Acta Cryst. (1961). **14**, 333

The Crystal Structure of Thymine Monohydrate*

BY RAYMOND GERDIL†

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

(*Received* 2 *July* 1959)

Crystals of thymine monohydrate are monoclinic with space group $P2_1/c$; the unit cell dimensions are:

$$a = 6{\cdot}077,\ b = 27{\cdot}862,\ c = 3{\cdot}816\ \text{Å};\ \beta = 94°\ 19'.$$

The crystal structure was refined on the basis of three-dimensional intensity data obtained with copper radiation. The structural parameters for the heavy atoms, including anisotropic temperature factors, were refined by the least-squares method; the positions of the hydrogen atoms were located from a three-dimensional difference synthesis. The final R factor was 0·078 for 1068 observed reflections and the standard deviations in the positional parameters for the heavy atoms are about 0·0023 Å.

A surprising feature is the amount of the water of crystallization, which is about 83% of the theoretical value. An arrangement of the water molecules is proposed which accounts for the observed X-ray data.

Introduction

The pyrimidine thymine is a fundamental component of nucleic acids; a detailed and accurate determination of its structure was therefore undertaken as a part of a program of research on the structure and configuration of polynucleotide chains now in progress at this Institute.

According to Furberg & Hordvik (1956), crystals of thymine grown from alcohol are needle-shaped, elongated along b, whereas parallelepipeds with all faces well developed are obtained from water. Both types are described as monoclinic with the same cell dimensions (see Table 1).

In our experience most of the crystals of thymine grown from alcohol or concentrated aqueous solutions were slightly elongated, well-shaped plates, with a pronounced tendency to form twins or clusters. Some single crystals suitable for preliminary X-ray investigation were found and their cell measurements confirmed the values published by Furberg. However, it was noticed that a second type of large single crystal with well-developed faces separated from dilute aqueous solutions allowed to stand at room temperature for many days.‡

A preliminary rotation photograph, of a crystal of the second type showed a unit translation of 3·8 Å along the principal morphological axis, a value very different from that found for the first type of crystal. The new crystals were further examined and identified as an hydrated compound of thymine.

Experimental

(i) *Unit cell and space group*

Single crystals of thymine monohydrate were grown by slow evaporation at room temperature of dilute aqueous solutions. They were parallelepipeds elongated in the c direction; the well defined forms were {010} and {120}. The crystals rapidly lost their transparency on storage at room temperature; this was followed by spontaneous cleavage parallel to (010). The explanation might be found in a partial or total loss of water of crystallization.

For X-ray examination crystals were mounted on glass fibers with shellac and then placed in thin-walled glass capillaries with a drop of mother liquor. The capillaries were sealed, and under these conditions crystals were kept for many weeks without noticeable change. One single crystal having roughly uniform cross section about 0·35 mm. in diameter was mounted with the c axis parallel to the Weissenberg rotation axis. A fragment obtained by cleavage of a larger crystal was used for recording Weissenberg photographs about the b axis.

The systematic extinctions observed on Weissenberg photographs established uniquely the space group to be $P2_1/c$. The lattice constants a, c, β, and b were fitted by the least-squares method to the spacings of the Cu $K\alpha$ components recorded for 20 $h0l$ reflections and for 22 $hk0$ reflections. The spacings were measured on zero-layer Weissenberg photographs and the films were calibrated by superimposing a powder pattern of NaCl on either edge. The resulting unit-cell dimensions are given in Table 1.

The density observed by flotation in benzene–carbon tetrachloride mixtures at 25 °C. is 1·454 ± 0·002 g.cm.$^{-3}$. The number of molecules per unit cell calculated was found to be 3·93; assuming the correct

---

* Contribution No. 2476 from the Gates and Crellin Laboratories of Chemistry. This work was supported, in part, by a Grant (CVRE No. 121) from The National Foundation.

† Present address: Cyanamid European Research Institute, 91, route de la Capite, Cologny, Geneva, Switzerland.

‡ The two types of crystal habit could be seen in the same mother liquor.

Table 1. *Unit cell dimensions*

Thymine monohydrate* (Cu $K\alpha$) = 1·54178 Å	Thymine (Furberg & Hordvik, 1956)
$a = 6{\cdot}077 \pm 0{\cdot}003$ Å	$a = 12{\cdot}87$ Å
$b = 27{\cdot}862 \pm 0{\cdot}007$	$b = 6{\cdot}83$
$c = 3{\cdot}816 \pm 0{\cdot}002$	$c = 6{\cdot}72$
$\beta = 94° 19' \pm 3{\cdot}1'$	$\beta = 105°$

* The reported confidence limits are twice the standard deviation for each measurement.

number to be 4, the apparent molecular weight of thymine monohydrate is 141·1 compared to a formula weight of 144·1. This discrepancy is presumably greater than the experimental error and indicates that the actual contribution of the water to the space lattice is about 83% of the theoretical value; in other words, approximately one out of every six water molecules is missing. This assumption was later confirmed by the results of the structure determination.

(ii) *Collection of intensity data*

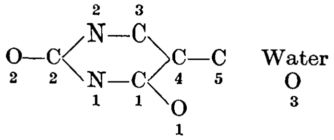
The intensity data were obtained from equi-inclination Weissenberg photographs taken with nickel-filtered Cu $K\alpha$ radiation. The multiple-film technique was used and the intensities were estimated visually by comparison against a calibrated scale prepared from the same crystal. Empirical film factors were obtained for each set of films. The intensities were corrected for Lorentz and polarization factors to obtain relative values of $F^2(hkl)$.

The structure analysis was carried out on the basis of the intensities measured for the layer-lines zero through three about the c axis. 1270 Reflections were covered, of which 1068 were strong enough to be observed. The values for the observed intensities ranged from about 2 to 60,000, the highest value (the $\bar{1}21$ reflection) being far above the next smaller one. The correlation factors to bring all exposures of the various layer lines to the same intensity scale were determined by means of Weissenberg photographs from a crystal rotated about b.

Determination of the atomic positions

(i) *Two-dimensional refinement*

For all further calculations the atoms were labelled as follows:



The shortness of the c axis made it almost certain that the (001) projection would be completely resolved and that the thymine molecule would not be tilted more than about 35° from the (001) plane. Accordingly, the $hk0$ reflections were placed on an absolute scale by Wilson's method and were then used to prepare a Patterson projection down the c axis. The peak at the origin was removed and the function sharpened; in addition the coefficients were modified by the function

$$(2 \sin \theta/\lambda)^4 \exp\left[-(4{\cdot}4 \sin \theta/\lambda)^2\right]$$

(Waser & Schomaker, 1953; Shoemaker, Donohue, Schomaker & Corey, 1950). The unobserved reflections were given half the threshold intensity in these calculations. The resulting Patterson projection is shown in Fig. 1.

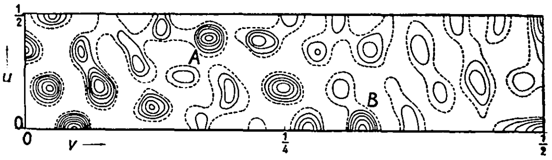
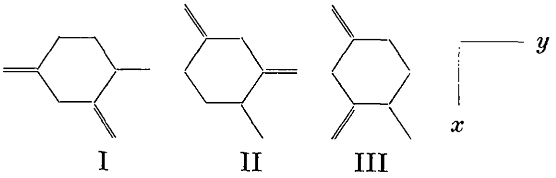


Fig. 1. Sharpened Patterson projection $P(u, v, 0)$. The contours are at equal but arbitrary intervals. The zero contour is dashed; the negative contours are not shown. The origin peak has been removed.

As a first approximation in interpreting the Patterson projection it was assumed that all bonds were about 1·35 Å in length and all bond angles were 120°. The hexagonal pattern of the peaks near the origin (Fig. 1) indicated that the thymine nucleus is oriented with one pair of covalent bonds parallel to b. There are then three different general orientations:



These orientations give three different sets of intramolecular interactions. Orientation III was eliminated by considering the relative heights of the Patterson peaks, but it was not possible to distinguish between I and II on this basis.

With the orientation of the nucleus fixed, an attempt was then made to locate the positions of the molecules relative to the origin (taken at a center of symmetry). On consideration that two molecules connected by hydrogen bonds across a center of symmetry should be separated by a distance of about 5 Å: peak A was regarded as providing a reasonable internuclear distance across the origin. Peak B at a distance of 9 Å from the origin was thought to reflect the intermolecular distance of two molecules related by a glide plane. Such a packing, studied with the aid of molecular models (Corey & Pauling, 1953), provides a satisfactory filling of the unit cell, provided the water molecules are assumed to occupy holes occurring between atoms O_1, C_5, and the glide plane.

Further information about the orientation of the molecule and the location of the water molecule was arrived at in the following manner: the unitary structure factors for six ($0k0$), two ($h00$), and three ($hk0$) planes were derived from the measured intensities. On the assumption that the intermolecular parameters obtained from the Patterson projection were correct, the unitary structure factors for 9 atoms in the postulated orientation I, II (and their mirror images in y) were calculated by means of structure-factor graphs. The objectionable discrepancies arising between the calculated and observed values were assumed to be due to the missing water molecule. For each of the mentioned reflections, those areas covered by the projected water positions which resulted in a minimization of the quantity $|U_o(hk0)-U_c(hk0)|$, were plotted on a single map. The common area outlined by the overlap of the various corrective areas was thought to indicate the most probable position of H_2O. It must be noted that for orientation II (and its mirror-image) and for the mirror-image of I no consistent results could be obtained. However, the orientation I yielded a small common area, the boundaries of which restricted roughly the projection of H_2O onto (001) to $x=0{\cdot}85$ and $y=0{\cdot}21$. On the further assumption that the flat molecule lay strictly parallel to (001), this trial structure was used in a structure-factor calculation. With an over-all thermal parameter $B=3{\cdot}4$ Å^2, as previously determined by the Wilson's method, the resulting reliability factor was 46%.

