

## New Techniques for the Determination of the Scale and Overall Isotropic Temperature Factors

BY S. PARTHASARATHY\* AND K. SEKAR†

*Department of Crystallography and Biophysics,‡ University of Madras, Guindy Campus, Madras 600 025, India*

(Received 1 April 1992; accepted 11 September 1992)

### Abstract

Intensity data collected by various methods constitute the raw material from which crystal structures are derived. It is necessary to put these intensities on an absolute basis and obtain an estimate of the overall isotropic temperature factor. New techniques have been developed to estimate the scale and temperature factors using the  $\langle I_{\text{calc}} \rangle$  data derived from another known structure whose molecule is similar to the given compound. Theoretical results have been tested in a few cases. It is found that the present methods generally lead to better estimates for  $k$  and  $B$  than the traditional Wilson-plot method.

### 1. Introduction

Intensity data collected by various methods constitute the raw material from which crystal structures are derived. The raw data are first corrected for Lorentz and polarization effects and absorption to obtain a set of intensities on the same relative scale. It is then necessary to put these intensities on an absolute basis and obtain an estimate of the overall isotropic temperature factor ( $B$ , say). This is generally done by the well known Wilson-plot method (Wilson, 1942), which requires *a priori* the information regarding the unit-cell contents (*i.e.* number and types of atoms in the unit cell). In organic and biological molecules, atoms are bonded at specific bond distances and angles. If this additional information is introduced in some reasonable way into the scaling procedure, we may expect to get values of the scale factor  $k$  and the temperature factor that are more accurate than those that could be obtained from the traditional Wilson plot. Standard direct-methods programs [*e.g.* MULTAN80 (Main *et al.*, 1980)] make use of the theoretical formula for molecular scattering in the scaling procedure. In this paper, we shall describe other methods of incorporating *a priori* structural information available from crystals composed of similar molecules in either the Cambridge Structural

Database (Allen, Kennard & Taylor, 1983) or the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977) to arrive at values of  $k$  and  $B$  that are more accurate than those that could be obtained by the traditional Wilson-plot method. We shall also compare the results obtained in actual cases using our new techniques with those obtained from MULTAN80 to assess their utility.

It should be noted that the method proposed here is applicable only to molecular crystals. Furthermore, the molecules in the given crystal and in the model§ should have comparable general features.

### 2. Notation

- $B$  = overall isotropic temperature factor for the given crystal
- $f$  = atomic scattering factor of a stationary atom
- $g$  = atomic scattering factor of an atom undergoing thermal vibration
- $I_a$  = absolute intensity of a reflection  $H (= hkl)$  for the given crystal
- $[\bar{I}_a]_i$  = local average value of  $I_a$  [*i.e.* average value of  $I_a$  for reflections within the  $i$ th  $(\sin\theta)/\lambda$  interval]
- $I_r$  = relative intensity of the reflection  $H$  for the given crystal
- $[\bar{I}_r]_i$  = local average value of  $I_r$  in the  $i$ th  $(\sin\theta)/\lambda$  interval
- $I'$  = absolute intensity of the reflection  $H$  for a model
- $[\bar{I}']_i$  = local average value of  $I'$  in the  $i$ th  $(\sin\theta)/\lambda$  interval
- $k$  = scale factor necessary to make  $I_r$  equal to  $I_a$
- $N$  = number of atoms in the unit cell of the given crystal
- $n$  = number of atoms in a single molecule of the given crystal

§ In crystallography, the term 'model' generally means a trial structure whose phases are used as an approximation to the phases of the given crystal in the structure-determination process. In this paper, we shall use the term model to denote a structure whose local average intensity data are closely related to those of the given crystal so that the information content of these data could be used for scaling the relative intensity data of the given crystal.

\* To whom correspondence should be addressed.

† Present address: Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

‡ DCB contribution no. 801.

- $P$  = group of extra atoms per unit cell in the model relative to the given crystal or *vice versa*  
 $p$  = number of extra atoms in the molecule of the model relative to the molecule in the given crystal and *vice versa*  
 $p_1 = 1/k$   
 $p_2 = 2B$   
 $Q$  = group of atoms per unit cell common to the given crystal and the model  
 $q$  = number of common atoms in the molecules of the given crystal and the model  
 $R_i$  =  $i$ th interval of  $(\sin\theta)/\lambda$ , namely,  $s_i \leq s \leq s_{i+1}$ ,  $i = 1$  to  $\nu$   
 $s = (\sin\theta)/\lambda$   
 $\bar{s}_i$  = mean value of  $(\sin\theta)/\lambda$  for reflections in the interval  $R_i$   
 $x_i = [\bar{I}']_i$   
 $y_i = [\bar{I}_r]_i$   
 $\delta$  = lattice-multiplicity factor for the given crystal\*  
 $\varepsilon = \delta\eta\zeta$   
 $\eta$  = point-group multiplicity factor for the given crystal†  
 $\zeta$  = number of molecules in the asymmetric unit of the given crystal,  
 $\nu$  = number of  $(\sin\theta)/\lambda$  intervals into which the intensity data of the given crystal is partitioned  
 $\pi_\nu$  = a partition of the reflections into  $\nu$  groups based on the  $(\sin\theta)/\lambda$  values  
 $\rho = \varepsilon/\varepsilon'$   
 $\sigma$  = sum of the squares of the atomic scattering factors of all the atoms in a single molecule of the given crystal,

$$\sigma_p^0 = \sum_{j=1}^p f_j^2, \quad \sigma_p = \sum_{j=1}^p g_j^2,$$

$$\sigma_q^0 = \sum_{j=1}^q f_j^2, \quad \sigma_q = \sum_{j=1}^q g_j^2.$$

We shall use the symbols  $N'$ ,  $n'$ ,  $\delta'$ ,  $\varepsilon'$ ,  $\zeta'$  and  $\sigma'$  to denote the quantities pertaining to the model and these are respectively the analogues of  $N$ ,  $n$ ,  $\delta$ ,  $\varepsilon$ ,  $\zeta$  and  $\sigma$  of the given crystal. We shall freely use the above symbols in the subsequent sections without further definition. We shall take all atoms in the model to be stationary and all atoms in the given crystal to have the same isotropic thermal parameter  $B$ .

### 3. Derivation of the basic equation for the determination of scale and temperature factors

Suppose that the given crystal has a lattice-multiplicity factor  $\delta$ , point-group multiplicity factor  $\eta$  and

$\zeta$  molecules of the same structural formula in the asymmetric unit. Let  $n$  be the number of atoms in each molecule of the given crystal. Let  $S_w = 1/a_{\min}$ , where  $a_{\min}$  is the minimum of the cell parameters  $a$ ,  $b$ ,  $c$  of the given crystal. Let  $S_{\max}$  be the maximum value of  $(\sin\theta)/\lambda$  for the observed reflections for the given crystal. Let  $I_a$  be the absolute intensity and  $I_r$  the relative intensity of a reflection of the given crystal. We shall assume that the observed intensity data of the given crystal are in the same relative scale, so that we can write

$$I_a = kI_r, \quad (1)$$

where  $k$  is the scale factor needed to convert the relative intensities of the given crystal into the absolute scale. The atomic scattering factor  $g$  of an isotropically vibrating atom is related to the scattering factor  $f$  of a stationary atom by

$$g = f \exp(-B \sin^2\theta/\lambda^2) = f \exp(-Bs^2). \quad (2)$$

Suppose that we have a crystal of another compound whose molecules have features similar to the molecules of the given crystal\* so that it can be chosen as a model. Let  $\zeta'$  be the number of molecules in the asymmetric unit,  $n'$  be the number of atoms in each molecule,  $\delta'$  be the lattice-multiplicity factor and  $\eta'$  be the point-group-multiplicity factor for the model. Let  $S'_w = 1/a'_{\min}$ , where  $a'_{\min}$  is the minimum value of the cell parameters  $a'$ ,  $b'$ ,  $c'$  of the model. We shall assume that the structure of the model is known completely so that we can compute its intensity data assuming the atoms to be stationary.

