Crystal Structure and Absolute Configuration of 1-β-D-Arabinofuranosylcytosine 5'-Monophosphate (ara-CMP) Trihydrate

By J. S. Sherfinski and R. E. Marsh

Department of Chemistry,* California Institute of Technology, Pasadena, California 91125, USA

and A. K. Chwang and M. Sundaralingam

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

Abstract

Two independent crystal structure analyses (UW and CIT) of the nucleotide 1-β-D-arabinofuranosylcytosine 5'-monophosphate (ara-CMP) trihydrate are reported. The compound crystallizes from water in the space group $P2_12_12_1$ with average unit-cell dimensions $a =$ 17·51 (2), $b =$ 18·24 (2), $c =$ 4·814 (2) Å. The final R factors after least-squares refinement were 0·061 for 945 reflections in the UW study and 0·046 for 1796 reflections in the CIT study. The average e.s.d.'s in atomic coordinates are about 0·009 Å (UW) and 0·004 Å (CIT). The bond distances and angles from the two studies agree reasonably well. The nucleotide displays the *anti* conformation for the base ($\chi = 30\cdot7°$), the *gauche*+ ($\psi = 53\cdot3°$) conformation about the exo-cyclic C(4')—C(5') bond and the classical twist conformation $^3T_2$ [C(3')-*endo*, C(2')-*exo*] for the arabino-furanose ring. These features are one of the favored conformational combinations for the ribonucleotides as well. The base sites N(3) and N(4) are engaged in a pair of hydrogen bonds to the anionic phosphate oxygens of an adjacent nucleotide. The three water molecules are arranged near a 2_1 screw axis to fill a rather large channel parallel to the c axis and are linked by hydrogen bonds to the sugar, base, and phosphate moieties of the surrounding nucleotides.

Introduction

As part of a program of research in our two laboratories (UW and CIT) on studies of compounds related to the nucleic acid constituents and their analogs, we have determined the crystal structure of the first nucleotide containing the arabinose sugar, *viz* ara-CMP.3H$_2$O (1-β-D-arabinofuranosylcytosine 5'-monophosphate trihydrate). The crystal structures of

the corresponding nucleoside ara-C (arabinosyl-cytosine) (Chwang & Sundaralingam, 1973; Lefebvre-Soubeyran & Tougard, 1973; Tougard & Lefebvre-Soubeyran, 1974) and of its hydrochloride salt (Sherfinski & Marsh, 1973) have already been reported.

Experimental – UW

Colorless crystals of ara-CMP (Terra-Marine Bioresearch) were obtained by slow evaporation from a water-enriched 2-propanol solution. Intensity data were collected on a Picker FACS-1 diffractometer. Crystal data and details of data collection are given in Table 1.

The structure was solved by direct methods using the program *MULTAN* (Main, Germain & Woolfson, 1970) and refined by least-squares techniques. The final

Table 1. *Crystal data for ara-CMP trihydrate,* $C_9H_{14}N_3O_8P.3H_2O$

Orthorhombic		Space group $P2_12_12_1$
λ (Cu $K\alpha$) = 1·5418 Å		$Z = 4$
$M_r = 377\cdot2$		$F(000) = 792$
	UW	CIT
a	17·498 (4) Å	17·531 (3) Å
b	18·225 (5)	18·252 (4)
c	4·812 (1)	4·816 (1)
V	1535 Å³	1541 Å³
d_c	1·633 Mg m^{-3}	1·626 Mg m^{-3}
d_o	1·63	1·62
Radiation	Ni-filtered Cu $K\alpha$	
Scan mode	Step scan	θ–2θ scan
Scan range	1·4°	1·91° + 0·00822 $(2\theta)°$
Scan rate	2° min^{-1}	1° min^{-1}
2θ range	5–127°	5–150°
Reflections scanned	1447	1859
Reflections retained	945	1859
Crystal dimensions	0·07 × 0·08 × 0·12 mm	0·14 × 0·13 × 0·05 mm

R index was 0·061 for the 945 observed reflections. Only 14 of the 20 H atoms were located in difference density maps; six H atoms [four belonging to water molecules, H'(9), H(10), H(11), H'(11), and the hydroxyl HO(2') and HO(3')] were not located. In the final cycles of anisotropic refinement of the non-hydrogen atoms, the H-atom parameters were kept fixed.