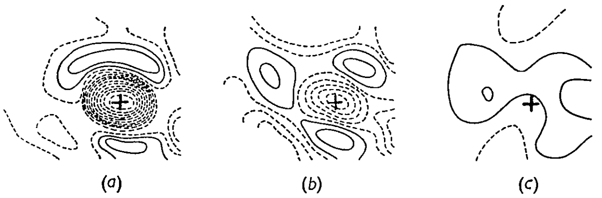


Fig. 2. A portion of the (001) difference map around O_3 at three stages of refinement. In the calculations, the contribution of the water was taken to be: (a) 100%; (b) 87·5%; (c) 83% of the theoretical value. Contours are at intervals of $\frac{1}{4}$ e.Å^{-3}; the zero and negative contours are dashed.

A first Fourier projection was calculated from 44 low order $F(hk0)$'s for which the above trial structure yielded reasonable values; although at this stage all atoms were not expected to be well resolved, the location of the water molecule was obviously confirmed. After five cycles of least-squares refinement (with all reflections weighted equally) the reliability factor dropped to 22%. A difference-synthesis computed at this stage showed that the temperature parameter had been overestimated. Accordingly, B was reduced to $2{\cdot}04$ Å^2; the R factor in the next least-squares cycle was 18%. A further least-squares cycle yielded no significant improvement. An $hk0$ difference synthesis calculated at this stage showed a large negative peak at the position of the water molecule, which indicated that the oxygen atom had been overweighted (Fig. 2(a)); therefore, the oxygen atomic scattering curve for O_3 was temporarily replaced by a nitrogen atomic scattering curve. Moreover, use was made of the weighting system of Hughes (1941)—with zero weight for the unobserved reflections—in all subsequent least-squares calculations. The agreement improved considerably after a further cycle of least-squares refinement, and the R factor dropped to 13·6%. A difference projection (Fig. 2(b)) now showed that the weight assigned to O_3 was still overestimated. (Yet, the use of the atomic form factor of carbon to represent the oxygen atom did not improve the agreement in a recalculation of the structure factors.) The information about the thermal movements was used to assign an isotropic temperature factor to each atom of the asymmetric unit. This resulted in a significant improvement for the structure-factor agreement, with $R=12{\cdot}4\%$. At this stage the attention was turned to the determination of the z parameters.

(ii) *Determination of the z parameters*

The outstandingly strong $\overline{1}21$ reflection, having a structure amplitude of more than $\frac{1}{3}$ of the maximum possible, suggested that the orientation of the molecules was close to that of the ($\overline{1}21$) plane. Moreover, if it is assumed that two molecules of thymine related by a center of symmetry are hydrogen bonded together, then these molecules are coplanar and the common plane is required to pass through the origin. Critical examination of the covalent bond lengths in the (001) Fourier projection showed that an apparently high degree of parallelism of the molecular plane was preserved with respect to b.* Hence it was legitimate to assume that the Fourier peaks on (010) would be evenly distributed close to a line whose direction would be determined by the prevalent orientation of the intramolecular vectors projected on to (010). Thus a sharpened Patterson projection down the b axis was calculated using the 25 observed $h0l$ reflections. The origin peak was removed and the coefficients were multiplied by the modification function

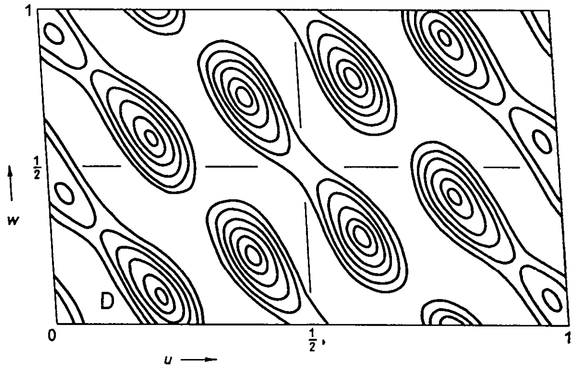


Fig. 3. Sharpened Patterson projection $P(u, 0, w)$. Contours at equal but arbitrary intervals, beginning with the zero contour; negative contours are not shown. The origin peak has been removed.

---

* Due to the length of the b axis this did not seem to involve any contradiction; even for a plane close to ($\overline{1}21$).

already mentioned. The function $P(u, 0, w)$ is shown in Fig. 3. The scarcity and the shape of the peaks indicated that, as expected, most of the vectors overlapped to a great extent and that vectors of different nature had their length equal in projection. The peak D about 1·3 Å from the origin was characteristic of the interaction between pairs of bonded atoms. Furthermore, from the structural features already asserted and from the symmetry inherent to the space group it was deduced that, in a direction conditioned by the orientation of the molecular plane, the length of the interatomic vectors projected upon (010) would be roughly equal to or a multiple of the distance between pairs of bonded atoms. Consequently, the linear arrangement of the peaks at constant spacing about 1·3 Å in the direction [2,0,1] was selected as providing structural information consistant with our criterion. Accordingly, this result was used to derive the following linear relations $z_i = 0{\cdot}5x_i$ between the unknown z and the known x parameters. In spite of its extremely simplified form—as evident in the elongation of the peaks in the $[\bar{1},0,2]$ direction—this relation proved to be a straightforward step to the determination of the z atomic parameters. A first set of z values defined as above was used in a structure-factor calculation including the $(hk1)$ and $(\bar{h}k1)$ planes only. The poor agreement between the observed and the calculated structure factors ($R = 61\%$) reflected the inaccuracy of the analytical relation; but the underlying assumption was sufficiently correct to result in the convergence of the reliability factor in the course of the next least-squares refinements. In each of the successive computations the x and y parameters were given the values obtained from the two-dimensional refinement and after six cycles R had dropped to 21%.

Three-dimensional refinement of the positional and thermal parameters

Like most of the preceding Fourier summations, the three-dimensional structure-factor and least-squares calculations were carried out on a Datatron digital computer. A program including individual anisotropic temperature factors in the structure-factor routine, and permitting least-squares adjustments on them, was used in the three-dimensional refinement. The temperature factor for one atom was expressed in the form

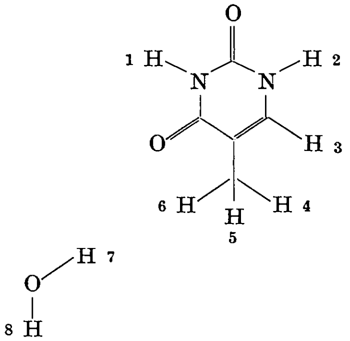
$$T_i = \exp -\{\alpha_i h^2 + \beta_i k^2 + \gamma_i l^2 + \delta_i hk + \varepsilon_i hl + \eta_i kl\} .$$

In addition, an independent scale factor, k_i, associated with F_c as required by the least-squares procedure, was allocated to each atom. The seven normal equations were solved to give six anisotropic shifts and one scale-factor shift per atom in addition to the positional corrections. Finally, an average of the independent scale factors over the number of atoms was taken to reduce the number of unknowns for each atom to six; the six normal equations involving only the temperature factors were then solved for the final temperature parameters.

Preliminary values for the constants γ and ε were obtained by an analytical process similar to the Wilson's method and introduced in a first cycle of least-squares refinement, making use of the zero and second layer only. Values for α and β were assigned on the basis of the known average isotropic B parameter; but the remaining δ and η coefficients were given zero value in the preliminary computation. The resulting R factor was 18·8%; the average change in the positional parameters x and y was about 0·008 Å and the average shift Δz was about twice as large. This difference was ascribed to the fact that the former had already been refined to a great extent in projection. An interesting feature was that the individual scale factor assigned to the water molecule came out 96% of the average value taken over the nine other atoms. The atomic scattering curve of O_3 had already been multiplied by a factor 0·875 (i.e., 7 instead of 8 electrons) in the structure factor calculations, and yet it appeared that the weight of the water molecule was still overestimated. The new 'weighting factor' of 0·84 was a satisfactory confirmation of the value 0·83 deduced from the unit-cell dimensions and the measured density. Accordingly, the atomic scattering curve of O_3 was multiplied by a factor 0·83 in all the subsequent structure-factor calculations and the individual scale factors of O_3 remained in line with the average value for the other atoms. In addition, this correction made the $(\varrho_o - \varrho_c)$-function approximately zero at the center of the atom O_3 in the difference maps calculated during the next steps of the refinement (see Fig. 2(c)).

The next least-squares computations were based on the complete set of three-dimensional data and after five cycles the R factor was reduced to 10·8%. The average shifts were of the order of 0·004 Å for the positional parameters. At this point of the investigation a search was made for the positions and the contribution of the hydrogen atoms.

The hydrogen atoms were labelled as follows:



A difference map projected on to (001) gave a clear indication of the positions of the three hydrogen atoms attached to the thymine nucleus. A region of high electron density corresponding to the location of the hydrogen atoms of the methyl group indicated that two of them overlapped partially. The peaks associated with the hydrogen atoms of the water molecule were not very distinctly resolved in the two-dimensional projection; nevertheless, the flattened peaks situated on both sides of O_3, on a line parallel to *b* (Fig. 2(*c*)), were tentatively identified as hydrogen atoms. The values of the *y* parameters of the hydrogen atoms determined from the two-dimensional map were then used in the three-dimensional computation of appropriate difference map sections perpendicular to the *b* axis. The positions of the eight hydrogen atoms were clearly apparent with the restriction that the peak for H_5 appeared to be anomalously small. A composite drawing of this three-dimensional difference map is shown in Fig. 4 and the hydrogen atom parameters are listed in Table 2.

