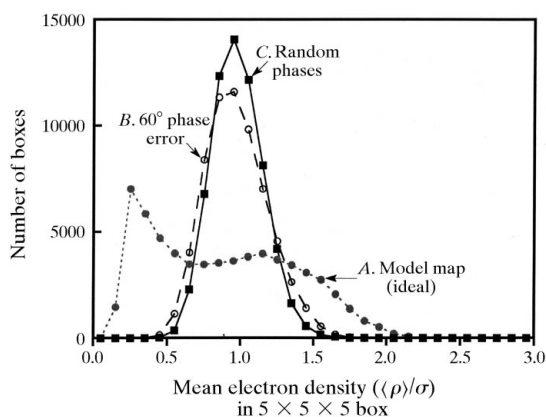
greater standard deviation of 0.21 (Fig. 1*b*; Fig. 2, curve *B*). It is this slight increase in standard deviation above that seen with a map calculated with random phases that we use to evaluate the quality of a map.

Fig. 3(*a*) illustrates the dependence of the standard deviation of r.m.s. density on the phase error of model maps calculated at a resolution of 2.5 Å. For maps with phase errors greater than about 80°, the standard deviation of r.m.s. density is essentially independent of phase error. For maps with phase errors up to 80°, however, the standard deviation of r.m.s. density decreases uniformly with increasing phase error. The box size used to calculate the standard deviation of r.m.s. electron density appears to have little overall effect on the calculation (compare the curves from boxes with sides 3, 5 and 9 units in Fig. 3*a*).

Fig. 3(*b*) illustrates the effect of resolution on the sensitivity of the method. The standard deviation of r.m.s. density at lower resolution (4 Å) has characteristics similar to those at higher resolution (2.5 Å), but it is much more noisy. Conse-
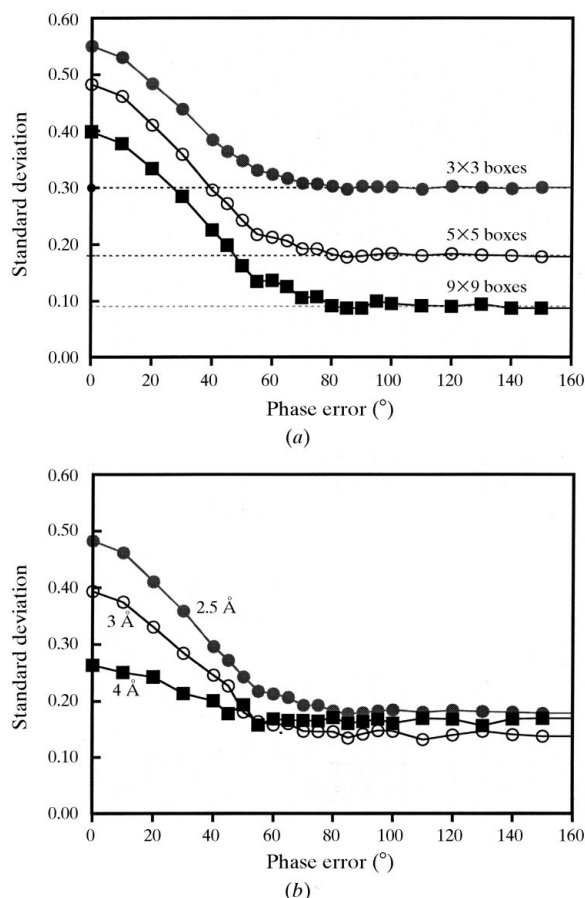

Figure 2
Distribution of r.m.s. density for the maps shown in Fig. 1. The r.m.s. electron density in local regions consisting of $5 \times 5 \times 5$ grid units was evaluated for each map in Fig. 1 and the number of local regions with each range of r.m.s. electron density is shown. Curve *A* is based on the map in Fig. 1 with no phase error, curve *B* on the map with a 60° phase error and curve *C* on the map with random phases.

Figure 3
Standard deviation of r.m.s. density as a function of mean phase error in the structure factors used to calculate the map. Amplitudes and phases of structure factors were calculated as in Fig. 1. Electron-density maps were calculated from these amplitudes and phases after adding random errors to the phases. (*a*) The standard deviation of r.m.s. density is plotted as a function of the mean phase error using box sizes of 3, 5 and 9 grid units on a side for maps calculated at a resolution of 2.5 Å. (*b*) The standard deviation of r.m.s. density is plotted as a function of the mean phase error using box size of 5 grid units on a side for maps calculated at resolutions of 2.5, 3.0 and 4.0 Å.

quently, this method has much more sensitivity at high resolution than low resolution.

