

## The Crystal Structure and Absolute Configuration of 2'-Deoxycytidine Hydrochloride\*

BY E. SUBRAMANIAN† AND D. J. HUNT‡

A. A. Noyes Laboratory of Chemical Physics, California Institute of Technology, Pasadena, California 91109, U.S.A.

(Received 31 December 1968)

2'-Deoxycytidine hydrochloride ( $C_9H_{14}N_3O_4Cl$ ) was crystallized from water as monoclinic needles, space group  $P2_1$ , with  $a=6.561$ ,  $b=17.659$ ,  $c=5.125$  Å,  $\beta=108.08^\circ$  and two molecules per cell. The crystal structure has been determined by a three-dimensional X-ray diffraction analysis. A complete hemisphere of data (positive and negative values of both  $h$  and  $k$ ) was collected with a General Electric XRD-5 automatic diffractometer and nickel-filtered Cu  $K\alpha$  radiation. The structure was solved by the heavy-atom method. Full-matrix least-squares refinement of both the D and L enantiomorphs has led to respective  $R$  values of 0.035 and 0.061; the configuration that gave the lower  $R$  value confirms previous conclusions regarding the absolute configuration of the D sugar. The atoms in the pyrimidine ring exhibit small but significant deviations from planarity. The torsion angle describing the conformation about the glycosidic bond is  $0^\circ$ . The least-squares planes through any four ring atoms of the sugar residue, normally used to describe its conformation, are unsatisfactory in that the deviations of the atoms from the planes are rather large. The conformation can, perhaps, be best described relative to the plane through C(1'), C(4'), and O(1'), with respect to which C(3') is displaced *endo* by 0.361 Å and C(2') is 0.245 Å *exo*.

### Introduction

Accurate molecular parameters as well as the nature of the hydrogen-bonding and packing systems in nucleosides and nucleotides provide useful information toward deducing the detailed structures of the nucleic acids. We have determined the crystal structure and absolute configuration of 2'-deoxycytidine hydrochloride ( $C_9H_{14}N_3O_4Cl$ ) as one in a series of nucleic acid constituents being studied in these laboratories.

### Experimental

2'-Deoxycytidine hydrochloride (1- $\beta$ -2'-deoxy-D-ribofuranosylcytosine hydrochloride) was purchased from Mann Research Laboratories Inc., New York. The compound, prepared from salmon sperm deoxyribonucleic acid (DNA) by enzymatic hydrolysis and ion-exchange purification,§ was recrystallized from water as colourless, monoclinic needles elongated along the  $c$  axis. Crystal data are shown in Table 1. The space group was determined as either  $P2_1$  or  $P2_1/m$  from Weissenberg photographs taken about the  $b$  and  $c$  axes and exhibiting systematic absences only amongst  $0k0$

for  $k=2n+1$ . That the molecule lacks a centre of symmetry and  $Z=2$  establishes the space group as  $P2_1$ . The unit-cell parameters were obtained from a least-squares procedure minimizing  $\sum w_n(\sin^2\theta_{\text{obs}} - \sin^2\theta_{\text{calc}})^2$  ( $4/\lambda_n^2$ ). Values of  $2\theta$  (either  $\alpha_1$  or  $\alpha_2$  reflections) for fourteen reflections were measured on a diffractometer and assigned weights,  $w_n$ , proportional to  $1/\sin^2 2\theta$ . The confidence limits cited are subjective, being approximately three times the estimated standard deviations obtained from the least-squares procedure.

Table 1. Crystal data

$a$	$6.561 \pm 0.001$ Å
$b$	$17.659 \pm 0.003$
$c$	$5.125 \pm 0.001$
$\beta$	$108.08 \pm 0.02^\circ$
$t$	$27 \pm 3^\circ\text{C}$

Formula:  $C_9H_{13}N_3O_4.HCl$   
 M.W. 263.5  
 $\rho_m = 1.548$  g.cm $^{-3}$  (by flotation)  
 $\rho_c = 1.551$  g.cm $^{-3}$   
 $F(000) = 276$   
 $Z = 2$   
 $\lambda(\text{Cu } K\alpha_1) = 1.54050$  Å  
 $\lambda(\text{Cu } K\alpha_2) = 1.54434$

\* Contribution No. 3607 from the A. A. Noyes Laboratory of Chemical Physics. This work was supported in part by research grant No. HE-02143 from the National Heart Institute, U.S. Public Health Service to the California Institute of Technology.

† Present address: Department of Physics, University of Madras, Madras-25, India.

‡ Present address: Physics Department, University of York, Heslington, York, England.

§ Private communication with Mann Research Laboratories, 1969.

Intensity data were measured on a General Electric XRD-5 automatic diffractometer with nickel-filtered Cu  $K\alpha$  radiation and a  $\theta$ - $2\theta$  scanning technique. The crystal was a square prism of dimensions  $0.10 \times 0.10 \times 0.13$  mm, sliced from a longer needle. A scanning speed of  $1^\circ$  per minute was employed and the scan range varied from  $2^\circ$  at  $2\theta=20^\circ$  to  $4^\circ$  at  $2\theta=120^\circ$ . Background was measured for thirty seconds at the beginning and end of each scan. As a check on the sta-

bility of the crystal and the instrument, the 250 reflection was remeasured every fifteen reflections: no significant variation in its intensity was noted. Two quadrants of data (positive and negative values of both  $h$  and  $k$ ) totalling 2382 reflections were measured, corresponding to approximately 96% of the total theoretically possible within the copper sphere. The intensities of 31 reflections turned out to be negative and were set to zero. Each intensity was assigned a variance derived from counting statistics with an additional term ( $0.0004S^2$ , where  $S$  is the scan count) to allow for systematic errors.

The intensities and their standard deviations were corrected for Lorentz and polarization factors and placed on an absolute scale by the method of Wilson (1942). The calculated value of the mass absorption coefficient ( $\mu_{Cu K\alpha}$ ) is  $31 \text{ cm}^{-1}$  leading to values of  $\mu R$  of the order of 0.3 for the crystal used for data collection. Thus, absorption corrections were deemed unnecessary.

### Structure determination and refinement

The coordinates of the chloride ion were determined from a three-dimensional, sharpened Patterson function. Four Fourier iterations led to a complete trial structure and an  $R$  value ( $(\sum |k|F_o - |F_c|) / \sum k|F_o|$ ) of 0.23. All calculations were performed on an IBM 7094 computer using sub-programs operating under the *CRYRM* crystallographic system (Duchamp, 1964).

