Acta Cryst. (1958). **11**, 83

The Structures of the Pyrimidines and Purines. VI.
The Crystal Structure of Theophylline

BY D. JUNE SUTOR

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 28 *June* 1957)

Theophylline crystallizes in the monoclinic space group $P2_1$, nearly $P2_1/a$, with unit-cell dimensions $a = 13{\cdot}3$, $b = 15{\cdot}3$, $c = 4{\cdot}50$ Å, $\beta = 99{\cdot}5°$. The x and y coordinates were determined by an application of the isomorphous-replacement method and the z coordinates by a consideration of the hydrogen bond system possible in the crystal; these coordinates were refined by difference syntheses and difference generalized projections. The standard deviation in a bond length is 0·01 Å. The intramolecular distances are similar to those in other compounds containing a purine or pyrimidine group, but bond lengths of substituents attached to the purine ring are equivalent to the sum of the single- or double-bond radii. The molecules are held together in the crystal by a network of hydrogen bonds. Molecules are linked in pairs across a pseudo centre of symmetry by two N–H $\cdots$ O bonds of length 2·76 Å, and each molecule is also linked to a water molecule by a N $\cdots$ H–O bond of 2·89 Å. Each water molecule also forms two further hydrogen bonds of length 2·70 Å and 2·76 Å with adjacent water molecules.

Introduction

In recent year, the crystal structures of several pyrimidines and purines have been determined. Chief interest has been in the bases obtained by the complete hydrolysis of nucleic acids, and studies have been made of the pyrimidine uracil (Parry, 1954) and the purines adenine hydrochloride (Broomhead, 1948; Cochran, 1951) and guanine hydrochloride (Broomhead, 1951). As yet, no structure determinations of the methyl hydroxy purines have been reported, and this investigation was undertaken to provide accurate information about the bond lengths in theophylline, 1:3 dimethyl 2:6 dihydroxy purine (see Fig. 1).

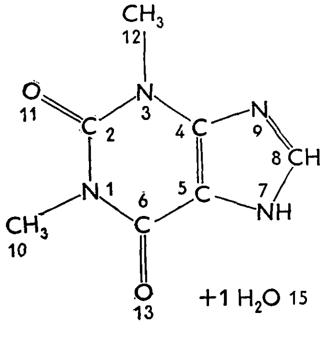

Fig. 1. Theophylline, showing the numbering system used.

Experimental

Theophylline was obtained commercially and recrystallized by dissolving in hot water and allowing the solution to cool gradually, when crystals of composition $C_7H_8N_4O_2.1\,H_2O$ were obtained as both needles and plates. The crystals used in this investigation were well formed needles elongated along c and of uniform cross-section.

Unit-cell dimensions were obtained from rotation and Weissenberg photographs:

$$a = 13{\cdot}3,\ b = 15{\cdot}3,\ c = 4{\cdot}50 \text{ Å},\ \beta = 99{\cdot}5°.$$

The density, measured by flotation, is 1·452 g.cm.$^{-3}$; the value calculated for four units of $C_7H_8N_4O_2.1\,H_2O$ per cell is 1·456 g.cm.$^{-3}$. Photographs show only the absence $0k0$ when k is odd; thus the space group is $P2_1$–C_2^2 and the asymmetric unit must contain two molecules. There are, however, only 18 reflexions of the type $h0l$ with $h = 2n+1$ present out of a possible 78 (the $h00$ reflexions with $h = 2n+1$ are entirely absent) and the largest of these is 1·2% of $F_{000}(F_{000} = 416)$. The departure from $P2_1/a$ must be very small, and throughout this work the space group has been taken as the centrosymmetric one.

