Stereochemistry of Nucleic Acids and Their Constituents. XVII.[1a] Crystal and Molecular Structure of Deoxycytidine 5'-Phosphate Monohydrate.[1b] A Possible Puckering for the Furanoside Ring in B-Deoxyribonucleic Acid[1c]

M. A. Viswamitra,*[1d] B. Swaminatha Reddy,[1d]
George Hung-Yin Lin,[1e] and M. Sundaralingam*[1e]

Contribution from the Department of Physics, India Institute of Science,
Bangalore 12, India, and the Department of Biochemistry,
University of Wisconsin, Madison, Wisconsin 53706.
Received November 30, 1970

Abstract: The crystal structure of deoxycytidine 5'-phosphate monohydrate (5'-dCMP), $C_9H_{14}N_3O_7P \cdot H_2O$, has been determined by X-ray diffraction techniques. Two entirely independent investigations were carried out. In one case, 1149 intensity data were collected by multiple-film equiinclination Weissenberg technique using Cu Kα radiation. In the other case, 1248 three-dimensional intensity data were collected on a four-circle diffractometer, also using Cu Kα radiation. In both cases the structure was refined by full-matrix least-squares techniques. The final R values for the two investigations were 0.070 (from film data) and 0.035 (from diffractometer data); in the latter case, when secondary extinction corrections were applied the R value was reduced to 0.023. The unit cell is orthorhombic, space group $P2_12_12_1$, with average $a = 6.786$, $b = 11.345$, and $c = 16.769$ Å. The nucleotide exists as a zwitterion with N(1) protonated by one of the phosphate protons. A comparison of the geometries of the N(1)-protonated and neutral cytosine derivatives shows that there are marked differences in the bond angles and bond distances involving N(1), C(2), and C(4). A novel glycosyl torsion angle, −6.0°, and sugar conformation $C(3')_{exo}$–$C(2')_{exo}$, where both C(3') and C(2') are on the opposite side of C(5'), are observed in 5'-dCMP. The conformation about the C(4')–C(5') bond is gauche–gauche, the only conformation observed so far for the known 5'-nucleotides. A correlation between the furanoside ring conformations and the glycosyl torsion angles is presented. All available hydrogens participate in the hydrogen bonds. The molecules related by the screw axis parallel to the a axis are linked together by pairs of hydrogen bonds, resulting in an infinitely extended spiral around the screw axis. The crystal structure can be considered essentially a close packing of these infinite spirals linked by hydrogen bonds to each other and the water of crystallization.

The stereochemistry of none of the four deoxyribonucleotides, the basic monomer units of DNA (deoxyribonucleic acid), has been reported in the literature, although the structure of the calcium salt of thymidylic acid was reported by Trueblood, Horn, and Luzzati.[2] This paper describes the results of two entirely independent investigations of the structure of deoxycytidine 5'-phosphate (5'-dCMP) monohydrate (Figure 1). Since different experimental techniques had been used, it was considered appropriate to report both sets of results in the same paper and to make the necessary comparison between them. For convenience, we shall divide the experimental portions and the results of this report into two parts, describing the work of Viswamitra and Reddy (VR) and that of Lin and Sundaralingam (LS).

Experimental Section

The crystals for both investigations were prepared by VR. Crystals of deoxycytidine 5'-phosphate monohydrate are colorless prisms; some of them are as long as 10 mm. They were grown by slow diffusion of acetone (both liquid and vapor) through water solutions containing a wide range of concentrations of deoxycytidine 5'-phosphate. Crystal data for the two investigations are compared in Table I.

Table I. Crystal Data for Deoxycytidine 5'-Phosphate Monohydrate

	VR	LS
Space group	$P2_12_12_1$	
a	6.776 ± 0.002 Å	6.796 ± 0.002 Å
b	11.340 ± 0.003 Å	11.349 ± 0.003 Å
c	16.775 ± 0.006 Å	16.763 ± 0.004 Å
v	1289.0 Å^3	1292.9 Å^3
z	4	4
d_{obsd}	1.687 g cm^{-3}	1.669 g cm^{-3}
d_{calcd}	1.675 g cm^{-3}	1.671 g cm^{-3}
μ(Cu Kα)	23.63 cm^{-1}	23.54 cm^{-1}

VR. The intensity data were collected on a nearly cylindrical crystal of 0.32 mm diameter and about 1 mm long, by the multiple-film equiinclination Weissenberg technique, for layers $0kl$–$5kl$ and $h0l$ using Cu Kα radiation. A total of 1149 reflections was collected, comprising about 80% of the copper sphere. These intensities were estimated visually by comparison with a calibrated film strip. The data were then corrected for Lorentz and polarization factors and placed on a common arbitrary scale using the $h0l$ data. Absolute scaling of reflections was done by a Wilson plot. No correction for absorption was made.

LS. Three-dimensional intensity data in the range $0° \leq 2 \leq 127°$ were collected on a Picker four-circle automated diffractometer with Ni-filtered Cu Kα radiation. The θ–2θ scan mode with a scan rate of 2°/min was employed for the data collection. Background counts of 20 sec were measured at each end of the scan. Three standard reflections were checked at an interval of every 100 reflections, and they showed fluctuations in intensity of only ±2%. The data were corrected for Lorentz and polarization effects, but no

---

[1] (a) Part XVI of this series of papers is by J. A. Carrabine and M. Sundaralingam, *Biochemistry*, 10, 292 (1971). (b) Abbreviations used: cytidine 3'-phosphate, 3'-CMP; cytidine 3'-phosphate, orthorhombic form, 3'-CMP(O); cytidine 3'-phosphate, monoclinic form, 3'-CMP(M); deoxyribonucleic acid, DNA. (c) A preliminary communication on this structure has been published: M. A. Viswamitra and B. S. Reddy, *Z. Kristallogr.*, 131, 237 (1970). (d) India Institute of Science. (e) University of Wisconsin.

[2] K. N. Trueblood, P. Horn, and V. Luzzati, *Acta Crystallogr.*, 14, 965 (1961).

Table II. Positional Parameters and Anisotropic Thermal Parameters of the Nonhydrogen Atoms[a–c]

