described in § 4. In such hypothetical cases, $|F_N|$ will be free from errors of observation. However, when one uses the observed overall value of an $R$ index for finding $\langle|\Delta\mathbf{r}|\rangle$, one must remember that the experimental errors in $|F_{obs}|$ would also contribute to the overall value of the $R$ index. Owing to this one has to note the following two points in estimating $\langle|\Delta\mathbf{r}|\rangle$ in actual cases. (*a*) Since the overall values of Booth-type indices for a complete model (*i.e.* $P = N$) are generally small, they would not be suitable for estimating the mean coordinate error of a complete model. (*b*) The $\langle|\Delta\mathbf{r}|\rangle$ value obtained for a model from the observed overall values of $R$ indices may be somewhat overestimated.

We thank Sekar and Geetha for providing structural data for two of the structures used in the test of the theoretical results. One of us (NE) thanks the University of Madras for financial assistance.

### References

KISTENMACHER, J. J., HUNT, D. J. & MARSH, R. E. (1972). *Acta Cryst.* **B28**, 3352-3361.

KROON, J. & KANTERS, J. A. (1973). *Acta Cryst.* **B29**, 1278-1283.

LISGARTEN, J. N. & PALMER, J. N. (1980). *Acta Cryst.* **B36**, 2345-2349.

LOW, J. N., HOWE, R. E., SCRIMGEOUR, C. M. & WATT, D. W. (1988). *Acta Cryst.* **C44**, 1762-1764.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802-810.

PARTHASARATHY, S. (1987). *Direct Methods: Macromolecular Crystallography and Crystallographic Statistics*, edited by H. SCHENK, A. J. C. WILSON & S. PARTHASARATHY, pp. 93-107. Singapore: World Scientific.

PARTHASARATHY, S. & PARTHASARATHI, V. (1975). *Acta Cryst.* **A31**, 178-181.

PARTHASARATHY, S. & VELMURUGAN, D. (1981). *Acta Cryst.* **A37**, 472-480.

PONNUSWAMY, M. N. & PARTHASARATHY, S. (1977). *Acta Cryst.* **A33**, 838-844.

SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. New York: Pergamon.

SRINIVASAN, R. & RAMACHANDRAN, G. N. (1966). *Acta Cryst.* **20**, 570-571.

SWAMINATHAN, P. & SRINIVASAN, R. (1975). *Acta Cryst.* **A31**, 310-318.

VELMURUGAN, D. & PARTHASARATHY, S. (1981*a*). *Acta Cryst.* **A37**, 480-485.

VELMURUGAN, D. & PARTHASARATHY, S. (1981*b*). *Pramana*, **17**, 499-508.

*Acta Cryst.* (1990). **A46**, 502-507

# The Minimum Probable Values of Residuals for the Best Centrosymmetric Model of an Approximately Centrosymmetric Crystal

BY N. ELANGO AND S. PARTHASARATHY

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

(*Received* 19 *December* 1989; *accepted* 24 *January* 1990)

### Abstract

Theoretically expected overall values of ten types of normalized $R$ indices are derived for the best centrosymmetric model of a non-centrosymmetric crystal which is approximately centrosymmetric as a function of the mean coordinate error $\langle|\Delta\mathbf{r}|\rangle$ of the best centrosymmetric model of the non-centrosymmetric crystal. The results obtained were tested in a few cases.

### 1. Introduction

The joint probability density function of the normalized structure-factor magnitude $y_N$ of a non-centrosymmetric crystal which is approximately centrosymmetric and that (*i.e.* $y_N^c$) calculated for the best centrosymmetric model (see later for an explanation of this nomenclature) has been worked out by Swaminathan & Srinivasan (1975) (SS, 1975). They have used this distribution to obtain the *local value* of the normalized Booth-type index ${}_BR_1(y_N)$ (see Table 1 for a definition) as a function of a parameter $D$ [see equation (12) for a definiton of $D$]. A number of other $R$ indices which are more efficient in the refinement stage are available in the literature [see Elango & Parthasarathy (1990) for their definitions and notation] and it is worthwhile to obtain the theoretical expressions for their overall values for the present situation. From a practical point of view, the overall value finds more general use than the local values of $R$ indices. Further, it is essential to take into account the effect of data truncation due to unobserved reflections into the theoretical treatment. We shall therefore derive the theoretical expressions for the *overall values* of different types of $R$ indices and use them to obtain the overall values of $R$ indices *directly* as a function of $\langle|\Delta\mathbf{r}|\rangle$ in the form of a table. In this paper we shall follow the notation and nomenclature used by Elango & Parthasarathy (1990). We

* Contribution No. 750.

0108-7673/90/060502-06$03.00 

shall restrict ourselves to the case $P = N$ owing to the theoretical complexity of the present situation. This does not, however, limit the utility of the results, since during the structure completion stage the model of a non-centrosymmetric crystal having a high degree of centrosymmetry will be generally taken to be centrosymmetric until the approximate locations of *all* atoms in the true structure get determined. It may be noted here that the results obtained in this paper represent the minimum probable value of the $R$ indices for the best centrosymmetric model (see § 2 for details).