For intensity data, the $hk0$, $hk1$, $hk2$ and $hk3$ layer lines were recorded on Weissenberg photographs with Cu $K\alpha$ radiation, using the multiple-film technique and the equi-inclination method where applicable. The intensity of each reflexion was estimated visually by comparison with a standard scale. A correlation factor for the multiple-film technique was calculated for each layer line by the method of Bullen (1953), and a correlation ratio between sets of films of different exposure times was determined by a comparison of the common reflexions. A correction for the variation in the area of the reflexions occurring on upper-level Weissenberg photographs was first made by the method of Phillips (1954). His formula was applied to the extended reflexions, and a correction curve for the contracted reflexions was constructed from a comparison of the intensities from the same plane occurring on both sides of the film. Later, it was shown that the differences between observed and calculated structure factors varied according to whether that reflexion had been extended or contracted, and the above method had not adequately corrected both types. As the correction to be applied is a function of $\sin\theta$, the calculated structure factors were independently scaled to the contracted and extended reflexions and a correction was effected in this way. The agreement between observed and calculated structure factors is consistently better for reflexions originally contracted, thus suggesting that measurements should be made on these and not on the extended ones. In practice, this is justifiable, for the eye can measure the integrated intensity only over a small area, the larger the area of the extended reflexion, i.e. the further out in reciprocal space the layer line, the greater will be the error in the visually estimated intensity. The discrepancy between R factors for extended and contracted reflexions increases from $hk1$ to $hk3$ (see Table 2).

Determination and refinement of the structure

The x and y coordinates were obtained from an application of the isomorphous-replacement method (Sutor, 1956), and refined by six Fourier projections to an R factor of 17%. The calculation of structure factors and Fourier series was carried out on the EDSAC. In the calculation of the structure factors, the individual scattering curves for carbon, nitrogen and oxygen of Hoerni & Ibers (1954) were used, and the calculated values were scaled to the observed. Further refinement was now accomplished by five difference syntheses (Cochran, 1951). The hydrogen atoms attached to C_8, C_{10} and N_7 were located, and round C_{12} was a positive region which could be ascribed to a rotating methyl group. The contribution of these hydrogen atoms was calculated by taking three-tenths of the contribution of ten atoms equally spaced on a circle and projected on to the $hk0$ plane. Several of the carbon, nitrogen and oxygen atoms were given additional isotropic or anisotropic temperature factors. The expression for the atomic scattering factor can then be represented by

$$f = f_0 \exp\left[-\{\alpha+\beta\cos^2(\varphi-\psi)\}\sin^2\theta\right],$$

(Hughes, 1941),

where α and β are constants, $(2\sin\theta, \varphi)$ are the polar coordinates of a point in the reciprocal lattice and ψ is the angle between the direction of maximum vibration and the a axis. Values of α, β and ψ are listed in Table 1, the direction of maximum vibration being measured anticlockwise from a. These corrections produced negligible changes in the atomic coordinates.

Table 1. *Additional temperature factors*

	*hk*0			*hk*1, *hk*2, *hk*3		
Atom	α	β	ψ (°)	α	β	ψ (°)
C_2	0·3	0·3	108	0·1	—	—
C_5	—	—	—	−0·2	—	—
C_6	0·2	0·2	0	—	—	—
C_8	0·3	0·3	121	0·1	—	—
C_{10}	0·8	0·8	45	0·8	0·8	45
C_{12}	0·8	0·8	176	0·8	0·8	176
N_1	0·2	0·2	53	—	—	—
N_3	0·2	0·2	177	—	—	—
N_7	−0·2	—	—	—	—	—
N_9	0·7	0·7	175	—	—	—
O_{11}	0·8	0·8	130	0·8	0·8	130
O_{13}	0·8	0·8	175	0·8	0·8	175
O_{15}	0·8	0·8	0	0·8	0·8	0

The scaling curve is of the form $F_o = F_c \exp[-0{\cdot}9\sin^2\theta]$. The final R factor is 7·9%.

