Acta Cryst. (1993). D**49**, 90–99

Electron-Density Histograms and the Phase Problem

BY V. YU. LUNIN

Institute of Mathematical Problems of Biology, Academy of Sciences of Russia, Pushchino, Moscow Region, 142292 Russia

(*Received* 29 *April* 1992; *accepted* 2 *September* 1992)

Abstract

The specific properties of the range of possible electron-density values may serve as useful additional information for the determination and refinement of structure-factor phases. Fourier synthesis histograms (showing the spectra of frequencies of different possible values) produce the most adequate representation of these properties. The mathematical background and practical uses of these histograms are discussed. The investigation provides new information on some traditional methods of phase refinement, including density-modification procedures, maximization of Cochran's integral value and other techniques.

1. Introduction

As conventional X-ray experiments do not measure all the quantities necessary to calculate the electron-density distribution, it is necessary to find additional information on the structure under investigation to obtain the desired supplementary data. Such information may result from additional experiments on the particular structure and its modifications (*e.g.* multi-wavelength experiments or use of intensities from heavy-atom derivatives) or from general features of a class of structures (*e.g.* standard values for bond lengths or bond angles) or even from basic scientific concepts such as the atomic nature of matter. In the last two cases the additional information may usually be expressed in the form of mathematical restrictions on parameters connected with the structure studied, *e.g.* penalty functions for bond-length deviations, Sayre's equations for structure-factor values and so on. This paper gives the mathematical background for a class of methods which use the additional information contained in restrictions on the range of possible values for the electron-density distribution function.

Here we consider the cases in which we do not know all the necessary structure factors $F_s\exp(i\varphi_s)$ to calculate the Fourier synthesis of desired resolution

$$\rho(\mathbf{r}) = \frac{1}{V_{\text{cell}}} \sum_{|\mathbf{s}| \le 1/d_{\min}} F_{\mathbf{s}}\exp(i\varphi_{\mathbf{s}})\exp[-2\pi i(\mathbf{s},\mathbf{r})] \quad (1)$$

and want to find out the unknown data. These cases may occur when (i) the structure-factor modulus $F_{\mathbf{s}}$ is known, but the phase value $\varphi_{\mathbf{s}}$ is known only approximately and we wish to refine its value; (ii) $F_{\mathbf{s}}$ is known, but $\varphi_{\mathbf{s}}$ is unknown and we want to determine its value; (iii) both $F_{\mathbf{s}}$ and $\varphi_{\mathbf{s}}$ are unknown. The methods described below are suitable with slight modification for all these cases, and so, rather than considering the particular cases, we will discuss the general problem of improvement of the structure-factor set.

For the last case, *i.e.* when both $F_{\mathbf{s}}$ and $\varphi_{\mathbf{s}}$ are unknown, two remarks are worthwhile. First, in this paper we refer to the situation when a relatively small number of structure factors are unknown both in modulus and phase (*e.g.* $\{F_{\mathbf{s}}\}$ are unmeasured for experimental reasons). Second, the calculation of both $F_{\mathbf{s}}$ and $\varphi_{\mathbf{s}}$ values for a number of reflections with unknown moduli is a common practice during improvement of an electron-density map $\rho(\mathbf{r})$, taking all ρ values (at the nodes of a fine grid) as independent variables and applying some restrictions. In this case the modified map contains non-zero structure factors not only at the nominal resolution but up to the resolution of double the grid spacing.

This paper discusses the additional information arising from the specific properties of the range of possible values for electron-density Fourier syntheses (1). A number of approaches to the use of these and other properties of the density range have been suggested. The most widely used property has been non-negativity of the electron-density distributions. It was used explicitly in the derivation of the Karle–Hauptman inequalities (Karle & Hauptman, 1950) in the maximum-determinant (Tsoucaris, 1970) and other methods (Biraud, 1969; Davies & Rollett, 1976), or implicitly in maximum-entropy approaches (Gull & Daniel, 1978). The more extended restrictions such as $\rho_{\min} < \rho(\mathbf{r}) < \rho_{\max}$ (with prescribed $\rho_{\min}$ and $\rho_{\max}$ values) or $\rho(\mathbf{r}) = \{0 \text{ or } 1\}$ (Vainstein & Khachaturyan, 1977; Cannillo, Oberti & Ungaretti, 1983) were also used.

Another class of methods utilizing the knowledge of which values must be present on electron-density maps are the numerous density-modification methods (Podjarny, Bhat & Zwick, 1987) in which ρ

0907-4449/93/010090-10$06.00

values are replaced by 'more correct' values and improved phase values are calculated. The modification may be performed in accordance with the particular ρ value [*e.g.* $\rho^{\text{mod}} = 0$ if $\rho < 0$, or $\rho^{\text{mod}} = 3\rho^2 - 2\rho^3$ (Collins, Brice, La Cour & Legg, 1976) if $0 < \rho < 1$] or be based on the grid-point position in the unit cell [*e.g.* $\rho^{\text{mod}}(\mathbf{r}) = 0$ if $\mathbf{r}$ belongs to the solvent region].

A better way to reflect the constraints on the possible values of the function $\rho(\mathbf{r})$ is to specify not only which values the function can have but also how frequently each of these values appears. Several different independent approaches to using the spectra of density frequencies corresponding to protein crystals ('histograms') have been suggested in recent years (Lunin, 1986, 1988; Harrison, 1988; Luzzati, Mariani & Delacroix, 1988; Zhang & Main, 1990*a*). Such histograms contain all the information used in the structure-factor improvement methods mentioned above and therefore have all the potential abilities of these methods. Furthermore, they hold much-more-detailed information, so that utilizing these methods can give additional routes to structure-factor improvement and supplement the most commonly used methods.

In §2 we introduce strict definitions of our methods. In §3 we discuss ways of obtaining the standard histogram for the desired synthesis before all the necessary structure-factor values have been determined. This is followed in §§4, 5 and 6 by a discussion of how to exploit the information inherent in a histogram for structure-factor improvement. Computational aspects of the suggested approaches are outlined in §7, and §8 covers some practical applications.

Although in practical applications the information contained in a histogram must be used together with all the other available additional information, in this paper we shall restrict ourselves solely to histogram information in order to show how histograms alone can improve structure factors.

