Acta Cryst. (1982). A**38**, 21–29

A Density-Modification Method for the Improvement of Poorly Resolved Protein Electron-Density Maps

BY T. N. BHAT AND D. M. BLOW

Blackett Laboratory, Imperial College of Science and Technology, London SW7 2AZ, England

(*Received* 23 *February* 1981; *accepted* 16 *June* 1981)

Abstract

An efficient computer procedure has been developed for the extraction of regions of contiguous, well-connected high density from a three-dimensional electron-density map. This procedure may be used to generate an extended model volume, from a smaller volume based on a starting atomic model. The starting model may, for example, only include main-chain atoms for a protein, or may omit uninterpretable segments of chain. The procedure may also be used to extract regions of contiguous density where no model exists. The extended model volume is used to produce a properly scaled model electron density. Calculated structure factors are obtained from the scaled model electron density by fast Fourier transform, and combined, with appropriate weights, with the existing phase information to give improved phase angles. Calculation of a new electron-density map with the new phase angles initiates the next step of a cyclic procedure which converges rapidly. The procedure has been applied to the structure determination of tyrosyl-*t*RNA synthetase. It has led to identification of most of the available amino-acid sequence in the electron density, and a revised tracing of the main polypeptide chain. Evidence for improvement in phase angles is obtained from electron-density difference maps for substrate and inhibitor binding, in which a reduction in background density is observed.

Introduction

In protein crystallography, the isomorphous replacement method becomes progressively more difficult to use, and the resulting phase angles less reliable, as resolution is extended. Once the point is reached where a detailed atomic interpretation of almost all the molecule can be made, more accurate phase angles may be obtained from calculated structure factors (Watenpaugh, Sieker, Herriott & Jensen, 1971), and the way is open for further refinement by a variety of techniques. The resolution may then be extended as far as diffracted intensities are observable (Deisenhofer & Steigemann, 1975), with a substantial increase in the accuracy of interpretation.

It is, however, often difficult to reach the point where a sufficiently accurate atomic interpretation can be made, so that refinement on the basis of calculated phases can begin. 'Direct' methods have only given a clearly improved interpretation when they start from good phases at relatively high resolution [2·8 Å, de Rango, Mauguen & Tsoucaris (1975); 2·5 Å, Sayre (1974); 2·0 Å, Hendrickson & Karle (1973)], and often at a high cost in computation. Other methods have involved the modification of calculated electron densities to give a more physically reasonable distribution of electron-density values (Hoppe & Gassmann, 1968; Collins, Cotton, Hazen, Meker & Morimoto, 1975; Podjarny & Yonath, 1977). These methods have had some success at lower resolution, though with limitations which are clearly shown in the work on yeast $t\mathrm{RNA}_f^{\mathrm{Met}}$ (Schevitz, Podjarny, Krishnamachari, Hughes, Sigler & Sussman, 1979).

The production of sufficiently accurate phase angles by calculation from an atomic model is particularly difficult when the amino-acid sequence of the protein is unknown, when it cannot be confidently related to the observed electron density, or when some parts of the structure remain uninterpretable. In the case of yeast hexokinase *B* (Anderson, McDonald & Steitz, 1978), the interpretation was improved without an amino-acid sequence by a series of iterative steps, involving: (1) the assignment of amino-acid side chains by careful examination of the electron density; (2) real-space refinement of coordinates (Diamond, 1974, 1976) to optimize the fit of the postulated structure; (3) use of calculated phases to perform a difference Fourier refinement, with restrained structural shifts (Konnert, 1976); followed by reassignment of side chains [step (1)].

This procedure is a method of improving a protein model when the amino-acid sequence is not known (or only partially known), but there is an interpretable electron density map. It is an extremely laborious and time-consuming activity.

Following a technique suggested by Agarwal & Isaacs (1977), Evans & Hudson (1979) assign atomic

0567-7394/82/010021-09$01.00

positions to points on a regular grid whenever the electron density is above a threshold. The grid spacings are chosen to be approximately equal to the average atomic distance. The atomic positions are refined and a new electron density map is calculated for further study. Though the technique is aimed at improving the phase angles when the electron-density map does not allow assignment of amino-acid side chains, the technique has limitations because of its arbitrariness.

The method described in this paper is suitable both when the correct type of side chain is not known and when the map is not good enough for any detailed interpretation; it exploits the capabilities of existing computer techniques effectively; and we think that it can be successfully initiated at an earlier stage in the interpretation of the structure. The procedure operates directly on the electron-density distribution and never converts it to an atomic interpretation. It relies on selection of features from an electron-density distribution which could reasonably correspond to protein structure, and elimination of the remaining features. As we have used it, the procedure begins from a starting model of protein structure specified by a partial set of atomic coordinates. It would also be possible to use the procedure without any starting model.

The operation of the density modification procedure at any one point depends on the nature of the electron-density distribution which surrounds it. In this respect the method differs from previous density modification procedures in which the new density at any point depends only on the value of the starting density at the same point.