Initial refinement was carried out using a full-matrix, least-squares routine on positional and individual isotropic temperature parameters and a scale factor. The quantity minimized was  $\sum w(|F_o|^2 - 1/k^2|F_c|^2)^2$ ,  $k$  being

the scale factor and  $w$  equal to  $1/\sigma^2(|F_o|^2)$ . Only one quadrant of data ( $k$  and  $l$  positive) was used in these early stages and no correction for anomalous scattering was applied to the scattering factor of the chloride ion. Scattering factors were taken from *International Tables for X-ray Crystallography* (1962). When the  $R$  value had been reduced to 0.078, difference Fourier syntheses were calculated in the planes where hydrogen atoms were expected to lie. Peaks ranging in height from 0.2 to  $0.5 \text{ e. \AA}^{-3}$  were observed for all fourteen hydrogen atoms. These atoms were included in subsequent structure factor calculations with isotropic temperature factors of  $3.0 \text{ \AA}^2$ , but their coordinates were not refined. Scattering factors used for the hydrogen atoms were those calculated by Stewart, Davidson & Simpson (1965). Refinement was continued with anisotropic temperature factors applied to the non-hydrogen atoms; when parameter shifts were less than one quarter of their e.s.d.'s, positional and isotropic thermal parameters for only the hydrogen atoms were refined.

After four cycles of hydrogen-atom refinement both the D and L enantiomorphs were dealt with separately, applying real and imaginary anomalous dispersion corrections of  $0.3e$  and  $0.7e$  respectively to the scattering factor for chlorine (*International Tables for X-ray Crystallography*, 1962), and including the complete data set ( $k$  both positive and negative). For space group  $P2_1$  with one or more anomalous scatterers,  $F_{hkl} \neq F_{\bar{h}\bar{k}l}$ . Refinement was then continued alternately on heavy-atom coordinates with anisotropic temperature factors, and hydrogen-atom coordinates with isotropic temperature factors, until all parameter shifts were less than one tenth of their e.s.d.'s for both enantiomorphs.

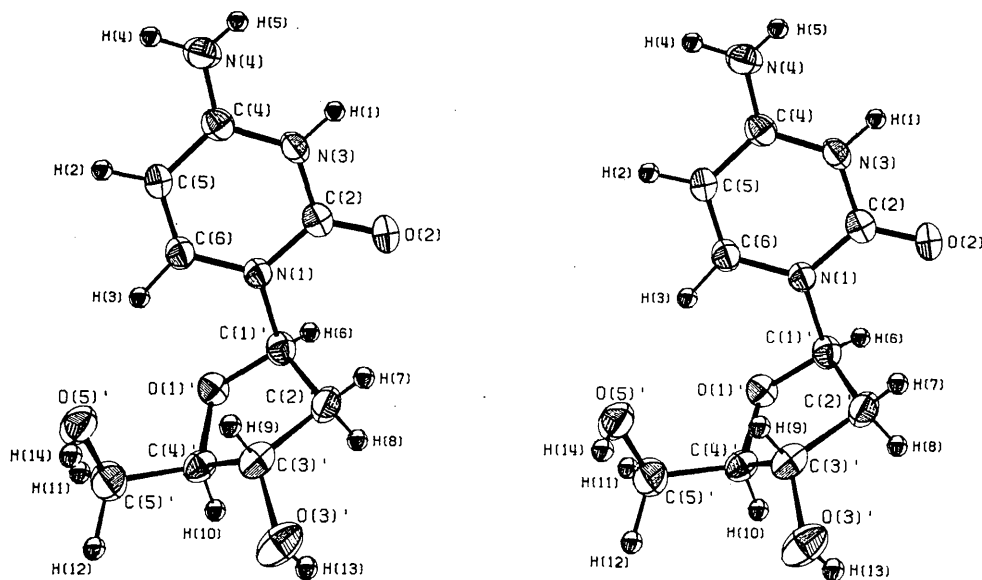


Fig. 1. A stereoscopic view of the 2'-deoxycytidine cation. The heavy atoms are represented by ellipsoids (drawn at the 50% probability level) defined by the principal axes of thermal motion.

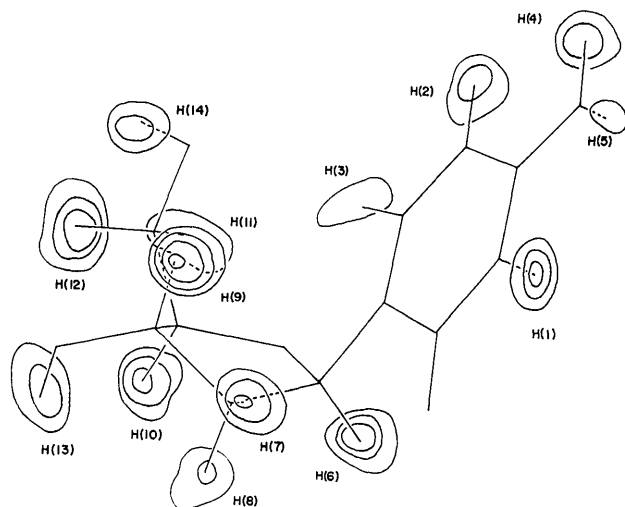


Fig. 2. The composite difference electron-density map, for which the contributions of the hydrogen atoms were excluded from the calculated structure factors. Contours are drawn at intervals of 0.1 e.Å<sup>-3</sup> starting at 0.2 e.Å<sup>-3</sup>.

## Results and discussion

### The absolute configuration of the molecule

In Table 2 are listed the final  $R$  values and the 'goodness of fit' for the D and L enantiomorphs; the agreement is conclusively better for the D enantiomorph. A stereoscopic view of the deoxycytidine cation is shown in Fig. 1, for which the ORTEP program was used (Johnson, 1965). This configuration for the D sugar is in agreement with previous conclusions as to its absolute configuration.

Table 2. The final  $R$  values and 'goodness of fit' for both enantiomorphs

	D	L
$R = (\sum  k F_o  -  F_c ) / \sum k F_o $	0.035	0.061
$[\sum w(k^2 F_o ^2 -  F_c ^2)] / \sum w F_o ^4$	0.003	0.013
'Goodness of fit' = $[\frac{\sum w( F_o ^2 -  F_c ^2/k^2)^2}{m-s}]^{1/2}$	1.31	2.62

In Table 3 are listed the final atomic parameters together with their e. s. d.'s for the non-hydrogen atoms;

Table 3. The final positional and thermal parameters and their e. s. d.'s for the heavy atoms

The temperature factors are in the form

$$T = \exp [-(b_{11}h^2 + b_{22}k^2 + b_{33}l^2 + b_{12}hk + b_{13}hl + b_{23}kl)].$$

All values have been multiplied by 10<sup>5</sup>.

