

## Crystal and Molecular Structure of Cytidylic Acid *b*

EYVIND ALVER and SVEN FURBERG

*Chemical Institute, University of Bergen, Bergen, Norway*

Cytidylic acid *b*,  $C_9H_{14}N_3O_8P$ , forms crystals of symmetry  $P2_12_12$  with unit cell dimensions  $a = 8.80 \text{ \AA}$ ,  $b = 21.7 \text{ \AA}$ ,  $c = 6.85 \text{ \AA}$ . The structure was derived from electron density and difference syntheses along the  $a$  and  $c$  axes.

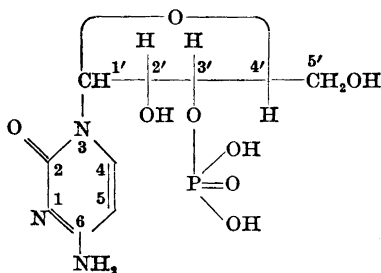
The molecule contains no unusual bond lengths, but some of the bond angles in the ribose ring differ significantly from "normal" values. The phosphate group is attached to the 3'-position of the sugar. The puckering of the ribose ring is different from what was found in crystals of cytidine, different ring atoms being out of the plane of the four other atoms in the two cases. The ribose and the phosphate group are arranged so as to make the molecule as extended as possible, the P—O<sub>3'</sub> bond having the same direction as the N<sub>3</sub>—C<sub>1'</sub> bond. This structure differs considerably from the structure hitherto assumed for nucleotides when building models of nucleic acids.

As constituents of the nucleic acids, the nucleosides and nucleotides are compounds of great biological interest, and it is important to have a detailed knowledge of their stereochemistry. The crystal structure of a nucleoside, cytidine, has been determined by one of us<sup>1</sup>, and since then a number of heavy atom derivatives of nucleosides have been studied by X-ray methods<sup>2,3</sup>. As for the nucleotides, a note on the determination of a projection of adenylic acid has appeared in the literature<sup>4</sup>. Furthermore, a nucleotide-like component forms part of the molecule of vitamin B<sub>12</sub>, the crystal structure of which has been determined by Hodgkin *et al*<sup>5</sup>. In the present paper the determination of the complete crystal structure of a nucleotide, cytidylic acid *b*, the 3' phosphate ester of cytidine, is described. A preliminary note on this work has been published<sup>6</sup>.

### EXPERIMENTAL. CRYSTAL DATA

A specimen of cytidylic acid *b* was kindly supplied by Dr. Waldo E. Cohn, of the Oak Ridge National Laboratory, Tennessee. Large, well-developed crystals were obtained by evaporation of an aqueous solution, generally with (010) dominating and elongated along  $c$ . Forms {011}, {001}, {100} and {110} were also observed. (100) is a cleavage plane.

An examination by the polarising microscope showed the crystals to be biaxial positive, with (001) as the optic axial plane, and with  $a \parallel a$ ,  $\gamma \parallel b$  and  $\beta \parallel c$ . The optic axial angle is great.



Structural formula, cytidylic acid b.

Oscillation and Weissenberg photographs were taken about the  $a$  and  $c$  axes, using filtered copper radiation ( $\lambda = 1.542 \text{ \AA}$ ). The crystals are orthorhombic, with unit cell dimensions  $a = 8.80 \text{ \AA}$ ,  $b = 21.7 \text{ \AA}$  and  $c = 6.85 \text{ \AA}$  ( $\pm 0.5 \%$ ). The only systematic absences occur in the  $h00$  reflections for  $h$  odd and in the  $0k0$  reflections for  $k$  odd, and the space group is therefore  $P2_12_12_1$ . Some of the odd-order  $0k0$  reflections appear with low intensity on long-exposure photographs, but their form indicate that they are Renninger reflections. The density, measured by flotation, was found to be  $1.66 \text{ g/cm}^3$ . There are four molecules  $C_9H_{14}N_3O_8P$  in the unit cell, which gives a calculated density of  $1.64 \text{ g/cm}^3$ .

For the structure determination, Weissenberg photographs of the zones  $hk0-hk4$  and  $0kl-5kl$  were taken, and about 1300 independent reflections were recorded. The intensities were estimated visually and corrected for Lorentz-polarization effects<sup>7</sup> and for the variation of spot shape on equi-inclination photographs<sup>8</sup>. Small crystals were used and no correction for absorption was applied. The  $hk0$  reflections were recorded from two different crystals and the mean value of the two sets of measurements was used. The calculations were carried out on desk machines and Bevers-Lipson strips were employed for the Fourier summations.

#### DETERMINATION OF THE STRUCTURE

Projections of the structure along the  $a$ - and  $c$ -axes were determined. The  $a$ -axis projection was first solved, as the strongest reflections are found in this zone. Approximate positions for most of the 21 atoms were derived by Patterson methods, inequalities and "trial and error", use being made of the Patterson function to determine phases. The refinement was carried out mainly by means of  $(F_o - F_c)$  syntheses.

*Location of the phosphorus atom.* Patterson projections along the  $a$  and  $c$  axes were first calculated (Fig. 1). The highest maxima were supposed to represent P—P vectors, and it can be seen from Fig. 1b that the phosphorus atom probably has the approximate coordinates  $y_p = 0.25$ ,  $z_p = 0.10$ . The maxima in Fig. 1a are too diffuse to give reliable coordinates, but indicate that the phosphorus atom is situated near  $x = 0$  or  $x = 0.25$ .

The Harker sections  $(u, v, 0)$ ,  $(1/2, v, w)$  and  $(u, 1/2, w)$  were calculated in the hope that they might give more reliable coordinates for the phosphorus atom. The section  $u = 1/2$  has its highest maximum at  $(0, 0.23)$ , corresponding to coordinates  $y_p = 0.25$  and  $z_p = 0.115$ . This section was "sharpened" by dividing  $F^2$  by  $f^2$ . In the Harker section  $w = 0$  the P—P vector  $(2x, 2y)$  should then be found at  $(2x, 0.5)$ . However, two maxima of equal heights appeared at  $v = 1/2$ , leaving the two possibilities  $x_p = 0.06$  and  $x_p = 0.25$ . This difficulty was overcome by an application of the "vector convergence"

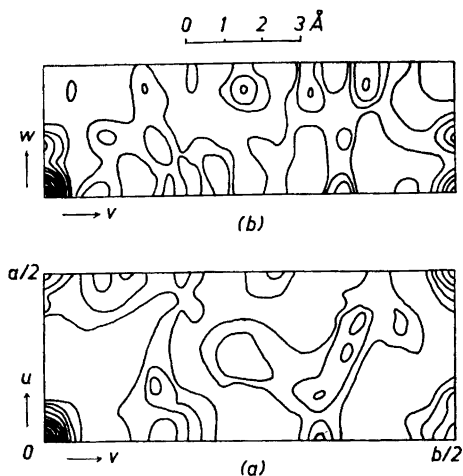


Fig. 1. Patterson projections on the (001) plane (a) and the (100) plane (b).