The fundamental principle is that an electron-density map of a protein at moderate resolution (3·5 to 2·5 Å) must consist of a contiguous band of electron density representing the polypeptide backbone, along which the side-chain densities form short projections. This principle is used to modify the phase information from isomorphous replacement, treating the contiguous density as a 'known' part of the structure for calculation of a phase probability distribution according to Sim (1959). The phase information is combined with that derived from isomorphous replacement, following the proposals of Rossmann & Blow (1961) and Hendrickson & Lattman (1970), with the technique of Bricogne (1976). The method can be iterated automatically.

The method has been used in attempts to make a more complete and more accurate interpretation of the structure of tyrosyl-*t*RNA synthetase from *B. stearothermophilus*, for which a partial and tentative structure has been given by Irwin, Nyborg, Reid & Blow (1976) at a resolution of 2·7 Å. To test the validity of the method, we have assessed the quality of derivative electron-density maps, and the interpretability of the protein electron-density map, with various sets of phase angles.

A particular difficulty in the interpretation of this structure is an unusually wide variation of the degree of order in different regions of the polypeptide chain. In this respect the structure does not conform to the physical characteristics which underlie all conventional direct methods – namely that the squared electron density resembles the actual electron density, modified only by convolution with an appropriate shape function.

Procedure

It is convenient to begin by describing the procedure used when a tentative atomic model has been made (this may well contain only main-chain coordinates without side-chain assignments). The procedure adopted without a model is described later.

The procedure includes seven steps which form an iterative cycle:

(1) Determination of 'occupancy' for each residue in the tentative model of the molecule as observed in the current electron-density map.

(2) Calculation of electron density for the tentative model, with the occupancies determined in step (1).

(3) Extraction from the current electron-density map of (*a*) features corresponding to the tentative model used in steps (1) and (2), or linked to it through regions of high electron density [and (*b*), optionally, other electron-density features which probably correspond to real features of the structure], to form the extended model volume.

(4) Generation of the extended model density on a finer grid, appropriate scaling of the model density to the electron-density map.

(5) Calculation of structure factors based on the extended model density.

(6) Modification of the phase information for each reflection, based on the calculated phase, pre-existing phase information derived from isomorphous replacement or any other source, and on the global agreement between observed and calculated structure factors.

(7) Calculation of a revised electron-density distribution with these modified phases.

This new electron-density map is used for reassignment of occupancies (step 1), and the whole procedure is iterated, using the same tentative model of the molecule. The process is found to converge within a few cycles, and then the latest revised electron-density distribution forms the basis for building a new tentative model. If this model contains a significant number of new features, a further application of the whole iterative procedure may be made.

Many of the steps mentioned above are performed by published procedures. Step (2) uses a procedure described by Agarwal (1978); steps (5) and (7) are performed by the fast Fourier transform technique

(Ten Eyck, 1973); step (6) employs the phase combination algorithm of Bricogne (1976). The novel parts of the procedure, to be described in detail, are steps (1), (3) and (4).

Step (1): *Determination of occupancy*

It frequently happens that the electron-density map of a protein shows a marked variation in the electron-density values observed at different points in the polypeptide chain. Such variations are generally due to disorder, which may be either thermal motion or 'static' disorder (differences between the time-averaged contents of different unit cells). For our purposes it is convenient to describe these density variations by determination of a suitable 'occupancy' for each amino-acid residue which provides a reasonable representation of the electron-density map.

An alternative way would be to define a Debye–Waller factor for each atom or group of atoms on the basis of the existing electron-density map. Phases determined by isomorphous replacement tend to be poorer at higher resolution and increase the apparent Debye–Waller factor. We are particularly concerned to improve phases at higher resolution using the available structural information. Debye–Waller factors tend to decrease the contribution by the model at higher resolution and we preferred to use occupancies which give a weight independent of $\sin\theta$. In a later step of the procedure, the tails of the electron-density distribution will be truncated, and this may be considered as giving a degree of 'sharpening' in reciprocal space.

The following procedure is used to determine an occupancy for each amino-acid residue in the starting model from the values of the starting electron density $\rho(\mathbf{x})$. The density is calculated on an orthogonal array of points $\mathbf{x}$, whose spacing is approximately one quarter of the resolution limit.

A logical array $E(\mathbf{x})$ is set up from the starting electron density $\rho(\mathbf{x})$, such that $E(\mathbf{x})$ is **true** at all values of $\mathbf{x}$ where $\rho(\mathbf{x})$ exceeds a chosen threshold value c_1. The tentative model consists of a list of $\mathbf{x}$ values for atomic centres. A fixed atomic radius is used to create a volume map $V(\mathbf{x})$ for each amino-acid residue, which is **true** for all values of $\mathbf{x}$ within an atomic radius of any atom in the residue. The fraction of **true** values of $V(\mathbf{x})$, which are also **true** in $E(\mathbf{x})$ gives the occupancy for the amino-acid residue.

By computing the overlap of two *logical* arrays, we are measuring the *volume* of the model density for the amino-acid residue in which the starting electron density is above threshold. It would also be possible to integrate the starting electron density over the model volume. The overlap of the logical arrays was felt to be a more appropriate measure of the desired quantity; in particular, for assessment of the quality or reliability of the interpretation of the density.