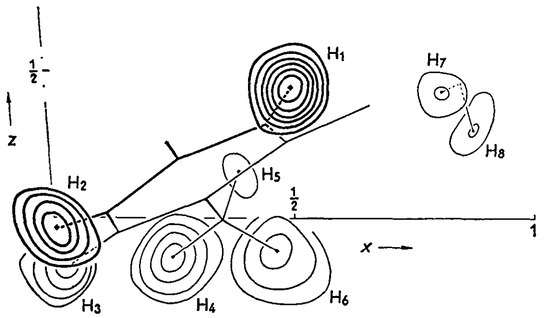


Fig. 4. A composite drawing of the electron density associated with the hydrogen atoms. The contours are at intervals of 0·1 e.Å^{-3}, beginning with the 0·3 e.Å^{-3} contour.

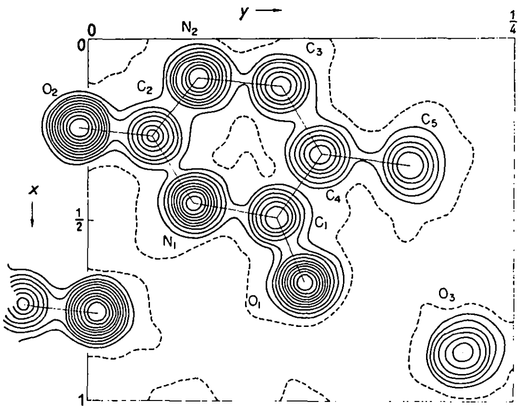


Fig. 5. Final (001) Fourier projection. The contours are drawn at intervals of 1 e.Å^{-2}; the dashed lines represent the 1 e.Å^{-2} contour.

A new set of structure factors and a least-squares refinement were then calculated; the hydrogen atom parameters of Table 2 were included and an isotropic thermal parameter, assumed equal to that of the adjacent heavy atom, was assigned to each hydrogen atom. The *R* factor was reduced from 10·8 to 8·5%.

Table 2. *The hydrogen-atom parameters*

Atom	*x*	*y*	*z*
H_1	0·532	0·044	0·430
H_2	0·003	0·049	−0·035
H_3	0·016	0·124	−0·157
H_4	0·242	0·210	−0·127
H_5	0·392	0·210	0·162
H_6	0·452	0·196	−0·109
H_7	0·824	0·197	0·424
H_8	0·885	0·237	0·302

Two subsequent cycles of least-squares refinement indicated shifts in the hydrogen atom parameters which were clearly oscillating with an order of magnitude of 0·06 Å about the values derived from the difference map; therefore, only the positions derived from the difference synthesis were retained. The final set of structure factors led to an *R* factor of 7·8%, an improvement which is to be mainly associated with the further adjustment of the anisotropic thermal parameters. The maximum positional parameter shift for the heavy atoms was 0·0039 Å and the average shift was 0·0012 Å. The final positional parameters, which include the shifts indicated in the last least-squares refinement, are listed with their standard deviations in Table 4. The final anisotropic thermal parameters are listed in Table 5.

Discussion of the structure

(i) *The molecular arrangement*

The molecular arrangement in the crystal is illustrated by the clinographic projection in Fig. 6. The main features of the molecular packing may be visualized in terms of stacks extending indefinitely in the *a* and *c* directions. Adjacent stacks lie parallel to and are related by the glide planes. The stacks consist of narrow ribbons (a portion of a 'ribbon' is depicted by the sequence of molecules M_1, M_2, M_3, and M_4 in Fig. 6), approximately two molecules in width, which are inclined about 30° to (001) and superimposed at one unit translation in the *c* direction. Within a ribbon, each molecule is hydrogen-bonded to two adjacent molecules across two different centers of symmetry; these hydrogen bonds involve the two N–H groups in the ring and the single oxygen atom O_2. The other oxygen atom O_1, accepts only a single hydrogen bond from a water molecule. Fig. 7 schematically shows a portion of the hydrogen-bond network, with hydrogen-bond distances and angles. The best plane of the ten atoms C_2, N_1, N_2, O_2, C_2', N_2', O_2', C_2'', N_2'', O_2'' is defined by the equation

$$-3·2824x+5·7695y+3·2595z=0·0066;$$

the distance from this plane to the origin is 0·0066 Å and the average perpendicular deviation of the atoms from the plane is 0·04 Å. The lengths of the two N–H $\cdots$ O bonds are equal within experimental error, the average value being 2·840 Å; the average length of the hydrogen bond between the water molecule and

Table 3. *Observed and calculated structure factors*

The three columns in each group contain the values, reading from left to right of k, $10F_o$, and $10F_c$

0,k,0		
4	261	-251
6	76	-90
8	214	192
10	142	137
12	508	-514
14	230	228
16	490	472
18	254	248
20	55	-44
22	21	18
24	58	-58
26	59	50
28	94	103
30	75	76
32	106	99
34	< 12	-12

1,k,0		
0	294	-303
1	185	-204
2	95	-89
3	414	-436
4	327	326
5	116	-125
6	131	-129
7	296	-308
8	< 10	-3
9	167	155
10	17	18
11	140	138
12	110	104
13	83	62
14	96	95
15	238	238
16	176	-179
17	15	15
18	87	91
19	167	181
20	37	-33
21	29	-28
22	18	-17
23	26	23
24	30	32
25	160	180
26	< 19	7
27	95	99
28	21	23
29	< 17	15
30	23	-30
31	< 15	-20
32	14	-1
33	< 24	28
34	11	9
35	24	18

2,k,0		
0	320	-280
1	176	187
2	98	-84
3	456	-480
4	244	-239
5	116	117
6	465	-472
7	267	274
8	371	-367
9	120	-127
10	118	-107
11	< 13	16
12	22	15
13	< 13	-13
14	89	-91
15	150	152
16	26	-23
17	49	-50
18	120	119
19	115	-116
20	< 18	3
21	65	-60
22	197	-208
23	113	117
24	161	-161
25	< 19	-10
26	65	-66
27	23	-18
28	< 17	-17
29	23	29
30	40	-42
31	< 14	-10
32	24	-26
33	32	28
34	38	34

3,k,0		
0	77	-64
1	73	-66
2	114	-102
3	263	-257
4	161	-152
5	87	-74
6	302	299
7	29	15
8	126	114
9	141	135
10	111	-100
11	110	97
12	104	-99
13	20	19
14	< 16	3
15	66	74
16	73	76
17	151	-154
18	270	-288
19	185	-204
20	40	-41
21	190	-210
22	160	170
23	81	-78
24	88	99
25	25	23
26	35	-33
27	< 16	-5
28	67	-71
29	21	23
30	35	32
31	< 12	9
32	24	19

4,k,0		
0	26	-23
1	< 16	7
2	242	242
3	117	96
4	152	145
5	15	-6
6	134	127
7	105	-90
8	111	105
9	17	-10
10	52	37
11	< 17	8
12	87	87
13	51	-48
14	59	-55
15	66	-64
16	34	-30
17	22	11
18	30	30
19	102	100
20	89	93
21	48	51
22	36	34
23	48	-50
24	67	73
25	22	22
26	51	56
27	< 13	6
28	48	43
29	47	-40
30	37	30

5,k,0		
0	261	-250
1	33	-36
2	49	-47
3	49	-49
4	142	136
5	92	-94
6	78	78
7	< 18	-11
8	40	-41
9	47	-40
10	< 19	29
11	40	40
12	107	116
13	102	106
14	< 18	4
15	< 18	17
16	48	-54
17	48	57
18	35	-37
19	< 17	-6
20	30	34
21	55	60
22	37	39
23	57	47
24	31	27
25	< 12	-3
26	< 10	4
27	44	45

6,k,0		
0	37	32
1	43	43
2	110	-117
3	146	154
4	169	-176
5	57	60
6	< 18	-23
7	< 18	6
8	77	87
9	25	17
10	27	21
11	37	-41
12	73	-88
13	43	-60
14	39	-40
15	44	-61
16	24	24
17	< 14	13
18	29	-31
19	49	42
20	50	-49
21	58	55
22	< 10	-1

7,k,0		
0	21	23
1	< 14	3
2	24	25
3	26	29
4	22	19
5	16	17
6	31	-29
7	< 13	-6
8	24	-26
9	17	11
10	43	-44
11	< 12	14
12	26	-26
13	< 10	-3
14	< 10	1
15	35	-38

0,k,1		
1	267	308
2	545	-567
3	108	-113
4	301	-292
5	110	-88
6	15	2
7	15	-22
8	21	-1
9	214	-208
10	156	-144
11	133	132
12	206	190
13	157	-153
14	264	258
15	87	80
16	127	120
17	34	24
18	< 13	3
19	41	-36
20	21	26
21	85	82
22	83	82
23	127	-137
24	59	60
25	17	-16
26	82	89
27	61	62
28	72	78
29	34	40
30	33	34
31	< 11	-9
32	< 11	-9
33	30	-26
34	< 9	-2

1,k,1		
1	85	-57
2	234	-225
3	285	-271
4	332	342
5	657	-664
6	120	-122
7	163	-145
8	39	39
9	88	-80
10	144	-142
11	186	-175
12	41	40
13	146	-142
14	54	52
15	23	26
16	105	-97
17	149	141
18	48	-46
19	39	-36
20	119	119
21	185	-190
22	60	67
23	113	-114
24	46	-46
25	34	-33
26	< 14	11
27	93	-96
28	29	-33
29	54	-50
30	23	26
31	15	-15
32	48	-38
33	25	22
34	25	-22

