COLEMAN, D. E. (1988). PhD thesis. Univ. of North Carolina at Chapel Hill, USA.
COLEMAN, D. E. & CARTER, C. W. JR (1984). *Biochemistry*, **23**, 381-385.
COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.
CRUMLEY, K. V. (1989). MS thesis. Univ. of North Carolina at Chapel Hill, USA.
DUMAS, C. (1988). Thèse DSN. Univ. de Paris-Sud, Centre D'Orsay, France.
FERSHT, A. R., ASHFORD, J. S., BRUTON, C. J., JAKES, R., KOCH, G. L. E. & HARTLEY, B. S. (1975). *Biochemistry*, **14**, 1-4.
GILMORE, C. J. (1984). *J. Appl. Cryst.* **17**, 42-46.
HAGE, F. (1986). *PROFL: a Computer Program for Integration and Display of Virtual Area Detector Films.* Department of Biochemistry, CB No. 7260, Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7260, USA.
HARKER, D. (1953). *Acta Cryst.* **6**, 731-736.
HARRISON, S. C. (1969). *J. Mol. Biol.* **42**, 457-483.
HARRISON, S. C. & JACK, A. (1975). *J. Mol. Biol.* **97**, 173-191.
IBEL, K. & STUHRMANN, H. B. (1975). *J. Mol. Biol.* **93**, 255-265.
IRWIN, M. J., NYBORG, J., REID, B. R. & BLOW, D. M. (1976). *J. Mol. Biol.* **105**, 577-586.
JACK, A., HARRISON, S. C. & CROWTHER, R. A. (1975). *J. Mol. Biol.* **97**, 163-172.
JACROT, B. (1976). *Rep. Prog. Phys.* **39**, 911-953.
MERLE, M., TREZEGUET, V., GRAVES, P. V., ANDREWS, D., MUENCH, K. H. & LABOUESSE, B. (1986). *Biochemistry*, **25**, 1115-1123.
MORAS, D., LORBER, B., ROMBY, P., EBEL, J.-P., GIEGE, R., LEWIT-BENTLY, A & ROTH, M. (1983). *J. Biomol. Struct. Dynam.* **1**, 209-223.
NARAYAN, R. & NITYANANDA, R. (1982). *Acta Cryst.* A**38**, 122-128.
PIRO, O. E. (1983). *Acta Cryst.* A**39**, 61-83.
PODJARNY, A., BHAT, T. N. & ZWICK, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351-374, and references contained therein.
PODJARNY, A., REES, B., THIERRY, J. C., CAVARELLI, J., JESIOR, J. C., ROTH, M., LEWIT-BENTLEY, A., KAHN, R., LORBER, B., EBEL, J. P., GIEGE, R. & MORAS, D. (1987). *J. Biomol. Struct. Dynam.* **5**, 187-198.
PRINCE, E., SJOLIN, L. & SVENSSON, L. A. (1988). *Acta Cryst.* A**44**, 218-222.
ROTH, M., LEWIT-BENTLEY, A. & BENTLEY, G. A. (1984). *J. Appl. Cryst.* **17**, 77-84.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60-65.
STUHRMANN, H. B. & KIRSTE, R. G. (1965). *Z. Phys. Chem.* **46**, 247-270.
WEBSTER, T. A., LATHROP, R. H. & SMITH, T. F. (1987). *Biochemistry*, **26**, 6950-6957.
WEISSMANN, L. (1982). In *Computational Crystallography*, edited by D. SAYRE, pp. 56-63. Oxford: Clarendon Press.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A**39**, 47-60.
WINTER, G. P., HARTLEY, B. S., MCLACHLAN, A. D., LEE, M. & MUENCH, K. H. (1977). *FEBS Lett.* **82**, 348-350.
WORCESTER, D. L. & FRANKS, N. P. (1976). *J. Mol. Biol.* **100**, 359-368.

Acta Cryst. (1990). A**46**, 68-72

R Factors in X-ray Fiber Diffraction. III. Asymptotic Approximations to Largest Likely *R* Factors

BY R. P. MILLANE

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

(*Received* 13 *June* 1989; *accepted* 21 *August* 1989)

Abstract

The largest likely R factor is useful for evaluating the significance of R factors obtained in structure determinations, and is smaller in fiber diffraction than in traditional crystallography. Very simple approximations to functions used to calculate the largest likely R factor in fiber diffraction are derived. For example, the largest R factor (R_m) for m overlapping terms is very well approximated by $R_m \simeq (2/\pi)^{1/2} m^{-1/2}$. These are a useful alternative to the exact, but quite complicated, expressions derived previously. More significantly, they provide insight into the behavior of R factors in fiber diffraction and may be useful in further analysis.