Atom	x	y	z	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
N(1)	3003 (11)	−198 (6)	−73 (4)	124 (18)	53 (5)	23 (2)	2 (14)	1 (10)	6 (5)
	3009 (3)	−209 (2)	−71 (1)	111 (6)	43 (2)	21 (1)	−4 (3)	1 (2)	5 (1)
C(2)	4312 (12)	505 (7)	−476 (4)	100 (20)	55 (5)	25 (2)	−10 (18)	−2 (11)	−3 (6)
	4308 (4)	494 (2)	−479 (1)	120 (7)	48 (2)	19 (1)	−13 (4)	−3 (2)	1 (1)
N(3)	3540 (4)	1527 (5)	−795 (4)	152 (19)	49 (4)	30 (2)	−13 (14)	−2 (10)	−7 (5)
	3537 (3)	1532 (2)	−794 (1)	113 (6)	45 (2)	23 (1)	−4 (3)	0 (2)	6 (1)
C(4)	1582 (12)	1822 (8)	−787 (4)	133 (20)	59 (5)	22 (2)	−3 (7)	5 (11)	−3 (6)
	1585 (4)	1807 (2)	−788 (1)	128 (7)	49 (2)	20 (1)	−2 (4)	−5 (2)	−2 (1)
C(5)	296 (13)	1052 (7)	−383 (4)	175 (22)	43 (5)	25 (2)	−8 (17)	−1 (10)	1 (6)
	293 (4)	1039 (2)	−380 (1)	116 (7)	53 (2)	23 (1)	−4 (3)	1 (2)	0 (1)
C(6)	1056 (13)	54 (7)	−34 (4)	157 (24)	62 (6)	22 (2)	−9 (20)	−6 (12)	0 (6)
	1049 (4)	49 (2)	−45 (1)	106 (7)	50 (2)	21 (1)	−12 (4)	2 (2)	0 (1)
O(2)	6039 (9)	242 (5)	−552 (3)	147 (16)	68 (4)	33 (2)	0 (14)	4 (9)	11 (5)
	6043 (3)	245 (2)	−559 (1)	105 (5)	62 (2)	30 (1)	2 (3)	5 (2)	9 (1)
N(4)	918 (12)	2760 (6)	−1159 (4)	165 (20)	61 (5)	36 (3)	14 (17)	7 (12)	11 (6)
	951 (4)	2759 (2)	−1160 (1)	140 (6)	51 (2)	33 (1)	5 (3)	2 (2)	12 (1)
C(1′)	3802 (13)	−1360 (3)	241 (4)	118 (22)	50 (5)	25 (2)	−4 (16)	4 (10)	2 (6)
	3813 (4)	−1359 (2)	244 (1)	136 (7)	39 (2)	22 (1)	−2 (3)	1 (2)	0 (1)
C(2′)	5474 (13)	−1182 (7)	846 (4)	183 (23)	53 (5)	26 (2)	−11 (18)	4 (13)	7 (6)
	5455 (4)	−1203 (2)	858 (1)	145 (7)	43 (2)	24 (1)	−8 (4)	−4 (2)	2 (1)
C(3′)	4839 (13)	−2003 (7)	1551 (5)	138 (22)	58 (6)	27 (3)	−1 (17)	−6 (12)	7 (7)
	4847 (4)	−2011 (2)	1539 (1)	154 (7)	42 (2)	23 (1)	11 (4)	−4 (2)	1 (1)
C(4′)	2641 (14)	−2070 (7)	1466 (5)	207 (25)	50 (6)	25 (3)	17 (18)	24 (13)	5 (7)
	2633 (4)	−2062 (2)	1464 (1)	169 (8)	41 (2)	26 (1)	0 (4)	12 (3)	7 (1)
O(1′)	2262 (9)	−1963 (5)	615 (3)	186 (16)	53 (4)	26 (2)	−23 (13)	−5 (8)	8 (4)
	2256 (2)	−1965 (1)	613 (1)	133 (5)	48 (2)	27 (1)	−20 (3)	−1 (2)	4 (1)
O(3′)	5680 (10)	−3165 (5)	1412 (3)	197 (17)	52 (4)	33 (2)	16 (14)	−5 (10)	1 (5)
	5691 (3)	−3160 (1)	1402 (1)	157 (5)	46 (2)	30 (1)	15 (3)	0 (2)	4 (1)
C(5′)	1495 (15)	−1161 (7)	1935 (5)	298 (31)	47 (5)	34 (3)	27 (21)	47 (16)	9 (7)
	1499 (5)	−1152 (2)	1928 (2)	237 (9)	44 (2)	35 (1)	20 (4)	35 (3)	12 (2)
O(5′)	2151 (10)	5 (5)	1700 (3)	215 (18)	47 (4)	30 (2)	10 (13)	6 (9)	4 (4)
	2147 (3)	19 (1)	1706 (1)	160 (6)	41 (2)	26 (1)	11 (3)	17 (2)	3 (1)
O(6)	−1296 (10)	832 (5)	1819 (4)	161 (18)	62 (4)	38 (2)	2 (14)	−3 (10)	7 (5)
	−1306 (3)	833 (2)	1817 (1)	124 (5)	65 (2)	30 (1)	3 (3)	−13 (2)	6 (1)
O(7)	1034 (10)	1216 (5)	2902 (3)	199 (19)	81 (5)	26 (2)	1 (16)	3 (9)	−2 (5)
	1049 (3)	1219 (2)	2892 (1)	152 (5)	79 (2)	22 (1)	1 (3)	2 (2)	−6 (1)
O(8)	1676 (10)	2145 (5)	1525 (4)	231 (18)	49 (4)	36 (2)	−3 (14)	6 (10)	5 (5)
	1664 (3)	2155 (2)	1530 (1)	182 (6)	48 (2)	33 (1)	−1 (3)	3 (2)	10 (1)
P	832 (4)	1122 (2)	1963 (1)	154 (8)	49 (1)	23 (1)	10 (4)	−2 (3)	2 (2)
	830 (1)	1121 (1)	1961 (0)	115 (2)	42 (1)	21 (0)	9 (1)	0 (1)	2 (0)
O(W)	3571 (10)	−4977 (5)	1963 (4)	221 (18)	67 (4)	33 (2)	−4 (15)	8 (9)	−1 (6)
	3580 (3)	−4980 (2)	1973 (1)	185 (6)	65 (2)	30 (1)	−4 (3)	−2 (2)	−5 (1)

[a] All parameters and their standard deviations given in parentheses have been multiplied by 10^4. [b] For each atom the results from VR and LS are given in the first line and second line, respectively. [c] The temperature factor is of the form $\exp[-(\beta_{11}h^2 + \ldots + 2\beta_{12}hk + \ldots)]$.

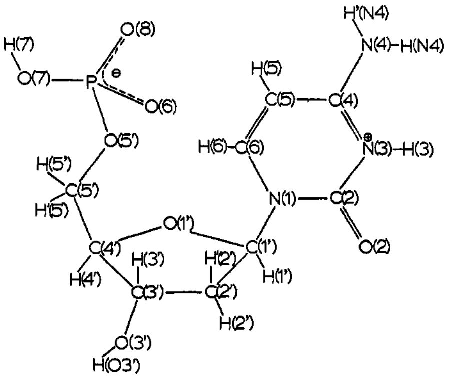


Figure 1. Numbering in deoxycytidine 5′-phosphate.

absorption corrections were made (μ = 23.5 cm^{-1} for Cu Kα radiation). A reflection was considered unobserved if the intensity was smaller than 1.5 times its standard deviation. On this basis, 1197 reflections were considered to be observed out of the total of 1248 reflections.

Structure Determination. VR. The position of the phosphorus atom (x = 0.083, y = 0.112, z = 0.196) was determined unambiguously from the Harker sections of a sharpened three-dimensional Patterson synthesis. A three-dimensional minimum function computed on the basis of the position of the phosphorus atom showed clearly the four oxygen atoms of the PO_4 group. However, the peaks for the remaining 16 heavy atoms could not be identified unambiguously since the minimum function had far too many peaks. A three-dimensional Fourier synthesis based on the phases of the phosphate group showed about the same number of peaks as found in the minimum function. It was noticed that among the 0*kl* structure factors calculated using the PO_4 group, the reflection (024) was calculated almost zero while it was observed to be the strongest. A Bragg–Lipson chart of this reflection, when superposed on the (100) Fourier projection computed with the phases of the PO_4 group indicated the sign of (024) to be positive. From this map we also picked out the ten strongest peaks likely to represent atom sites. A (100) Fourier projection computed incorporating this new information, however, did not show the molecule unambiguously. At this stage a three-dimensional model of the molecule was made with the atoms located at the peaks common to both the 3d minimum function and the 3d Fourier synthesis and also satisfying the (100) Fourier projection. This procedure gave in one single step the positions of all the remaining nonhydrogen atoms including that of the water oxygen.

LS. A sharpened three-dimensional Patterson map was computed and the position of the phosphorus atom was derived from the Harker peaks. Based on the contribution of the phosphorus atom to the structure the phase angles were calculated. Then the phases were refined by applying Karle and Hauptman's tangent formula.[3,4] 267 reflections with normalized structure factors

---

(3) J. Karle and H. Hauptman, *Acta Crystallogr.*, **9**, 635 (1956).

(4) J. Karle and I. L. Karle, *ibid.*, **21**, 869 (1966).

$E > 1.2$ were used for the refinement. The R value was 0.24, where $R = \Sigma||E_o| - |E_c||/\Sigma|E_o|$. A three-dimensional E map was then computed and it revealed the complete molecular structure.

Refinement of the Structure. VR. Atomic coordinates, individual layer scale factors, and isotropic temperature factors were refined with a block-diagonal-matrix least-squares program on an Elliot 803 computer to an R value of 0.11 (using a program written by G. A. Mair, Royal Institution, London,). The refinement of the heavy-atom parameters was continued on a CDC 3600 computer, using a full-matrix least-squares program (LALS, Trueblood, *et al.*). The total number of parameters in an anisotropic refinement of the heavy atoms would be 189, plus 9 layer scale factors. The refinement was carried out in two blocks of 10 and 11 atoms each, since a maximum of 160 parameters only could be refined simultaneously with the available computer program. The refinement dropped the R value to 0.082.

