Acta Cryst. (1981). B**37**, 1561–1564

The Structure of the Hemihydrate of 5-Methylcytosine, a Nucleic Acid Base

By C. T. Grainger

School of Physics, The University of New South Wales, Kensington 2033, *Australia*

and D. Bailey

School of Physics and Materials, The New South Wales Institute of Technology, Broadway 2077, *Australia*

(*Received* 21 *November* 1979; *accepted* 28 *January* 1981)

Abstract

The crystal structure of 5-methylcytosine hemihydrate, $C_5H_7N_3O.\frac{1}{2}H_2O$, has been determined using X-rays and direct methods of solution, and refined to a conventional R of 0·047 for all reflexions. The intensities of 1200 independent reflexions were measured on an automatic single-crystal diffractometer using counter methods. The crystals are monoclinic, space group $C2/c$, a = 18·269 (7), b = 8·252 (3), c = 10·632 (4) Å, β = 127·09 (4)°, Z = 8. Measured and calculated densities are 1·390 (15) and 1·394 (1) Mg m^{-3}. The crystals were supplied by Koch–Light Laboratories. The nine non-hydrogen atoms in the molecule are planar within 0·023 Å. The molecules of 5-methylcytosine are hydrogen bonded to each other in such a way as to form two-stranded ribbons. The water of hydration plays an important part by forming bonds within each ribbon and also by forming the only links, apart from van der Waals contacts, between neighbouring ribbons.

Introduction

The title compound (I) occurs naturally in the DNA and RNA (particularly tRNA) of many plants and animals where it comprises a small fraction (up to 12%) of the total pyrimidine base content (see *e.g.*, Davidson, 1976). There is evidence (Coulondre, Miller, Farabaugh & Gilbert, 1978) that in DNA 5-methylcytosine residues are associated with spontaneous base substitution 'hotspots'. A crystallographic study is of obvious interest.

Experimental

Crystals of (I) obtained from Koch–Light Laboratories, Colnbrook, Buckinghamshire, England, were used for X-ray diffraction without recrystallization. Their density was measured by flotation in a mixture of cyclohexane and 1,2-dibromoethane. The conditions for reflexion were $h + k = 2n$ for all reflexions and $l = 2n$ for $h0l$ reflexions, which correspond to space groups $C2/c$ and Cc. Since the intensity statistics were centric the space group $C2/c$ was tried first and this yielded the solution. The cell referred to above (and in the *Abstract*) was derived from an earlier one having the same b and c edges but with a = 14·578 (5) Å and β = 91·516 (30)°, measured at 295 K based on λ = 1·5406 Å for Cu $K\alpha_1$.

The intensities of 1200 independent reflexions were measured using a Siemens automatic single-crystal diffractometer with Ni-filtered Cu $K\alpha$ radiation and a $\theta/2\theta$ scan for $\theta < 70°$. Lorentz, polarization and absorption corrections were applied. The data included 75 reflexions classed as unobserved whose measured intensities were less than three times the e.s.d. based on counting statistics.

The structure was solved using the centrosymmetric direct-methods program *LSAM* (Main, Woolfson & Germain, 1969). All the non-hydrogen atoms appeared in the successful E map (along with a weaker 'ghost' structure). With the aid of least-squares refinements and difference Fourier syntheses all the H atoms were eventually located. Final refinement was by full-matrix least squares (Busing, Martin & Levy, 1962) with all positional parameters, anisotropic and isotropic temperature parameters for non-hydrogen and H atoms respectively, and an extinction parameter refined without constraints. The hydrogen parameters were well-behaved.

The atomic scattering factors, from *International Tables for X-ray Crystallography* (1974), were not corrected for anomalous dispersion. The reflexions were given weights equal to the reciprocal of the variance based on counting statistics.

The final difference Fourier map showed no discrepancies greater than 0·23 e Å^{-3}. R was 0·047 for all reflexions, and 0·044 omitting those measured as zero.