2,k,1		
1	< 10	0
2	74	-73
3	36	30
4	378	-365
5	15	-3
6	232	-223
7	139	-117
8	340	329
9	37	31
10	168	141
11	13	-17
12	92	93
13	58	47
14	118	-117
15	32	-33
16	100	101
17	33	33
18	87	-80
19	68	61
20	452	-455
21	35	-39
22	164	-172
23	32	-27
24	28	23
25	17	-14
26	92	94
27	13	7
28	22	-21
29	< 12	-6
30	54	-44
31	< 11	-5
32	54	42
33	< 8	7

3,k,1		
1	115	116
2	106	-100
3	117	106
4	81	73
5	438	415
6	37	32
7	134	108
8	< 13	3
9	110	105
10	75	-73
11	59	56
12	18	20
13	68	62
14	61	54
15	41	-33
16	64	-60
17	104	-96
18	43	-42
19	< 14	-2
20	33	32
21	157	168
22	50	48
23	114	125
24	19	-19
25	20	22
26	23	-22
27	110	100
28	< 11	-3
29	44	31
30	17	13
31	14	16

4,k,1		
1	20	-16
2	241	-232
3	< 13	-5
4	129	-124
5	23	22
6	47	45
7	34	28
8	37	-38
9	74	68
10	< 14	-13
11	99	-86
12	23	25
13	43	45
14	153	165
15	26	20
16	65	70
17	< 14	-0
18	< 14	2
19	< 14	-3
20	39	40
21	33	-33
22	12	16
23	99	86
24	67	62
25	< 11	-1
26	90	78
27	65	-55
28	57	52
29	14	-19

5,k,1		
1	275	-265
2	261	252
3	112	-110
4	24	23
5	18	14
6	72	-69
7	< 17	-10
8	46	-39
9	25	23
10	98	106
11	24	29
12	< 14	-11
13	26	-25
14	113	-119
15	88	-91
16	19	-19
17	102	-106
18	86	83
19	33	-40
20	34	32
21	11	-14
22	< 11	0
23	30	25
24	10	-3
25	11	14

6,k,1		
1	34	31
2	17	15
3	< 14	0
4	58	57
5	< 14	0
6	71	70
7	24	-23
8	< 13	6
9	62	-64
10	13	-10
11	< 12	-11
12	53	-48
13	11	-14
14	12	-17
15	21	16
16	42	-40
17	20	-9
18	50	-44
19	< 9	-2
20	10	-8

7,k,1		
1	59	59
2	11	18
3	37	35
4	45	40
5	< 10	8
6	< 10	1
7	95	93
8	61	-55
9	65	67
10	30	-32
11	< 8	-1
12	14	14

$\bar{1}$,k,1		
1	648	635
2	1060	1039
3	652	629
4	155	140
5	114	-114
6	184	-167
7	80	81
8	90	-92
9	155	-157
10	290	297
11	26	16
12	60	-49
13	98	90
14	361	-369
15	125	119
16	48	-38
17	172	174
18	156	168
19	141	141
20	66	64
21	21	17
22	14	13
23	35	-44
24	< 15	-14
25	32	36
26	66	-76
27	33	36
28	< 14	-12
29	90	104
30	12	-10
31	35	39
32	37	37
33	36	40
34	< 8	7

$\bar{2}$,k,1		
1	149	-125
2	44	-46
3	< 9	-13
4	146	158
5	58	-52
6	182	166
7	144	140
8	89	-64
9	104	106
10	45	40
11	56	52
12	161	-164
13	< 9	10
14	21	30
15	25	27
16	102	-109
17	37	42
18	107	-106
19	43	-45
20	18	20
21	43	-44
22	126	-133
23	51	55
24	56	-61
25	54	60
26	43	-53
27	38	-42
28	23	-25
29	35	-43
30	12	-12
31	14	15
32	25	-21
33	18	18

$\bar{3}$,k,1		
1	146	-140
2	86	86
3	206	-187
4	101	96
5	21	23
6	122	133
7	233	-253
8	179	-189
9	195	-191
10	111	-124
11	50	-63
12	< 13	-1
13	36	31
14	< 13	2
15	39	37
16	77	-86
17	44	-49
18	23	25
19	45	-43
20	150	170
21	27	26
22	76	86
23	64	-70
24	48	-65
25	112	-128
26	84	-96
27	36	-38
28	12	17
29	30	-32
30	< 11	-3
31	13	10
32	33	-30

$\bar{4}$,k,1		
1	167	-166
2	101	94
3	13	12
4	88	-70
5	132	129
6	36	-36
7	64	64
8	61	56
9	57	45
10	< 14	17
11	69	-64
12	103	-104
13	15	20
14	54	-62
15	107	-124
16	142	164
17	138	-163
18	118	129
19	32	-38
20	29	-27
21	32	38
22	28	26
23	102	119
24	95	109
25	30	32
26	41	49
27	< 11	2
28	45	-41
29	11	6
30	32	-36

$\bar{5}$,k,1		
1	100	101
2	40	-36
3	64	67
4	51	53
5	55	55
6	< 14	-3
7	32	33
8	33	37
9	67	64
10	36	32
11	< 15	0
12	29	30
13	20	-21
14	27	33
15	< 14	4
16	26	-29
17	41	48
18	44	-54
19	23	28
20	< 13	-10
21	12	-10
22	29	-27
23	41	43
24	17	-18
25	47	46
26	11	14

$\bar{6}$,k,1		
1	54	-64
2	101	-107
3	45	-47
4	< 14	-6
5	34	36
6	96	109
7	47	53
8	43	42
9	54	59
10	74	-85
11	42	58
12	13	-11
13	< 13	1
14	22	23
15	< 12	-8
16	< 11	-4
17	23	-22
18	51	-48
19	< 10	-4
20	59	-57
21	33	31
22	< 7	-9

$\bar{7}$,k,1		
1	< 11	0
2	14	18
3	33	-34
4	67	-69
5	70	-73
6	114	-113
7	< 11	-11
8	< 11	-15
9	< 11	-3
10	55	58
11	21	-20
12	39	39
13	< 9	-9
14	< 8	-5
15	< 6	0

0,k,2		
1	130	138
2	248	-253
3	94	90
4	202	-184
5	40	-29
6	86	-72
7	138	-118
8	160	-147
9	37	-28
10	67	-67
11	68	62
12	88	-89
13	58	-60
14	27	26
15	91	-90
16	41	46
17	21	20
18	82	-90
19	109	120
20	61	-70
21	36	37
22	< 15	-3
23	15	-15
24	66	-77
25	< 14	5
26	39	-41
27	< 13	17
28	30	-40
29	23	-26
30	24	-26
31	45	-51
32	12	11

1,k,2		
0	69	54
1	120	-121
2	145	142
3	178	-174
4	25	16
5	41	31
6	260	-236
7	< 9	-13
8	156	-130
9	67	59
10	95	81
11	55	47
12	90	89
13	47	51
14	29	-31
15	59	58
16	24	27
17	173	-191
18	217	236
19	118	-120
20	48	52
21	91	-107
22	150	-177
23	73	-86
24	89	-106
25	< 13	-3
26	18	19
27	< 12	-4
28	48	48
29	22	24
30	26	-26
31	< 8	6
32	7	-11

2,k,2		
0	103	95
1	31	28
2	139	138
3	141	-137
4	94	78
5	48	-35
6	213	198
7	133	124
8	218	222
9	25	18
10	28	32
11	25	22
12	28	-22
13	< 13	12
14	21	18
15	53	53
16	< 14	1
17	< 14	8
18	50	-50
19	112	-131
20	15	13
21	70	-79
22	93	102
23	59	65
24	132	144
25	13	10
26	42	44
27	< 12	-3
28	< 10	12
29	12	14
30	28	33
31	11	17

3,k,2		
0	136	151
1	39	-39
2	37	35
3	67	-66
4	89	-93
5	92	-93
6	27	27
7	51	-51
8	< 13	-0
9	52	42
10	< 13	-9
11	31	36
12	37	-50
13	< 14	17
14	43	-45
15	53	60
16	61	77
17	74	81
18	25	-29
19	48	58
20	22	25
21	< 13	-4
22	14	17
23	27	28
24	52	-46
25	102	105
26	< 10	1
27	50	55
28	8	10
29	< 6	-10

4,k,2		
0	434	-463
1	83	-88
2	147	-169
3	55	-55
4	68	72
5	< 14	5
6	18	15
7	< 14	-0
8	60	-65
9	15	-17
10	20	-16
11	15	-10
12	79	98
13	44	51
14	38	-45
15	25	31
16	162	-168
17	23	-26
18	86	-89
19	< 13	-11
20	< 12	2
21	13	14
22	22	23
23	23	-26
24	25	27
25	10	15
26	23	-29

5,k,2		
0	15	23
1	52	55
2	< 14	-12
3	81	88
4	30	-36
5	40	45
6	< 14	-5
7	61	60
8	35	31
9	< 13	-9
10	< 13	-3
11	39	-41
12	24	-28
13	50	-55
14	< 12	-1
15	65	-67
16	18	17
17	< 11	-1
18	< 10	1
19	43	-43
20	< 9	-9
21	10	-10
22	< 6	6

6,k,2		
0	61	60
1	26	25
2	24	24
3	60	51
4	67	67
5	29	-31
6	99	101
7	53	-54
8	52	53
9	24	26
10	< 10	5
11	25	29
12	< 9	-9
13	11	-12
14	21	21
15	27	-31

