was supported by NSF grant BBS8717949 and NIH grant GM33265.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.
BRAGG, W. L. & PERUTZ, M. F. (1952). *Proc. R. Soc. London Ser. A*, **213**, 425-435.
CASPAR, D. L. D. (1956). *Nature (London)*, **177**, 928.
FINCH, J. T. (1965). *J. Mol. Biol.* **12**, 612-619.
FRANKLIN, R. E. (1956). *Nature (London)*, **177**, 938-930.
FRANKLIN, R. E. & KLUG, A. (1955). *Acta Cryst.* **8**, 777-780.
KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.
MAKOWSKI, L. (1984). In *Biological Macromolecules and Assemblies.* Vol. I. *Virus Structures*, edited by F. A. JURNAK & A. MCPHERSON, pp. 203-253. New York: Wiley.
NAMBA, K., PATTANAYEK, R. & STUBBS, G. (1989). *J. Mol. Biol.* In the press.
WASER, J. (1955). *Acta Cryst.* **8**, 142-150.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318-321.
WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 397-399.
WINTER, W. T., ARNOTT, S., ISAAC, D. H. & ATKINS, E. D. T. (1978). *J. Mol. Biol.* **125**, 1-19.

Acta Cryst. (1989). A**45**, 258-260

R Factors in X-ray Fiber Diffraction. I. Largest Likely *R* Factors for *N* Overlapping Terms

BY R. P. MILLANE

The Whistler Center for Carbohydrate Research, Smith Hall, Purdue University, West Lafayette, Indiana 47907, *USA*

(*Received* 2 *June* 1988; *accepted* 19 *September* 1988)

Abstract

Simple expressions are obtained for the largest likely *R* factor in X-ray fiber diffraction recently derived by Stubbs [*Acta Cryst.* (1989), A**45**, 254-258]. These generalize the largest likely *R* factors obtained by Wilson [*Acta Cryst.* (1950), **3**, 397-399] for centric and acentric crystals. Expressions are obtained in terms of special functions and as finite series that simplify the calculation of *R* factors. These may be useful for further analysis and understanding of the effects of particle diameter and symmetry and diffraction data resolution on the reliability of structure determinations.

1. Introduction

The *R* factor is used routinely in crystallography to measure the reliability of structure determinations. Interpretation of the *R* factor obtained for a particular structure determination is aided by comparing it with the value for a completely wrong structure, *i.e.* a structure that is uncorrelated with the correct structure. This is referred to as the 'largest likely *R* factor', and Wilson (1950) showed that its value is $2\sqrt{2}-2=0{\cdot}828$ for a centric crystal and $2-\sqrt{2}=0{\cdot}586$ for an acentric crystal.

Recent advances in data collection and structure refinement in X-ray (and neutron) fiber diffraction analysis (Millane, 1988) have led to determinations of the structures of complex fibrous molecules and assemblies (Millane, Walker, Arnott, Chandrasekaran & Ratliff, 1984; Namba & Stubbs, 1985; Park, Arnott, Chandrasekaran, Millane & Campagnari, 1987; Stark, Glucksman & Makowski, 1988), and the *R* factor is used as a measure of the reliability of these structures also. The molecules in a fiber specimen are randomly oriented about the fiber axis so that the diffraction pattern is cylindrically averaged. The measured intensity is therefore equal to the sum of a number of different intensity terms diffracted by a single molecule. The number of terms in the sum depends on the maximum diameter and symmetry of the molecule, and the position in reciprocal space at which the intensity is measured. Since the measured intensities are sums of individual structure intensities, the *R* factor is in general smaller than in conventional crystallography. Stubbs (1989) has recently determined the largest likely *R* factor in fiber diffraction analysis as a function of the number of overlapping intensity terms. This allows the maximum value of the *R* factor for a particular structure determination to be estimated by averaging the values over the recorded diffraction pattern where the number of overlapping terms varies. This can be applied to both continuous diffraction from non-crystalline specimens and Bragg diffraction from polycrystalline specimens. The values obtained allow the *R* factor to be used for an objective assessment of the quality of structures determined by fiber diffraction.

Here, simple analytical and algebraic forms of Stubbs's (1989) expression for the largest likely fiber diffraction *R* factor are derived. These may be useful for further theoretical analysis of the dependence of the *R* factor on the number of overlapping intensity terms, the particle size and symmetry, and resolution of the diffraction data.