Step (2): *Creation of the model density array*

The model density array $M(\mathbf{x})$ is calculated from the coordinates of the tentative model, using a convenient electron-density distribution about each atom (Agarwal, 1978) and applying the occupancies determined in step (1). $M(\mathbf{x})$ and $\rho(\mathbf{x})$ are brought approximately to the same scale by equating the sum of electron densities within them at all points where both exceed a chosen value. [A more precise scaling is made in step (4).]

Step (3): *Extraction of selected features from the electron density*

This is the most important step of the new procedure. Its aim is to recognize features in the starting electron-density map which correspond to an image of protein structure at moderate resolution. In step (3*a*), contiguous regions of density are selected, provided they connect with any part of the tentative model, and provided they are not connected to the model through a 'neck' which is too narrow. In the optional step (3*b*) other large continuous regions of density are also selected, again subject to the rejection of densities connected by unduly narrow 'necks' of linking density.

For economy in array size in step (3), it is convenient to work on a coarser grid, whose spacing can be about 1·5 times the grid spacing used in other steps. Thus, working from a 2·7 Å resolution electron-density map, we have used a grid spacing of about 1 Å in this step, reducing the number of grid points to less than one third. The symbol $\mathbf{y}$ will be a position vector on this coarser grid.

Step (3*a*): *Recording the outline of the extended model volume*

Three logical arrays $S(\mathbf{y})$, $E(\mathbf{y})$, $O(\mathbf{y})$ are used. S and O can be considered as 'searching' and 'output' arrays. $E(\mathbf{y})$ is a starting electron density array, similar to $E(\mathbf{x})$ in step (1) but on the coarser grid.

Throughout the procedure, any point in the electron-density map may be marked as S, E or O, or none of these. No point is ever more than one of S, E or O. The whole procedure may be envisaged as the placing of one of three kinds of marker, S, E or O, at points of the electron-density map.

In each step of the procedure, S represents points in the map which are the growing surface of the extended model volume. One by one, each point in S is deleted from S and entered into O. We then consider every neighbouring point which is not yet part of O or S. Such a neighbouring point becomes part of S if the starting density is large (*i.e.* it is part of E), *and* if a sufficient number of its neighbours are part of S, E or O (*i.e.* if it is sufficiently well connected to points where the starting density or the model density is large).

To start with, all points in the model density, $M(\mathbf{y})$, which exceed a chosen value c_2 are accepted as part of S. Any of these points which are already marked E [where the starting density $\rho(\mathbf{y})$ exceeds c_1] are removed from E.

This procedure may be expressed formally as follows:

Initially, for all $\mathbf{y}$:

[$O(\mathbf{y}) =$ **false**
$S(\mathbf{y}) =$ **true** *if* $M(\mathbf{y}) \geq c_2$, *else* $S(\mathbf{y}) =$ **false**
$E(\mathbf{y}) =$ **true** *if* ($\rho(\mathbf{y}) \geq c_1$ *and* $S(\mathbf{y}) =$ **false**),
else $E(\mathbf{y}) =$ **false**.]

Repetitive procedure for each $\mathbf{y}$ where $S(\mathbf{y})$ is **true**:

[$O(\mathbf{y})$ becomes **true** *and* $S(\mathbf{y})$ becomes **false**.
For each $\mathbf{y}' =$ *adjacent* $\mathbf{y}$, where $O(\mathbf{y}')$ is **false**:
{*If* $E(\mathbf{y}')$ is **true** *and* more than n **trues** are found amongst ($E(\mathbf{y}'')$ *or* $S(\mathbf{y}'')$ *or* $O(\mathbf{y}'')$, where $\mathbf{y}'' =$ *adjacent* $\mathbf{y}'$) then $S(\mathbf{y}')$ becomes **true** *and* $E(\mathbf{y}')$ becomes **false**}.]

The repetitive procedure is continued over each remaining **true** point in S, until all members of S are **false**. The array O then defines the extended model volume.

A point $\mathbf{y}'$ is defined as *adjacent* $\mathbf{y}$ if $|\mathbf{y}' - \mathbf{y}| \leq$ some predetermined neighbour distance d. In our work d has been taken as the spacing of the grid on which the density values $\rho(\mathbf{y})$ are defined.

Step (3*b*): *Inclusion of other significant features in extended model volume*

This step is optional. It is designed to allow the addition of other features to the extended model volume, even when they are not linked to the points represented in the starting model, M. The criterion for inclusion is that the feature should form a contiguous, well-connected region of density above the threshold level, which exceeds a certain minimum volume.

After the array O has been set up, defining the extended model volume as described for step (3*a*), the only remaining **true** values in E are at points which are not adjacent to **true** values in O, or, if adjacent, are not well connected to other points of O. In step (3*b*) the remaining **true** values in E are tested to see whether they can serve as a nucleus for a contiguous, well-connected region of high density, containing at least v grid points.

One remaining true value in $E(\mathbf{y})$ is transferred to $S(\mathbf{y})$, and is used as a starting point for a repetitive procedure similar to that described for step (3*a*). The only difference is that instead of using the array O, a new array O' is used, which is made all **false** before the repetitive procedure begins. During the repetitive procedure, other **true** values can be transferred from E to S and from S to O'. When there are no more **true** values in S, the total number of **true** values in O' is checked. If this number exceeds v, the minimum acceptable volume, all the **true** values in O' are transferred to O and form part of the extended model volume. If any further **true** values remain in E, the array O' is set all **false**, and the procedure is repeated.