## 2. Derivation of the theoretical expressions for the overall values of $R$ indices

Consider a non-centrosymmetric crystal of space group $P1$ which is approximately centrosymmetric containing a sufficiently large number ($N$, say) of similar atoms in the unit cell. If we choose the origin to be at the position of the approximate centre of symmetry, the position vectors of these $N$ atoms can be taken to be $\mathbf{r}_{Nj}+\Delta\mathbf{r}_{Nj}$, $-\mathbf{r}_{Nj}+\Delta\mathbf{r}'_{Nj}$, $j=1$ to $N/2$. Following Luzzati (1953), we shall take $\Delta\mathbf{r}_{Nj}$ and $\Delta\mathbf{r}'_{Nj}$ to be mutually independent random vectors obeying a Gaussian distribution law. Suppose that we denote a group of $P$ atoms with scattering powers $f_{Pj}$ ($j=1, 2, \ldots, P$) at locations $\mathbf{r}_{Pj}$ symbolically as $\bigcup_{j=1}^{P}(f_{Pj}, \mathbf{r}_{Pj})$. We can then represent the given crystal structure symbolically as

true structure:

$$\bigcup_{j=1}^{N/2}(f_{Nj}, \mathbf{r}_{Nj}+\Delta\mathbf{r}_{Nj})+\bigcup_{j=1}^{N/2}(f_{Nj}, -\mathbf{r}_{Nj}+\Delta\mathbf{r}'_{Nj}). \quad (1)$$

In our study, we define the following centrosymmetric structure as the model for the true structure defined in (1):

model structure:

$$\bigcup_{i=1}^{N/2}(f_{Ni}, \mathbf{r}_{Ni})+\bigcup_{i=1}^{N/2}(f_{Ni}, -\mathbf{r}_{Ni}). \quad (2)$$

Let us define $\langle|\Delta\mathbf{r}|\rangle$ to be

$$\langle|\Delta\mathbf{r}|\rangle=(1/N)\sum_{j=1}^{N/2}[|\Delta\mathbf{r}_{Nj}|+|\Delta\mathbf{r}'_{Nj}|]. \quad (3)$$

$\langle|\Delta\mathbf{r}|\rangle$ represents the mean coordinate error for the centrosymmetric model of the non-centrosymmetric structure. The value of the mean coordinate error for any other centrosymmetric model of the given structure [see (1)] would always be greater than that of the model defined in (2). Hence we shall refer to the model defined in (2) as the best centrosymmetric model (BCM, hereafter) of the given non-centrosymmetric structure. If we confine ourselves to only centrosymmetric models of the given non-centrosymmetric structure [see (1)], the value of any given type of $R$ index may be expected to be lowest for the BCM. Hence the value of the $R$ index pertaining to the BCM may be referred to as the minimum probable value. In this paper we shall derive the theoretical overall values of the different types of $R$ indices expected for a BCM of a non-centrosymmetric structure which is approximately centrosymmetric.

Table 1. *Theoretical expressions for the overall values of various types of R indices for truncated data*

| Index | Expression | Index | Expression |
|---|---|---|---|
| $\bar{R}_1(F_t)$ | $\frac{\sum_s f_s\langle\lvert y_N-y_N^c\rvert\rangle_s^t}{\sum_s f_s\langle y_N\rangle_s^t}$ | $\bar{R}_1(I_t)$ | $\frac{\sum_s f_s^2\langle\lvert y_N^2-(y_N^c)^2\rvert\rangle_s^t}{\sum_s f_s^2\langle y_N^2\rangle_s^t}$ |
| $\bar{R}_1(y_t)$ | $\frac{\sum_s\langle\lvert y_N-y_N^c\rvert\rangle_s^t}{\sum_s\langle y_N\rangle_s^t}$ | $\bar{R}_1(z_t)$ | $\frac{\sum_s\langle\lvert y_N^2-(y_N^c)^2\rvert\rangle_s^t}{\sum_s\langle y_N^2\rangle_s^t}$ |
| ${}_B\bar{R}_1(F_t)$ | $\frac{\sum_s f_s^2\langle(y_N-y_N^c)^2\rangle_s^t}{\sum_s f_s^2\langle y_N^2\rangle_s^t}$ | ${}_B\bar{R}_1(I_t)$ | $\frac{\sum_s f_s^4\langle[y_N^2-(y_N^c)^2]^2\rangle_s^t}{\sum_s f_s^4\langle y_N^4\rangle_s^t}$ |
| ${}_B\bar{R}_1(y_t)$ | $\frac{\sum_s\langle(y_N-y_N^c)^2\rangle_s^t}{\sum_s\langle y_N^2\rangle_s^t}$ | ${}_B\bar{R}_1(z_t)$ | $\frac{\sum_s\langle[y_N^2-(y_N^c)^2]^2\rangle_s^t}{\sum_s\langle y_N^4\rangle_s^t}$ |
| $\bar{R}_1^f(F_t)$ | $2\sum_s\left\langle\left\lvert\frac{y_N-y_N^c}{y_N+y_N^c}\right\rvert\right\rangle_s^t$ | $\bar{R}_1^f(I_t)$ | $2\sum_s\left\langle\left\lvert\frac{y_N^2-(y_N^c)^2}{y_N^2+(y_N^c)^2}\right\rvert\right\rangle_s^t$ |