Atomic parameters are given in Table 1.*

---

* Lists of structure factors and anisotropic thermal parameters have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 35802 (8 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

0567-7408/81/081561-04$01.00

Table 1. *Fractional coordinates* ($\times 10^4$ *for non-hydrogen atoms*, $\times 10^3$ *for hydrogen atoms*), *and isotropic temperature parameters B and* B_{eq}

$$B_{eq} = \tfrac{4}{3}\textstyle\sum_i B_{ii}/a_i^{*2}.$$

	x	y	z	B/B_{eq} (Å^2)
N(1)	2037 (1)	1915 (2)	1019 (1)	3·1
C(2)	1316 (1)	1680 (2)	1045 (2)	3·0
N(3)	1479 (1)	1258 (2)	2445 (2)	3·6
C(4)	2346 (1)	1048 (2)	3786 (2)	3·3
C(5)	3088 (1)	1312 (2)	3835 (2)	2·9
C(6)	2893 (1)	1775 (2)	2352 (2)	2·8
O(7)	508 (1)	1830 (2)	−153 (1)	3·8
C(8)	4049 (1)	1084 (3)	5300 (2)	3·8
N(9)	3576 (1)	2050 (2)	2253 (2)	3·7
O(10)	0	233 (2)	2500	3·6
H(11)	21 (2)	−42 (3)	331 (3)	3·9
H(12)	238 (2)	62 (3)	471 (3)	2·0
H(13)	403 (2)	75 (3)	616 (3)	3·5
H(14)	439 (2)	30 (4)	517 (4)	5·0
H(15)	440 (2)	212 (3)	558 (3)	3·4
H(16)	342 (2)	233 (3)	131 (3)	2·2
H(17)	419 (2)	206 (3)	318 (3)	2·5
H(18)	101 (2)	91 (3)	245 (3)	2·7

Description of the structure

Fig. 1 shows details of the asymmetric unit and its intermolecular bonding. The final atomic parameters are in Table 1. Superscripts *a*, *b*, *c*, ... refer to the symmetry codes in Table 2. Distances and angles are in Tables 3, 4 and 5. The non-hydrogen atoms of the molecule are coplanar within 0·023 Å.

The water molecules are situated on diad axes. Each is hydrogen-bonded to four molecules of (I), two bonds being of the form O(10)···H(18)–N(3) and two of the form O(10)–H(11)···O′(7), where the prime indicates an equivalent position.

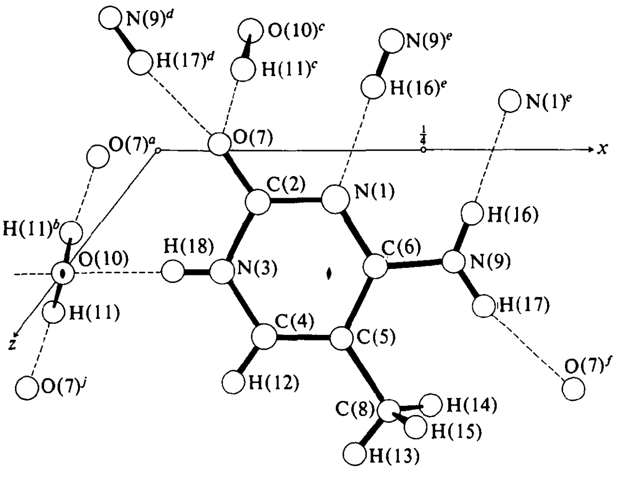

Fig. 1. Intermolecular bonding for 5-methylcytosine hemihydrate projected down the *b* axis.

Table 2. *Symmetry code*

(*a*)	$-x$,	$-y$,	$-z$	(*f*)	$\frac{1}{2}+x$,	$\frac{1}{2}-y$,	$\frac{1}{2}+z$
(*b*)	$-x$,	y,	$\frac{1}{2}-z$	(*g*)	$\frac{1}{2}-x$,	$\frac{1}{2}-y$,	$1-z$
(*c*)	x,	$-y$,	$-\frac{1}{2}+z$	(*h*)	$-x$,	y,	$-\frac{1}{2}-z$
(*d*)	$-\frac{1}{2}+x$,	$\frac{1}{2}-y$,	$-\frac{1}{2}+z$	(*j*)	x,	$-y$,	$\frac{1}{2}+z$
(*e*)	$\frac{1}{2}-x$,	$\frac{1}{2}-y$,	$-z$				