	$x$	$y$	$z$	$b_{11}$	$b_{22}$	$b_{33}$	$b_{12}$	$b_{13}$	$b_{23}$
N(1)	65477 (24)	56324 (10)	13943 (36)	1296 (40)	169 (5)	2406 (80)	-77 (24)	701 (95)	-52 (34)
C(2)	76371 (32)	60340 (12)	37243 (46)	1616 (52)	207 (7)	2327 (92)	-173 (31)	946 (118)	24 (43)
N(3)	63697 (25)	65067 (10)	47324 (36)	1624 (45)	179 (5)	2154 (79)	-168 (25)	281 (99)	-190 (34)
C(4)	42444 (31)	66316 (11)	35179 (44)	1736 (52)	164 (6)	2755 (97)	-186 (29)	1327 (119)	44 (42)
C(5)	32530 (30)	62368 (13)	10539 (48)	1241 (49)	232 (7)	3286 (103)	-117 (30)	437 (117)	-352 (47)
C(6)	44134 (29)	57536 (12)	00837 (44)	1397 (48)	190 (6)	2802 (98)	-119 (29)	638 (118)	-59 (45)
O(2)	95523 (21)	59770 (9)	48607 (35)	1369 (37)	290 (5)	3419 (75)	-105 (23)	-147 (93)	-281 (34)
N(4)	32145 (28)	71005 (10)	46346 (41)	2128 (51)	230 (6)	3408 (94)	-97 (30)	2034 (115)	-460 (40)
C(1')	78957 (30)	51335 (12)	01950 (46)	1391 (48)	227 (7)	2577 (98)	-104 (30)	1156 (116)	-78 (44)
C(2')	88022 (32)	44614 (14)	19966 (46)	1485 (48)	265 (7)	2705 (98)	329 (31)	507 (118)	1 (47)
C(3')	70407 (33)	38752 (13)	09416 (45)	1734 (53)	224 (7)	2422 (95)	299 (30)	1705 (117)	243 (43)
C(4')	62750 (30)	40361 (12)	-21163 (42)	1561 (49)	194 (7)	2314 (92)	174 (29)	1498 (111)	38 (42)
C(5')	39631 (35)	38365 (13)	-36182 (50)	1851 (61)	276 (8)	3377 (117)	-107 (35)	1538 (136)	-338 (51)
O(1')	65792 (22)	48502 (9)	-23247 (29)	1782 (37)	194 (4)	1967 (64)	-73 (22)	1023 (81)	107 (29)
O(3')	76293 (24)	31115 (9)	16086 (35)	2406 (46)	248 (5)	4439 (91)	446 (26)	3124 (107)	833 (37)
O(5')	25471 (22)	40940 (9)	-21958 (35)	1553 (38)	236 (5)	4303 (89)	97 (23)	1932 (98)	174 (37)
Cl <sup>-</sup>	83510 (7)	75000	99685 (11)	1866 (13)	236 (2)	3161 (25)	-80 (9)	1175 (30)	-264 (13)

Table 4. Final positional and thermal parameters and their e. s. d.'s for the hydrogen atoms

The coordinates have been multiplied by 10<sup>4</sup>.

	$x$	$y$	$z$	$B$
H(1)	7090*(32)	6782*(11)	6249*(48)	3.0 (0.5)
H(2)	1749*(35)	6289 (15)	0078*(53)	5.1 (0.6)
H(3)	3798*(37)	5446 (13)	-1789*(53)	4.3 (0.6)
H(4)	1749 (39)	7141 (15)	4051 (57)	6.1 (0.7)
H(5)	3800 (30)	7311 (11)	6065 (44)	2.9 (0.5)
H(6)	9000 (34)	5463 (13)	-0147 (52)	3.7 (0.5)
H(7)	9241 (33)	4561 (12)	3982 (48)	3.8 (0.5)
H(8)	10108 (33)	4281 (12)	1519 (47)	3.6 (0.5)
H(9)	5875 (27)	4001 (11)	1764 (38)	1.9 (0.4)
H(10)	7147 (29)	3771 (11)	-3013 (41)	2.5 (0.4)
H(11)	3561 (33)	4062 (13)	-5588 (47)	3.9 (0.5)
H(12)	3828 (30)	3240 (12)	-3902 (47)	3.4 (0.5)
H(13)	8600 (37)	3000 (15)	1353 (54)	5.1 (0.6)
H(14)	2274 (39)	3718 (14)	-1455 (56)	5.4 (0.7)

Table 5. Final observed and calculated structure factors

The columns are k, 10|F\_o|, 10σ(F\_c|), 10|F\_c|, and the phase angle in degrees. The notation --- for |F\_o| means that |F\_o| was assigned the value zero; \*\* for the standard deviation indicates that the reflection was not included in the least-squares refinement.

Table with multiple columns containing numerical data for structure factors, including observed values (F\_o), calculated values (F\_c), and phase angles. The table is organized into several vertical sections, each starting with a small header (e.g., 'k', '10|F\_o|', '10σ(F\_c|)', '10|F\_c|', 'Phase'). The data points are arranged in rows, with some rows containing multiple values for a single reflection, indicating different measurements or calculations. The values range from small integers to several hundred, and the phase angles are given in degrees, often with a sign indicating the quadrant.

Table 5 (cont.)