Note: Let $\varphi(y_N, y_N^c)$ be a function of $y_N$ and $y_N^c$ and $\psi(y_N)$ be a function of $y_N$ then

$$\langle\varphi(y_N, y_N^c)\rangle_s^t=\left[\frac{2(1+D^2)}{\pi(1-D^2)}\right]^{1/2}\left\{\int_0^{\infty}\int_{y_t}^{\infty}\varphi y_N\exp\left[\frac{-[2y_N^2+(1+D^2)(y_N^c)^2]}{2(1-D^2)}\right]\times I_0\left[\frac{2Dy_Ny_N^c}{1-D^2}\right]dy_N\,dy_N^c\right\}\times\left\{\int_{y_t}^{\infty}y_N\exp\left[\frac{-y_N^2}{1-D^4}\right]I_0\left[\frac{D^2y_N^2}{1-D^4}\right]dy_N\right\}^{-1}$$

and

$$\langle\psi(y_N)\rangle_s^t=\left\{\int_{y_t}^{\infty}\psi y_N\exp\left[\frac{-y_N^2}{1-D^4}\right]I_0\left[\frac{D^2y_N^2}{1-D^4}\right]dy_N\right\}\times\left\{\int_{y_t}^{\infty}y_N\exp\left[\frac{-y_N^2}{1-D^4}\right]I_0\left[\frac{D^2y_N^2}{1-D^4}\right]dy_N\right\}^{-1}$$

It is relevant to note here that the model we are dealing with is of the complete type (*i.e.* $P=N$ and consequently $\sigma_1^2=1$). For such a model, an unnormalized $R$ index based on $|F|$ or $I$ and the corresponding index based on $|F|/\sigma_1$ or $I/\sigma_1^2$ become one and the same. Thus, for the present case (*i.e.* $P=N$), we would have

$$\begin{aligned}\bar{R}(F_t)&=\bar{R}_1(F_t), & {}_B\bar{R}(F_t)&={}_B\bar{R}_1(F_t),\\ \bar{R}^f(F_t)&=\bar{R}_1^f(F_t), & \bar{R}(I_t)&=\bar{R}_1(I_t),\\ {}_B\bar{R}(I_t)&={}_B\bar{R}_1(I_t), & \bar{R}^f(I_t)&=\bar{R}_1^f(I_t).\end{aligned} \quad (4)$$

Hence we need consider only one of the above pairs and we shall therefore confine ourselves to the normalized $R$ indices. Further, it can be shown that $\bar{R}_1^f(y_t)=\bar{R}_1^f(F_t)$ and $\bar{R}_1^f(z_t)=\bar{R}_1^f(I_t)$. Hence we need consider only one of the pairs $\bar{R}_1^f(y_t)$, $\bar{R}_1^f(F_t)$ and

$\bar{R}_1^f(z_t)$, $\bar{R}_1^f(I_t)$. Thus we need consider only ten types of $R$ indices which are given in Table 1.

The procedure for deriving the theoretical expressions of the overall values of $R$ indices for the present situation is similar to that used by Parthasarathy & Velmurugan (1981). It consists of expressing the defining relationships of $R$ indices in terms of the normalized variables $y_N$ and $y_N^c$, dividing the reflections into groups based on $(\sin\theta)/\lambda$ values, then summing over reflections in each given range of $(\sin\theta)/\lambda$ and then over the various ranges of $(\sin\theta)/\lambda$ into which the interval 0 to $[(\sin\theta)/\lambda]_{\max}$ has been partitioned. However, since some modifications in the procedure are necessary, we shall briefly outline the procedure for the index $\bar{R}_1(F_t)$ and summarize the final results for the other indices.

*Theoretical expression for* $\bar{R}_1(F_t)$

Let $|F_N|$ be the magnitude of the structure factor of a reflection $H$ $(=hkl)$ for the given structure and let $|F_N^c|$ be the corresponding calculated value for the model. Let $y_N$ and $y_N^c$ be the normalized structure-amplitude variables associated with $|F_N|$ and $|F_N^c|$ respectively. That is,

$$y_N = |F_N|/\sigma_N, \qquad y_N^c = |F_N^c|/\sigma_N, \tag{5}$$

where $\sigma_N$ is defined as

$$\sigma_N^2 = \langle|F_N|^2\rangle = \langle|F_N^c|^2\rangle. \tag{6}$$

By definition,

$$\bar{R}_1(F_t) = \sum_{hkl}^{t} \big||F_N| - |F_N^c|/\sigma_1\big| \Big/ \sum_{hkl}^{t} |F_N|. \tag{7}$$

Thus, following the procedure used by Parthasarathy & Velmurugan (1981) for obtaining equation (9) from equation (1) of their paper, we can show that (7) yields

$$\bar{R}_1(F_t) = \sum_s \sigma_{Ns} n_s \langle|y_N - y_N^c|\rangle_s^t \Big/ \sum_s \sigma_{Ns} n_s \langle y_N\rangle_s^t, \tag{8}$$

where $n_s$ is the number of reflections in the $s$th range of $(\sin\theta)/\lambda$. For further simplification of (8), we need to consider the properties of the probability density functions $P_t(y_N, y_N^c)$ and $P_t(y_N)$ valid for the present situation. If we partition the reciprocal space into equi-volume shells, then the different shells into which the reciprocal space has been partitioned would contain practically the same number of observed reflections. Hence we can rewrite (8) as

$$\bar{R}_1(F_t) = \sum_s \sigma_{Ns} \langle|y_N - y_N^c|\rangle_s^t \Big/ \sum_s \sigma_{Ns} \langle y_N\rangle_s^t. \tag{9}$$

For structures with similar atoms (9) can be rewritten as

$$\bar{R}_1(F_t) = \sum_s f_s \langle|y_N - y_N^c|\rangle_s^t \Big/ \sum_s f_s \langle y_N\rangle_s^t. \tag{10}$$

Unlike the situation in Parthasarathy & Velmurugan (1981), for the present situation $\langle y_N\rangle_s^t$ depends on $[(\sin\theta)/\lambda]$ [see (16)] and hence cannot be taken out of the summation symbol in the denominator of (10). Hence, in order to obtain the overall value $\bar{R}_1(f_t)$, we have to evaluate the numerator and the denominator in (10) separately. We shall consider this aspect presently.