Let  $S_w^0 = \max\{S_w, S'_w\}$ . Suppose that the reflections of the given crystal in the interval  $S_w^0 < s < S_{\max}$  are partitioned into  $\nu$  groups based on their  $(\sin\theta)/\lambda$  values by the points  $s = s_i$ ,  $i = 1, 2, \dots, \nu + 1$ , where  $s_1 = S_w^0$  and  $s_{\nu+1} = S_{\max}$ . We use the symbol  $\pi_\nu$  to denote the set of these  $\nu$  subintervals of  $(\sin\theta)/\lambda$ . The number of reflections in each  $(\sin\theta)/\lambda$  interval for the given crystal could in general be different from that of the model in the corresponding interval.

The local average values of  $I_a$  and  $I_r$  corresponding to the  $i$ th  $(\sin\theta)/\lambda$  interval may be written as (Wilson, 1942; Rogers, 1965)

$$[\bar{I}_a]_i = \delta\eta\zeta\sigma = \varepsilon\sigma_i, \quad [\bar{I}']_i = \delta'\eta'\zeta'\sigma' = \varepsilon'\sigma'_i, \quad (3)$$

where  $\varepsilon$ ,  $\varepsilon'$  are defined to be

$$\varepsilon = \delta\eta\zeta, \quad \varepsilon' = \delta'\eta'\zeta'. \quad (4)$$

\*  $\delta$  (or  $\delta'$ ) is 1 for a  $P$  lattice, 2 for an end-centred or  $I$  lattice and 4 for an  $F$  lattice.

† This is the total multiplicity arising from the point-group symmetry operations of the given crystal, e.g.  $\eta = 4$  for a crystal of space group  $I222$ .

\* Examples: Cambridge Structural Database version 3.1 (1988) contains information of 69 691 compounds. A number of crystal structures of steroids are already in the data file. If a new steroid is crystallized then it would not be difficult to choose an appropriate steroid from the data file to constitute a 'model' for scaling purposes. In macromolecular crystallography, the native protein may be chosen as the 'model' to scale the data of its derivatives. Insulin of one species may be used as a model to scale the intensity data from an insulin crystal of another species.

From (1) we can also write

$$[\bar{I}_a]_i = k[\bar{I}_r]_i, \quad i = 1 \text{ to } \nu. \quad (5)$$

We shall now consider the following three cases: (i)  $n = n'$ , (ii)  $n > n'$  and (iii)  $n < n'$ .

### 3.1. Result for the case $n = n'$

We can rewrite (3) as

$$\sigma_i = [\bar{I}_a]_i / \varepsilon, \quad \sigma'_i = [\bar{I}'_i] / \varepsilon'. \quad (6)$$

Since the molecules in the given crystal and in the model are assumed to have the same composition, we can write

$$\sigma_i = \sigma'_i \exp(-2Bs^2). \quad (7)$$

For (7) we have used the fact that the atoms in the model are stationary while those in the given crystal are vibrating with the same isotropic temperature factor. Making use of (6) in (7), we obtain

$$[\bar{I}_a]_i = \rho[\bar{I}'_i] \exp(-2B\bar{s}_i^2). \quad (8)$$

Making use of (5) in (8), we obtain

$$[\bar{I}_r]_i = (\rho/k)[\bar{I}'_i] \exp(-2B\bar{s}_i^2), \quad (9)$$

which can be rewritten compactly as

$$y_i = \rho x_i p_1 \exp(-p_2 \bar{s}_i^2), \quad i = 1 \text{ to } \nu. \quad (10)$$

In (10),  $x_i$ ,  $y_i$  and  $\bar{s}_i^2$  are known quantities while  $p_1$  and  $p_2$  are the unknowns to be estimated. Equation (10) is the basic equation from which all the methods discussed in this paper are developed. In deriving (10), we have used (7), which involves the sums of the squares of the atomic scattering factors of the atoms in the molecules of the given crystal and the model. Thus, (10) is derived on the assumption that the atoms are uniformly distributed in the unit cell. However, in the final expression (10) for the basic equation,  $\sigma$  and  $\sigma'$  do not appear. We shall hence assume that (10) could be taken as a reasonably good approximation, even for actual crystals where the concept of random distribution of atoms in the unit cell fails due to the definite structure of the molecule, since the molecules composing the crystal and the model are assumed to have considerable structural similarity. This comment also applies to the basic equations derived in §§ 3.2 and 3.3 for the other two cases.

### 3.2. Result for the case $n > n'$

We shall use the symbol  $Q$  to denote the common part of the unit cell of the given crystal and the model\* and the symbol  $P$  for the extra atoms in the unit cell

of the given crystal. We can write the structure factor of a reflection  $H (=hkl)$  for the given crystal in terms of the contributions from the  $P$  and  $Q$  atoms as

$$F = F_Q + F_P \quad (11)$$

so that the intensity of the reflection can be written as

$$|F|^2 = |F_Q|^2 + |F_P|^2 + 2|F_P||F_Q|\Delta, \quad (12)$$

where  $\Delta$  is the cosine of the angle between  $F_P$  and  $F_Q$  if the crystal is noncentrosymmetric and the product of the signs of  $F_P$  and  $F_Q$  if the crystal is centrosymmetric. Since the  $P$  and  $Q$  groups are independent, the local average value of  $\Delta$  for reflections in any given  $(\sin\theta)/\lambda$  interval is zero. Let us denote  $|F_Q|^2$  and  $|F_P|^2$  by  $I_Q$  and  $I_P$ , respectively. Furthermore,  $|F|^2 = I_a$ . With these facts, we can obtain from (12)

$$[\bar{I}_a]_i = [\bar{I}_Q]_i + [\bar{I}_P]_i. \quad (13)$$

We can also write

$$[\bar{I}_Q]_i = \varepsilon[\sigma_q^0]_i \exp(-2B\bar{s}_i^2), \quad (14)$$

$$[\bar{I}_P]_i = \varepsilon[\sigma_p^0]_i \exp(-2B\bar{s}_i^2). \quad (15)$$

Using (5) on the left-hand side of (13) and (14) and (15) on the right-hand side of (13), we obtain

$$[\bar{I}_r]_i = (\varepsilon/k)([\sigma_q^0]_i + [\sigma_p^0]_i) \exp(-2B\bar{s}_i^2). \quad (16)$$

For the present situation where the model is assumed to consist of only the  $Q$  atoms, we can also write

$$[\bar{I}'_i] = \varepsilon'[\sigma_q^0]_i. \quad (17)$$

Making use of (17) in (16), we obtain

$$y_i = (\rho x_i + \varepsilon[\sigma_p^0]_i) p_1 \exp(-p_2 \bar{s}_i^2), \quad (18)$$

which is the required basic equation for this case.