2. The histogram of a finite-resolution Fourier synthesis

This paper discusses the information which may be extracted from the range of available values of the electron-density distribution function. The most adequate representation of this information for a function $\rho(\mathbf{r})$ defined in a unit cell is the Lebesgue measure generated by this function. We start, however, with a more simple approach which can be easily realized in computer practice.

2.1. *Values frequency spectra*

The most direct approach to represent how frequently a function takes each of its values is as follows. Let us introduce a uniform grid in the unit cell V, and let $\{\rho_j\}_{j=1}^{N}$ be the set of values calculated at the grid points. Let us split the interval $(\rho_{\min}, \rho_{\max})$ of the values into K equal parts (bins) and for each bin determine how frequently values $\{\rho_j\}$ fall into the bin

$$\hat{\nu}_k = n_k/N. \tag{2}$$

Here n_k is the number of grid points such that the corresponding ρ_j value belongs to the kth bin, *i.e.*

$$\rho_{\min} + (j-1)\frac{\rho_{\max} - \rho_{\min}}{K} \le \rho_j \le \rho_{\min} + j\frac{\rho_{\max} - \rho_{\min}}{K}, \tag{3}$$

and N is the total number of grid points. We call the spectrum (distribution) of frequencies $\{\hat{\nu}_k\}_{k=1}^{K}$ the histogram corresponding to the function $\rho(\mathbf{r})$.

Sometimes it is more convenient to deal with the normalized frequencies

$$\nu_k = n_k/(\Delta_k N). \tag{4}$$

Here Δ_k denotes the length of the kth bin. In this case the probability of the ρ value at a randomly chosen grid point falling into the kth bin is $\nu_k\Delta_k$.

2.2. *Histograms and cumulative functions*

The frequencies ν_k and $\hat{\nu}_k$ depend, strictly speaking, not only on the $\rho(\mathbf{r})$ function, but also on the unit-cell grid and the manner in which the bins are introduced. To avoid this dependence we may introduce the Lebesgue measure on the range of $\rho(\mathbf{r})$ values, *i.e.* introduce the cumulative function

$$N(t) = \text{mes}\{\mathbf{r}:\rho(t) \le t\}/V_{\text{cell}}. \tag{5}$$

(Hereafter $\{\mathbf{r}: \checkmark\}$ denotes all the points in the unit cell satisfying the conditions $\checkmark$, mes Ω is the volume of the set Ω in the unit cell and $V_{\text{cell}} = \text{mes}\, V$ is the unit-cell volume.) We will consider further $\rho(\mathbf{r})$ functions to be represented as sums of finite Fourier series (1). In this case there is a derivative

$$\nu(t) = \mathrm{d}N(t)/\mathrm{d}t, \quad N(t) = \int_{-\infty}^{t} \nu(\tau)\mathrm{d}\tau. \tag{6}$$

Hence, an equivalent way of introducing the Lebesgue measure is to define the density of the measure $\nu(t)$.

To avoid confusion between different uses of the term 'density' we will define the density of the measure $\nu(t)$ as a histogram corresponding to the $\rho(\mathbf{r})$ function studied. This notation is not in contradiction with the previous use of the term histogram, as the normalized frequencies (4) are simply approximations of the values $\nu(t_k)$ calculated at the middle t_k of the bins

$$\nu(t_k) = \lim_{\Delta_k \to 0, N \to \infty} \nu_k. \tag{7}$$

The use of $\nu(t)$ to represent information on the range of possible values is more convenient when considering theoretical questions while in practice we have to deal with an approximate representation, *e.g.* with the frequencies $\{\nu_k\}$.

Using (4) to calculate approximate values for $\nu(t)$ has some disadvantages. The most important is that the frequencies $\{\nu_k\}$ depend on the ρ bin and not on the precise ρ value. This means that small deviations in $\{\rho_j\}$ which do not change the bins result in the same frequency values, and therefore the gradient of ν_k with respect to $\{\rho_j\}$ is equal to zero almost everywhere, so it cannot be used when improving the structure-factor set from a minimal principle (see §7.1 below). However, we can improve the situation by using another formula to calculate approximate $\nu(t)$ values. For any function $\lambda(\rho)$ and any t the following formula is valid

$$(1/V_{\text{cell}})\int_V \lambda[t-\rho(\mathbf{r})]\mathrm{d}V_{\mathbf{r}} = \int_{-\infty}^{\infty} \lambda(t-\tau)\nu(\tau)\mathrm{d}\tau, \quad (8)$$

which expresses the equality of averaging through space and with respect to the measure. As $\lambda(t)$ tends to Dirac's δ function we obtain

$$\nu(t) = (1/V_{\text{cell}}) \lim_{\lambda(t)\to\delta(t)} \int_V \lambda[t-\rho(\mathbf{r})]\mathrm{d}V_{\mathbf{r}}. \quad (9)$$

Introducing a uniform grid in the unit cell and applying the simplest numerical formula to calculate the integral we get a set of approximate formulae for $\nu(t)$ values

$$\nu(t) \simeq (1/N)\sum_{j=1}^{N} \lambda(t-\rho_j) \quad (10)$$

if $\lambda(t)$ is close enough to the δ function. Formula (4) is a particular case of (10) corresponding to

$$\lambda(t) = \begin{cases} 1/\Delta & \text{for } 0 \le t \le \Delta, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

If $\lambda(t)$ has a non-zero derivative inside the interval, the values $\nu(t)$ defined by (10) have non-zero gradients with respect to $\{\rho_j\}$. In practice we used (Lunin, 1988) a simple $\lambda(t)$ form

$$\lambda_\kappa(t) = \begin{cases} -(1/\kappa^2)|t| + 1/\kappa & \text{for } |t| \le \kappa \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

2.3. *Histograms corresponding to the molecular region*

Zhang & Main (1990*a*,*b*) suggested calculating frequencies (2) or (4) for the points belonging to the molecular region only. This has some advantages when predicting the standard histogram and improving phases in the case of known molecular boundaries (see §3.3 below). Nevertheless it is difficult to use such histograms at the early stages of structure determination when the molecular boundaries are not known reliably.

2.4. *Histograms for finite-resolution electron-density syntheses*

Fig. 1 shows a typical histogram corresponding to a middle-resolution electron-density synthesis for a protein. The asymmetry of such histograms was clearly suggested as a criterion of phase correctness by Podjarny & Yonath (1977). It must be noted that this property was used indirectly by Cochran as early as 1952 (see §6.2 below).