*Expression for* $\langle y_N\rangle_s^t$

The probability density function (=p.d.f.) of $y_N$ is known to be (SS, 1975)

$$P(y_N) = \frac{2y_N}{(1-D^4)^{1/2}} \exp\left[\frac{-y_N^2}{1-D^4}\right] I_0\left(\frac{D^2 y_N^2}{1-D^4}\right), \qquad 0 \le y < \infty, \tag{11}$$

where $D$ is defined to be (Luzzati, 1953)

$$D = \langle\cos 2\pi \mathbf{H}\,.\,\Delta\mathbf{r}\rangle. \tag{12}$$

Luzzati (1953) also showed that $D$ can be written as

$$D = \exp\{-\pi^3 \langle|\Delta\mathbf{r}|\rangle^2 [(\sin\theta)/\lambda]^2\}. \tag{13}$$

The p.d.f. of $y_N$ for the truncated data $y_t \le y_N < \infty$ can be obtained from (11) as

$$P_t(y_N) = P(y_N)/\beta \tag{14}$$

where $\beta$ is defined to be

$$\beta = \int_{y_t}^{\infty} P(y_N)\, dy_N. \tag{15}$$

$\beta$ is a function of $y_t$, $\langle|\Delta\mathbf{r}|\rangle$ and $(\sin\theta)/\lambda$. For a given model (*i.e.* for a given $\langle|\Delta\mathbf{r}|\rangle$) and for a value of $y_t$, $\beta$ would depend only on $(\sin\theta)/\lambda$ [see (11) to (15)]. Hence, if we confine ourselves to reflections in a narrow region of $(\sin\theta)/\lambda$, $\beta$ can be taken to be a constant ($\beta_s$, say).

Using (14) we can obtain $\langle y_N\rangle_s^t$:

$$\langle y_N\rangle_s^t = \int_{y_t}^{\infty} y_N P_t(y_N)\, dy_N = (1/\beta_s) \int_{y_t}^{\infty} y_N P(y_N)\, dy_N. \tag{16}$$

Making use of (11) in (16), we obtain

$$\langle y_N\rangle_s^t = \left\{\int_{y_t}^{\infty} y_N^2 \exp\left[\frac{-y_N^2}{1-D^4}\right] I_0\left[\frac{D^2 y_N^2}{1-D^4}\right] dy_N\right\} \times \left\{\int_{y_t}^{\infty} y_N \exp\left[\frac{-y_N^2}{1-D^4}\right] I_0\left[\frac{D^2 y_N^2}{1-D^4}\right] dy_N\right\}^{-1}, \tag{17}$$

where we have cancelled out the common factors from the denominator and numerator. Equation (17) is to be evaluated numerically for any given value of the parameter $D$.

*Expression for* $\langle|y_N - y_N^c|\rangle_s^t$

The joint p.d.f. of $y_N$ and $y_N^c$ is known to be (SS, 1975)

$$P(y_N, y_N^c) = \left(\frac{2}{\pi}\right)^{1/2} \frac{2y_N}{1-D^2} \times \exp\left[\frac{-[2y_N^2 + (1+D^2)y_N^{c2}]}{2(1-D^2)}\right] \times I_0\left[\frac{2Dy_N y_N^c}{1-D^2}\right], \quad 0 \le y_N, y_N^c < \infty. \quad (18)$$

The joint p.d.f. of $y_N$ and $y_N^c$ valid for the truncated data $y_t \le y_N < \infty$ can be obtained from (18) as

$$P_t(y_N, y_N^c) = (1/\beta)P(y_N, y_N^c), \quad y_t \le y_N < \infty, \quad 0 \le y_N^c < \infty, \quad (19)$$

where $\beta$ is defined in (15) and $P(y_N, y_N^c)$ in (18). The expectation value of $|y_N - y_N^c|$ for the truncated data corresponding to a narrow range of $(\sin\theta)/\lambda$ will therefore be given by

$$\langle|y_N - y_N^c|\rangle_s^t = \int_0^\infty \int_{y_t}^\infty |y_N - y_N^c| P_t(y_N, y_N^c)\, dy_N\, dy_N^c = (1/\beta_s) \int_0^\infty \int_{y_t}^\infty |y_N - y_N^c| P(y_N, y_N^c)\, dy_N\, dy_N^c. \quad (20)$$

Making use of (15) and (18) in (20), we obtain

$$\langle|y_N - y_N^c|\rangle_s^t = \left[\frac{2(1+D^2)}{\pi(1-D^2)}\right]^{1/2} \left\{ \int_0^\infty \int_{y_t}^\infty \left\{ |y_N - y_N^c| y_N \times \exp\left\{\frac{-[2y_N^2 + (1+D^2)(y_N^c)^2]}{2(1-D^2)}\right\} \times I_0\left[\frac{2Dy_N y_N^c}{1-D^2}\right] dy_N\, dy_N^c \right\} \times \left\{ \int_{y_t}^\infty y_N \exp[-y_N^2/(1-D^4)] \times I_0[D^2 y_N^2/(1-D^4)]\, dy_N \right\}^{-1}, \quad (21)$$

where we have cancelled out the common factors from the denominator and numerator. Equation (21) is to be evaluated numerically.