method<sup>9</sup>. The vector convergence map obtained by overlapping four transcriptions of the Patterson  $c$ -projection was derived for the two possible  $x$ -coordinates. In the case of  $x = 0.25$ , the map showed strong overlapping of peaks in the region near the axis of rotation, where no atoms can be situated. In fact, no atomic centre should be placed at a shorter distance from this axis than its van der Waals radius. The vector convergence map corresponding to  $x = 0.06$  had a very low value in the "forbidden" region. Consequently, the phosphorus atom probably has the position (0.06, 0.25, 0.115). The agreement with the coordinates determined later on by Fourier methods is excellent. The "correct" vector convergence map also showed that the molecule probably should lie roughly diagonally in the  $c$ -projection.

Very little information on the positions of the lighter atoms could be derived from the Harker sections because of strong masking by non-Harker maxima.

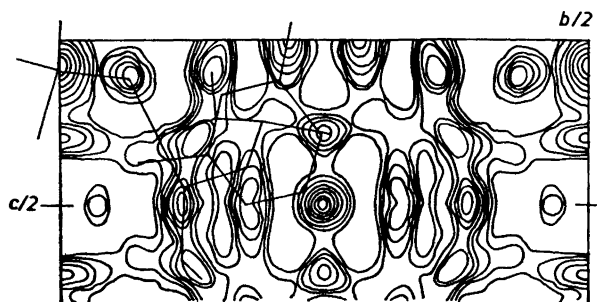


Fig. 2. The  $a$ -projection vector convergence map.

*The a-projection.* Attempts to solve the structure by the standard heavy atom technique were not successful. The phosphorus atom carries less than 10 % of the electrons in the molecule and is hardly heavy enough to determine the phases. Furthermore, its contribution to many structure factors is zero, as  $y_P = 0.25$ . Some information on the position of the lighter atoms was obtained from the Patterson functions in two ways. Firstly, a vector convergence map (Fig. 2) was derived by adding four transcriptions of the Patterson  $a$ -projection with their origins transferred to (0,0), (0.5, 0), (0, 0.23) and (0.5, 0.23). The map did not show sharp maxima which could be interpreted as individual atoms. Its symmetry is also too high. However, it indicated roughly the distribution of the atoms in the unit cell, and was used in conjunction with Bragg-Lipson charts to determine the probable signs of a number of structure factors. The second application of the Patterson function was based on the assumption that the lines  $w = 0$  in the Patterson  $a$ -projection and  $u = 1/2$  in the Patterson  $c$ -projection probably to a large extent represent vectors between equivalent atoms. By combining these two curves the vectors between non-equivalent atoms can be partly eliminated, and the curves may be expected to indicate the  $y$ -coordinates of the atoms. The probable signs of the four strongest  $0k0$ -reflections were derived in this way. The procedure may possibly be improved by sharpening the Patterson function.

The structure factors were put on an approximately absolute scale by Wilson's method<sup>10</sup>. The unitary structure factors ( $U$ ) all had values below 0.30, except for  $U_{0,14,1} = 0.40$  and  $U_{005} = 0.34$ . The signs of the strong reflections 005, 060, 080, 0,14,0, 0,16,0, 0,14,1 and 0,25,2 were derived from the vector convergence map and the distribution curve described above. As the unitary structure factors are small, the usual inequality  $(U_H \pm U_K)^2 \leq (1 \pm U_{H+K})(1 \pm U_{H-K})$  could give no more signs. However, it was used to determine seven more probable signs after multiplying the  $U$  values by  $\exp(1.8 \sin^2 \theta / \lambda^2)$ <sup>11</sup>. The probable signs of eight additional reflections were indicated by the "multiplication rule"<sup>12</sup>. The  $F_o$  map calculated from these 22 important reflections was unclear, but indicated the positions of the oxygen atoms in the phosphate group. This was confirmed by calculating the three-dimensional Patterson function in this region and by the Harker sections. A number of other terms could now be included in the  $F_o$  synthesis, but in spite of this, the rest of the molecule had to be placed mainly by trial and error procedures. This was rather difficult, as freedom of rotation about a number of single bonds and the flexibility of the sugar ring make a great variety of molecular shapes possible. Extensive use was made of Bragg-Lipson charts and evidence from the Patterson, vector convergence, and  $F_o$  maps considered. The vector convergence map was particularly useful at this stage, as it seemed improbable that any atom should be lying in a strongly negative area. In effect, it represents the correct structure remarkably well, as will be seen from Fig. 2. An important clue is furthermore provided by the optical properties of the crystal, which indicate that the plane of the pyrimidine ring is roughly perpendicular to the  $a$ -axis. Reasonable qualitative agreement between observed and calculated structure factors for reflections of lower orders was finally obtained with a structure, in which the sugar and pyrimidine parts of different molecules overlapped very badly. By  $F_o$  refinement the reliability factor

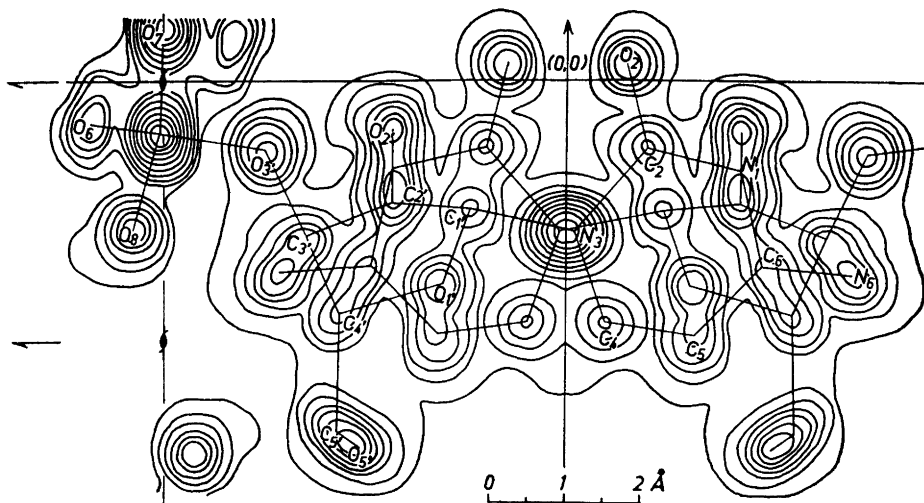


Fig. 3. Electron density projection in direction of  $a$  axis.  
Contours at 2, 4, 6, .....  $e. \text{Å}^{-3}$