This additional procedure is relatively expensive in computer time, as it requires a fresh start of the repetitive procedure for each small 'island' in the starting electron density which is above the threshold.

It may be noted that this procedure could conceivably be used without any starting model. In this case, steps (1), (2) and (3*a*) would be omitted. Such a method might be useful in the first interpretation of a relatively straightforward electron-density map, but we have not tried to use it in this way.

Step (4): *Generation of the scaled, extended model density on a finer grid*

The procedure described above is designed to be executed rather efficiently in a large computer, with a coarse grid so the three-dimensional logical arrays can be accommodated in the available storage. But to avoid the effects of series termination in the subsequent Fourier transform steps, and to describe the model density as accurately as possible, it is necessary to use a finer grid. In step (4) the extended model volume is used to generate the extended model density at finer intervals. At the same time the model density is rescaled to give a better fit to the starting density ρ.

The extended model density will contain high values of electron density throughout the model volume, surrounded by regions of low uniform density c_3, with sharp gradients of density between them. These sharp gradients will be rounded by the Fourier transform steps, but it is essential that the sampling of the final extended model density grid is sufficiently fine. Otherwise strong contributions to the Fourier transform, generated by the abrupt changes of electron density, will have effects within the domain of reciprocal space where crystallographic data exist. For this purpose, a grid spacing less than about 0·25 of the crystallographic limit of resolution is required for $\rho(\mathbf{y})$ and $M(\mathbf{y})$ in step (4).

The symbol $\mathbf{x}$ will represent a coordinate on the fine grid. Every point on the fine grid can be mapped to a non-integral point on the coarse grid by a transformation $\mathbf{x} = \mathsf{T}\mathbf{y}$. A logical array $L(\mathbf{x})$ is constructed which is **true** *if* $O(\mathsf{T}^{-1}\mathbf{x})$ is **true**, *else* $L(\mathbf{x})$ is **false**. $O(\mathsf{T}^{-1}\mathbf{x})$ means $O(\mathbf{y})$ for the value of $\mathbf{y}$ which most nearly satisfies $\mathbf{x} = \mathsf{T}\mathbf{y}$. L therefore represents the extended model volume on the fine grid.

The next step is the creation of the extended model density M' from the extended model volume. It is impor-

tant that M be correctly scaled, since M' at each point of L is the larger of ρ and the scaled M.

The further scaling operation is carried out on the fine grid, so as to increase the number of values used, and so improve the statistical aspect of the scaling. A least-squares fitting operation is carried out to minimize $\sum [\rho(\mathbf{x}) - (KM(\mathbf{x}) + d)]^2$, summed over all values of $\mathbf{x}$ where $\rho(\mathbf{x}) \geq c_1$ and $M(\mathbf{x}) \geq c_2$. The values of K and d derived from this procedure are used to form the scaled extended model density $M'(\mathbf{x})$. For every value of $\mathbf{x}$ where $L(\mathbf{x})$ is **true**, $M'(\mathbf{x})$ is chosen as $max\ [\rho(\mathbf{x}), KM(\mathbf{x}) + d]$. If $L(\mathbf{x})$ is **false**, $M'(\mathbf{x})$ is assigned a background value c_3, which in our work has generally been chosen as zero.

It is found necessary to make a careful examination of the scaled extended model density. The scale factor determined by the procedure described above will give good results when the shapes of the electron-density peaks in the two maps are similar, and when there is a good overlap between the peaks of the two maps. If these conditions are not satisfied it may be necessary to make final adjustments from a detailed examination of the relative values of adjacent density contributions from the two maps. If, for example, the density corresponding to main chain comes from the model density M, and the density for the side chain comes from the starting density ρ, the densities of the main chain and the side chain should be made similar.

The extended model density includes the density in the model volume together with high densities of ρ which are contiguous and well connected with the model volume. Outside this it has a background level $\leq c_3$. The extended model density has the property that from any grid point that is not at the background density it is possible to reach at least one grid point where $M'(\mathbf{x}) > c_2$, by stepping through adjacent grid points which have non-background values of $M'(\mathbf{x})$, and each of which have at least n adjacent points where either $M'(\mathbf{x}) > c_2$ or $\rho(\mathbf{x}) > c_1$. The extended model density therefore forms continuous regions of density above the threshold level, in which narrow necks of density are eliminated (Fig. 1).

Steps (5), (6) *and* (7): *Fourier transforms and phase combination*

Structure factors are calculated from the scaled extended model density $M'(\mathbf{x})$ by an inverse fast Fourier transform in the appropriate space group (Ten Eyck, 1973). Calculated structure amplitudes may be compared with the experimental X-ray data by evaluation of an R factor. A low R factor is not necessarily a criterion of success: it may simply mean that M' is derived almost entirely from the starting electron density ρ. Similarly, the phase angles may be compared with the phase angles used for the starting electron density, or with phase angles derived from the model M. If they agree too closely with either this is an indication that the procedure is not achieving anything new. The program *RFACTOR* (J. L. Ladner & R. C. Ladner, personal communication) is convenient for these comparisons.