$\bar{1}$,k,2		
0	440	403
1	70	83
2	156	142
3	32	-32
4	213	-176
5	205	-161
6	192	-167
7	70	-63
8	73	65
9	< 10	-9
10	112	-102
11	85	83
12	182	-171
13	74	66
14	33	37
15	23	26
16	65	63
17	91	92
18	74	81
19	15	15
20	50	-57
21	47	49
22	53	-59
23	62	71
24	< 14	-12
25	< 14	-4
26	18	-19
27	27	31
28	32	37
29	53	53
30	34	40
31	25	30
32	< 7	6

$\bar{2}$,k,2		
0	137	-110
1	30	15
2	178	150
3	292	267
4	424	376
5	241	204
6	73	67
7	24	23
8	163	-156
9	22	23
10	15	-9
11	91	-93
12	162	160
13	67	-63
14	62	54
15	64	-57
16	62	-58
17	51	-51
18	38	38
19	64	69
20	45	50
21	85	93
22	36	38
23	19	-25
24	33	38
25	32	-42
26	21	29
27	57	-56
28	11	15
29	< 10	-4
30	11	15
31	25	29

$\bar{3}$,k,2		
0	11	9
1	12	-0
2	79	-72
3	41	39
4	39	-33
5	< 12	-4
6	63	54
7	< 12	1
8	19	18
9	< 13	-1
10	69	66
11	19	14
12	40	34
13	< 14	0
14	19	23
15	68	-76
16	43	47
17	< 14	-3
18	30	-32
19	14	-12
20	18	23
21	51	-57
22	22	25
23	68	-70
24	33	-33
25	22	-26
26	< 12	-8
27	11	19
28	< 9	-5
29	19	-22
30	< 6	-3

$\bar{4}$,k,2		
0	120	-112
1	47	40
2	68	-63
3	38	30
4	59	-60
5	14	20
6	14	5
7	25	25
8	14	11
9	61	-62
10	109	-107
11	89	-84
12	25	-32
13	59	-66
14	60	64
15	< 14	3
16	< 14	-7
17	66	66
18	71	-78
19	24	28
20	31	-35
21	50	49
22	34	39
23	54	54
24	25	30
25	14	-15
26	72	-77
27	38	-45

$\bar{5}$,k,2		
0	86	-73
1	68	-62
2	< 15	2
3	79	-67
4	42	38
5	< 15	-5
6	< 15	5
7	15	12
8	< 14	8
9	19	21
10	< 12	49
11	33	-33
12	< 14	-1
13	< 14	13
14	125	-125
15	86	79
16	86	-89
17	12	10
18	13	11
19	17	-16
20	20	21
21	10	-11
22	12	-12
23	21	26
24	19	23

$\bar{6}$,k,2		
0	36	36
1	< 13	8
2	19	18
3	< 13	7
4	32	31
5	35	31
6	36	32
7	25	30
8	< 12	-6
9	< 12	7
10	15	4
11	22	25
12	< 11	4
13	< 11	11
14	< 10	11
15	21	-24
16	20	20
17	18	-16
18	< 7	-7
19	< 6	-6

$\bar{7}$,k,2		
0	20	21
1	< 9	-4
2	21	-21
3	25	-25
4	62	-70
5	21	-20
6	< 8	1
7	< 8	4
8	83	92
9	10	12
10	30	38

0,k,3		
1	108	131
2	8	-3
3	64	-65
4	54	48
5	87	-75
6	35	27
7	45	-40
8	67	-60
9	15	-18
10	27	-22
11	13	7
12	93	91
13	12	-5
14	< 8	3
15	114	123
16	124	-124
17	104	114
18	57	-61
19	< 10	-2
20	39	42
21	13	-15
22	< 8	1
23	69	-68
24	76	-86
25	46	-48
26	45	-46
27	5	-9
28	25	31

Table 3 (*cont.*)

1,k,3		
1	77	75
2	63	-51
3	101	90
4	37	-28
5	30	31
6	51	-53
7	49	40
8	81	76
9	84	86
10	97	103
11	40	41
12	< 8	0
13	26	-21
14	29	26
15	43	-43
16	19	19
17	40	43
18	43	-49
19	48	53
20	89	-85
21	20	-25
22	76	-80
23	20	20
24	25	27
25	67	71
26	55	64
27	34	45

2,k,3		
1	43	42
2	43	42
3	40	37
4	59	-57
5	16	-13
6	37	-41
7	34	-30
8	< 8	-3
9	25	-21
10	< 10	6
11	24	-32
12	31	32
13	10	-8
14	< 10	-7
15	36	-38
16	39	44
17	17	-15
18	43	43
19	41	40
20	< 8	1
21	< 8	6
22	71	69
23	32	-25
24	42	45
25	15	-16
26	< 4	4

3,k,3		
1	81	-86
2	182	-197
3	131	-143
4	85	-92
5	19	-24
6	45	49
7	22	18
8	12	6
9	13	16
10	52	-57
11	17	-16
12	18	19
13	43	-48
14	92	105
15	< 8	-4
16	47	31
17	43	-42
18	47	-47
19	67	-65
20	48	-48
21	14	16
22	19	-18
23	13	11

4,k,3		
1	74	-85
2	66	68
3	14	16
4	69	71
5	19	19
6	< 8	3
7	12	16
8	8	7
9	28	27
10	22	24
11	25	27
12	48	-48
13	20	23
14	60	-55
15	30	-33
16	37	-38
17	14	-16
18	8	-7
19	10	10
20	< 4	5

5,k,3		
1	10	11
2	10	14
3	78	81
4	37	-34
5	89	90
6	12	10
7	25	26
8	31	28
9	8	-10
10	13	14
11	17	21
12	5	-5
13	30	31
14	8	-12

$\bar{1}$,k,3		
1	89	-94
2	29	32
3	28	-24
4	53	-46
5	50	-48
6	30	-28
7	30	-26
8	10	-8
9	57	-55
10	32	-26
11	36	-34
12	22	-22
13	46	45
14	31	-33
15	14	-10
16	41	39
17	53	-55
18	40	43
19	17	-14
20	11	-10
21	< 10	-6
22	37	33
23	12	-14
24	20	21
25	29	-32
26	10	-7
27	18	-21
28	5	-7

$\bar{2}$,k,3		
1	59	49
2	130	111
3	100	89
4	66	52
5	14	2
6	127	-117
7	102	-94
8	34	-34
9	97	-93
10	86	80
11	30	-27
12	< 10	-12
13	< 11	-2
14	29	-26
15	11	-12
16	< 11	-3
17	18	21
18	40	47
19	15	17
20	73	77
21	23	-26
22	< 10	6
23	17	-18
24	< 8	3
25	< 8	-0
26	< 7	2
27	< 5	-1

$\bar{3}$,k,3		
1	< 10	-2
2	43	-40
3	< 10	10
4	68	67
5	89	88
6	156	160
7	47	41
8	30	29
9	< 11	8
10	78	-82
11	23	24
12	26	-27
13	< 11	4
14	< 11	5
15	21	21
16	14	-16
17	23	-25
18	< 10	-8
19	39	-39
20	13	9
21	39	41
22	< 8	4
23	44	47
24	12	-15
25	21	24

$\bar{4}$,k,3		
1	24	-24
2	18	22
3	< 11	-7
4	15	-14
5	15	-11
6	< 11	-0
7	20	20
8	< 11	-4
9	13	10
10	< 11	-2
11	< 11	-0
12	< 11	4
13	49	55
14	11	-12
15	17	24
16	< 10	3
17	< 9	-9
18	14	-11
19	15	16
20	25	-24
21	13	12
22	6	-1

$\bar{5}$,k,3		
1	54	-52
2	13	-13
3	10	-7
4	< 10	-1
5	13	8
6	< 10	-11
7	< 10	-2
8	17	18
9	11	5
10	18	-19
11	14	11
12	75	-80
13	15	-16
14	< 8	-7
15	33	-37
16	47	53
17	14	-14
18	23	29

$\bar{6}$,k,3		
1	26	26
2	25	-29
3	12	-11
4	18	-15
5	7	-7
6	< 7	6
7	< 7	4
8	14	-15
9	17	20
10	< 6	-3
11	< 5	-5

Table 4. *Final positional parameters and their standard deviations*

Atom	x	y	z	σ_x	σ_y	σ_z
C_1	0·4934	0·11049	0·2602	0·0024 Å	0·0024 Å	0·0026 Å
C_2	0·2666	0·03762	0·2003	0·0023	0·0023	0·0027
C_3	0·1330	0·11262	−0·0431	0·0026	0·0024	0·0028
C_4	0·3176	0·13674	0·0685	0·0024	0·0024	0·0026
C_5	0·3492	0·18920	−0·0040	0·0032	0·0026	0·0033
N_1	0·4519	0·06210	0·3158	0·0019	0·0019	0·0023
N_2	0·1087	0·06419	0·0216	0·0021	0·0021	0·0024
O_1	0·6707	0·12724	0·3808	0·0018	0·0020	0·0023
O_2	0·2446	−0·00565	0·2570	0·0018	0·0018	0·0021
O_3	0·8665	0·21960	0·4554	0·0040	0·0035	0·0058

Table 5. *Final anisotropic thermal parameters*

Atom	$10^4\alpha$	$10^4\beta$	$10^4\gamma$	$10^4\delta$	$10^4\varepsilon$	$10^4\eta$
C_1	75	3·2	351	1	−57	−4
C_2	51	3·0	391	5	−140	−7
C_3	65	3·6	357	10	−133	18
C_4	98	3·2	249	2	−13	10
C_5	222	3·2	513	1	−71	24
N_1	58	2·6	545	6	−223	−8
N_2	60	3·2	479	3	−236	6
O_1	75	3·7	718	−16	−249	0
O_2	104	2·9	785	−5	−394	15
O_3	238	4·7	2743	−33	−438	−13

O_1 is also 2·84 Å. The water molecules are hydrogen bonded together across the glide planes, at an apparent O–H · · · O distance of 2·55 Å, and thus form zig-zag chains, which run parallel to the c axis. The marginal hydrogen atoms at the vertices of the chains form hydrogen bonds with the atoms O_1 of successive thymine molecules: the water molecules are therefore essentially responsible for the packing of the molecular stacks in the b direction. The apparent H–O–H bond angle of water as determined from the coordinates of atoms H_7 and H_8 (Table 2) was found to be 121°. This magnitude is very close to the value of 123° for the angle measured between the average directions of the hydrogen bonds carbonyl–water and water–water. No particular significance can be assigned to the C=O · · · H–O hydrogen-bond length and to the short (2·55 Å) water–water approach since they are only average distances associated with the random disorder of the water molecules (this particular case is discussed below in section (iii)).