It is worth noting that a negative-value wing is present on the histogram corresponding to Fourier synthesis calculated with exact structure-factor values. Negative-value regions appear in the syntheses due to series-termination effects and cannot be removed without changing the resolution or distorting the phase values. Furthermore, the deepest negative 'pits' are mainly concentrated in the molecular region and may be used in the search for molecular boundaries (Urzhumtsev, Lunin & Luzyanina, 1989).

The most valuable feature of the histogram is that it depends on the level of phase error when calculating the synthesis. Fig. 2 shows histograms corresponding to different error levels. This sensitivity of histograms to phase errors and to the lack of data (Lunin, 1988) used in calculating the synthesis allows the histogram to be used as an indicator of the correctness of the phases. We discuss ways in which histograms may be used for structure-factor improvement in §4 below.

3. Histogram prediction

To use the information inherent in an electron-density histogram for phase determination or refinement we must know the 'true' histogram corre-

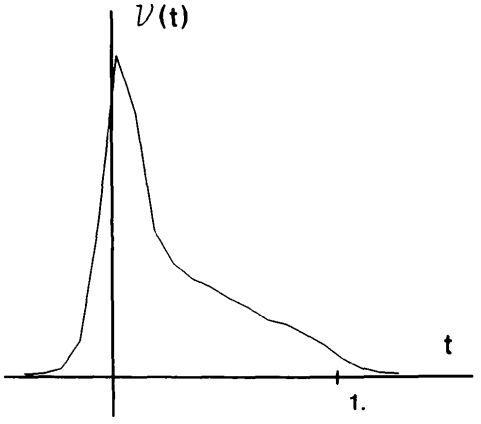

Fig. 1. A typical histogram for protein electron-density synthesis at 4 Å resolution.

sponding to the synthesis (1). As there is no direct experimental way of determining this 'true' histogram, it has to be predicted from general features of the class of maps to which the structure under study belongs. We briefly discuss below the main approaches proposed for protein electron-density histogram prediction. Two comments are relevant: (1) the histogram depends strongly on the resolution of the synthesis, so the resolution employed when using or predicting the histogram must be kept in mind; (2) the histogram also depends on the overall temperature factor.

3.1. *Gaussian model for the histogram*

Harrison (1988) suggested the use of a Gaussian curve as a model for the histogram. At first sight this approximation seems too crude and incapable of reproducing the asymmetry of the real histogram. Nevertheless, the solvent influence makes the low-resolution histograms more similar to Gaussian curves than the high-resolution histograms. Therefore, at low resolution such a histogram model may possibly be useful since it restricts density values that are too high or too low.

3.2. *Use of similar (homologous) objects to predict the histogram*

The initial concept of histogram prediction (Lunin, 1988; Luzzati, Mariani & Delacroix, 1988) was to use the histogram calculated from a similar sample. If the atomic coordinates for such a sample are known, then it is possible to calculate the corresponding structure factors, the synthesis at the desired resolution and its histogram. The most difficult question when applying this approach is 'What is a similar object when predicting the histogram?' In our experience (see §3.3 below) 'similar' in this context means that the known sample has the same fraction of the unit-cell volume per atom of the protein molecule and the unit-cell dimensions are similar to those of the crystal under investigation.

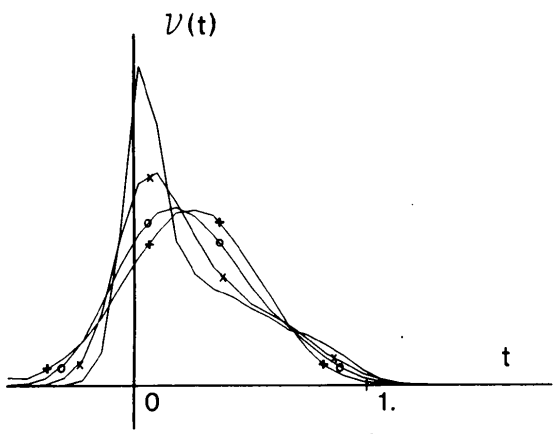

Fig. 2. Histograms corresponding to 4 Å resolution syntheses calculated with different mean phase-error levels: — exact phases; —+— −15°; —o— −36°; —×— −90°.

3.3. *Prediction of histograms for the molecular region only*

It has been established that if we restrict ourselves to the molecular region only then all histograms of protein electron-density syntheses are similar (Main, 1990*a*,*b*; Lunin & Skovoroda, 1991). (Here, of course, we mean syntheses at equal resolutions.) An empirical formula describing this histogram has been suggested by Main (1990*a*).

3.4. *The two-component histogram model*

It has been shown for proteins (Lunin & Skovoroda, 1991) that a large variety of histograms calculated for the whole unit cell may be described by the two-component formula

$$\nu(t) = (F_{000}/V_{\text{cell}})\nu^0(t) + q^0(t). \qquad (13)$$

Here $\nu^0(t)$ and $q^0(t)$, for any particular resolution, are standard distributions, similar for all the Fourier syntheses. The standard distributions $\nu^0(t)$ and $q^0(t)$ may be determined from the condition that (13) should be as accurate as possible for a set of histograms corresponding to proteins of known atomic structures. These standard distributions may be assumed as being calculated for the molecular and solvent regions separately.

4. Structure-factor improvement as a minimal principle

Let us now assume that we know the histogram $\nu^{\text{true}}(t)$ which corresponds to the synthesis (1) calculated with the true structure-factor values. We call this ideal histogram the *standard histogram*. We will discuss below how this histogram may be used for improvement of a structure-factor set.

4.1. *Structure-factor improvement from a minimal principle*

The most direct way to exploit the histogram information is to find out which unknown structure-factor values result in a Fourier synthesis histogram closest to the standard one. Let us specify some numerical criterion of histogram closeness, *e.g.*

$$Q_{\text{hist}} = \int_{-\infty}^{\infty} w(t)[\nu^{\text{calc}}(t) - \nu^{\text{true}}(t)]^2 \mathrm{d}t, \qquad (14)$$

where $w(t)$ denotes a weighting function. Then for any set of trial values for unknown structure factors we can merge trial and known values, calculate the corresponding Fourier synthesis and its histogram, and compare this histogram with the standard one

by means of (14). In this way we define the function $Q_{\rm hist}$ depending on trial structure-factor values, and the best values may be defined as ones which minimize the discrepancy (14). The set of trial values may be different in different situations. It may include phase values for some reflections or both moduli and phases for others, but the main consideration is the same, *i.e.* to minimize the discrepancy $Q_{\rm hist}$.