The phase angles calculated from the scaled extended model density are combined with other existing phase information (*e.g.* isomorphous replacement) by the method of Bricogne (1976). We have not been able to define any theoretical basis for the weight to be given to the calculated phases: a weight was chosen which gave phases about equally different from the model phases and the isomorphous replacement phases.

A fast Fourier transform with the combined phases so generated produces an improved electron density map $\rho(\mathbf{x})$ which can be compared with the model in step (1) on a further cycle.

Iteration of the procedure

The cyclic application of steps (1)–(7) is found to be highly convergent, and in our experience the procedure is complete after three or four cycles. It is then necessary to make a detailed examination of the new calculated density, to see whether it suggests that a new set of model coordinates should be used, significantly different from the model used in steps (1) and (2) of each cycle of the previous procedure. This may include the interpretation of new stretches of polypeptide backbone which were not sufficiently recognizable in the previous map; the inclusion of specific amino-acid side chains which can now be confidently recognized; the deletion of features corresponding to an earlier incorrect interpretation; or the more accurate inter-

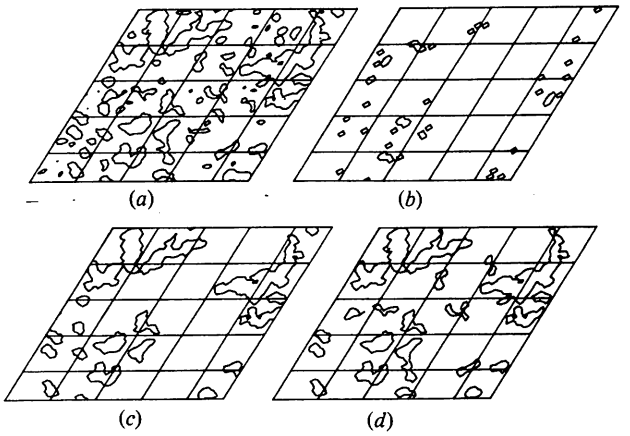


Fig. 1. A two-dimensional example of generating an extended model volume (step 3). (*a*) Starting density $\rho(\mathbf{x})$; (*b*) model density $M(\mathbf{x})$; (*c*) outline of extended model volume $O(\mathbf{y})$ = **true** produced by step (3*a*). This includes all the densities of map (*a*) which are well connected to the densities of map (*b*). Map (*d*) is produced by step (3*b*) on map (*a*) without reference to the model. It has all well-connected densities of map (*a*) whose area exceeds an arbitrarily chosen critical area.

pretation of a structure already recognized. Some of the newly interpreted features may arise from features which were not included in the previous model, due to the use of step (3*b*), which automatically includes large features in the calculated density, even though they do not correspond to any feature of the model. If a significantly improved model can be arrived at, a new iteration of the procedure can be initiated, resulting in a further improved electron-density map.

Although our application of the procedure has always used a starting model, one can envisage starting from a map in which no interpretation can be given in terms of atomic coordinates. In this case, strong positive features on the electron-density map can be picked up as structural features in step (3*b*) and steps (5), (6) and (7) can be pursued to generate a new electron-density map.

Application

The method was applied to tyrosyl-*t*RNA synthetase (Irwin *et al.*, 1976), a crystalline protein whose asymmetric unit contains one subunit (molecular weight 45 000 Daltons) of the dimeric enzyme molecule. Three isomorphous derivatives had been used to obtain phase information to a resolution of 2·7 Å (MIR phases), though beyond 3 Å measured structure amplitudes were weak and there was a severe drop in the phasing power of two isomorphous derivatives. The resulting electron-density map had given rise to a tentative tracing of the polypeptide backbone covering about 65% of the expected length of the polypeptide chain: however, there were several breaks in the continuity of the proposed polypeptide chain as represented in the electron-density map. Two substantial fragments of the amino-acid sequence have been determined (Winter, 1979) but these had not been recognized in the electron density.

Structure amplitudes had also been obtained for crystals which had been treated with the inhibitor tyrosinyl adenylate (Monteilhet & Blow, 1978), and for crystals in which the intermediate on the catalytic pathway, tyrosyl adenylate, was formed (Rubin & Blow, 1981). Tyrosinyl adenylate gave an electron-density difference map which was clearly interpretable, with the isomorphous replacement phase angles, while the difference map for tyrosyl adenylate was much more noisy.

A chain of polyalanine was built to follow the putative chain tracing, and its coordinates were subjected to two rounds of real-space refinement (Diamond, 1974). Structure factors calculated from the resulting coordinates had a mean phase difference of almost 90° when compared to MIR phases. A difference map for the inhibitor, with these calculated phases, gave no significant peak at the inhibitor binding site, indicating that the phases were much less accurate than the MIR phases.