A three-dimensional difference Fourier synthesis computed at this stage gave the positions of 13 hydrogen atoms. Keeping the parameters of heavy atoms fixed, the parameters of the hydrogen atoms with isotropic temperature factors of 5 Å^2 were refined to an R value of 0.074. The positions of H(O-3′), H(O-7), and H(N-3) were located from a difference Fourier map subsequently. Further refinement of all the hydrogen atoms gave a final R value of 0.070.

The function minimized in the refinement was $\Sigma w(F_o - F_c)^2$, weighting function w employed being $1/(a + b|F_o| + c|F_o|^2 + d|F_o|^3)$,[5] where $a = 2F_{\min}$, $b = 1$, $c = 2/F_{\max}$, $d = 5/F_{\max}{}^2$. For hydrogen atoms the scattering factors of Stewart, Davidson, and Simpson[6] were used. For the other atoms scattering factors were computed using a function developed by Cromer and Waber.[7]

LS. The refinement was carried through on a UNIVAC 1108 computer using the full-matrix least-squares program of Busing, Martin, and Levy.[8] The Evans[9] weighting scheme was used.

Two cycles of isotropic least-squares refinement followed by two anisotropic cycles for the nonhydrogen atoms reduced the reliability index, R, from 0.26 to 0.064, where $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$; F_o is the observed and F_c is the calculated structure factor. At this stage the positional parameters of the hydrogen atoms were obtained from a difference Fourier map. The structure was subjected to three further least-squares cycles with anisotropic temperature factors for nonhydrogen atoms and isotropic temperature factors for hydrogen atoms. The R value for the 1197 observed reflections was 0.035 and the final shift, σ, was less than 0.11 for all nonhydrogen parameters and less than 0.40 for all hydrogen parameters. The data were finally corrected for secondary extinction according to Zacharisen[10] and the structure was refined by one least-squares cycle with anisotropic temperature factors for nonhydrogen atoms and isotropic temperature factors for hydrogen atoms. The R value was reduced to 0.023. The changes in bond distances and bond angles due to secondary extinction corrections were less than 0.67σ.

The scattering factors for P, O, N, and C atoms are from Cromer and Waber[7] and for H atoms from Stewart, Davidson, and Simpson.[6]

Results and Discussion

The final positional and thermal parameters for nonhydrogen atoms and hydrogen atoms are given in Tables II and III, respectively.[11] The bond distances and bond angles not involving hydrogen atoms are given in Table IV. All distances and angles involving hydrogen atoms are normal and they are not tabulated.

Table III. Positional Parameters and Isotropic Thermal Parameters of the Hydrogen Atoms[a,b]

Atom	x	y	z	B, Å^2
H(3)	450 (15)	201 (8)	−106 (5)	
	451 (4)	204 (2)	−108 (1)	3.0 (0.5)
H(N-4)	179 (14)	325 (8)	−143 (6)	
	186 (4)	320 (2)	−142 (2)	3.9 (0.6)
H′(N-4)	−58 (16)	289 (8)	−120 (6)	
	−57 (5)	284 (3)	−125 (2)	4.5 (0.7)
H(5)	−129 (14)	132 (8)	−28 (5)	
	−116 (4)	117 (2)	−39 (1)	3.1 (0.5)
H(6)	24 (14)	−62 (8)	23 (5)	
	20 (4)	−44 (2)	19 (1)	2.7 (0.6)
H(1′)	414 (14)	−192 (7)	−24 (5)	
	403 (4)	−183 (2)	−22 (1)	3.0 (0.5)
H(2′)	703 (14)	−136 (7)	76 (5)	
	674 (4)	−147 (2)	60 (1)	3.1 (0.5)
H′(2′)	559 (14)	−18 (8)	105 (5)	
	543 (4)	−44 (2)	103 (1)	3.1 (0.5)
H(3′)	570 (14)	−177 (8)	210 (5)	
	541 (4)	−173 (2)	209 (1)	4.1 (0.6)
H(4′)	219 (15)	−286 (8)	160 (4)	
	207 (4)	−288 (2)	164 (1)	4.1 (0.6)
H(O-3′)	468 (14)	−387 (8)	145 (5)	
	471 (5)	−379 (3)	170 (2)	6.6 (0.8)
H(5′)	1 (14)	−119 (8)	184 (5)	
	−9 (5)	−117 (3)	186 (2)	5.0 (0.7)
H′(5′)	191 (13)	−126 (7)	250 (5)	
	186 (5)	−127 (2)	253 (2)	5.6 (0.8)
H(7)	229 (15)	152 (7)	301 (5)	
	228 (6)	152 (3)	309 (2)	7.3 (0.8)
H(W)	244 (17)	−484 (8)	169 (6)	
	296 (4)	−526 (2)	172 (2)	4.2 (0.6)
H′(W)	298 (14)	−503 (7)	233 (6)	
	300 (5)	−463 (3)	241 (2)	6.2 (0.8)

[a] Positional parameters and their standard deviations in parentheses have been multiplied by 10^3; $B = 5$ Å^2 for all hydrogen atoms in VR. [b] For each atom the parameters from VR and LS are given in the first line and second line, respectively.

In general the bond distances and bond angles show good agreement in the two investigations. However, the standard deviations in bond distances and bond angles from LS ($\sigma(l)_{av}$ 0.003 Å, $\sigma(\theta)_{av}$ 0.2°) are about three times smaller than those from VR ($\sigma(l)_{av}$ 0.01 Å, $\sigma(\theta)_{av}$ 0.6°). In the following discussion the results of VR are given in parentheses alongside the results of LS.

Molecular Conformation and Geometry. Glycosyl Bond. An important stereochemical parameter in nucleosides and nucleotides is the glycosyl torsion angle χ_{CN} which describes the relative orientation of the base with respect to the sugar.[12a,b,13] The angle (as defined by Sundaralingam[13]) is −5.9° (−6.1°) in 5′-dCMP. Therefore, the conformation about the glycosyl bond is anti[12a,13] and is the first case with a small negative torsion angle. The distribution of the glycosyl torsion angles in the known β-pyrimidine glycosides is given as a function of the furanoside ring pucker in Figure 2. It is seen that, with the exception of the present investigation, the χ_{CN} values in the anti conformation lie between 0 and 70°. Moreover, for the C(3′)$_{endo}$ puckering $0° \leq \chi \leq 42°$, whereas for the C(2′)$_{endo}$ puckering $36° \leq \chi \leq 65°$.

As a consequence of the small χ_{CN} value, the steric interaction between the base and sugar increases; therefore, the glycosyl C(1′)-N(1) bond distance,

---

(5) D. W. J. Cruickshank, D. E. Pilling, A. Bujosa, F. M. Lovell, and M. R. Truter, "Computing Methods and the Phase Problem in X-Ray Crystal Analysis," Pergamon Press, New York, N. Y., 1961, p 32.

(6) R. F. Stewart, E. R. Davidson, and W. T. Simpson, *J. Chem. Phys.*, **42**, 3175 (1965).

(7) D. T. Cromer and J. T. Waber, *Acta Crystallogr.*, **18**, 104 (1965).

(8) W. R. Busing, K. O. Martin, and H. A. Levy, "ORFLS, A Fortran Crystallographic Least-Squares Program," ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1962.

(9) H. T. Evans, Jr., *Acta Crystallogr.*, **14**, 689 (1961).

(10) W. H. Zacharisen, *ibid.*, **16**, 1139 (1963).

(11) The observed and calculated structure amplitudes for the two determinations have been deposited as Document No. NAPS-01370 with the ASIS National Auxiliary Publication Service, c/o CCM Information Corp., 909 3rd Ave., New York, N. Y. 10022. A copy may be secured by citing the document number and by remitting $2.00 for microfiche or $5.00 for photocopies. Advance payment is required. Make checks or money orders payable to: CCMIC-NAPS.

(12) (a) J. Donohue and K. N. Trueblood, *J. Mol. Biol.*, **2**, 363 (1960); (b) M. Sundaralingam and L. H. Jensen, *ibid.*, **13**, 914 (1965).

(13) M. Sundaralingam, *Biopolymers*, **7**, 821 (1969).

