Acta Cryst. (1959). **12**, 66

The Stereochemistry of Deoxyribonucleic Acid. II. Hydrogen-Bonded Pairs of Bases

By M. Spencer

Medical Research Council Biophysics Research Unit, King's College, London, England

(*Received* 3 *July* 1958)

Probable structures for the purines and pyrimidines found in deoxyribonucleic acid are used in an attempt to define the limits of the pairs of bases assumed by Watson & Crick (1953) to exist in the structure. It is found that the limits depend very markedly on whether or not slight distortions are allowed in making the pairs equivalent. Probable base pairs for use in a model can, however, be quite accurately defined. The alternative base pairs suggested by Donohue (1956) are also investigated, and an estimate made of the distortion involved in forming equivalent pairs. Other considerations are summarized which indicate that the pairs are unlikely to be present in DNA. In the course of the discussion it is shown that one cannot reject on stereochemical grounds the possibility that a pyrimidine nucleoside may take up a different configuration to that normally assumed in models of DNA. There is, however, no evidence that it exists.

Introduction

Part I of this paper (Spencer, 1959) has summarized the reasons for requiring accurate prior knowledge of covalent bond lengths and angles in the components of a model of crystalline deoxyribonucleic acid (DNA), and likely values were suggested. Although the lengths of covalent bonds may be fairly accurately predicted, it is impossible at present to predict exact lengths for the hydrogen bonds which are assumed to be formed between the bases. One can only say that they probably lie within certain limits (Donohue, 1952). It is important to know the effect of this uncertainty on the range of models that can be built. In the first part of this paper an attempt is made to define the probable limits of the pairs of bases proposed by Watson & Crick (1953).

There is every indication that it will ultimately be possible to predict exactly all the observed X-ray diffraction from models similar to that of Watson & Crick (1953). The information available from X-ray diffraction patterns of DNA is, however, limited. It is not necessarily possible to deduce from it alone that certain alternative structures might not be made to explain the observed diffraction. One of the alternatives is that the bases are joined together by different systems of hydrogen bonds to those proposed by them. It is clear that structures embodying this alternative would not necessarily contradict any of the known facts about the structure of DNA, nor would they affect the mechanism by which the molecule is believed to carry and to transmit genetic information. The question has been considered by Donohue (1956) and discussed briefly by Crick (1957*a*). Donohue showed that an alternative way of pairing adenine with thymine and guanine with cytosine is possible. He stated that the formation of equivalent pairs involved some distortion from the regular structures assumed for the bases. The second part of this paper is concerned with making a roughly quantitative estimate of the degree of distortion involved in this scheme. Finally, certain other considerations which affect the likelyhood of Donohue's scheme as a basis for the structure of DNA are set down.

Watson-Crick pairs of bases

It will be assumed in what follows that the four bases found in DNA have the probable structures set down in Part I of this paper. It will also be assumed that hydrogen bonds may have lengths lying in the ranges found by Donohue (1952) in a survey of experimental results. The ranges found were 2·68 to 3·17 Å for N–H · · · O and 2·94 to 3·37 Å for N–H · · · N, and a large number of values which have since been reported also lie within these ranges. It might be thought that the wide range of possible lengths for the hydrogen bonds would render pointless the use in a model of highly accurate data for covalent bonds. However, diffraction patterns indicate that DNA molecules have a highly regular structure. It has been pointed out by Crick & Watson (1954) that this implies that the glycosidic links of one pair are equivalent to those of the other, and that the links in each pair are related by an axis of symmetry or diad. These requirements impose restrictions on the lengths of the hydrogen bonds in a model of DNA.

The configuration of a pair of bases will be described by the distance d between the carbon atoms C_1 and C_1' of the deoxyribose rings which are linked to the bases, and by the angles θ_1 and θ_2 between the glycosidic links and the line C_1C_1'. For exact equivalence of the links θ_1 and θ_2 must be equal. In the Watson–Crick pairs the diad lies in the plane of the bases. The first problem to be investigated concerns the range of values of d and θ over which exactly equivalent pairs of bases may be formed. The second problem concerns