-17 28 3 28 160	1 26 5 18 234	-12 125 2 124 7	12 275 5 31 887	-1 32 5 11 112	-12 15 5 10 124	-16 31 2 10 104	0 8 5 5	-3 97 3 43 184	-2 4 6 6
-18 86 2 155 176	4 116 5 117 166	-11 36 5 155 296	13 36 5 155 296	-10 32 5 11 216	-11 16 5 10 116	-11 16 5 10 116	-11 16 5 10 116	-8 36 5 26 81	-8 36 5 26 81
-19 58 2 75 144	6 106 3 86 102	-10 66 5 155 296	14 43 2 155 296	-9 32 5 11 216	-10 16 5 10 116	-10 16 5 10 116	-10 16 5 10 116	-7 26 7 19 229	-7 26 7 19 229
-20 58 2 75 144	7 101 2 155 176	-9 96 5 155 296	15 50 2 155 296	-8 32 5 11 216	-9 16 5 10 116	-9 16 5 10 116	-9 16 5 10 116	-6 16 5 10 116	-6 16 5 10 116
-21 58 2 75 144	8 96 2 155 176	-8 126 5 155 296	16 57 2 155 296	-7 32 5 11 216	-8 16 5 10 116	-8 16 5 10 116	-8 16 5 10 116	-5 16 5 10 116	-5 16 5 10 116
-22 58 2 75 144	9 91 2 155 176	-7 156 5 155 296	17 64 2 155 296	-6 32 5 11 216	-7 16 5 10 116	-7 16 5 10 116	-7 16 5 10 116	-4 16 5 10 116	-4 16 5 10 116
-23 58 2 75 144	10 86 2 155 176	-6 186 5 155 296	18 71 2 155 296	-5 32 5 11 216	-6 16 5 10 116	-6 16 5 10 116	-6 16 5 10 116	-3 16 5 10 116	-3 16 5 10 116
-24 58 2 75 144	11 81 2 155 176	-5 216 5 155 296	19 78 2 155 296	-4 32 5 11 216	-5 16 5 10 116	-5 16 5 10 116	-5 16 5 10 116	-2 16 5 10 116	-2 16 5 10 116
-25 58 2 75 144	12 76 2 155 176	-4 246 5 155 296	20 85 2 155 296	-3 32 5 11 216	-4 16 5 10 116	-4 16 5 10 116	-4 16 5 10 116	-1 16 5 10 116	-1 16 5 10 116
-26 58 2 75 144	13 71 2 155 176	-3 276 5 155 296	21 92 2 155 296	-2 32 5 11 216	-3 16 5 10 116	-3 16 5 10 116	-3 16 5 10 116	0 16 5 10 116	0 16 5 10 116
-27 58 2 75 144	14 66 2 155 176	-2 306 5 155 296	22 99 2 155 296	-1 32 5 11 216	-2 16 5 10 116	-2 16 5 10 116	-2 16 5 10 116	1 16 5 10 116	1 16 5 10 116
-28 58 2 75 144	15 61 2 155 176	-1 336 5 155 296	23 106 2 155 296	0 32 5 11 216	-1 16 5 10 116	-1 16 5 10 116	-1 16 5 10 116	2 16 5 10 116	2 16 5 10 116
-29 58 2 75 144	16 56 2 155 176	0 366 5 155 296	24 113 2 155 296	1 32 5 11 216	0 16 5 10 116	0 16 5 10 116	0 16 5 10 116	3 16 5 10 116	3 16 5 10 116
-30 58 2 75 144	17 51 2 155 176	1 396 5 155 296	25 120 2 155 296	2 32 5 11 216	1 16 5 10 116	1 16 5 10 116	1 16 5 10 116	4 16 5 10 116	4 16 5 10 116
-31 58 2 75 144	18 46 2 155 176	2 426 5 155 296	26 127 2 155 296	3 32 5 11 216	2 16 5 10 116	2 16 5 10 116	2 16 5 10 116	5 16 5 10 116	5 16 5 10 116
-32 58 2 75 144	19 41 2 155 176	3 456 5 155 296	27 134 2 155 296	4 32 5 11 216	3 16 5 10 116	3 16 5 10 116	3 16 5 10 116	6 16 5 10 116	6 16 5 10 116
-33 58 2 75 144	20 36 2 155 176	4 486 5 155 296	28 141 2 155 296	5 32 5 11 216	4 16 5 10 116	4 16 5 10 116	4 16 5 10 116	7 16 5 10 116	7 16 5 10 116
-34 58 2 75 144	21 31 2 155 176	5 516 5 155 296	29 148 2 155 296	6 32 5 11 216	5 16 5 10 116	5 16 5 10 116	5 16 5 10 116	8 16 5 10 116	8 16 5 10 116
-35 58 2 75 144	22 26 2 155 176	6 546 5 155 296	30 155 2 155 296	7 32 5 11 216	6 16 5 10 116	6 16 5 10 116	6 16 5 10 116	9 16 5 10 116	9 16 5 10 116
-36 58 2 75 144	23 21 2 155 176	7 576 5 155 296	31 162 2 155 296	8 32 5 11 216	7 16 5 10 116	7 16 5 10 116	7 16 5 10 116	10 16 5 10 116	10 16 5 10 116
-37 58 2 75 144	24 16 2 155 176	8 606 5 155 296	32 169 2 155 296	9 32 5 11 216	8 16 5 10 116	8 16 5 10 116	8 16 5 10 116	11 16 5 10 116	11 16 5 10 116
-38 58 2 75 144	25 11 2 155 176	9 636 5 155 296	33 176 2 155 296	10 32 5 11 216	9 16 5 10 116	9 16 5 10 116	9 16 5 10 116	12 16 5 10 116	12 16 5 10 116
-39 58 2 75 144	26 6 2 155 176	10 666 5 155 296	34 183 2 155 296	11 32 5 11 216	10 16 5 10 116	10 16 5 10 116	10 16 5 10 116	13 16 5 10 116	13 16 5 10 116
-40 58 2 75 144	27 1 2 155 176	11 696 5 155 296	35 190 2 155 296	12 32 5 11 216	11 16 5 10 116	11 16 5 10 116	11 16 5 10 116	14 16 5 10 116	14 16 5 10 116
-41 58 2 75 144	28 26 2 155 176	12 726 5 155 296	36 197 2 155 296	13 32 5 11 216	12 16 5 10 116	12 16 5 10 116	12 16 5 10 116	15 16 5 10 116	15 16 5 10 116
-42 58 2 75 144	29 31 2 155 176	13 756 5 155 296	37 204 2 155 296	14 32 5 11 216	13 16 5 10 116	13 16 5 10 116	13 16 5 10 116	16 16 5 10 116	16 16 5 10 116
-43 58 2 75 144	30 36 2 155 176	14 786 5 155 296	38 211 2 155 296	15 32 5 11 216	14 16 5 10 116	14 16 5 10 116	14 16 5 10 116	17 16 5 10 116	17 16 5 10 116
-44 58 2 75 144	31 41 2 155 176	15 816 5 155 296	39 218 2 155 296	16 32 5 11 216	15 16 5 10 116	15 16 5 10 116	15 16 5 10 116	18 16 5 10 116	18 16 5 10 116
-45 58 2 75 144	32 46 2 155 176	16 846 5 155 296	40 225 2 155 296	17 32 5 11 216	16 16 5 10 116	16 16 5 10 116	16 16 5 10 116	19 16 5 10 116	19 16 5 10 116
-46 58 2 75 144	33 51 2 155 176	17 876 5 155 296	41 232 2 155 296	18 32 5 11 216	17 16 5 10 116	17 16 5 10 116	17 16 5 10 116	20 16 5 10 116	20 16 5 10 116
-47 58 2 75 144	34 56 2 155 176	18 906 5 155 296	42 239 2 155 296	19 32 5 11 216	18 16 5 10 116	18 16 5 10 116	18 16 5 10 116	21 16 5 10 116	21 16 5 10 116
