

The Crystal and Molecular Structure of a Hydrogen-Bonded Complex Between 1-Methylthymine and 9-Methyladenine*

BY KARST HOOGSTEEN

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California, U.S.A.

(Received 25 September 1962)

Crystals of a 1:1 hydrogen-bonded complex between 1-methylthymine and 9-methyladenine can be grown from an aqueous solution containing equimolecular quantities of the two compounds. The crystals are monoclinic, with $a = 8.304$, $b = 6.552$, $c = 12.837$ Å, and $\beta = 106^\circ 50'$. The space group is $P2_1/m$, with two base-pair complexes in the unit cell. The structure was refined with three-dimensional data taken with copper radiation. The positional coordinates and anisotropic temperature factors of the heavy atoms were obtained by least-squares analyses. The hydrogen atoms, except those of two methyl groups, were located from a three-dimensional difference Fourier synthesis. The 1-methylthymine and 9-methyladenine molecules form a planar base pair lying in a mirror plane and are connected to each other by two nearly linear hydrogen bonds, from the NH_2 group of 9-methyladenine to O(9) of 1-methylthymine (2.846 Å) and from N(3) of 1-methylthymine to N(7) of 9-methyladenine (2.924 Å).

This structure differs from the adenine-thymine pairing proposed by Watson & Crick, where N(3) of thymine is hydrogen bonded to N(1) of adenine. The distance between the methyl group at N(1) of 1-methylthymine and the one at N(9) of 9-methyladenine is 8.645 Å, whereas this distance is 11.1 Å in the pairing proposed by Watson & Crick.

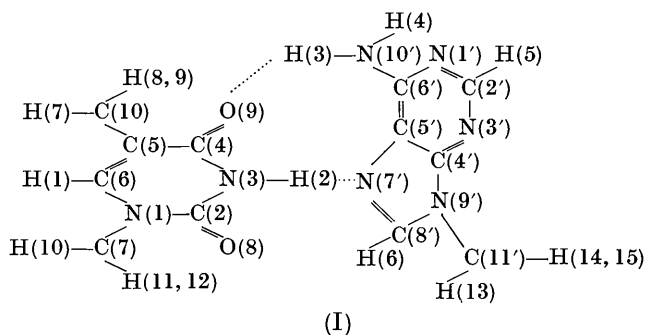
Introduction

As part of a program of research on the structure of nucleic acids, attempts have been made to grow crystals containing hydrogen-bonded pairs of purine and pyrimidine derivatives. A preliminary report on the structure of a 1:1 complex between 1-methylthymine and 9-methyladenine (I) has already been given (Hoogsteen, 1959). The manner in which these

hydrogen-bonded to the amino group of 9-methyladenine. This paper describes the refinement of the approximate structure with the use of three-dimensional data and discusses the conclusions that can be drawn from the geometry of the base pair as it occurs in the crystals.

Experimental

Equimolecular quantities of 1-methylthymine and 9-methyladenine were dissolved in hot water. Slow cooling and subsequent evaporation of this solution to dryness at room temperature gave good crystals of the 1:1 complex in the form of well developed needles, elongated in the direction of the b axis. A needle bounded by the forms $\{100\}$, $\{001\}$, and $\{010\}$, with a cross section of 0.2×0.2 mm, was mounted with the rotation axis parallel to the b axis. This crystal was used in the preparation of multiple-film equi-inclination Weissenberg photographs of layers $h0l$ through $h5l$. By cleaving a needle parallel to (010) , a fragment with approximately the same cross section as the first crystal was obtained and mounted parallel to the a axis. This crystal was used to collect the intensities of layers $0kl$ through $7kl$. Ni-filtered $\text{Cu } K\alpha$ radiation was used. The intensities were estimated visually with the aid of a calibration strip. After correction for the Lorentz and polarization factors, the two sets of data were correlated to the same relative scale. In all, 1361 reflections were measured with an intensity different from zero, of which 856 were obtained about both the a and the b axes. These reflections were used to obtain an experimental



two molecules are hydrogen-bonded to each other was established with the use of $h0l$ and $h1l$ intensities. They form a planar base pair with N(3) of 1-methylthymine connected to N(7) of 9-methyladenine, and the C(4)=O carbonyl group of 1-methylthymine

* Contribution No. 2888 from the Gates and Crellin Laboratories of Chemistry. This investigation was supported, in part, by two grants to the California Institute of Technology, a grant CVRE 121 from the National Foundation, and a grant No. G-9467 from the National Science Foundation.

estimate of the standard deviation in the observed structure factors, $\sigma(F_o)$. The values of

$$\Delta F = \frac{1}{2}(|F_o^a| - |F_o^b|)$$

where F_o^a and F_o^b are the structure factors measured along the a and b axes, were averaged over 6 regions of similar F_o values and plotted against the average value of F_o for that region. These points were then fitted by the least-squares method to the equation $\Delta F = A + BF_o + CF_o^2$, giving $A = 0.27$, $B = 0.012$, $C = 0.0012$. The coefficients A and B were used in the final stages of the refinement to calculate the weights of the reflections.