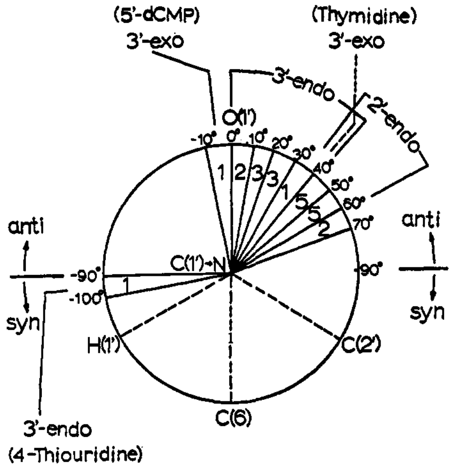


Figure 2. Distribution of the glycosyl torsion angles in the known β-pyrimidine glycosides. The numbers indicated in the sectors of 10° represent numbers of known compounds. For the $C(3')_{endo}$ conformation of a sugar χ is between 0 and 42°, while for the $C(2')_{endo}$ conformation χ is between 36 and 65° (see also A. E. V. Haschemeyer and A. Rich, *J. Mol. Biol.*, **27**, 369 (1967)). The subtle differences in the furanoside ring conformation of 5'-dCMP and thymidine (see text) may explain the differences in their χ angles. 4-Thiouridine (W. Saenger and K. H. Scheit, *ibid.*, **50**, 153 (1970)) is the only nucleoside with a syn χ angle and the sugar has the $C(3')_{endo}$ puckering. The definition of the glycosyl torsion angle, χ, as used throughout this paper is that given in ref 13. Note that the angles 180–360° in ref 13 are referred to as −180 to 0° here; angles 0–180° have the same meaning.

1.510 Å (1.519 Å), is the longest observed so far. The correlation between the χ angle and the C(1')–N bond distance has already been noted in previous work.[14,15]

The Deoxyribose. Sugar rings in nucleosides and nucleotides are usually puckered with either C(2') or C(3') out of the plane formed by the other four atoms.[16,17] In 5'-dCMP the best four-atom plane in the deoxyribose ring is defined by C(1'), C(2'), C(4'), and O(1') (Table V, plane II). C(3') is displaced 0.477 Å, lying on the opposite side of the sugar plane to C(5'). The conformation of the deoxyribose ring is therefore $C(3')_{exo}$. The only other known pyrimidine nucleoside which is $C(3')_{exo}$ is thymidine.[18] The only purine nucleoside in the $C(3')_{exo}$ conformation is deoxyadenosine,[19] which was the first case with this conformation. It is noteworthy that all of the three nucleosides with the $C(3')_{exo}$ conformation possess a deoxyribose. This observation suggests that the 3' exo conformation may be energetically more favored for the deoxyribonucleic acids than the ribonucleic acids. It also has for the first time provided information that the absence of the 2'-hydroxy group can give the 2'-deoxyribose a conformation, $C(3')_{exo}$, which is probably more favorable than for the ribose. When referred to the three-atom plane C(1'), O(1'), and C(4') (Table V, plane III), the conformation of 5'-dCMP is ($C(3')_{exo}$–$C(2')_{exo}$) which is different from those of thymidine ($C(3')_{exo}$–$C(2')_{endo}$) and deoxyadenosine ($C(3')_{exo}$–$C(2')_{endo}$). Therefore, 5'-dCMP is the first compound where both C(3') and C(2') are displaced on the same side of the three-atom plane and opposite to C(5'),[20] Figure 3. The torsion angles about the bonds of the sugar ring are compared for the three compounds in Table VI.

The torsion angles involving the ring hydrogen atoms are also given in Table VI. These angles are of con-

Table IV. Bond Distances and Bond Angles Not Involving Hydrogen Atoms[a]

	VR	LS	$\lvert\Delta\rvert/\sigma_{12}$[b]
	Bond Distance, Å		
C(2)–O(2)	1.214 (10)	1.220 (3)	0.6
C(4)–N(4)	1.313 (10)	1.320 (3)	0.7
N(1)–C(2)	1.371 (10)	1.372 (3)	0.1
C(2)–N(3)	1.381 (10)	1.394 (3)	1.2
N(3)–C(4)	1.368 (11)	1.363 (3)	0.4
C(4)–C(5)	1.408 (11)	1.413 (4)	0.4
C(5)–C(6)	1.374 (11)	1.349 (3)	2.1
C(6)–N(1)	1.352 (11)	1.367 (3)	1.3
N(1)–C(1')	1.519 (10)	1.510 (3)	0.9
O(1')–C(1')	1.397 (10)	1.404 (3)	0.7
C(1')–C(2')	1.534 (11)	1.528 (4)	0.5
C(2')–C(3')	1.565 (11)	1.522 (3)	3.8
C(3')–C(4')	1.498 (13)	1.511 (4)	1.0
C(4')–O(1')	1.456 (10)	1.454 (3)	0.2
C(3') O(3')	1.455 (10)	1.443 (3)	1.1
C(4')–C(5')	1.512 (12)	1.505 (4)	0.6
C(5')–O(5')	1.450 (10)	1.449 (3)	0.1
P–O(5')	1.612 (6)	1.597 (2)	2.4
P–O(6)	1.499 (7)	1.508 (2)	1.2
P–O(7)	1.585 (5)	1.572 (2)	2.1
P–O(8)	1.488 (7)	1.490 (2)	0.3
	Bond Angle, Deg		
C(2)–N(1)–C(6)	122.2 (0.6)	120.9 (0.2)	2.1
C(2)–N(1)–C(1')	116.4 (0.6)	116.4 (0.3)	0.0
C(6)–N(1)–C(1')	121.0 (0.7)	122.2 (0.2)	1.6
N(1)–C(2)–N(3)	115.8 (0.6)	116.0 (0.2)	0.3
N(1)–C(2)–O(2)	122.2 (0.6)	122.8 (0.2)	0.9
N(3)–C(2)–O(2)	122.0 (0.5)	121.2 (0.2)	1.5
C(2)–N(3)–C(4)	124.6 (0.5)	123.9 (0.2)	1.3
N(3)–C(4)–C(5)	117.0 (0.6)	117.8 (0.2)	1.3
N(3)–C(4)–N(4)	121.6 (0.6)	120.1 (0.2)	0.8
C(5)–C(4)–N(4)	121.3 (0.6)	122.0 (0.2)	1.1
C(4)–C(5)–C(6)	118.9 (0.6)	118.3 (0.2)	0.9
C(5)–C(6)–N(1)	121.3 (0.6)	122.8 (0.2)	2.4
N(1)–C(1')–C(2')	112.2 (0.6)	113.6 (0.2)	2.2
N(1)–C(1')–O(1')	108.3 (0.5)	107.8 (0.2)	0.9
O(1')–C(1')–C(2')	108.6 (0.5)	108.1 (0.2)	0.9
C(1')–C(2')–C(3')	102.6 (0.6)	103.7 (0.2)	1.7
C(2')–C(3')–C(4')	103.4 (0.6)	103.3 (0.2)	0.2
C(2')–C(3')–O(3')	108.1 (0.6)	108.4 (0.2)	0.5
C(4')–C(3')–O(3')	109.2 (0.6)	110.4 (0.2)	1.9
C(3')–C(4')–O(1')	105.3 (0.6)	104.7 (0.2)	0.9
C(3')–C(4')–C(5')	115.3 (0.6)	116.1 (0.2)	1.3
O(1')–C(4')–C(5')	111.3 (0.6)	111.5 (0.2)	0.3
C(4')–O(1')–C(1')	110.5 (0.5)	109.7 (0.2)	1.5
C(4')–C(5')–O(5')	108.8 (0.6)	109.9 (0.2)	1.7
C(5')–O(5')–P	118.2 (0.5)	118.6 (0.2)	0.7
O(5')–P–O(6)	108.5 (0.3)	109.1 (0.2)	1.7
O(5')–P–O(7)	106.1 (0.3)	105.5 (0.2)	1.7
O(5')–P–O(8)	105.3 (0.3)	105.9 (0.2)	1.7
O(6)–P–O(7)	105.0 (0.3)	105.4 (0.2)	1.1
O(6)–P–O(8)	117.5 (0.4)	117.3 (0.2)	0.4
O(7)–P–O(8)	113.9 (0.3)	113.0 (0.2)	2.5

[a] Standard deviations are given in parentheses; those for bond distances have been multiplied by 10^3. [b] $\lvert\Delta\rvert/\sigma_{12}$ denotes (difference in bond distances or bond angles)/$(\sigma_1^2 + \sigma_2^2)^{1/2}$, where σ_1 and σ_2 are standard deviations from the two investigations.