-48 58 2 75 144	35 61 2 155 176	19 936 5 155 296	43 246 2 155 296	20 32 5 11 216	19 16 5 10 116	19 16 5 10 116	19 16 5 10 116	22 16 5 10 116	22 16 5 10 116
-49 58 2 75 144	36 66 2 155 176	20 966 5 155 296	44 253 2 155 296	21 32 5 11 216	20 16 5 10 116	20 16 5 10 116	20 16 5 10 116	23 16 5 10 116	23 16 5 10 116
-50 58 2 75 144	37 71 2 155 176	21 996 5 155 296	45 260 2 155 296	22 32 5 11 216	21 16 5 10 116	21 16 5 10 116	21 16 5 10 116	24 16 5 10 116	24 16 5 10 116
-51 58 2 75 144	38 76 2 155 176	22 1026 5 155 296	46 267 2 155 296	23 32 5 11 216	22 16 5 10 116	22 16 5 10 116	22 16 5 10 116	25 16 5 10 116	25 16 5 10 116
-52 58 2 75 144	39 81 2 155 176	23 1056 5 155 296	47 274 2 155 296	24 32 5 11 216	23 16 5 10 116	23 16 5 10 116	23 16 5 10 116	26 16 5 10 116	26 16 5 10 116
-53 58 2 75 144	40 86 2 155 176	24 1086 5 155 296	48 281 2 155 296	25 32 5 11 216	24 16 5 10 116	24 16 5 10 116	24 16 5 10 116	27 16 5 10 116	27 16 5 10 116
-54 58 2 75 144	41 91 2 155 176	25 1116 5 155 296	49 288 2 155 296	26 32 5 11 216	25 16 5 10 116	25 16 5 10 116	25 16 5 10 116	28 16 5 10 116	28 16 5 10 116
-55 58 2 75 144	42 96 2 155 176	26 1146 5 155 296	50 295 2 155 296	27 32 5 11 216	26 16 5 10 116	26 16 5 10 116	26 16 5 10 116	29 16 5 10 116	29 16 5 10 116
-56 58 2 75 144	43 101 2 155 176	27 1176 5 155 296	51 302 2 155 296	28 32 5 11 216	27 16 5 10 116	27 16 5 10 116	27 16 5 10 116	30 16 5 10 116	30 16 5 10 116
-57 58 2 75 144	44 106 2 155 176	28 1206 5 155 296	52 309 2 155 296	29 32 5 11 216	28 16 5 10 116	28 16 5 10 116	28 16 5 10 116	31 16 5 10 116	31 16 5 10 116
-58 58 2 75 144	45 111 2 155 176	29 1236 5 155 296	53 316 2 155 296	30 32 5 11 216	29 16 5 10 116	29 16 5 10 116	29 16 5 10 116	32 16 5 10 116	32 16 5 10 116
-59 58 2 75 144	46 116 2 155 176	30 1266 5 155 296	54 323 2 155 296	31 32 5 11 216	30 16 5 10 116	30 16 5 10 116	30 16 5 10 116	33 16 5 10 116	33 16 5 10 116
-60 58 2 75 144	47 121 2 155 176	31 1296 5 155 296	55 330 2 155 296	32 32 5 11 216	31 16 5 10 116	31 16 5 10 116	31 16 5 10 116	34 16 5 10 116	34 16 5 10 116
-61 58 2 75 144	48 126 2 155 176	32 1326 5 155 296	56 337 2 155 296	33 32 5 11 216	32 16 5 10 116	32 16 5 10 116	32 16 5 10 116	35 16 5 10 116	35 16 5 10 116
-62 58 2 75 144	49 131 2 155 176	33 1356 5 155 296	57 344 2 155 296	34 32 5 11 216	33 16 5 10 116	33 16 5 10 116	33 16 5 10 116	36 16 5 10 116	36 16 5 10 116
-63 58 2 75 144	50 136 2 155 176	34 1386 5 155 296	58 351 2 155 296	35 32 5 11 216	34 16 5 10 116	34 16 5 10 116	34 16 5 10 116	37 16 5 10 116	37 16 5 10 116
-64 58 2 75 144	51 141 2 155 176	35 1416 5 155 296	59 358 2 155 296	36 32 5 11 216	35 16 5 10 116	35 16 5 10 116	35 16 5 10 116	38 16 5 10 116	38 16 5 10 116
-65 58 2 75 144	52 146 2 155 176	36 1446 5 155 296	60 365 2 155 296	37 32 5 11 216	36 16 5 10 116	36 16 5 10 116	36 16 5 10 116	39 16 5 10 116	39 16 5 10 116
-66 58 2 75 144	53 151 2 155 176	37 1476 5 155 296	61 372 2 155 296	38 32 5 11 216	37 16 5 10 116	37 16 5 10 116	37 16 5 10 116	40 16 5 10 116	40 16 5 10 116
-67 58 2 75 144	54 156 2 155 176	38 1506 5 155 296	62 379 2 155 296	39 32 5 11 216	38 16 5 10 116	38 16 5 10 116	38 16 5 10 116	41 16 5 10 116	41 16 5 10 116
-68 58 2 75 144	55 161 2 155 176	39 1536 5 155 296	63 386 2 155 296	40 32 5 11 216	39 16 5 10 116	39 16 5 10 116	39 16 5 10 116	42 16 5 10 116	42 16 5 10 116
-69 58 2 75 144	56 166 2 155 176	40 1566 5 155 296	64 393 2 155 296	41 32 5 11 216	40 16 5 10 116	40 16 5 10 116	40 16 5 10 116	43 16 5 10 116	43 16 5 10 116
-70 58 2 75 144	57 171 2 155 176	41 1596 5 155 296	65 400 2 155 296	42 32 5 11 216	41 16 5 10 116	41 16 5 10 116	41 16 5 10 116	44 16 5 10 116	44 16 5 10 116
-71 58 2 75 144	58 176 2 155 176	42 1626 5 155 296	66 407 2 155 296	43 32 5 11 216	42 16 5 10 116	42 16 5 10 116	42 16 5 10 116	45 16 5 10 116	45 16 5 10 116
-72 58 2 75 144	59 181 2 155 176	43 1656 5 155 296	67 414 2 155 296	44 32 5 11 216	43 16 5 10 116	43 16 5 10 116	43 16 5 10 116	46 16 5 10 116	46 16 5 10 116
-73 58 2 75 144	60 186 2 155 176	44 1686 5 155 296	68 421 2 155 296	45 32 5 11 216	44 16 5 10 116	44 16 5 10 116	44 16 5 10 116	47 16 5 10 116	47 16 5 10 116
-74 58 2 75 144	61 191 2 155 176	45 1716 5 155 296	69 428 2 155 296	46 32 5 11 216	45 16 5 10 116	45 16 5 10 116	45 16 5 10 116	48 16 5 10 116	48 16 5 10 116
-75 58 2 75 144	62 196 2 155 176	46 1746 5 155 296	70 435 2 155 296	47 32 5 11 216	46 16 5 10 116	46 16 5 10 116	46 16 5 10 116	49 16 5 10 116	49 16 5 10 116
-76 58 2 75 144	63 201 2 155 176	47 1776 5 155 296	71 442 2 155 296	48 32 5 11 216	47 16 5 10 116	47 16 5 10 116	47 16 5 10 116	50 16 5 10 116	50 16 5 10 116
-77 58 2 75 144	64 206 2 155 176	48 1806 5 155 296	72 449 2 155 296	49 32 5 11 216	48 16 5 10 116	48 16 5 10 116	48 16 5 10 116	51 16 5 10 116	51 16 5 10 116
-78 58 2 75 144	65 211 2 155 176	49 1836 5 155 296	73 456 2 155 296	50 32 5 11 216	49 16 5 10 116	49 16 5 10 116	49 16 5 10 11		