$R = \Sigma (||F_o| - |F_c||) / \Sigma |F_o|$  was brought down to about 0.45 (ignoring reflections with  $\sin \Theta > 0.75$ ), but the lack of resolution made further improvement by this method difficult. The refinement was continued by successive  $(F_o - F_c)$  syntheses, about twenty of these being calculated. In the beginning only terms for which  $|F_c|$  was considerably greater than  $|F_o|$  were included. The refinement process did not seem to be automatic at the earliest stages and a certain election amongst the shifts indicated by the difference maps had to be made. The positions of most of the fourteen hydrogen atoms were indicated by the final maps. The others were placed by stereochemical considerations. Their inclusion did not improve the agreement significantly. Furthermore, the eight strongest reflections were corrected for extinction effects, following the procedure described by Lipson *et al.*<sup>19</sup> The final value for  $R$  for all  $0kl$  reflections observable by copper radiation is 0.11. In calculating  $R$  the non-observed reflections are only taken into account when  $F_c$  exceeds the observable limit of intensity. In the temperature factor  $\exp(-B \sin^2 \Theta / \lambda^2)$  the value of  $B$  is  $2.2 - 0.5 \sin^2 \varphi$  for the phosphorus atom and  $2.7 - 0.5 \sin^2 \varphi$  for all the other atoms,  $\varphi$  being the angle between the  $c$ -axis and the normal to the reflecting plane. The atomic scattering curves of McWeeny<sup>13</sup> (for C, N, O and H) and Viervoll and Øgrim<sup>14</sup> (for P) were used. The final electron density map is reproduced in Fig. 3.

*The c-projection.* Approximate  $x$ -coordinates were derived by means of the  $a$ -projection and approximately known interatomic distances, assuming that  $x_p$  was correctly determined from the Harker section and that there were hydrogen bonds between the  $\text{NH}_2$ -group and the neighbouring  $\text{PO}_4$ -group. The postulated structure gave a factor  $R$  of 0.50, which was brought down to 0.11 by about ten successive  $(F_o - F_c)$  refinements. Some very radical shifts in the

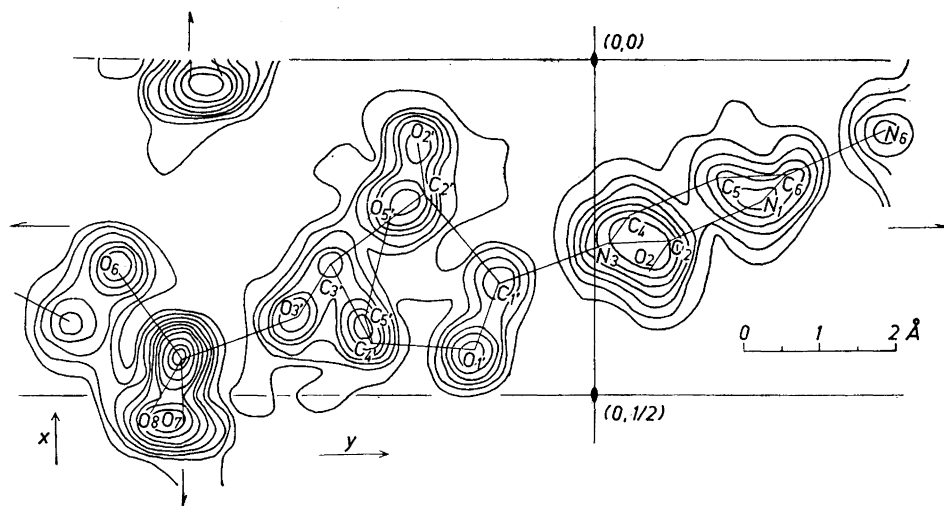


Fig. 4. Electron density projection in direction of  $c$  axis.  
Contours at 2, 4, 6, ..... e. Å<sup>-2</sup>.

rotation about single bonds were clearly indicated by the difference maps in the course of this process. The hydrogen atoms were located in the same way as in the  $a$ -projection and their inclusion made  $R$  drop to 0.10. Correction for extinction led to a final value of 0.09 for  $R$ . The final value chosen for  $B$  was 1.7 for all the atoms, but a comparison of  $F_o$  and  $F_c$  indicated that the vibrations of the phosphorus atom and the keto oxygen atom were anisotropic. This was taken into account by applying a factor  $B$  of  $1.7 - 0.5 \sin^2 \varphi$  to the phosphorus atom and of  $1.7 + 1.0 \cos^2 \varphi$  to the keto oxygen atom,  $\varphi$  being the angle between the  $a$ -axis and the normal to the reflecting plane. Further refinements of the thermal parameters were not carried out. The final electron density map is given in Fig. 4. In Table 4 observed and calculated structure factors are compared for the  $hk0$  reflections. A corresponding list for the  $0kl$  reflections can be supplied by the authors on request.

In Table 1 the atomic coordinates derived from the final difference maps in the two projections are given. The two sets of  $y$ -coordinates were in good agreement and the mean value was taken. In Table 2 the coordinates employed in the calculation of the contribution from the hydrogen atoms are given.

#### DESCRIPTION AND DISCUSSION OF THE STRUCTURE

The bond lengths and bond angles calculated from the coordinates in Table 1 are given in Table 3 and Fig. 5. A rough estimate of the accuracy of the structure was obtained by comparing the two sets of independently determined  $y$ -coordinates. For several atoms, inclusive the phosphorus atom, the two values are identical. The greatest deviation, 0.04 Å, occur for the atom  $C_6$ , and the

Table 1. Atomic coordinates as fractions of the corresponding cell edge.

Atom	$x$	$y$	$z$
P	0.060	0.2466	0.114
O <sub>6</sub>	0.195	0.206	0.066
O <sub>7</sub>	-0.042	0.2495	-0.069
O <sub>8</sub>	-0.025	0.2278	0.291
O <sub>3'</sub>	0.119	0.3162	0.136
O <sub>2'</sub>	0.382	0.392	0.086
O <sub>1'</sub>	0.064	0.426	0.397
O <sub>5'</sub>	0.279	0.3755	0.711
O <sub>4</sub>	0.187	0.5342	-0.030
C <sub>2'</sub>	0.195	0.337	0.309
C <sub>1'</sub>	0.300	0.395	0.260
C <sub>1'</sub>	0.168	0.443	0.255
C <sub>4'</sub>	0.076	0.362	0.459
C <sub>3'</sub>	0.120	0.358	0.673
C <sub>4</sub>	0.266	0.520	0.474
C <sub>5</sub>	0.320	0.579	0.511
C <sub>6</sub>	0.330	0.622	0.366
C <sub>2</sub>	0.235	0.546	0.135
N <sub>3</sub>	0.223	0.505	0.289
N <sub>1</sub>	0.280	0.603	0.181
N <sub>2</sub>	0.388	0.679	0.387

standard deviation of the  $y$ -coordinates from their means, all atoms taken together, is 0.012 Å. It is assumed that the accuracy of the  $x$ - and  $z$ -coordinates is about the same as that of the  $y$ -coordinates from one projection only, corresponding to a standard deviation of about 0.017 Å. This gives a standard deviation of 0.015–0.025 Å in the bond lengths and of 1–2° in the bond angles. The errors within the phosphate group probably correspond to the lower limit, whereas somewhat greater errors may occur where the atoms overlap most badly.