---

(14) M. Sundaralingam, *Acta Crystallogr.*, **21**, 495 (1966), and unpublished results.

(15) G. H.-Y. Lin, M. Sundaralingam, and S. K. Arora, *J. Amer. Chem. Soc.*, **93**, 1235 (1971).

(16) M. Spencer, *Acta Crystallogr.*, **12**, 59 (1959).

(17) M. Sundaralingam, *J. Amer. Chem. Soc.*, **87**, 599 (1965).

(18) D. W. Young, P. Tollin, and H. R. Wilson, *Acta Crystallogr., Sect. B*, **25**, 1423 (1969).

(19) D. G. Watson, D. J. Sutor, and P. Tollin, *ibid.*, **19**, 111 (1965).

(20) The case where both C(3') and C(2') are displaced on the same side of C(5') is referred to as the $C(3')_{endo}C(2')_{endo}$ conformation, and has been discussed in earlier papers.[14,17]

Table V. Least-Squares Planes through the Base and Sugar[a,b]

Atom	Plane I dev, Å	Atom	Plane II dev, Å	Plane III dev, Å
N(1)	−0.016 (−0.011)*	C(1′)	−0.047 (−0.041)*	0.000 (0.000)*
C(2)	0.030 (0.023)*	C(2′)	0.027 (0.024)*	0.222 (0.199)
N(3)	−0.028 (−0.020)*	C(3′)	0.476 (0.462)	0.647 (0.615)
C(4)	0.012 (0.006)*	C(4′)	−0.029 (−0.025)*	0.000 (0.000)*
C(5)	0.002 (0.006)*	O(1′)	0.048 (0.043)*	0.000 (0.000)*
C(6)	0.001 (−0.003)*	C(5′)	−1.407 (−1.409)	−1.379 (−1.385)
O(2)	0.103 (0.097)	N(1)	−1.333 (−1.336)	−1.311 (−1.315)
N(4)	0.069 (0.072)			
C(1′)	0.103 (0.119)			
Rms Δ	0.019 (0.015)		0.039 (0.034)	0.000 (0.000)
σ(rms Δ)	0.002 (0.007)		0.002 (0.008)	0.003 (0.008)

[a] Atoms used in the calculation of plane are indicated by asterisks. [b] Data from VR are given in parentheses.

Table VI. Torsion Angles in Deoxyribose of 5′-dCMP Compared with Those of Thymidine and Deoxyadenosine Monohydrate

		5′-dCMP, deg LS	VR	Thymidine, deg	Deoxyadenosine monohydrate, deg
		(a) Involving Ring Atoms			
τ_0	C(4′)-O(1′)-C(1′)-C(2′)	8.8	7.9	−7	−2
τ_1	O(1′)-C(1′)-C(2′)-C(3′)	12.0	11.7	28	24
τ_2	C(1′)-C(2′)-C(3′)-C(4′)	−26.9	−25.9	−37	−34
τ_3	C(2′)-C(3′)-C(4′)-O(1′)	32.6	31.1	33	34
τ_4	C(3′)-C(4′)-O(1′)-C(1′)	−26.3	−25.2	−17	−20
		(b) Involving Ring Hydrogen Atoms			
	H(1′)-H(2′)	18	11	22	28
	H(1′)-H′(2′)	138	134	154	143
	H(2′)-H(3′)	91	89	76	82
	H′(2′)-H(3′)	−45	−41	−48	−38
	H(3′)-H(4′)	−89	−87	−88	−91
	Conformation	$C(3')_{exo}$–$C(2')_{exo}$		$C(3')_{exo}$–$C(2')_{endo}$	$C(3')_{exo}$–$C(2')_{endo}$
		(c) Involving Exocyclic Bond C(4′)-C(5′)			
	H(4′)-H(5′)	−59	−63	56	45
	H(4′)-H′(5′)	69	62	174	173
	Conformation	Gauche–gauche		Gauche–trans	Gauche–trans

siderable importance in deriving the conformations of ribose and deoxyribose through their relation with the coupling constants obtained from nuclear magnetic resonance studies. The H–C–O–H torsion angle is H(3′)-H(O3′), 87° (91°), and has the usual gauche angle.[21]

Both the $C(3')_{exo}$–$C(2')_{exo}$ conformation[17,19,22] for the furanoside ring and the small negative glycosyl torsion angle may be of consequence in connection with defining the detailed conformation of DNA and other polynucleotides.

A detailed analysis of the torsion angles in the B-DNA model[23] has shown that the molecule exhibits torsional strains,[13] for several of the torsion angles are outside the preferred range.[13] Furthermore, many of the bond distances and bond angles show deviations of about 0.05 Å and 5°, respectively, from standard values.[12b] The deoxyribose in this model has an improbable "$C(2')_{endo}$" conformation with C(2′) and C(3′) only 0.19 and −0.10 Å, respectively, out of the C(1′)-O(1′)-C(4′) plane (compare these values with those cited in ref 17). Much of the torsional strain in this manually built model does not persist in the recent linked-atom least-squares model.[24] In this new model for B-DNA the deoxyribose has the standard $C(2')_{endo}$–$C(3')_{exo}$ conformation. Since alternative conformations, in particular $C(3')_{exo}$–$C(2')_{endo}$ and $C(3')_{exo}$–$C(2')_{exo}$ have also been found among the known deoxyribonucleosides and deoxyribonucleotides, we suggest that these novel conformations should be considered in future model-building studies, DNA in particular, and

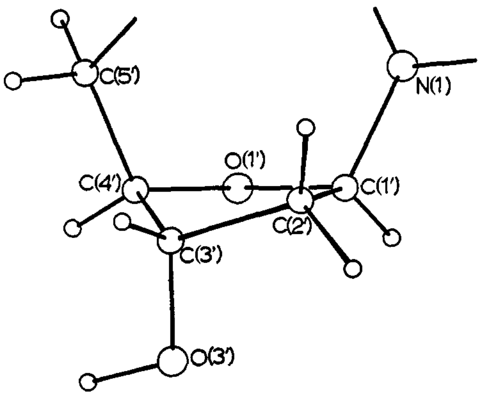


Figure 3. A view of the deoxyribose showing the puckering of C(2′) and C(3′) with respect to the C(1′)-O(1′)-C(4′) plane.

---

(21) M. Sundaralingam, *Biopolymers*, **6**, 189 (1968).

(22) H. R. Wilson, A. Rahman, and P. Tollin, *J. Mol. Biol.*, **46**, 585 (1969).

(23) R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins, and L. D. Hamilton, *ibid.*, **2**, 2 (1960).

(24) S. Arnott, S. D. Dover, and A. J. Wonacott, *Acta Crystallogr., Sect. B.*, **25**, 2192 (1969).

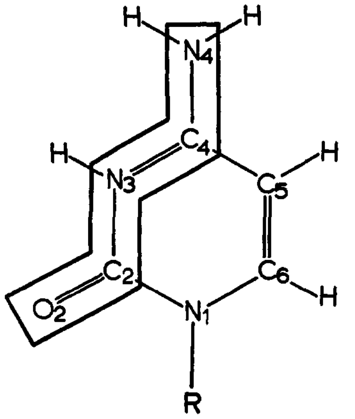


Figure 4. The cytosine cation. The major differences between the cytosine cation and the neutral base are confined to the site of protonation, *i.e.*, the bond distances and bond angles indicated within the solid lines. R represents a ribose, a deoxyribose, or a methyl group.

nucleic acids in general. In fact an analysis of the furanoside conformations has shown that the conformational transition $C(2')_{endo} \rightleftarrows C(3')_{exo}$ or *vice versa* is energetically not unfavorable. While there are several points of similarity between these conformations, there are also noticeable differences.[17]

The conformation about the C(4')–C(5') bond is gauche–gauche and the angles O(1')–C(4')–C(5')–O(5') and C(3')–C(4')–C(5')–O(5') are −63.3° (−62.6°) and 56.5° (57.1°), respectively. However, in thymidine and deoxyadenosine which are the two other nucleosides in the $C(3')_{exo}$ conformation, the conformation is gauche–gauche about the C(4')–C(5') bond. The gauche–gauche conformation is the only conformation that has been observed for the known 5'-nucleotides. It is also the conformation[13,25] found in the refined models of the proposed structures of the double-helical nucleic acids and polynucleotides.[26] For instance, the torsion angle C(3')–C(4')–C(5')–O(5') has been determined to be 50, 59, 49, 50, 50, 33, and 56°, respectively, in A'-RNA, poly(rI·rC), β-A-RNA (rA–rU), β-A-RNA (reovirus), α-A-RNA (reovirus), A-DNA, B-DNA, and poly(rI·dC).[25]

The bond lengths and bond angles in the deoxyribose moiety, excepting the C(3')–O(3') bond length, are generally in good agreement with similar structures.[12b,17] The C(3')–O(3') bond of 1.443 Å (1.455 Å) is significantly longer than those found in other structures. The increase in length of this bond is attributed to the conformation of the sugar. In the $C(3')_{exo}$ conformation, the C(3')–O(3') bond is in the axial orientation and the nonbonded interactions between O(3') and the adjacent atoms of the deoxyribose ring are increased in comparison with structures in the alternative conformations.