The above considerations can be extended to the other types of $R$ indices. The final expressions obtained for the overall values of the various $R$ indices are therefore listed in Table 1 without details of derivation.

## 3. Method of obtaining overall value $\bar{R}$ as a function of $\langle|\Delta\mathbf{r}|\rangle$

The theoretical overall value of any $R$ index for a given $\langle|\Delta\mathbf{r}|\rangle$ and $y_t$ may be computed by the following procedure: (i) Partition the reciprocal space into $\nu$ equi-volume shells. (ii) Calculate the mean values of $(\sin\theta)/\lambda$ for these shells ($\bar{s}_i$, $i = 1, 2, \ldots, \nu$, say). (iii) Compute the values of $D$ corresponding to the various shells by using the values of $\bar{s}_i$, $i = 1$ to $\nu$ [see (13)]. Let these values of $D$ be denoted by $D_i$, $i = 1$ to $\nu$. (iv) Make use of the values of $D$ thus computed to calculate the local values corresponding to the different shells of the numerator and denominator quantities in the expression for $\bar{R}$. This involves numerical integration of the appropriate integrals shown at the bottom of Table 1. (v) Using the local values obtained in step (iv) calculate the overall values of the numerator and denominator quantities in the expression for $\bar{R}$. (vi) Make use of the overall values obtained in step (v) to obtain the overall value of the $R$ index using the appropriate expression in Table 1. The theoretical overall values of the various $R$ indices thus obtained are given as a function of $\langle|\Delta\mathbf{r}|\rangle$ in Table 2 for $y_t = 0{\cdot}0$, $0{\cdot}15$, $0{\cdot}30$, $0{\cdot}45$ and $0{\cdot}60$. It may be noted that the partititioning of the reciprocal space into $\nu$ equi-volume shells and the evaluation of $\bar{s}_i$, $i = 1$ to $\nu$, can be carried out by the procedure outlined by Elango & Parthasarathy (1990). The scattering factor needed as weights in the expressions for $\bar{R}$ is taken to be the same as that in the previous paper (Elango & Parthasarathy, 1990).

## 4. Test of the theoretical results

The theoretical results obtained in Table 2 were tested in a few cases. Two centrosymmetric structures (called *S*1 and *S*2) belonging to space group $P\bar{1}$ were taken to construct three different hypothetical non-centrosymmetric structures which are approximately centrosymmetric as follows: The structure of the dimethyl ester of *meso*-tartaric acid (Kroon & Kanters, 1973) (referred to as structure *S*1) was taken and random positional errors (following a Gaussian distribution) were introduced *independently* on the $(x, y, z)$ coordinates of all 24 atoms in the unit cell. This resulted in a non-centrosymmetric structure which is approximately centrosymmetric with 24 atoms in the asymmetric unit (the space group being $P1$ after the introduction of coordinate errors). Three independent sets of errors with $\langle|\Delta\mathbf{r}|\rangle = 0{\cdot}064$, $0{\cdot}162$ and $0{\cdot}227$ Å were thus introduced to the centrosymmetric structure *S*1 resulting in three different non-centrosymmetric structures (*A*1, *A*2 and *A*3, say) which are approximately centrosymmetric. Thus the

Table 2. *Theoretical overall values of the minimum probable values of R indices* (×1000) *for the best centrosymmetric model of an approximately centrosymmetric crystal as a function of* $\langle|\Delta\mathbf{r}|\rangle$ *corresponding to different values of* $y_t$