### 3.3. Result for the case $n < n'$

We shall use the symbol  $Q$  to denote the common part of the given crystal\* and the model and  $P$  to denote the extra atoms in the unit cell of the model. Starting from the structure-factor relation  $F' = F'_Q + F'_P$  for the model and following a procedure similar to that used to derive (13) from (11), we obtain

$$[\bar{I}'_i] = [\bar{I}'_Q]_i + [\bar{I}'_P]_i. \quad (19)$$

Since the atoms in the model are assumed to be stationary, we can write

$$[\bar{I}'_Q]_i = \varepsilon'[\sigma_q^0]_i, \quad [\bar{I}'_P]_i = \varepsilon'[\sigma_p^0]_i \quad (20)$$

so that (19) can be rewritten as

$$[\sigma_q^0]_i = x_i / \varepsilon' - [\sigma_p^0]_i. \quad (21)$$

\* In this case, the model is assumed to consist only of  $Q$  atoms.

\* In this case, the given crystal is assumed to consist only of  $Q$  atoms.

We can also write

$$\begin{aligned} y_i &= [\bar{I}_r]_i = (1/k)[\bar{I}_a]_i \\ &= (1/k)[\bar{I}_Q]_i \\ &= (1/k)\varepsilon[\sigma_q]_i \\ &= (1/k)\varepsilon[\sigma_q^0]_i \exp(-2B\bar{s}_i^2). \end{aligned} \quad (22)$$

Eliminating  $[\sigma_q^0]_i$  between (21) and (22), we obtain

$$y_i = (\rho x_i - \varepsilon[\sigma_p^0]_i) p_1 \exp(-p_2 \bar{s}_i^2). \quad (23)$$

### 3.4. Basic equations in a common format

We can represent (10), (18) and (23) as a single equation

$$y_i = p_1 \exp(-p_2 \bar{s}_i^2) X_i \quad (24)$$

provided we define  $X_i$  as

$$X_i = \rho x_i + \omega \varepsilon[\sigma_p^0]_i$$

and assign the value 0, 1 or  $-1$  to  $\omega$  according to whether  $n = n'$ ,  $n > n'$  or  $n < n'$ ;  $p_1$  and  $p_2$  are the only unknowns in (24). It is obvious from (24) that the problem of determination of  $p_1$  and  $p_2$  may be approached using optimization techniques based on suitable objective functions derived from (24).

The success of an optimization method depends critically on (i) the quality of modelling and (ii) the proper choice of the objective function (Hamming, 1989). In this section we have given a procedure to take care of (i). We shall consider (ii) in § 4. Since the Newton-Raphson algorithm is convenient for tackling the present optimization problem, we shall give the essential results of that algorithm in § 4.1.

## 4. Optimization methods for determining $k$ and $B$

### 4.1. Newton-Raphson algorithm for optimization

Suppose that  $x$  and  $y$  are two observables that are known to satisfy an equation of the form

$$f(x, y; \mathbf{p}) = 0, \quad (25)$$

where  $\mathbf{p} = (p_1, p_2, \dots, p_\mu)$  is the unknown-parameter vector with  $\mu$  components. Suppose that we know  $\nu$  ( $> \mu$ ) pairs of empirical values  $(x_i, y_i)$ ,  $i = 1$  to  $\nu$ , that satisfy (25) for some  $\mathbf{p}$ . In this case, owing to the errors of observation in  $(x_i, y_i)$ , the value of  $f(x_i, y_i; \mathbf{p})$  will be close but not exactly equal to zero. Hence the positive quantity  $O_f$  (also called the objective function) defined by

$$O_f = \sum_{i=1}^{\nu} z_i^2, \quad (26)$$

where

$$z_i = f(x_i, y_i; \mathbf{p}), \quad (27)$$

will also be close but not exactly equal to zero. However, under the circumstance stated above, it is

reasonable to assume that the vector  $\mathbf{p}$  that minimizes (26) in the neighbourhood of some approximate solution  $(\mathbf{p})_0$  would be the best value for the unknown parameters that may be derived from the given data  $(x_i, y_i)$ ,  $i = 1$  to  $\nu$ . In connection with the optimization of (26) by the Newton-Raphson method, it is convenient to define a  $\nu \times \mu$  matrix  $M$  and a  $\nu \times 1$  matrix  $\mathbf{z}$  by

$$M = \begin{bmatrix} \partial z_1 / \partial p_1 & \partial z_1 / \partial p_2 & \cdots & \partial z_1 / \partial p_\mu \\ \partial z_2 / \partial p_1 & \partial z_2 / \partial p_2 & \cdots & \partial z_2 / \partial p_\mu \\ \vdots & \vdots & \ddots & \vdots \\ \partial z_\nu / \partial p_1 & \partial z_\nu / \partial p_2 & \cdots & \partial z_\nu / \partial p_\mu \end{bmatrix}, \quad (28)$$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_\nu \end{bmatrix}.$$

The Newton-Raphson algorithm for minimizing (26) is given by (Everitt, 1987)

$$(\mathbf{p})_{i+1} = (\mathbf{p})_i - (M^T M)^{-1} M^T \mathbf{z}, \quad i = 0, 1, 2, \dots, \quad (29)$$

where  $M^T$  is the transpose of  $M$ .

### 4.2. Derivation of the theoretical results for optimization using different objective functions

In our study we shall consider five objective functions based on five types of differences between  $y_i$  and  $p_1 \exp(-p_2 \bar{s}_i^2) X_i$  [see (24)]. The objective functions and the corresponding quantities needed for computing the matrices  $M$  and  $\mathbf{z}$  that are required in the optimization procedure are given in Table 1 for these cases. For convenience, the methods based on the various objective functions are given symbolic names as described in Table 1. We shall study the performance of the optimization based on these objective functions in § 6.2. We shall presently describe five other methods of using (24) for determining  $k$  and  $B$  from the known values of  $x_i$ ,  $y_i$  and  $\bar{s}_i$ .

## 5. Other methods of determining $k$ and $B$

The five methods considered here will be referred to as (i) LLSQ (=linearized least-squares method), (ii) ZSM( $B$ ) (=zero-slope method using  $B$  as the variable parameter), (iii) ZSML( $B$ ) (=zero-slope method involving a logarithm, using  $B$  as the variable parameter), (iv) ZSM( $k$ ) (=zero-slope method using  $k$  as the variable parameter) and (v) KBCIPM (=  $k$  and  $B$  curves intersection-point method).

### 5.1. Theory of LLSQ method

We can rewrite (24) as

$$p_1 \exp(-p_2 \bar{s}_i^2) = y_i / X_i, \quad i = 1 \text{ to } \nu. \quad (30)$$

Table 1. The quantities needed to form the matrices  $z$  and  $M$  corresponding to the different objective functions

In each case, the objective function is defined to be  $O_f = \sum_{i=1}^n z_i^2$ .  $h_i$  is defined to be  $h_i = p_1 \exp(-p_2 \bar{s}_i^2) X_i$ .

AD Method in which the objective function is based on the absolute difference between  $y_i$  and  $h_i$ .

FD Method in which the objective is based on the fractional difference between  $y_i$  and  $h_i$  in the standard form (i.e. the form in which the denominator of the fraction contains the sum of the quantities  $y_i$  and  $h_i$  pertaining to the crystal and to the model, respectively).

FDC Method in which the objective function is based on the fractional difference between  $y_i$  and  $h_i$  such that the denominator of the fraction contains the quantity  $y_i$  pertaining to the given crystal.

FDM Method in which the objective function is based on the fractional difference between  $y_i$  and  $h_i$  such that the denominator of the fraction contains the quantity  $h_i$  pertaining to the model.