The quantity minimized during the refinement was $\sum w(|F_o| - k|F_c|)^2$. The weighting scheme used in the refinement was $1/\sqrt{w} = \sigma(F) = 2\cdot075 - 0\cdot333F_o$ for $|F_o| \leq 4\cdot9$; and $1/\sqrt{w} = 0\cdot45 + 0\cdot0025F_o$ for $F_o > 4\cdot9$. The scattering factors for P, O, N, C were from Cromer & Waber (1965), and those for H were from Stewart, Davidson & Simpson (1965).

Experimental – CIT

Colorless crystals were obtained from a water solution of ara-CMP (Terra-Marine Bioresearch). Two sets of data were collected, one utilizing graphite-mono-chromatized Mo radiation and used in the initial structural solution and the second utilizing Ni-filtered Cu radiation and used for the final refinement. The details are summarized in Table 1.

The structure was solved using information from both a three-dimensional Patterson function and direct methods. Least-squares refinement with anisotropic temperature factors for the non-hydrogen atoms and isotropic temperature factors for the H atoms led to an R index of 0·047. Difference maps calculated at this point failed to give a clear indication of the hydrogen-bond arrangement between O(10) and O'(10) (at $\frac{1}{2} - x$, $1 - y$, $\frac{1}{2} + z$); accordingly, the H atom involved was presumed to be twofold disordered, and two half-H atoms were placed on the O(10)—O'(10) vector $\frac{1}{3}$ and $\frac{2}{3}$ the distance from O(10) to O'(10). These partial H atoms and the other H atom attached to O(10) were not refined, but were repositioned periodically during the course of the refinement and included in all structure-factor calculations.

The quantity minimized during the refinement was $\sum [(|F_o|^2 - |F_c|^2)^2/\sigma^2(F_o^2)]$, where the variances $\sigma^2(F_o^2)$ were derived from counting statistics plus a term $(0\cdot02S)^2$, where S is the scan count. Form factors for neutral atoms (P, C, N, O: *International Tables for X-ray Crystallography*, 1962; H: Stewart, Davidson & Simpson, 1965) were corrected for anomalous dispersion (Hope, 1969; for O, $f' = 0\cdot047$ and $f'' = 0\cdot032$; for N, $f' = 0\cdot032$, $f'' = 0\cdot019$ e) near the end of the refinement, and structure factors for both enantiomers were calculated; the D configuration was significantly favored at the $\alpha = 0\cdot005$ level (Hamilton, 1965). Final least-squares refinement, with the heavy-atom coordinates in one matrix and their anisotropic temperature parameters, the coordinates and isotropic temperature factors of the H atoms, and the scale factor in a second matrix led to an R index of 0·046 and a goodness-of-fit of 1·69.

Table 2. *Positional parameters of the nonhydrogen atoms and their standard deviations*

The x and y values have been multiplied by 10^5 and z by 10^4.