Table 2. Coordinates assumed for the hydrogen atoms.

Atom	$x$	$y$	$z$
H <sub>O<sub>6</sub></sub>	0.170	0.168	0.934
H <sub>O<sub>7</sub></sub>	0.840	0.267	0.930
H <sub>3'</sub>	0.237	0.302	0.435
H <sub>2'</sub>	0.366	0.420	0.350
H <sub>O<sub>1'</sub></sub>	0.330	0.384	0.943
H <sub>1'</sub>	0.107	0.442	0.110
H <sub>4'</sub>	0.970	0.334	0.430
H <sub>5'</sub>	0.055	0.395	0.755
H <sub>5'</sub>	0.086	0.326	0.750
H <sub>O<sub>4</sub></sub>	0.346	0.340	0.735
H <sub>4</sub>	0.280	0.500	0.560
H <sub>5</sub>	0.350	0.584	0.610
H <sub>N<sub>3</sub></sub>	0.450	0.683	0.510
H <sub>N<sub>1</sub></sub>	0.430	0.692	0.286

Table 3. Bond lengths (in Å units) and bond angles.

Bond lengths.					
P—O <sub>8</sub>	1.52	C <sub>1</sub> '—O <sub>1</sub> '	1.39	C <sub>5</sub> —C <sub>6</sub>	1.36
P—O <sub>7</sub>	1.54	O <sub>1</sub> '—C <sub>4</sub> '	1.46	C <sub>6</sub> —N <sub>1</sub>	1.40
P—O <sub>8</sub>	1.48	C <sub>4</sub> '—C <sub>3</sub> '	1.56	N <sub>1</sub> —C <sub>2</sub>	1.34
P—O <sub>3</sub> '	1.60	C <sub>3</sub> '—C <sub>5</sub> '	1.52	C <sub>2</sub> —N <sub>3</sub>	1.38
O <sub>3</sub> '—C <sub>3</sub> '	1.43	C <sub>5</sub> '—O <sub>5</sub> '	1.47	C <sub>2</sub> —O <sub>2</sub>	1.23
C <sub>3</sub> '—C <sub>2</sub> '	1.60	C <sub>5</sub> '—N <sub>3</sub>	1.45	C <sub>6</sub> —N <sub>6</sub>	1.35
C <sub>2</sub> '—O <sub>2</sub> '	1.40	N <sub>3</sub> —C <sub>4</sub>	1.36		
C <sub>4</sub> '—C <sub>1</sub> '	1.56	C <sub>4</sub> —C <sub>5</sub>	1.39		

Bond angles.					
O <sub>8</sub> —P—O <sub>8</sub>	115°	C <sub>2</sub> '—C <sub>1</sub> '—O <sub>1</sub> '	107°	C <sub>1</sub> '—N <sub>3</sub> —C <sub>2</sub>	120°
O <sub>8</sub> —P—O <sub>7</sub>	112°	C <sub>1</sub> '—O <sub>1</sub> '—C <sub>4</sub> '	114°	N <sub>3</sub> —C <sub>4</sub> —C <sub>5</sub>	119°
O <sub>8</sub> —P—O <sub>3</sub> '	110°	C <sub>4</sub> '—C <sub>3</sub> '—C <sub>5</sub> '	105°	C <sub>4</sub> —C <sub>5</sub> —C <sub>6</sub>	122°
O <sub>6</sub> —P—O <sub>7</sub>	108°	O <sub>1</sub> '—C <sub>4</sub> '—C <sub>3</sub> '	101°	C <sub>5</sub> —C <sub>6</sub> —N <sub>1</sub>	116°
O <sub>6</sub> —P—O <sub>3</sub> '	108°	C <sub>3</sub> '—C <sub>5</sub> '—O <sub>5</sub> '	110°	C <sub>6</sub> —N <sub>1</sub> —C <sub>2</sub>	125°
O <sub>7</sub> —P—O <sub>3</sub> '	103°	O <sub>1</sub> '—C <sub>4</sub> '—C <sub>5</sub> '	111°	N <sub>1</sub> —C <sub>2</sub> —N <sub>3</sub>	116°
P—O <sub>3</sub> '—C <sub>3</sub> '	121°	C <sub>3</sub> '—C <sub>4</sub> '—C <sub>5</sub> '	116°	C <sub>2</sub> —N <sub>3</sub> —C <sub>4</sub>	122°
O <sub>3</sub> '—C <sub>3</sub> '—C <sub>2</sub> '	110°	C <sub>4</sub> '—C <sub>5</sub> '—O <sub>5</sub> '	113°	N <sub>1</sub> —C <sub>2</sub> —O <sub>2</sub>	121°
C <sub>3</sub> '—C <sub>2</sub> '—O <sub>2</sub> '	116°	O <sub>1</sub> '—C <sub>1</sub> '—N <sub>3</sub>	111°	N <sub>3</sub> —C <sub>2</sub> —O <sub>2</sub>	123°
O <sub>2</sub> '—C <sub>2</sub> '—C <sub>1</sub> '	113°	C <sub>2</sub> '—C <sub>1</sub> '—N <sub>3</sub>	112°	C <sub>5</sub> —C <sub>6</sub> —N <sub>6</sub>	125°
C <sub>3</sub> '—C <sub>2</sub> '—C <sub>1</sub> '	96°	C <sub>1</sub> '—N <sub>3</sub> —C <sub>4</sub>	118°	N <sub>1</sub> —C <sub>6</sub> —N <sub>6</sub>	119°

*The molecular structure.* It is seen from the projections of the structure that the N-glycosidic linkage is of the  $\beta$ -type and that the phosphate group is bound to ribose in the 3'-position. Cytidylic acid *b* is consequently the 3'-phosphate ester of cytidine, in agreement with the conclusions drawn from more indirect physical studies<sup>15</sup>.

