NUESSLEIN, V. & SCHRODER, U. (1967). *Phys. Status Solidi*, **21**, 309-314.

PARKINSON, D. H. & QUARRINGTON, J. E. (1954). *Proc. Phys. Soc. (London) Sect. A*, **67**, 569-579.

PATHAK, P. D. & TRIVEDI, J. M. (1973). *Acta Cryst.* A**29**, 45-49.

PATOMÄKI, L. K. & LINKOAHO, M. V. (1969). *Acta Cryst.* A**25**, 304-305.

PIESBERGEN, U. (1966). In *Semiconductors and Semimetals*, edited by R. K. WILLARDSON & A. C. BEER, Vol. 2, pp. 49-60. New York: Academic Press.

REDMOND, A. D. & YATES, B. (1972). *J. Phys. C*, **5**, 1589-1603.

REEBER, R. R. (1974). *Phys. Status Solidi A*, **26**, 253-260.

ROEDHAMMER, P., WEBER, W., GMELIN, E. & REIDER, K. H. J. (1976). *J. Chem. Phys.* **64**, 581-585.

ROWE, J. M., NICKLOW, R. M., PRICE, D. L. & ZANIO, K. (1974). *Phys. Rev. B*, **10**, 671-675.

RUSAKOV, A. P., VEKILOV, Y. K. & KADYSHEVICH, A. E. (1970). *Fiz. Tverd. Tela*, **12**, 3238-3243. [*Sov. Phys. Solid State* (1971), **12**, 2618-2621.]

SHARKO, A. V. & BOTAKI, A. A. (1970). *Fiz. Tverd. Tela*, **12**, 2247-2249. [*Sov. Phys. Solid State* (1971), **12**, 1796-1798.]

SKELTON, E. L., RADOFF, P. L., BOLSAITIS, P. & VERBALIS, A. (1972). *Phys. Rev. B*, **5**, 3008-3012.

SMIRNOV, I. A. (1972). *Phys. Status Solidi A*, **14**, 363-404.

SUBHADRA, K. G. & SIRDESHMUKH, D. B. (1977). *Pramana*, **9**, 223-227.

SUBHADRA, K. G. & SIRDESHMUKH, D. B. (1978*a*). *Pramana*, **10**, 357-360.

SUBHADRA, K. G. & SIRDESHMUKH, D. B. (1978*b*). *Pramana*, **10**, 597-600.

VETELINO, J. F., MITRA, S. S. & NAMHOSKI, K. V. (1970). *Phys. Rev. B*, **2**, 2167-2175.

WALFORD, L. K. & SCHOEFFEL, J. A. (1970). *Philos. Mag.* **21**, 375-384.

WILLIS, B. T. M. & PRYOR, A. W. (1975). *Thermal Vibrations in Crystallography.* Cambridge Univ. Press.

ZENER, C. & BILINSKY, S. (1936). *Phys. Rev.* **50**, 101-104.

ZUBIK, K. & VALVODA, V. (1975). *Czech. J. Phys.* B**25**, 1149-1154.

Acta Cryst. (1988). A**44**, 142-144

Simple Statistics for Intensity Data from Twinned Specimens

BY T. O. YEATES

UCLA Department of Chemistry and Biochemistry and Molecular Biology Institute, Los Angeles, CA 90024, USA

(*Received* 6 *May* 1987; *accepted* 29 *September* 1987)

Abstract

The statistics of intensity data from hemihedrally twinned specimens are analyzed in terms of a new parameter and are shown to take a simple form in both the centrosymmetric and non-centrosymmetric cases. This analysis provides a sensitive method for determining the twinning fraction. The effects of intensity measurement errors on the observed statistics are discussed.