	x		y		z	
	UW	CIT	UW	CIT	UW	CIT
N(1)	3237 (44)	3215 (16)	34528 (46)	34607 (16)	7981 (16)	7937 (6)
C(2)	10301 (69)	10348 (21)	31709 (64)	31695 (20)	7360 (24)	7396 (9)
O(2)	16221 (41)	16136 (14)	33755 (47)	33744 (17)	8483 (20)	8491 (7)
N(3)	10481 (48)	10357 (17)	26082 (49)	26174 (17)	5418 (19)	5404 (7)
C(4)	4309 (63)	4133 (20)	23586 (54)	23550 (17)	4155 (25)	4093 (8)
N(4)	4838 (51)	4885 (17)	18562 (50)	18621 (17)	2180 (19)	2137 (8)
C(5)	−2880 (66)	−3005 (20)	26555 (64)	26549 (20)	4898 (28)	4832 (10)
C(6)	−3076 (65)	−3145 (20)	31819 (57)	31970 (19)	6792 (23)	6711 (9)
P	−24436 (15)	−24472 (5)	34548 (15)	34556 (5)	8854 (6)	8810 (2)
O(5')	−18621 (37)	−18626 (14)	41011 (37)	41016 (14)	8160 (12)	8144 (5)
O(6)	−25007 (45)	−24964 (15)	30109 (35)	30119 (13)	6265 (16)	6209 (6)
O(7)	−31775 (39)	−31694 (14)	37604 (42)	37620 (15)	9972 (17)	9887 (6)
O(8)	−20264 (41)	−20247 (15)	29790 (38)	29873 (14)	11098 (17)	11055 (6)
C(1')	2996 (60)	2848 (20)	41059 (58)	41043 (19)	9840 (23)	9821 (8)
O(1')	−4177 (39)	−4260 (14)	41105 (35)	41129 (12)	11242 (14)	11156 (6)
C(2')	3703 (62)	3560 (21)	48174 (55)	48357 (19)	8261 (20)	8247 (8)
O(2')	7819 (38)	7674 (15)	47868 (42)	47845 (15)	5767 (16)	5736 (6)
C(3')	−4710 (57)	−4784 (21)	50271 (53)	50304 (19)	7783 (20)	7763 (8)
O(3')	−5521 (39)	−5522 (14)	57823 (39)	57879 (14)	7353 (16)	7327 (7)
C(4')	−8172 (58)	−8460 (21)	47759 (58)	47706 (19)	10429 (22)	10418 (7)
C(5')	−16778 (61)	−16831 (21)	45979 (61)	46096 (20)	10429 (24)	10365 (8)
O(9)	31526 (51)	31673 (16)	34269 (52)	34259 (16)	3920 (19)	4029 (7)
O(10)	23092 (57)	23258 (18)	46705 (54)	46555 (18)	5034 (24)	5033 (8)
O(11)	39420 (48)	39610 (16)	37322 (44)	37423 (15)	9078 (22)	9120 (7)

Results

Coordinates of the heavy atoms obtained from both refinements are given in Table 2.* The estimated standard deviations in bond distances and angles involving the heavy atoms are approximately 0·014 Å and 0·7° respectively in the UW work while they are 0·005 Å and 0·3° respectively in the CIT work. Since the two results show no disagreements which are significant in terms of the combined r.m.s. standard deviations, and the CIT analysis is twice as precise as the UW, only the CIT results are discussed hereafter.

Molecular dimensions and conformation

Ara-CMP exists as a zwitterion with the base nitrogen atom N(3) protonated. The dimensions (Fig. 1) of this protonated cytosine base are in good agreement with those found in other structural studies (Voet & Rich, 1970). The dimensions of the arabinose sugar agree quite well with the average values found in other arabinonucleotides and nucleosides having the C(3')-endo puckering (Sundaralingam, 1975).

The molecular conformation of ara-CMP is shown in Fig. 1. The disposition of the base is anti (Donohue & Trueblood, 1960; Sundaralingam, 1969) and the glycosyl torsion angle χ is 30·7°. The arabinose ring exhibits the twist pucker $^3T_2$ [C(3')-endo, C(2')-exo] with a phase angle of pseudorotation P of 12·6° and a maximum amplitude of puckering τ_m of 38·7° (Altona & Sundaralingam, 1972). The conformation around the exocyclic C(4')—C(5') bond is gauche⁺ ($\psi = 53\cdot3°$),