The dimensions of the phosphate group agree well with those found in dibenzylphosphate<sup>17</sup> and in crystalline phosphoric acid<sup>16</sup>. Values between 1.48 Å and 1.60 Å are derived for the P—O bond lengths and all of them evidently have a large amount of double bond character. Correspondingly the angle P—O<sub>3</sub>'—C<sub>3</sub>' is 121°, close to the value usually found for angles between double and single bonds. The longest P—O bond (1.60 Å) is the P—O<sub>3</sub>' ester bond. It runs roughly in the direction of the *y*-axis and should thus be expected to be more accurately determined than the other bonds. The lengthening of this bond relative to P—O<sub>6</sub> and P—O<sub>7</sub> is therefore probably significant. No such effect was, however, observed in the case of dibenzylphosphate. The shortest P—O distance (P—O<sub>8</sub>) is 1.48 Å, indicating that O<sub>8</sub> is the "lone" oxygen atom. Correspondingly, all the angles involving P—O<sub>8</sub> are somewhat greater than the tetrahedral angle and the angles between other P—O bonds somewhat smaller, the mean values being 112° and 106°, respectively, as was also found in phosphoric acid.

Four of the atoms of the ribose ring lie nearly in one plane, namely C<sub>1</sub>', O<sub>1</sub>', C<sub>3</sub>' and C<sub>4</sub>', but the fifth member of the ring, C<sub>2</sub>', is out of this plane by about 0.5 Å in such a direction that its oxygen atom O<sub>2</sub>' is brought close to the plane. In cytidine a different atom, C<sub>3</sub>', was displaced from the plane of the other atoms by about 0.5 Å. The bond lengths are normal within the limits



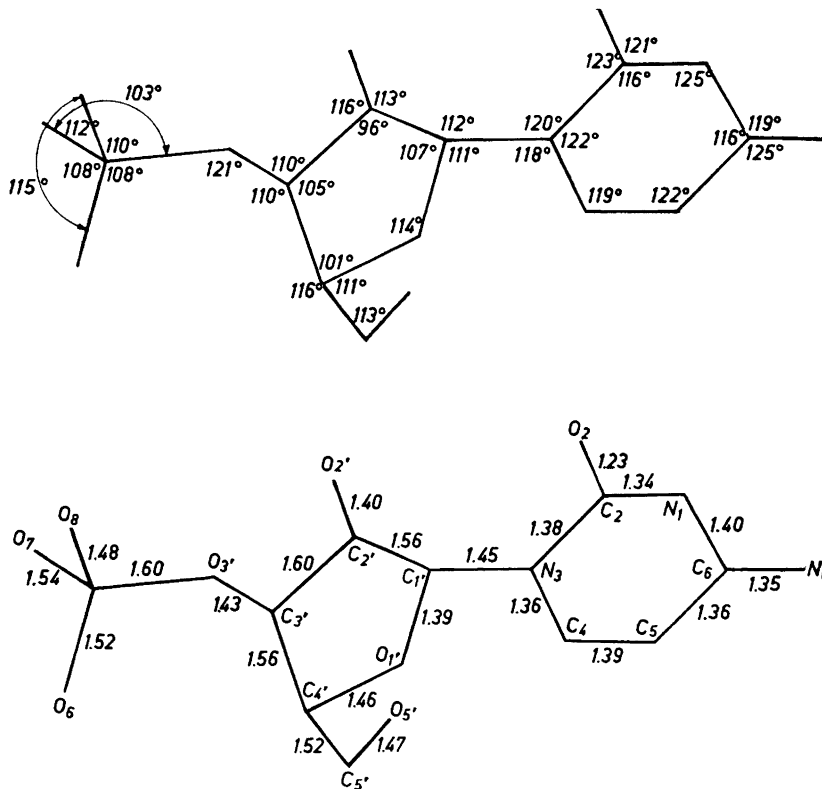


Fig. 5. Bond lengths and bond angles in cytidylic acid *b*.

of error, but some of the bond angles differ significantly from the expected values. Thus the ring angles at C<sub>4'</sub> and C<sub>2'</sub> are found as small as 101° and 96°, respectively. By a deformation of the furanose ring from a hypothetical planar structure the sum of the ring angles must become smaller. If it is deformed so as to bring base and phosphate into the relative orientation found in this work, it is seen by means of models that the most reasonable way of obtaining this is to make the angles at C<sub>2'</sub> and C<sub>4'</sub> smaller. Unfortunately the coordinates of C<sub>2'</sub> are very difficult to determine by two-dimensional methods because of overlapping with other atoms in both projections. The hydrogen and oxygen atoms attached to the ring are in a nearly staggered position with respect to the C—C ring bonds, corresponding to a stable conformation of the sugar.

The bond N<sub>3</sub>—C<sub>1'</sub> is found to be 1.45 Å and is thus essentially a single bond. It lies in the plane of the pyrimidine ring and forms angles of 111° and 112° with the adjacent ring bonds in the ribose. The bond angles and bond lengths in the pyrimidine part of the molecule are all normal, being in close agreement with those derived by Parry for uracil<sup>18</sup>. Thus the ring angles at the nitrogen

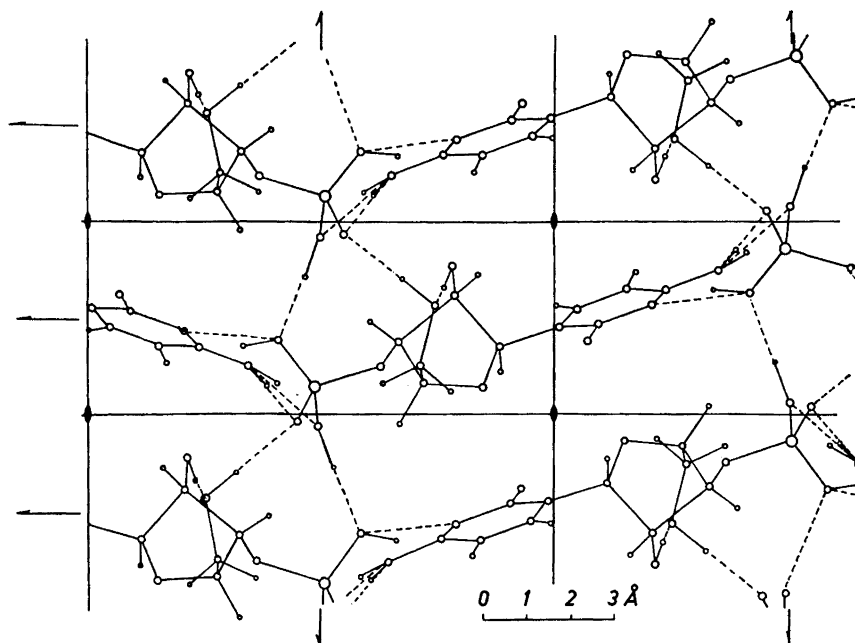


Fig. 6. The *c*-projection of the structure. Broken lines indicate hydrogen bonds. The hydrogen atoms are shown as small circles.