FDCM Method in which the objective function is based on the fractional difference between  $y_i$  and  $h_i$  such that the denominator of the fraction contains the square root of the product of the quantities  $y_i$  and  $h_i$  pertaining to the given crystal and to the model, respectively.

Method	$z_i$	$\partial z_i / \partial p_1$	$\partial z_i / \partial p_2$
AD	$y_i - p_1 \exp(-p_2 \bar{s}_i^2) X_i$	$-\exp(-p_2 \bar{s}_i^2) X_i$	$\bar{s}_i^2 X_i p_1 \exp(-p_2 \bar{s}_i^2)$
FD	$\frac{y_i - p_1 \exp(-p_2 \bar{s}_i^2) X_i}{y_i + p_1 \exp(-p_2 \bar{s}_i^2) X_i}$	$\frac{-2X_i y_i \exp(-p_2 \bar{s}_i^2)}{[y_i + p_1 \exp(-p_2 \bar{s}_i^2) X_i]^2}$	$\frac{2X_i y_i \bar{s}_i^2 p_1 \exp(-p_2 \bar{s}_i^2)}{[y_i + p_1 \exp(-p_2 \bar{s}_i^2) X_i]^2}$
FDC	$[y_i - p_1 \exp(-p_2 \bar{s}_i^2) X_i] / y_i$	$(-X_i / y_i) \exp(-p_2 \bar{s}_i^2)$	$[[X_i \bar{s}_i^2 p_1 \exp(-p_2 \bar{s}_i^2)] / y_i$
FDM	$1 - \frac{y_i \exp(p_2 \bar{s}_i^2)}{X_i p_1}$	$\frac{y_i \exp(p_2 \bar{s}_i^2)}{X_i p_1^2}$	$\frac{-y_i \bar{s}_i^2 \exp(p_2 \bar{s}_i^2)}{X_i p_1}$
FDCM	$\frac{y_i - p_1 \exp(-p_2 \bar{s}_i^2) X_i}{[y_i p_1 \exp(-p_2 \bar{s}_i^2) X_i]^{1/2}}$	$-\frac{1}{2} \left[ \frac{y_i}{p_1^2 \exp(-p_2 \bar{s}_i^2) X_i} \right]^{1/2}$	$\frac{\bar{s}_i^2 [y_i + p_1 \exp(-p_2 \bar{s}_i^2) X_i]}{2[y_i p_1 \exp(-p_2 \bar{s}_i^2) X_i]^{1/2}}$

Taking natural logarithms of both sides of (30), we obtain

$$\ln p_1 - p_2 \bar{s}_i^2 = \ln (y_i / X_i). \quad (31)$$

We can rewrite (31) compactly as

$$A + B u_i = v_i, \quad (32)$$

where  $u_i$  and  $v_i$  are defined to be

$$u_i = 2\bar{s}_i^2 \quad \text{and} \quad v_i = \ln (X_i / y_i) \quad \text{with} \quad A = \ln k. \quad (33)$$

In (32),  $u_i$  and  $v_i$  are known quantities. Using the method of linear least squares, we obtain the unknowns  $A$  and  $B$ :

$$A = \left( \sum_i u_i^2 \sum_i v_i - \sum_i u_i \sum_i u_i v_i \right) / \left[ \nu \sum_i u_i^2 - \left( \sum_i u_i \right)^2 \right],$$

$$B = \left( \nu \sum_i u_i v_i - \sum_i u_i \sum_i v_i \right) / \left[ \nu \sum_i u_i^2 - \left( \sum_i u_i \right)^2 \right]. \quad (34)$$

From (33), we obtain

$$k = \exp A. \quad (35)$$

### 5.2. Theory of ZSM(B)

We can rewrite (24) as

$$k = (X_i / y_i) \exp(-2B\bar{s}_i^2), \quad (36)$$

where we have substituted for  $p_1$  and  $p_2$ . Suppose that  $B_0$  is the true value of the temperature factor for the given crystal. From (36) it is clear that the quantities  $(X_i / y_i) \exp(-2B_0 \bar{s}_i^2)$  computed for different intervals of  $(\sin \theta) / \lambda$  would be nearly equal. That is, if we plot a scatter diagram with  $\bar{s}_i$  as abscissa and  $(X_i / y_i) \exp(-2B_0 \bar{s}_i^2)$  as ordinate, we would expect to obtain a straight line with zero slope. However, since

the value of  $B_0$  is not known *a priori*, we cannot plot such a scatter diagram. In the absence of any information regarding the value of  $B_0$ , we can treat  $B$  in (36) as a variable and define a new quantity  $K_i(B)$  by

$$K_i(B) = (X_i / y_i) \exp(-2B\bar{s}_i^2). \quad (37)$$

If we give different values to  $B$  and compute  $\nu$  ordered pairs  $\{[\bar{s}_i, K_i(B)], i = 1 \text{ to } \nu\}$  corresponding to each value of  $B$ , we can then obtain a scatter diagram for each value of  $B$ . The scatter diagrams corresponding to different values of  $B$  would have different slopes.\* For example, the scatter diagram for which  $B > B_0$  would have negative slope while the one for which  $B < B_0$  would have positive slope. Thus, if we study the slope of the scatter diagrams for different values of  $B$  (say, 2, 3, 4, 5, ...), we would find the slope to change from a negative value to a positive value in some interval  $B' < B < B''$  of  $B$  [see Fig. 1(a) for an example]. The value  $B_0$  of  $B$  corresponding to zero slope is hence expected to be in the interval  $B' < B < B''$ . We shall now consider a convenient method for estimating the values of  $k$  and  $B$  from the slopes of the scatter diagrams corresponding to  $B'$  and  $B''$ .

The slope  $m$  of the scatter diagram is a function of  $B$ . That is,  $m = m(B)$ . The Taylor-series expansion of the function  $m(B)$  about  $B = B_0$  is given by

$$m(B) = m(B_0) + (dm/dB)_{B_0} (B - B_0) + \dots \quad (38)$$

Since the value of the slope of the scatter diagram for  $B = B_0$  is zero, we have  $m(B_0) = 0$ . Neglecting the

\* We shall fit a least-squares straight line to the scatter diagram corresponding to a given  $B$  and refer to the slope of the least-squares straight line thus obtained as the slope of the scatter diagram corresponding to the particular  $B$ .

terms of higher powers in  $B - B_0$ , we obtain from (38)

$$m(B) \approx (dm/dB)_{B_0}(B - B_0). \quad (39)$$

Putting  $B = B'$  in (39), we arrive at

$$B_0 \approx B' - [m(B') / (dm/dB)_{B_0}]. \quad (40)$$

Since the values of  $B'$  and  $B''$  are close to  $B_0$ , we can approximate the value of  $(dm/dB)_{B_0}$  by  $(m'' - m') / (B'' - B')$ . That is,

$$(dm/dB)_{B_0} \approx (m'' - m') / (B'' - B'), \quad (41)$$

where  $m' = m(B')$  and  $m'' = m(B'')$ . Making use of (41) in (40), we obtain

$$B_0 \approx B' - [m' / (m'' - m')](B'' - B'). \quad (42)$$

Hence, we can determine the value of  $B_0$ . Using this value, we can compute the value of  $k$  from [see (36)]

$$k = (1/\nu) \sum_{i=1}^{\nu} (X_i/y_i) \exp(-2B_0\bar{s}_i^2). \quad (43)$$

### 5.3. Theory of ZSML(B)

This method is similar to the ZSM(B). Here the natural logarithm of  $(X_i/y_i) \exp(-2B\bar{s}_i^2)$  is taken as the ordinate instead of  $(X_i/y_i) \exp(-2B\bar{s}_i^2)$ . Scatter diagrams obtained in a typical situation are shown in Fig. 1(b).