The z coordinates were derived from a consideration of the hydrogen bond system possible in the crystal. The c axis is short and in the $hk0$ projection the water molecules lie close to the pseudo centres of symmetry (see Fig. 3). In order to form hydrogen bonds of the maximum length across these centres, the oxygen of the water molecule must have a z coordinate of approximately 0·25. The water molecule can form a further hydrogen bond with N_9 of the iminazole ring. N_9 was given a z coordinate such that this bond was 2·86 Å in length; the remaining z coordinates were calculated by assuming a planar molecule with bond lengths in the purine group similar to those in adenine hydrochloride. Bond lengths outside the rings were taken as the normal C–N and C=O distances. These coordinates were refined by two $hk1$ generalized projections (Cochran & Dyer, 1952). At this stage, no further sign changes were observed and further refinement was made by difference generalized projections. In these functions, the height of the difference maps at an atomic centre is due not only to an error in the z coordinate but also to errors in the scaling factor and temperature factor, and the slope is caused by errors in the x and y coordinates. Throughout, the slope at atomic centres was neglected and the x and y coordinates from the $hk0$ projection used for these had been accurately refined. A correction Δz_j to the z coordinate of the jth atom was made according to the equation

$$\Delta z_j = \sum_L \frac{\partial}{\partial z}\left(C_L \cos 2\pi Lz + S_L \sin 2\pi Lz\right)_{z=z_j} / C_j\,,$$

where C_L and S_L are the expressions for the cosine and sine functions for the layer line hkL as given by Cochran & Dyer, the coefficients now being $[F_o(hkL)-F_c(hkL)]$, and C_j is the curvature of the jth atom. Additional temperature factors can be determined from a consideration of the expressions $(C_L \cos 2\pi Lz_j + S_L \sin 2\pi Lz_j)$, which must be evaluated at several points round each atom to determine whether the vibration is isotropic or anisotropic. It can be shown that this method of interpreting difference generalized projections is equivalent to calculating the difference density along a line through the x and y coordinates.

Four $hk1$ and $hk3$ difference generalized projections were calculated, the $hk2$ layer line not being used till the last stage of refinement because of the large amount of work already involved. Approximate z coordinates were obtained for the hydrogen atoms of the methyl group from these difference maps, but more accurate values were calculated for the hydrogens of the iminazole ring by assuming a planar configuration with the rest of the molecule. The z coordinates of the rotating methyl group were calculated by taking the effective C–H distance to be 0·85 Å. Several of the atoms were given additional temperature factors which are listed in Table 1. The final R factors for the $hk1$, $hk2$ and $hk3$ layer lines are given in Table 2 together with individual R factors for reflexions originally extended (hkL) and contracted ($hk\bar{L}$), and the constant α in the scaling curve $F_o = F_c \exp\,[-\alpha \sin^2 \theta]$. It is fortuitous that this is a Gaussian curve since it takes into account many factors including partial correction for spot shape.

Table 2. *R factors and scaling curves for the hk1, hk2 and hk3 layer lines*

	$hk1$	$hk2$	$hk3$
R (overall) (%)	10	12·4	16·9
$R(hk\bar{L})$ (%)	9	10·7	14·1
$R(hkL)$ (%)	12	13·2	20·8
$\alpha(hk\bar{L})$	0·7	1·2	1·5
$\alpha(hkL)$	0·9	1·2	1·5

The final difference maps were used to evaluate x and y coordinates for each layer line, and the individual values were combined to find the 'best' three-dimensional coordinates according to the equations:

$$x_j = \sum_L \left(x_{CL} \cos^2 2\pi L z_j + x_{SL} \sin^2 2\pi L z_j\right) + \text{a similar expression for } y_j\,,$$

where x_{CL} and x_{SL} are the x coordinates from the difference generalized projections C_L and S_L, and the summation is carried out over all the layer lines. The average change in x and y from the final coordinates for the $hk0$ projection was found to be 0·0002 and 0·0001 respectively. These small changes would not affect the R factors and so structure factors were not recalculated. The x and y coordinates from the $hk0$ projection were also used in the calculation of the bond lengths, for it is thought that this projection is the most accurate. The final atomic coordinates are listed in Table 3, x^* and y^* parameters represent the 'best' three-dimensional coordinates. The final R factor for the 1051 observed reflexions is 11·3%.