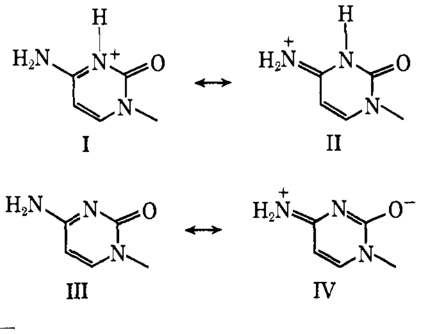
The Phosphate. The phosphate group is in the characteristic staggered conformation about the C(5')–O(5') bond; the torsion angles are C(5')–O(5')–P–O(6), −42.9° (−44.0°), C(5')–O(5')–P–O(7), 69.9° (68.3°), and C(5')–O(5)–P–O(8), −170.0° (−170.6°). The ester linkage, C(5')–O(5'), shows the elongated conformation as usual, the torsion angle C(4')–C(5')–O(5')–P being 165.8° (166.4°).

The bond distances and bond angles in the phosphate group agree well with those of similar monoanionic phosphates.[28]

The Cytosine Cation. The cytosine base was found to be a cation; N(3) was the site of protonation.[29] The protonation results from the migration of the proton from the phosphate oxygen O(8). It has already been established by Sundaralingam and Jensen[12b] in an earlier crystal-structure analysis of cytidine 3'-phosphate (3'-CMP(O)) that the cytosine nucleotide prefers the zwitterion structure. These authors also observed that significant changes in the bond distances and bond angles occurred between the cytosine cation and the neutral base. The major differences were generally confined to the site of protonation, and the adjacent bonds and angles as indicated within the solid lines in Figure 4.

A comparison of bond distances and bond angles in protonated and unprotonated cytosine structures is given in Table VII. The average bond distances and bond angles of the neutral cytosine bases, the N(1)-substituted neutral cytosine bases, and the N(1)-substituted protonated cytosine bases are given in Figure 5.

The bond lengths of C(2)–O(2), 1.220 Å (1.214 Å), and C(2)–N(3), 1.394 Å (1.381 Å), are in good agreement with those of the cytosine cation in 3'-CMP[12b,30] and deoxycytidine hydrochloride,[31] while the C(2)–O(2) bond is significantly shorter and the C(2)–N(3) bond is significantly longer than those observed in the neutral cytosine derivatives, 1-methylcytosine,[32] 1-methylcytosine in complex with 9-ethylguanine,[33] and cytidine.[34] For comparison, the values for the cytosine bases, cytosine monohydrate[35] and cytosine[36] are also given in Table VII. These differences can be attributable in part to the steric effect of the proton at N(3), and in part to the contributions from the resonance structures. Although many resonance structures can



---

(25) S. Arnott, *Progr. Biophys. Mol. Biol.*, **21**, 267 (1970).

(26) It may be pointed out that in the Crick and Watson[27] model of B-deoxyribonucleic acid the conformation about C(4')–C(5') was gauche–trans. However, in the refined model[23] the conformation was gauche–gauche.[13]

(27) F. H. C. Crick and J. D. Watson, *Proc. Roy. Soc., Ser. A*, **223**, 80 (1954).

(28) M. Sundaralingam and E. Putkey, *Acta Crystallogr., Sect. B*, **26**, 790 (1970).

(29) Similarly, the crystal-structure analysis of adenosine 3'-phosphate[14] has demonstrated that N(1) was the site of protonation. In 5'-dCMP the bond distance and bond angles involving H(3) are: N(3)–H(3), 1.00 Å (0.96 Å); C(2)–N(3)–H(3), 115° (114°); and C(4)–N(3)–H(3), 121° (122°).

(30) C. E. Bugg and R. E. Marsh, *J. Mol. Biol.*, **25**, 67 (1967).

(31) E. Subramanian and D. J. Hunt, *Acta Crystallogr., Sect. B*, **26**, 303 (1970).

(32) F. S. Mathews and A. Rich, *Nature (London)*, 201, 179 (1964).

(33) E. J. O'Brien, *Acta Crystallogr.*, **23**, 92 (1967).

(34) S. Furberg, C. S. Petersen, and C. Romming, *ibid.*, **18**, 313 (1965).

(35) G. A. Jeffrey and Y. Kinoshita, *ibid.*, **16**, 20 (1963).

(36) D. L. Barker and R. E. Marsh, *ibid.*, **17**, 1581 (1964).

Table VII. Comparison of Bond Distances (Ångströms) and Bond Angles (Degrees) in Protonated and Unprotonated Cytosine Structures

	Unprotonated					Protonated			
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*
				Distances					
C(2)–O(2)	1.234	1.260	1.240	1.233	1.246	1.214	1.201	1.207	1.220 (1.214)
C(4)–N(4)	1.330	1.332	1.346	1.337	1.333	1.308	1.323	1.315	1.320 (1.313)
N(1)–C(2)	1.374	1.376	1.400	1.401	1.379	1.382	1.401	1.389	1.372 (1.371)
C(2)–N(3)	1.364	1.354	1.359	1.349	1.361	1.386	1.382	1.403	1.394 (1.381)
N(3)–C(4)	1.337	1.351	1.338	1.354	1.335	1.357	1.339	1.350	1.363 (1.368)
C(4)–C(5)	1.424	1.432	1.436	1.419	1.420	1.412	1.420	1.403	1.413 (1.408)
C(5)–C(6)	1.342	1.348	1.343	1.345	1.345	1.338	1.351	1.335	1.349 (1.374)
C(6)–N(1)	1.357	1.361	1.373	1.362	1.368	1.369	1.350	1.361	1.367 (1.352)
σ (*l*)	0.003	0.004	0.005	0.007	0.006	0.003	0.007	0.009	0.003 (0.010)
				Angles					
N(1)–C(2)–N(3)	118.1	120.1	118.7	119.0	119.7	114.9	113.4	114.0	116.0 (115.8)
C(2)–N(3)–C(4)	119.9	118.9	120.8	120.5	119.4	125.2	126.0	124.5	123.9 (124.6)
N(3)–C(4)–C(5)	122.0	122.0	122.1	121.4	121.6	116.9	118.3	118.5	117.8 (117.0)
C(4)–C(5)–C(6)	117.3	117.1	114.7	117.2	118.2	119.3	117.2	118.4	118.3 (118.9)
C(5)–C(6)–N(1)	120.1	120.6	122.9	121.8	12011	122.1	122.7	122.6	122.8 (121.3)
C(6)–N(1)–C(2)	122.7	121.3	119.7	119.9	120.9	121.3	122.3	122.0	120.9 (122.2)
N(3)–C(4)–N(4)	118.2	117.8	118.8	118.4	116.8	119.9	119.9	120.6	120.1 (121.6)
C(5)–C(4)–N(4)	119.9	120.2	119.2	120.3	121.6	123.2	121.8	120.9	122.0 (121.3)
N(1)–C(2)–O(2)	119.8	118.4	118.1	117.5	119.2	123.5	124.5	123.9	122.8 (122.2)
N(3)–C(2)–O(2)	112.2	121.5	123.3	123.4	121.2	121.5	122.1	122.9	121.2 (122.0)
C(1′)–N(1)–C(2)			118.2	118.7	117.2	116.7	117.1	118.2	116.4 (116.4)
C(1′)–N(1)–C(6)			122.1	121.4	122.0	122.1	120.4	119.6	122.2 (121.0)
$\sigma(\theta)$	0.2	0.4	0.4	0.4	0.4	0.2	0.6	1.0	0.2 (0.6)

[a] Cytosine.[36] [b] Cytosine monohydrate.[35] [c] 1-Methylcytosine.[32] [d] 1-Methylcytosine (in complex with 9-ethylguanine).[33] [e] Cytidine.[34] [f] Deoxycytidine hydrochloride.[31] [g] 3′-CMP(O).[12b] [h] 3′CMP(M).[30] [i] 5′CMP; this work (values of VR in parentheses).