### 5.4. Theory of ZSM(k)

We can rewrite (24) as

$$\exp(p_2\bar{s}_i^2) = p_1 X_i/y_i. \quad (44)$$

Taking natural logarithms of both sides, we can rewrite (44) as

$$B = \ln [X_i/(ky_i)] / 2\bar{s}_i^2. \quad (45)$$

For a given crystal, the left-hand side of (45) is a constant. From an argument similar to that used in § 5.2, it is clear that the graph of  $\ln [X_i/(ky_i)] / 2\bar{s}_i^2$  versus  $\bar{s}_i$  will have zero slope when the true value for  $k$  ( $k_0$ , say) is used.

The value of  $k$  obtained from the linearized least-squares method ( $k_L$ , say) may be taken to be the initial value of  $k$  about which the scatter diagrams are to be obtained. From (45), it is clear that the scatter diagram is expected to have positive slope for  $k < k_0$  and negative slope for  $k > k_0$ . Fig. 1(c) shows the results obtained in a typical situation. The value of  $k$  for which the slope of the scatter diagram is expected to be zero may then be obtained by the procedure used in § 5.2 by studying the scatter diagrams obtained for  $k = \alpha k_L$ , where  $\alpha = 0.8, 1.0, 1.2, 1.4, 1.6$  etc. If  $k_0$  is the value of the scale factor obtained corresponding to zero slope of the scatter diagram, the best estimate of  $B$  may then be obtained

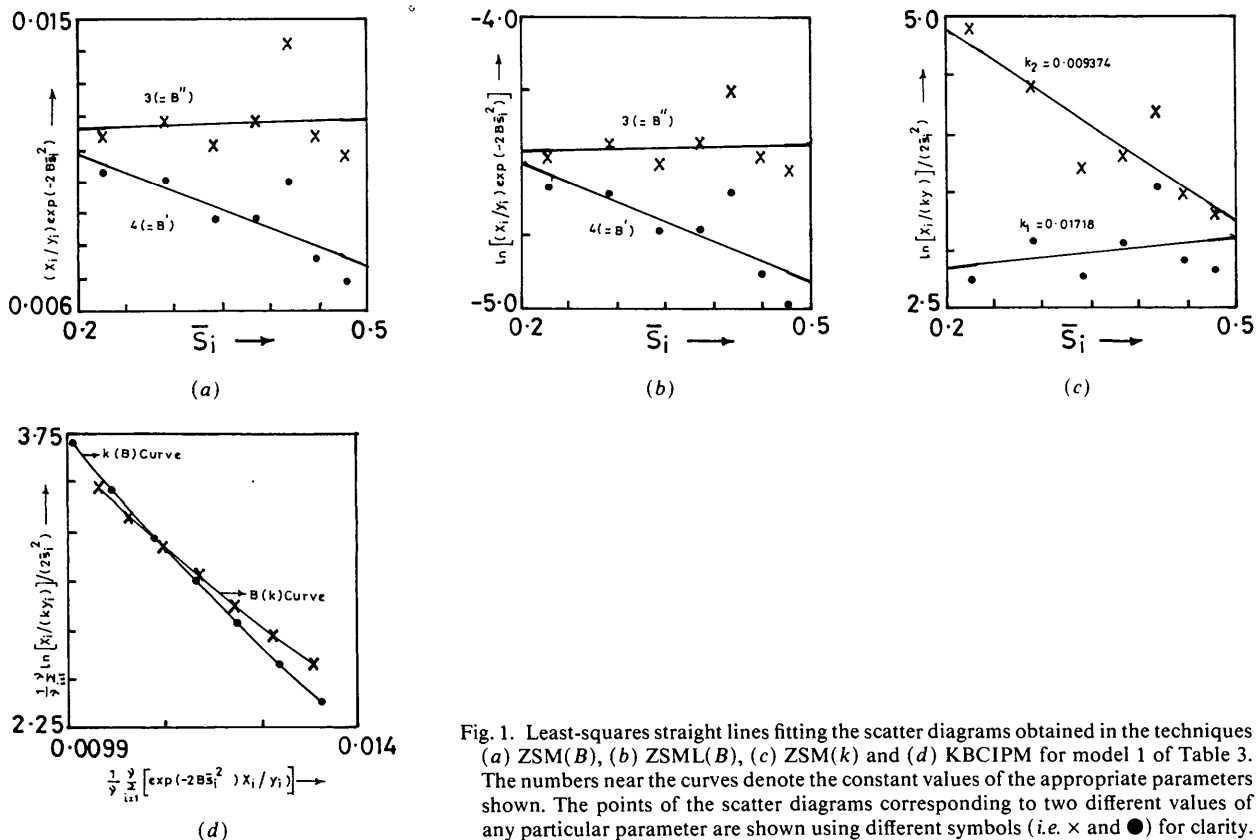


Fig. 1. Least-squares straight lines fitting the scatter diagrams obtained in the techniques (a) ZSM(B), (b) ZSML(B), (c) ZSM(k) and (d) KBCIPM for model 1 of Table 3. The numbers near the curves denote the constant values of the appropriate parameters shown. The points of the scatter diagrams corresponding to two different values of any particular parameter are shown using different symbols (i.e.  $\times$  and  $\bullet$ ) for clarity.

Table 2. *Relevant data for the crystals used in the tests*

Crystal codes:

- I L-Prolyl-L-leucine monohydrate (Panneerselvam & Chacko, 1990).  
 II L-Prolyl-L-isoleucine monohydrate (Panneerselvam, Chacko & Veena, 1989).  
 III 5 $\alpha$ -Androstan-3 $\alpha$ -ol-17-one (High & Kraut, 1966).  
 IV 5 $\alpha$ -Androstan-3 $\beta$ -ol-17-one (Weeks, Cooper, Norton, Hauptman & Fisher, 1971).  
 V L-Arginine hydrochloride monohydrate (Mazumdar & Srinivasan, 1964).  
 VI L-Arginine hydrobromide monohydrate (Mazumdar & Srinivasan, 1964).  
 VII 2-(Chlorophenyl)thiazolidine (Eswaramoorthy, Ponnuswamy & Raju, 1991).  
 VIII 2-Phenylthiazolidine (Eswaramoorthy, Ponnuswamy & Raju, 1991).  
 IX 3-Ethyl-1-methyl-2,6-diphenyl-4-piperidone (Sekar & Parthasarathy, 1992).  
 X 1,3,3-Trimethyl-2,6-diphenyl-4-piperidone (Sekar, Parthasarathy & Radhakrishnan, 1992).  
 XI 3-Methyl-2,6-diphenyl-4-piperidone (Sekar, Parthasarathy & Rajalingam, 1990).  
 XII 2,6-Bis(*p*-methoxyphenyl)-3,5-dimethyl-4-piperidone (Sekar, Parthasarathy & Radhakrishnan, 1990).