* Lists of structure factors and anisotropic thermal parameters from both refinements and the atomic parameters of the H atoms from the CIT refinement have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 34450 (18 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.
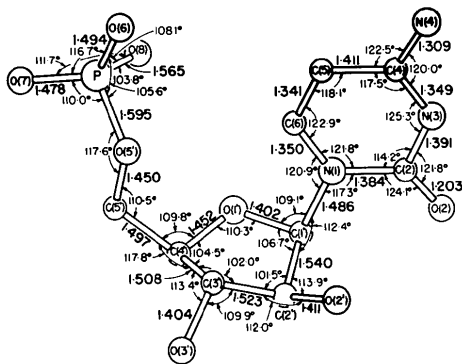

Fig. 1. Bond distances (Å) and angles (°) involving the non-hydrogen atoms in ara-CMP (CIT).

and the C(5')—O(5') torsion angle is 137·2° which is quite considerably distorted from the ideal trans. The overall conformation of ara-CMP falls within one of the most preferred conformational states for the arabinose nucleotides and the nucleotide building blocks of nucleic acids. A listing of the torsion angles in ara-CMP is given in Table 3.

Molecular packing and hydrogen bonding

A view of the crystal structure of ara-CMP is shown in Fig. 2, and the hydrogen-bond distances and angles are

Table 3. *Torsion angles* (°) *in ara-CMP trihydrate* (*CIT values*)

In ara-CMP, ω' and φ' involve the 'backbone' atom sequences H—O(8)—P—O(5') and C(4')—C(3')—O(3')—HO(3') respectively. The e.s.d.'s in the non-hydrogen atom torsion angles are about 0·6°.

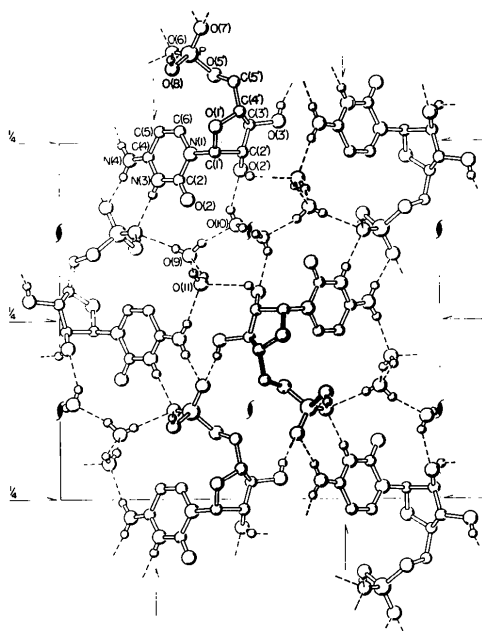| Base–sugar glycosyl bond		Endocyclic torsion angles	
χ	30·7	τ_0	4·0
Backbone bonds		τ_1	−26·0
		τ_2	37·0
ω	−64·2	τ_3	−35·7
φ	137·2	τ_4	20·2
ψ	53·3		
ψ'	84·0	Pseudorotation parameters	
φ'	−76	Phase angle	
ω'	132	P	12·6
Sugar pucker	$^3T_2$	Maximum puckering amplitude	
		τ_m	38·7

Fig. 2. The crystal structure of ara-CMP.3H₂O. The origin is at the upper left, with a directed downward, b from left to right, and c toward the viewer.

given in Table 4. The hydrogen-bond data presented here are based on the CIT refinement, since the UW study failed to unambiguously determine the positions of all H atoms.

The three water molecules are hydrogen-bonded to each other to form a 'stream' of water along the c axis, and they interact through an elaborate network of hydrogen bonds to the base, sugar, and the phosphate moieties of adjacent nucleotides. All three of the independent water molecules exhibit their full complement of hydrogen bonding, the two donor and two acceptor hydrogen bonds being distributed tetrahedrally about each water O atom. As is usual, the carbonyl O(2) and sugar O(1') do not participate in any true hydrogen bonds; other possible hydrogen-bond acceptors which do not contribute to the crystal-packing scheme are O(3'), O(8), and O(5'). The two anionic oxygens O(6) and O(7) of the phosphate group accept a pair of hydrogen bonds from the Watson–Crick base-pairing sites N(3) and N(4) [N(3) is protonated in this compound]. These base–phosphate interactions link adjacent nucleotides along the screw axis parallel to the a

axis. The molecular conformation assumed by ara-CMP allows two water molecules to form a hydrogen-bonding bridge between its phosphate group and N(4) of the cytosine base. The only phosphate hydroxyl group, O(8)—H, is hydrogen-bonded to the anionic phosphate O(6) of a symmetry-related molecule, forming the shortest hydrogen bond in the structure (2·62 Å).