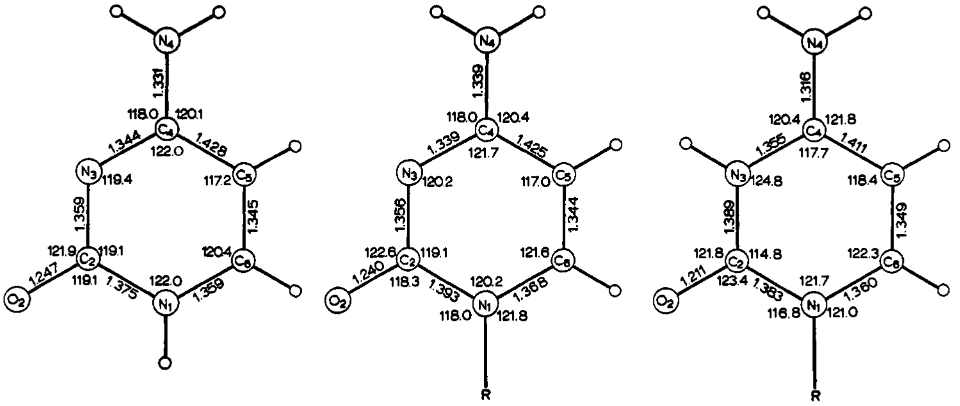


Figure 5. Average bond distances and bond angles of (left) the neutral cytosine bases (*a* and *b* in Table VII), (center) the N(1)-substituted neutral cytosine bases (*c*–*e* in Table VII), and (right) the N(1)-substituted protonated cytosine bases (*f*–*i* in Table VII).

be written for the cytosine cation[12b] and cytosine,[36] the observed bond distances strongly suggest that the contribution of structure II to the overall resonance of the cytosine cation is much greater than the contribution of structure IV to the overall resonance of the neutral cytosine.

The carbonyl C(2)–O(2) bond in the present structure is longer than that observed in 3′-CMP. This can be explained from the fact that O(2) is hydrogen bonded to the water molecule in the present structure while O(2) in both the orthorhombic and monoclinic crystal forms of 3′-CMP and deoxycytidine hydrochloride is not hydrogen bonded. Therefore, 5′-dCMP is the only protonated cytosine derivative that displays hydrogen bonding at O(2); in contrast, all of the known neutral derivatives of cytosine are hydrogen bonded at O(2), Figure 6.

It should be pointed out that differences in the bond distances and bond angles among the neutral cytosine bases or the protonated cytosine bases may perhaps be partly attributable to (in addition to errors in the structure analysis and thermal vibrations) the differences in the type, number, and strengths of the hydrogen bonds to the bases in the crystal. Similar differences in bond distances and bond angles are seen in the adenine rings of the many known crystal structures of adenine derivatives.[37] Craven[38] in his exhaustive studies on the crystal structures of the barbiturates first noted the effect of hydrogen bonding on the bond distances and bond angles of the pyrimidine ring.

The greatest changes in bond angles arising from the protonation at N(3) occur at N(3) and at the adjacent positions, C(2) and C(4). The angle C(2)–N(3)–C(4), average 125°, in the protonated derivatives is about 5° greater than that (120.5) in the neutral cytosine deriv-

---

(37) S. T. Rao and M. Sundaralingam, *J. Amer. Chem. Soc.*, **92**, 4963 (1970).

(38) B. Craven, *Acta Crystallogr., Sect. A*, **25**, 4138 (1969).

Table VIII. Hydrogen Bonds and Angles

Atom A	Atom B	Symmetry code[a]	Distance, Å	Bond	Angle, deg
O(2)	O(W)	I	2.946 (2.939)	C(2)–O(2)···O(W)	132.7 (132.2)
O(2)	H(W)	I	2.34 (2.13)	C(2)–O(2)···H(W)	129 (122)
				O(W)–H(W)···O(2)	150 (146)
O(3′)	O(7)	II	2.609 (2.603)	C(3′)–O(3′)···O(7)	120.6 (120.6)
				P–O(7)···O(3)	123.4 (122.1)
O(3′)	H(O-7)	II	1.67 (1.72)	C(3′)–O(3′)···H(7)	117 (114)
				O(7)–H(7)···O(3′)	167 (157)
O(6)	O(W)	III	2.713 (2.719)	P–O(6)···O(W)	120.2 (119.9)
O(6)	H′(W)	III	1.81 (2.04)	P–O(6)···H′(W)	124 (122)
				O(W)–H′(W)···O(6)	165 (154)
O(8)	N(3)	IV	2.874 (2.877)	C(2)–N(3)···O(8)	108.8 (109.0)
				C(4)–N(3)···O(8)	127.2 (126.4)
				P–O(8)···N(3)	109.6 (109.5)
O(8)	H(3)	IV	1.88 (1.92)	P–O(8)···H(3)	106 (107)
				N(3)–H(3)···O(8)	171 (173)
O(6)	N(4)	IV	2.691 (2.709)	C(4)–N(4)···O(6)	116.9 (115.7)
				P–O(6)···N(4)	127.2 (127.4)
O(6)	H(N-4)	IV	1.78 (1.79)	P–O(6)···H(N4)	127 (129)
				N(4)–H(N-4)···O(6)	173 (171)
O(8)	N(4)	V	2.981 (2.941)	C(4)–N(4)···O(8)	116.4 (117.6)
				P–O(8)···N(4)	119.9 (120.5)
O(8)	H′(N-4)	V	1.94 (1.94)	N(4)–H′(N-4)···O(8)	172 (164)
				P–O(8)···H′(N-4)	119 (119)
O(W)	O(3′)	VI	2.691 (2.668)	C(3′)–O(3′)···O(W)	115.1 (115.6)
O(W)	H(O-3′)	VI	1.61 (1.70)	O(3′)–H(O3′)···O(W)	164 (152)

[a] Symmetry codes: I, $\frac{1}{2} + x, -\frac{1}{2} - y, -z$; II, $1 - x, -\frac{1}{2} + y, \frac{1}{2} - z$; III, $-x, \frac{1}{2} + y, \frac{1}{2} - z$; IV, $-\frac{1}{2} + x, \frac{1}{2} - y, -z$; V, $\frac{1}{2} + x, \frac{1}{2} - y, -z$; VI, x, y, z.

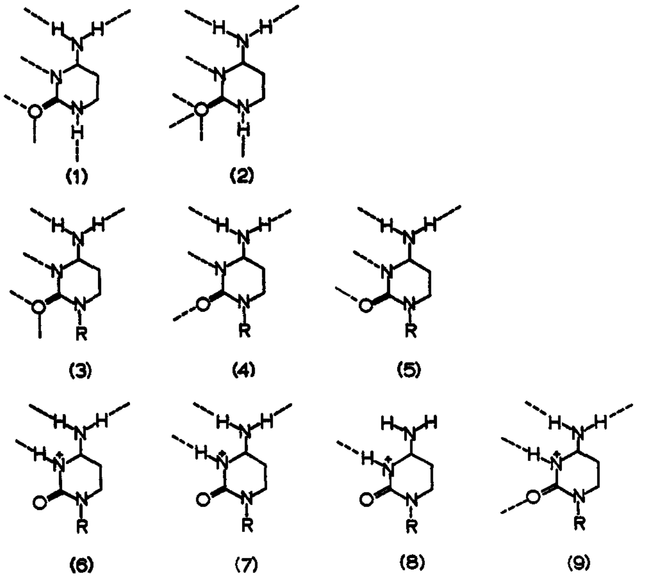


Figure 6. Hydrogen bonds of protonated and unprotonated cytosine structures. Numbers (1)–(9) correspond to footnotes *a–i* given in Table VII.

atives.[12b] Similar changes in valence angle at nitrogen have been observed in a large number of protonated and unprotonated nitrogen heterocycles.[39] On the other hand, the angles N(1)–C(2)–N(3) and N(3)–C(4)–C(5) are about 4–5° smaller in the protonated than in the unprotonated derivatives. A further point of interest is that the average exocyclic angle N(1)–C(2)–O(2), 123.4°, in the protonated cytosine, regardless of whether O(2) is hydrogen bonded or not, is about 5.5° greater than in the unprotonated species. The remaining angles of the neutral and protonated cytosine moieties show differences of less than 2°.