Accurate cell dimensions were determined from zero-layer Weissenberg photographs taken around the a and b axes. A Weissenberg camera with the Straumanis method of film loading and a radius of 10 cm was used. Observed values of $\sin^2 \theta$ for 6 $0kl$ and 12 $h0l$ reflections, occurring in the back reflection region where the splitting of the α_1 and α_2 reflections was complete, were used as input to a least-squares treatment; the cell dimensions are:

$$\begin{aligned} a &= 8.304 \pm 0.002, \quad b = 6.552 \pm 0.002, \\ c &= 12.837 \pm 0.003 \text{ \AA}; \quad \beta = 106^\circ 50' \pm 3'; \\ D_x &= 1.438, \quad D_m = 1.433 \text{ g.cm}^{-3} \end{aligned}$$

for λ Cu $K\alpha_1 = 1.54051 \text{ \AA}$.

The standard deviations were derived from the least-squares calculation in the usual way.

Refinement of the structure

In the preliminary paper (Hoogsteen, 1959) an account was given of the structure determination and the partial refinement with the $h0l$ and $h1l$ data until the disagreement index was 0.165. Two additional least-squares analyses were calculated in which the positional parameters x and z together with the individual isotropic temperature factors were adjusted. The disagreement index decreased to 0.145. The refinement was then continued with three-dimensional data on a Burroughs 220 computer.

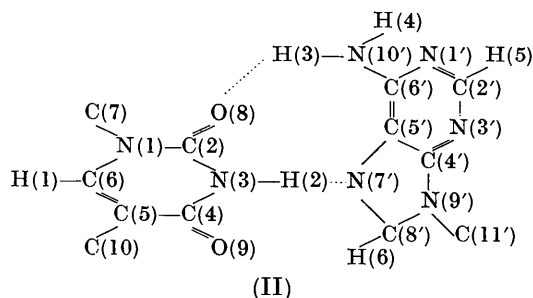
The monoclinic structure-factor and least-squares program allows for the refinement of the positional parameters, six anisotropic temperature factors and an individual scale factor for each atom. It determines the minimum value of the expression $\sum w(|F_o^2| - |F_c^2|)^2$ and uses the scattering factors of Berghuis, Haanappel, Potters, Loopstra, MacGillavry & Veenendaal (1955) for C, N, O, and of McWeeny (1951) for H. The normal equations are solved with neglect of all interatomic cross terms. For each atom only cross terms between the x and z parameters and between the six anisotropic temperature factors together with a scale factor are collected. The space group of the crystals had been found to be $P2_1$ or $P2_1/m$ (Hoogsteen, 1959). If it is assumed to be $P2_1/m$, all carbon, nitrogen, and oxygen atoms of the base pairs must lie in the mirror planes

at $y = \frac{1}{4}$ and $y = \frac{3}{4}$. The positional parameters x and z , the anisotropic temperature factors, α , β , γ , and ϵ and one scale factor remain to be refined. The scale factor was adjusted by taking the average value of the indicated corrections to the individual atomic scale factors. The weights were calculated according to the Hughes (1941) weighting scheme:

$$\begin{aligned} 1/w &= 1/F_o^2 \quad \text{if } F_o > 4F_{\min}, \\ 1/w &= 1/4F_o F_{\min} \quad \text{if } F_o \leq 4F_{\min}. \end{aligned}$$

These weights do not accurately reflect the uncertainties of F_o^2 values, but they were considered to be adequate at this stage of refinement. Accidental absences were included in the refinement calculations only if their calculated values exceeded the minimum observable ones. After five refinement cycles the disagreement factor decreased to 0.125.

At this point it was decided to attempt the resolution of an ambiguity resulting from the configuration of the 1-methylthymine molecule. With only minor changes in the atomic positions, the crystal structure could accommodate base pairs (II), of which O(9)



instead of O(8) is hydrogen bonded to H(4) and O(8) instead of O(9) is hydrogen bonded to H(3). The principal difference between this structure and the one assumed until so far is an interchange of the atoms N(1) and C(5) of the 1-methylthymine molecule.

The same type of structural ambiguity is possible in the structure of 1-methylthymine (Hoogsteen, 1963), but the positions of N(1) and C(5) could be identified at a similar stage of refinement by means of a difference Fourier synthesis, and subsequent calculations gave no indications of disorder. A three-dimensional difference Fourier map of the present structure was therefore calculated. In this difference map the electron densities at the positions of N(1) and C(5) were not significantly different from those at other atomic positions, indicating that the configuration initially chosen (I) was the correct one. The positions of all hydrogen atoms directly attached to the ring systems of the adenine and thymine molecules were also indicated in this map by distinct maxima. One hydrogen atom of each of the three methyl groups could also be located in the mirror plane, and it was therefore assumed that the orientation of the methyl groups obeyed the symmetry restriction of the mirror

Table 1. Final coordinates and calculated estimated standard deviations of the $P2_1/m$ least-squares refinements, and anisotropic temperature factors