gen atom, N(4), forms at best only very weak hydrogen bonds. The N(4)-O(3') ( $1-x, \frac{1}{2}+y, 1-z$ ) distance of 2.804 Å is favourable but H(5) lies 39° from the N-O line; H(4) lies 26° from the line N(4)-Cl<sup>-</sup> ( $-1+x, y, -1+z$ ) and the N-Cl<sup>-</sup> distance is long (3.415 Å). Also, the distance N(4)-O(2) ( $-1+x, y, z$ ) is 3.146 Å but H(4) lies 48° from the N-O line.

Table 6. *Some intermolecular close contacts and hydrogen-bond angles*

Proposed hydrogen bond distances are identified by \*.

	Distance
O(1') (I) —N(1)	3.497 Å
O(1') (I) —C(2)	3.134
O(1') (I) —N(3)	3.274
O(1') (I) —O(2)	3.405
N(4) (II) —O(2)	3.146
C(6) (III) —O(2)	3.492
O(5') (III) —C(2')	3.283
O(3') (IV) —N(4)	2.804
N(3) —Cl <sup>-</sup>	3.131*
O(5') (IV) —Cl <sup>-</sup>	3.161*
O(3') (V) —Cl <sup>-</sup>	3.178*
N(4) (III) —Cl <sup>-</sup>	3.415

	Angle
C(4) —N(3) —Cl <sup>-</sup>	113.3°
C(2) —N(3) —Cl <sup>-</sup>	121.3
C(5') (IV) —O(5') (IV) —Cl <sup>-</sup>	97.2
C(3') (V) —O(3') (V) —Cl <sup>-</sup>	116.6
C(4) —N(4) —O(3') (IV)	160.5
C(4) (III) —N(4) (III) —Cl <sup>-</sup>	108.8
C(4) (II) —N(4) (II) —O(2)	97.8
H(4) —N(4) —O(3') (IV)	76.0
H(5) —N(4) —O(3') (IV)	39.0
H(14) (IV) —O(5') (IV) —Cl <sup>-</sup>	8.0
H(1) —N(3) —Cl <sup>-</sup>	6.0
H(13) (IV) —O(3') (V) —Cl <sup>-</sup>	6.0
H(4) (II) —N(4) (II) —O(2)	48.0
H(4) (III) —N(4) (III) —Cl <sup>-</sup>	26.0

(I) =  $x, y, 1+z$   
 (II) =  $1+x, y, z$   
 (III) =  $1+x, y, 1+z$   
 (IV) =  $1-x, \frac{1}{2}+y, 1-z$   
 (V) =  $2-x, \frac{1}{2}+y, 1-z$

Each chloride ion is surrounded by two oxygen atoms [O(3') ( $2-x, \frac{1}{2}+y, 1-z$ ) and O(5') ( $1-x, \frac{1}{2}+y, 1-z$ )] and a nitrogen atom, N(3), at distances of 3.178, 3.161, and 3.131 Å respectively. In each case the distances are less than the sum of the van der Waals radii for oxygen or nitrogen and chlorine atoms (Pauling, 1960) and the hydrogen atoms are directed towards the chloride ion; the deviations from linearity are given in Table 6.

In 2'-deoxycytidine hydrochloride, as in both the orthorhombic and monoclinic crystal structures of cytidylic acid *b* (Sundaralingam & Jensen, 1965; Bugg & Marsh, 1967), the carbonyl oxygen atom, O(2), fails to participate in hydrogen bonding.

The pyrimidine rings of molecules related by one unit-cell translation in the *c* direction lie in planes approximately 3.2 Å apart, but there is very little overlap when viewed perpendicular to the planes. The

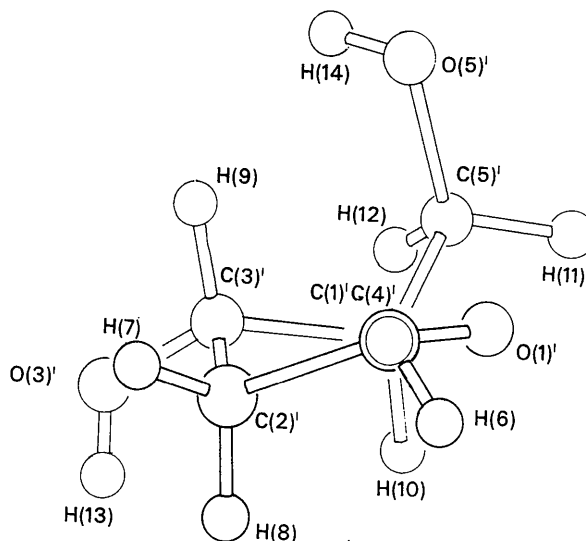


Fig. 4. The sugar residue viewed in the direction from C(1') to C(4').

Table 7. *Least-squares planes through the ring atoms of the sugar residue*

	Deviations (Å)					
	(a)	(b)	(c)	(d)	(e)	(f)
C(1')	-0.050	0.052	0.172	0.137	0.313†	0.000
O(1')	0.080	-0.054	-0.110	0.039†	0.120	0.000
C(4')	-0.075	0.032	-0.386†	-0.137	-0.188	0.000
C(3')	0.045	0.557†	0.098	0.219	0.181	0.361†
C(2')	-0.541†	-0.030	-0.159	-0.220	-0.113	-0.245†
C(5')	0.942†	0.881†	0.271†	0.698†	0.543†	0.935†

The equations of the least-squares planes, with coefficients equal to the direction cosines of the plane normal with respect to the crystallographic axes, are:

$$(a) \quad -0.9518X + 0.2715Y + 0.4306Z = -2.377$$

$$(b) \quad -0.9519X + 0.1158Y + 0.6495Z = -3.919$$

$$(c) \quad -0.8431X + 0.3570Y + 0.6437Z = -1.239$$

$$(d) \quad -0.9072X + 0.2905Y + 0.5707Z = -2.117$$

$$(e) \quad -0.8652X + 0.3778Y + 0.5817Z = -1.312$$

$$(f) \quad -0.9427X + 0.1727Y + 0.5636Z = -3.262$$

† Means an atom excluded from the calculation of the plane.

closest contact between the rings is 3.675 Å from C(6) ( $x, y, 1+z$ ) to C(4). Hence, it seems that base-stacking interactions play little part in stabilizing the crystal structure.

### The molecular conformation

The torsion angle (Sundaralingam & Jensen, 1965) describing the relative orientations of the sugar and base residues is  $0^\circ$ ;<sup>\*</sup> the atoms N(1), C(6), C(1'), and O(1') lie in a plane, within the accuracy of this determination. This conformation is significantly different from those reported to date for all other pyrimidine nucleosides and nucleotides with the possible exception

\* Two other, slightly different definitions for the torsion angle are commonly used (Donohue & Trueblood, 1960; Haschemeyer & Rich, 1967), resulting in small differences in values. For example, using the definition of Haschemeyer & Rich the torsion angle for 2'-deoxycytidine hydrochloride is  $-3.5^\circ$ .