Table 3. *Distances* (Å) *and angles* (°) *for non-hydrogen atoms*

1–2	1·349 (2)	6–1–2	119·9 (2)
2–3	1·375 (2)	1–2–3	118·8 (2)
3–4	1·359 (2)	1–2–7	122·4 (2)
4–5	1·344 (2)	3–2–7	118·8 (2)
5–6	1·443 (2)	2–3–4	121·6 (2)
6–1	1·339 (2)	3–4–5	122·0 (2)
2–7	1·245 (2)	4–5–6	115·0 (2)
5–8	1·497 (2)	4–5–8	122·9 (2)
6–9	1·333 (2)	6–5–8	122·1 (2)
		1–6–5	122·7 (2)
		5–6–9	120·3 (2)
		1–6–9	116·9 (2)

Table 4. *Some distances* (Å) *and angles* (°) *involving hydrogen atoms*

3–18	0·91 (3)	5–8–13	109 (2)
4–12	1·01 (2)	5–8–14	113 (2)
8–13	0·98 (3)	5–8–15	110 (2)
8–14	0·96 (3)	13–8–14	109 (3)
8–15	1·00 (3)	14–8–15	106 (3)
9–16	0·89 (3)	13–8–15	110 (2)
9–17	0·94 (3)	6–9–16	116 (2)
10–11	0·88 (3)	6–9–17	120 (2)
		16–9–17	123 (3)
2–3–18	120 (2)	11–10–11^b	105 (3)
4–3–18	118 (2)	11^c–7–17^d	87 (1)
3–4–12	115 (2)	18–10–11	111 (2)
5–4–12	123 (2)	18–10–11^b	90 (2)

Table 5. *Hydrogen bonding: distances* (Å) *and angles* (°)

$D\cdots H\cdots A$	$H\cdots A$	$D\cdots A$	Angle $D\cdots H\cdots A$
10–11–7^j	1·80 (3)	2·682 (2)	174 (3)
9–16–1^e	2·18 (3)	3·064 (3)	177 (2)
9–17–7^f	2·16 (3)	3·029 (3)	154 (2)
3–18–10	1·95 (3)	2·866 (2)	178 (2)

The molecules of (I) are hydrogen bonded to each other and to the water molecules in such a way as to form two-stranded ribbons as shown in Fig. 2. The molecule (*A*), belonging to the asymmetric unit shown in Fig. 1, forms two hydrogen bonds with molecule (*e*), *viz*, N(1)···H(16)^e–N(9)^e and the corresponding centrosymmetrically related one N(9)–H(16)···N(1)^e.

Atom O(7) forms bonds $O(7)\cdots H(17)^d–N(9)^d$ with molecule (*d*) and $O(7)\cdots H(11)^c–O(10)^c$ with a water molecule which in turn forms a similar bond with molecule (*h*). Such hydrogen bonds interlink molecules (*d*), (*h*), (*A*), (*e*), (*f*), (*b'*), (*A''*), (*g'*), . . ., together to form a somewhat puckered two-stranded ribbon. A neighbouring ribbon on the same level consists of molecules (*b*), (*d'*), (*g*), (*A'*), (*b''*), (*f'*), (*g''*), (*A'''*) . . .. Because of the *c*-glide plane, the ribbons shown in Fig. 2 half overlap those on the levels above and below.

Between neighbouring ribbons there are only van der Waals contacts (Fig. 3), although in Fig. 2 in projection they may appear close together. However, neighbouring ribbons on the same level are linked indirectly *via* water molecules as seen for the pairs of molecules (*A*) and (*b*), (*f*) and (*g*), (*A''*) and (*b''*).

In Fig. 2 each water molecule displays only two of its four hydrogen bonds. The two bonds not shown are directed, as one moves along a ribbon, alternately to ribbons in the layers above and below, thereby holding the structure together. Thus each water molecule plays a dual role by forming bonds within and between ribbons.