The coordinates are listed in fractions of the unit-cell axes; $y = \frac{1}{2}$

	x	$\sigma(x)$ (Å)	z	$\sigma(z)$ (Å)	$10^4\alpha$	$\sigma(\alpha)$	$10^4\beta$	$\sigma(\beta)$	$10^4\gamma$	$\sigma(\gamma)$	$10^4\epsilon$	$\sigma(\epsilon)$
N(1)	0.7485	0.003	0.4370	0.003	143	(6)	285	(11)	77	(3)	79	(6)
C(2)	0.7598	0.004	0.5461	0.004	125	(7)	275	(12)	70	(3)	35	(6)
N(3)	0.6078	0.003	0.5693	0.003	124	(5)	284	(10)	52	(2)	33	(5)
C(4)	0.4506	0.004	0.4847	0.004	129	(6)	261	(11)	65	(3)	35	(6)
C(5)	0.4481	0.004	0.3835	0.004	147	(7)	290	(13)	64	(3)	38	(6)
C(6)	0.5936	0.004	0.3591	0.004	164	(7)	293	(13)	60	(3)	55	(7)
C(7)	0.9024	0.004	0.4043	0.005	152	(8)	399	(17)	101	(4)	116	(9)
O(8)	0.8905	0.003	0.6184	0.003	127	(5)	444	(12)	89	(2)	18	(5)
O(9)	0.3234	0.003	0.5265	0.002	138	(5)	408	(11)	59	(2)	53	(5)
C(10)	0.2781	0.004	0.2991	0.004	167	(8)	394	(15)	52	(3)	17	(6)
N(1')	0.2659	0.003	0.9123	0.003	171	(7)	299	(11)	69	(3)	76	(6)
C(2')	0.3687	0.004	1.0164	0.004	191	(8)	318	(13)	63	(3)	73	(7)
N(3')	0.5324	0.004	1.0542	0.003	181	(7)	312	(11)	58	(2)	45	(6)
C(4')	0.6012	0.004	0.9712	0.003	157	(7)	204	(10)	62	(3)	38	(6)
C(5')	0.5184	0.004	0.8618	0.004	155	(7)	259	(12)	61	(3)	38	(6)
C(6')	0.3416	0.004	0.8319	0.004	162	(7)	236	(11)	63	(3)	51	(6)
N(7')	0.6310	0.003	0.8010	0.003	145	(6)	335	(12)	63	(2)	56	(6)
C(8')	0.7807	0.004	0.8742	0.004	155	(7)	314	(13)	75	(3)	49	(7)
N(9')	0.7714	0.003	0.9785	0.003	146	(6)	279	(11)	67	(2)	16	(5)
N(10')	0.2461	0.003	0.7285	0.003	128	(6)	465	(14)	61	(2)	25	(3)
C(11')	0.9088	0.004	1.0789	0.004	160	(8)	423	(17)	74	(3)	-21	(8)

plane. The coordinates of the hydrogens lying out of the mirror plane were then calculated, and these positions were found to correspond to extended and weak maxima in the difference map. The coordinates of the hydrogens thus obtained were then inserted in the subsequent least-squares refinements; they were assigned one isotropic temperature factor with $B=4.0 \text{ \AA}^2$.

The refinement was completed by 8 more least-squares cycles in which the weights were calculated according to the expression

$$1/w = \frac{1}{0.50 + 0.54F_o + 0.024F_o^2}$$

The coefficients of F_o and F_o^2 were derived from A and B of the experimentally determined equation for $\sigma(F_o)$. The constant 0.50 was added to decrease the weights of the weak and unobserved reflections. The calculated corrections to the input parameters of the hydrogen atoms were also applied. The average correction for the positional parameters of the heavy atoms for the final calculation was 0.0004 Å with a maximum value of 0.001 Å. None of the corrections of the other parameters was larger than 0.2 times their standard deviations, and refinement based on the centrosymmetric space group was terminated at this point. The disagreement index R was 0.081, with exclusion of the unobserved reflections. The final positional parameters x and z , together with the anisotropic temperature factors α, β, γ , and ϵ^* for the carbon, nitrogen, and oxygen atoms obtained in this way, are listed in Table 1.

During the refinement the interatomic distances

* These temperature parameters are the coefficients of the expression $T = \exp -(\alpha h^2 + \beta k^2 + \gamma l^2 + \delta hk + \epsilon hl + \eta kl)$.

and bond angles involving the hydrogen atoms were calculated as a check on their positional coordinates.

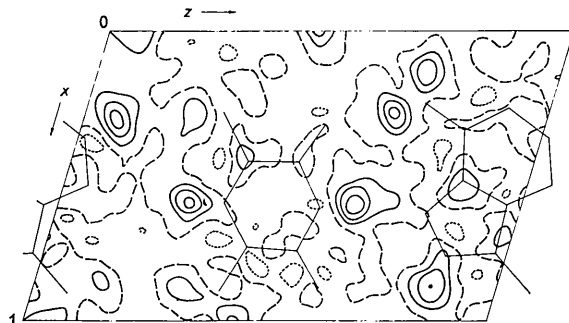


Fig. 1. The section parallel to (010) at $y = \frac{1}{2}$ of the final difference-Fourier synthesis. Contours are drawn at intervals of $0.2 \text{ e.}\text{\AA}^{-3}$. Broken and dotted lines indicate the zero and negative contours.