0108-7673/89/030258-03$03.00

2. Preliminaries

Putting Wilson's (1950) analysis in the context of fiber diffraction (Stubbs, 1989), one finds that the R factor for a random distribution of atoms is given by

$$R = 2 - 4\langle H(\mathscr{G})\rangle / \langle \mathscr{G} \rangle \tag{1}$$

where $\langle\ \rangle$ denotes the ensemble average in reciprocal space and $\mathscr{G}$ is the length of a $2N$-dimensional vector $\boldsymbol{\mathscr{G}}$ whose components are the real and imaginary parts of the complex Fourier-Bessel structure factor G_n (Klug, Crick & Wyckoff, 1958; Millane, 1988). N is the number of significant terms G_n that contribute to the recorded intensity and depends on the maximum diameter of the molecule and its symmetry (*via* the selection rule), and the position in reciprocal space of the intensity measurement. Any G_n terms that are real contribute only one dimension to $\boldsymbol{\mathscr{G}}$. In general, therefore, $\boldsymbol{\mathscr{G}}$ belongs to an m-dimensional space where

$$m = 2N - Q \tag{2}$$

and Q of the G_n terms are real. $H(\mathscr{G})$ is defined by

$$H(\mathscr{G}) = \int_0^{\mathscr{G}} xP(x)\,dx \tag{3}$$

where $P(\mathscr{G})$ is the probability density function for $\mathscr{G}$ so that $\langle \mathscr{G} \rangle = H(\infty)$ and

$$\langle H(\mathscr{G})\rangle = \int_0^{\infty} H(x)P(x)\,dx. \tag{4}$$

Stubbs (1989) showed that $P(\mathscr{G})$ is given by

$$P(\mathscr{G}) = (\pi\varepsilon)^{-m/2} V_m \mathscr{G}^{m-1} \exp(-\mathscr{G}^2/\varepsilon), \tag{5}$$

where ε can be estimated from the atomic scattering factors, and V_m is the $(m-1)$-dimensional surface area of the unit sphere in m-dimensional space. Using (1)-(5), Stubbs (1989) has calculated the largest likely R factor by numerical integration for m up to 16.

3. An analytical expression for the largest likely R factor

The expressions for $\langle \mathscr{G} \rangle$ and $\langle H(\mathscr{G})\rangle$ can be evaluated as follows. From (3) and (5),

$$\langle \mathscr{G} \rangle = \pi^{-m/2}\varepsilon^{1/2} V_m \int_0^{\infty} x^m \exp(-x^2)\,dx \tag{6}$$

and evaluation of the integral [Gradshteyn & Ryzhik (1980), equation (3.461)] gives

$$\begin{aligned} \langle \mathscr{G} \rangle &= \frac{\pi^{-m/2+1/2}\varepsilon^{1/2} V_m 2^{-m}(m-1)!}{[(m/2)-1]!}, && m \text{ even} \\ &= \tfrac{1}{2}\pi^{-m/2}\varepsilon^{1/2} V_m [(m/2)-1/2]!, && m \text{ odd.} \end{aligned} \tag{7}$$

Similarly, (3)-(5) give

$$\langle H(\mathscr{G})\rangle = \pi^{-m}\varepsilon^{1/2} V_m^2 \int_0^{\infty} x^{m-1} \exp(-x^2) \times \int_0^{x} y^m \exp(-y^2)\,dy\,dx. \tag{8}$$

Treating (x, y) as Cartesian coordinates and changing to polar coordinates (r, θ) allows (8) to be put in the form

$$\langle H(\mathscr{G})\rangle = \pi^{-m}\varepsilon^{1/2} V_m^2 \int_0^{\infty} r^{2m} \exp(-r^2)\,dr \times \int_0^{\pi/4} \cos^{m-1}\theta \sin^m\theta\,d\theta. \tag{9}$$

The r integral can be evaluated as above, and the θ integral by a trigonometric substitution, although it can be written more concisely in terms of the incomplete beta function $B_x(m, n)$ [Gradshteyn & Ryzhik (1980), equation (8.391)], giving

$$\langle H(\mathscr{G})\rangle = \pi^{-m+1/2}\varepsilon^{1/2} 2^{-2m-1} V_m^2 [(2m-1)!/(m-1)!] \times B_{1/2}[(m/2)+1/2, m/2]. \tag{10}$$

The surface area V_m can be determined by repeated application of Wallis's formula [Stubbs (1989); Abramowitz & Stegun (1972), equation (6.1.49)], giving

$$\begin{aligned} V_m &= 2\pi^{m/2}/[(m/2)-1]!, && m \text{ even} \\ &= 2^{m-1}\pi^{m/2-1/2}[(m/2)-3/2]!/(m-2)!, && m \text{ odd.} \end{aligned} \tag{11}$$

Combining (1), (7), (10) and (11), and denoting the largest likely R factor for m overlapping terms by R_m, gives (for any m)

$$R_m = 2 - 2^{-m+2} m \binom{2m-1}{m} B_{1/2}[(m/2)+1/2, m/2], \tag{12}$$

where the binomial coefficient

$$\binom{m}{n} = m!/[n!(m-n)!].$$

Equation (12) is a concise expression for the largest likely R factor. Evaluation of the special function can be avoided by using the algebraic expressions derived in the next section.