1. Introduction

The largest likely R factor (that for a structure uncorrelated with the correct structure) is a useful yardstick for evaluating the significance of R factors obtained in structure determinations. The largest likely R factor for single crystals was determined by Wilson (1950) and has recently been determined for fiber diffraction (Stubbs, 1989; Millane, 1989*a*, *b*). R factors in fiber diffraction are generally smaller than in single-crystal analyses because the diffraction pattern is cylindrically averaged. The R factor depends on the number of overlapping complex Fourier-Bessel structure factors at different positions in reciprocal space, and therefore on the diameter and symmetry of the diffracting particle and the maximum resolution of the diffraction data. The largest likely R factor in fiber diffraction, while easily calculated, is a rather complicated expression involving special functions (Millane, 1989*b*), making its interpretation obscure. Approximations to largest likely R factors in fiber diffraction are derived here by developing asymptotic approximations to components of this expression.

0108-7673/90/010068-05$03.00

This approximation is very simple and applications to typical structure determinations show that it is quite accurate.

Essential results for largest likely R factors in fiber diffraction are recalled in § 2 and asymptotic approximations to these expressions are derived in the following section. In § 4, the approximations are applied to a number of fiber diffraction analyses and the results discussed.

2. Preamble

The diffracting particles or crystallites in a fiber specimen are randomly rotated so that the diffraction pattern is cylindrically averaged. This means that R factors tend to be lower in fiber diffraction than in traditional macromolecular crystallography of single crystals, since each datum contains more degrees of freedom.

The largest likely R factor associated with a fiber diffraction pattern is given by (Millane, 1989*b*)

$$R = \sum_{m=1}^{M} N_m R_m S_m \Big/ \sum_{m=1}^{M} N_m S_m \qquad (1)$$

where the sums are over the different numbers of overlapping complex Fourier–Bessel structure factors (both real and imaginary parts) G_n (Klug, Crick & Wyckoff, 1958) that contribute to the different intensity measurements. N_m of the intensity measurements have m overlapping terms and M is the maximum value of m on the diffraction pattern. For a noncrystalline specimen, the intensity measurements are samples (along the layer lines) of the cylindrically averaged continuous transform of the diffracting particle. For a polycrystalline specimen, each measurement is a set of composite crystalline intensities (Bragg reflections). The R_m are the largest likely R factors if every measurement contained m overlapping terms (Stubbs, 1989), and are given by (Millane, 1989*a*)

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

where $\binom{m}{n}$ is the binomial coefficient and $B_x(m, n)$ is the incomplete beta function. Note that there is a minus sign missing in equation (6) of Millane (1989*b*). The S_m are proportional to the mean values of the amplitudes that contain m overlapping terms and are given by (Millane, 1989*b*)

$$S_m = \Gamma[(m/2)+1/2]/\Gamma(m/2) \qquad (3)$$

where $\Gamma(x)$ is the gamma function. For a particular diffraction pattern, the N_m can be easily determined and the largest likely R factor calculated using tabulated values of R_m and S_m (Millane, 1989*b*).

3. Asymptotic analysis

The expression (1) for the largest likely R factor contains the quantities R_m and S_m that are rather complicated functions of m. The object here is to obtain approximate expressions for R_m and S_m that are considerably simpler than the exact expressions (2) and (3), thereby providing insight into the behavior of R factors and easing any subsequent analysis. This is achieved by developing series for R_m and S_m that are asymptotic in m for $m \to \infty$.

