accurate prescription for structure-factor estimation (Mathieson, 1979, 1984*b*).

The small $CuInSe_2$ single crystal used in this study has proved to be a convenient, but in no way special, specimen with which to demonstrate some of the possibilities of the $\Delta\omega, \Delta 2\theta$ technique. The experimental observations have general significance for the modelling of crystals, since such models are mostly based on homogeneous mosaic spread (by contrast see Boehm, Prager & Barnea, 1974; Le Page & Gabe, 1978). The characteristics of the local mosaic distributions have proved to be quite diverse, even for the limited number of aspects in which the crystal has been viewed. Clearly, further work, with a variety of crystal specimens, is required to realize and appreciate the full potential of this technique.

One of us (AWS) acknowledges the financial support of a CSIRO Postdoctoral Award.

Note added in proof: A more accurate estimate of the diameter of Fig. 8(*c*) (see the end of § 3) can be obtained by carrying out a two-dimensional convolution of the functional forms associated with (μ), σ, λ, c and the detector aperture, *i.e.* a simulation of an $I(\Delta\omega, \Delta 2\theta)$ distribution. Such a calculation yields a value of ~0·051°.

References

BOEHM, J. M., PRAGER, P. R. & BARNEA, Z. (1974). *Acta Cryst.* A**30**, 335-337.

COMPTON, A. H. & ALLISON, S. K. (1935). *X-rays in Theory and Experiment.* New York: Van Nostrand.

DUISENBERG, A. J. M. (1983). *Acta Cryst.* A**39**, 211-216.

HOYT, A. (1932). *Phys. Rev.* **40**, 477-483.

LE PAGE, Y. & GABE, E. J. (1978). *J. Appl. Cryst.* **11**, 254-256.

MATHIESON, A. MCL. (1979). *Acta Cryst.* A**35**, 50-57.

MATHIESON, A. MCL. (1982). *Acta Cryst.* A**38**, 378-387.

MATHIESON, A. MCL. (1984*a*). *J. Appl. Cryst.* **17**, 207-209.

MATHIESON, A. MCL. (1984*b*). *Acta Cryst.* A**40**, 355-363.

MATHIESON, A. MCL. (1984*c*). *Aust. J. Phys.* **37**, 55-61.

MATHIESON, A. MCL. & STEVENSON, A. W. (1984). *Aust. J. Phys.* **37**, 657-665.

MATHIESON, A. MCL. & STEVENSON, A. W. (1985). *Acta Cryst.* A**41**, 290-296.

PARKES, J., TOMLINSON, R. D. & HAMPSHIRE, M. J. (1973). *J. Appl. Cryst.* **6**, 414-416.

Acta Cryst. (1986). A**42**, 230-240

On Absolute Scaling in Protein Crystallography using Sums of Low-Resolution Intensities and Wilson Statistics at Low Resolution

BY MICHEL ROTH

Institut Laue-Langevin, 156*X*, 38042 *Grenoble CEDEX, France*

(*Received* 11 *July* 1985; *accepted* 23 *December* 1985)

Abstract

A method of absolute scaling of diffraction data is proposed, based on the calculation of the sum of the intensity diffracted at low resolution (Bragg d spacing >15 Å). This sum is proportional to the mean-square deviation of the scattering-length density in the unit cell, and this property is used to determine the scale factor. The method is applied to the case of neutron diffraction using contrast variation experiments with biological molecules, and it is used to check the validity of some assumptions concerning the system under study, such as the global rate of H/D exchange or the uniformity of scattering-length density in the molecules. The use of this method requires an asymptotic correction of the sum of intensity. This correction is based on Porod's law, whose application to diffraction experiments is discussed, in particular for contrast variation experiments. An analysis of the spherical average of the diffracted intensity as a function of the scattering vector, compared to isotropic solution scattering, allows the conditions of applicability of Wilson statistics to be specified at low and medium resolution, *i.e.* the random statistical model underlying the Wilson statistics in this scattering range to be defined.

1. Introduction

In a structural study of complex molecules, such as biological macromolecules, by low-resolution neutron crystallography using H_2O/D_2O contrast variation, it is essential to know all the data concerning the contrast of all components of the system accurately. In practice, these values are usually calculated from the available information on the chemical composition of the components, making assumptions on the degree of H/D exchange on different sites within the components. Looking for a simple method

0108-7673/86/040230-11$01.50

that will allow one to check the consistency of these contrast calculations, we tried to use the variation with D_2O concentration in the solvent of the sum of the intensities measured at different contrasts after relative scaling (Roth, Lewit-Bentley & Bentley, 1984). A first sketch of the method was presented in the paper by Moras, Lorber, Romby, Ebel, Giege, Lewit-Bentley & Roth (1983). It turned out to be a method of absolute scaling of diffraction data, not necessarily restricted to contrast variation experiments.

The use of the sum of diffracted intensities as a means of absolute scaling of diffraction data has already been proposed by several authors (Kartha, 1953; Krogh-Moe, 1956; Norman, 1957; Rothbauer, 1978), as recalled by Giacovazzo (1980). Its usual application assumes that the measurements are performed to atomic resolution, *i.e.* to the limit of reciprocal space where the scattering power of atoms tends to zero because of their form factors (in the X-ray case) and because of the temperature factors. This is essential not only to ensure the convergence of the sums but also because the method relies on an atomic description of the scattering density. In the present paper we use the sum of intensities limited to low resolution only; this quantity depends mainly on the mean scattering density of the components of the molecule and of the solvent.

The intensity sum in low-resolution diffraction is equivalent to the integrated intensity in small-angle scattering, and its use for absolute scaling was proposed by Luzzati, Tardieu, Mateu & Sturhmann (1976) for X-ray small-angle scattering on biological solutions. The method was never widely used because the measurement of the forward-scattered intensity $I(0)$, made by extrapolating the intensity scattered at very small angles to zero angle using the Guinier law, leads to simpler results. It is unfortunately not transposable to diffraction experiments because of the paucity of data measured in the Guinier region. Diffraction is subject to sampling of intensity in reciprocal space on a grid too large with respect to the Guinier region (see Fig. 1).

Even though the intensities become, in general, weak towards the outer limit of the low-resolution region (*i.e.* around 0·3–0·4 $Å^{-1}$ in reciprocal space) compared to the intensities measured at smaller Bragg angles, their contribution to the sum is not negligible. The mean intensity at the limit of low resolution can be considered as consisting of two terms, one due to the global shape of the molecule and the other due to internal fluctuations of scattering-length density, provided there are no strong correlations between the two. The first intensity term decreases with Q^{-4} (Porod's law) under certain conditions and it is necessary to add the corresponding contribution extrapolated to infinite Q values to the sum of intensities actually measured. A by-product of the application of Porod's law in a contrast variation experiment is the possibility to determine the areas of interfaces between the constituents of the lattice. This type of analysis is currently used in small-angle-scattering applications in material sciences and was proposed a while ago by Luzzati, Witz & Nicolaieff (1961) for proteins. The second term of the mean intensity varies little with Q in this region of reciprocal space and its order of magnitude corresponds fairly well to what one would obtain from a random mixture of amino acids (whose scattering-length densities are very different from one another) in a protein. The contribution of this term has to be subtracted in the definition of the sum of low-resolution intensities we propose to use.