Although the distances were not unreasonable, it became clear that the hydrogen atoms lying above and below the mirror plane were moving towards the mirror plane. The angles which the carbon atoms subtend to those hydrogen atoms were as a result too small, about 80° for the last refinement calculation. A final difference-Fourier synthesis was therefore calculated based upon the parameters of Table 1; the contributions of the hydrogen atoms to the structure factors were omitted. The section at $y = \frac{1}{2}$, parallel to (010), *i.e.* through the mirror plane, is shown in Fig. 1. The hydrogen atoms that are directly attached to the ring systems of the thymine and adenine molecules all show up as maxima with a height of about $0.5 \text{ e.}\text{\AA}^{-3}$. Their coordinates were determined from this map by the interpolation method of Booth (1948); they are listed in Table 2. In Fig. 2

Table 2. *Coordinates of hydrogen atoms determined from the final difference-Fourier analysis with space group $P2_1/m$*

The temperature parameters were derived from the final least-squares calculation

	<i>x</i>	<i>y</i>	<i>z</i>	<i>B</i>
H(1)	0.596	0.250	0.642	3.6 Å ²
H(2)	0.595	0.250	0.282	2.0
H(3)	0.275	0.250	0.652	3.7
H(4)	0.137	0.250	0.711	2.9
H(5)	0.287	0.250	1.061	2.9
H(6)	0.881	0.250	0.857	3.4
H(13)	0.018	0.250	0.060	3.7
H(14)	0.884	0.127	0.122	5.7

are shown sections through the planes containing the hydrogen atoms of the three methyl groups at C(7), C(10), and C(11'). Although the three maxima around C(11') in Fig. 2(c) are low (about 0.4 e.Å⁻³) their coordinates were determined and accepted as the coordinates of the hydrogen atoms of C(11'). They are also listed in Table 2. The hydrogen coordinates obtained so far did not differ significantly from those derived from the least-squares analysis, and the temperature parameters of the last least-squares calculation are also listed in Table 2.

Table 3. *The assumed coordinates of the hydrogen atoms at C(7) and C(10)*

	<i>x</i>	<i>y</i>	<i>z</i>
H(7)	0.287	0.180	0.232
H(8)	0.191	0.192	0.332
H(9)	0.225	0.392	0.286
H(10)	0.874	0.213	0.326
H(11)	0.959	0.386	0.421
H(12)	0.989	0.158	0.452

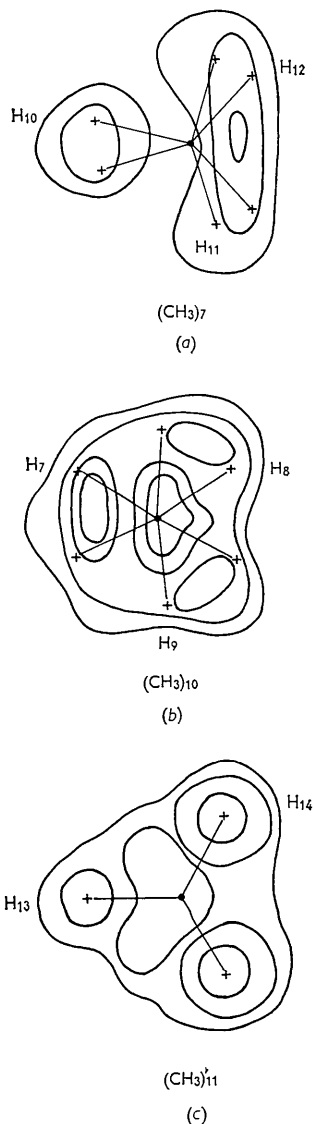


Fig. 2. Sections through the electron density associated with the hydrogen atoms of the three methyl groups. Contours are at intervals of 0.2 e.Å⁻³, beginning with the 0.2 e.Å⁻³ contour.

It is clear that the electron distribution around C(7) and C(10) cannot be represented by single positions of three hydrogen atoms. An approximation for these electron density distributions was therefore made by putting six hydrogen atoms, with half weight, around each of the two carbon atoms in the manner indicated in Fig. 2(a) and (b). These locations were chosen with a C-H distance of 1.02 Å, which is the average value found for other C-H distances in this structure determination. The calculated coordinates of the six hydrogen atoms are listed in Table 3.

One additional structure-factor and least-squares calculation was done with the coordinates listed in Tables 1, 2, and 3 as input parameters. The six hydrogens with half weight were assigned a temperature factor $B=4.0$ Å². This calculation did not show any improvement in the agreement between the observed and calculated structure factors compared with the previous one; no significant corrections to the positions of the heavy atoms were found. The structure factors obtained from this calculation together with their observed values are listed in Table 4.

*Attempts to refine the *y* coordinates*

To this point the structure has been refined on the basis of the space-group symmetry $P2_1/m$, so that all the carbon, nitrogen, and oxygen atoms were lying in the mirror plane. Relatively small deviations of the atoms from the mirror plane, however, may occur if the planes of the two molecules do not coincide with (010) or if the bases are not planar molecules. Such departures of the atoms from the mirror planes would be incorporated in the anisotropic temperature factors β_i . The principal axes of the thermal ellipsoids (Table 8) normal to (010) are in all cases except for C(6') and N(9') the largest ones.