These results indicate that the standard deviation of r.m.s. density might be a useful measure of the quality of a map for maps with up to about an 80° mean phase error.

3.2. Application to structure determination of a dehalogenase enzyme

In order to test the idea that the non-randomness of native Fourier maps can be used effectively to distinguish correct from incorrect heavy-atom solutions, we examined the Fourier maps calculated during the progress of structure determination (J. Newman, unpublished work) of a dehalogenase enzyme from *Rhodococcus* species ATCC 55388 (American Type Culture Collection, 1992). We have incorporated an evaluation of the non-randomness of native Fourier maps as described here into our automated structure-determination program (*SOLVE*; Terwilliger & Berendzen, manuscript in preparation) which was used to determine the dehalogenase structure. As each potential refined heavy-atom solution for this structure was evaluated, a native Fourier was calculated at a resolution of 2.5 Å and the standard deviation of its local variation was determined. In order to obtain an objective measure of the quality of these trial solutions, the native Fourier was also compared with a Fourier calculated from the model for the dehalogenase, which has now been refined at a resolution of 1.5 Å. In order to carry out this comparison of Fourier maps, the heavy-atom solutions were translated to match the origin used for the model structure. Additionally, trial solutions were separated into two matching groups related by inversion. Maps calculated using the group with the correct hand could be compared directly with the correct map, while those with the inverse hand could not be compared readily. Consequently, we analyzed the groups separately. First the group with the correct hand was examined to compare map correlations with the standard deviation of local variation of the native Fourier. The pairs of maps obtained from matching heavy-atom solutions with inverted handedness were then compared.

Fig. 4(*a*) shows the correlation coefficient between the trial map and the map calculated from the refined model as a function of the standard deviation of the local variation of electron-density maps for the dehalogenase, using heavy-atom solutions of the correct hand. For maps with standard deviation of normalized r.m.s. electron density below about 0.26 in this example, the non-randomness of the native Fourier is only weakly correlated with the quality of the map. For maps with standard deviation of normalized r.m.s. electron density above 0.26, however, the non-randomness of the native Fourier is very strongly correlated with the quality of the map. It is clear that the non-randomness of the native Fourier can be used effectively as a measure of the relative quality of different test heavy-atom solutions in this case. The solutions with a high degree of non-randomness are the solutions with a high correlation to the map based on the refined model.

In cases where anomalous differences have been measured, the non-randomness of the native Fourier can be used not only to evaluate the overall quality of a heavy-atom solution, but also to determine the correct handedness of the heavy-atom sites. Fig. 4(*b*) shows the non-randomness of the native Fouriers calculated for the dehalogenase structure using
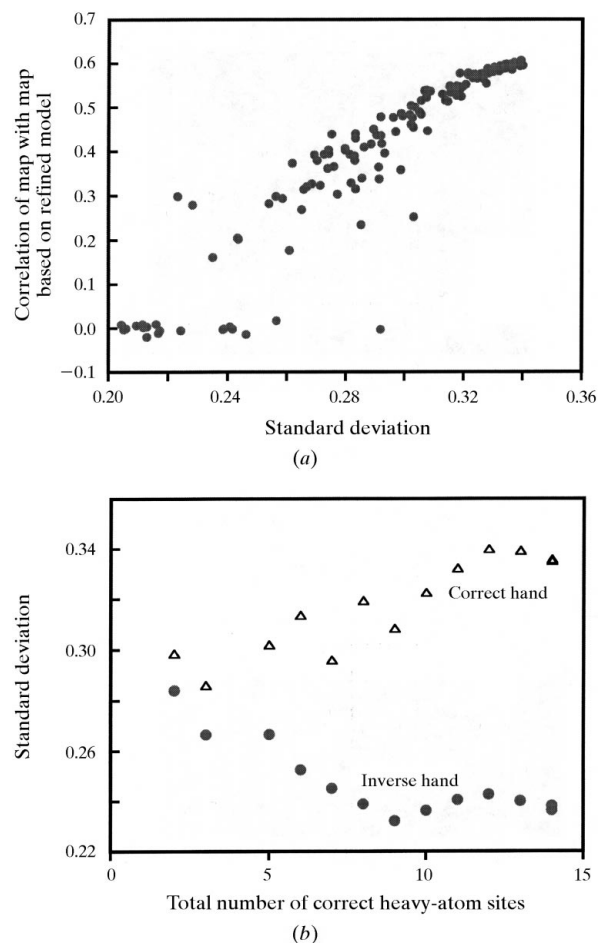