The procedure decribed in the previous section was carried out on electron-density arrays containing some 108 000 points at a spacing of about 1·0 Å (coarse grid). The model electron density was constructed from a polyalanine chain of 267 residues and the starting array $S(\mathbf{y})$ contained about 4000 **true** values. The procedure propagated these into a contiguous volume map of nearly 30 000 values in $O(\mathbf{y})$, using about 1 min of CPU of an IBM 370/165. Step (3*b*) was not used at this stage. The scaled extended model density was calculated in an array of 500 000 points at a spacing of about 0·6 Å. After calculation of the fast Fourier transform, combination of the calculated phase angles with the isomorphous phases required 40 s CPU time. Only very small changes were made during a second cycle of the same procedure, in which the new starting electron density was based on the combined phases, and the procedure was judged to have converged. Fig. 2 summarizes the results for the first cycle. For the second cycle the new phases recorded a r.m.s. difference of 22° from the calculated phases and 48° from the starting or MIR phases. Examination of phase changes as a function of figure of merit derived from the MIR phases show that smaller phase changes have occurred for the reflections which are already accurately phased by isomorphous replacement (Fig. 2).

The combined phase angles derived from this first application of the procedure were used to calculate a new difference map for tyrosyl adenylate, which had significantly less noise.

The phases were used to compute a revised electron-density map, whose detailed study led to a substantial re-interpretation of the structure. The interpretation was made in collaboration with Dr Peter Brick and Dr Jens Nyborg. Two new long α-helices were discovered

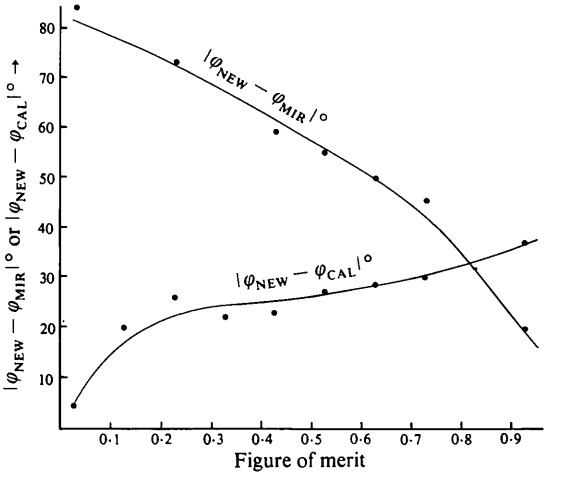


Fig. 2. The phase change after the first cycle of the density modification procedure as a function of figure of merit of the starting phases.

Table 1. *Progress of density modification procedure*

	Starting model	Cycle*	$\langle\alpha_{comb} - \alpha_{MIR}\rangle$	$\langle\alpha_{comb} - \alpha_{calc}\rangle$
Round 1	Tentative chain tracing, 276 amino acids	Cycle 1†	31°	45°
	Wire model, real-space refinement	Cycle 2	48°‡	22°‡
Round 2	Revised chain tracing, 318 amino acids, 78 side chains	Cycle 1	36°	37°
	Wire model, using computer graphics for some α-helical regions,	Cycle 2	35°	36°
	model regularization for remainder	Cycle 3	34°	34°

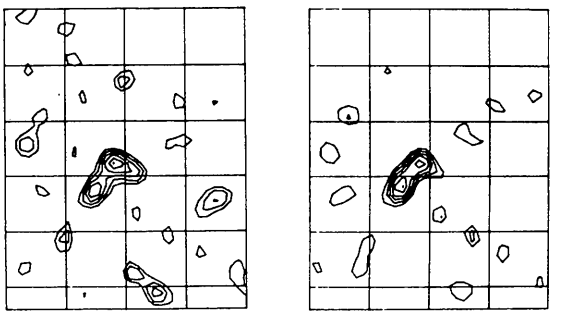
* In each cycle starting electron-density map ρ obtained by combining phase information of α_{MIR} and α_{calc} derived from extended model density, by the method of Bricogne (1976).

† Starting density map from MIR phases.

‡ The increase of $\langle\alpha_{comb} - \alpha_{MIR}\rangle$ and the very low value of $\langle\alpha_{comb} - \alpha_{calc}\rangle$ in this cycle are both a consequence of assigning a high weight to α_{calc} in the phase combination.

