However, the signal-to-noise ratio can be improved by rejecting intensity pairs if $(p+q)$ is below some limit. Rejection of pairs of observations on this basis does not bias the expected distribution for $S(H)$. This can be verified by evaluating the integrals (6) and (13) within these new limits. The $S(H)$ take the same form as before, multiplied by a scale factor corresponding to the fraction of the joint probability distribution within the new limits.

Errors in the estimation of α may also be introduced if individual reflections of low intensity are rejected (as may occur with an I/σ cutoff). When α is small, p may be very small while q is large, or *vice versa.* Omission of reflection pairs of this type will result in an overestimation of α. For this reason, an I/σ cutoff should not be imposed on data prior to determination of the twinning fraction.

As previously mentioned, we have assumed that the true untwinned intensities of two reflections related by the twinning operation are statistically independent. In crystals where a non-crystallographic symmetry operation nearly coincides with the twinning operation (Rees & Lipscomb, 1980), this assumption is not valid. Application of the statistics presented here to such a case leads to an overestimation of the twinning fraction.

Concluding remarks

The statistics of a new parameter, H, take a simple form for intensity data from a hemihedrally twinned specimen, and are sensitive to the twinning fraction, α. The observed statistics for H may be compared with the predicted statistics, allowing one to obtain an accurate value for the twinning fraction. Rejection of weak pairs of observations on the basis of the sum of the twin-related intensities allows a more accurate determination of α, without introducing bias.

In addition to the application to twinning, H can be defined in terms of the intensities of equivalent reflections from two different data sets. The values of $\langle|H|\rangle$ and $\langle(H^2)\rangle$ could provide measures of similarity between data sets, in a fashion similar to the crystallographic R factor. The expected values of these terms for unrelated data sets are given by ($8a$), ($8b$), ($14a$) and ($14b$) with $\alpha=0$. These terms differ from the R factor in that they are not dominated by the reflections of highest intensity. Instead, all reflections contribute similarly to the average.

This work was supported in part by NIH training grant GM07185. Assistance from D. C. Rees, B. T. Hsu and M. C. Yeates is appreciated.

References

Britton, D. (1972). *Acta Cryst.* **A28**, 296-297.
Fisher, R. G. & Sweet, R. M. (1980). *Acta Cryst.* **A36**, 755-760.
Murray-Rust, P. (1973). *Acta Cryst.* **B29**, 2559-2566.
Rees, D. C. (1980). *Acta Cryst.* **A36**, 578-581.
Rees, D. C. & Lipscomb, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277-280.
Srinivasan, R. & Parthasarathy, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

Acta Cryst. (1988). A**44**, 144-150

Use of the Information on Electron Density Distribution in Macromolecules

By V. Yu. Lunin

Research Computing Centre, USSR Academy of Sciences, Pushchino, Moscow Region, 142292 *USSR*

(*Received* 29 *May* 1987; *accepted* 1 *October* 1987)

Abstract

A new type of information on the distribution of electron density in crystals of biological macromolecules is proposed. This is a quasihistogram of the image of the function of electron density distribution at a finite resolution. It is shown how this information should be used to restore the values of low-angle structure factors whose amplitudes have not been measured during X-ray experiments.

Introduction

X-ray analysis of the spatial organization of macromolecules implies searching for a function $\rho(\mathbf{r})$ which is the sum of truncated Fourier series

$$\rho(\mathbf{r}) = V^{-1} \sum_{|\mathbf{s}|\le 1/d_{\min}} F(\mathbf{s}) \exp i[\varphi(\mathbf{s}) - 2\pi(\mathbf{s},\mathbf{r})]. \quad (1)$$

We shall call this function the 'image', or more precisely 'image of the function of the electron density

0108-7673/88/020144-07$03.00

distribution at a resolution $d_{\min}$'. In contrast to the 'true' electron density distribution corresponding to the infinite series in (1), the properties of image $\rho(\mathbf{r})$ may strongly depend on the resolution that is produced. For example, the sum (1) may take negative values, several atoms or atomic groups may not be separated and so on. To obtain an accurate image, we should use exact values of all the structure factors needed to calculate (1). In practice there is always something that prevents this, either due to unknown phases or even to unknown phases and amplitudes for some of the structure factors. This lack of necessary information may produce maps with practically indiscernible molecule configurations. An example of such a map is shown in Fig. 1(*b*) where about 17% of terms are excluded from the sum (1).