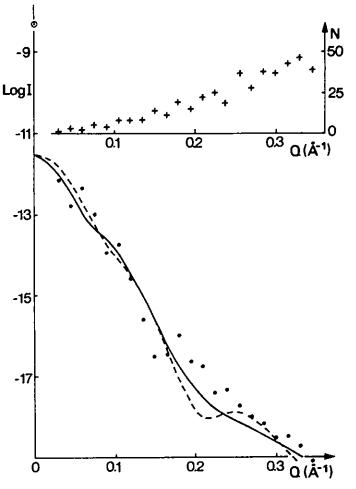

Fig. 1. Average intensity at low resolution. Example of averaged diffraction intensity as a function of the scattering vector (spherical average, see text § 3) compared to calculated small-angle scattering intensity. The experimental data (•) are from the neutron diffraction of a tRNA Asp-aspartyl tRNA synthetase crystal at match point of tRNA (Moras, Lorber, Romby, Ebel, Giege, Lewit-Bentley & Roth, 1983). The structure is body centered (*I*432). The systematically absent reflections of this space group were not taken into account in the average. The insert at the top indicates the number of unique reflections contributing to the average at each point. The point 0 on the ordinate scale represents the intensity diffracted coherently by the $N_m = 48$ molecules in the unit cell, calculated using (6.3). The intensity scattered incoherently at angle 0 by the same number of molecules is $N_m/2$ times smaller. The factor 1/2 comes from the body-centered symmetry [$M_p = 2$, see (6.4)]: the diffraction of two molecules related by a centering translation remains coherent in a spherical average, therefore each independent scatterer is constituted by M_p molecules, their number is N_m/M_p instead of N_m and their scattering power is M_p^2 times larger. This incoherent intensity at angle 0 is represented by the common origin of the solid and dashed lines. The two curves correspond to the isotropic small-angle scattering intensity calculated with two tentative models for the synthetase. These models were determined by a search in the unit cell and least-squares fit against the unaveraged diffraction data (agreement factor $R \simeq 40\%$). The conclusion is that to find a common scale between the averaged diffraction intensity and the isotropic small-angle scattering of a single molecule, one has to multiply the latter by M_pN_m.

In the first part, § 2, we shall establish the relationship between the sums of intensities and the parameters of contrast in the system. The second part, § 3, will discuss the application of Porod's law to a diffraction experiment with a multicomponent system, using contrast variation. The third part, § 4, deals with the corrections just mentioned and presents some tests of the method using model calculations. The fourth part, § 5, describes the method of calculation of the absolute scale factor and presents some results obtained in two experimental cases. In § 6, we will present a method of calculation of $F_{obs}(0)$ and discuss the application of Wilson's statistics to low resolution in protein crystallography, and terminate with a few remarks in § 7.

2. Theoretical outline – sum of intensity and contrast variation

2.1. *Sum of intensity and mean-square deviation of density*

We define the structure factor per unit cell as usual by

$$F(\mathbf{H}) = \int_{\text{unit cell}} \rho(\mathbf{r}) \exp(i\mathbf{H}\,.\,\mathbf{r})\,\mathrm{d}\mathbf{r}, \tag{2.1}$$

where $\mathbf{H}$ is a reciprocal-lattice vector and $\rho(\mathbf{r})$ is the scattering-length density, and the intensity by

$$I(\mathbf{H}) = F(\mathbf{H})F(\mathbf{H})^*. \tag{2.2}$$

We consider the sum s of the intensities over all reciprocal-lattice vectors except the origin:

$$s = \sum_{\mathbf{H}\neq 0} I(\mathbf{H}). \tag{2.3}$$

By using the inverse relation

$$\rho(\mathbf{r}) = (1/V) \sum_{\mathbf{H}} F(\mathbf{H}) \exp(-i\mathbf{H}\,.\,\mathbf{r}), \tag{2.4}$$

it can be shown that this sum is equal to

$$s = V^2 \overline{[\rho(\mathbf{r}) - \overline{\rho(\mathbf{r})}]^2}, \tag{2.5}$$

where the averages indicated by the bars are taken over the unit cell whose volume is equal to V.

2.2. *Introduction of step functions to describe the geometry of the components*

We shall assume that the system is a three-component system. It could be a biological molecule consisting of nucleic acid and protein in a solvent as a third component, or a protein associated with a detergent in a solvent, or a three-component three-dimensionally-ordered liquid crystal. The subsequent relations can be easily generalized to a larger number of components. With three components, the scattering-length density can be written in the following way (Sturhmann, 1982):

$$\rho(\mathbf{r}) = \sum_{i=1}^{3} [\rho_i + \delta\rho_i(\mathbf{r})]\Gamma_i(\mathbf{r}), \tag{2.6}$$

where $\Gamma_i(\mathbf{r})$ is a step function equal to 1 inside the component i and to 0 outside it, ρ_i is the average scattering-length density of the component i and $\delta\rho_i(\mathbf{r})$ is the local deviation of $\rho(\mathbf{r})$ with respect to ρ_i. With the assumption that there are no empty regions in the unit cell [*i.e.* $\sum_{i=1}^{3} \Gamma_i(\mathbf{r}) = 1$], the sum s can be expressed, according to (2.4), as

$$s = V^2\Big(\sum_{i=1}^{2} (\rho_i - \rho_3)^2 f_i(1-f_i) - 2(\rho_1-\rho_3)(\rho_2-\rho_3)f_1f_2 + \delta\rho^2 \Big) \tag{2.7}$$

with

$$\delta\rho^2 = \sum_{i=1}^{3} f_i\sigma_i^2, \tag{2.8}$$

where σ_i^2 is the average value of $\delta\rho_i(\mathbf{r})^2$ in the component i, and

$$f_i = (1/V) \int_{\text{unit cell}} \Gamma_i(\mathbf{r})\,\mathrm{d}\mathbf{r}, \tag{2.9}$$

f_i being the volume fraction of the component i in the unit cell. The factor $\rho_i - \rho_3$ is the contrast of the component i with respect to the solvent.

2.3. *Contrast variation*

In D_2O/H_2O contrast variation experiments, the structure factor $F(\mathbf{H})$ is a linear function of the D_2O concentration (Worcester, 1976) and is given by

$$F(\mathbf{H}) = F_0(\mathbf{H}) + c\Delta F(\mathbf{H}) \tag{2.10}$$

with

$$c = c_{D_2O}/(c_{H_2O} + c_{D_2O}), \tag{2.11}$$

where c_{D_2O} and c_{H_2O} are the molar concentrations of D_2O and H_2O respectively. As a consequence, s is a quadratic function of c, and can be written according to (2.7) as

$$s = V^2\Big(\sum_{i=1}^{2} \Delta\rho_i^2 f_i(1-f_i)(c-c_i)^2 - 2\Delta\rho_1\Delta\rho_2 f_1f_2(c-c_1)(c-c_2) + \delta\rho^2 \Big). \tag{2.12}$$

The following notations are used:

the linear variation of the contrast of the component i as a function of c is written

$$\rho_i - \rho_3 = \rho_{0i} + c\Delta\rho_i; \tag{2.13}$$

the corresponding concentration of zero contrast (match point) is given by

$$c_i = -\rho_{0i}/\Delta\rho_i. \tag{2.14}$$

As shown by Roth, Lewit-Bentley & Bentley (1984), the intensity $I_\Delta(\mathbf{H})$, which is equal to the square of the partial structure factor $\Delta F(\mathbf{H})$ in (2.10), can be determined directly from the diffraction data. It is the structure factor corresponding to the sites of exchangeable hydrogens in the system, *i.e.* the Fourier transform of $\Delta\rho_i\Gamma_i(\mathbf{r})$ [see (2.6), (2.13)] plus a minor contribution from the $\delta\rho_i(\mathbf{r})\Gamma_i(\mathbf{r})$ terms. The expression of the sum of these intensities $I_\Delta(\mathbf{H})$ as a function of contrast parameters is given by the coefficient of c^2 in (2.12).