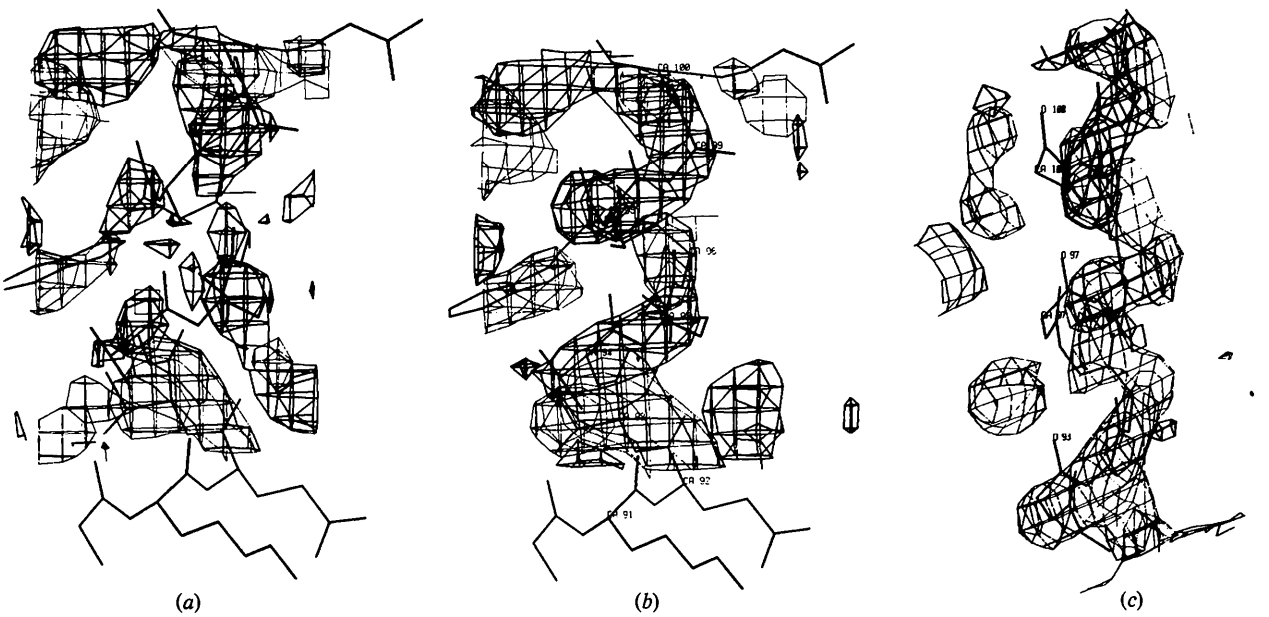
in regions of density which we had previously failed to interpret, and it also appeared that several errors of connectivity had been made in the linking of chains recognized in the MIR map. In some other regions where breaks had been noted in the original map, density was now found to be continuous. In order to interpret the polypeptide chain in this new map, it was still necessary to apply the criteria that the main chain should be continuous, and that it should contain no closed loops. Even with the careful application of these criteria, it was impossible to be certain that the polypeptide chain density had been interpreted in an unambiguous manner.

This interpretation led to a second atomic model, in which some known segments of amino-acid sequence could be confidently aligned to the electron density. The second atomic model included 318 amino-acid residues,



(*a*) (*b*)

Fig. 3. Electron-density difference maps for the intermediate tyrosyl adenylate computed (*a*) with MIR phases, (*b*) with phases obtained at the end of the procedure. Contours are drawn at constant intervals starting at 2/7 of the phosphate peak height.



(*a*) (*b*) (*c*)

Fig. 4. A comparison of the electron density in a helical segment of the molecule, as found in the MIR map, and using the final phases. (*a*) MIR phases, final model. (*b*) Final phases, final model. (*c*) A slightly different view of the density from the final phases, showing the model coordinates which were used when this segment was first recognized as an α-helix, at the beginning of the second round of density modification.

for which the polypeptide chain had been completely rebuilt. There was little electron density in the map for 80 or more residues at the carboxyl terminus of the chain, and this part of the structure was not interpretable. The coordinates were recorded from a Kendrew wire model and bond lengths and angles were regularized by an energy refinement procedure (Dodson, Isaacs & Rollett, 1976). In this model side-chain atoms were included for 78 amino acids, as their positions were fairly convincing. These coordinates were used to generate the model density M, and the entire procedure of density modification was repeated and iterated for three cycles. Phase angles obtained in the third cycle were used to calculate a difference map for tyrosyl adenylate. This difference map (Fig. 3) showed a further increase in the contrast of the tyrosyl adenylate density over background, indicating that the phase angles are further improved. Phase changes from cycle to cycle are given in Table 1.

Part of the electron-density map obtained with these phases is shown in Fig. 4. This is an α-helical part of the molecule which was unrecognizable in the original isomorphous replacement map, and was only crudely represented in the second atomic model. In this region we were still unable to place the atoms of a regular helix precisely into the density (Fig. 4*c*). The second application of the procedure has produced an improved helical array of density, but has not followed the position given by the atomic model. This demonstrates that the procedure does not simply reproduce the features of an assumed atomic model. This is also shown by the identification of new links and the deletion of wrongly traced linkages (as many as ten residues) in the revised structural model.

This electron-density map was used for the construction of a third atomic model, for which an Evans & Sutherland graphical display system was used. In this model some 200 amino-acid side chains were introduced, the alignment of the amino-acid sequence being recognized from the electron density with reasonable confidence. There are still at least 80 amino acids missing at the carboxyl terminus of the polypeptide chain. The density fades abruptly in an α-helical section, and it seems likely that the disappearance of density is due to some kind of considerable disorder in this part of the polypeptide chain. Polyacrylamide gel analysis of protein from dissolved crystals confirms that the crystalline material is undegraded.

In a further application of the procedure, step (3*b*) was used. The electron-density map contained about 18 000 Å^3 above the threshold density. With about 30 s of IBM 370/165 CPU time, the program selected a large region of continuous density, covering approximately 1500 Å^3, which probably represents a significant fraction of the missing polypeptide chain, but which is not clearly enough resolved to allow any interpretation to be made.

Conclusion

The procedure presented here is not completely automatic. It requires the assignment of some parameters on a rather arbitrary basis, and careful monitoring of the effects of these choices. In particular, we have not found a completely satisfactory method for bringing the calculated electron density ρ and the model density M onto the same scale. The relative weighting of the newly calculated phase angles to the previously existing (isomorphous replacement) phase angles lacks any objective basis.