Although this feature is common for a layer structure, some of these axes, especially those of the annular substituents, are larger than could be expected from temperature movement alone, indicating that deviations from the mirror plane might exist. Furthermore, the observed electron distributions due to the hydrogen atoms of C(7) and C(10') (Fig. 2(a) and (b)) cannot be described by single positions conforming to the symmetry restrictions of the mirror plane. The question now arises whether or not those deviations from the mirror plane are ordered in the sense that they conform to the symmetry operations of the twofold screw axes. In the first case, a better description of the structure can be given by assigning it the space group $P2_1$, but in the second case no real advantage can be obtained by the omission of the mirror plane; the deviations in this state occur in a disordered way throughout the crystal. In order to find an answer to this question, the refinement was continued with the calculations based on the space group $P2_1$.

In order to remove the mirror plane from the space group $P2_1/m$, some atoms had to be given y coordinates different from $\frac{1}{2}$. In the structure of adenine hydrochloride (Broomhead, 1948; Cochran, 1951) the z coordinates were calculated on the basis of a planar molecule, so that no deviations from planarity were known for the 9-methyladenine molecule. For the structure of 1-methylthymine (Hoogsteen, 1963) the following departures from the best plane of the molecule were found: N(1)=0.014 Å, C(2)=0.005 Å, N(3)=0.005 Å, C(4)=-0.009 Å, C(5)=0.002 Å, C(6)=0.009 Å, C(7)=-0.035 Å, O(8)=-0.017 Å, O(9)=-0.035 Å, and C(10)=0.028 Å. The atoms C(7), O(8), O(9), and C(10) were assigned y coordinates that gave the same distances for these atoms to (010) in the AT structure. In the following structure-factor and least-squares calculations only the positional and temperature parameters of the heavy atoms were refined. The weighting scheme used in the final stages of the centrosymmetric refinement was applied. The calculations were terminated after six refinement cycles, since no shift was larger than 0.03 times the standard deviation.

A comparison between the coordinates obtained in this way and those of Table I showed that none of the final parameters of the centrosymmetric refinement was significantly different from those acquired by the acentric refinement. The largest shift in the y coordinate was for C(7): 0.061 Å, with a calculated standard deviation of 0.01 Å. Although the disagreement index R decreased from 0.081 to 0.073, the sum of the squares of the weighted residuals, $\sum w[\Delta F^2]^2$, remained constant. Since the acentric structure contains 63 parameters more than the centrosymmetric one the standard deviation for an observation of unit weight actually increased, indicating that the acentric refinement did not converge to a set of coordinates significantly better than those

derived from the refinement with space group $P2_1/m$. The following discussion is therefore based on the parameters of the centrosymmetric structure listed in Tables 1, 2, and 3.

Discussion

In Fig. 3 a view of the structure down the b axis is presented. The molecules in heavy lines belong to the same layer. Part of an adjacent layer, one-half of the b axis down and related to the upper layer by the twofold screw axes, is also shown in the top half of the figure. The structure is a layer structure with

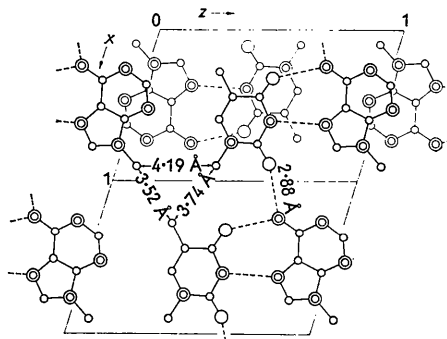


Fig. 3. A view of the structure viewed down the b axis. The small circles represent carbon atoms, the large single circles oxygen atoms and the double circles nitrogen atoms.

layers of base pairs parallel to (010) stacked over each other. The amino group of 9-methyladenine is hydrogen-bonded to O(8) of the 1-methylthymine molecule that lies one unit cell up along the negative direction of the a axis. Each layer is thus formed by columns of base pairs running parallel to a . The cohesion of the structure within each layer is provided only by the van der Waals interaction between these columns.

The three methyl groups at C(7), C(10), and C(11') are pointing towards each other, with the following distances:

$$\begin{aligned} C(7)-C(11') &= 4.19 \text{ \AA}, & C(7)-C(10) &= 3.74 \text{ \AA}, \\ C(10)-C(11') &= 3.52 \text{ \AA}. \end{aligned}$$

There are no short intermolecular distances between molecules in the same layer. The packing of the base pairs within a layer is shown in Fig. 4.

Since the b axis is 6.552 Å in length, no interatomic distance between molecules of adjacent layers is shorter than 3.276 Å. The shortest distance across two layers occurs between C(4') of 9-methyladenine and C(2') of the molecule related to it by a twofold screw axis; it is 3.285 Å.