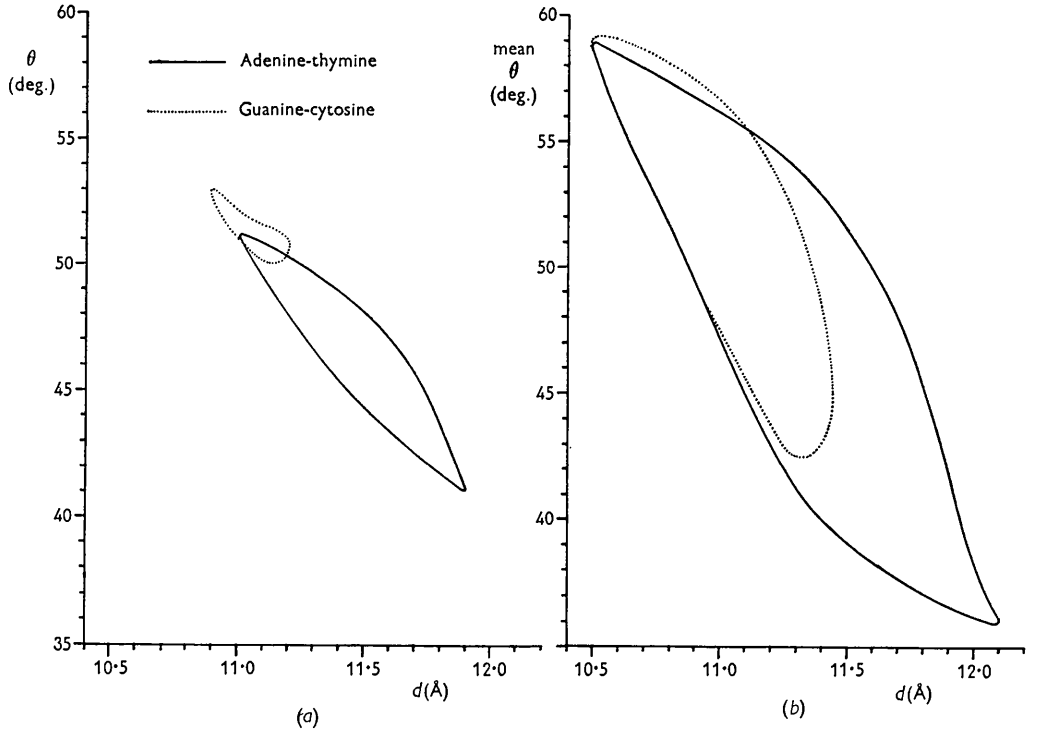

Fig. 1. Limiting dimensions of Watson–Crick base pairs (*a*) assuming invariable structures, (*b*) allowing slight distortions. d = distance between deoxyribose carbon atoms linked to bases; θ = mean angle between glycosidic links and line joining these atoms.

the extension of this range when certain slight distortions are permitted.

Exactly equivalent pairs

It will be assumed that guanine and cytosine are joined by three hydrogen bonds. Watson & Crick (1953) originally proposed that they formed two, but they mentioned the possibility that three were formed, and Pauling & Corey (1956) have since shown that it is reasonable that three should form. It will also be assumed that the N–H bond involved in a hydrogen bond may make an angle of up to 10° with the line of the hydrogen bond. No other distortion will be allowed.

The range of values over which the two pairs can independently form is illustrated in Fig. 1(*a*). The curves show, for each pair, the limiting values of θ for each possible value of d; a pair may thus be formed with any combination of values of d and θ which lie inside the appropriate closed curve. The range of values over which exactly equivalent pairs may form is given by the area over which the two closed curves overlap. It will be seen that the area of overlap is small; θ must lie between 50° and 51°, and d between 11·0 and $11{\cdot}1_5$ Å. It would, of course, be considerably increased if guanine formed only two hydrogen bonds with cytosine. The area of overlap is not increased by allowing N–H bonds to deviate further from the lines of their hydrogen bonds; it is limited only by the lengths of the hydrogen bonds.

Pairs of bases for which d = 11·0 Å and θ = 51° are illustrated in Fig. 2. It is suggested that the dimensions given are those most suitable for use in a model of DNA. The N–H · · · O bonds are all somewhat long, though they lie within the range reported by Donohue. They could be shortened by slight distortion of certain bond angles, but one cannot at present decide if this is likely to occur.