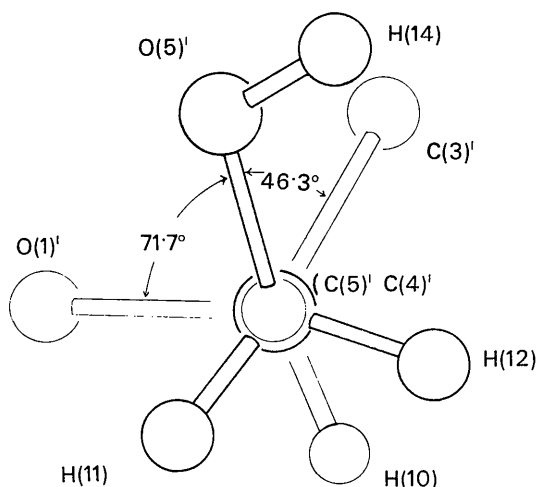
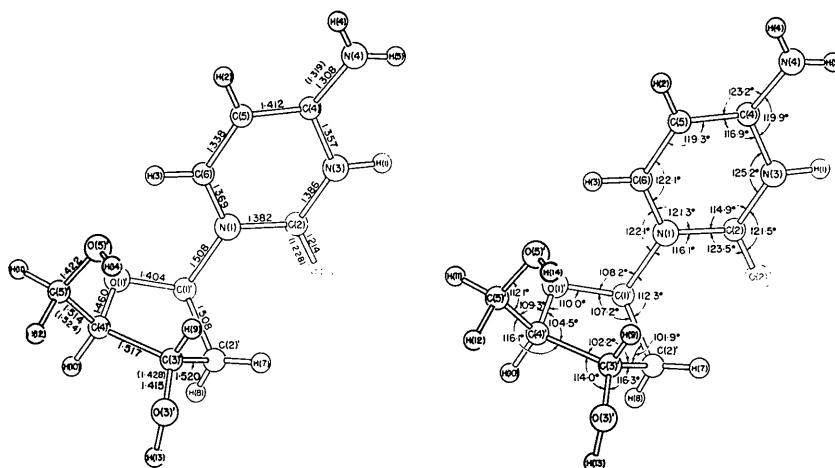


Fig. 5. The conformation of the bond C(5')-C(4').

of the uridine residue in  $\beta$ -adenosine-2'- $\beta$ -uridine-5'-phosphoric acid (Shefter, Barlow, Sparks & Trueblood, 1964) for which the torsion angle is reported as  $-5^\circ$ . For all other pyrimidine nucleosides and nucleotides the range of torsion angles is from  $-20$  to  $-65^\circ$ . The conformation adopted in 2'-deoxycytidine hydrochloride would seem to be unfavourable sterically, in that it results in O(1') being at its minimum possible distances from C(6) (2.679 Å) and H(3) (2.20 Å), which are respectively 0.17 and 0.15 Å smaller than the 'normal' intramolecular distances listed by Haschemeyer & Rich (1967). Usually, base-to-sugar close contacts in pyrimidine nucleosides and nucleotides are between the hydrogen atom attached to C(1') and the atoms C(2) and O(2). In the orthorhombic and monoclinic forms of cytidylic acid *b* (Sundaralingam & Jensen, 1965; Bugg & Marsh, 1967) and in 5-fluorodeoxyuridine (Harris & MacIntyre, 1964) these H-O distances are 2.23, 2.27, and 2.19 Å respectively, not significantly different from the O(1')-H(3) distance in 2'-deoxycytidine hydrochloride.

Table 7 shows all possible least-squares planes through four ring atoms of the sugar residue in 2'-deoxycytidine hydrochloride. Previously, for most pyrimidine nucleosides and nucleotides, the conformation of the sugar residue has been best described relative to one of these planes with either C(2') or C(3') significantly displaced from the plane. However, in 2'-deoxycytidine hydrochloride, the deviations of the atoms from these two planes (Table 7, planes *a, b*) are rather large;<sup>\*</sup> thus, the conformation is perhaps best described relative to the plane through the atoms C(1'), C(4'), and O(1'). Sundaralingam (1965) also makes use of this plane in a review article on sugar conformations. With respect to this plane in 2'-deoxycytidine hydrochloride

\* The root mean square deviations of the atoms from the planes excluding C(2') and C(3') are 0.064 and 0.043 Å respectively.



(Table 7, plane *f*), C(3') is displaced *endo* [to the same side of the plane as C(5')] by 0.361 Å and C(2') is 0.245 Å *exo*. The sugar residue viewed in the direction C(1')–C(4') is shown in Fig. 4. The puckering of the five-membered ring can also be described by individual torsion angles for the ring bonds – the dihedral angle between the planes defined by a bond and its two adjacent ring bonds (Brown & Levy, 1963). The values for 2'-deoxycytidine hydrochloride are C(1') → O(1') =  $-9.9^\circ$ , O(1') → C(4') =  $14.2^\circ$ , C(4') → C(3') =  $-31.9^\circ$ , C(3') → C(2') =  $36.9^\circ$ , and C(2') → C(1') =  $-29.5^\circ$ .

The conformation of the bond C(5')–C(4') can be seen in Fig. 5 which also shows the two projected valency angles  $\varphi_{OO} = 71.7^\circ$  and  $\varphi_{OC} = 46.3^\circ$  as defined by Shefter & Trueblood (1965). The conformation is *gauche-gauche*, the bond C(5')–O(5') being on the same side of the bonds O(1')–C(4') and C(3')–C(4'). This conformation occurs most frequently in nucleosides and nucleotides.

#### The molecular geometry

As in cytidylic acid *b* (Sundaralingam & Jensen, 1965; Bugg & Marsh, 1967), the deoxycytidine cation

is protonated at N(3). Intramolecular bond lengths and angles are shown in Fig. 6; the corresponding e.s.d.'s are 0.003 Å and  $0.2^\circ$  respectively. The values in parentheses in this Figure include estimated corrections for thermal motion on the basis of a riding model (Busing & Levy, 1964). However, discussion will be restricted to the uncorrected values.

The bond lengths and angles in the sugar residue agree quite well with those reported previously for other nucleosides and nucleotides and the values observed for the cytosine moiety are in very good agreement with those reported for cytidylic acid *b*. Probably the only significant difference is for the bond N(1)–C(1'), 1.508 (0.003) Å in 2'-deoxycytidine hydrochloride and 1.475 (0.006) Å in the orthorhombic form of cytidylic acid *b* and 1.485 (0.009) Å in the monoclinic form. The lengthening of the bond in 2'-deoxycytidine hydrochloride might be explained by repulsive forces between the electrons in the bonds N(1)–C(6) and C(1')–O(1'), which lie in the same plane. In the orthorhombic and monoclinic forms of cytidylic acid *b*, the angles between the analogous bonds are about  $42^\circ$  and  $39^\circ$  respectively.