The only direct contact between adjacent stacks of thymine molecules takes place between the protruding methyl groups which are separated by 3·89 Å. The shortest intermolecular approach, aside from the hydrogen-bond distances, is 3·05 Å and occurs between the carbonyl oxygen O_1 of one molecule and the carbon atom C_3 of the adjoining molecule related by the unit translation along a. The superimposed molecular layers are probably held together mainly by van der Waals forces, though there is structural evidence that the sharing of hydrogen bonds with the water molecules in their present arrangement reinforces the cohesion of the layers.

(ii) *The thymine molecule*

The intramolecular bond lengths and angles are listed in Table 6 and shown in Fig. 8.

The nine atoms of the thymine molecule are closely coplanar. The best plane through the six atoms of the pyrimidine ring, as calculated by the least-squares method, is defined by the equation

$$-2{\cdot}8451x + 6{\cdot}0337y + 3{\cdot}3947z = 0{\cdot}1527\,;$$

the distance from this plane to the origin is 0·1527 Å. The average deviation of the six ring atoms from the plane is 0·004 Å, with a maximum deviation of 0·008 Å for atom N_1; the departures from this plane

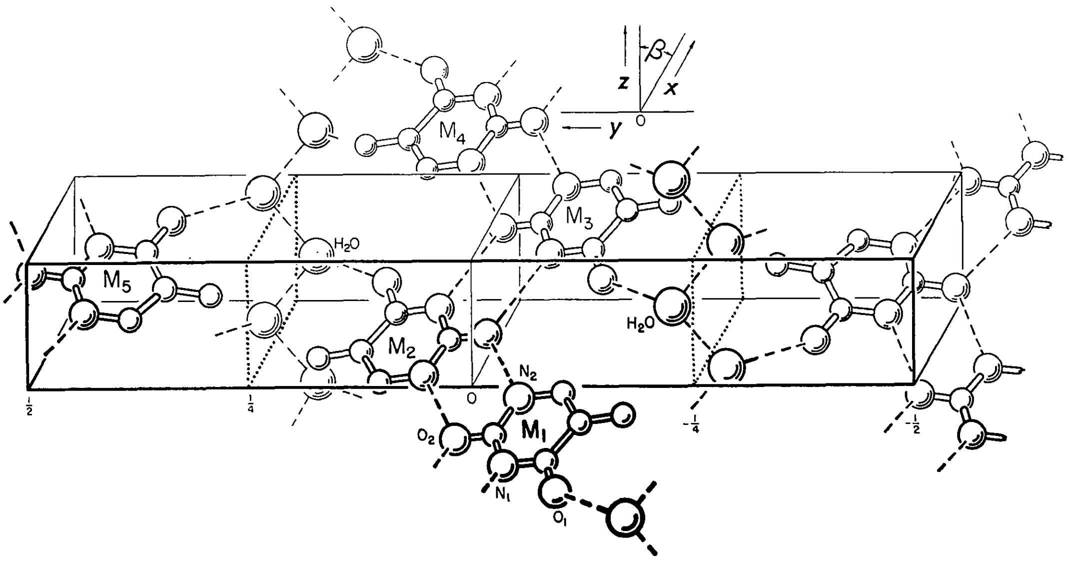


Fig. 6. A clinographic projection of the crystal structure of thymine monohydrate. Molecules M_1, M_2, M_3, and M_4 are successively related by centers of symmetry at (0, 0, 0), ($\frac{1}{2}$, 0, $\frac{1}{2}$), and (101), respectively. M_5 is related to M_2 by the glide plane (dotted line) at $y=\frac{1}{4}$, and to M_3 by a screw diad axis at ($\frac{1}{2}$, y, $\frac{3}{4}$). Coordinates of the atoms of molecule M_2 are listed in Table 3.

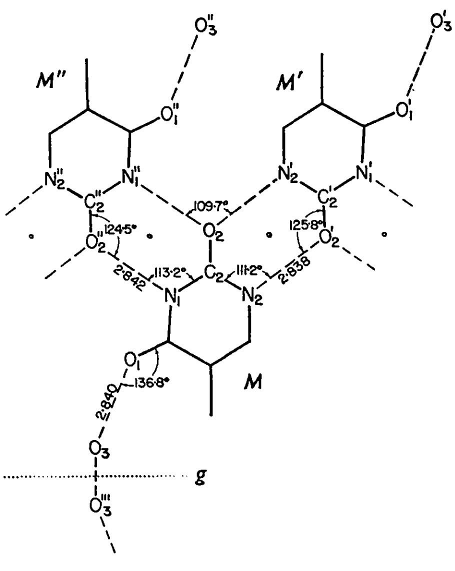


Fig. 7. A portion of the hydrogen bond network, showing bond lengths and angles associated with the hydrogen bonds. Molecule M is related to M' and to M'' by two different centers of symmetry. The dotted line g represents the trace of a glide plane.

Table 6. *Intramolecular bond distances and bond angles*

The reported confidence limits are twice the standard deviation for each bond. The calculated standard deviation of the bond angles vary from 0·3° to 0·5°

C_1–N_1	1·391 ± 0·006 Å	N_1–C_1–C_4	115·6°
N_1–C_2	1·361 ± 0·008	C_2–N_1–C_1	126·3
C_2–N_2	1·355 ± 0·009	N_2–C_2–N_1	115·2
N_2–C_3	1·382 ± 0·007	C_3–N_2–C_2	122·8
C_3–C_4	1·349 ± 0·009	C_4–C_3–N_2	121·8
C_4–C_1	1·447 ± 0·009	C_1–C_4–C_3	118·2
C_4–C_5	1·503 ± 0·008	N_1–C_1–O_1	118·3
C_1–O_1	1·231 ± 0·008	C_4–C_1–O_1	126·1
C_2–O_2	1·234 ± 0·007	N_1–C_2–O_2	122·1
		N_2–C_2–O_2	122·7
		C_1–C_4–C_5	119·0
		C_3–C_4–C_5	122·8

of the three atoms attached to the ring are as follows: O_1, −0·001 Å; O_2, −0·010 Å; C_5, −0·018 Å. Owing to the relatively small standard deviations (Table 4), the deviations for O_2 and the methyl carbon atom are probably significant.

In a recent paper, Spencer (1959) has derived from the most reliable data available in the literature average values for the ring bond lengths and angles in pyrimidines and related compounds. Spencer makes the assumption that each atom has a 'ring radius' which depends on the type of bond formed by the atom outside the ring, and that accordingly the length of a ring bond may be predicted as a first approximation by adding the radii of two adjacent atoms. A diagrammatic comparison of the bond distances observed in the pyrimidine ring of thymine with those derived by adding 'ring radii' (Spencer, 1959) is shown in Fig. 9. It is evident that Spencer's predicted values (*c*) differ considerably from the observed ones (*a*). On one hand, the significance of the deviations of the individual observed values from Spencer's predicted values might be assessed by referring to the circle of radius 2σ associated with each observed value in Fig. 9; however, on the other hand, the standard deviations of Spencer's values are so much larger than the mean standard deviation for the thymine bond lengths that any meaningful comparison of the observed and the predicted values becomes difficult, if

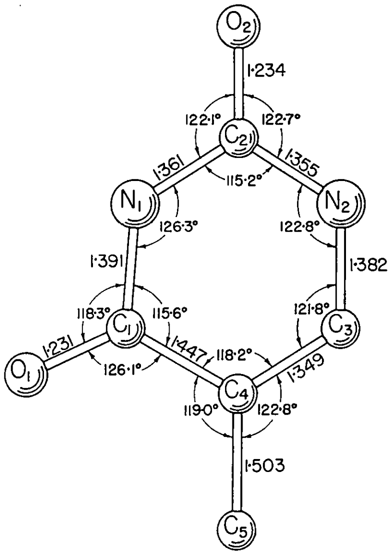


Fig. 8. Dimensions of the thymine molecule.

not impossible. In the C–N bonds of thymine the nitrogen atoms form a single bond outside the ring and two carbon atoms, C_1 and C_2, form a double bond outside the ring. On consideration of the experimental bond lengths (Fig. 9), it does not appear that the attachments to the ring affect the bond lengths in such a way that the interatomic distances could be predicted within the limit of error by the simple summation of average bond radii.

The magnitude of the measured angles support the arguments of Pauling & Corey (1956) regarding the dependence of ring angles on annular attachments.