Interatomic bond distances and angles

The interatomic bond distances and angles of the base pair are shown in Fig. 5. In Table 5 are listed

Table 5. *Bond distances and angles of 1-methylthymine*

	AT	NMT	Δ	$\Delta/\sigma(\Delta)$
	($\sigma =$ 0.005 Å)	($\sigma =$ 0.004 Å)		
N(1)-C(2)	1.376 Å	1.379 Å	0.003 Å	0.5
C(2)-N(3)	1.378	1.379	0.001	0.2
N(2)-C(4)	1.377	1.375	0.002	0.3
C(4)-C(5)	1.422	1.432	0.010	1.6
C(5)-C(6)	1.333	1.346	0.013	2.0
C(6)-N(1)	1.382	1.383	0.001	0.2
C(7)-N(1)	1.456	1.470	0.014	2.2
O(8)-C(2)	1.207	1.214	0.007	1.1
O(9)-C(4)	1.238	1.237	0.001	0.2
C(10)-C(5)	1.510	1.497	0.013	2.0
$(\sigma = 0.2^\circ)$				
N(1)-C(2)-N(3)	115.0°	115.4°	0.4°	1.1
C(2)-N(3)-C(4)	126.3	126.3	0.0	0.0
N(3)-C(4)-C(5)	115.7	116.1	0.4	1.1
C(4)-C(5)-C(6)	119.0	118.3	0.7	1.9
C(5)-C(6)-N(1)	123.2	123.3	0.1	0.3
C(6)-N(1)-C(2)	120.4	120.6	0.2	0.6
C(7)-N(1)-C(6)	120.1	121.2	1.1	3.1
C(7)-N(1)-C(2)	119.4	118.2	1.2	3.3
N(1)-C(2)-O(8)	124.3	123.3	1.0	2.8
N(3)-C(2)-O(8)	120.7	121.3	0.6	1.7
N(3)-C(4)-O(9)	119.9	120.0	0.1	0.3
C(5)-C(4)-O(9)	124.4	123.9	0.5	1.4
C(10)-C(5)-C(4)	117.3	119.3	2.0	5.6
C(10)-C(5)-C(6)	123.6	122.4	1.2	3.3

Table 6. *Bond distances and angles of 9-methyladenine*

	AT	A+HCl'
	($\sigma = 0.005$ Å)	
N(1')-C(2')	1.361 Å	1.37 Å
C(2')-N(3')	1.304	1.30
N(3')-C(4')	1.347	1.36
C(4')-C(5')	1.373	1.37
C(5')-C(6')	1.406	1.40
C(6')-N(1')	1.355	1.38
C(5')-N(7')	1.381	1.37
N(7')-C(8')	1.323	1.35
C(8')-N(9')	1.363	1.33
N(9')-C(4')	1.389	1.36
C(6')-N(10')	1.335	1.30
N(9')-C(11')	1.453	—
$(\sigma = 0.3^\circ)$		
N(1')-C(2')-N(3')	130.9°	124.5
C(2')-N(3')-C(4')	110.0	112.5
N(3')-C(4')-C(5')	127.4	128.0
C(4')-C(5')-C(6')	116.9	118.0
C(5')-C(6')-N(1')	118.0	114.0
C(6')-N(1')-C(2')	116.8	123.0
C(5')-N(7')-C(8')	104.4	102.0
N(7')-C(8')-N(9')	112.9	115.0
C(8')-N(9')-C(4')	106.2	105.0
N(9')-C(4')-C(5')	105.5	107.0
C(4')-C(5')-N(7')	111.0	111.0
N(10')-C(6')-C(5')	123.0	126.5
N(10')-C(6')-N(1')	119.0	119.5
C(8')-N(9')-C(11')	128.2	—
C(11')-N(9')-C(4')	125.6	—

A discussion of the dimensions of 1-methylthymine in terms of valence bond structures has already been given for the structure of 1-methylthymine and this discussion seems to apply to the molecule as it occurs in the structure of the base pair as well. The dimensions of the 9-methyladenine molecule are listed in Table 6, together with those of adenine hydrochloride as determined in the structure of this compound (Broomhead, 1948; Cochran, 1951). The standard deviations in the bond lengths for the latter structure were estimated not to exceed 0.01 Å. A comparison of the bond lengths of this molecule shows that the addition of a proton to N(1) apparently does not alter the bond distances appreciably. Larger differences are found in the bond angles. The angle C(6)-N(1)-C(2) is 116.8° in 9-methyladenine, whereas it is 123.0° in the adenine ion. The attachment of a hydrogen to the N(1) atom apparently increases the ring bond angle by about 6°. This difference is in good agreement with the one given by Pauling & Corey (1956).

The most important feature of this structure is the manner in which the two molecules are hydrogen-bonded together. O(9) of 1-methylthymine is hydrogen-

Table 7. *Distances and angles involving the hydrogen atoms for which the coordinates were determined from the final difference-Fourier synthesis*

C(6)-H(1)	0.99 Å
N(3)-H(2)	0.97
N(10')-H(3)	1.08
N(10')-H(4)	0.87
C(2')-H(5)	1.01
C(8')-H(6)	0.92
C(11')-H(13)	1.00
C(11')-H(14)	1.03
C(4)-N(3)-H(2)	109.3°
C(2)-N(3)-H(2)	124.3
C(5)-C(6)-H(1)	116.4
N(1)-C(6)-H(1)	120.5
C(6')-N(10')-H(3)	133.0
C(6')-N(10')-H(4)	122.2
N(1')-C(2')-H(5)	102.9
N(3')-C(2')-H(5)	126.2
N(7')-C(8')-H(6)	123.9
N(9')-C(8')-H(6)	123.2
N(9')-C(11')-H(13)	108.5
N(9')-C(11')-H(14)	104.3

bonded to the NH₂ group of 9-methyladenine with N-H...O = 2.847 Å. N(3) of 1-methylthymine is hydrogen-bonded to N(7) of 9-methyladenine at a distance of 2.923 Å. In a review given by Fuller (1959), two examples of a ring NH to ring N hydrogen bond are given: xanthazol monohydrate (Nowacki & Bürcki, 1955) with a distance of 2.88 Å, and isocyanic acid (Senko & Templeton, 1958) with N-H...N = 2.90 Å. These distances are rather short. Although no attempt was made to locate the hydrogen atoms in these two compounds, disorder of their positions