If we know amplitudes and phases only for some of the structure factors needed we can set the unknown remainder arbitrary to obtain 'allowable' images, those which do not contradict the information we possess on structure factors. Let L be a set of all these allowable images. There are several approaches to determining the unknown structure factors, *i.e.* to choosing an image in the class L. The most common way is to make the unknown factors zero so that the sum (1) does not include them at all (Fig. 1*b*). A more accurate choice requires additional restrictions on the function $\rho(\mathbf{r})$ which would allow for its specific properties.

This paper is an attempt to account accurately for the specificity of the range of values of images $\rho(\mathbf{r})$ corresponding to the distributions of electron densities in macromolecular crystals. Some examples of the images obtained are shown in Figs. 1(*c*) and (*d*).

1. Histogram corresponding to image $\rho(\mathbf{r})$ as a kind of additional information on the object

Assume $\rho(\mathbf{r})$ is calculated at points of a unit-cell grid and $\{\rho_j\}$ are its values at these points. Let us construct a histogram for these values. For this purpose we

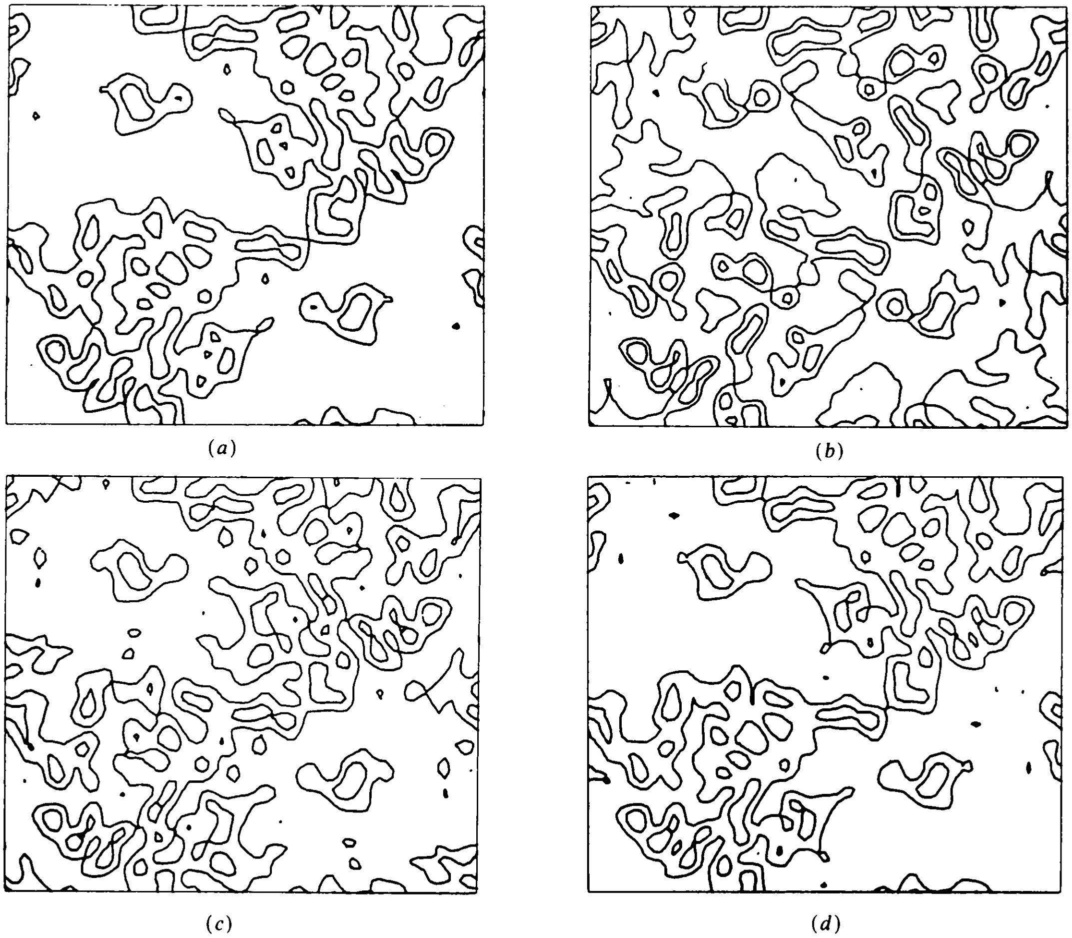

Fig. 1. (*a*) Section $z=12/48$ of image $\rho(\mathbf{r})$ at resolution 4 Å calculated from the subtilisin model; (*b*) the same section for the sum (1) with 17% of terms excluded; (*c*) restored image without account of the information on the amplitudes of unknown structure factors; (*d*) restored image with account of the information on the amplitudes of unknown structure factors. The lower contour level in all maps isolates 40% of unit-cell volume.