A comparison of the bond distances in pyrimidine rings of thymine and of uracil (Parry, 1954) shows significant differences in all bonds of the same type, with a possible exception for bond C_2–N_2. These differences are astonishing in view of the fact that the molecules of thymine and uracil differ chemically only in the extra-annular attachment of a methyl group, a structural feature which alone cannot reasonably explain the large observed discrepancies in the bond distances.

In terms of the resonance theory, the bond lengths observed in thymine are satisfactorily accounted for if the canonical formulae shown in Fig. 10 are assumed to contribute to the resonant state in the following proportions:

I, 34%; II–III, 17%; IV, 10%; V, 14%; VI–VII, 4%.

The familiar equation of Pauling (1948) was used in the calculation, with C–C = 1·54 Å; C=C = 1·33 Å; C–N = 1·47 Å; $C=N^+$ = 1·24 Å; $C–O^-$ = 1·45 Å; and C=O = 1·20 Å.

The $C=N^+$ and $C–O^-$ bond distances given here are respectively modified in the manner suggested by Pauling (1948) to take into account the effect of the effective charges which are present on the N and O atoms in the corresponding cannonical formulae. The values of the bonds for the neutral pairs of atoms were taken to be C=N = 1·27 Å (Pauling, 1948) and C–O = 1·42 Å (Schomaker & Stevenson, 1941). A far better agreement with the observed results for the keto groups is obtained if the revised value C=O = 1·20 Å (Vaughan & Donohue, 1952; Penfold, 1953) is used instead of the value of 1·22 Å given by Pauling (1948). The calculated bond lengths are as follows:

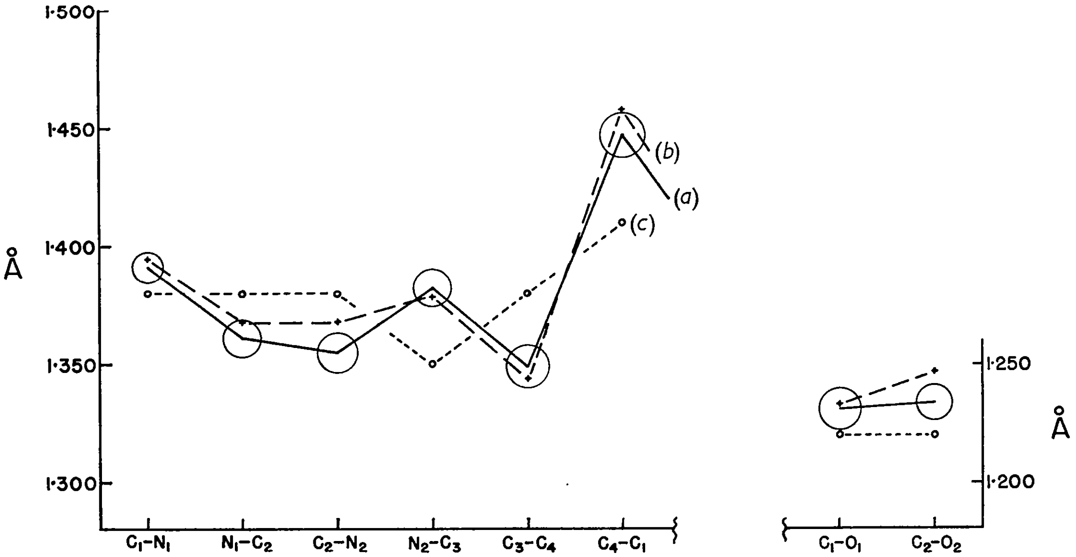


Fig. 9. Diagrammatic comparison of the bond lengths of thymine. (*a*) Observed in thymine monohydrate (the radius of each circle is taken equal to twice the standard deviation of the associated bond distance). (*b*) Calculated by the valence bond method in combining the canonical formulae I–VII, shown in Fig. 10. (*c*) predicted (Spencer, 1959).

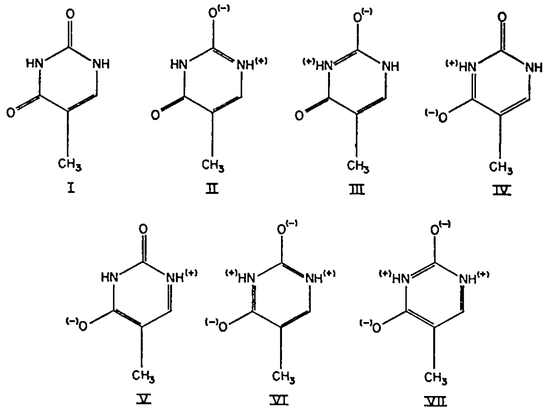


Fig. 10. Some canonical formulae of thymine.

C_1–N_1	1·395 Å	C_3–C_4	1·344 Å
N_1–C_2	1·368	C_1–C_4	1·458
C_2–N_2	1·368	C_1–O_1	1·233
N_2–C_3	1·379	C_2–O_2	1·247

In Fig. 9, these calculated values (*b*) are compared graphically with the experimental ones (*a*). In view of the approximations involved the agreement is surprisingly good. This might be, however, fortuitous in part, because no consideration has been given to the singly- and doubly-charged Dewar forms and to the ionic forms with positively or negatively charged carbon atoms.

Table 7. *Bond distances and angles involving the hydrogen atoms*

N_1–H_1	0·81 Å	C_1–N_1–H_1	125°
N_2–H_2	0·79	C_2–N_1–H_1	108
C_3–H_3	0·86	C_2–N_2–H_2	112
C_5–H_4	0·97	C_3–N_2–H_2	125
C_5–H_5	0·88	N_2–C_3–H_3	111
C_5–H_6	0·80	C_4–C_3–H_3	127
O_3–H_7	0·71	C_4–C_9–H_4	126
O_3–H_8	0·77	C_4–C_9–H_5	123
		C_4–C_9–H_6	118
		H_4–C_9–H_5	96
		H_4–C_9–H_6	96
		H_5–C_9–H_6	89
		H_7–O_3–H_8	121

The apparent bond distances and angles involving the hydrogen atoms, as calculated from the parameters of Table 2, are listed in Table 7. All the observed bond lengths are considerably shorter than the normal internuclear distances. An explanation of this apparent bond contraction has been given by Tomiie (1958) in terms of a displacement of the maximum electron density around the proton towards the heavier atom; furthermore, the shortening increases with the increase of the ionic character of the bond.

The departures of the three hydrogen atoms attached to the ring atoms from the median plane of the molecule are as follows: H_1, +0·059; H_2, +0·016; H_3, +0·017. If significant, the reason for the systematic trend of the three C–H bonds of the methyl group to be bent away from the bond attached to the ring at angles much larger than in the normal tetrahedral configuration is not readily apparent.

(iii) *The structural anomalies associated with the water of crystallization*

The importance of the water molecules to the stability of the crystal structure of thymine monohydrate is manifest in the fact that the framework collapses on removal of the water of crystallization. Although it is possible that crystals of thymine monohydrate contain variable proportions of water, the fact that the crystal selected for the structure analysis apparently contained a relative amount of water—0·83—equal to that determined on the basis of the measured density indicates that the range of water content is probably quite small. No subsidiary reflections were observed on the diffraction pattern, and it was concluded that there is no periodic packing fault in the arrangement of the water molecules but rather that the water molecules are statistically distributed over the available sites.

It perhaps should be pointed out that thymine monohydrate cannot properly be described as an inclusion compound. The dominant feature which causes such a classification to be inappropriate lies in the fact that the included species (in this case water) cannot be considered as 'trapped' in cavities formed by the other molecules, because the interaction between the two components is predominently of electrostatic nature (hydrogen bonding) and not based on van der Waals forces (Schlenk, 1951).

A surprising result of the least-squares refinement is the extremely large value of the thermal parameter γ of the oxygen atom of the water molecule; it corresponds to a B of 16 Å^2 and an implied root-mean vibrational amplitude of 0·45 Å in the c direction, obviously a prohibitive value. This anomalously large value was tentatively attributed to a localized disorder in the arrangement of the water molecules; that is, it was presumed that the z coordinate for the oxygen atom does not correspond to its actual position but rather represents the average value of a distribution of parameters. In an attempt to obtain more information about this statistical disorder a three-dimensional difference synthesis was calculated with F_c values from which the contribution of the water molecules had been omitted. A section of this difference map, passing through O_3, is shown in Fig. 11(*a*); as expected, the electron density associated with the water molecule is considerably elongated in a direction closely parallel to the c axis. The shape of the electron density distribution along the main axis of the ellipsoid is approximately Gaussian and accordingly can be adequately fitted by a single atom with a large anisotropic temperature factor; it seems more reasonable, however, to presume that it represents—as stated above—the sum of a number of smaller peaks. For the purpose of semi-quantitative discussion the electron density along the radius of a spherical oxygen atom was calculated for different temperature factors. In order to provide reasonable agreement with the empirical curve the calculation was carried out over 1000 terms of the copper sphere of reflection. In simplifying the problem it is seen that, according to the water content, at least five constituent atoms of weight $\frac{1}{6}$ are necessarily involved in the compound peak. In Fig. 11(*b*) it is shown that, with a suitably chosen temperature factor of 8 Å^2, the summation of five separate peaks (III)—each peak being one-sixth of the transform of an oxygen atom—placed at distances $0, \pm\frac{1}{2}(\overline{u^2})^{\frac{1}{2}}, \pm(\overline{u^2})^{\frac{1}{2}}$ from the average position results in curve (II) which resembles closely the observed profile (I). The value of B is still large, but in view of the environment of the water molecule it is not improbable.