| Index | $y_t$ \ $\langle\|\Delta\mathbf{r}\|\rangle$ | 0.00 | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 | 0.12 | 0.14 | 0.16 | 0.18 | 0.20 | 0.22 | 0.24 | 0.26 | 0.28 | 0.30 | 0.35 | 0.40 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\bar{R}_1(F_t)$ | 0.00 | 0 | 4 | 25 | 57 | 91 | 127 | 161 | 195 | 227 | 258 | 288 | 315 | 342 | 367 | 390 | 411 | 459 | 499 |
| | 0.15 | 0 | 35 | 61 | 92 | 126 | 160 | 193 | 225 | 255 | 283 | 310 | 336 | 360 | 383 | 404 | 424 | 468 | 504 |
| | 0.30 | 0 | 31 | 57 | 87 | 120 | 151 | 182 | 211 | 239 | 266 | 292 | 317 | 340 | 362 | 382 | 401 | 444 | 479 |
| | 0.45 | 0 | 19 | 49 | 79 | 109 | 138 | 167 | 196 | 224 | 251 | 276 | 300 | 323 | 344 | 364 | 382 | 423 | 456 |
| | 0.60 | 0 | 27 | 45 | 65 | 90 | 118 | 146 | 174 | 202 | 229 | 254 | 279 | 302 | 323 | 343 | 362 | 403 | 435 |
| $\bar{R}_1(y_t)$ | 0.00 | 0 | 5 | 33 | 72 | 114 | 155 | 194 | 232 | 269 | 303 | 335 | 366 | 394 | 420 | 444 | 467 | 515 | 553 |
| | 0.15 | 0 | 40 | 70 | 108 | 148 | 188 | 226 | 261 | 294 | 325 | 355 | 382 | 408 | 432 | 454 | 475 | 519 | 554 |
| | 0.30 | 0 | 35 | 66 | 103 | 141 | 178 | 212 | 245 | 277 | 307 | 335 | 361 | 386 | 410 | 431 | 451 | 494 | 527 |
| | 0.45 | 0 | 23 | 59 | 94 | 128 | 162 | 196 | 229 | 261 | 291 | 319 | 345 | 369 | 392 | 413 | 432 | 472 | 503 |
| | 0.60 | 0 | 32 | 50 | 75 | 107 | 140 | 174 | 207 | 239 | 269 | 298 | 325 | 350 | 372 | 393 | 412 | 452 | 482 |
| $\bar{R}_1(I_t)$ | 0.00 | 0 | 0 | 9 | 29 | 57 | 91 | 129 | 168 | 209 | 251 | 292 | 333 | 374 | 413 | 451 | 488 | 574 | 650 |
| | 0.15 | 0 | 44 | 70 | 97 | 131 | 169 | 210 | 251 | 292 | 332 | 372 | 410 | 448 | 484 | 519 | 552 | 630 | 698 |
| | 0.30 | 0 | 43 | 69 | 102 | 141 | 179 | 217 | 254 | 291 | 327 | 363 | 398 | 433 | 466 | 498 | 529 | 601 | 664 |
| | 0.45 | 0 | 23 | 65 | 107 | 146 | 182 | 217 | 251 | 285 | 318 | 351 | 384 | 416 | 447 | 477 | 506 | 573 | 633 |
| | 0.60 | 0 | 33 | 59 | 83 | 113 | 148 | 183 | 219 | 254 | 289 | 323 | 356 | 388 | 419 | 449 | 478 | 543 | 599 |
| $\bar{R}_1(z_t)$ | 0.00 | 0 | 1 | 21 | 61 | 112 | 169 | 229 | 290 | 352 | 412 | 471 | 527 | 581 | 631 | 679 | 723 | 819 | 896 |
| | 0.15 | 0 | 58 | 88 | 136 | 195 | 258 | 319 | 378 | 434 | 488 | 539 | 589 | 636 | 680 | 722 | 761 | 846 | 914 |
| | 0.30 | 0 | 53 | 92 | 153 | 212 | 265 | 316 | 366 | 416 | 466 | 515 | 562 | 607 | 649 | 689 | 727 | 807 | 871 |
| | 0.45 | 0 | 37 | 105 | 159 | 208 | 255 | 305 | 355 | 404 | 452 | 498 | 543 | 585 | 625 | 662 | 696 | 770 | 827 |
| | 0.60 | 0 | 49 | 74 | 116 | 172 | 228 | 282 | 333 | 382 | 429 | 474 | 516 | 557 | 595 | 630 | 662 | 729 | 779 |
| ${}_B\bar{R}_1(F_t)$ | 0.00 | 0 | 0 | 1 | 5 | 12 | 20 | 30 | 42 | 55 | 70 | 86 | 102 | 119 | 137 | 155 | 173 | 217 | 259 |
| | 0.15 | 0 | 0 | 3 | 7 | 14 | 23 | 33 | 45 | 58 | 72 | 87 | 103 | 120 | 137 | 154 | 171 | 214 | 254 |
| | 0.30 | 0 | 0 | 3 | 7 | 13 | 21 | 30 | 41 | 53 | 66 | 80 | 95 | 110 | 126 | 142 | 159 | 199 | 237 |
| | 0.45 | 0 | 0 | 2 | 6 | 11 | 18 | 27 | 37 | 48 | 60 | 73 | 87 | 101 | 116 | 131 | 146 | 183 | 218 |