The advantage of such an approach is not only the simplicity of its formulation, but also the possibility of including all the other additional information represented as a minimal principle for some target function in the process of structure-factor improvement. For example, Sayre's equations (Sayre, 1952, 1972, 1974) may be taken into account together with the histogram information if we define the best values for unknown structure factors as those minimizing the compound criterion

$$Q^{\rm com} = \alpha Q_{\rm hist} + \beta \sum_{\mathbf{s}} |F_{\mathbf{s}} \exp(i\varphi_{\mathbf{s}}) - \theta(\mathbf{s}) \times \sum_{\mathbf{u}} F_{\mathbf{s}-\mathbf{u}} \exp(i\varphi_{\mathbf{s}-\mathbf{u}}) F_{\mathbf{u}} \exp(i\varphi_{\mathbf{u}})|^2. \quad (15)$$

4.2. *The use of a priori phase and histogram information*

Another possible type of additional information which can be used together with the histogram information is the probability distributions either for individual phases or for some phase invariants. There are two known ways of representing this information as a minimal principle. The first is to introduce a likelihood function (supposing mutual independence of the distributions) and maximize this function or its logarithm. Examples of this approach are minimization of the compound criterion (Lunin & Urzhumtsev, 1985)

$$Q^{\rm com} = \alpha Q_{\rm hist} + \beta \sum_{\mathbf{s}} (A_{\mathbf{s}} \cos\varphi_{\mathbf{s}} + B_{\mathbf{s}} \sin\varphi_{\mathbf{s}} + C_{\mathbf{s}} \cos 2\varphi_{\mathbf{s}} + D_{\mathbf{s}} \sin 2\varphi_{\mathbf{s}}) \quad (16)$$

where $A_{\mathbf{s}}$, $B_{\mathbf{s}}$, $C_{\mathbf{s}}$, $D_{\mathbf{s}}$ are the coefficients of Hendrickson & Lattman (1970) for the corresponding probability distribution, or the proposal of Hauptman (1989) for using phase-invariant distributions in a similar manner. [See also Bricogne & Gilmore (1990) for a more thorough analysis of likelihood-function-based approaches.]

The other approach suggested by Hašek (1974, 1984) involves (instead of trying to obtain the most probable value for every phase invariant) the use of phase-invariant probability distributions on the whole in a manner similar to that described above for Fourier synthesis histograms. The main aim of the method is to find out the phase values which minimize the discrepancy between the calculated invariant-value distribution and the theoretical one. This means that the best phases do not produce the most probable invariant values, but every invariant value appears as frequently as it is prescribed by the theoretical distribution. This results in a criterion for the minimal problem

$$Q^{\rm com} = \alpha \int_{-\infty}^{\infty} w(t)[\nu^{\rm calc}(t) - \nu^{\rm true}(t)]^2 dt + \beta \int_{-\infty}^{\infty} u(t)[\mu^{\rm calc}(t) - \mu^{\rm true}(t)]^2 dt \quad (17)$$

where $\nu(t)$ denotes the histogram for electron-density values and $\mu(t)$ the histogram for phase-invariant values.

4.3. *The use of positional information together with histogram information*

If the information on the molecular boundary is present it can be taken into account by making use of the two separate histograms for the Fourier synthesis calculated through the molecular and solvent regions (Harrison, 1988). In this case the minimized function may have the form

$$Q^{\rm com} = \alpha \int_{-\infty}^{\infty} w(t)[\nu_{\rm mol}^{\rm calc}(t) - \nu_{\rm mol}^{\rm true}(t)]^2 dt + \beta \int_{-\infty}^{\infty} u(t)[\nu_{\rm sol}^{\rm calc}(t) - \nu_{\rm sol}^{\rm true}(t)]^2 dt \quad (18)$$

where $\nu_{\rm mol}(t)$ and $\nu_{\rm sol}(t)$ are histograms describing the frequencies of Fourier synthesis values in the molecular and the solvent regions respectively.

5. A system of equations for structure factors: fixed-point representation

In this section we will deduce equations which restrict the values of structure factors of a function with the prescribed histogram $\nu^{\rm true}(t)$ and discuss ways of solving them.

5.1. *The functional equation for functions possessing the prescribed histogram*

First, we will find out the functional relationship for functions having the given histogram. To be more precise, we will show that a function $\rho(\mathbf{r})$ has the histogram $\nu^{\rm true}(t)$ if and only if the function satisfies the equation $\rho = \tau[\rho]$ where τ is the special functional transformation. The transformation is defined by the function $\nu^{\rm true}(t)$ and varies for different histograms.

Suppose the histogram $\nu^{\rm true}(t)$ is known and we look for a transformation τ: $\rho \rightarrow \rho'''$ such that for any function $\rho(\mathbf{r})$ the modified function ρ''' has the histogram $\nu^{\rm true}(t)$. There may be a number of such transformations, but if we demand additionally that the transformation should not distort the isosurfaces of

$\rho(\mathbf{r})$ [*i.e.* $\rho'''(\mathbf{r}_1) > \rho'''(\mathbf{r}_2)$ if and only if $\rho(\mathbf{r}_1) > \rho(\mathbf{r}_2)$] then (Lunin & Vernoslova, 1991) a unique transformation of this kind exists which has the form

$$\rho(\mathbf{r}) \rightarrow \rho'''(\mathbf{r}) = \lambda_\rho[\rho(\mathbf{r})]. \tag{19}$$

The modifying function $\lambda_\rho(t)$ satisfies the equation

$$N^{\text{true}}(\lambda_\rho) = N^{\text{calc}}(t), \tag{20}$$

where the cumulative functions $N^{\text{true}}(t)$ and $N^{\text{calc}}(t)$ are calculated as

$$N^{\text{obs}}(t) = \int_{-\infty}^{t} \nu^{\text{true}}(x)\mathrm{d}x,$$
$$N^{\text{calc}}(t) = \int_{-\infty}^{t} \nu^{\text{calc}}(x)\mathrm{d}x. \tag{21}$$

Here $\nu^{\text{calc}}(t)$ is the histogram corresponding to the function under modification $\rho(\mathbf{r})$. It must be emphasized that in contrast to common density modification (Podjarny, Bhat & Zwick, 1987) the modifying function $\lambda_\rho(t)$ depends on both the desired histogram $\nu^{\text{true}}(t)$ and the function $\rho(\mathbf{r})$ being modified, *i.e.* for different functions $\rho(\mathbf{r})$ the modifying functions $\lambda_\rho(t)$ are different. [It must also be noted that it is possible to consider transformations of a function $\rho(\mathbf{r})$ distorting its isosurfaces and to produce other types of equations for structure factors, but such considerations lie beyond the scope of this paper.]