Table 8. *Principal axes and direction cosines with respect to a^* , b , and c of thermal ellipsoids*

	Axis	$g_{i(1)}$	$g_{i(2)}$	$g_{i(3)}$	$B_{i/4}$
N(1)	1	0	1	0	1.225
	2	0.435	0	0.900	1.167
	3	0.900	0	-0.435	0.842
C(2)	1	0	1	0	1.180
	2	-0.178	0	0.984	1.138
	3	0.984	0	0.178	0.778
N(3)	1	0	1	0	1.219
	2	-0.587	0	0.810	0.882
	3	0.810	0	0.587	0.734
C(4)	1	0	1	0	1.119
	2	-0.273	0	0.962	1.051
	3	0.962	0	0.273	0.794
C(5)	1	0	1	0	1.243
	2	-0.499	0	0.866	1.076
	3	0.866	0	0.499	0.879
C(6)	1	0	1	0	1.255
	2	0.962	0	-0.273	1.046
	3	0.273	0	0.962	0.911
C(7)	1	0	1	0	1.713
	2	0.459	0	0.888	1.544
	3	0.888	0	-0.459	0.807
O(8)	1	0	1	0	1.904
	2	-0.204	0	0.979	1.520
	3	0.979	0	0.204	0.768
O(9)	1	0	1	0	1.752
	2	0.534	0	0.845	0.885
	3	0.845	0	-0.531	0.869
C(10)	1	0	1	0	1.690
	2	0.812	0	-0.584	1.221
	3	0.584	0	0.812	0.735
N(1')	1	0	1	0	1.252
	2	0.872	0	0.488	1.115
	3	-0.488	0	0.873	0.975

	Axis	$g_{i(1)}$	$g_{i(2)}$	$g_{i(3)}$	$B_{i/4}$
C(2')	1	0	1	0	1.365
	2	0.999	0	0.037	1.204
	3	-0.037	0	0.999	0.926
N(3')	1	0	1	0	1.339
	2	0.919	0	-0.394	1.196
	3	0.394	0	0.919	0.872
C(4')	1	-0.707	0	0.707	1.096
	2	0	1	0	0.880
	3	0.707	0	0.707	0.883
C(5')	1	0	1	0	1.113
	2	-0.703	0	0.711	1.078
	3	0.711	0	0.703	0.878
C(6')	1	0.855	0	-0.518	1.055
	2	0	1	0	1.013
	3	0.518	0	0.855	0.941
N(7')	1	0	1	0	1.436
	2	0.272	0	0.962	0.953
	3	0.962	0	-0.272	0.916
C(8')	1	0	1	0	1.348
	2	-0.215	0	0.976	1.184
	3	0.976	0	0.215	0.971
N(9')	1	-0.509	0	0.864	1.255
	2	0	1	0	1.198
	3	0.861	0	0.509	0.802
N(10')	1	0	1	0	1.997
	2	-0.432	0	0.902	1.060
	3	0.902	0	0.432	0.755
C(11')	1	0	1	0	1.817
	2	-0.543	0	0.840	1.645
	3	0.840	0	0.543	0.741

along the N-H...N hydrogen bond was considered to be possible. The peak attributed to H(2) in the final difference map of the AT base pair is elongated in the direction of N(7') and there is an isolated maximum of 0.30 e.Å⁻³ 1.00 Å from N(7') in a position where the hydrogen could be expected, if attached to this atom. This, however, cannot be considered as significant evidence for a partial proton transfer along the NH...N bond because spurious peaks with similar heights appear elsewhere in the difference map around other atoms.

It is also of interest to note the deviation from linearity of the atoms A-H...B involved in the hydrogen bonds, as indicated in Fig. 5. None of the deviations from linearity of the atoms involved in the hydrogen bonding between the two bases is exceptionally large. Donohue & Trueblood (1960) observed from a study of molecular models that for N-H...O linear, a deviation of about 25° for N-H...N could be expected. Apparently the departures from the assumed dimensions cooperate in making the experimentally found hydrogen bonding system quite satisfactory. The only other hydrogen

bond present in the structure is the one between the NH₂ group of 9-methyladenine and O(8) of the 1-methylthymine molecule, one unit cell up along the direction of the a axis. Its length is 2.881 Å, and there is no large departure from linearity in this hydrogen bond.