| | 0.60 | 0 | 0 | 2 | 4 | 9 | 15 | 23 | 32 | 42 | 53 | 66 | 79 | 92 | 106 | 120 | 134 | 169 | 202 |
| ${}_B\bar{R}_1(y_t)$ | 0.00 | 0 | 0 | 4 | 12 | 24 | 40 | 58 | 79 | 103 | 128 | 155 | 182 | 210 | 237 | 264 | 290 | 351 | 403 |
| | 0.15 | 0 | 1 | 5 | 15 | 27 | 43 | 62 | 83 | 105 | 129 | 155 | 181 | 207 | 234 | 260 | 285 | 344 | 393 |
| | 0.30 | 0 | 1 | 5 | 14 | 25 | 39 | 56 | 76 | 97 | 120 | 144 | 169 | 194 | 219 | 244 | 267 | 322 | 367 |
| | 0.45 | 0 | 1 | 5 | 12 | 22 | 35 | 51 | 70 | 90 | 112 | 135 | 158 | 181 | 205 | 227 | 249 | 299 | 340 |
| | 0.60 | 0 | 1 | 3 | 9 | 18 | 30 | 45 | 62 | 81 | 102 | 124 | 146 | 169 | 191 | 213 | 234 | 280 | 317 |
| ${}_B\bar{R}_1(I_t)$ | 0.00 | 0 | 0 | 0 | 0 | 2 | 5 | 9 | 16 | 24 | 33 | 45 | 58 | 73 | 89 | 107 | 127 | 181 | 241 |
| | 0.15 | 0 | 0 | 2 | 4 | 7 | 13 | 20 | 28 | 39 | 51 | 65 | 80 | 96 | 114 | 133 | 153 | 208 | 269 |
| | 0.30 | 0 | 0 | 1 | 4 | 9 | 15 | 23 | 32 | 43 | 54 | 67 | 80 | 95 | 110 | 127 | 145 | 194 | 250 |
| | 0.45 | 0 | 0 | 2 | 6 | 11 | 17 | 24 | 31 | 40 | 50 | 61 | 74 | 87 | 102 | 118 | 136 | 184 | 240 |
| | 0.60 | 0 | 0 | 1 | 3 | 6 | 12 | 19 | 27 | 36 | 47 | 60 | 73 | 88 | 103 | 120 | 137 | 182 | 232 |
| ${}_B\bar{R}_1(z_t)$ | 0.00 | 0 | 0 | 1 | 8 | 21 | 42 | 71 | 109 | 154 | 206 | 264 | 326 | 393 | 462 | 533 | 604 | 776 | 929 |
| | 0.15 | 0 | 1 | 6 | 19 | 39 | 68 | 101 | 140 | 183 | 231 | 286 | 345 | 410 | 477 | 546 | 615 | 782 | 928 |
| | 0.30 | 0 | 1 | 9 | 26 | 45 | 67 | 93 | 127 | 169 | 217 | 272 | 332 | 395 | 460 | 526 | 592 | 748 | 883 |
| | 0.45 | 0 | 2 | 12 | 24 | 39 | 61 | 90 | 126 | 167 | 213 | 264 | 319 | 378 | 438 | 499 | 560 | 701 | 822 |
| | 0.60 | 0 | 1 | 5 | 18 | 39 | 65 | 94 | 126 | 164 | 207 | 254 | 306 | 360 | 416 | 472 | 527 | 653 | 757 |
| $\bar{R}_1'(F_t)$ | 0.00 | 0 | 19 | 92 | 165 | 227 | 280 | 326 | 367 | 403 | 436 | 466 | 493 | 518 | 540 | 560 | 578 | 616 | 645 |
| | 0.15 | 0 | 59 | 124 | 188 | 246 | 296 | 339 | 378 | 413 | 444 | 472 | 498 | 521 | 543 | 562 | 579 | 616 | 644 |
| | 0.30 | 0 | 47 | 97 | 152 | 210 | 263 | 310 | 352 | 389 | 422 | 452 | 479 | 503 | 525 | 545 | 562 | 599 | 628 |
| | 0.45 | 0 | 28 | 70 | 116 | 167 | 220 | 270 | 317 | 359 | 396 | 430 | 459 | 486 | 510 | 531 | 550 | 589 | 618 |
| | 0.60 | 0 | 38 | 64 | 95 | 135 | 180 | 228 | 275 | 320 | 361 | 399 | 433 | 464 | 491 | 515 | 537 | 581 | 614 |
| $\bar{R}_1'(I_t)$ | 0.00 | 0 | 36 | 165 | 289 | 392 | 477 | 548 | 611 | 666 | 714 | 758 | 797 | 831 | 862 | 890 | 914 | 965 | 1003 |
| | 0.15 | 0 | 116 | 232 | 343 | 438 | 517 | 585 | 644 | 696 | 741 | 781 | 818 | 850 | 879 | 905 | 929 | 978 | 1014 |
| | 0.30 | 0 | 94 | 190 | 287 | 382 | 466 | 539 | 602 | 657 | 706 | 749 | 787 | 821 | 851 | 879 | 903 | 953 | 990 |
| | 0.45 | 0 | 56 | 138 | 226 | 314 | 400 | 479 | 549 | 612 | 666 | 714 | 756 | 794 | 827 | 856 | 882 | 934 | 973 |
| | 0.60 | 0 | 77 | 127 | 188 | 260 | 337 | 414 | 487 | 554 | 615 | 669 | 717 | 759 | 797 | 830 | 860 | 919 | 962 |