The effect of distortions

Although the models used in constructing base pairs are believed to be accurate, there is no direct evidence that the dimensions are correct. Furthermore, bond angles outside the rings might well be slightly distorted from the symmetrical positions assumed. The effect on the range of values of d and θ of allowing reasonable distortion in certain bond angles is illustrated in Fig. 1(*b*). The following assumptions have been made:

1) N–H bonds may deviate by up to 10° from the lines of their hydrogen bonds.
2) All bond angles except angles in rings may vary by up to 4° from their assumed values.
3) The angles θ may deviate by ±4° from the mean for a pair of bases.

It will be seen that the curves overlap over a wide

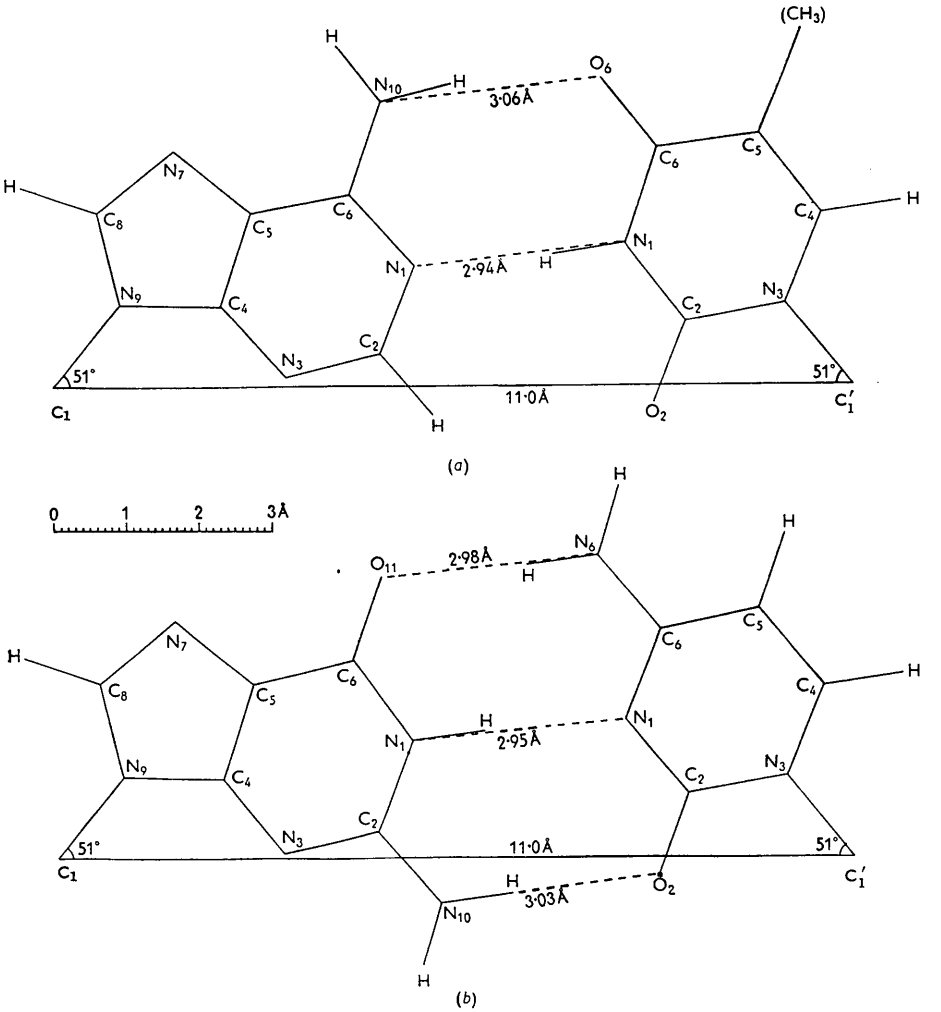

Fig. 2. Watson–Crick base pairs for use in a model of DNA; (*a*) adenine–thymine, (*b*) guanine–cytosine. Covalent bond lengths and angles used are those given in Part I.

range of values of d and θ; d may lie between 10·5 and 11·5 Å, and the mean value of θ between 42° and 59°. In addition, pairs of bases might form with slightly different values of d; this would further extend the range of possible structures.

One is forced to the conclusion that, although the *likely* form of the base pairs may be quite accurately described, the dimensions might be somewhat different if inter- or intra-molecular van der Waals forces, inter-ionic forces, or forces due to water binding were to influence the configuration of the DNA molecule.