subdivide the interval $(\rho_{\min}, \rho_{\max})$ into K equal parts (bins) and determine how frequent are occurrences of ρ_j in each of the bins:

$$\tau_k = n_k \Big/ \sum_i n_i, \qquad k = 1, \ldots, K.$$

Here n_k is the number of ρ_j in the bin k, which is a number of such ρ_j that

$$\rho_{\min} + (k-1)\frac{\rho_{\max} - \rho_{\min}}{K} \leq \rho_j < \rho_{\min} + k\frac{\rho_{\max} - \rho_{\min}}{K}.$$

Analysis of histograms for images $\rho(\mathbf{r})$ corresponding to proteins with known structures suggests that they have characteristic features distinguishing them from histograms of randomly chosen functions. This fact is illustrated by Fig. 2 showing histograms for the following three functions of form (1) at $d_{\min} = 4$ Å:

$\rho_e(\mathbf{r})$	is determined by (1), where $F(\mathbf{s})\exp[i\varphi(\mathbf{s})]$ are exact structure factors found from an atomic model of subtilisin;
$\rho_{sp}(\mathbf{r})$	$F(\mathbf{s})$ in (1) are exact values from the subtilisin model and $\varphi(\mathbf{s})$ are determined by a randomizer;
$\rho_r(\mathbf{r})$	about 17% of reflections are eliminated from the synthesis $\rho_e(\mathbf{r})$.

Histograms for exact images $\rho_e(\mathbf{r})$ may have some dissimilarities for different proteins or different resolutions, so that determination of the 'standard' histogram $\{\tau_k^0\}$ for a still unknown object is a separate problem. In this paper we want to confine ourselves to an important particular case when the standard histogram is known. This is the case when a homologous protein with known structure exists. Its histogram may serve as the 'standard'.

Assume now that we know this standard histogram, $\{\tau_k^0\}$, for the object we are studying. Let us use this information to choose a function in the above class L. Let $\rho^c(\mathbf{r})$ be a function of this class. We can calculate the histogram $\{\tau_k^c\}$ for this function and compare it with the standard.

$$Q(\rho^c) = K^{-1}\sum_{k=1}^{K}(\tau_k^c - \tau_k^0)^2/\tau_k^0.$$

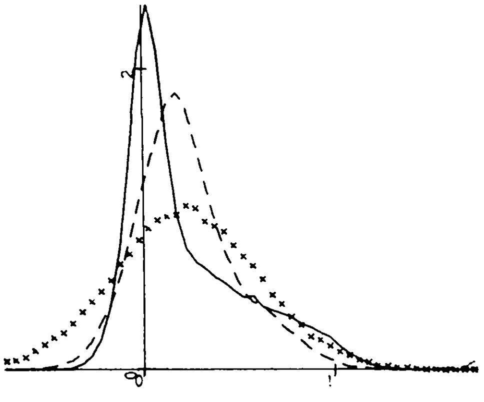

Fig. 2. Histograms for functions $\rho(\mathbf{r})$(——), $\rho_{\mathrm{sp}}(\mathbf{r})$ (× × ×), $\rho_r(\mathbf{r})$ (----).

Let us now formulate the criterion for choosing the function $\rho(\mathbf{r})$ in the class L as follows:

Criterion 1. We seek a function $\rho(\mathbf{r})$ of the class L for which the value $Q(\rho)$ is minimal.

This means that the unknown structure factors are allowed to take any (not necessarily zero) values, but it is required that the histogram of $\rho(\mathbf{r})$ should be as close to the standard as possible.

The result of application of this type of criterion is shown in Figs. 1(*c*) and (*d*).