It follows from (19)–(21) that if the function $\rho(\mathbf{r})$ itself has the histogram $\nu^{\text{true}}(t)$ then $\lambda_\rho(t) \equiv t$ and the transformation $\tau[\rho]$ does not change this function, *i.e.* the function satisfies the functional equation $\rho = \tau[\rho]$. Conversely, if the last equation is valid, then $\lambda_\rho(t) \equiv t$ and therefore $N^{\text{true}}(t) \equiv N^{\text{calc}}(t)$ and hence $\nu^{\text{true}}(t) \equiv \nu^{\text{calc}}(t)$. This means that the function $\rho(\mathbf{r})$ has the histogram $\nu^{\text{true}}(t)$. Thus, in order that the function $\rho(\mathbf{r})$ has the histogram $\nu^{\text{true}}(t)$, the functional equation

$$\rho(\mathbf{r}) = \tau[\rho](\mathbf{r}) \tag{22}$$

has to be satisfied.

5.2. *Histogram-specification and histogram-matching methods*

Equation (22), satisfied only by functions with the $\nu^{\text{obs}}(t)$ histogram, is equivalent to the set of corresponding equations for structure factors

$$F_{\mathbf{s}}\exp(i\varphi_{\mathbf{s}}) = \mathcal{F}_{\mathbf{s}}\{\rho'''(\mathbf{r})\}, \tag{23}$$

where $\rho'''(\mathbf{r}) = \lambda_\rho[\rho(\mathbf{r})]$ is the transformation [(19)–(21)] of the function (1) and $\mathcal{F}_{\mathbf{s}}\{\rho'''(\mathbf{r})\}$ denotes the **s**-indexed structure factor of $\rho'''(\mathbf{r})$. Here the right-hand side of every equation depends on all the structure factors of $\rho(\mathbf{r})$. [Note that the left-hand side of (23) is equal to zero for $\mathbf{s} > 1/d_{\min}$ if the histogram $\nu^{\text{true}}(t)$ corresponds to the resolution $d_{\min}$.] Assuming moduli values to be known from X-ray experiments, these equations may be considered as phase-value determination equations.

The equations (23) may be split into 'phase' and 'radial' parts

$$\varphi_{\mathbf{s}} = \arg[\mathcal{F}_{\mathbf{s}}\{\rho'''(\mathbf{r})\}], \tag{24}$$

$$|\mathcal{F}_{\mathbf{s}}\{\rho'''(\mathbf{r})\}| = F_{\mathbf{s}}. \tag{25}$$

(Here arg[z] means the phase of the complex number z.) The usual method of solution of such equations (Lunin, 1985) is to use the phase part (24) only and to solve them by the simple iteration method, *i.e.* the current phase values are used to calculate the right-hand side of (24) and the calculated values are taken as improved phases. In our case this means that the following procedure is repeated: (i) the Fourier synthesis is calculated with the current phase values; (ii) the synthesis modification [(19)–(21)] is performed; (iii) the phases of the modified function $\rho'''(\mathbf{r})$ are taken for the next cycle.

It is easy to see that this scheme coincides in its main features with the 'histogram-specification' (Harrison, 1988) and 'histogram-matching' (Zhang & Main, 1990*a*,*b*) methods of phase refinement.

It must be noted that while the complete system [(24)–(25)] is equivalent to the Fourier synthesis having the prescribed histogram $\nu^{\text{true}}(t)$, (24) alone does not have such a feature. This means that false 'self-consistent' solutions which do not result in the desired histogram may arise when solving the set of equations (24) alone. However, we have not met such a situation in our tests.

The iterative procedure is easier to implement on a computer than the minimization procedure but has a poorer convergence rate. Nevertheless, the inclusion of additional information in the iterative procedure may cause serious difficulties and necessitate complicated methods (Main, 1990*a*,*b*).

5.3. *The density-modification method*

The known histogram $\nu^{\text{true}}(t)$ provides a stricter base for numerous density-modification methods (Podjarny, Bhat & Zwick, 1987). The computational algorithm of these methods is the iterative phase-redetermination procedure described above, but with the modifying function $\lambda(t)$ not depending on the function $\rho(\mathbf{r})$ to be modified, or depending on its maximum value only. Different modifying functions have been suggested, but the two most widespread are the attenuation of negative values [$\lambda(t) = \kappa t$ for $t < 0$, where $\kappa < 1$ (Podjarny & Yonath, 1977)] and empirical transformation of the positive values [*e.g.* $\lambda(t) = 3(t/\rho_{\max})^2 - 2(t/\rho_{\max})^3$ (Hoppe & Gassmann, 1968; Collins, Brice, La Cour & Legg, 1976)].

To demonstrate the correlation with histogram-based methods the following test was performed. The 4 and 1.5 Å sets of structure factors were calculated

from atomic coordinates for cytochrome b_5 (Mathews, Levine & Argos, 1971). A series of syntheses was then calculated with the exact moduli values, but with phase values containing random errors. The modifying functions $\lambda_\rho(t)$ for the transformation [(19)–(21)] restoring the histogram $\nu^{\text{true}}(t)$ are shown in Fig. 3 for the case of the largest phase errors. The curves for lower error levels occupy intermediate positions between this transform and the identity one. For the 4 Å resolution syntheses these curves have the two main features of the density-modification methods mentioned above, *i.e.* attenuation of the lowest values and a $3\rho^2 - 2\rho^3$-type transformation for middle and high values, but the strength of these features decreases with decrease of the phase errors. For the 1.5 Å resolution syntheses these transformations set density values up to a $0.2\rho_{\max}$ value nearly equal to zero. Such a transformation is close to the '0 or 1' transformation (Cannillo, Oberti & Ungaretti, 1983) and to the transformation suggested by Shiono & Woolfson (1992). Thus 'classical' density-modification methods may be considered as those using (in an implicit form) the additional information contained in Fourier synthesis histograms, but which are less adapted to the particular error level.