The expression of the sum of intensity (2.7) differs from that given in an earlier paper (Moras, Lorber, Romby, Ebel, Giege, Lewit-Bentley & Roth, 1983), in addition to the fact that the $\delta\rho^2$ term was neglected there, because here the first term, $I(0)$, is excluded from the sum. The sum considered in the first paper is equal to $v^2\rho^2(\mathbf{r})$ instead of (2.5) because $I(0) = V^2\overline{\rho(\mathbf{r})}^2$, according to (2.1).

3. Porod's law in low-resolution diffraction

3.1. *Asymptotic behavior of mean intensity diffracted at low resolution*

As is well known in small-angle experiments on disordered systems, the small-angle scattering intensity of a system with sharp and randomly oriented interfaces decreases asymptotically with Q^{-4} for large values of Q, Q being the scattering vector [$Q = 4\pi \sin(\theta)/\lambda$ with 2θ = scattering angle and λ = neutron wavelength]. This is reasonably true on the condition that (1) the product QD_m, D_m being the smallest diameter of the scattering object, is larger than about 5 in the Q range under consideration (typically: $0{\cdot}25 < Q < 0{\cdot}4$ Å^{-1}, *i.e.* $25 > d > 15$ Å in Bragg d spacing), (2) the scattering density is homogeneous inside the components of the system, or, if not, the internal inhomogeneity does not produce rapid intensity variations in this Q range, (3) the interfaces do not show fractal properties (Bale & Schmidt, 1984). Experimental evidence of this asymptotic behavior of the intensity in small-angle-scattering experiments was given by Luzzati, Witz & Nicolaieff (1961), Tardieu, Mateu, Sardet, Weiss, Luzzati, Aggerbeck & Scanu (1976) and Gulik, Montheilhet, Dessen & Fayat (1976) for biological systems, Hendrikx & Charvolin (1981) and Cabane & Duplessix (1982) for lyotropic and micellar phases.

The same asymptotic behavior is observed in low-resolution diffraction experiments, not for the individual reflection intensities $I(\mathbf{H})$, but for their spherical averages, $I(Q)$. This average is calculated in spherical shells defined by a mean Q value and a width ΔQ:

$$I(Q) = \langle I(\mathbf{H}) \rangle \qquad (3.1)$$

for $Q - \Delta Q/2 < |\mathbf{H}| < Q + \Delta Q/2$.

The spherical averaging transforms the intensity diffracted coherently by the molecules in the lattice into an intensity distribution equivalent to that given by the small-angle scattering of randomly oriented single molecules free of interparticle interface. This is demonstrated in Fig. 1 and explained in more detail in the caption and discussed in § 6. This type of averaging is used currently for Wilson plots (Giacovazzo, 1980). The applicability of Porod's law to averaged diffraction data is shown in Figs. 2, 3 and 4.

At the origin of Porod's law is the fact that the Patterson function $P(R)$ in a system with sharp interfaces is a first-order linear decreasing function of R which can be written (spherical average)

$$P(R) = P(0) - (R/4) \sum_i \sum_j (\rho_i - \rho_j)^2 S_{ij}. \qquad (3.2)$$

Here the two sums are to be taken over all components, $(\rho_i - \rho_j)$ and S_{ij} are respectively the contrast and the area of the interface between the components i and j. The Q^{-4} variation of the intensity is obtained by a Fourier transformation of (3.2). For non-primitive lattices, *i.e.* lattices with centering translation symmetries, one also has to take into account similar first-order expansions of $P(R)$ in the vicinity of the points equivalent to the origin $R = 0$, by centering translation. The asymptotic expression of $I(Q)$ is then given by

$$I(Q) = (2\pi M_p Q^{-4}) \sum_i \sum_i (\rho_i - \rho_j)^2 S_{ij}, \qquad (3.3)$$

where M_p = (number of centering translations + 1). A similar expression for solution scattering was introduced first by Luzzati, Tardieu, Mateu & Sturhmann (1976) (see also Hendrikx & Charvolin, 1981).

3.2. *Contrast variation*

The application of (3.3) to a contrast variation experiment is straightforward. With a three-component system it reads:

$$I(Q) = (2\pi M_p Q^{-4}) \Big[\sum_{i=1}^{2} \Delta\rho_i^2 S_i (c - c_i)^2 - 2\Delta\rho_1 \Delta\rho_2 S_{12} (c - c_1)(c - c_2) \Big], \qquad (3.4)$$

where S_1 and S_2 are the external surface areas of the components 1 and 2 respectively, *i.e.*

$$S_1 = S_{13} + S_{12}, \qquad S_2 = S_{23} + S_{12}. \qquad (3.5)$$

The asymptotic expression for $I_\Delta(Q)$ is deduced from (3.4): it is given by the coefficient of c^2 in (3.4).

As can be seen from (3.4), the variations of $Q^4 I(Q)$ as a function of c are parabolic. Comparison with the parabolic variations of s with respect to c, (2.12), shows that the latter are governed by the values of

the volumes of the components (*i.e.* the volume fractions f_i), whereas the former are governed by the values of the interface areas. These two variations are usually not homotectic in experimental cases (see Figs. 5 and 6), which is direct evidence of a difference in the surface/volume ratio of the two components of the molecule. The concentration c of the minimum of the sum of intensities is, in most real cases where

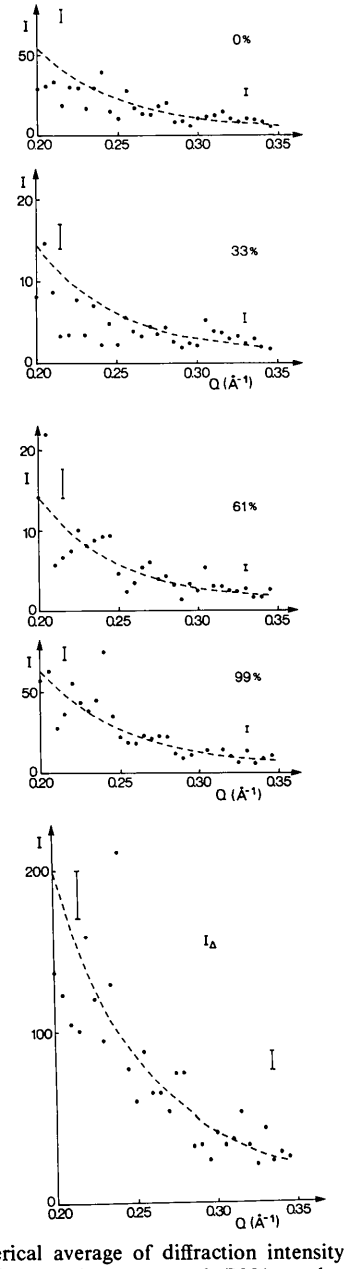

Fig. 2. Spherical average of diffraction intensity from different crystals of tRNA Asp - aspartyl tRNA synthetase complex at various contrasts. The contrasts are specified by the concentration c (2.11) indicated (% D_2O). The values of 33 and 61% are close to the concentrations of zero contrast of the synthetase and the tRNA respectively. The intensity $I_\Delta(\mathbf{H})$ is defined in § 2. The dashed curves represent the average Q^{-4} variations after fitting of (4.5), range of fit: $0{\cdot}25 < Q < 0{\cdot}35$ Å^{-1}. The magnitude of errors on intensity is given by the two error bars.