Nevertheless, the method has achieved a substantial improvement in the interpretability of a protein electron-density map, and in the accuracy of phase angles as demonstrated by the quality of electron-density difference maps. This has been achieved without laborious model building and amino-acid assignment decisions at each cycle of refinement, and with very modest use of computer time.

Computing facilities and financial support for TNB were provided by the SERC Daresbury Laboratory.

References

AGARWAL, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
AGARWAL, R. C. & ISAACS, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.
ANDERSON, C. M., MCDONALD, R. C. & STEITZ, T. A. (1978). *J. Mol. Biol.* **123**, 1–13.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
COLLINS, D. M., COTTON, F. A., HAZEN, E. E. JR, MEKER, E. F. JR & MORIMOTO, C. N. (1975). *Science*, **190**, 1047–1053.
DEISENHOFER, J. & STEIGEMANN, W. (1975). *Acta Cryst.* B**31**, 238–250.
DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.
DIAMOND, R. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLÁČEK, pp. 293–301. Copenhagen: Munksgaard.
DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 311–315.
EVANS, P. R. & HUDSON, P. J. (1979). *Nature (London)*, **279**, 500–504.
HENDRICKSON, W. A. & KARLE, J. (1973). *J. Biol. Chem.* **248**, 3327–3334.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–107.
IRWIN, M. J., NYBORG, J., REID, B. R. & BLOW, D. M. (1976). *J. Mol. Biol.* **105**, 577–586.
KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.
MONTEILHET, C. & BLOW, D. M. (1978). *J. Mol. Biol.* **122**, 407–417.
PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* A**33**, 655–661.

RANGO, C. DE, MAUGUEN, Y. TSOUCARIS, G. (1975). *Acta Cryst.* A**31**, 227–233.
ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641–647.
RUBIN, J. R. & BLOW, D. M. (1981). *J. Mol. Biol.* **145**, 489–500.
SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.
SCHEVITZ, R. W., PODJARNY, A. D., KRISHNAMACHARI, N., HUGHES, J. J., SIGLER, P. B. & SUSSMAN, J. L. (1979). *Nature (London)*, **278**, 188–190.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 183–191.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1971). *Acta Cryst.* B**29**, 943–956.
WINTER, G. (1979). In *Transfer RNA: Structure, Properties and Recognition*, edited by P. SCHIMMEL, D. SÖLL & J. ABELSON, pp. 255–265. Cold Spring Harbor Monograph 9A.

Acta Cryst. (1982). A**38**, 29–33

Lattice Dynamical Calculations of the Mean Square Amplitudes of Crystalline Biphenyl

BY H. BONADEO* AND E. BURGOS*

División Física del Sólido, Comisión Nacional de Energía Atómica, A. del Libertador 8250, 1429 *Beunos Aires, Argentina*

(*Received* 14 *November* 1980; *accepted* 6 *May* 1981)

Abstract

The molecular mean-square-amplitude matrix of crystalline biphenyl, $C_{12}H_{10}$, is calculated with an intermolecular potential of the atom–atom type. The effect of the presence of the low-lying torsional mode, which interacts with translational modes, is discussed. The Born **S**-matrix method is used, and proves to be an excellent approximation, which takes about $\frac{1}{5}$ of the computer time of the exact calculation. The resulting amplitudes are in fair agreement with experiment, and show that the extremely high amplitude of libration about the long molecular axis may be satisfactorily explained without assuming a double-well shape for the torsional potential.

Introduction

The study of the thermal motions of the atoms in molecular crystals is of considerable interest for researchers in the field of X-ray and neutron diffraction and NQR spectroscopy, among others. Crystallographers have developed several models to interpret atomic displacements in terms of molecular motions, using diffraction data. Cruickshank (1956*a*,*b*) proposed his **TL** model of thermal motion, where the molecules are supposed to be rigid, and no interactions between translations and rotations are allowed; when this restriction is lifted, the widely used **TLS** model of Schomaker & Trueblood (1968) results.

A further complication arises when low-lying internal vibrations, which mix with the external modes, like the butterfly mode of naphthalene and the torsional mode of biphenyl, are present.

Although the usual crystallographic rigid-body model can be partially successful in accounting also for the mean square displacements relative to these internal modes, the treatment is nearly always incomplete: a practical demonstration of the limits of such a procedure is given here for biphenyl.

There are several difficulties associated with the correct interpretation of molecular thermal motions from diffraction data; in this sense, a lattice dynamical approach is particularly useful, since it allows correlation with other experimental data, like vibrational frequencies (see Filippini, Gramaccioli, Simonetta & Suffritti, 1974). A series of lattice dynamical calculations of the molecular mean-square-amplitude tensors **L**, **T** and **S** have been published (McKenzie & Pryor, 1971; Luty, 1972; Filippini, Gramacioli, Simonetta & Suffritti, 1973; Cerrini & Pawley, 1973; Pawley, 1972; for instance); the correct values of the **L** tensors are particularly important to obtain bond-length corrections.

In the present work we present a lattice dynamical calculation of the molecular mean-square-amplitude tensor of crystalline biphenyl at two temperatures, in the harmonic approximation. Although the semi-empirical atom–atom parameters due to Williams (1966) used in the calculations have not been adjusted

* Fellow of the Consejo Nacional de Investigaciones Científicas y Técnicas.

0567-7394/82/010029-05$01.00