Crystal	Space group	Asymmetric unit	<i>z</i>	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	$\alpha$ (°)	$\beta$ (°)	$\gamma$ (°)
I	<i>P</i> 2 <sub>1</sub>	C <sub>11</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub> ·H <sub>2</sub> O	2	6.492	5.417	20.491	90.0	96.6	90.0
II	<i>P</i> 2 <sub>1</sub>	C <sub>11</sub> H <sub>20</sub> N <sub>2</sub> O <sub>3</sub> ·H <sub>2</sub> O	2	6.601	5.413	19.128	90.0	98.1	90.0
III	<i>P</i> 2 <sub>1</sub>	C <sub>19</sub> H <sub>30</sub> O <sub>2</sub>	2	9.560	7.900	11.780	90.0	111.4	90.0
IV	<i>P</i> 2 <sub>1</sub>	C <sub>19</sub> H <sub>30</sub> O <sub>2</sub>	2	6.596	21.521	6.313	90.0	109.4	90.0
V	<i>P</i> 2 <sub>1</sub>	(C <sub>6</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub> ·HCl·H <sub>2</sub> O) <sub>2</sub>	2	11.220	8.500	11.070	90.0	91.0	90.0
VI	<i>P</i> 2 <sub>1</sub>	(C <sub>6</sub> H <sub>14</sub> N <sub>2</sub> O <sub>2</sub> ·HBr·H <sub>2</sub> O) <sub>2</sub>	2	11.260	8.650	11.250	90.0	91.3	90.0
VII	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C <sub>9</sub> H <sub>10</sub> CINS	4	5.529	8.486	20.432	90.0	90.0	90.0
VIII	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C <sub>9</sub> H <sub>11</sub> NS	4	8.388	18.114	5.625	90.0	90.0	90.0
IX	<i>C</i> 2/ <i>c</i>	C <sub>20</sub> H <sub>23</sub> NO	8	21.526	6.096	25.958	90.0	97.3	90.0
X	<i>P</i> $\bar{1}$	C <sub>20</sub> H <sub>23</sub> NO	2	6.486	10.921	12.542	79.9	88.9	73.5
XI	<i>C</i> 2/ <i>c</i>	C <sub>18</sub> H <sub>19</sub> NO	8	19.266	6.999	21.653	90.0	94.4	90.0
XII	<i>P</i> 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C <sub>21</sub> H <sub>25</sub> NO <sub>3</sub>	4	7.031	11.675	22.783	90.0	90.0	90.0

from (45) as

$$B = (1/\nu) \sum_{i=1}^{\nu} \ln(X_i/k_0 y_i)/2\bar{s}_i^2. \quad (46)$$

### 5.5. Theory of KBCIPM

We have shown that *B* and *k* can be written as [see (46) and (43)]

$$B = (1/\nu) \sum_{i=1}^{\nu} \ln(X_i/ky_i)/2\bar{s}_i^2, \quad (47)$$

$$k = (1/\nu) \sum_{i=1}^{\nu} X_i \exp(-2B\bar{s}_i^2)/y_i. \quad (48)$$

Suppose that approximate values of *k* and *B* are known to be *k<sub>a</sub>* and *B<sub>a</sub>*, respectively. For example, we may obtain the approximate values of *k* and *B* from the LLSQ method (see § 5.1). One can vary *k* in the neighbourhood of *k<sub>a</sub>* and use (47) to calculate the corresponding values of *B*. Thus, taking the values of *k* to be  $\alpha k_a$ , where  $\alpha = 0.90, 0.95, 1.0, 1.05$  and  $1.1$ , the corresponding values of *B* may be obtained from (47). The various points (*k*, *B*) thus obtained define a curve in the (*k*, *B*) plane and we shall refer to this curve as the *B*(*k*) curve. Similarly, one can vary *B* in the neighbourhood of *B<sub>a</sub>* and use (48) to calculate the corresponding value of *k*. The set of points obtained by taking *B* to be *B<sub>a</sub>* - 0.3, *B<sub>a</sub>* - 0.15, *B<sub>a</sub>*, *B<sub>a</sub>* + 0.15 and *B<sub>a</sub>* + 0.30 defines a curve in the (*k*, *B*) plane and we shall refer to this curve as the *k*(*B*) curve. The point where the *k*(*B*) curve and *B*(*k*) curve intersect may be taken to be a more accurate

estimate of *k* and *B*. Fig. 1(*d*) shows the result obtained in a typical situation.

### 6. Test of the theoretical results

The above methods of determining the scale and temperature factors were tested on the intensity data of a few known crystals (for details of these crystals see Table 2). The results obtained from these tests are given in Table 3 for crystal I. The first model for crystal I was obtained by deleting the C atom marked 1 in Fig. 2(*b*). The second model for crystal I was obtained by deleting the C atom marked 1 in Fig. 2(*b*) as well as replacing the H atom at the C atom marked 2 by a C atom at a stereochemically correct position. Crystal II itself was taken as the third model for crystal I. The results obtained for the other crystals are provided in Tables I-VIII of the deposited material\* and the method of obtaining the models for these crystals are also indicated symbolically in these tables. For each crystal, the value of *k* and the average value of the isotropic equivalents of the temperature factors of the various non-hydrogen atoms as obtained in the final cycle of the least-squares refinement were taken to be the true values of *k* and *B*, respectively, and these are denoted by *k<sub>t</sub>* and *B<sub>t</sub>*.

\* Tables and other relevant material have been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 55489 (13 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

Table 3. Determination of the values of  $k$  and  $B$  for crystal I

Results obtained using model 1				Results obtained using model 2			
No.	Method	$k \times 10^{-3}$	$B$	No.	Method	$k \times 10^{-3}$	$B$
1	AD	11.218 (12.2)	3.20 (-38.6)	1	AD	11.094 (10.9)	3.25 (-37.7)
2	FD	11.719 (17.2)	3.02 (-42.0)	2	FD	11.538 (15.4)	3.09 (-40.7)
3	FDC	11.644 (16.4)	3.08 (-40.8)	3	FDC	11.497 (15.0)	3.13 (-40.0)
4	FDM	11.755 (17.5)	2.98 (-42.8)	4	FDM	11.558 (15.6)	3.06 (-41.3)
5	FDCM	11.182 (11.8)	3.09 (-40.6)	5	FDCM	11.154 (11.5)	3.14 (-39.7)
6	LLSQ	11.718 (17.2)	3.02 (-42.0)	6	LLSQ	11.538 (15.4)	3.09 (-40.7)
7	ZSM( $B$ )	11.537 (15.4)	3.09 (-40.7)	7	ZSM( $B$ )	11.342 (13.4)	3.16 (-39.4)
8	ZSML( $B$ )	11.627 (16.3)	3.05 (-41.5)	8	ZSML( $B$ )	11.463 (14.6)	3.11 (-40.3)
9	ZSM( $k$ )	11.404 (14.0)	3.12 (-40.2)	9	ZSM( $k$ )	11.269 (12.7)	3.17 (-39.1)
10	KBCIPM	11.259 (12.6)	3.17 (-39.1)	10	KBCIPM	11.161 (11.6)	3.21 (-38.3)
	Mean	11.506 (15.1)*	3.08 (-40.8)*		Mean	11.362 (13.6)*	3.14 (-39.7)*