Donohue's alternative base pairs

Construction of equivalent base pairs

Donohue (1956) showed that an alternative way of pairing adenine with thymine and guanine with cytosine is possible, and that roughly equivalent pairs can be so formed. The pairs differ from those proposed by Watson & Crick (1953) in the hydrogen bonds formed, and in the fact that the glycosidic links are related by a diad perpendicular to the plane of the bases. If one uses the same base structures as in the previous section, it is found that exactly equivalent pairs cannot be formed. It is, however, possible to form approximately equivalent pairs if one introduces slight distortions similar to those referred to in the previous section. The pairs believed to involve least distortion are shown in Fig. 3. There is one short contact (non-bonded interatomic distance) between hydrogen atoms attached to guanine and cytosine, but the corresponding N–N distance is not short. The maximum distortion of bond angle is 4° and the maximum H–N $\cdots$ X angle 17°.

One may roughly estimate the energy involved in the distortion of bond angles from a knowledge of approximate force constants. Using the largest force constant for bending of a single bond reported by Westheimer (1956) one obtains an energy of 600 cal.mole^{-1} for each 4° distortion, or 5·4 kcal.mole^{-1} for

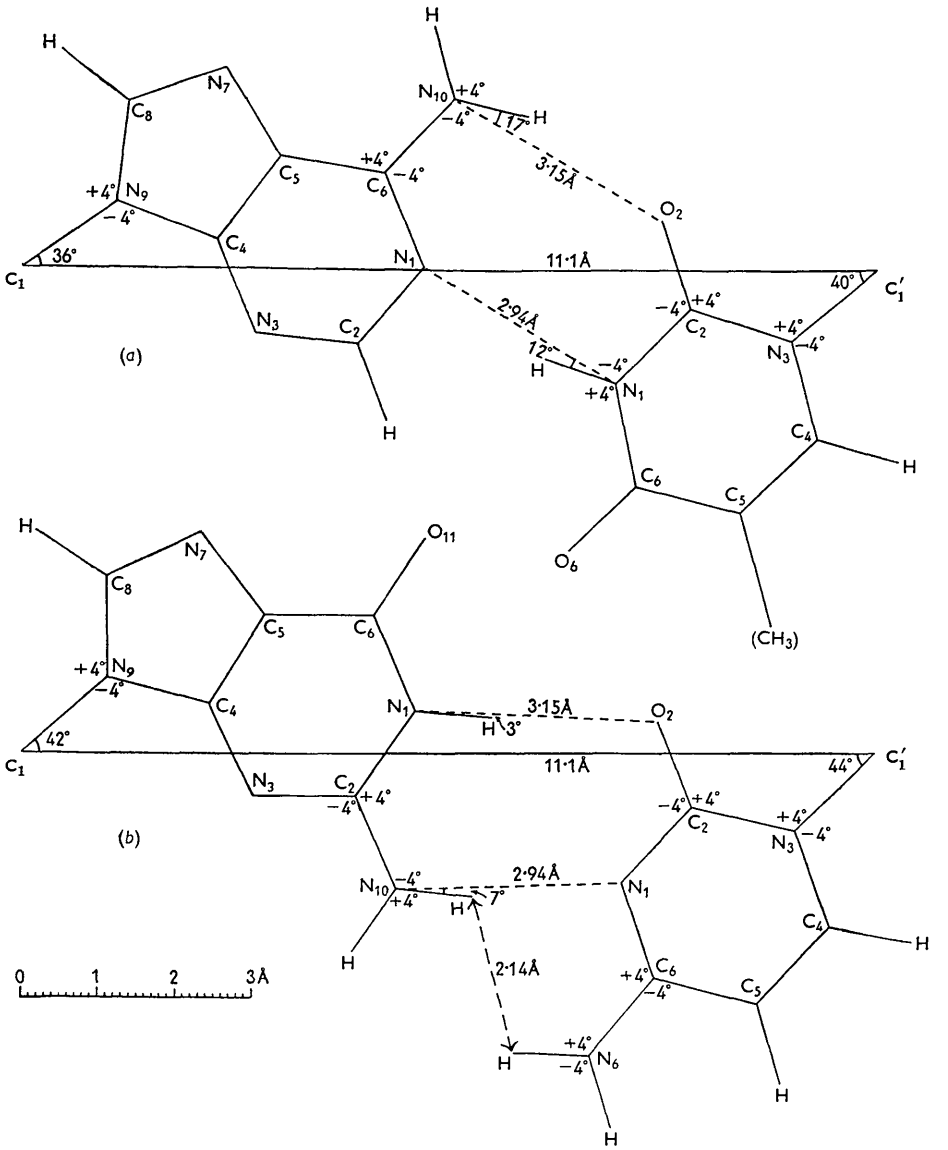

Fig. 3. Donohue base pairs involving least distortion; (*a*) adenine–thymine, (*b*) guanine–cytosine. Covalent bond lengths are as in Part I. Bond angles are as in Part I except where indicated; the distortion introduced is shown in each case.

each base pair. This is of the order of the maximum energy of one hydrogen bond. It is therefore possible that formation of two hydrogen bonds by a pair of bases could occur at the expense of the distortions described.

The existence of Donohue's pairs in DNA

Although there appears to be no *a priori* reason why Donohue's pairs should not occur in DNA, the following pieces of indirect evidence all suggest that they do not:

(i) Inosine can apparently replace guanine but not adenine in synthetic DNA (Adler *et al.*, 1958). On Donohue's scheme inosine can form only one hydrogen bond with cytosine, and pairing specificity is lost because it can form the same bond with thymine. On the Watson–Crick scheme, however, inosine can pair with cytosine but not with thymine.

(ii) The degree of crystallinity of DNA is unusually high for a fibrous material. This suggests that each molecule has a diad perpendicular to its length, so that it is 'reversible'.* Furthermore, the sodium salt of DNA crystallises in a lattice whose *b* axis makes an angle very close to 90° with the *ac* plane (Wilkins, Seeds, Stokes & Wilson, 1953); this suggests that the