Table 3. *The fractional coordinates referred to the monoclinic crystal axes*

	x	y	z	x^*	y^*
C_2	0·2271	0·2241	0·7277	0·2273	0·2240
C_4	0·0924	0·2340	0·0035	0·0923	0·2341
C_5	0·0738	0·1464	0·9757	0·0736	0·1468
C_6	0·1310	0·0904	0·8195	0·1309	0·0905
C_8	0·4710	0·2900	0·2205	0·4707	0·2902
C_{10}	0·2768	0·0818	0·5529	0·2767	0·0818
C_{12}	0·1885	0·3665	0·9158	0·1884	0·3663
N_1	0·2076	0·1340	0·7100	0·2072	0·1341
N_3	0·1688	0·2731	0·8807	0·1692	0·2731
N_7	0·4946	0·3666	0·1189	0·4944	0·3666
N_9	0·0299	0·2744	0·1601	0·0298	0·2745
O_{11}	0·2940	0·2552	0·6156	0·2941	0·2551
O_{13}	0·1204	0·0115	0·7895	0·1203	0·0118
O_{15}	0·4742	0·0454	0·2330	0·4740	0·0453
H_1	0·415	0·284	0·334	—	—
H_2	0·455	0·418	0·162	—	—
H_3	0·314	0·105	0·705	—	—
H_4	0·299	0·031	0·486	—	—
H_5	0·235	0·044	0·655	—	—
H_9	0·483	0·098	0·230	—	—
H_{10}	0·007	0·488	0·300	—	—

x^* and y^* are the coordinates from the weighted $hk0$, $hk1$ and $hk3$ layer lines.

The coordinates of the ten hydrogen atoms used in calculating the contribution of the rotating group are

x	y	z
0·173	0·383	0·071
0·173	0·383	0·734
0·148	0·392	0·012
0·148	0·392	0·768
0·210	0·382	0·092
0·210	0·382	0·749
0·237	0·378	0·055
0·237	0·378	0·812
0·137	0·393	0·885
0·247	0·377	0·938

Accuracy of the analysis

The standard deviation of the x and y coordinates was calculated by the method of Cruickshank (1949):

$$\sigma x_r = \left\{\overline{\left(\frac{\partial D}{\partial x}\right)^2}\right\}^{\frac{1}{2}} \Big/ C_j\,.$$

It was found that

$$\begin{aligned} \sigma x_r = \sigma y_r &= 0{\cdot}0056\ \text{Å for the } hk0 \text{ projection,} \\ &= 0{\cdot}0064\ \text{Å for the } hk1 \text{ projection,} \\ &= 0{\cdot}0064\ \text{Å for the } hk3 \text{ projection.} \end{aligned}$$

σz_r cannot be estimated by this method, but it is likely to be of the same order of magnitude, and many of the bond lengths are only slightly dependent on the z parameter. The standard deviation of a bond length is then 0·010 Å.

The standard deviation of a bond angle was calculated from the equation of Ahmed & Cruickshank (1953), and was found to be 0·5°.

Discussion of the structure

The plane of the atoms in the five- and six-membered rings is given by the equation $z = -1{\cdot}957x + 0{\cdot}737y + 1{\cdot}011$, and the average deviation of the atoms from this plane is 0·012 Å, the maximum being 0·027 Å. Substituents in the pyrimidine ring deviate from the plane by the following amounts:

C_{10}, 0·101 Å; C_{12}, 0·014 Å; O_{11}, 0·037 Å; O_{13}, 0·023 Å.

The molecule is therefore planar with the exception of C_{10}; other deviations are not significant. The non-planarity of this atom may be due to the presence of oxygen substituents on either side.