An asymptotic approximation to R_m is obtained by first developing asymptotic expansions for the binomial coefficient and the incomplete beta function in (2). The binomial coefficient is given by

$$\binom{2m-1}{m} = \Gamma(2m)/\{m[\Gamma(m)]^2\}. \qquad (4)$$

Replacement of the gamma function by Stirling's formula and development of (4) as an asymptotic series gives

$$\binom{2m-1}{m} = 2^{2m-1}\pi^{-1/2}m^{-1/2}[1-(1/8m)+O(m^{-2})], \quad m \to \infty \qquad (5)$$

where $O(x)$ indicates terms of order x. The incomplete beta function is defined by [Gradshteyn & Ryzhik (1980), equation (8.391)]

$$B_{1/2}[(m/2)+1/2, m/2] = \int_0^{1/2} x^{m/2-1/2}(1-x)^{m/2-1}\,dx \qquad (6)$$

and with $p = m/2$ and $y = 1-2x$ it can be written as

$$B_{1/2}(p+1/2, p) = 2^{-2p+1/2}\int_0^1 (1-y)^{1/2}(1-y^2)^{p-1}\,dy. \qquad (7)$$

As $p \to \infty$, the integrand has significant value only in the vicinity of the origin, so that (7) can be put in the form

$$B_{1/2}(p+1/2, p) = 2^{-2p+1/2}\int_0^{\varepsilon} (1-y)^{1/2}(1-y^2)^{-1} \times \exp[p \ln(1-y^2)]\,dy, \quad \varepsilon \to 0, p \to \infty. \qquad (8)$$

Expansion of the terms in the integrand as power series and with $x = p^{1/2}y$ the integral can be written as

$$B_{1/2}(p+1/2, p) = 2^{-2p+1/2}p^{-1/2}\int_0^{\infty} H(x, p) \times \exp(-x^2)\,dx, \quad p \to \infty, \qquad (9)$$

where $H(x, p)$ is a polynomial in x (see Appendix), and the domain of integration can be made infinite because the integrand is exponentially small for large x. The integral in (9) can be evaluated [Gradshteyn

& Ryzhik (1980), equation (3.461)], for each term of $H(x, p)$, and changing from p back to m gives

$$B_{1/2}[(m/2)+1/2, m/2] = 2^{-m}m^{-1/2}[\pi^{1/2}-2^{-1/2}m^{-1/2} + (1/8)\pi^{1/2}m^{-1}+O(m^{-3/2})], \quad m\to\infty. \quad (10)$$

Substitution of (5) and (10) into (2) and development of an asymptotic series gives

$$R_m = (2/\pi)^{1/2}m^{-1/2}+O(m^{-3/2}), \quad m\to\infty, \quad (11)$$

which is the required approximation for R_m. Note that the m^{-1} term vanishes. The exact and leading-order asymptotic approximations to R_m are shown in Fig. 1. The maximum error is 0·01 for $m>3$ and 0·03 for all m. Extending the expansion to include the $m^{-3/2}$ term decreases its accuracy, a phenomenon quite common with asymptotic series, since they are not convergent.

An asymptotic series for S_m is easily calculated using Stirling's formula in (3), giving

$$S_m = 2^{-1/2}[m^{1/2}-(1/4)m^{-1/2}+O(m^{-3/2})], \quad m\to\infty. \quad (12)$$

The exact and leading-order ($m^{1/2}$) approximations to S_m are compared in Fig. 1. The maximum error for the leading-order approximation is 0·05 for $m>10$ and 0·14 for all m. Inclusion of the second term in (12) improves the accuracy (Fig. 1), the maximum error being 0·03 for all m.

With the leading-order behavior for R_m and S_m in (11) and (12), the leading-order approximation to R_mS_m is

$$R_mS_m = \pi^{-1/2}+O(m^{-1}), \quad m\to\infty, \quad (13)$$

which is also shown in Fig. 1, the maximum error being 0·01 for $m>5$ and 0·1 for all m. Since higher-order approximations to R_m do not improve the accuracy, useful higher-order approximations to R_mS_m cannot be derived formally. However, numerical calculations show that the approximation

$$R_mS_m \simeq \pi^{-1/2}[1-(1/8)m^{-1}] \quad (14)$$

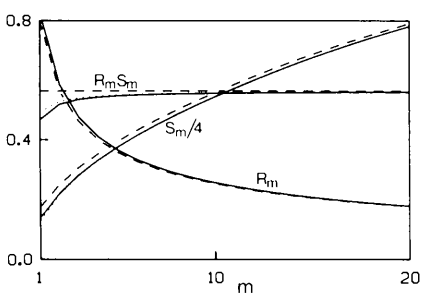

Fig. 1. Comparison of R_m, $S_m/4$ and R_mS_m (——) with their first-order (---) and second-order (· · ·) asymptotic approximations, as a function of the number of overlapping terms m.

gives increased accuracy over (13) as shown in Fig. 1, the maximum error being 0·03 for all m.