2. Computational statement of the problem. Quasihistograms

2.1. *General formulation of the problem*

Assume that for a set of reflections S_d we know amplitudes $F^o(\mathbf{s})$ and phases $\varphi^o(\mathbf{s})$ of structure factors and that for a set of reflections S_u we do not know either phases or amplitudes and phases. Then, to choose a concrete image in the class L, we need to determine the values

$$F^c(\mathbf{s})\exp[i\varphi^c(\mathbf{s})] \quad \text{for } \mathbf{s} \in S_u.$$

Following the above considerations, we can try to find them so as to make minimal the function Q where $\{\tau_k^0\}$ are the values prescribed and $\{\tau_k^c\}$ are calculated as follows:

(*a*) for a unit-cell grid we calculate the values of $\rho^c(\mathbf{r})$

$$\begin{aligned}\rho_j^c &= \rho^c(\mathbf{r}_j)\\ &= V^{-1}\{\sum_{\mathbf{s}\in S_d} F^o(\mathbf{s})\exp i[\varphi^o(\mathbf{s}) - 2\pi(\mathbf{s}, \mathbf{r}_j)]\\ &\quad + \sum_{\mathbf{s}\in S_u} F^c(\mathbf{s})\exp i[\varphi^c(\mathbf{s}) - 2\pi(\mathbf{s}, \mathbf{r}_j)]\};\end{aligned} \tag{2}$$

(*b*) we determine how frequent are occurrences of ρ_j^c in the given bins:

$$\tau_k^c = N^{-1}\sum_{j=1}^{N}\mu^0(\rho_j^c - t_k), \tag{3}$$

where N is the total number of grid points, t_k is the middle of the kth bin $[t_k = \rho_{\min} + (k - 0{\cdot}5)\Delta]$, Δ is the length of bin $[\Delta = (\rho_{\max} - \rho_{\min})/K]$, and

$$\mu^0(t) = \begin{cases} 1 & \text{if } |t| \leq \Delta/2 \\ 0 & \text{if } |t| < \Delta/2. \end{cases}$$

Minimization of such a function Q is a hard computational problem. This stems from the fact that Q is a step function such that small variations in the structure factors do not, generally speaking, allow ρ_j to 'pass over' from one bin into another. Stated differently, frequencies τ_k remain unchanged. Hence, no gradient methods can be used to minimize Q (the gradient of Q is zero for almost all values of variables). To cope with this task we shall reformulate the problem, introducing quasifrequencies and quasihistograms.

2.2. *Quasifrequencies and quasihistograms*

'Bad' properties of Q result from the fact that the frequencies τ_k are calculated in (3) using the step function $\mu^0(t)$.

Definition. Let $\mu(t)$ be an arbitrary function such that $\mu(t) \geq 0$ and $\int_{-\infty}^{\infty}\mu(t)\,\mathrm{d}t = \Delta$. By quasifrequencies we mean the values

$$\hat{\tau}_k = N^{-1}\sum_{j=1}^{N}\mu(\rho_j^c - t_k) \qquad (4)$$

and by a quasihistogram the set of quasifrequencies $\{\hat{\tau}_k\}$.

We introduce quasifrequencies to indicate that the contribution of a point ρ_j may actually be distributed over some bins, relating quasifrequencies $\hat{\tau}_k$ not only to the values ρ_j^c in a given bin but also to the values in the neighbouring bins. This dependence may be made smooth, which enables gradient methods to be used in order to minimize the function

$$\hat{Q} = K^{-1}\sum_{k=1}^{K}(\hat{\tau}_k^c - \hat{\tau}_k^0)^2/\hat{\tau}_k^0. \qquad (5)$$

Here $\hat{\tau}_k^0$ are the standard quasifrequencies which in our case are found from image $\rho(\mathbf{r})$ for a homologous protein and $\hat{\tau}_k^c$ are given by (4) and (2).

Now the problem of restoration of image $\rho(\mathbf{r})$ from an incomplete set of structure factors can be formulated as follows.

Problem 1. We seek a function $\rho_{\rm opt}(\mathbf{r})$ of the class L for which the value $\hat{Q}(\rho_{\rm opt})$ is minimal, *i.e.* the quasifrequencies are closest to the standard.

In this paper we shall use the simplest function $\mu(t)$ of the following form:

$$\mu(t) = \begin{cases} -(1/\beta^2\Delta)|t| + 1/\beta & \text{for } |t| < \beta\Delta \\ 0 & \text{for } |t| \geq \beta\Delta. \end{cases} \qquad (6)$$

2.3. *Treatment of noncentrosymmetric (ncs) reflections*

If the amplitude of a structure factor is unknown, this can be presented in the general form as

$$F^c(\mathbf{s})\exp[i\varphi^c(\mathbf{s})] = f_R(\mathbf{s}) + if_I(\mathbf{s}).$$

When solving Problem 1, we should find two independent free parameters for this structure factor, f_R and f_I. It is natural to take zeros as the initial values of these variables in order that the structure factor can be 'grown out' from zero.