Table 3. *Test of the theoretical results for the best centrosymmetric models*

| | (*A*1, *S*1) | | (*A*2, *S*1) | | (*A*3, *S*1) | | (*B*1, *S*2) | | (*B*2, *S*2) | | (*B*3, *S*2) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *R* index | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ | $\bar{R}$ | $\langle\|\Delta\mathbf{r}\|\rangle$ |
| $\bar{R}_1(F_t)$ | 9·0 | 0·061 | 26·6 | 0·179 | 34·7 | 0·246 | 5·9 | 0·041 | 12·6 | 0·084 | 21·6 | 0·143 |
| $\bar{R}_1(y_t)$ | 10·4 | 0·061 | 30·0 | 0·175 | 39·3 | 0·245 | 6·8 | 0·041 | 14·2 | 0·081 | 24·2 | 0·138 |
| $\bar{R}_1(I_t)$ | 11·4 | 0·066 | 34·0 | 0·187 | 44·7 | 0·249 | 6·4 | 0·036 | 13·0 | 0·074 | 23·5 | 0·130 |
| $\bar{R}_1(z_t)$ | 15·9 | 0·062 | 46·7 | 0·180 | 65·3 | 0·262 | 9·1 | 0·040 | 18·7 | 0·071 | 33·8 | 0·129 |
| ${}_B\bar{R}_1(F_t)$ | 0·9 | 0·067 | 7·3 | 0·190 | 12·1 | 0·253 | 0·3 | 0·040 | 1·4 | 0·083 | 4·1 | 0·140 |
| ${}_B\bar{R}_1(y_t)$ | 1·6 | 0·063 | 11·9 | 0·178 | 20·8 | 0·250 | 0·5 | 0·036 | 2·3 | 0·075 | 6·7 | 0·131 |
| ${}_B\bar{R}_1(I_t)$ | 0·6 | 0·064 | 5·4 | 0·177 | 8·4 | 0·225 | 0·2 | 0·040 | 0·7 | 0·068 | 2·1 | 0·113 |
| ${}_B\bar{R}_1(z_t)$ | 3·2 | 0·066 | 24·8 | 0·191 | 53·8 | 0·283 | 0·6 | 0·031 | 2·7 | 0·061 | 10·1 | 0·124 |
| $\bar{R}_1'(F_t)$ | 15·2 | 0·060 | 43·0 | 0·185 | 52·1 | 0·256 | 11·1 | 0·045 | 25·0 | 0·095 | 39·6 | 0·164 |
| $\bar{R}_1'(I_t)$ | 28·5 | 0·059 | 70·9 | 0·181 | 83·7 | 0·250 | 21·4 | 0·045 | 44·2 | 0·094 | 65·7 | 0·160 |
| $\langle\|\Delta\mathbf{r}\|\rangle_{\text{est}}$ | | 0·063 | | 0·182 | | 0·252 | | 0·039 | | 0·079 | | 0·137 |
| $\langle\|\Delta\mathbf{r}\|\rangle_{\text{true}}$ | | 0·064 | | 0·162 | | 0·227 | | 0·041 | | 0·079 | | 0·131 |

Note: $R$ is in % and $\langle|\Delta\mathbf{r}|\rangle$ is in Å. $\langle|\Delta\mathbf{r}|\rangle_{\text{est}}$ is the average value in the respective columns.

centrosymmetric structure *S*1 would be the BCM for all these three non-centrosymmetric structures *A*1, *A*2 and *A*3. In each case the structure factor calculated using the known coordinates (of *A*1, *A*2 or *A*3) were taken to correspond to $F_N$. The structure factors calculated using the known coordinates of structure *S*1 were taken to correspond to $F_N^c$. The overall values of various *R* indices for the three cases, namely (i) (*A*1, *S*1), (ii) (*A*2, *S*1) and (iii) (*A*3, *S*1), were computed omitting reflections for which $y_N < 0{\cdot}3$ $(=y_t)$ and these are given in columns 2, 4 and 6 of Table 3. The theoretically expected values of $\langle|\Delta\mathbf{r}|\rangle$ were then estimated from the respective *R* values by interpolation using the results in Table 2 and the results thus obtained are given in columns 3, 5 and 7 respectively of Table 3. The average of the $\langle|\Delta\mathbf{r}|\rangle$ values thus obtained from the 'observed' overall values of the *R* indices are given in the row marked $\langle|\Delta\mathbf{r}|\rangle_{\text{est}}$ under the respective columns. The true values of $\langle|\Delta\mathbf{r}|\rangle$ for the three cases, namely (*A*1, *S*1), (*A*2, *S*1) and (*A*3, *S*1), are given in the last row marked $\langle|\Delta\mathbf{r}|\rangle_{\text{true}}$. A similar procedure was used in the case of the centrosymmetric structure of Hanson & Rohrl (1972) (referred to as structure *S*2) and three non-centrosymmetric structures (called *B*1, *B*2 and *B*3) with $\langle|\Delta\mathbf{r}|\rangle = 0{\cdot}041$, 0·079 and 0·131 Å, respectively, were generated. The relevant final results obtained for the three cases, namely (*B*1, *S*2), (*B*2, *S*2) and (*B*3, *S*2), are also summarized in Table 3. It is seen from Table 3 that there is reasonably good agreement between the corresponding values of $\langle|\Delta\mathbf{r}|\rangle_{\text{est}}$ and $\langle|\Delta\mathbf{r}|\rangle_{\text{true}}$ in all cases.

One of the authors (NE) thanks the University of Madras for financial assistance.

### References

ELANGO, N. & PARTHASARATHY, S. (1990). *Acta Cryst.* A**46**, 495–502.
HANSON, A. W. & ROHRL, M. (1972). *Acta Cryst.* B**28**, 2287–2291.
KROON, J. & KANTERS, J. A. (1973). *Acta Cryst.* B**29**, 1278–1283.
LUZZATI, V. (1953). *Acta Cryst.* **6**, 550–552.
PARTHASARATHY, S. & VELMURUGAN, D. (1981). *Acta Cryst.* A**37**, 472–480.
SWAMINATHAN, P. & SRINIVASAN, R. (1975). *Acta Cryst.* A**31**, 310–318.

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

# A Formula for Electron Density Histograms for Equal-Atom Structures

BY PETER MAIN

*Department of Physics, University of York, Heslington, York YO1 5DD, England*

(*Received* 21 *September* 1989; *accepted* 29 *January* 1990)

## Abstract

A formula is developed which gives the histogram of electron density values for polypeptide structures. The six parameters of the formula have been evaluated and are given for a range of resolutions from 4·5 to 0·9 Å. The formula may be used in density modification techniques of map improvement for small proteins.

## Introduction

Some recent papers (Zhang & Main, 1990*a*, *b*; Main, 1990) have described a method of determination and

0108-7673/90/060507-03$03.00 