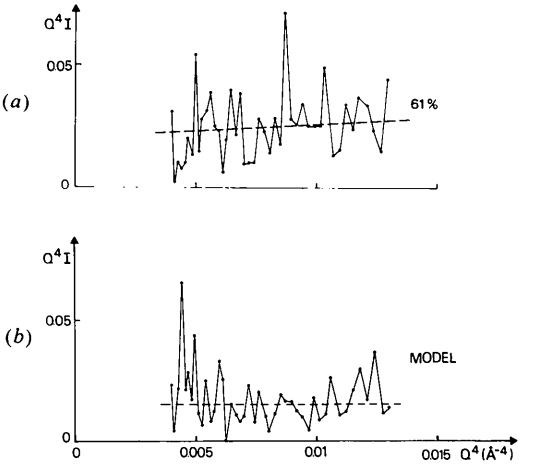

Fig. 3. $Q^4I(Q)$ *vs* Q^4 diagram of Porod's law with (*a*) experimental data (same as Fig. 1). The finite slope of the dashed straight line comes from the non-vanishing value of I_∞, (4.5). (*b*) Model data corresponding to the experimental data, calculated with a two-ellipsoid model for the molecule ($R \simeq 40\%$).

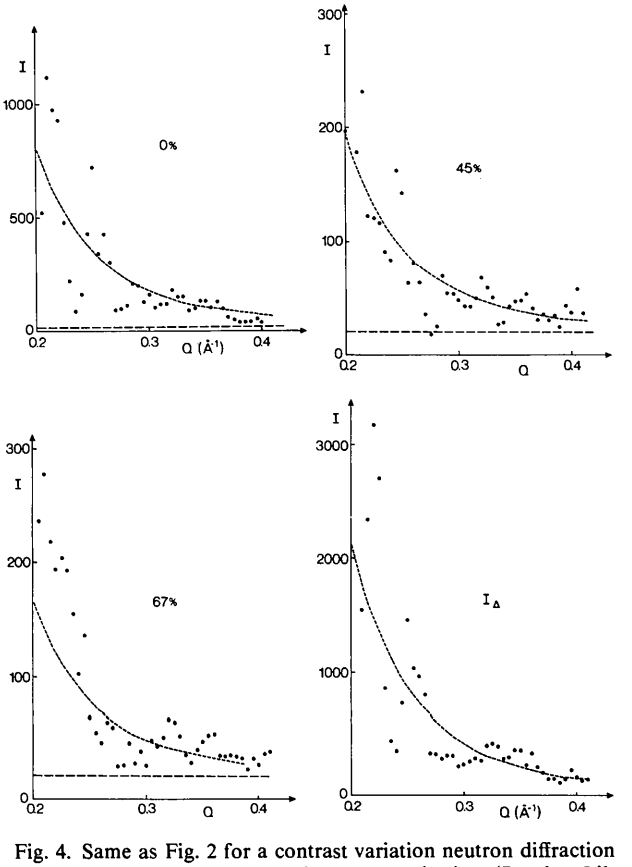

Fig. 4. Same as Fig. 2 for a contrast variation neutron diffraction experiment on the satellite tobacco necrosis virus (Bentley, Liljas, Lewit-Bentley, Roth, Skoglund & Unge, 1986). The oscillations of the average intensity are more pronounced than in the case of Fig. 2. They are due to the icosahedral shape of the protein capsid of the virus. The concentrations of 45 and 67% are close to the concentrations of zero contrast of the protein and the RNA respectively. The horizontal straight line represents the average value of I_∞.

$\Delta\rho_1 \simeq \Delta\rho_2$, approximately equal to the volume-weighted average $(c_1f_1+c_2f_2)/(f_1+f_2)$ of the zero-contrast concentrations c_1 and c_2 of the two components of the molecule, whereas the concentration of the minimum of the $Q^4I(Q)$ parabola is given by the average $(c_1S_{13}+c_2S_{23})/(S_{13}+S_{23})$, *i.e.* by an average weighted by the components/solvent interface areas. With the same approximation ($\Delta\rho_1 \simeq \Delta\rho_2$), the quantity $Q^4I_\Delta(Q)$ depends only on the external surface of the whole molecule $S_{13}+S_{23}$.

4. Practical use of the intensity sums

4.1. *Correction for series truncation*

Data collections are generally performed in a limited resolution range. The sum (2.3) has thus to be rewritten from a practical point of view:

$$s = \sum_{\mathbf{H}\neq 0}^{Q_2} I(\mathbf{H}) + C_p. \tag{4.1}$$

The summation over **H** is made inside the resolution shell $|\mathbf{H}| < Q_2$, where Q_2 is the resolution limit of data collection, and C_p is a truncation correction term. Assuming that Porod's law holds starting from a lower limit Q_1 smaller than Q_2, *i.e.* that according to (3.3) or (3.4)

$$I(Q) = A/Q^4 \tag{4.2}$$

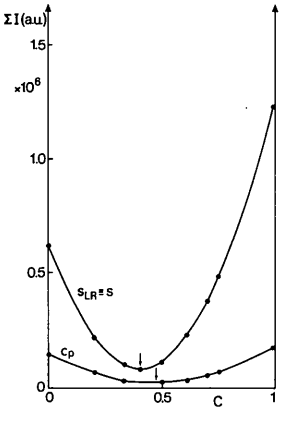

Fig. 5. Parabolic variation of the two sums of intensity s_{LR} (4.6) and s [(5.4), (5.6), (5.7) and (5.8)] after scaling using (5.9) (then $s_{LR}=s$), and of Porod's complement C_p (4.7) of s_{LR}, as a function of the concentration c (2.11) in the case of the neutron diffraction experiment with contrast variation on the tRNA Asp - aspartyl tRNA synthetase complex by Moras, Lorber, Romby, Ebel, Giege, Lewit-Bentley & Roth (1983). In that paper, a parabolic variation of the sum of intensity was shown, but the sum of intensity was not corrected for truncation by the C_p term, which is why the results differ somewhat from the present s_{LR} variations. At the time, the importance of this truncation correction was not well recognized and especially the fact that its variation with c is not proportional to that of the full sum s_{LR}.

for $Q > Q_1$, one finds, on the one hand,

$$C_p = [Q_1/(Q_1+Q_2)] \sum_{Q_1}^{Q_2} I(\mathbf{H}), \tag{4.3}$$

where the summation is made in the spherical shell, in the reciprocal space, defined by $Q_1 < |\mathbf{H}| < Q_2$ and, on the other hand,

$$C_p = 4\pi nA/Q_2. \tag{4.4}$$

The expression for A is given by (3.3) or (3.4) and n is the average density of reciprocal-lattice points in reciprocal space.