* Polyadenylic acid (Watson, 1957) and polyalanine (Elliott & Malcolm, 1958) are helical polymers which lack such a diad, and in both cases the packing is less perfect than in the case of DNA.

crystal is monoclinic, in which case there would again have to be a diad perpendicular to the length of the molecule. In a DNA structure involving Donohue's pairs the sugar-phosphate chains could not be related by such a diad. Crystalline packing might, however, occur if the phosphate groups controlled the packing and were related by the diad.

(iii) A study of X-ray diffraction data and molecular models suggests strongly that the sugar-phosphate chains cannot be related by a diad coincident with the helix axis (Crick & Watson, 1954). If this is so the chains must either be parallel but not symmetrically disposed about the axis, which seems unlikely, or they must be antiparallel. If they were antiparallel the deoxyribose rings attached to the bases of Donohue's pairs would have to adopt two different configurations, related by a rotation of about 180° about the glycosidic link. Only one configuration has been observed in crystals of pyrimidine nucleosides (Furberg, 1950; Huber, 1957; Woolfson, 1956). It has been implied (Crick, 1957*b*) that the second configuration is unlikely to occur with pyrimidines because of short contacts between base and sugar atoms. Construction of a model shows that it involves slightly short contacts which may be removed by slight distortion of the base, so that it cannot really be ruled out.* Since this point is of interest in certain other applications, details of the model are given in the Appendix. One may conclude that, although two nucleoside configurations *could* occur, the one which has been observed in crystals is stereochemically favourable compared with the other, and it is likely that DNA contains only this configuration.

Conclusions

The range of positions of the glycosidic links of the Watson–Crick base pairs is quite limited if it is assumed that the base structures are invariable, that guanine and cytosine form three hydrogen bonds, and that the pairs of bases must be exactly equivalent. The degree of latitude is then of the same order as the probable error of 0·1 Å in the known distances between molecules in a crystallite (Langridge *et al.*, 1957). It is also comparable with the cumulative error in the dimensions of a molecular model due to uncertainty in the assumed values of covalent bond lengths and angles (Part I of this paper). If, however, certain reasonable distortions are allowed, the range is considerably increased. From this it has been concluded that stabilizing forces acting on the DNA molecule might fix the positions of the links anywhere within this range. If one knows more of these forces, it would help in producing a highly accurate model of DNA from the data at present available.