Equations (12), (13) and (14) give simple, but quite accurate, approximations to the quantities used in (1) to calculate largest likely R factors.

4. Examples and discussion

Almost all high-resolution studies of fibrous macromolecules involve at least two complex Fourier-Bessel terms ($m=4$) on the layer lines, and usually four to six terms ($m=8$ to 12).The asymptotic expressions derived here should therefore be accurate enough to calculate accurate largest likely R factors in such studies. Even low-resolution (8-5 Å) studies often involve $m\geq 4$ so that these approximations should be useful here also. To assess the utility of these approximations in actual applications, largest likely R factors for some typical structures were calculated using the asymptotic approximations derived here and compared with the values calculated exactly (Millane, 1989*b*). The structures used were the polysaccharide chondroitin 4-sulfate (Millane, Mitra & Arnott, 1983), a nucleic acid (Park, Arnott, Chandrasekaran, Millane & Campagnari, 1987), the helical virus TMV (Namba & Stubbs, 1985) and the bacteriophage Pf1 (Stark, Glucksman & Makowski, 1988). Three of these are based on continuous diffraction and one (chondroitin 4-sulfate) on a polycrystalline specimen. The number of overlapping terms at a particular position in reciprocal space was determined as described by Millane (1989*b*). To assess the use of the simplest expressions derived above, the approximate R factors were caluclated using the leading terms only for S_m and R_mS_m in (12) and (13) respectively. The exact (R) and approximate ($\hat{R}$) largest likely R factors for two resolutions for each structure are listed in Table 1. The error in using the approximate expressions is less than 0·01 in all cases, which is ample accuracy in applications. Additional calculations show that using the (more accurate) second-order approximations in (12) and (14) does not necessarily increase the accuracy of calculated R factors. This is because the errors in the leading-order approximations to S_m and R_mS_m have the same sign and therefore tend to cancel in the quotient in (1), whereas the second-order terms, although more accurate, have opposite signs and the errors tend to add in (1).

The accuracy of the expressions derived here is further assessed by comparing the exact and approximate R factors for a hypothetical noncrystalline specimen with 10_1 helix symmetry, a maximum radius of 10 Å and c repeat of 20 Å, as a function of diffraction data resolution ρ_{max}. The results of these calculations (Fig. 2) show that the asymptotic approximations predict the largest likely R factor

Table 1. *Exact (R) and approximate ($\hat{R}$) largest likely R factors for four structures*

Molecule	Helix symmetry	Maximum radius (Å)	c repeat (Å)	Minimum resolution (Å)	Maximum resolution (Å)	M	R	$\hat{R}$
K⁺ C-4-S	3_2	7·0	27·8	∞	4·0	4	0·519	0·507
K⁺ C-4-S	3_2	7·0	27·8	∞	3·0	6	0·489	0·478
DNA	10_1	10·0	32·3	∞	3·0	10	0·413	0·407
DNA	10_1	10·0	32·3	∞	2·5	10	0·387	0·382
TMV	49_3	90·0	69·0	10·0	5·0	10	0·373	0·367
TMV	49_3	90·0	69·0	10·0	3·0	16	0·307	0·304
Pf1	27_5	30·0	75·6	10·0	5·0	6	0·458	0·448
Pf1	27_5	30·0	75·6	10·0	3·0	10	0·381	0·375

The K⁺ C-4-S data are based on a polycrystalline specimen (trigonal unit cell with $a = b = 13{\cdot}8$ Å and space group $P3_221$), and the other structures on noncrystalline specimens (continuous diffraction). References for these structures are listed in the text.

with an error not exceeding 0·01, except at very low resolution where only a few Bessel terms are involved.

5. Concluding remarks

Very simple approximations to the functions used to calculate largest likely R factors in fiber diffraction have been obtained. Although these are asymptotic expansions valid for large numbers of overlapping terms, they are very accurate, even for small numbers of terms. Largest likely R factors for typical structures calculated using the (simplest) leading-order approximations are quite accurate. The utility of these results is perhaps more that they provide insight into the behavior of fiber diffraction R factors and provide a simpler basis for further analysis than for actual calculation of R factors in particular cases, since this can be done accurately using tabulated values for the functions involved (Millane, 1989*b*).