Results of model tests of this method of evaluation of the intensity sums, using (4.1) and (4.3), are given in Table 1. The models used were simple geometrical models with uniform density. The agreement is quite good. As can be seen in Fig. 3 (bottom), the average value of $Q^4I(Q)$ may show some oscillations as a function of Q. The choice of Q_1 and Q_2 has therefore to be made carefully in order to fit into such a true or pseudo periodicity by averaging over one or several (pseudo)periods.

4.2. *Correction for non-uniformity of the density*

In small-angle scattering experiments, it is often necessary to give the following form to the asymptotic behavior of $I(Q)$:

$$I(Q) = A/Q^4 + I_\infty \tag{4.5}$$

instead of (4.2). This takes into account the fact that the true measured intensity does not always tend to 0 when the small-angle-scattering effects vanish, but towards a finite value I_∞, representing background diffuse scattering from the sample (*e.g.* incoherent scattering, scattering from the solvent, or short-range atomic disorder). In diffraction experiments,

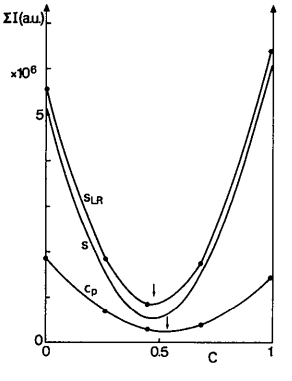

Fig. 6. Same as Fig. 5, for the case of the contrast variation experiment on STNV (Bentley, Liljas, Lewit-Bentley, Roth, Skoglund & Unge, 1986; same data series as Fig. 4). In this case s was scaled with respect to s_{LR} using the factor $K = K_a$ (see text § 5.2). The deviation between the curves s_{LR} and s is ascribed to the inhomogeneity of the scattering density in the components of the virus.

Table 1. *Results of tests of the method of calculation of the sum of the intensity diffracted at low resolution and of the determination of the external surface area, using geometrical models with uniform density*

Column (3) gives the radius of the spheres or the 1/2 axes of the ellipsoids constituting the model of the molecule.
Column (4) gives the volume fraction occupied by all these model molecules in the unit cell.
Column (5) gives the true values s_{th} of the sum of low-resolution intensity, in arbitrary units.
Column (6) gives the value of s calculated using (4.1) and (4.3).
Column (7) gives the value of the correction term C_p included in s.
Column (8) gives the Q_1, Q_2 range used in applying (4.3).
Column (9) gives the true external surface area of the model per sphere or ellipsoid.
Column (10) gives the same area determined using (4.9) and (5.2).

(1) Model	(2) Space group	(3) Dimensions (Å)	(4) f	(5) s_{th}	(6) s	(7) C_p	(8) Q_1-Q_2 (Å^{-1})	(9) S_{th} (Å^2)	(10) S (Å^2)
1 sphere	$I432$	25	0·071	65 800	66 700	9600	0·20-0·32	7850	8780
2 spheres	$I432$	27	0·178	551 200	555 400	77 600	0·20-0·33	9160	9750
2 ellipsoids	$I432$	16·9, 25·2, 39·2	0·151	201 200	205 800	33 400	0·25-0·31	8930	9200
	$I432$	22, 22, 45	0·197	551 200	555 400	75 200	0·20-0·33	10 600	10 460
1 ellipsoid	$P2_12_12_1$	20, 25, 50	0·177	812 900	815 900	105 500	0·25-0·35	11 850	12 250

however, with a proper background-correction procedure, all such background scattering is eliminated. It is nevertheless necessary to consider a finite I_∞, but here I_∞ represents the average intensity diffracted beyond low resolution, say at medium resolution. Its origin is related to the scattering-density fluctuations in the crystal on a scale finer than that of the overall molecular dimensions. These fluctuations contribute to the value of $\delta\rho_i(\mathbf{r})$ in (2.6), *i.e.* to $\delta\rho^2$ in (2.7). One way of taking this effect into account is to assume that these short-range fluctuations are distributed at random within the components of the molecule. Their average contribution to the intensity in the low-resolution range is then constant and equal to I_∞. This leads to the definition of a sum of low-resolution intensity, s_{LR}, which will be used in practice instead of s, as

$$s_{LR} = \sum_{\mathbf{H}\neq 0}^{Q_2} I(\mathbf{H}) + C_p - (4\pi/3) I_\infty n Q_2^3, \qquad (4.6)$$

by redefining C_p as

$$C_p = [Q_1/(Q_1+Q_2)] \times \left[\sum_{Q_1}^{Q_2} I(\mathbf{H}) - (4\pi/3) I_\infty n (Q_2^3 - Q_1^3)\right]. \qquad (4.7)$$

In so far as this assumption is valid, the values of s_{LR} will correspond to expressions like (2.7) or (2.12) where the term $\delta\rho^2$ is discarded. Their quantitative comparison allows the calculation of an absolute scale factor K of the diffraction experiment. This will be discussed in more detail in § 5.

4.3. *A random model to estimate* I_∞

At medium resolution, a protein or a nucleic acid can be considered as consisting of a large number of a few rigid units, the amino-acid residues and the components of the nucleotides (*i.e.* the sugar-phosphate groups and the bases or the base pairs). The effect of averaging the intensity in spherical shells is to cancel interparticle interference. The contribution to the average intensity of these units at medium resolution is given by the mean-square deviation of their total scattering lengths, the contribution of the mean value of these quantities (*i.e.* the contribution of the average scattering-length density of the components of the molecule) disappearing towards the end of the low-resolution range of diffraction. This is a consequence of the expression of the density (2.6). According to this model, I_∞ is given by

$$I_\infty = K[N_r\overline{(b_r - \overline{b_r})^2} + N_n\overline{(b_n - \overline{b_n})^2}], \qquad (4.8)$$

where K is the absolute scale factor of the data (§ 5), N_r and N_n represent the number of amino-acid residues and nucleotide components in the unit cell respectively and b_r and b_n the total scattering length of a given residue or nucleotide component respectively. The values of the b_r's and b_n's can be found in the papers of Worcester, Gillis, O'Brien & Ibel (1976) and Jacrot (1976). Their mean-square deviations, which appear in (4.8), depend on the composition of the given protein or nucleic acid. For numerical evaluations we used the example of the satellite necrosis tobacco virus protein coat (Ysebaert, van Emmelo & Fiers, 1980) as a typical representative of a protein and took an equal mixture of the four types of nucleotides *A*, *U*, *C*, *G* to calculate a value for a RNA. The values of these mean-square deviations were similar for both cases, *i.e.* 1·15 and 0·98 barn respectively, in H_2O (1 barn = 100 fm^2). As will be shown in §5.2, (4.8) gives the right order of magnitude compared to the values of I_∞ determined experimentally. The value of I_∞ is, in fact, dominated by the amino-acid residue contribution in these examples because $N_r \gg N_n$.