* The question of whether the configuration is likely to *change* to this one from that observed in crystals depends on the energy required to pass through the intermediate configurations, which involve shorter contacts.

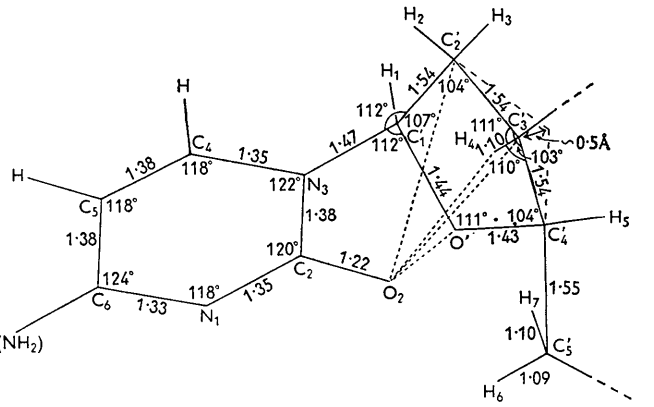

Fig. 4. A possible model of deoxycytidine. Bond lengths and angles are based on those given in Part I.

The alternative base pairs suggested by Donohue have been discussed in detail, because one cannot be sure on X-ray evidence alone that they might not form the basis of the structure of DNA. There is strong indirect evidence that they do not, in fact, occur in DNA, but it has proved impossible to reject them altogether. Further work on molecular models will be necessary if the question is to be settled conclusively. There is, meanwhile, no reason to doubt that the Watson–Crick pairing scheme is correct.

I am indebted to Dr M. H. F. Wilkins and Dr G. Zubay for many helpful discussions.

APPENDIX

A possible configuration of deoxycytidine (see Fig. 4).

Coordinates

	Atom	*x* (Å)	*y* (Å)	*z* (Å)
	C_1'	0·00	0·00	0·00
	H_1	−0·48	−0·98	−0·11
	C_2'	0·81	0·06	1·31
	H_2	0·33	0·70	2·06
	H_3	0·92	−0·95	1·72
	C_3'	2·17	0·64	0·87
	H_4	2·17	1·74	0·95
	C_4'	2·27	0·22	−0·61
	H_5	2·67	−0·80	−0·68
	C_5'	3·15	1·20	−1·43
	H_6	2·61	1·50	−2·33
	H_7	3·41	2·08	−0·83
	O'	0·92	0·26	−1·08
	N_3	−1·13	0·94	0·00
Base undistorted	C_2	−0·87	2·29	0·00
	O_2	0·28	2·69	0·00
Base distorted	C_2	−0·96	2·31	0·00
	O_2	0·12	2·87	0·00

Bond angles and Van der Waals contact distances

	$C_1'N_3C_2$ (°)	$N_3C_2O_2$	O_2H_4 (Å)	O_2C_2'	O_2C_3'	O_2O'
Base undistorted	119	120	2·32	2·99	2·92	2·74
Base distorted	123	124	2·53	3·18	3·15	2·94

References

ADLER, J., BESSMAN, M. J., LEHMAN, I. R., SCHACHMAN, H. K., SIMMS, E. S. & KORNBERG, A. (1958). *Fed. Proc.* **17**, 178.

CHARGAFF, E. (1950). *Experientia* **6**, 201.

CHARGAFF, E. (1955). *The Nucleic Acids* (ed. E. Chargaff & J. N. Davidson), **1**, 307. New York: Academic Press.

CRICK, F. H. C. (1957*a*). (Special Publications.) *N.Y. Acad. Sci.* **5**, 175.

CRICK, F. H. C. (1957*b*). *The Chemical Basis of Heredity* (ed. W. D. McElroy & B. Glass), p. 536. Baltimore: The John Hopkins Press.

CRICK, F. H. C. & WATSON, J. D. (1954). *Proc. Roy. Soc.* A, **223**, 80.

DONOHUE, J. (1952). *J. Phys. Chem.* **56**, 502.

DONOHUE, J. (1956). *Proc. Nat. Acad. Sci., Wash.* **42**, 60.

ELLIOTT, A. & MALCOLM, B. R. *Proc. Roy. Soc.* A. In press.

FURBERG, S. (1950). *Acta Cryst.* **3**, 325.