Results obtained using model 3				Results obtained using $\bar{X}$ data			
No.	Method	$k \times 10^{-3}$	$B$	No.	Method	$k \times 10^{-3}$	$B$
1	AD	11.198 (12.0)	3.23 (-38.0)	1	AD	11.170 (11.7)	3.22 (-38.1)
2	FD	11.403 (14.0)	3.15 (-39.4)	2	FD	11.553 (15.5)	3.09 (-40.7)
3	FDC	11.319 (13.2)	3.23 (-38.0)	3	FDC	11.484 (14.8)	3.15 (-39.6)
4	FDM	11.442 (14.4)	3.11 (-40.4)	4	FDM	11.587 (15.9)	3.05 (-41.5)
5	FDCM	11.379 (13.8)	3.25 (-37.7)	5	FDCM	11.051 (10.5)	3.16 (-39.4)
6	LLSQ	11.402 (14.0)	3.16 (-39.4)	6	LLSQ	11.552 (15.5)	3.09 (-40.7)
7	ZSM( $B$ )	11.154 (11.5)	3.25 (-37.7)	7	ZSM( $B$ )	11.339 (13.4)	3.17 (-39.2)
8	ZSML( $B$ )	11.333 (13.3)	3.18 (-39.0)	8	ZSML( $R$ )	11.474 (14.7)	3.11 (-40.2)
9	ZSM( $k$ )	11.219 (12.2)	3.21 (-38.3)	9	ZSM( $k$ )	11.297 (13.0)	3.17 (-39.2)
10	KBCIPM	11.018 (10.2)	3.29 (-36.8)	10	KBCIPM	11.147 (11.5)	3.23 (-38.1)
	Mean	11.287 (12.9)*	3.20 (-38.5)*		Mean	11.365 (13.7)*	3.14 (-39.7)*

Deviations in % from the refined values  $k_i$  and  $B_i$  are given within parentheses [cf. equation (51)]

$k_i = 10.000$	$B_i = 5.21$
$\langle k \rangle = 11.385 (13.9)$	$\langle B \rangle = 3.14 (-39.7)$
$\bar{k} = 11.365 (13.7)$	$\bar{B} = 3.14 (-39.7)$
$k_I = 15.187 (51.9)$	$B_I = 1.74 (-66.6)$
$k_{II} = 14.568 (45.7)$	$B_{II} = 1.91 (-63.3)$
$k_{III} = 12.433 (24.3)$	$B_{III} = 2.40 (-53.9)$

\* These error values are calculated using the corresponding mean values directly and they are not the average of the quantities in parentheses in the respective columns of the table.

### 6.1. Some explanatory remarks

(i) All the techniques require the local average intensity data (also called the  $\bar{I}$  data) for at least one model. The known crystal structure of a compound whose molecule is expected to have reasonable similarity to that of the given crystal may be chosen as a model. We shall refer to the known crystal structure

from which a model is to be obtained (by making some appropriate minor modifications to the asymmetric unit) as the potential model crystal. In many situations, it is possible to obtain more than one model from a particular potential model crystal and hence the methods of determining  $k$  and  $B$  can be implemented on more than one  $\bar{I}$ -data set. We shall refer to the different models that can be obtained from a particular potential model crystal as model 1, model 2 etc.

(ii) The  $\bar{I}$  data of each model were obtained from the structure-factor magnitudes of the reflections calculated from the known positions of the atoms in the model by taking the atoms to be stationary.

(iii) The mean value of  $k$  obtained by different methods [i.e. AD, FD etc. (see Table 1)] from the  $\bar{I}$  data of each model is given at the bottom of Table 3 corresponding to that model and the mean of these mean values, denoted by  $\langle k \rangle$ , is also given at the bottom of the table. A similar situation applies to  $B$ .

(iv) If we have  $m$  different models for a given crystal, instead of calculating the mean of the mean values of  $k$  and  $B$  as described in (iii), we can average the  $X$  data ( $X_1, X_2, \dots, X_m$ ) obtained from the  $m$

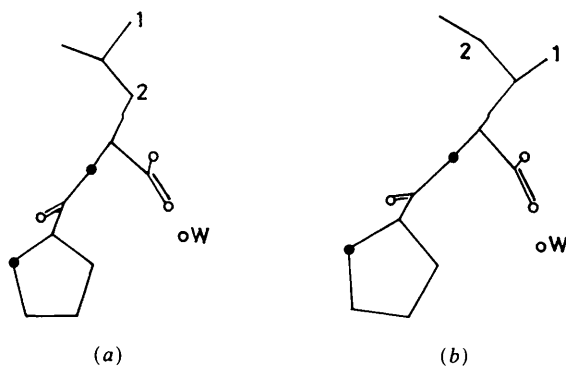


Fig. 2. Representation of the molecules defining the asymmetric units of crystals I and II described in Table 2.



Table 4. Summary of the percentage errors in the values of  $k$  and  $B$  estimated by the various procedures for different crystals

Results for crystal I are taken from Table 3 and those for the others from Tables I-VIII of the deposited material.

Crystal	% error in					% error in				
	$\langle k \rangle$	$\bar{k}$	$k_I$	$k_{II}$	$k_{III}$	$\langle B \rangle$	$\bar{B}$	$B_I$	$B_{II}$	$B_{III}$
I	13.9	13.7	51.9	45.7	24.3	-39.7	-39.7	-66.6	-63.3	-53.9
II	-3.6	-4.0	64.5	53.3	24.5	-11.7	-11.8	-58.9	-53.3	-39.8
III	5.3	*	12.3	24.1	23.2	-0.3	*	5.6	-3.0	-2.5
IV	-7.2	*	9.0	19.5	19.2	8.4	*	7.5	-4.4	-3.4†
V	1.4	1.5	23.0	19.6	19.6	-7.0	-7.0	-40.1	-38.0	-37.4
VI	2.3	3.0	11.4	9.6	10.1	1.6	1.0	5.3	7.5	7.0
VII	5.8	5.7	15.6	15.2	12.5	-2.4	-2.5	-6.9	-6.9	-5.3
VIII	-8.5	-8.3	10.8	10.3	6.5†	-8.8	-9.1	-19.5	-19.5	-17.9
IX	-12.0	-11.0	16.2	16.5	16.5	-9.2	-9.6	-16.7	-16.9	-18.5
Mean	6.7	6.7	23.9	23.8	17.4	9.9	11.5	25.2	23.6	20.6

\* The  $\bar{X}$ -data method cannot be applied here since only one model was used.† The case where the values obtained by a *MULTAN80* procedure (i.e. I, II or III) are better than the values obtained by the new procedures.

models [see (24)] and use the resulting average  $X$  data (called  $\bar{X}$  data) for determining the scale and temperature factors of the given crystal. For this, the same partition  $\pi_i$  should be used for all the models. The basic equation for this procedure (called the  $\bar{X}$ -data method) can be obtained from (24) as

$$y_i = p_i \exp(-p_2 \bar{s}_i^2) \bar{X}_i, \quad (49)$$

where

$$\bar{X}_i = (X_1 + X_2 + \dots + X_m) / m. \quad (50)$$

In Table 3, the values of  $k$  and  $B$  obtained by this method are denoted by  $\bar{k}$  and  $\bar{B}$ , respectively.

(v) It is convenient to define the percentage error  $r$  in the measured value  $v_m$  of a quantity  $v$  ( $=k$  or  $B$ ) to be

$$r = 100(v_m - v_t) / v_t, \quad (51)$$

where  $v_t$  is the true value of  $v$ . A positive value of  $r$  denotes overestimation while a negative value denotes underestimation.