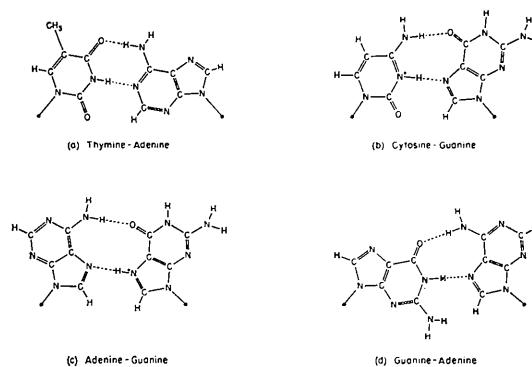


Fig. 6. Some hydrogen-bonded base pairs. For explanation see text.

In Fig. 6(a) is reproduced the adenine-thymine pair as proposed by Watson & Crick (1953). In this model O(9) and N(3) of thymine are hydrogen-bonded to the amino group and to N(1) respectively, of adenine. It is not obvious why in the crystal structure N(3) of thymine is hydrogen-bonded to N(7) and not to N(1) of adenine. Also, the question whether the base pair occurring in the crystal is present in DNA remains unanswered here. We have found it possible to build a DNA molecule from the atomic models of Corey & Pauling (1953) that is similar to the double helix proposed by Watson & Crick, but in which the base pairs are hydrogen-bonded in the configuration occurring in the AT crystals. The distance between the two deoxyribose carbon atoms that are attached to the base pairs in this model is 8.645 Å, much shorter than the corresponding distance of about 11.1 Å in the model proposed by Watson & Crick. The two deoxyribose-phosphate chains at the outside of the molecule have, as a result of this, alternatingly smaller and larger vertical separation than they have in the Watson-Crick model.

The arrangement of a guanine-cytosine pair that has the same distance between the deoxyribose carbon atoms attached to the base pairs as in the adenine-thymine pair is shown in Fig. 6(b). In the tautomeric forms that are accepted for these molecules neither N(3) of cytosine nor N(7) of guanine has a hydrogen atom available for a hydrogen bond between these two atoms. If the configuration occurring in the crystal structure of AT is also present in DNA, there should be added a proton to N(3) of cytosine or to N(7) of guanine in the base pairs of these two molecules. This would be possible if, for example, one of the corresponding nucleotides is incorporated in a zwitterion configuration. The question as to which one might be a zwitterion can presumably be answered on the basis of the assumption that it is primarily the dimensions of the adenine-thymine and guanine-cytosine base pairs which exclude the occurrence of other pairs in the DNA molecule. If deoxyguanylic acid is incorporated in the zwitterion configuration with the proton residing on N(7), the formation of an adenine-guanine pair as shown in Fig. 6(c) would be possible, and the dimensions are such that this pair could replace the adenine-thymine or guanine-cytosine pairs with the configuration occurring in the AT crystals. In this case the specificity of the base pairing would be lost. If, however, deoxycytidilic acid is present in the zwitterion form, no other pairs

can be formed and the specificity is preserved. In this configuration, cytidine would have two donors, guanine two acceptors, adenine and thymidine one donor and one hydrogen acceptor each. Donohue & Trueblood (1960) have already pointed out that it is possible to form a base pair of adenine and guanine as shown in Fig. 6(d). The dimensions of this base pair are such that it presumably could be inserted into the double helix proposed by Watson & Crick, again destroying the specificity.

Numerous attempts have been made to grow crystals containing other combinations of purine and pyrimidine derivatives. Aqueous solutions containing equimolar quantities of all possible pairs of the four compounds: 1-methylthymine, 9-methyladenine, 1-methylcytosine, and 9-methylguanine have so far failed to yield crystals containing hydrogen-bonded pairs other than the thymine-adenine pair. In all other cases only crystals of the separate compounds could be detected.

I should like to thank Prof. R. B. Corey and Dr R. E. Marsh for their continuous interest during this investigation and the reading of the manuscript.

References

- BERGHUIS, J., HAANAPPEL, IJ. M., POTTERS, M., LOOPSTRA, B. O., MACGILLAVRY, C. H., & VEENENDAAL, A. L. (1955). *Acta Cryst.* **8**, 478.
- BOOTH, A. D. (1948). *Fourier Technique in X-ray Organic Structure Analysis*. Cambridge University Press.
- BROOMHEAD, J. M. (1948). *Acta Cryst.* **1**, 324.
- COCHRAN, W. (1951). *Acta Cryst.* **4**, 81.
- COREY, R. B. & PAULING, L. (1953). *Rev. Sci. Instrum.* **24**, 621.
- DONOHUE, J. & TRUEBLOOD, K. N. (1960). *J. Mol. Biol.* **2**, 363.
- FULLER, W. (1959). *J. Phys. Chem.* **63**, 1705.
- HOOGSTEN, K. (1959). *Acta Cryst.* **12**, 822.
- HOOGSTEN, K. (1963). *Acta Cryst.* **16**, 28.
- HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
- MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.
- NOWACKI, W. & BÜRCKI, H. (1955). *Z. Kristallogr.* **106**, 339.
- PAULING, L. & COREY, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164.
- SENKO, M. E. & TEMPLETON, D. H. (1958). *Acta Cryst.* **11**, 808.
- WATSON, J. D. & CRICK, F. H. C. (1953). *Nature, Lond.* **171**, 737.