An ideal result would describe (analytically) the dependence of the largest likely R factor on parameters such as symmetry and resolution. Although the analysis presented here substantially simplifies this problem, it falls short of a complete solution because of the complicated form of N_m. Work on this problem is continuing.

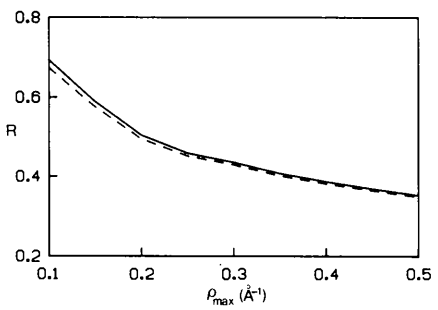

Fig. 2. Variation of the largest likely R factor, for a structure with 10_1 helix symmetry, a maximum radius of 10 Å and c repeat of 20 Å, with maximum resolution of the diffraction data ρ_{max}. The solid curve is the exact R factor, and the broken curve the R factor calculated using the leading-order asymptotic approximations.

I am grateful to the US National Science Foundation for support (DMB-8606942) and to Deb Zerth for word processing.

APPENDIX
Derivation of $H(x, p)$

The derivation of $H(x, p)$ in (9) is outlined here. Referring to the integrand in (8), we use the following approximations:

$$(1-y)^{1/2} = 1-(1/2)y-(1/8)y^2+O(y^3), \qquad y\to 0 \qquad (A.1)$$

$$(1-y^2)^{-1} = 1+y^2+O(y^4), \qquad y\to 0 \qquad (A.2)$$

$$\ln(1-y^2) = -y^2-(1/2)y^4-(1/3)y^6+O(y^8), \quad y\to 0. \qquad (A.3)$$

Use of (*A*.3) shows that

$$\exp[p\ln(1-y^2)] = \exp(-py^2)[1-p(y^4/2+y^6/3)+\ldots], \quad y\to 0. \qquad (A.4)$$

Letting $x = p^{1/2}y$, referring to (8) and (9), and using the above results, we find that

$$H(x,p) = 1-(1/2)p^{-1/2}x+p^{-1}[(7/8)x^2-(1/2)x^4] + O(p^{-3/2}), \quad x\to 0, \quad p\to\infty. \qquad (A.5)$$

References

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

MILLANE, R. P. (1989*b*). *Acta Cryst.* A**45**, 573-576.

MILLANE, R. P., MITRA, A. K. & ARNOTT, S. (1983). *J. Mol. Biol.* **169**, 903-920.

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.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

Acta Cryst. (1990). A**46**, 72

On integrating the techniques of direct methods with anomalous dispersion: the one-phase structure seminvariants in the monoclinic and orthorhombic systems. III. Primitive non-centrosymmetric space groups of type 1*P*220. Erratum.

By D. VELMURUGAN and HERBERT A. HAUPTMAN, *Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, New York* 14203-1196, *USA*

(*Received* 16 *October* 1989)

Abstract

There are two errors in the short communication by Velmurugan & Hauptman [*Acta Cryst.* (1989). A**45**, 656]. On line 9 in the *Abstract*, $\Phi_{2h'2k'0}$ should read $\Phi_{2h,2k,0}$ and on line 3 in *Summary of final results*, $\Phi_{2h'2k'0}$ should read $\Phi_{2h,2k,0}$.

All relevant information is given in the *Abstract*.

Acta Cryst. (1990). A**46**, 72

The International Union of Crystallography: its formation and early development. Erratum.

By HARMKE KAMMINGA,* *Department of History and Philosophy of Science, King's College London, Chelsea Campus, Manresa Road, London SW3 6LX, England*

(*Received* 14 *November* 1989)

Abstract

There is an unfortunate error in the caption to the figure on page 585 of the article by Kamminga [*Acta Cryst.* (1989). A**45**, 581-601]. The photograph was taken at the University of Leeds during the symposium held on 18-19 July 1946 and not 1948 as stated in the article.

* Present address: Department of History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge CB2 3RH, England.

All relevant information is given in the *Abstract*.