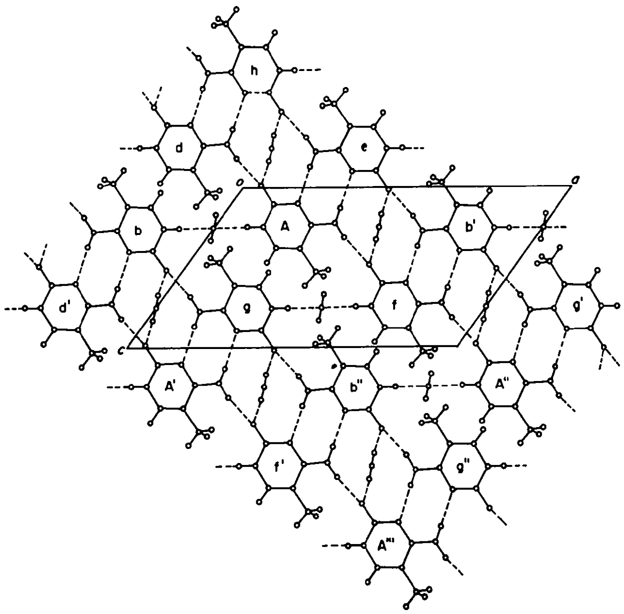

Fig. 2. Hydrogen-bonding pattern showing two-stranded ribbons. This diagram shows molecules in the range $0 < y < \frac{1}{2}b$ projected down the *b* axis. The letter *A* denotes the chosen asymmetric unit. The symmetry codes of Fig. 1 identify other molecules. The remainder, identified by primes, are translational-equivalent molecules.

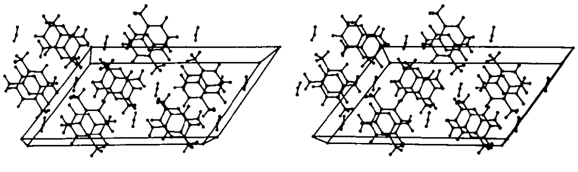

Fig. 3. Stereoscopic view of cell contents and adjacent molecules viewed approximately down the *b* axis.

Discussion

The structure of 5-methylcytosine hemihydrate (I) may be compared with those of anhydrous cytosine (II) (Barker & Marsh, 1964; McClure & Craven, 1973), cytosine monohydrate (III) (Jeffrey & Kinoshita, 1963; McClure & Craven, 1973), 5-cytosineacetic acid (IV) (Marsh, Bierstedt & Eichhorn, 1962), and *N*-methylcytosine (V) (Matthews & Rich, 1964).

In respect of the main molecule there are no surprises. A detailed comparison of the dimensions of the molecules of (I), (II) and (III), whose refinements are of comparable accuracy, shows very close agreement; the linear dimensions of (I) usually lie between those of (II) and (III) except in the vicinity of the added methyl group of (I). The distances, angles and standard deviations were calculated using *ORFFE* (Busing, Martin & Levy, 1964).

In respect of the overall arrangement of molecules, (I) resembles (IV) and (V) in that it forms double-stranded ribbons whereas (II) and (III) form essentially single-stranded chains with weaker bonding, direct or indirect, between chains. However, some elements of the bonding in (I) recall those observed in (II)–(V). Thus, the bonding in (I) of N(1) and N(9) to $N(9)^e$ and $N(1)^e$, respectively, across a symmetry centre is similar to that in (V) except that in the latter the two molecules are not coplanar as they are in (I). The parallel three-bond pattern in (I) for O(7), N(1) and N(9) resembles (IV), in a general way, but not in detail as the latter involves different pairs of bonded atoms. In (II) and (III) the same atoms as in (I) are again involved in bonding, but again not with the corresponding atoms. Both (II) and (III) form single-stranded chains involving bonds $N(3)–H(18)\cdots N'(1)$ and $O(7)\cdots H'(16)–N'(9)$ at one end, and the reverse bonds at the other. The difference between them is that in (II) each chain is directly linked to two adjacent chains *via* two additional bonds $O(7)\cdots H''(17)–N''(9)$ and $N(9)–H(17)\cdots O'''(7)$, whereas in (III) each chain is indirectly linked to three adjacent chains by additional bonds to three water molecules, two from O(7) and one from N(9)–H(17). Indirect linking between chains *via* water molecules is thus a feature of both (III) and (I).

From the point of view of individual atoms or groups the following qualitative remarks may be made about the intermolecular hydrogen bonding found in these five structures.

(*a*) Where water is present it serves as a bridge between chains of molecules by bonding to O(7), N(3) or N(9).

(*b*) The amino group always forms two hydrogen bonds in such a way as to link three molecules together.

In (II), (III) and (IV) all are directed to O atoms, usually O(7) of adjacent molecules, but in one case to the O of a water molecule. In both (I) and (V) one of the bonds is with the N(1) atom of a second molecule and the other with the O(7) of a third molecule.

(*c*) N(1) is always hydrogen-bonded to another N atom, variously equivalents of N(9), N(3) and N(1) [in (IV) a half-hydrogen is involved].