In general the bases themselves (cytosine and cytosine monohydrate) and the unprotonated derivatives show the same differences from the protonated derivatives.

Least-Squares Plane of Cytosine Cation. The least-squares plane through the cytosine ring atoms is given in Table V (plane I). The pyrimidine ring is not strictly planar; the deviations of C(2) and N(3) from the plane are significant and are in the opposite directions. The distortion in the plane of the cytosine ring is attributable to the hydrogen bonding pattern (see below). Similar deviations in the base plane have been observed in cytosine monohydrate,[35] 3′-CMP(O),[12b] deoxycytidine hydrochloride,[31] and cytidine,[34] while in cytosine[36] and 3′-CMP(M)[30] the base rings are planar.

Molecular Packing and Hydrogen Bonding. All available hydrogen atoms participate in hydrogen bonding. The hydrogen bond distances and angles are given in Table VIII. An interesting aspect of the hydrogen bond scheme is that the molecules related by the screw axis parallel to the *a* axis are linked together through hydrogen bonds involving the amino hydrogen atoms, the base hydrogen atom, H(N$^+$-3), and the phosphate group, resulting in an infinitely extended right-handed spiral around the screw axis (Figures 7 and 8). There is no base stacking in this crystal since adjacent bases related by the 2_1 axes parallel to *a* are sufficiently far apart. This is consistent with the lack of base stacking in other protonated cytosine derivatives.[40,41] The crystal structure can be considered to be a close packing of the infinite spirals linked by hydrogen bonds. The voids between the spirals are filled by the crystal water which is linked through three hydrogen bonds to

---

(39) C. Singh, *Acta Crystallogr.*, **19**, 861 (1965).

(40) M. Sundaralingam, S. T. Rao, and C. E. Bugg, Abstracts, American Crystallographic Association Meeting, Seattle, Wash., March 1969.

(41) C. E. Bugg, J. Thomas, M. Sundaralingam, and S. T. Rao, *Biopolymers*, **10**, 175 (1971).

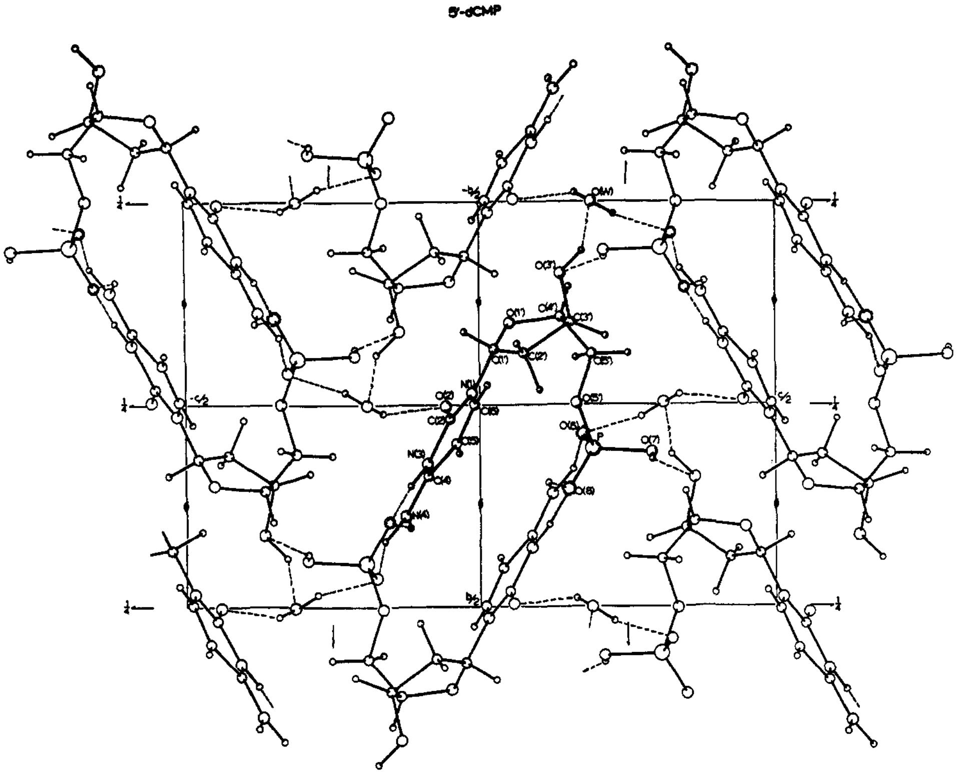


Figure 7. Intermolecular hydrogen bonding scheme viewed along *a* axis. The molecules related by the screw axis parallel to *a* are linked together through hydrogen bonds, resulting in an infinitely extended right-handed spiral around the screw axis.

the phosphate, deoxyribose, and base of three different neighboring nucleotide molecules forming a trigonal pyramid. The hydrogen bond angles at the water molecule are: O(3′)···O(W)···O(2), 96.0° (96.5°), O(3′)···O(W)···O(6), 108.0° (108.3°), and O(2)···O(W)···O(6), 107.6° (108.0°).

The involvement of the carbonyl oxygen O(2) in the hydrogen bonding to water results in a displacement of the atoms O(2) and C(2) from the plane of the base and toward the water. Hence, the mode of puckering of the base in 5′-dCMP and the other nucleic acid constituents is mainly influenced by the hydrogen bonding pattern[15,42] and the molecular packing.

The amino group donates its protons to the phosphate oxygen atoms O(6) and O(8) of two different molecules related by translation along the *a* axis. The only hydrogen on the phosphate group, H(O-7), is donated to O(3′), and it is the shortest hydrogen bond in the structure. The second short hydrogen bond also involves O(3′), and the molecule of water.

The phosphate–base pairing scheme, O(8)···N(1) and O(6)···N(4), is extremely characteristic of the cytidine and adenosine nucleotides[14] and may be important in nucleic acid–nucleic acid interactions.

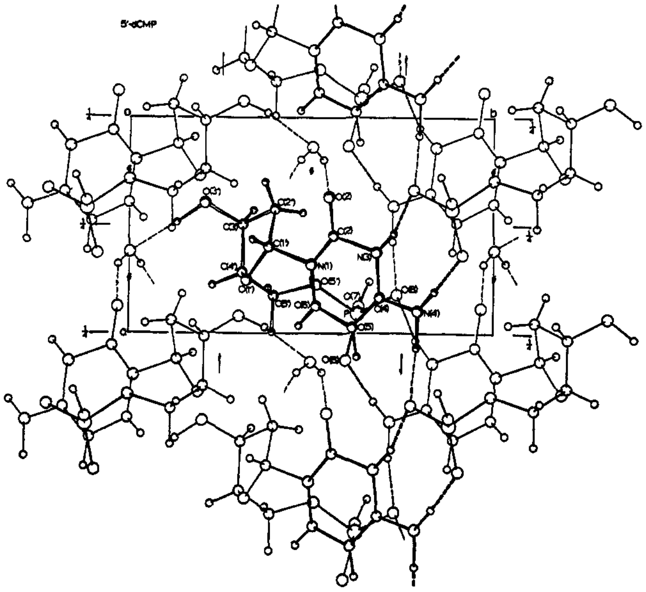


Figure 8. Projection along the *c* axis showing the intermolecular hydrogen bonding scheme.

Acknowledgment. The authors (VR) thank Professor R. S. Krishnan for his keen interest in this work and Dr. Venkatesan and Dr. R. Chidambaram for making available many of the programs used in the computation. The authors (LS) thank the National Institutes of Health of the United States Public Health Service for research support through Grant No. GM 17378 and the University Research Committee for the award of computer time for the use of the University computer. They also thank Dr. S. T. Rao for the crystallographic computer programs.

---

(42) M. Sundaralingam and S. K. Arora, *Proc. Nat. Acad. Sci. U. S.*, 64, 1021 (1969).