6. Moments-based representation of equations for structure factors

6.1. *Histograms and moments*

Another way of exploiting the information contained in a histogram is not to use the histogram itself, but rather some of its characteristics, *e.g.* the series of central moments

$$m_k = \int_{-\infty}^{\infty} t^k \nu(t)\mathrm{d}t, \; k = 0, 1, \ldots . \qquad (26)$$

As $\nu(t)$ has non-zero values in the finite interval $(\rho_{\min}, \rho_{\max})$ only, the moment of any order has a finite value and $m_k < C(\rho_{\max} - \rho_{\min})^k$. Conversely, if the series $\{m_k\}_{k=1}^{\infty}$ is given and $m_k < C\kappa^k$ for some C and κ, then there is a unique distribution $\nu(t)$ of such moments, which may be formally represented as

$$\nu(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\left[\sum_{k=0}^{\infty}\frac{(ix)^k}{k!}m_k\right]\mathrm{d}x. \qquad (27)$$

So the definition for a function $\rho(\mathbf{r})$, the standard histogram $\nu^{\text{true}}(t)$ or the standard values $\{m_k^{\text{true}}\}$ of the moments are equivalent ways of presenting the information on ρ-value frequencies.

Instead of using (26) to calculate the statistical moments, the integration of $\rho(\mathbf{r})$ through the unit cell may be performed:

$$m_k = (1/V_{\text{cell}})\int_V \rho(\mathbf{r})^k \mathrm{d}V_{\mathbf{r}}, \quad k = 0, 1, \ldots . \qquad (28)$$

If $\rho(\mathbf{r})$ has the form of Fourier synthesis (1) and the standard values $\{m_k^{\text{true}}\}$ of the moments are known these produce a system of equations for structure-factor values, which are an equivalent representation of the property of $\rho(\mathbf{r})$ to have the histogram $\nu^{\text{true}}(t)$

$$F_{000} = m_1^{\text{true}} V_{\text{cell}}$$

$$\sum_{\mathbf{s}} F_{\mathbf{s}}^2 = m_2^{\text{true}}(V_{\text{cell}})^2$$

$$\sum_{\mathbf{s}_1+\mathbf{s}_2+\mathbf{s}_3=0} F_{\mathbf{s}_1}F_{\mathbf{s}_2}F_{\mathbf{s}_3}\exp[i(\varphi_{\mathbf{s}_1}+\varphi_{\mathbf{s}_2}+\varphi_{\mathbf{s}_3})] = m_3^{\text{true}}(V_{\text{cell}})^3$$

$$\sum_{\mathbf{s}_1+\ldots+\mathbf{s}_k=0} F_{\mathbf{s}_1}\ldots F_{\mathbf{s}_k}\exp[i(\varphi_{\mathbf{s}_1}+\ldots+\varphi_{\mathbf{s}_k})] = m_k^{\text{true}}(V_{\text{cell}})^k. \qquad (29)$$

The first two equations in (29) do not contain phase values, so the first two moments do not restrict possible phase values and cannot be used for phase determination. Nevertheless, these moments allow the F_{000} value and the scale factor for $\{F_{\mathbf{s}}\}$ values to be defined, if they were measured on a relative scale.

6.2. *Simplified approaches: Cochran's integral maximization*

A simplified way of using the standard-moment values for phase improvement is to employ a small

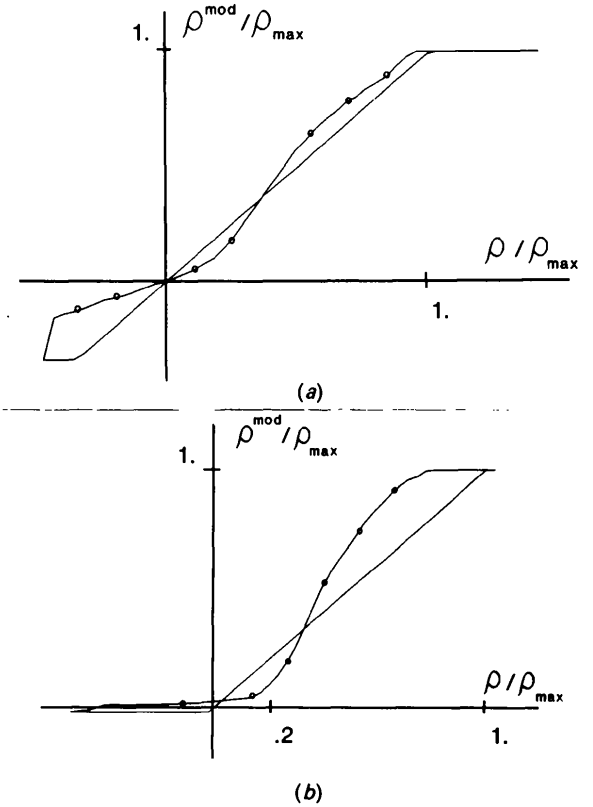

Fig. 3. Identity transform and density-modification function (—o—) for (*a*) 4 Å and (*b*) 1.5 Å resolution syntheses calculated with random phase values.

number of moments or even just one of them. For example, Podjarny & Yonath (1977) established that the quality of a phase set is correlated with the value of the third moment m_3 which is connected to the skewness of the histogram. So, they suggested the use of this value as an indicator of phase correctness. A more active approach is to maximize m_3 by changing the phase values. This method is directly related to maximization of Cochran's (1952) integral

$$(1/V_{\text{cell}})\int_V \rho(\mathbf{r})^3 \mathrm{d}V_{\mathbf{r}} = m_3. \tag{30}$$

So, maximization of the integral in (30) may be considered as maximization of the skewness of the electron-density histogram. It must be noted that for arbitrary images the assumption of maximum skewness is not generally valid. Hence a more careful approach is to fit the calculated m_3 value to the standard one.

Another approach based on the single fourth central moment μ_4

$$\mu_4 = \int_{-\infty}^{\infty} (t-m_1)^4 \nu(t)\mathrm{d}t = m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4 \tag{31}$$

was suggested by Luzzati, Mariani & Delacroix (1988). They suggested the determination of the standard μ_4^{true} value from compounds with known structures and then the use of the similarity of μ_4^{calc} and μ_4^{true} as the indicator of correctness of the phase-problem solution.

7. Computational aspects

Here we briefly discuss some computational problems which arise when solving a minimal problem.