(vi) For each crystal, the values of  $k$  and  $B$  were also computed using *MULTAN80* by the following three procedures, which differ in data input. For the first procedure (called procedure I), the input data are the type and number of atoms of each type in the unit cell. The second and third procedures (procedures II and III) use the additional data in the form of the coordinates of the atoms of a part of the molecule (e.g. benzene ring or skeleton of piperidone ring etc.) and those for the full molecule, respectively, for implementing the Debye formula for molecular scattering. The values of  $k$  and  $B$  obtained by procedures I, II and III are denoted by  $k_I$ ,  $k_{II}$  and  $k_{III}$  and  $B_I$ ,  $B_{II}$  and  $B_{III}$ , respectively.

(vii) Fig. 1 graphically illustrates the principle involved in the methods *ZSM(B)*, *ZSML(B)*, *ZSM(k)* and *KBCIPM*. These methods are, however, implemented in terms of numerical computations in our study.

## 6.2. Results obtained from the new techniques and *MULTAN80*

From the present study on the various crystals, we obtain the following results. (i) No particular technique is uniformly best or worst for the determination of both  $k$  and  $B$ . (ii) For a particular model of a given crystal, the method that yields the best value of  $k$  does not necessarily yield the best value for  $B$  as well. (iii) For a given crystal using different models, the technique that gives the best value of  $k$  with one model does not necessarily give the best value of  $k$  with another model. A similar statement is valid for the evaluation of  $B$ . Furthermore, random errors in the input data may influence the various techniques differently. Hence, it is thought useful to employ the different techniques for any given model and compute the averages of the values of  $k$  and  $B$  obtained from the different techniques.\* The mean values thus obtained are given in Table 3 along with the percentage errors in the estimates (shown within parentheses). The mean values of  $k$  and  $B$  of these mean values obtained from the different models (denoted by  $\langle k \rangle$  and  $\langle B \rangle$ ) are also given in Table 3 together with the values  $\bar{k}$  and  $\bar{B}$  obtained by the  $\bar{X}$ -data method. For each crystal, the values of  $\bar{k}$  and  $\langle k \rangle$  and  $\bar{B}$  and  $\langle B \rangle$  are found to be equally good. However, since the  $\bar{X}$ -data method is computationally somewhat simpler, it is to be preferred.

Table 4 gives a summary of the percentage errors in the mean values  $\bar{k}$ ,  $\langle k \rangle$ ,  $\bar{B}$  and  $\langle B \rangle$  for the given crystals. For comparison, the percentage errors in the estimates of  $k$  and  $B$  obtained by the three *MULTAN80* procedures are also given. From Table 4, we obtain the following results.

\* The calculation of the  $\bar{I}$  data for the model takes the major portion of the calculation time; the evaluation of  $k$  and  $B$  by all ten new techniques takes much less time. For example, for the typical case reported in Table 3, the total time taken for the job in a MicroVAX-II system is 3.5 min of which the calculation of the  $k$  and  $B$  values by all the ten techniques takes only 30 s.

(i) Of the three *MULTAN80* procedures, procedure III is found to be the best in one out of nine cases for the determination of  $k$  and in one out of nine cases for the determination of  $B$ . It is interesting to note that procedure I turns out to be preferable to procedure III in three out of nine cases for the determination of  $k$  and two out of nine cases for the determination of  $B$ .

(ii) The new techniques are found to lead to significantly better values of  $k$  and  $B$  than *MULTAN80* procedures in eight out of nine cases. Even in the rare situations where *MULTAN80* values are better, the new techniques are only slightly inferior. Thus, it appears that the new techniques could be used profitably.

(iii) To obtain an overall idea, the global average values of the percentage-error magnitudes in the estimated values of  $k$  and  $B$  for the various crystals are also given in Table 4. These also show that the new techniques are in general preferable to the *MULTAN80* procedures.

(iv) It may incidentally be noted that all ten new techniques and the three *MULTAN80* procedures underestimate the value of  $B$  in most cases. Furthermore, the new techniques overestimate the values of  $k$  in most cases while the *MULTAN80* procedures lead to overestimated values in all the cases.

### 7. Concluding remarks

From the above considerations it appears that the  $\bar{X}$ -data method could be used profitably in practice. The new procedure is likely to be particularly useful in the field of protein crystallography where one has access to intensity data from crystals of homologous proteins. However, further work needs to be carried out on the data from actual protein crystals to confirm this conclusion.

One can also employ the maximum-entropy principle to tackle the present problem instead of the procedures discussed in this paper. Work in this direction is in progress.

The authors thank Dr M. N. Ponnuswamy and Mr S. Eswaramoorthy for providing raw intensity data for some crystals used in the test of the theoretical results. Thanks are also due to the Council of Scientific and Industrial Research, India, for the award of a Senior Research Fellowship to KS.

### References

- ALLEN, F. H., KENNARD, O. & TAYLOR, R. (1983). *Acc. Chem. Res.* **16**, 146–151.
- BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, D. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- ESWARAMOORTHY, S., PONNUSWAMY, M. N. & RAJU, K. S. (1991). Personal communication.
- EVERITT, B. S. (1987). *Introduction to Optimization Methods and their Application in Statistics*. London, New York: Chapman & Hall.
- HAMMING, R. W. (1989). *Introduction to Applied Numerical Analysis*. New York: Hemisphere.
- HIGH, D. F. & KRAUT, J. (1966). *Acta Cryst.* **21**, 88–96.
- MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1980). *MULTAN80. A System of Computer Programs for the Automatic Solution of Crystal Structures for X-ray Diffraction Data*. Univ. of York, England, and Louvain, Belgium.
- MAZUMDAR, S. K. & SRINIVASAN, R. (1964). *Curr. Sci. (India)*, **33**, 573–575.
- PANNEERSELVAM, K. & CHACKO, K. K. (1990). *Int. J. Pept. Protein Res.* **35**, 460–464.
- PANNEERSELVAM, K., CHACKO, K. K. & VEENA, R. (1989). *Int. J. Pept. Protein Res.* **33**, 191–194.
- ROGERS, D. (1965). In *Computing Methods in Crystallography*, edited by J. S. ROLLETT. Oxford: Pergamon Press.
- SEKAR, K. & PARTHASARATHY, S. (1992). Personal communication.
- SEKAR, K., PARTHASARATHY, S. & RADHAKRISHNAN, T. R. (1990). *Acta Cryst.* **C46**, 1338–1340.
- SEKAR, K., PARTHASARATHY, S. & RADHAKRISHNAN, T. R. (1993). *Acta Cryst.* **C49**, 93–95.
- SEKAR, K., PARTHASARATHY, S. & RAJALINGAM, P. (1990). *Acta Cryst.* **C46**, 1153–1155.
- WEEKS, C. M., COOPER, A., NORTON, D. A., HAUPTMAN, H. & FISHER, J. (1971). *Acta Cryst.* **B27**, 1562–1572.
- WILSON, A. J. C. (1942). *Nature (London)*, **150**, 152–154.

*Acta Cryst.* (1993). **A49**, 398–405

## Edge Contributions to the Kirste–Porod Formula: the Truncated Circular Right Cone Case

BY S. CICCARIELLO

*Dipartimento di Fisica ‘G. Galilei’ and sez. INFN, via Marzolo 8, I-35131 Padova, Italy*

(Received 5 June 1992; accepted 23 September 1992)

### Abstract

The wide-angle correlation function  $\gamma(r)$  of a dilute system of particles, shaped as truncated circular right

cones, is evaluated in closed form. The angular average of the contribution proportional to  $r^3$  gives  $\gamma'''(0^+)$ , the value at the origin of the third derivative with respect to  $r$  of the corresponding small-angle