(*a*)

(*b*)

Figure 4
Standard deviation of local r.m.s. electron density during structure determination of *Rhodococcus* dehalogenase. The structure solution of *Rhodococcus* dehalogenase was carried out using the program *SOLVE* (Terwilliger & Berendzen, in preparation) based on data from a native and five derivatives (Au, Au, Hg, Pt and Sm heavy atoms) with anomalous differences measured for each derivative (J. Newman, unpublished data). *SOLVE* evaluated a total of 186 potential heavy-atom solutions during the course of structure determination. Each heavy-atom solution was compared with the final solution and an origin shift or inversion was applied if necessary to match the heavy-atom positions. As discussed in the text, (*a*) shows only solutions with the correct hand and (*b*) compares matching solutions with inverted handedness. (*a*) Non-randomness of native Fourier *versus* map quality. The abscissa is the standard deviation of the local r.m.s. electron density in the test native Fourier. The ordinate is the correlation coefficient between the native Fourier calculated from the trial-refined heavy atoms and the final refined model of the dehalogenase. (*b*) Non-randomness of the native Fourier as a function of the number of correct heavy-atom sites in test solutions for solutions of correct or inverted hand. The abscissa is the total number of correct heavy-atom sites in the five derivatives used in phasing, where a site was considered correct if it was within 1.5 Å of a heavy-atom site in the final solution in the appropriate derivative. The ordinates are the standard deviations of local r.m.s. density for native Fouriers calculated with correct and inverted handedness.

heavy-atom solutions that have the correct and inverted hands as a function of the number of correct heavy-atom sites used in phasing. Two heavy-atom solutions that are related by simple inversion will have identical phasing statistics and cannot be distinguished on that basis. Fig. 4(*b*) illustrates that the non-randomness of the native Fouriers calculated with the correct hand are readily distinguishable from those with an inverted hand.

4. Discussion and conclusions

The standard deviation, over the unit cell, of local r.m.s. density is a reasonable quantity to consider as a measure of the global quality of an electron-density map because it reflects an important component of the information in a map: the separation of solvent and macromolecule. The examples shown in Figs. 3 and 4 illustrate that it is indeed useful both in principle and when applied to actual structure determination. The non-randomness of the native Fourier discriminates most strongly between correct and incorrect solutions (*e.g.*, correct and inverted handedness) when phase calculation is most precise (Fig. 4). This is because when the phasing is very weak, the level of noise in the map can be so high that it masks any differences between the location of solvent and protein regions in the map.

The procedure described here will not be useful in every case, as some macromolecular crystals have very little solvent and others have very high solvent content. These crystals at the extremes of solvent fraction are not likely to have as clear a differentiation of solvent and macromolecule as those with about 50% solvent content. Consequently, the measure of non-randomness of the native Fourier used here might not be as useful as other algorithms that use connectivity of electron density or other measures of non-randomness.

As mentioned above, the evaluation of non-randomness of the native Fourier is based upon much the same criteria as identification of solvent and protein regions in density-modification procedures (*e.g.* Abrahams *et al.*, 1994). This means that successful identification of a correct heavy-atom solution is likely to be a good indication of the likelihood of successful application of density modification to the resulting electron-density map. This could provide a useful link in future auto-mated procedures that combine heavy-atom solutions with density modification.

References

Abrahams, J. P., Leslie, A. G. W., Lutter, R. & Walker, J. E. (1994). *Nature (London)*, **370**, 621–628.

American Type Culture Collection (1992). *Catalogue of Bacteria and Bacteriophages*, 18th ed., pp. 271–272.

Baker, D., Krukowski, A. E. & Agard, D. A. (1993). *Acta Cryst.* D**49**, 186–192.

Chang, G. & Lewis, M. (1994). *Acta Cryst.* D**50**, 667–674.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 210–220.

Podjarny, A. D., Bhat, T. N. & Zwick, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351–373.

Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.

Skinner, M. M., Zhang, H., Leschnitzer, D. H., Bellamy, H., Sweet, R. M., Gray, C. M., Konings, R. N. H., Wang, A. H.-J. & Terwilliger, T. C. (1994). *Proc. Natl Acad. Sci. USA*, **91**, 2071–2075.

Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 1–5.

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

Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* A**46**, 41–46.