4.4. *Determination of interface areas*

There are several ways to determine the values of S_1, S_2 and S_{12}, using (3.4), (4.4) and (4.7), in practice. The simplest is to consider the value of C_p at the two

points of zero contrast c_1 and c_2 and at their mid-point $c_0=(c_1+c_2)/2$. One thus obtains:

$$S_1 = K'C_p(c_2)/\Delta\rho_1^2 \tag{4.9}$$

$$S_2 = K'C_p(c_1)/\Delta\rho_2^2 \tag{4.10}$$

$$S_{12} = K'\{2C_p(c_0) - [C_p(c_1)+C_p(c_2)]/2\}/(\Delta\rho_1\Delta\rho_2) \tag{4.11}$$

with

$$K' = KQ_2/[8\pi^2 M_p n(c_1-c_2)^2]. \tag{4.12}$$

These determinations depend in a sensitive manner on the values of the contrast parameters ρ_{0i} and $\Delta\rho_i$ and the volume fractions f_i, both directly and through the values of K, c_1 and c_2.

5. Absolute scale factor and discussion of results

5.1. *Absolute scale factor*

A direct application of the expression for the sums of intensity is the estimation of a scale factor for calculated intensities, I_{calc}, corresponding to a given model, with respect to the measured ones, I_{obs}. This integral scale factor is given by

$$K = V^2\overline{[\rho(\mathbf{r})-\overline{\rho(\mathbf{r})}]^2}/s, \tag{5.1}$$

where $\rho(\mathbf{r})$ is the model density. If it is the *real* density then K is the *absolute* scale factor. This method of calculating an absolute scale factor was first proposed by Kartha (1953) and then taken over by others (see *Introduction*). As a reference they used, for the case of X-rays, a density $\rho(\mathbf{r})$ consisting of the individual electron densities of the atoms, supposed to be non-overlapping. In the field of small-angle scattering, in material science, this kind of scaling has also been known for some time. Luzzati, Tardieu, Mateu & Stuhrmann (1976) proposed it in a form similar to (5.1) for biological solution scattering experiments. The expression for K is particularly simple at zero-contrast concentration of one component. For instance, for $c=c_2$,

$$K = V^2\Delta\rho_1^2(c_2-c_1)^2 f_1(1-f_1)/s_{\text{LR}}(c_2) \tag{5.2}$$

and K refers then to the model density $\rho(\mathbf{r})=\Delta\rho_1(c_2-c_1)\Gamma_1(\mathbf{r})$ of component 1. This kind of scale factor depending only on few simple parameters of the model (contrast and volume fraction) is sometimes very useful, for instance in the case of a systematic search of position, orientation and shape of a geometrical model in the unit cell of the crystal. In addition to speeding up the computing, it introduces an extra constraint in the calculation of the agreement factors between data and model, which makes the search more selective. This method was used as a first step in the location and modeling of the synthetase molecule (as a double ellipsoid) in the study of the complex tRNA Asp - aspartyl tRNA synthetase (Moras, Podjarny, Thierry, Rees, Giege, Lewit-Bentley & Roth 1986).

5.2. *Results for contrast variation*

In a contrast variation experiment, the experimental sum of the intensity s_{LR} (4.6) and its theoretical counterpart s (2.12) are quadratic functions of c, each defined by three coefficients:

$$s_{\text{LR}}(c) = A_e c^2 + B_e c + D_e \tag{5.3}$$

$$s(c) = A_t c^2 + B_t c + D_t. \tag{5.4}$$

If there are no errors in the values of the experimental concentrations c, and using the hypothesis that the scattering-density inhomogeneities inside the components of the macromolecule are short-range randomly distributed fluctuations, then the scaling of these two parabolas leads to the following equations:

$$K_a = A_t/A_e;\ K_b = B_t/B_e;\ K_d = D_t/D_e \tag{5.5}$$

and

$$A_t = V^2\left[\sum_{i=1}^{2}\Delta\rho_i^2 f_i(1-f_i) - 2\Delta\rho_1\Delta\rho_2 f_1 f_2\right] \tag{5.6}$$

$$B_t = -2V^2\left[\sum_{i=1}^{2}\rho_{0i}\Delta\rho_i f_i(1-f_i) - 2(\rho_{01}\Delta\rho_2+\rho_{02}\Delta\rho_1)f_1 f_2\right] \tag{5.7}$$

$$D_t = V^2\left[\sum_{i=1}^{2}\rho_{0i}^2 f_i(1-f_i) - 2\rho_{01}\rho_{02}f_1 f_2\right], \tag{5.8}$$

where K_a, K_b and K_d are three different determinations of the absolute scale factor K. In this ideal case, one should have of course

$$K_a = K_b = K_d = K. \tag{5.9}$$

Errors in the experimental concentrations c can be detected to some extent by comparing the concentrations of the minima of the two parabolas. In fact, although the values of the physical parameters in (5.6), (5.7) and (5.8) are not known with a very high precision, it is nevertheless possible to predict a rather narrow range (<10%) of values for the concentration of the minimum of (5.4). If such an error is suspected because of a large discrepancy between the minimum concentrations, one can use the least-squares procedure of relative scaling of the measured data to correct some of them (Roth, Lewit-Bentley & Bentley, 1984).

The parameters whose values are most uncertain in contrast variation experiments are $\Delta\rho_1$ and $\Delta\rho_2$ because in order to calculate them one needs to know the rate of H/D exchange on the labile H sites of the macromolecule, which depends on their accessibility to solvent as well as other factors such as pH and composition of the solvent. Equations (5.5) to (5.9) allow the determination of three unknowns. They can

thus be solved very easily with respect to $\Delta\rho_1$, $\Delta\rho_2$ and K.

This has been applied to the low-resolution neutron diffraction of crystals of the complex tRNA Asp - aspartyl tRNA synthetase described by Moras, Lorber, Romby, Ebel, Giege, Lewit-Bentley & Roth (1983). The quadratic variations of s_{LR} and C_p are shown in Fig. 5. The values of $\Delta\rho_1$ and $\Delta\rho_2$ found correspond to a rate of exchange of 100% for the tRNA and 72% for the synthetase. This differs from the results published in the above paper, for the reason explained in the caption of Fig. 5. The calculation of the concentrations of zero contrast gives $c_1 = 0{\cdot}674$, $c_2 = 0{\cdot}333$ and $c_0 = 0{\cdot}407$ for the tRNA, the synthetase, and the ensemble respectively. The total surface areas, S_1 and S_2, are found to correspond to an area of 6800 Å^2 per tRNA and 17 300 Å^2 per synthetase molecule, corresponding to a surface/volume ratio of 0·32 and 0·11 Å^{-1}, respectively. The interface between one tRNA and the synthetase is found to be equal to 1600 Å^2, *i.e.* about 25% of the total tRNA external surface. The absolute scale factor K corresponding to the data of Figs. 2 and 5 is equal to 0·336 Mbarn (using as scattering-length unit 10 fm). The value of I_∞ of about 0·4 (±0·4), found on the average for the different contrasts (Fig. 2) agrees quite well with the value 0·43 estimated using (4.8) (there are about 1200 residues per molecule of synthetase, and 75 nucleotides per tRNA).