atoms are greater than  $120^\circ$ , the values  $124^\circ$  and  $125^\circ$  being found in uracil and  $122^\circ$  and  $125^\circ$  in the present structure. The  $C=O$  distance of  $1.23 \text{ \AA}$  agrees with the values found for keto groups in similar compounds. All the ring atoms and substituents lie in the same plane within the limits of error. The  $C_6-N_6$  distance of  $1.35 \text{ \AA}$  indicates a considerable amount of double bond character. The peaks corresponding to the two hydrogen atoms of the amino group are well defined in both projections, and show that the three bonds from  $N_6$  form a flattened pyramid, as in other amines. This is also in agreement with the directions of the two hydrogen bonds from  $N_6$ .

The mutual orientation of the pyrimidine and ribose rings is very nearly as in crystals of cytidine, the two ring "planes" being roughly perpendicular to each other. The midpoints of the bonds  $C_3'-C_4'$  and  $C_5'-O_5'$  lie nearly in the pyrimidine plane. The distance  $C_4-O_5'$  is  $3.53 \text{ \AA}$ , corresponding to normal van der Waals separation between  $H_4$  and  $O_5'$ . In cytidine this distance was found to be as short as  $3.24 \text{ \AA}^*$ .

*Intermolecular forces.* The chemical formula of the compound shows that six of the hydrogen atoms in the molecule would be expected to engage in the

\* Note added in proof. In a recent paper<sup>22</sup> Spencer has discussed most of the available information on the structure of nucleic acid components and related compounds. The present work seems to support his views regarding the conformations of the ribose ring.

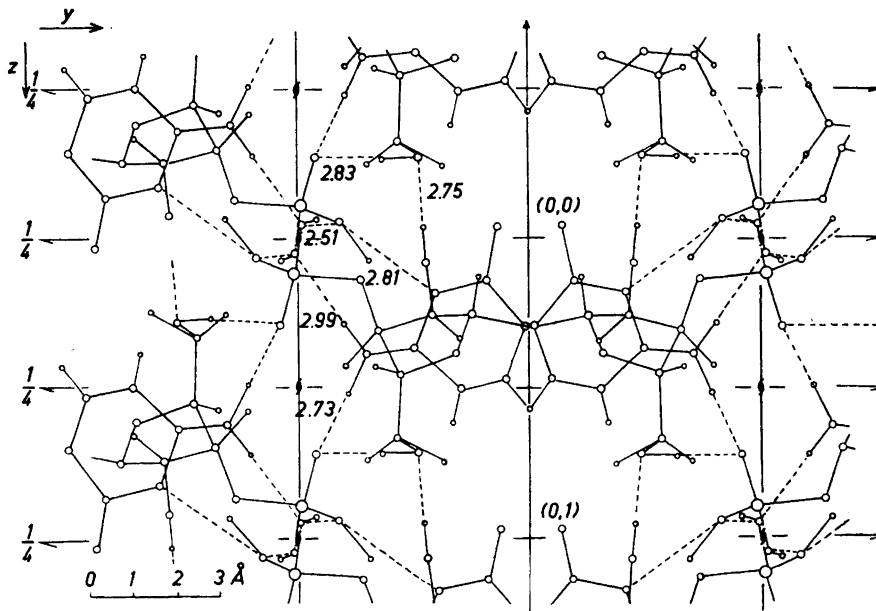


Fig. 7. The *a*-projection of the structure. The lengths of the hydrogen bonds are indicated (Ångström units).

formation of hydrogen bonds as there are two P—OH groups, one NH<sub>2</sub> group and two sugar hydroxyl groups in the molecule. Correspondingly only six intermolecular distances are found to be shorter than 3.1 Å and it appears therefore reasonable to interpret these interactions as hydrogen bonds.

Two hydrogen bonds of lengths 2.73 Å and 2.99 Å form linkages from the amino group to O<sub>8</sub> and O<sub>7</sub> in different phosphate groups. They are situated symmetrically with respect to C<sub>6</sub>—N<sub>6</sub> and lie in a plane which forms an angle of about 30° with the plane of the pyrimidine ring.

The oxygen atom O<sub>6</sub> of the phosphate group is linked by a hydrogen bond (length 2.81 Å) to the nearest ring nitrogen atom. This bond lies nearly in the ring plane. Furthermore, the atom O<sub>6</sub> is engaged in a very short hydrogen bond (length 2.51 Å) to O<sub>7</sub> in a neighbouring phosphate group. The two hydroxyl groups in the ribose, (OH)<sub>2'</sub> and (OH)<sub>5'</sub>, are connected by an O—H...O bond of length 2.75 Å running approximately in the *z* direction and (OH)<sub>5'</sub> takes also part in a similar bond (2.83 Å) to O<sub>8</sub>.

The amino derivatives of purines and pyrimidines are known to occur in crystals in the amino form. Thus there are two hydrogen atoms at N<sub>6</sub> and none at N<sub>1</sub>. Taking this into account, a consideration of the hydrogen bonding scheme shows that there are hydrogen atoms attached to O<sub>6</sub> and O<sub>7</sub>, but not to O<sub>8</sub>. Consequently O<sub>8</sub> is the "lone" oxygen atom of the phosphate group, as was also deduced above. The hydrogen bonds conform remarkably well to

the general stereochemistry of the molecule, and the angles between the hydrogen bonds and the adjacent intramolecular bonds do not deviate much from normal valency angles.

Neither the C=O group nor the ring oxygen atom of the ribose take part in hydrogen bonds. The number of "active" hydrogen atoms is not sufficient for hydrogen bond formation between all the electronegative atoms in the molecule.

None of the other intermolecular distances are shorter than normal van der Waals separations. There are "holes" in the structure around the twofold rotation axes. Every alternate one of these axes is surrounded by pairs of equivalent oxygen atoms at distances 3.4 Å ( $O_{1'}-O_{1'}$ ) and 3.6 Å ( $O_2-O_2$ ) whereas the packing around the others is even less close, the shortest distance across the axis being 3.9 Å ( $H_4-H_4$ ).

The general arrangement of the molecules and the hydrogen bonds are shown in Figs. 6 and 7.

Because of the great amount of overlap in both projections the structure can hardly be determined very accurately by two-dimensional methods, and probably not much more accurately than in this work. It is planned to carry out further refinements in three dimensions.