Table 4. *Distances (Å) and angles (°) for the hydrogen bonds D—H···A (CIT refinement)*

D	H	A	D–A ($\sigma \simeq 0.005$ Å)	H···A ($\sigma \simeq 0.05$ Å)	∠D–H···A ($\sigma \simeq 2°$)
N(3)	HN(3)	O(6)(e)	2·923	2·08	176
N(4)	HN(4)	O(7)(e)	2·790	1·98	154
N(4)	H'N(4)	O(11)(a)	2·959	2·11	167
O(2')	HO(2')	O(11)(b)	2·839	2·47	137
O(3')	HO(3')	O(7)(c)	2·661	1·88	160
O(8)	H(8)	O(6)(d)	2·617	2·15	144
				1·93	139
O(9)	H(9)	O(6)(e)	2·873	2·12	173
				1·87	159
O(9)	H'(9)	O(10)	2·279	1·87	162
O(10)	H(10)	O(2')	2·763	2·07	147
O(10)	H(10a)	O(10)(f)	2·784	1·83	175
					168
O(10)	H(10b)	O(10)(g)	2·784	1·86	175
O(11)	H(11)	O(9)(d)	2·803	2·07	150
O(11)	H'(11)	O(9)	2·878	2·19	156

Symmetry code: (a) $x - \frac{1}{2}, \frac{1}{2} - y, 1 - z$; (b) $\frac{1}{2} - x, 1 - y, z - \frac{1}{2}$; (c) $-\frac{1}{2} - x, 1 - y, z - \frac{1}{2}$; (d) $x, y, z + 1$; (e) $x + \frac{1}{2}, \frac{1}{2} - y, 1 - z$; (f) $\frac{1}{2} - x, 1 - y, z + \frac{1}{2}$; (g) $\frac{1}{2} - x, 1 - y, z - \frac{1}{2}$.

References

ALTONA, C. & SUNDARALINGAM, M. (1972). *J. Am. Chem. Soc.* **94**, 8205–8212.

CHWANG, A. K. & SUNDARALINGAM, M. (1973). *Nature (London) New Biol.* **243**, 78–80.

CROMER, D. T. & WABER, J. T. (1965). *Acta Cryst.* **18**, 104–109.

DONOHUE, J. & TRUEBLOOD, K. N. (1960). *J. Mol. Biol.* **2**, 363–371.

HAMILTON, W. C. (1965). *Acta Cryst.* **18**, 502–510.

HOPE, H. (1969). Private communication.

International Tables for X-ray Crystallography (1962). Vol. III, pp. 202–203. Birmingham: Kynoch Press.

LEFEBVRE-SOUBEYRAN, O. & TOUGARD, P. (1973). *C. R. Acad. Sci.* **276**, 403–406.

MAIN, P., GERMAIN, G. & WOOLFSON, M. M. (1970). *MULTAN. A System of Computer Programs for the Automatic Solution of Noncentrosymmetric Crystal Structures.* Univs. of York, England, and Louvain, Belgium.

SHERFINSKI, J. S. & MARSH, R. E. (1973). *Acta Cryst.* B**29**, 192–198.

STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.

SUNDARALINGAM, M. (1969). *Biopolymers*, **7**, 821–860.

SUNDARALINGAM, M. (1975). *Trans. NY Acad. Sci.* **255**, 3–42.

TOUGARD, P. & LEFEBVRE-SOUBEYRAN, O. (1974). *Acta Cryst.* B**30**, 86–89.

VOET, D. & RICH, A. (1970). *Prog. Nucleic Acid Res. Mol. Biol.* **10**, 183–265.