The pyrimidine ring is slightly but significantly non-planar. The deviations from the least-squares plane

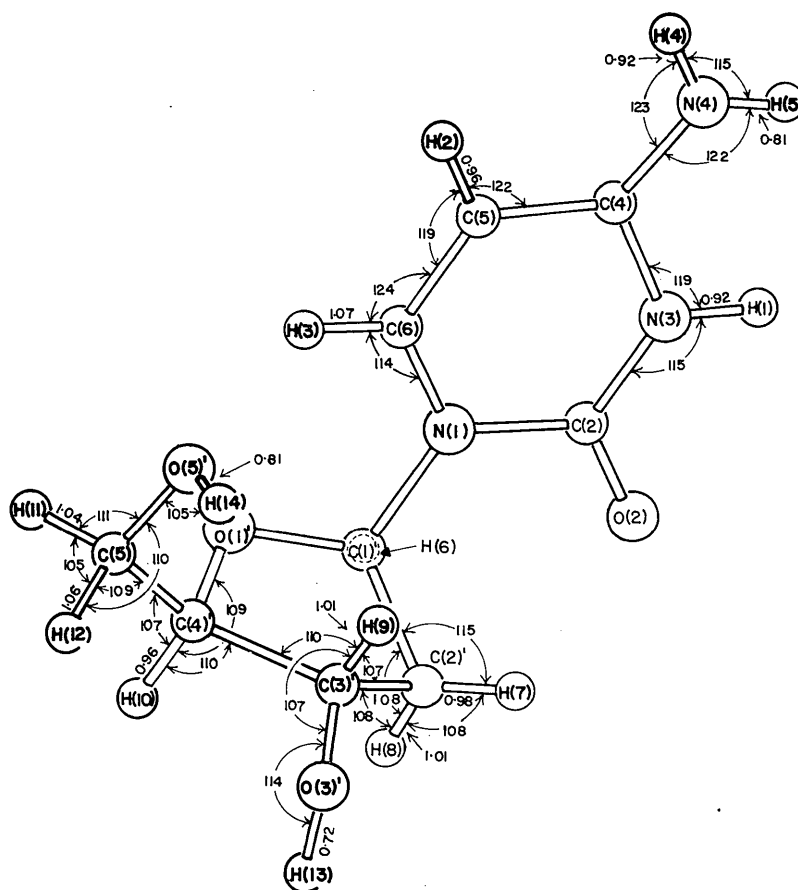


Fig. 7. Intramolecular bond distances and angles involving the hydrogen atoms. C(1')–H(6) = 0.99 Å, N(1)–C(1')–H(6) =  $107^\circ$ , O(1')–C(1')–H(6) =  $109^\circ$ , C(2')–C(1')–H(6) =  $114^\circ$ , C(3')–C(2')–H(7) =  $116^\circ$ , and C(4')–C(5')–H(11) =  $109^\circ$ .



through the six ring atoms are given in Table 8; the exocyclic atoms O(2), N(4), and C(1') are considerably displaced from this plane.

Table 8. *The least-squares plane through the six ring atoms of the pyrimidine base*

	Deviation
N(1)	0.024 Å
C(2)	-0.029
N(3)	0.012
C(4)	0.010
C(5)	-0.015
C(6)	-0.002
O(2)	-0.086†
N(4)	0.029†
C(1')	-0.065†
H(1)	-0.045†
H(2)	-0.009†
H(3)	-0.048†
H(4)	0.168†
H(5)	0.065†
Cl <sup>-</sup>	-0.105†

The equation of the least-squares plane, with coefficients equal to the direction cosines of the plane normal with respect to the crystallographic axes is

$$-0.3993X - 0.7422Y + 0.6354Z = -8.667.$$

† Means an atom excluded from the calculation of the plane.

The bond lengths and angles involving the hydrogen atoms are shown in Fig. 7; the e.s.d.'s in bond lengths and angles range respectively from 0.02 to 0.03 Å and from 1 to 2°. As is now commonly observed in X-ray diffraction investigations where electron density rather than nuclear density is examined, several bond lengths are significantly shorter than the accepted values (Pauling, 1960).

#### *Thermal motion*

It appears that the motions of all the heavy atoms are significantly anisotropic in that at least one of their  $b_{ij}$  terms (Table 3) differs by more than three e.s.d.'s

from the value it would have assuming isotropic motion with  $B$  equal to the mean principal axis  $B_i$ . Although corrections to several bond lengths were estimated on the basis of a simple riding model (Busing & Levy, 1964) and are shown in Fig. 6, no attempt was made to correct bond lengths for other librational effects.

We are grateful to Dr R. E. Marsh for helpful discussion and to Mrs J. Westphal for adapting certain computer programs to deal with the anomalous dispersion effect.

#### References

- BROWN, G. M. & LEVY, H. A. (1963). *Science*, **141**, 921.  
 BUGG, C. E. & MARSH, R. E. (1967). *J. Mol. Biol.* **25**, 67.  
 BUSING, W. R. & LEVY, H. A. (1964). *Acta Cryst.* **17**, 142.  
 DONOHUE, J. & TRUEBLOOD, K. N. (1960). *J. Mol. Biol.* **2**, 363.  
 DUCHAMP, D. J. (1964). A.C.A. Meeting, Bozeman, Montana, paper B-14, p. 29.  
 HARRIS, D. R. & MACINTYRE, W. M. (1964). *Biophys. J.* **4**, 203.  
 HASCHEMEYER, A. E. V. & RICH, A. (1967). *J. Mol. Biol.* **27**, 369.  
*International Tables for X-ray Crystallography* (1962). Vol. III. Birmingham: Kynoch Press.  
 JOHNSON, C. K. (1965). *A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations*. ORNL-3794, Oak Ridge, Tennessee.  
 PAULING, L. (1960). *The Nature of the Chemical Bond*, 3rd ed. Ithaca: Cornell Univ. Press.  
 SHEFTER, E., BARLOW, M., SPARKS, R. & TRUEBLOOD, K. (1964). *J. Amer. Chem. Soc.* **86**, 1872.  
 SHEFTER, E. & TRUEBLOOD, K. N. (1965). *Acta Cryst.* **18**, 1067.  
 STEWART, R. F., DAVIDSON, F. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175.  
 SUNDARALINGAM, M. (1965). *J. Amer. Chem. Soc.* **87**, 599.  
 SUNDARALINGAM, M. & JENSEN, L. H. (1965). *J. Mol. Biol.* **13**, 914.  
 WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.