HUBER, M. (1957). *Acta Cryst.* **10**, 129.

LANGRIDGE, R., SEEDS, W. E., WILSON, H. R., HOOPER, C. W., WILKINS, M. H. F. & HAMILTON, L. D. (1957). *J. Biophys. Biochem. Cytol.* **3**, 767.

PAULING, L. & COREY, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164.

SPENCER, M. (1959). *Acta Cryst.* **12**, 59.

WATSON, J. D. (1957). *The Chemical Basis of Heredity* (ed. W. D. McElroy & B. Glass), p. 552. Baltimore: The John Hopkins Press.

WATSON, J. D. & CRICK, F. H. C. (1953). *Nature, Lond.* **171**, 737.

WESTHEIMER, F. H. (1956). *Steric Effects in Organic Chemistry* (ed. M. S. Newman), p. 529. New York: Wiley.

WILKINS, M. H. F., SEEDS, W. E., STOKES, A. R. & WILSON, H. R. (1953). *Nature, Lond.* **172**, 759.

WOOLFSON, M. M. (1956). *Acta Cryst.* **9**, 974.

Short Communications

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 500 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible; and proofs will not generally be submitted to authors. Publication will be quicker if the contributions are without illustrations.

Acta Cryst. (1959). **12**, 71

Calculation of scattering intensity from a cylindrically symmetrical system.

By I. M. STUART, *Physics and Engineering Unit, Wool Textile Research Laboratories, Commonwealth Scientific and Industrial Research Organization, The Hermitage, 338 Blaxland Road, Ryde, N.S.W., Australia.*

(*Received* 30 *July* 1958)

G. Oster & D. P. Riley (1952) give an expression for the amplitude of scattering F in the equatorial plane by a cylindrically symmetrical system

$$F = \frac{\int_0^\infty rG(r)J_0(kr)\,dr}{\int_0^\infty rG(r)\,dr} .$$

Here F is normalised to be unity at zero scattering angle and $k = (4\pi/\lambda)\sin\theta$ where λ is the wavelength of the incident radiation and 2θ is the scattering angle. $G(r)da$ is the probability that scattering material lies in the element of area da at distance r from the centre of the equatorial section.

When

$$G(r) = \cos^2\frac{m\pi r}{R} \qquad r < R$$
$$= 0 \qquad r \geq R$$

the structure is said to be radially periodic. In this case $2m$ is the number of 'corrugations' in the distribution of scattering material across a diameter $2R$ of the equatorial section. If we define a new parameter $p = kR/(2m\pi)$, we can write

$$F(p) = \frac{\int_0^{2m\pi} x(1+\cos x)J_0(px)\,dx}{\int_0^{2m\pi} x(1+\cos x)\,dx} .$$

We discuss the evaluation of F when $2m$ is an odd integer, which is the case when there is a whole number of corrugations across a diameter. Oster & Riley (1952) state that F can be reduced to 'a complicated algebraic expression involving k, $2m\pi/R$, $J_0(kR)$, $J_1(kR)$', and from this expression appear to have inferred a maximum of F^2 at $p = 1$. The author has been unable to find such an expression; and to calculate F for $0 < p < \infty$ three series expansions have been derived, each suitable for calculation on a limited range. Further it can be shown that a maximum of F^2 occurs near, but not at, $p = 1$. For $2m = 7$ this maximum is at $p = 1{\cdot}038$, and the value there of F^2 exceeds that at $p = 1$ by 38%.

We list these three series below, giving a rough guide to their appropriate ranges and the number of terms needed to obtain $\sim 1\%$ accuracy.

A. A series expansion of $F(p)$, centred on $p = 1$ and convergent for all values of p. For 1% accuracy no more than 10 terms are needed in the range $1-1/(2m) < p < 1+1/(2m)$

$$F = 2\left[1-\frac{1}{(m\pi)^2}\right]^{-1}\sum_{n=0}^{n=\infty}\frac{[m\pi(1-p^2)]^n}{n!}\cdot\frac{J_{n+2}(2m\pi)}{2n+3} .$$

B. A series expansion of $F(p)$, centred on $p = 0$ and convergent for $p < 1$. For 1% accuracy we can neglect terms of order p and higher, in the range $0 < p < 1/(2m)$.