The same method has been applied to data from a neutron diffraction experiment at low resolution on the satellite tobacco necrosis virus (STNV) (Bentley, Liljas, Lewit-Bentley, Roth, Skoglund & Unge, 1986) (see Figs. 4 and 6). Although it gave a reasonable value for $\Delta\rho_2$, the value found for $\Delta\rho_1$ was much too high. One can interpret this aberration by the fact that the hypothesis of a random distribution of scattering fluctuations does not apply in this case. The equalities (5.9) have to be abandoned: s_{LR} contains a contribution from these inhomogeneities not completely eliminated by the I_∞ correction (4.6) and not taken into account in (5.6), (5.7) and (5.8). Among the three possible determinations of K (5.5), the first one, involving the factors A_e and A_t is very likely to be the least biased by this effect. Indeed, the factors A_e and A_t correspond to the sum of intensities $I_\Delta(\mathbf{H})$ defined in § 2.3 and depend only on the partial structure factors $\Delta F(\mathbf{H})$ (2.10) of the exchangeable H sites. The influence on $I_\Delta(\mathbf{H})$ of scattering-density inhomogeneities in the molecule cannot be larger, in relative terms, than the square of the ratio of the number of exchangeable H in the molecules (*i.e.* about twice the number of amino-acid residues and nucleotides per unit cell) to the number of exchangeable H in the solvent (*i.e.* about twice the number of water molecules per unit cell). In the case of STNV, this upper limit is equal to about 2·6%. A confirmation of this property of $I_\Delta(\mathbf{H})$ is given by the fact that the value of $I_{\Delta\infty}$ found with experimental data is always practically equal to 0. The result of scaling with $K = K_a$ is shown in Fig. 6 ($K = 11{\cdot}6$ kbarn) assuming 100% exchange for the evaluation of $\Delta\rho_1$ and $\Delta\rho_2$. One finds thus 175 000 Å^2 as the area of one RNA and 5200 Å^2 as the external area of the protein per protein subunit [this value seems to be compatible with the subunit size estimation given by Unge, Liljas, Strandberg, Vaara, Kannan, Fridborg, Nordman & Lentz (1980) from electromicrographs] and 75 000 Å^2 as the area of the RNA/protein interface: this corresponds to a surface/volume ratio of 0·53 Å^{-1} for the RNA and 0·21 Å^{-1} for the protein and a RNA/protein contact of about 40% of the RNA surface and 25% of the protein surface. These numerical values represent only rough estimations of these areas because of the uncertainty in the contrast parameters. Again the order of magnitude, $I_\infty \simeq 20 \pm 20$, agrees quite well with the value of 10·5 predicted by (4.8).

6. Determination of $F_{obs}(0)$ and Wilson's statistics

Another application of the low-resolution sum of intensity calculation is the determination of $F_{obs}(0)$ for a given experimental data set $F_{obs}(\mathbf{H})$. This quantity is dependent on the model used to describe the system. If one uses a model where the solvent is eliminated from the calculation of $F(\mathbf{H})$ by considering only the density difference $\rho(\mathbf{r}) - \rho_3$ between the molecule and solvent [to define $F(\mathbf{H})$ with respect to $\rho(\mathbf{r}) - \rho_3$ instead of $\rho(\mathbf{r})$ (2.1) does not change the value of $F(\mathbf{H})$ for $\mathbf{H}$ not equal to 0], the corresponding value of $F_{obs}(0)$ is given by

$$F_{obs}(0) = V\left[\sum_{i=1}^{2} (\rho_i - \rho_3) f_i\right] K^{1/2} \qquad (6.1)$$

or

$$F_{obs}(0) = V\left[\sum_{i=1}^{2} \Delta\rho_i (c - c_i) f_i\right] K^{1/2} \qquad (6.2)$$

according to (2.6), (2.13) and (2.14). At the concentration of zero contrast for one of the components, $c = c_2$ for instance, this expression becomes simply

$$F_{obs}(0) = [s_{LR}(c_2) f_1 / (1 - f_1)]^{1/2}. \qquad (6.3)$$

This relation can be obtained directly using a convolution (Sayre's) equation applied to a system with two uniform density levels.

The use of a Wilson plot [*i.e.* the spherical average $I(Q)$ *vs* Q^2] to determine the same $F_{obs}(0)$ is theoretically possible. An example of spherical averaging of diffraction data is shown in Fig. 1. The result, as explained in its caption, can be interpreted in the following way. By contrast with the usual Wilson's statistics, one has to consider here that the elementary scatterer, assumed to be randomly distributed in the unit cell, is the entire macromolecule (or that part of

it that is not matched out by the solvent) and not the individual atoms. The corresponding coherent scattering factor is equal to the square root of the spherical average of the squared modulus of the Fourier transform of the difference between its scattering density and that of the solvent, *i.e.* it is the usual average small-angle-scattering form factor of the molecule in the solvent. Consequently, the role of the atomic form factor and temperature factor is taken over by this form factor. As explained in the caption to Fig. 1, the spherical average $I(Q)$ of the intensity $I(\mathbf{H})$ diffracted by the N_m molecules in the unit cell scales with the isotropic small-angle scattering of the N_m molecules, multiplied by the number of centering translation symmetry elements M_p. Let us call I_0 the value at $Q=0$ of $I(Q)$, obtained by a linear extrapolation to 0 of $I(Q)$ as a function of Q^2. This quantity is related to $F_{obs}(0)$ (6.1) by

$$I_0 = M_p N_m [F_{obs}(0)/N_m]^2. \qquad (6.4)$$

The practical use of this method is, however, very inaccurate because of scarcity and scatter of data near the origin of reciprocal space. The use of a relation like (6.1) or (6.2) and a value of K given by one of the relevant relations in § 5 [an example of this combination is given by (6.3)] will probably give a better result although the value of K may only be approximate.

7. Remarks

Finally, we would like to add the following remarks:

1. *Low-resolution diffraction and small-angle scattering.* According to kinematical theory of scattering, the proper definition of $F(\mathbf{H})$ should read, instead of (2.1):

$$F(\mathbf{H}) \int_{\text{unit cell}} [\rho(\mathbf{r}) - \langle\rho(\mathbf{r})\rangle_s] \exp(i\mathbf{H}\cdot\mathbf{r})\, d\mathbf{r} \qquad (7.1)$$

with

$$I(\mathbf{H}) = \langle F(\mathbf{H}) F^*(\mathbf{H})\rangle_s, \qquad (7.2)$$

where $\langle\ldots\rangle_s$ is the statistical or thermal average over the whole sample. The reason is that the first-order perturbation calculation, which is one of the basic approximations of the kinematical theory, cannot be applied to the average interaction potential over the whole sample (except for very small samples) (Landau & Lifshitz, 1981), but has to be applied to the difference $\rho(r) - \langle\rho(r)\rangle_s$. The expression (7.1) is the same as the expression for the scattering amplitude in small-angle scattering. The difference between (2.1) and (7.1) has no practical consequence because, for H not equal to zero, both relations give the same value of $F(\mathbf{H})$. The advantage of (7.1) is that it gives a unified expression for diffraction (and especially for low-resolution diffraction) and small-angle scattering on the same crystal. Its drawback is to introduce some useless complications in model calculations. The corresponding value of $F_{obs}(0)$ is related to imperfections in the crystal. It is proportional to the standard deviation of the statistical density fluctuations in the sample.