If the amplitude of a structure factor is known and equals $F^o(\mathbf{s})$, this can be approached in the two following ways. The first is to present the structure factor in the general form as

$$F^c(\mathbf{s})\exp[i\varphi^c(\mathbf{s})] = F^o(\mathbf{s})\exp[i\varphi(\mathbf{s})]$$

so that only parameter φ is needed to solve Problem 1. This approach requires a choice of the initial values of the phase φ.

The second method consists of describing the structure factor as before by two independent parameters f_R and f_I, but a penalty must be added to this criterion for the deviation of the amplitude from the prescribed value. An example of such a penalty is

$$(F^c - F^o)^2.$$

In this case the natural starting values of f_R and f_I are zero.

Mixed tactics may be used so that for each such factor two parameters, f_R and f_I, should be varied in order to use the phases obtained as starting values in minimizing the function (5) with fixed amplitudes. But here a definite accuracy is required (see Appendix). In § 3 we give a comparison of the two approaches.

2.4. *Treatment of centrosymmetric (cs) reflections*

For cs reflections the phase may only assume one of two values distinguished by π. If neither amplitude nor phase of such a structure factor is known, it can be presented in the general form as

$$F^c(\mathbf{s})\exp[i\varphi^c(\mathbf{s})] = l(\mathbf{s})[\cos\varphi^*(\mathbf{s}) + i\sin\varphi^*(\mathbf{s})].$$

Here $\varphi^*(\mathbf{s})$ is one of two phase values allowed for the given reflection and $l(\mathbf{s})$ is an arbitrary (positive or negative) real number. In this way, a cs reflection with unknown amplitude can be described by a single parameter l with zero as the natural starting value.

If for a cs reflection the amplitude is known and equals $F^o(\mathbf{s})$, the reflection is described in the general form by one variable $\omega(\mathbf{s})$ either $+1$ or -1 so that

$$F^c(\mathbf{s})\exp[i\varphi^c(\mathbf{s})] = \omega(\mathbf{s})F^o(\mathbf{s})[\cos\varphi^*(\mathbf{s}) + i\sin\varphi^*(\mathbf{s})].$$

If the variables are discrete, the problem of minimization becomes more complicated. To avoid the difficulties that may arise, cs reflections should be described as before by the real variable $l(\mathbf{s})$, but a penalty must be added to the criterion minimized for the deviation of l from the allowable values. It may be of the form

$$[|l(\mathbf{s})| - 1]^2. \qquad (7)$$

3. Testing

The test object was an atomic model of subtilisin in a $73 \times 64 \times 48$ Å unit cell in space group $P2_12_12_1$. The

Table 1. *Results of tests on subtilisin*

	Minimization runs	$\hat{Q}_{\text{start}}$	$\hat{Q}_{\text{finish}}$	Mean phase error (°) (182 ncs reflections)	Terms with changed signs (170 cs reflections)	R*
Test 1	15	$0\cdot7\times10^{-5}$	$0\cdot5\times10^{-7}$	7	1	0·06
Test 2	10	$0\cdot3\times10^{-2}$	$0\cdot4\times10^{-5}$	37	19	0·46
	20	$0\cdot3\times10^{-2}$	$0\cdot5\times10^{-6}$	36	19	0·46
Test 3	5	$0\cdot8\times10^{-5}$	$0\cdot6\times10^{-6}$	33	19	0·00
Test 4	33	$0\cdot4\times10^{-2}$	$0\cdot9\times10^{-5}$	33	21	0·02

* $R=\sum|F^c-F^o|/\sum|F^c|$, the sum includes only those reflections for which the amplitude can be changed by minimization of $\hat{Q}$.

centre of mass was chosen so that molecules did not overlap one another. Atomic coordinates were used to calculate the values of structure factors $F^o(\mathbf{s})\exp[i\varphi^o(\mathbf{s})]$, which were regarded as exact in further tests. About 17% of the structure factors, that is 352 out of 2104, were declared unknown (set S_u), and we attempted to determine them by minimizing (5).

In order to obtain the 'standard' quasihistogram, we calculated structure factors corresponding to a protein molecule with the orientation modified relative to the starting one. These structure factors were used to determine the Fourier synthesis needed to find quasifrequencies ($\hat{\tau}_k^0$) which were further used to calculate the function (5). We subdivided an interval $(-0\cdot5, 1\cdot5)$ into 30 bins and applied the function $\mu(t)$ of form (6) with $\beta=5$. This method of calculating the quasihistogram simulated the corresponding calculation for a homologous protein.