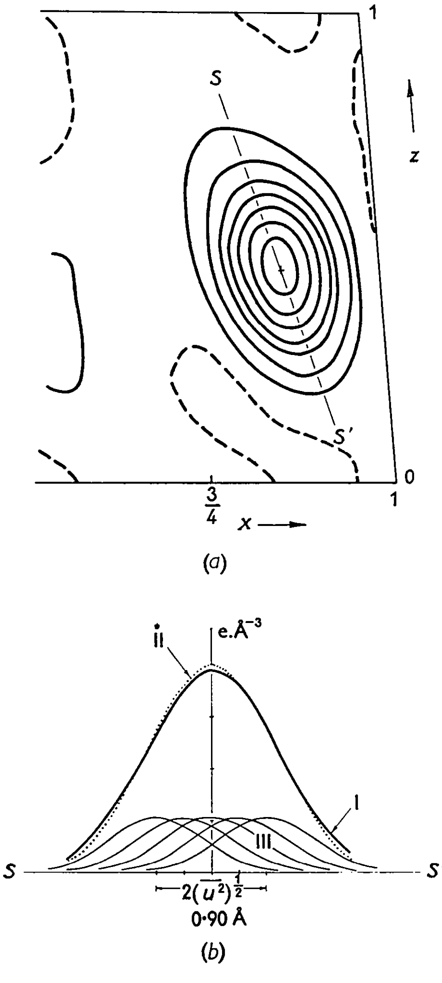


Fig. 11. (*a*) A section through the water molecule of a three-dimensional difference map in which the water contribution had been omitted from the F_c values. The contours are at intervals of 0·5 e.Å^{-3}; the zero contour is dashed. (*b*) The electron density distribution (I) along the line *ss'* of Fig. 11(*a*) as compared with the compound peak (II) obtained by the summation of five transform profiles (III) of an oxygen atom. The transforms are multiplied by a factor $\frac{1}{6}$.

An attempt is now made to find a simple model which would afford a suitable picture for the actual arrangement of the water molecules. In the first place, the water content suggests that one out of every six water molecules is missing; accordingly we reduce the

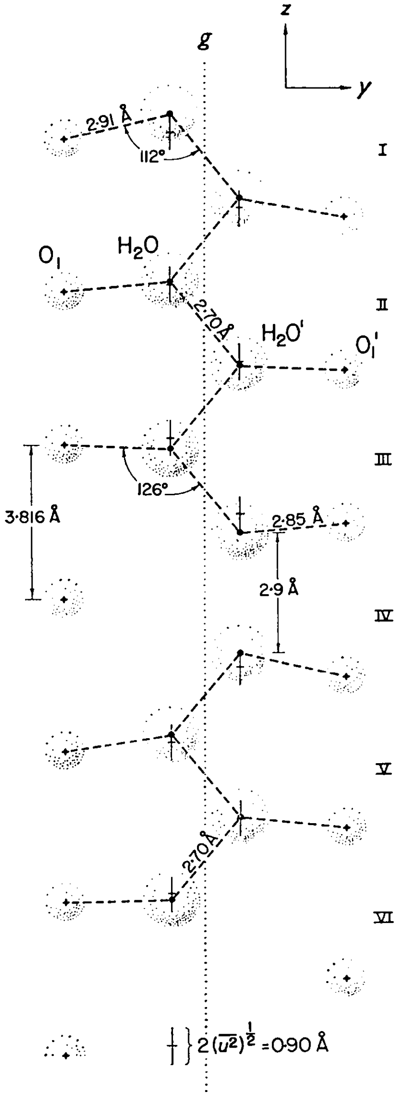


Fig. 12. Orthogonal projection onto (100) of one of the possible distributions of the water molecules. The crosses at O_1 and O_1' represent the centers of oxygen atoms of two thymine molecules related by the glide plane *g*. The heavy dots represent the oxygen atoms of water molecules, e.g. H_2O and H_2O'. The horizontal strokes, one at each occupied water site, denote the average positions related by the glide plane. The hydrogen bonds are indicated by dashed lines. The general orientation of a water molecule is as follows: e.g. the hydrogen atoms of H_2O (cell II) are engaged in the $O_1 \cdots HOH$ and $HOH \cdots OH_2'$ hydrogen bonds (also see atoms H_7 and H_8 in Fig. 4). On each side of the glide plane, five water molecules are distributed over six unit cells (I–VI) to form two short zig-zag chains of six and four elements respectively. The water molecules in a chain are separated by a constant distance of 2·70 Å, and their centers are assumed to be within a range of ±0·45 Å from the average position.

problem to finding how five water molecules on each side of the glide plane can be consistently distributed over six consecutive unit cells. As previously pointed out, any arrangement implicating a periodic pattern is to be avoided. According to our basic assumption of a localized disorder, the z coordinate of each water molecule is not unequivocally fixed, provided the oxygen atom is placed within a range of $\pm 0{\cdot}45$ Å from the average position. As a consequence, it is easily seen that the hydrogen bond linking two water molecules across a glide plane is no longer restricted to the small value of 2·55 Å, but may assume a larger and more normal value. In the arrangement shown in Fig. 12 the O–H · · · O bond between water molecules is increased to 2·70 Å, and this value is adopted as a reference to built 'rigid' zig-zag chains. The highest-membered chain that can be constructed in that fashion contains six water molecules. The remaining four water molecules are also hydrogen-bonded together to form a four-membered chain whose vertices, as a whole, are placed at random within the permissible sites. In the layer portion depicted in Fig. 12, an unoccupied site is located on each side of the glide plane (in cells IV and VI). It is interesting to note that adjacent chains have their extremities separated by an average distance of 2·9 Å,* which is about twice the van der Waals radius for oxygen. The slight change in the orientation of H_2O according to its displacement from the average position makes the bond to the hydrogen atom at one extremity of a chain point to the first oxygen atom of the adjacent chain rather than to the site of a missing water molecule (see cell IV in Fig. 12); therefore it is attractive to imagine that the protruding hydrogen atom might be engaged in a hydrogen bond between the extremities of two adjacent chains.

One may ask what causes the sequential arrangement of the water molecules to be interrupted. On the reasonable assumption that each water molecule in a chain must be hydrogen bonded to a thymine molecule, an answer is provided by the following considerations. In Fig. 12 there is evidence that localized structural modifications are associated with the displacement of a water molecule from the average position. On the basis of the proposed model, two types of alterations are most likely to produce discontinuities in the water configuration. First, when a water molecule is shifted toward positive values of z, the length of the C=O · · · H–O bond increases and may finally assume values at which a hydrogen bond is not likely to occur. The bond distance of 2·91 Å indicated in Fig. 12 (cell I) lies probably close to such a transition value. Secondly, when the central molecule is moved toward negative values of z, the hydrogen-bond angle C=O · · · HOH · · · OH_2 is increased and this variation appears to be at least as critical as the simultaneous, but slower, lengthening of the C=O · · · H–O bond. In the present case, 126° seems to be an upper angular limit above which a hydrogen bond is no longer formed. If one accepts the value 105° for the actual bond angle of water, one might tentatively say that a hydrogen bond is no longer formed if the angle, at the central donor oxygen, between one of the hydrogen bonds and the bond to the hydrogen atom is larger than 20°.

It is again pointed out that the proposed distribution of the water molecules in Fig. 12 is arbitrarily selected so as to provide a representative example for our assumption and does not reflect a periodic situation systematically repeated throughout the crystal lattice.

In conclusion we would like to emphasize the fact that if one accepts a normal hydrogen-bond length of 2·70 Å between the water molecules as prevalent throughout the crystal lattice, a configuration of the water molecules results which is consistent with the observed X-ray data. This configuration consists of indefinite sequences of short chains of hydrogen-bonded water molecules. Furthermore, it is seen that no extensive spacing necessarily intervenes between successive chains, but rather that adjacent chains have their extremities in close contact.

I should like to express my thanks to Prof. Robert B. Corey for his continuing interest and encouragement during the course of this investigation. Grateful acknowledgements are made for numerous helpful discussions with Dr R. E. Marsh on the present structure analysis.

Editorial Note:

For other hydrate structures, similar to thymine hydrate in that there are variable water contents and short water–water distances, see 'The Crystal Structure of Caffeine' by D. J. Sutor, *Acta Cryst.* **11**, 453 (1958) and 'The Crystal Structure of Biuret Hydrate' by E. W. Hughes, H. L. Yakel & H. C. Freeman, *Acta Cryst.* **14**, 345 (1961).

References

Corey, R. B. & Pauling, L. (1953). *Rev. Sci. Instrum.* **24**, 621.
Furberg, S. & Hordvik, A. (1956). *Acta Chem. Scand.* **10**, 135.
Hughes, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
Parry, G. S. (1954). *Acta Cryst.* **7**, 313.
Pauling, L. (1948). *The Nature of the Chemical Bond.* Ithaca: Cornell University Press.
Pauling, L. & Corey, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164.
Penfold, B. R. (1953). *Acta Cryst.* **6**, 591.
Schlenk, W. (1951). *Fortschr. chem. Forsch.* **2**. 92.
Schomaker, V. & Stevenson, D. P. (1941). *J. Amer. Chem. Soc.* **63**, 37.
Shoemaker, D. P., Donohue, J., Schomaker, V. & Corey, R. B. (1950). *J. Amer. Chem. Soc.* **72**, 2328.
Spencer, M. (1959). *Acta Cryst.* **12**, 59.
Tomiie, Y. (1958). *J. Phys. Soc., Jap.* **13**, 1030.
Vaughan, P. & Donohue, J. (1952). *Acta Cryst.* **5**, 530.
Waser, J. & Schomaker, V. (1953). *Rev. Mod. Phys.* **25**, 671.

---

* That is, one unit translation (3·8 Å) minus twice the r.m.s. displacement (0·45 Å) of O_3 in the c direction.