(*d*) The N(3)–H(18) group is always bonded. In two cases it bonds to an N(1) equivalent. In (I) the bond is with the O of a water molecule.

(*e*) The carbonyl oxygen O(7) forms one, two or three bonds. If water is absent all bonds are with N(9) equivalents. If water is present (I and III) O(7) forms hydrogen bonds with it.

As the above discussion indicates, the hydrogen bonding exhibited in the present structure is directionally compatible with the triple hydrogen-bond scheme for the pairing of guanine and cytosine in DNA as first suggested by Pauling & Corey (1956). However, in the present structure we find instead a double hydrogen-bond pattern between molecules within a ribbon, as occurs in crystalline *N*-methylcytosine (V). This preference could be associated with packing considerations due to the bulk of the methyl group and/or the tendency noted for the carbonyl O to form hydrogen bonds with water where present, as it is in most biological contexts.

We wish to express thanks to Mr D. C. Craig of the School of Chemistry for collecting the intensity data and assisting in computing, Dr A. G. Mackinlay of the School of Biochemistry for helpful discussions, and Miss Y. Rodriguez of the School of Physics for preparing the figures from diagrams drawn by *ORTEP* (Johnson, 1976). We thank the University of New South Wales for the provision of computing time on the University CDC Cyber computer.

References

BARKER, D. L. & MARSH, R. E. (1964). *Acta Cryst.* **17**, 1581–1587.

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-305. Oak Ridge National Laboratory, Tennessee.

BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1964). *ORFFE*. Report ORNL-TM-306. Oak Ridge National Laboratory, Tennessee.

COULONDRE, C., MILLER, J. H., FARABAUGH, P. J. & GILBERT, W. (1978). *Nature (London)*, **274**, 775–780.

DAVIDSON, J. F. (1976). *The Biochemistry of the Nucleic Acids*, p. 83 ff. London: Chapman & Hall.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

JEFFREY, G. A. & KINOSHITA, Y. (1963). *Acta Cryst.* **16**, 20–28.

JOHNSON, C. K. (1976). *ORTEP* II. Report ORNL-5138. Oak Ridge National Laboratory, Tennessee.

MCCLURE, R. J. & CRAVEN, B. M. (1973). *Acta Cryst.* B**29**, 1234–1238.

MAIN, P., WOOLFSON, M. M. & GERMAIN, G. (1969). *LSAM. A System of Computer Programs for the Automatic Solution of Centrosymmetric Crystal Structures.* Univs. of York, England, and Louvain, Belgium.

MARSH, R. E., BIERSTEDT, R. & EICHHORN, E. L. (1962). *Acta Cryst.* **15**, 310–316.

MATTHEWS, F. S. & RICH, A. (1964). *Nature (London)*, **201**, 179–180.

PAULING, L. & COREY, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164–181.

Acta Cryst. (1981). B**37**, 1564–1568

X-ray Structure Analysis and Molecular Conformation of *tert*-Butyloxycarbonyl-L-prolylproline (Boc-Pro-Pro)

BY M. E. KAMWAYA, O. OSTER AND H. BRADACZEK

Institut für Kristallographie, Freie Universität Berlin, Takustrasse 6, D-1000 Berlin 33, Federal Republic of Germany

(*Received* 18 *December* 1980; *accepted* 4 *February* 1981)

Abstract

$C_{15}H_{24}N_2O_5$ is orthorhombic, space group $P2_12_12_1$, with $a = 16{\cdot}968$ (1), $b = 14{\cdot}490$ (7), $c = 6{\cdot}664$ (9) Å, $V = 1638{\cdot}45$ Å^3, $M_r = 312{\cdot}39$, $D_o = 1{\cdot}2555$, $D_c = 1{\cdot}266$ Mg m^{-3}; $F(000) = 672$, $Z = 4$. The final $R = 3{\cdot}6\%$ for 1492 reflexions. Molecules are joined in pairs by a hydrogen bond O···OH, 2·613 Å. C^α is *cis* relative to the O atom of Boc and is *trans* relative to the C^α of the following ring. The absolute configuration of the non-centrosymmetric structure and the molecular conformation were determined by anomalous dispersion. The groups $NC^\alpha C^\gamma C^\delta$ of the two five-membered rings are fairly planar. In both cases C^β is

0567-7408/81/081564-05$01.00