For such a 'standard' quasihistogram, structure factors $F^o(\mathbf{s})\exp[i\varphi^o(\mathbf{s})]$ no longer correspond to the exact minimum of (5). Therefore, we checked in the first test what the deviation of structure factors determined as the solution to Problem 1 from their accurate values might be. We performed 15 runs of steepest descent using a special program. Minimization was made by varying the parameters $f_R(\mathbf{s})$ and $f_I(\mathbf{s})$ for ncs reflections, and $l(\mathbf{s})$ for cs reflections starting with the accurate values of these parameters. The results are listed in Table 1.

In the second test we attempted to restore structure factors from the set S_u without amplitudes or phases predetermined. As in the first test we varied the parameters $f_R(\mathbf{s}), f_I(\mathbf{s})$ and $l(\mathbf{s})$. The starting values were zero (Fig. 1*b*). The results are listed in Table 1, one of the sections of the synthesis with the structure factors restored is shown in Fig. 1(*c*).

In the third and fourth tests it was assumed that for $\mathbf{s}\in S_u$ only phase values are unknown but structure-factor amplitudes are known and equal $F^o(\mathbf{s})$. In the third test the starting values for the phases of ncs reflections were the results of the second test (minimization without account of the amplitudes). For cs reflections we took the terms determined in the second test; these structure factors were fixed during minimization. Only the parameters $\varphi(\mathbf{s})$ for ncs structure factors were varied. The result is shown in Fig. 1(*d*).

In the fourth test the starting-phase values were generated by a randomizer. The parameters $\varphi(\mathbf{s})$ for ncs reflections and the parameters $l(\mathbf{s})$ for cs reflections were varied. The starting values of $l(\mathbf{s})$ were zero. Penalty functions of type (7) were used to attach the values of cs structure-factor amplitudes to the prescribed ones.

The results of the third and fourth tests were compared. The mean phase discrepancy for ncs reflections was 28° and for cs reflections 18 terms reversed sign.

4. Estimation of the accuracy of isolating the regions from the images deformed

The data listed in Table 1 give a formal estimate for the accuracy of the image 'restored'. In actuality, the quality of image $\rho(\mathbf{r})$ depends on the informational 'content' we are able to elicit from the object. Electron density images $\rho(\mathbf{r})$ obtained in some way are used at different stages in the structure determination to recognize the location of the molecule, the path of the polypeptide chain, sites of side groups *etc.* In all these cases we proceed by isolating a region in the unit cell of the form

$$\Omega_\alpha=\{\mathbf{r}\colon \rho(\mathbf{r})\ge\alpha\},$$

which is composed of points r, where $\rho(\mathbf{r})\ge\alpha$, and is characterized by a certain value of level α. Hence, the quality of the approximate image $\rho^c(\mathbf{r})$ depends directly on how accurately regions Ω_α corresponding to the exact image $\rho(\mathbf{r})$ can be restored by means of the function $\rho^c(\mathbf{r})$. We introduce a criterion characterizing the accuracy of isolating the regions Ω_α as follows.

Let $\rho(\mathbf{r})$ be the accurate image (1) and $\rho^c(\mathbf{r})$ be some approximation of it. For a value of $t\in(0,1)$ we find the critical values $\alpha(t)$ and $\alpha^c(t)$ such that the regions

$$\Omega_{\alpha(t)}=\{r\colon \rho(\mathbf{r})\ge\alpha(t)\}$$

and

$$\Omega^c_{\alpha(t)}=\{\mathbf{r}\colon \rho^c(\mathbf{r})\ge\alpha^c(t)\}$$

have identical volumes equal to tV, where V is the volume of the unit cell. We can then estimate the accuracy of coincidence of the regions Ω_α and Ω_α^c by determining the volume of an area composed of points belonging to Ω_α and not to Ω_α^c. Let this volume be denoted $M(t)$. It is clear that the smaller is $M(t)$, the more Ω_α and Ω_α^c overlap.

Now let us normalize $M(t)$ as follows. Assume that the choice of the region Ω of volume tV is quite random. Then an 'average' region $\Omega_{\alpha(t)}$ will not contain a volume $(1-t)tV$, which is the degree of noncoincidence between the regions Ω and $\Omega_{\alpha(t)}$. So the relative error when isolating the region $\Omega_{\alpha(t)}$ by the function $\rho^c(\mathbf{r})$ is determined to be

$$m(t)=M(t)/[(1-t)tV].$$

In this way, the value $m(t)$ indicates how accurate is the region $\Omega_{\alpha(t)}$ restored by means of $\rho^c(\mathbf{r})$ compared with the one for which the set of points was chosen randomly.