2. *External surface and Porod's law.* The application of Porod's law to such systems raises the question: which external surface is really measured? The answer may be found by remarking that Porod's law actually provides a measurement of the mean area of the projection of the object on a plane, assuming the object to be oriented at random and convex (if it is not convex, things are a bit more complicated but not basically different). The external area is taken as equal to twice the mean projected area, as is the case for a convex object oriented at random. From this, it is clear that surface details small compared to the Porod characteristic length, which is equal to four times the volume/surface ratio [see (3.2)], are washed out. But, on the other hand, this shows also that modeling a molecule by a smooth surface is a reasonable approximation at low resolution.

3. *Structural hierarchy and Wilson statistics.* The use of spherical averaging of the diffracted intensity has revealed a structural hierarchy in biological molecules from the diffraction point of view. If the results are interpreted in terms of Wilson statistics of elementary randomly distributed scatterers, in the low-resolution diffraction range ($Q < 0{\cdot}4$ Å^{-1}, *i.e.* $d >$ 15 Å), the elementary scatterers are the whole macromolecules. This is followed by a domain where the elementary scatterers are the amino-acid residues and the sugar–phosphate groups and bases or base pairs of the nucleic acids as shown by the results and discussion of I_∞ in §§ 4 and 5. The recognition of this hierarchy may help in modeling a macromolecule using data at progressively increasing resolution.

The neutron diffraction data referred to in this paper have been collected at the Institut Laue–Langevin (Grenoble, France) on the low-resolution diffractometer D17.

The author is very grateful to Drs D. Moras, R. Giege and co-workers, and to Drs L. Liljas, G. Bentley and co-workers for making their data available to him for the calculations presented here.

The author wishes to acknowledge sincerely the help given by Dr A. Lewit-Bentley and Dr G. Bentley during many discussions on these problems and for a critical reading of the manuscript.

References

BALE, H. D. & SCHMIDT, P. W. (1984). *Phys. Rev. Lett.* **53**, 596–599.

BENTLEY, G. A., LILJAS, L., LEWIT-BENTLEY, A., ROTH, M., SKOGLUND, M. & UNGE, T. (1986). In preparation.

CABANE, B. & DUPLESSIX, R. (1982). *J. Phys. (Paris)*, **43**, 1529–1542.

GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*, ch. 1. London: Academic Press.
GULIK, A., MONTEILHET, C., DESSEN, P. & FAYAT, G. (1976). *Eur. J. Biochem.* **64**, 295-300.
HENDRIKX, Y. & CHARVOLIN, J. (1981). *J. Phys. (Paris)*, **42**, 1427-1440.
JACROT, B. (1976). *Rep. Prog. Phys.* **39**, 911-953.
KARTHA, G. (1953). *Acta Cryst.* **6**, 817-820.
KROGH-MOE, J. (1956). *Acta Cryst.* **9**, 951-953.
LANDAU, L. D. & LIFSHITZ, E. M. (1981). *Quantum Mechanics*, 3rd ed., pp. 159-160. Oxford: Pergamon Press.
LUZZATI, V., TARDIEU, A., MATEU, L. & STUHRMANN, H. (1976). *J. Mol. Biol.* **101**, 115-127.
LUZZATI, V., WITZ, J. & NICOLAIEFF, A. (1961). *J. Mol. Biol.* **3**, 367-378, 379-392.
MORAS, D., LORBER, B., ROMBY, P., EBEL, J. P., GIEGE, R., LEWIT-BENTLEY, A. & ROTH, M. (1983). *J. Biomol. Struct. Dyn.* **1**, 209-223.
MORAS, D., PODJARNY, M., THIERRY, J. C., REES, B., GIEGE, R., LEWIT-BENTLEY, A. & ROTH, M. (1986). In preparation.
NORMAN, N. (1957). *Acta Cryst.*, **10**, 370-373.
ROTH, M., LEWIT-BENTLEY, A. & BENTLEY, G. A. (1984). *J. Appl. Cryst.* **17**, 77-84.
ROTHBAUER, R. (1978). *Acta Cryst.* A**34**, 528-533.
STUHRMANN, H. B. (1982). *Small Angle X-ray Scattering*, ch. 6: *Contrast Variation*, edited by O. GLATTER & O. KRATKY. London: Academic Press.
TARDIEU, A., MATEU, L., SARDET, C., WEISS, B., LUZZATI, V., AGGERBECK, L. & SCANU, A. M. (1976). *J. Mol. Biol.* **101**, 129-153.
UNGE, T., LILJAS, L., STRANDBERG, B., VAARA, I., KANNAN, K. K., FRIDBORG, K., NORDMAN, C. E. & LENTZ, P. J. JR (1980). *Nature (London)*, **285**, 373-377.
WORCESTER, D. L. (1976). *Brookhaven Symp. Neutron Scattering in Biology*, edited by B. SCHOENBORN, pp. III 37-III 57. New York: Brookhaven National Laboratory.
WORCESTER, D. L., GILLIS, J. M. O'BRIEN, E. J. & IBEL, K. (1976). *Brookhaven Symp. Neutron Scattering in Biology*, edited by B. SCHOENBORN, pp. III 101-III 113. New York: Brookhaven National Laboratory.
YSEBAERT, M., VAN EMMELO, I. J. & FRIERS, W. (1980). *J. Mol. Biol.* **143**, 273-287.

Acta Cryst. (1986). A**42**, 240-246

Exact Joint Distribution of E_h, E_k and E_{h+k}, and the Probability for the Positive Sign of the Triple Product in the Space Group $P\bar{1}$

BY URI SHMUELI

School of Chemistry, Tel Aviv University, Ramat Aviv, 69 978 Tel Aviv, Israel

AND GEORGE H. WEISS

Physical Sciences Laboratory, Division of Computer Research and Technology, National Institutes of Health, Bethesda, Maryland 20205, USA

(*Received* 29 *August* 1985; *accepted* 23 *December* 1985)

Abstract

A recently formulated method of deriving exact Fourier-series representations of joint probability density functions (p.d.f.'s) of several normalized structure factors is applied to the derivation of an exact expression for the conditional probability that the sign of the triple product $E_hE_kE_{h+k}$ is positive. The relevant joint and conditional probabilities are derived for the space group $P\bar{1}$. The Fourier coefficients of the p.d.f. are given by rapidly convergent series of Bessel functions, and the convergence properties of the Fourier summations are also found to be favourable. The exact conditional probability is compared with the currently employed approximate one, well known as the hyperbolic tangent formula, for several hypothetical structures. The examples illustrate the effects of the number of atoms in the unit cell, the magnitude of the E values and the atomic composition on the exact and approximate probabilities. It is found, in agreement with previous studies, that the hyperbolic tangent formula may indeed significantly underestimate the probability when the number of equal atoms is small and the E values are only moderately large, and when the structure contains outstandingly heavy atoms. The opposite behaviour, *i.e.* the approximate probability overestimating the exact one, was not observed in the present calculations. For large values of the triple product in equal-atom and heterogeneous models, the agreement between the approximate and exact probabilities is usually good.

Introduction

The well known hyperbolic-tangent formula, from which the probability for the positive sign of a triple-product structure invariant is conventionally estimated, is one of the most important relationships in applications of direct methods to sign determination. The current version of the relationship is based on

0108-7673/86/040240-07$01.50