4. An algebraic expression for the largest likely R factor

An algebraic expression for R_m may be obtained by expressing the incomplete beta function in terms of the hypergeometric function $F(a, b; c; z)$

Table 1. *Exact values of the largest likely R factor, R_m, for m overlapping (real and imaginary) terms*

m	Exact	Approximate
1	$2\sqrt{2}-2$	0·828
2	$2-\sqrt{2}$	0·586
3	$\frac{7}{4}\sqrt{2}-2$	0·475
4	$2-\frac{9}{8}\sqrt{2}$	0·409
5	$\frac{107}{64}\sqrt{2}-2$	0·364
6	$2-\frac{151}{128}\sqrt{2}$	0·332
7	$\frac{835}{512}\sqrt{2}-2$	0·306
8	$2-\frac{1241}{1024}\sqrt{2}$	0·286
9	$\frac{26291}{16384}\sqrt{2}-2$	0·269
10	$2-\frac{40427}{32768}\sqrt{2}$	0·255

[Gradshteyn & Ryzhik (1980), equation (8.391)], giving

$$R_m = 2 - 2^{-3m/2+5/2} m (m+1)^{-1} \binom{2m-1}{m} \times F[(m/2)+1/2, 1-(m/2); (m/2)+3/2; 1/2]. \quad (13)$$

The power series for the hypergeometric function [Gradshteyn & Ryzhik (1980), equation (9.100)] in (13) terminates for m even, giving

$$R_m = 2 - 2^{-3m/2+5/2} m \binom{2m-1}{m} \times \sum_{n=0}^{m/2-1} (-1)^n (m+2n+1)^{-1} \binom{(m/2)-1}{n} 2^{-n}, \quad m \text{ even}. \quad (14)$$

For m odd, the series for the hypergeometric function in (13) does not terminate. However, using the relationships $B_x(m, n) = B(m, n) - B_{1-x}(n, m)$ and $B(m, n) = B(n, m)$, where $B(m, n) = B_1(m, n)$ is the beta function, one can put (12) in the form

$$R_m = 2 - 2^{-m+2} m \binom{2m-1}{m} \{B[m/2, (m/2)+1/2] - B_{1/2}[m/2, (m/2)+1/2]\}. \quad (15)$$

The beta functions in (15) can now be written as terminating hypergeometric series for m odd, and combined into a single series giving

$$R_m = 2 - 2^{-m+3} m \binom{2m-1}{m} \sum_{n=0}^{(m-1)/2} (-1)^n (m+2n)^{-1} \times \binom{(m/2)-1/2}{n} (1 - 2^{-m/2-n}), \quad m \text{ odd}. \quad (16)$$

Exact values for R_m can be calculated from (14) and (16) and are listed for m up to 10 in Table 1. The approximate numerical values so obtained agree with those calculated by Stubbs (1989) using numerical integration.

5. Concluding remarks

The R factors determined by Stubbs (1989) for a random distribution of atoms will be useful for assessing the reliability of fiber diffraction structure determinations. They may also prove useful in other applications where overlapped intensity data are used in structure determinations, such as in powder diffraction, solution scattering and others. Simple analytic and algebraic expressions for these R factors have been obtained. These may be useful for further theoretical development and quantitative understanding of the effects of particle size and symmetry, and resolution of the diffraction data, on the reliability of structure determinations. The algebraic expressions can be used to calculate values of R_m more easily than by using numerical integration, facilitating simpler calculation of R factors in particular cases.

I am grateful to Dr Gerald Stubbs for showing me his manuscript prior to publication. I also thank the US National Science Foundation for support (DMB-8606942), and Becky Hitt for word processing.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Table of Integrals, Series, and Products.* New York: Academic Press.
KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.
MILLANE, R. P. (1988). *Computing in Crystallography* 4: *Techniques and New Technologies*, edited by N. W. ISAACS & M. R. TAYLOR, pp. 169-186. IUCr/Oxford Univ. Press.
MILLANE, R. P., WALKER, J. K., ARNOTT, S., CHANDRASEKARAN, R. & RATLIFF, R. L. (1984). *Nucleic Acid Res.* **12**, 5475-5493.
NAMBA, K. & STUBBS, G. (1985). *Acta Cryst.* A**41**, 252-262.
PARK, H. S., ARNOTT, S., CHANDRASEKARAN, R., MILLANE, R. P. & CAMPAGNARI, F. (1987). *J. Mol. Biol.* **197**, 513-523.
STARK, W., GLUCKSMAN, M. J. & MAKOWSKI, L. (1988). *J. Mol. Biol.* **199**, 171-182.
STUBBS, G. (1989). *Acta Cryst.* A**45**, 254-258.
WILSON, A. J. C. (1950). *Acta Cryst.* **3**, 397-399.