Introduction

When a crystal lattice posesses a rotational symmetry axis which is not a symmetry element of the space group of the crystal, crystal specimens may grow as merohedral twins. In this case, the reciprocal lattices of the different crystal twin domains of the specimen exactly overlap. The resulting diffraction intensities are given by linear combinations of the true untwinned intensities of reflections which are related by the twinning operation. In order to extract the true intensities from the observed intensities, one must be able to determine the twinning fraction (fractional volume of the specimen) for each of the separate twins. Several methods have been described for approximating the twinning fraction (Fisher & Sweet, 1980; Murray-Rust, 1973; Britton, 1972; Rees, 1980). A statistical treatment for hemihedral twinning (two twin domains) has been developed by Rees (1980). In that treatment, the statistical distribution for centrosymmetric reflections cannot be determined analytically. In addition , the intensities must be normalized prior to analysis. In the present treatment, these problems are avoided by deriving the statistics of a parameter H, which is a function of the two twin-related intensity measurements in hemihedral twinning. In the discussions which follow, we assume that the untwinned intensities obey Wilson's (1949) statistics and that intensities for untwinned reflections are independent.

Statistics for centrosymmetric reflections

Let

$$H=(q-p)/(q+p) \tag{1}$$

where p and q are intensity measurements of reflections related by the twinning operation. (Criteria by which weak pairs of reflections may be rejected without bias are discussed in the *Errors* section.) With a twinning fraction of α ($0<\alpha<\frac{1}{2}$),

$$p=(1-\alpha)|F_1|^2+\alpha|F_2|^2 \tag{2a}$$

and

$$q=\alpha|F_1|^2+(1-\alpha)|F_2|^2 \tag{2b}$$

0108-7673/88/020142-03$03.00

and

$$H=(|F_2|^2-|F_1|^2)(1-2\alpha)/(|F_2|^2+|F_1|^2) \quad (3a)$$

with

$$(2\alpha-1)<H<(1-2\alpha). \quad (3b)$$

The structure factor for a single reflection may be considered as a random walk of atomic scattering factors. In the continuous limit, the centrosymmetric structure factor F_1, and its twin-related reflection F_2, are random Gaussian variables (Srinivasan & Parthasarathy, 1976) of equal variance and zero mean. The joint probability distribution for $|F_1|$ and $|F_2|$, assuming statistical independence, is given by

$$P(|F_1|,|F_2|)=(4k/\pi)\exp[-k(|F_1|^2+|F_2|^2)] \quad (4)$$

where k is equal to $1/(2\langle|F|^2\rangle)$.

From (3a),

$$|F_2|=|F_1|[(1-2\alpha+H)/(1-2\alpha-H)]^{1/2} \quad (5)$$

and the cumulative distribution function for H, $S(H)$, may be obtained by integration of

$$S(H)=\int_{|F_1|=0}^{|F_1|=\infty}\int_{|F_2|=0}^{|F_2|=|F_1|[(1-2\alpha+H)/(1-2\alpha-H)]^{1/2}}(4k/\pi)$$
$$\times\exp[-k(|F_1|^2+|F_2|^2)]\,\mathrm{d}|F_2|\,\mathrm{d}|F_1| \quad (6)$$

in polar coordinates to give

$$S(H)=\arccos[H/(2\alpha-1)]/\pi. \quad (7)$$

The expected cumulative distribution $S(H)$ for centrosymmetric reflections is plotted for different values of the twinning fraction α in Fig. 1. The distribution of the parameter H can be calculated from observed intensity data and compared with the predicted distributions to obtain a value for α. Alternatively, the average of the absolute value of H or the average of the square of H may be calculated over all reflection pairs and compared with the expected results

$$\langle|H|\rangle=2(1-2\alpha)/\pi \quad (8a)$$

and

$$\langle(H^2)\rangle=(1-2\alpha)^2/2. \quad (8b)$$

Statistics for non-centrosymmetric reflections

We define H as before, and express p and q in terms of the untwinned intensities I_1 and I_2:

$$p=(1-\alpha)I_1+\alpha I_2 \quad (9a)$$

and

$$q=\alpha I_1+(1-\alpha)I_2 \quad (9b)$$

$$H=(q-p)/(q+p)=(I_2-I_1)(1-2\alpha)/(I_2+I_1) \quad (10)$$

with the range for H given by (3b). The non-centrosymmetric intensities I_1 and I_2 obey exponential statistics (Srinivasan & Parthasarathy, 1976). The joint probability distribution for I_1 and I_2 is given by

$$P(I_1,I_2)=k^2\exp[-k(I_1+I_2)], \quad (11)$$

where k is equal to $1/\langle|F|^2\rangle$.

From (10),

$$I_2=I_1(1-2\alpha+H)/(1-2\alpha-H). \quad (12)$$

The cumulative distribution for H, $S(H)$, can be obtained by integration:

$$S(H)=\int_{I_1=0}^{I_1=\infty}\int_{I_2=0}^{I_2=I_1[(1-2\alpha+H)/(1-2\alpha-H)]}k^2$$
$$\times\exp[-k(I_1+I_2)]\,\mathrm{d}I_2\,\mathrm{d}I_1$$
$$=[1+H/(1-2\alpha)]/2. \quad (13)$$

The expected cumulative distribution $S(H)$ for non-centrosymmetric reflections is linear and is plotted for different values of α in Fig. 2. Comparison with an observed distribution allows a determination of α. As before, the average of the absolute value of H and the average of the square of H can be calculated and compared with the expected results for non-centrosymmetric reflections.

$$\langle|H|\rangle=(1-2\alpha)/2 \quad (14a)$$

and

$$\langle(H^2)\rangle=(1-2\alpha)^2/3. \quad (14b)$$

Errors

Random errors in the observed intensities lead to error in the estimation of the twinning fraction.

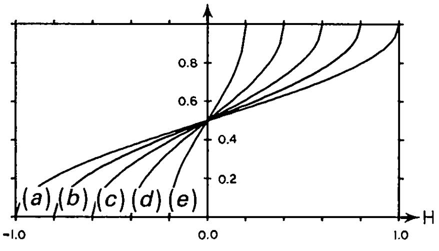

Fig. 1. Expected cumulative distribution of the parameter H for centrosymmetric reflections. (a) $\alpha=0{\cdot}0$, (b) $\alpha=0{\cdot}1$, (c) $\alpha=0{\cdot}2$, (d) $\alpha=0{\cdot}3$, (e) $\alpha=0{\cdot}4$. See text for definitions.

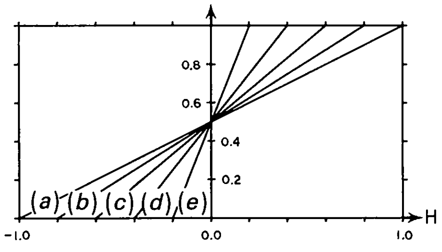

Fig. 2. Expected cumulative distribution of the parameter H for non-centrosymmetric reflections. (a) $\alpha=0{\cdot}0$, (b) $\alpha=0{\cdot}1$, (c) $\alpha=0{\cdot}2$, (d) $\alpha=0{\cdot}3$, (e) $\alpha=0{\cdot}4$. See text for definitions.

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

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

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

Concluding remarks

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

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

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

References

BRITTON, D. (1972). *Acta Cryst.* **A28**, 296-297.
FISHER, R. G. & SWEET, R. M. (1980). *Acta Cryst.* **A36**, 755-760.
MURRAY-RUST, P. (1973). *Acta Cryst.* **B29**, 2559-2566.
REES, D. C. (1980). *Acta Cryst.* **A36**, 578-581.
REES, D. C. & LIPSCOMB, W. N. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 277-280.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.

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

Use of the Information on Electron Density Distribution in Macromolecules

BY V. YU. LUNIN

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

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

Abstract

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

0108-7673/88/020144-07$03.00

Introduction

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

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

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