Fig. 3 shows the accuracy of isolating the regions $\Omega_{\alpha(t)}$ with the aid of different images. Curve *A* represents the accuracy of region isolation by means of image $\rho_{sp}(\mathbf{r})$ obtained from accurate amplitudes and random phases. It is seen in Fig. 3 that this synthesis contains almost no information on the object. Curve *B* represents the isolation accuracy for image $\rho_r(\mathbf{r})$ obtained from an incomplete set of reflections. It can be seen that this synthesis is good enough to reproduce the region of higher densities, which is 10–20% of the grid points, but it is not quite sufficient to characterize broader areas. This circumstance was used by Bukvetskaya, Shishova, Andrianov & Simonov (1977) to develop a method for phase refinement based on the analysis of regions corresponding to maximal values of electron densities.

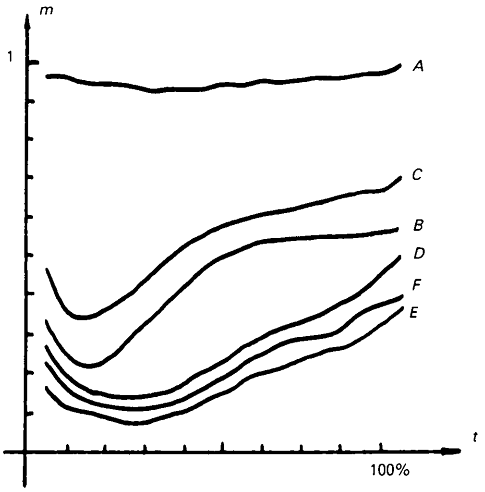

Fig. 3. Relative accuracy $m(t)$ of isolating the regions $\Omega_{\alpha(t)}$ by means of different functions (see text for explanation).

Curve *C* illustrates a synthesis which was calculated using true structure factors for set S_d and random phases and true amplitudes for S_u. It can be seen that this 'postdetermination' of unknown structure factors gives worse results than their simple removal from the synthesis (curve *B*).

Curves *D*, *E* and *F* represent the accuracy of isolating the regions $\Omega_{\alpha(t)}$ with the aid of images obtained in the second, third and fourth tests, respectively. Here we are faced with real progress.

If we compare curves *E* and *F*, we can see that although the formal errors of phase determinations in the third and the second tests are identical (see Table 1), the accuracy of isolating the regions $\Omega_{\alpha(t)}$ is higher when the technique of the third test is used.

5. Discussion

The attempt to improve electron density maps using information on the range of values of the function $\rho(\mathbf{r})$ is not new. There are various methods for restricting $\rho(\mathbf{r})\geq 0$ (see, for example, Qurashi, 1953; Sirota & Simonov, 1970; de Rango, Tsoucaris & Zelwer, 1969; Davies & Rollett, 1976; Navaza, Castellano & Tsoucaris, 1983). Attempts were made (Vainstein & Khachaturyan, 1977) to restrict the range of $\rho(\mathbf{r})$ to two values (which under normalization are 0 or 1). However, as can be seen from the histogram for the function $\rho(\mathbf{r})$ in Fig. 2, the condition $\rho(\mathbf{r})\geq 0$ may only be approximately satisfied for mean resolution. There are points at which the function $\rho_e(\mathbf{r})$ of the form (1) is negative even if the corresponding amplitudes and phases are determined absolutely accurately. The method proposed in this paper accounts more correctly for particular properties of the range of functions (1) relating to molecules. For example, it does not require that the function $\rho(\mathbf{r})$ should be nonnegative, it simply limits the volume of the region of negative values and the range of possible values. The fact that there is a large area occupied by the solvent is reflected in the histogram by a large peak near zero. The presence of a protein globule is reproduced by higher frequencies corresponding to large values of ρ compared with those appearing when the function $\rho_{sp}(\mathbf{r})$ with randomly chosen phases is analysed.

It remains to give a short mathematical interpretation of our approach. Conditions of types $\rho(\mathbf{r})\geq 0$ or $\rho(\mathbf{r})=\{0 \text{ or } 1\}$ determine only those values which the function $\rho(\mathbf{r})$ is allowed to take. Use of the histogram $\{\tau_k\}$ means that we additionally attempt to take account of the measure (Leberg function) generated by $\rho(\mathbf{r})$ for its range of values.