7.1. *Quasi-histograms*

For practical purposes the criterion (14) takes the form

$$Q_{\text{hist}} = \sum_{k=1}^{K} w_k(\nu_k^{\text{calc}} - \nu_k^{\text{true}})^2, \tag{32}$$

where $\{\nu_k\}$ are approximations for $\nu(t)$ values at the points t_k (bin middles). We assume that (10) is used to calculate $\{\nu_k^{\text{calc}}\}$ because of the reasons mentioned in §2.2.

The disagreement between ν^{calc} and ν^{true} values has three sources. The first is the errors in structure factors and the purpose of Q_{hist} minimization is to reduce these errors. The second source of disagreement is the replacement of the integral in (14) by the sum in (32). This difference is, however, small if the bin lengths are small. The third reason is the use of approximation (10) instead of (9). The accuracy of this approximation depends both on the grid in the unit cell for calculation of $\{\rho_j\}$ and on the closeness of $\lambda(t)$ to the δ function. If $\lambda(t)$ is far from $\delta(t)$ the difference may be large, but it can be eliminated if we change from a comparison of histograms to a comparison of quasi-histograms (Lunin, 1988)

$$Q_{\text{qh}} = \sum_{k=1}^{K} w_k(\mu_k^{\text{calc}} - \mu_k^{\text{true}})^2 \Longrightarrow \min. \tag{33}$$

Here $\{\mu_k^{\text{calc}}\}$ are calculated in accordance with (10) and $\{\mu_k^{\text{true}}\}$ are the values at the bin middles of the function

$$\mu^{\text{true}}(t) = \int_{-\infty}^{\infty} \lambda(t-\tau)\nu^{\text{true}}(\tau)\mathrm{d}\tau \tag{34}$$

calculated at the bin middles. This means that to check the structure-factor set we do not compare the histograms as they are, but rather some of their modifications.

7.2. *Fast differentiation algorithm*

It might seem that there could be significant difficulties in calculating the gradient for a complicated function depending on a large number of variables. However, a routine method exists (Kim, Nesterov & Cherkassky, 1984; Lunin & Urzhumtsev, 1985; Lunin, 1985) to develop for any function an algorithm allowing the calculation of all the components of the gradient in about the same time as a single-value calculation would take. Thus the only problem in local minimization is to find a fast algorithm for function-value calculation.

7.3. *The multi-minima problem*

The central questions in a minimal-problem solution are those connected with the starting-point choice and multi-minima functions. The minimal problems mentioned above are not unusual. If we have approximate phase values they may serve as a starting point for refinement, but if we determine unknown phases the result may depend strongly on the starting point. Ways of choosing starting values for centric and acentric reflections have been discussed previously (Lunin, 1988).

The multi-minima nature of the minimized function results in two problems. The first, namely the convergence to 'high' local minima, may sometimes be overcome by the use of the simulated-annealing method (Brünger, 1988) or by a descent from different starting points. A more difficult question arises when several very similar deep minima exist. As we compare calculated values with values containing some experimental errors or use some approximations when deducing theoretical relations, the deepest minima may not correspond to the true solution. So the problem is not only that of finding the global minimum but also of searching for the additional criteria needed to pick out the true solution from the admissible ones.

Our tests with phase refinement (Lunin & Vernoslova, 1991) have shown that if one starts from phases containing large errors it is easy to produce a new phase set resulting in a very 'good' histogram that is still incorrect. Similar results were obtained when discriminating random phase sets in accordance with the corresponding histograms (Lunin, Urzhumtsev & Skovoroda, 1990). It was established that the phase sets resulting in the true histogram may be incorrect. Therefore, the correct histogram is not the sole criterion guaranteeing the correct structure-factor value and some additional information is necessary.

8. Applications

We mention here briefly some examples of the use of the histogram technique in the structure-factor improvement problem. More detailed descriptions can be found in the original papers.

8.1. *Retrieval of both modulus and phase values*

Sometimes part of a low-resolution structure-factor modulus is not measured in the X-ray experiment. Elimination of such reflections from calculations may produce significant distortions in low- and middle-resolution syntheses (Urzhumtsev, 1991). The method suggested (Lunin, 1988) for restoring both moduli and phases for such reflections was the minimization of the criterion (14) with respect to the unknown structure-factor values. The method was tested on subtilisin and used in studies of γ-crystallin IIIb (Lunin & Skovoroda, 1991).

8.2. *Phase refinement and extension*

A test phase refinement for 2Zn pig insulin in the resolution range 3.0–1.5 Å and a comparison with other methods was performed by Zhang & Main (1990*a*,*b*). A test phase refinement for cytochrome b_5 is described by Lunin & Vernoslova (1991). Harrison (1988) has described the use of the histogram technique for phase refinement and extension for an asparaginase from *Acinetobacter glutamininasificans* at 4.2–3.2 Å resolution.

8.3. *Direct phasing of low-resolution reflections*

Two methods for direct phasing of a limited number of lowest resolution reflections have been proposed. Luzzati, Mariani & Delacroix (1988), when investigating lipid-containing systems, suggested consideration of all possible phase sets (for centrosymmetric structures) and selection of the best set in accordance with its fourth histogram moment. A different approach was suggested by Lunin, Urzhumtsev & Skovoroda (1990). It consists of generating a large number of random phase sets, selecting the admissible ones in accordance with their histograms, performing a cluster analysis to separate admissible variants into clusters and averaging the variants in every cluster to produce a few possible phase-problem solutions. This approach was tested with an artificial structure and the structures of cytochrome b_5 and Bence–Jones protein (Lunin, 1991), and was used in work on the elongation factor *G*.

Fig. 4 shows an example of the application of this approach to the experimental data for RNAse *Sa* kindly supplied by Mrs E. Dodson. The atomic model obtained by J. Sěvcik was used for verification of the results. The protein crystallized in space group $P2_12_12_1$ with $a = 64.9$, $b = 78.32$, $c = 38.79$ Å, and the asymmetric unit contained two RNAse molecules. In this test 39 experimentally measured structure-factor modulus values for RNAse corresponding to 16 Å resolution were used together with the histogram calculated using the atomic model of cytochrome b_5. About 100 000 random phase sets were generated and about 100 of these which resulted in good histograms were averaged to produce a Fourier synthesis. Both molecules contained in the asymmetric part of the unit cell can be clearly located in this synthesis.