*Relation to nucleic acid structure.* The structure of nucleotides and its relation to nucleic acid structure has been discussed by one of us on the basis of the crystal structure analysis of cytidine<sup>21</sup>. It was concluded that base and sugar were likely to be nearly perpendicular to one another and that in the most probable orientation of the 3' phosphate group the bond P—O<sub>3'</sub> was roughly perpendicular to the plane of the base. The distance from this plane to the phosphorus atom was about 3.4 Å, and it was shown how nucleotides of this structure could be combined to macromolecules of helical and other shapes. Roughly the same nucleotide structure has been maintained in later models of nucleic acids, such as the Crick-Watson double helix model for DNA<sup>20</sup>, and it would therefore appear to be of interest to compare it with the structure derived for cytidylic acid in the present work.

Firstly, some of the main features of the nucleotide model used are confirmed. Thus the relative orientation of pyrimidine and ribose is almost exactly as in crystals of cytidine, the two "planes" being roughly perpendicular to one another. Furthermore, the length of the glycosidic linkage N<sub>3</sub>—C<sub>1'</sub>, as well as the angles between this bond and the adjacent bonds in the ribose are also the same in the two cases within the limits of error. At present there would appear to be no reason to revise the nucleotide model in this respect.

As both base and phosphate are linked directly to the sugar, the structure of the latter is very important for the shape of the whole molecule. In the models hitherto constructed for nucleic acids the sugar ring is supposed to be puckered in the same way as in crystals of cytidine. As shown above, the ring has a different shape in crystals of cytidylic acid. In both conformations there is a nearly staggered arrangement of the substituent hydrogen and oxygen atoms. They would appear to be about equally favourable and should therefore both be considered when building models of nucleic acids.

As for the position of the phosphate group relative to the rest of the molecule, two directions of C<sub>3'</sub>—O<sub>3'</sub> differing by about 60° should be considered,

Table 4. Observed ( $F_o$ ) and calculated ( $F_c$ ) structure factors for the  $hk0$  zone (\* Values corrected for extinction).

$hk0$	$F_o$	$F_c$	$hk0$	$F_o$	$F_c$	$hk0$	$F_o$	$F_c$
0 20	41 *	-41	2 150	13	+ 8	4 160	13	-10
0 40	4	- 4	2 160	28	+27	4 170	30	-29
0 60	59 *	-65	2 170	45	-45	4 180	7	+ 6
0 80	65 *	+69	2 180	33	-32	4 190	22	+20
0 100	18	-19	2 190	8	+ 7	4 200	8	+ 6
0 120	22	+20	2 200	<4	+ 4	4 210	<4	+ 4
0 140	55	-55	2 210	<4	+ 0	4 220	11	- 7
0 160	52	+46	2 220	6	- 4	4 230	4	+ 5
0 180	18	-18	2 230	6	+ 4	4 240	12	+12
0 200	18	+14	2 240	<3	- 1	4 250	5	- 3
0 220	20	-18	2 250	13	-13	5 10	19	+15
0 240	14	+14	2 260	9	- 9	5 20	6	- 7
0 260	14	-13	2 270	11	+11	5 30	42	-47
1 10	21	-22	3 10	17	-17	5 40	<4	- 3
1 20	34	+32	3 20	59 *	+68	5 50	22	-23
1 30	28	-31	3 30	43	+43	5 60	12	-12
1 40	36	-38	3 40	42	+45	5 70	16	-16
1 50	5	- 1	3 50	74 *	+80	5 80	20	-16
1 60	98 *	-99	3 60	29	+32	5 90	33	+35
1 70	27	-31	3 70	8	+ 9	5 100	7	- 2
1 80	<3	- 4	3 80	3	- 3	5 110	8	- 8
1 90	78 *	+79	3 90	47	-48	5 120	15	-16
1 100	32	-28	3 100	15	+13	5 130	7	+ 1
1 110	<3	- 2	3 110	35	-35	5 140	9	- 7
1 120	30	-32	3 120	11	+10	5 150	23	+23
1 130	6	- 6	3 130	27	+25	5 160	17	+17
1 140	18	-18	3 140	5	+ 3	5 170	13	-11
1 150	37	-38	3 150	18	-19	5 180	6	+ 8
1 160	11	-12	3 160	26	+25	5 190	13	-13
1 170	21	+23	3 170	12	+ 8	5 200	<4	- 1
1 180	9	+10	3 180	9	-11	5 210	6	- 4
1 190	<4	+ 3	3 190	<4	- 5	5 220	25	-29
1 200	8	+ 7	3 200	7	+ 3	5 230	<3	- 2
1 210	<4	+ 0	3 210	25	+21	5 240	<3	- 4
1 220	7	+ 7	3 220	<4	+ 5	5 250	7	- 7
1 230	12	-11	3 230	4	+ 2	6 00	22	-23
1 240	5	+ 5	3 240	10	+ 9	6 10	<4	+ 3
1 250	12	+10	3 250	<2	- 2	6 20	5	+ 4
1 260	15	-12	3 260	16	-16	6 30	6	- 6
1 270	8	- 8	4 00	14	+14	6 40	11	- 7
2 00	78 *	-85	4 10	<3	- 1	6 50	16	-16
2 10	114 *	-114	4 20	17	+19	6 60	3	+ 1
2 20	154 *	-148	4 30	16	-17	6 70	28	+27
2 30	8	+ 7	4 40	58	+60	6 80	25	-25
2 40	66 *	+65	4 50	19	-14	6 90	11	+13
2 50	24	+26	4 60	21	-25	6 100	9	+ 9
2 60	75 *	-76	4 70	9	+ 6	6 110	11	+10
2 70	42	-42	4 80	44	+44	6 120	22	-20
2 80	17	+18	4 90	37	-40	6 130	12	- 9
2 90	30	+29	4 100	<4	- 5	6 140	13	+11
2 100	68	-67	4 110	20	+17	6 150	9	-10
2 110	50	+54	4 120	4	- 1	6 160	18	-19
2 120	62	+60	4 130	20	-18	6 170	<4	- 1
2 130	<4	+ 4	4 140	7	+ 7	6 180	18	+17
2 140	7	- 6	4 150	26	+22	6 190	8	+ 6