The author is grateful to O. M. Liginchenko for her help in preparing the manuscript.

APPENDIX

Replacement of the calculated structure factor amplitude by the experimental one

When for a reflection the exact value of the structure factor amplitude F^o is known and there is a 'calculated' value of the structure factor $F^c \exp(i\varphi^c)$, common practice is to use the 'mixed' value of the structure factor $F^o \exp(i\varphi^c)$ for further calculation.

It can be seen that the value $F^o \exp(i\varphi^c)$ is more exact than $F^c \exp(i\varphi^c)$ only when

$$F^c/F^o > 1, \text{ or when } F^c/F^o < 1 - 4\sin^2(\Delta\varphi/2).$$

Here $\Delta\varphi$ is the error in the calculation of phase φ^c.

In particular, if $\Delta\varphi > \pi/3$, then for all $F^c < F^o$ the value $F^o \exp(i\varphi^c)$ is less exact than $F^c \exp(i\varphi^c)$. This is why we must be careful in ascribing to the observed amplitude the calculated phase when F^c is smaller than F^o.

References

Bukvetskaya, L. V., Shishova, T. G., Andrianov, V. I. & Simonov, V. I. (1977). *Kristallografiya*, **22**, 494-497.

Davies, A. R. & Rollett, J. S. (1976). *Acta Cryst.* **A23**, 17-23.

Navaza, J., Castellano, E. E. & Tsoucaris, G. (1983). *Acta Cryst.* **A39**, 622-631.

Qurashi, M. M. (1953). *Acta Cryst.* **6**, 103.

Rango, C. de, Tsoucaris, G. & Zelwer, C. (1969). *Acta Cryst.* **A25**, S85.

Sirota, M. I. & Simonov, V. I. (1970). *Kristallografiya*, **15**, 681-685.

Vainstein, B. K. & Khachaturyan, A. G. (1977). *Kristallografiya*, **22**, 706-710.

Acta Cryst. (1988). **A44**, 150-157

Wide-Angle X-ray Diffraction of Materials Comprising Layer-Type Molecules

By M. Shioya and A. Takaku

Department of Textile and Polymeric Materials, Tokyo Institute of Technology, O-okayama, Meguro-ku, Tokyo, Japan

(*Received* 9 *January* 1987; *accepted* 1 *October* 1987)

Abstract

The shapes of the wide-angle X-ray reflexion profiles produced by materials comprising layer-type molecules, such as carbon fibres and pyrolytic graphites, are affected by distortions, sizes and preferential orientation of the crystallites. In the present study, the diffraction intensity distribution of layer-type materials has been deduced theoretically as a function of structural parameters and measuring direction. The reflexion profiles of carbon fibres have been simulated to investigate the effects of structural parameters on the modulation of the diffraction pattern.

1. Introduction

The crystallites in non-graphitic carbons such as carbon fibres and pyrolytic graphites are parallel stacks of individual graphite layers with no regularity of packing in mutual translations parallel to the layer. The wide-angle X-ray diffraction patterns of layer-type materials of this kind show 00*l* and *hk* reflexions, but do not exhibit general *hkl* reflexions. The *hk* reflexion profiles resulting from a random orientation of randomly stacked layers of finite size were first analysed by Warren (1941).

The *hk* reflexions are strongly asymmetric, and this peculiar peak shape is affected by preferential orientation of the crystallites. Guentert & Cvikevich (1964) have given a method to convert the *hk* reflexion profiles caused by randomly stacked layers of infinite size with preferential orientation into those with random orientation. Ruland & Tompa (1968, 1972) have expressed the *hk* reflexion profiles of randomly stacked layers as a function of the layer size and the degree of preferential orientation.

In certain layer-type carbons, there exists some degree of regularity in mutual translations of neighbouring layers (Franklin, 1951; Ruland, 1965; Fischer & Ruland, 1980). With increasing regularity, the asymmetric *hk* reflexions turn into the symmetric *hkl* reflexions (Houska & Warren, 1954). Ruland (1965) has evaluated the degree of regularity for powder samples of graphitic carbons.

This study analyses the wide-angle X-ray diffraction by finite-size crystallites with preferential orientation, comprising layer-type molecules stacked with Hosemann distortions of the second kind. The results of this study form the basis of trial-and-error evaluation of structural parameters. In the following discussion, the structure of layer-type materials is expressed by the electron density distribution. Then,

0108-7673/88/020150-08$03.00