9. Concluding remarks

The electron-density synthesis histogram carries the fullest information on the frequencies of possible density values. Therefore the use of such histograms is a further development of methods exploiting properties such as positivity, double-side boundedness, restriction of the density value (*e.g.* 0 or 1) and so on.

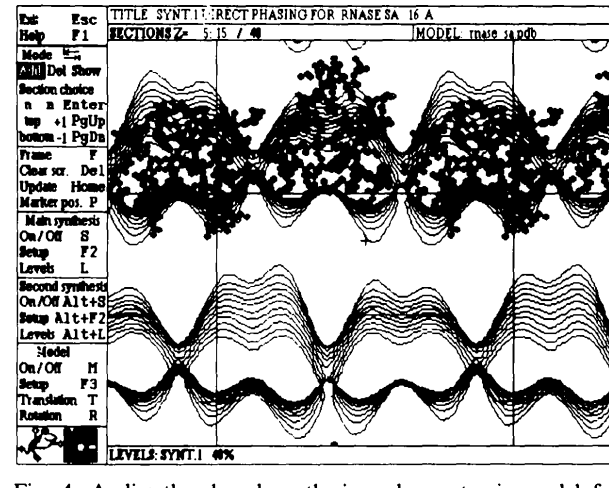

Fig. 4. A directly phased synthesis and an atomic model for RNAse *Sa*. Only atoms in the asymmetric part of the unit cell are shown.

The use of histograms as a limiting factor in the phase-determination process sheds new light on some of the most widespread methods developed semi-empirically and provides a strong mathematical background for them.

The use of histogram-closeness criteria [such as (14) or (17)] for some values connected with the structure under study instead of the request that every value should be as close as possible to the most probable one, produces a more statistically reliable distribution pattern.

Test calculations have shown that the histograms are sensitive to errors in structure-factor values, so that some incorrect phase sets may be discriminated against due to their poor histograms. However, a true histogram used on its own does not guarantee the correctness of phase values. Therefore, histograms do not allow the unambiguous determination of phases and further investigations of phase sets resulting in good histograms are necessary. It may be noted that a histogram gives a one-dimensional curve as additional information and this is, of course, inadequate information for producing a three-dimensional set of structure-factor phases immediately. These remarks do not discredit histograms as a tool for structure-factor improvement, but set reasonable boundaries for their use.

References

BIRAUD, Y. (1969). *Astron. Astrophys.* **1**, 124–127.

BRICOGNE, G. & GILMORE, C. J. (1990). *Acta Cryst.* A**46**, 284–297.

BRÜNGER, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

CANNILLO, E., OBERTI, R. & UNGARETTI, L. (1983). *Acta Cryst.* A**39**, 68–74.

COCHRAN, W. (1952). *Acta Cryst.* **5**, 65–67.

COLLINS, D. M., BRICE, M. D., LA COUR, T. F. M. & LEGG, M. J. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, K. HUML & B. SEDLACEK, pp. 330–335. Copenhagen: Munksgaard.

DAVIES, A. R. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 17–23.

GULL, S. F. & DANIELL, G. J. (1978). *Nature (London)*, **272**, 686–690.

HARRISON, R. W. (1988). *J. Appl. Cryst.* **21**, 949–952.

HAŠEK, J. (1974). *Acta Cryst.* A**30**, 576–579.

HAŠEK, J. (1984). *Acta Cryst.* A**40**, 338–352.

HAUPTMAN, H. (1989). XII Eur. Crystallogr. Meet., Moscow, Collected Abstracts, Vol. 1, p. 24.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–107.

KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.

KIM, K. V., NESTEROV, YU. E. & CHERKASSKY, B. V. (1984). *Dokl. Akad. Nauk SSSR*, **275**, 1306–1309.

LUNIN, V. YU. (1985). *Acta Cryst.* A**41**, 551–556.

LUNIN, V. YU. (1986). *Use of the Information on Electron Density Distribution in Proteins*. Preprint. Pushchino, Russia.

LUNIN, V. YU. (1988). *Acta Cryst.* A**44**, 144–150.

LUNIN, V. YU. (1991). *Use of the Electron-Density Synthesis Histograms for Solving of the Phase Problem in Protein Crystallography*. Preprint. Pushchino, Russia.

LUNIN, V. YU. & SKOVORODA, T. P. (1991). *Acta Cryst.* A**47**, 45–52.

LUNIN, V. YU. & URZHUMTSEV, A. G. (1985). *Acta Cryst.* A**41**, 327–333.

LUNIN, V. YU., URZHUMTSEV, A. G. & SKOVORODA, T. P. (1990). *Acta Cryst.* A**46**, 540–544.

LUNIN, V. YU. & VERNOSLOVA, E. A. (1991). *Acta Cryst.* A**47**, 238–243.

LUZZATI, V., MARIANI, P. & DELACROIX, H. (1988). *Makromol. Chem. Macromol. Symp.* **15**, 1–17.

MAIN, P. (1990*a*). *Acta Cryst.* A**46**, 372–377.

MAIN, P. (1990*b*). *Acta Cryst.* A**46**, 507–509.

MATHEWS, F. S., LEVINE, M. & ARGOS, P. (1971). *Nature (London) New Biol.* **233**, 15–16.

PODJARNY, A. D., BHAT, T. N. & ZWICK, M. (1987). *Ann. Rev. Biophys. Chem.* **16**, 351–373.

PODJARNY, A. D. & YONATH, A. (1977). *Acta Cryst.* A**33**, 655–661.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.

SAYRE, D. (1972). *Acta Cryst.* A**28**, 210–212.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.

SHIONO, M. & WOOLFSON, M. M. (1992). *Acta Cryst.* A**48**, 451–456.

TSOUCARIS, G. (1970). *Acta Cryst.* A**26**, 492–499.

URZHUMTSEV, A. G. (1991). *Acta Cryst.* A**47**, 794–801.

URZHUMTSEV, A. G., LUNIN, V. YU. & LUZYANINA, T. B. (1989). *Acta Cryst.* A**45**, 34–39.

VAINSTEIN, B. K. & KHACHATURYAN, A. G. (1977). *Kristallografiya*, **22**, 706–710.

ZHANG, K. Y. J. & MAIN, P. (1990*a*). *Acta Cryst.* A**46**, 41–46.

ZHANG, K. Y. J. & MAIN, P. (1990*b*). *Acta Cryst.* A**46**, 377–381.