$hkl$	$F_o$	$F_c$	$hkl$	$F_o$	$F_c$	$hkl$	$F_o$	$F_c$
6 20 0	8	-10	7 21 0	6	- 4	9 5 0	18	-17
6 21 0	21	-25	8 0 0	29	-31	9 6 0	12	+13
6 22 0	<3	+ 1	8 1 0	14	+13	9 7 0	<7	- 6
6 23 0	20	+21	8 2 0	20	+14	9 8 0	14	-14
7 1 0	5	+ 5	8 3 0	25	-25	9 9 0	<4	+ 2
7 2 0	16	-14	8 4 0	20	-18	9 10 0	7	+ 7
7 3 0	18	+16	8 5 0	<4	+ 1	9 11 0	4	+ 3
7 4 0	8	+ 5	8 6 0	36	+38	9 12 0	<3	- 1
7 5 0	<4	+ 4	8 7 0	<4	+ 4	9 13 0	9	-11
7 6 0	16	+16	8 8 0	15	-13	9 14 0	<3	- 0
7 7 0	7	+ 4	8 9 0	<4	- 1	10 0 0	<3	+ 1
7 8 0	11	+ 9	8 10 0	25	+24	10 1 0	21	+21
7 9 0	17	-16	8 11 0	7	- 6	10 2 0	6	+ 5
7 10 0	9	-10	8 12 0	16	-16	10 3 0	<3	+ 2
7 11 0	5	+ 1	8 13 0	10	+12	10 4 0	21	-23
7 12 0	9	- 8	8 14 0	<4	+ 1	10 5 0	18	+18
7 13 0	14	+14	8 15 0	8	+10	10 6 0	19	+20
7 14 0	15	+12	8 16 0	12	-10	10 7 0	21	-25
7 15 0	10	+ 6	8 17 0	<3	+ 0	10 8 0	11	-12
7 16 0	9	-12	8 18 0	15	+17	10 9 0	13	+11
7 17 0	<4	- 0	9 1 0	10	+12	10 10 0	18	+22
7 18 0	<3	+ 1	9 2 0	<4	- 1	10 11 0	9	-11
7 19 0	<3	+ 1	9 3 0	<4	- 2	10 12 0	11	-12
7 20 0	11	+10	9 4 0	7	- 8			

corresponding to the two conformations of the sugar. Furthermore, the angle  $C_{3'}-O_{3'}-P$  is  $121^\circ$ , in agreement with the results of the structure analysis of dibenzylphosphate<sup>17</sup>, whereas the tetrahedral angle has been assumed in the nucleic acid models. The bond  $O_{3'}-P$  is in a *trans* position with respect to the ribose ring and is almost parallel to the  $N_3-C_{1'}$  bond at a distance of about 0.6 Å, whereas these bonds are nearly perpendicular to one another in current nucleic acid models. In fact, the puckering of the ribose ring and the large angle at  $O_{3'}$  make the molecule a very extended one.

The question then arises if nucleotides with the sugar conformation found in the present work can be combined to form macromolecules by 3'-5' diester phosphate linkages, making the assumption that the bases are perpendicular to the molecular axis and separated by a distance of 3.4 Å corresponding to the dominating X-ray reflection. We have tried to study this problem by means of models.

In the helical structures proposed for DNA, such as our model I<sup>21</sup> and the Crick-Watson model<sup>20</sup>, the phosphate groups of the external sugar-phosphate spiral lie roughly in the plane of the bases, which form a central column. In order to build these structures using the sugar conformation found in the present work, we have found it necessary to place the phosphate group roughly midway between successive bases, a position which may possibly not be consistent with the strength of the 3.4 Å reflection. Even with this position of the phosphate group, some of the van der Waals separations are too short. This is also the case in the existing helical models, but on the whole it appears more promising to maintain the ribose conformation found in crystals of cytidine

as a basis for helical models. The possibility that different ribose residues within the same macromolecule may have different conformations should also be borne in mind.

Turning now to other possible structures for the macromolecules, such as the zig-zag model II described by one of us<sup>21</sup>, no difficulties are encountered when combining nucleotides of the shape found in the present work to models of this type. Evidently the two sugar conformations may both be used in this case to construct models which contain no unfavourably short van der Waals separations. However, there is considerable evidence that DNA has a double helix structure and other types of structures may therefore serve as models only for RNA. Very little is known about the space structure of this macromolecule, and so it does not seem worth while to discuss the models for it in more detail at the present stage.

We wish to express our thanks to *Norges Almenvitenskapelige Forskningsråd*, whose financial support made this work possible. We should also like to thank Dr. Waldo E. Cohn for providing us with a sample of cytidylic acid *b*, and the Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge, in particular Mrs. L. Barnett, for checking our calculations of bond lengths and bond angles.

#### REFERENCES

1. Furberg, S. *Acta Cryst.* **3** (1950) 325.
2. Zussmann, J. *Acta Cryst.* **6** (1953) 504.
3. Huber, M. *Acta Cryst.* **10** (1957) 129.
4. Brown, D. M., Fasman, G. D., Magarh, D. I., Todd, A. R., Cochran, W. and Woolfson, M. M. *Nature* **172** (1953) 1184.
5. Hodgkin, D. C., Kamper, J., Lindsey, J., MacKay, M., Pickworth, J., Robertson, J. H., Shoemaker, C. B., White, J. G., Prosen, R. J. and Trueblood, K. N. *Proc. Roy. Soc. London A* **242** (1957) 228.
6. Alver, E. and Furberg, S. *Acta Chem. Scand.* **11** (1957) 188.
7. Lu, C. S. *Rev. Sci. Instr.* **14** (1943) 331.
8. Philips, D. C. *Acta Cryst.* **7** (1954) 746.
9. Beevers, C. A. and Robertson, J. H. *Acta Cryst.* **3** (1950) 164; **4** (1951) 63.
10. Wilson, A. J. C. *Acta Cryst.* **2** (1949) 318.
11. Burbank, R. D. *Acta Cryst.* **5** (1952) 236.
12. Zachariassen, W. H. *Acta Cryst.* **5** (1952) 68.
13. McWeeny, R. *Acta Cryst.* **4** (1951) 513.
14. Viervoll, H. and Øgrim, O. *Acta Cryst.* **2** (1949) 277.
15. Fox, J. J., Cavalieri, L. F. and Chang, N. *J. Am. Chem. Soc.* **75** (1953) 4315.
16. Furberg, S. *Acta Chem. Scand.* **9** (1955) 1557.
17. Dunitz, J. D. and Rollet, J. S. *Acta Cryst.* **9** (1956) 327.
18. Parry, G. S. *Acta Cryst.* **7** (1954) 313.
19. Pinnock, P. R., Taylor, C. A. and Lipson, H. *Acta Cryst.* **9** (1956) 173.
20. Crick, F. H. C. and Watson, J. D. *Proc. Roy. Soc. London A* **223** (1954) 80.
21. Furberg, S. *Acta Chem. Scand.* **6** (1952) 634.
22. Spencer, M. *Acta Cryst.* **12** (1959) 59.

Received February 10, 1959.