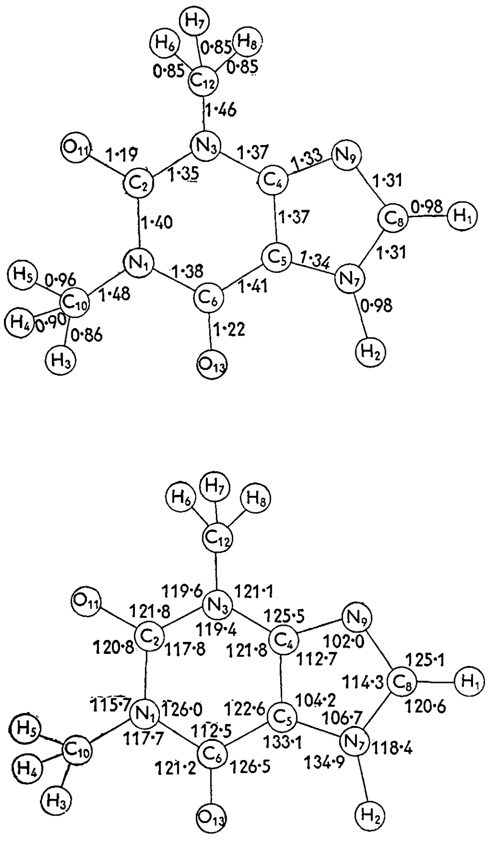

Fig. 2. Bond lengths (Å) and bond angles (°) in the theophylline molecule.

Bond lengths within the molecule are shown in Fig. 2. The lengths of the O_{15}–H_9 and O_{15}–H_{10} bonds are 0·82 Å and 0·72 Å. In a comparison of the bond lengths with similar bonds in adenine hydrochloride and uracil, use has been made of the significance tests of Cruickshank & Robertson (1953), and any difference greater than 2·5σ has been taken as significant. The only significant difference between like bonds is 0·05 Å for the C_2–N_3 bonds in theophylline and adenine hydrochloride, again probably due to the effect of substituents on these two atoms in theophylline. Uracil shows a similar shortening of this bond, and here the atom corresponding to C_2 has a substituted oxygen atom. A comparison of theophylline with uracil shows that the attachment of an iminazole ring to atoms corresponding to C_4 and C_5 produces a difference in this bond of 0·04 Å, which is only just significant, there being no significant differences in any of the other bonds.

There is a significant difference ($> 0{\cdot}03$ Å) between C–N bonds in the pyrimidine and iminazole rings of theophylline, the former being the longer. The exception is the C_2–N_3 bond, but this is also short in adenine hydrochloride and uracil. The C–N bonds outside the rings have an average of 1·47 Å, corresponding to the C–N single bond distance of 1·47 Å.

The C=O bonds are shorter, though not significantly, than C=O bonds in other compounds containing a keto group not a member of a carboxyl group. In uracil, the average of these bonds is 1·236 Å, in α-pyridone (Penfold, 1953) it is 1·236 Å, and in xanthazol (Nowacki & Burki, 1955) the two bonds are 1·21 Å and 1·24 Å. In these three compounds resonance forms containing C–O^- as a contributor are possible and these forms tend to lengthen these bonds. The average of the C = O bonds in theophylline, 1·205 Å, is in good agreement with the value of 1·20 Å postulated for this bond by Vaughan & Donohue (1952). The bond length of a substituent in a purine or pyrimidine ring is not, however, always equal to the sum of the single- or double-bond radii. Contrast adenine hydrochloride, where the substituent is NH_2 and the C–N distance is 1·30 Å, while in uracil the C=O bonds are 1·230 and 1·241 Å.

Any attempt at interpreting the length of the C–C and C–N bonds within the rings in terms of resonance contributions is limited by the fact that bonds outside the rings are pure double and single bonds. Plausible resonance forms which explain the double-bond character of C_4–N_9, C_5–N_7 and C_8–N_7 cannot be drawn. As in adenine hydrochloride, the method cannot be used to explain the bond lengths, and this is probably true of all relatively complicated molecules, where the bond lengths are determined by the geometry of the molecule as a whole.

The theophylline molecules are held together in the crystal by an arrangement of hydrogen bonds, shown in Fig. 3. Molecules are linked in pairs across an assumed centre of symmetry by two hydrogen bonds of length 2·76 Å between O_{13} and N_7 of related molecules. These bonds are somewhat shorter than the usual N · · · O hydrogen bond distance, but they are in agreement with 2·79 Å in DL-serine (Shoemaker, Barieau, Donohue & Lu, 1953) and 2·69 Å in hydroxy-L-proline (Donohue & Trueblood, 1952). Each theophylline molecule is linked to a water molecule by an O · · · N hydrogen bond of length 2·89 Å in which the water molecule donates the hydrogen atom. Each water molecule also forms two further hydrogen bonds of length 2·70 Å and 2·76 Å by donating a hydrogen atom to and receiving a hydrogen atom from

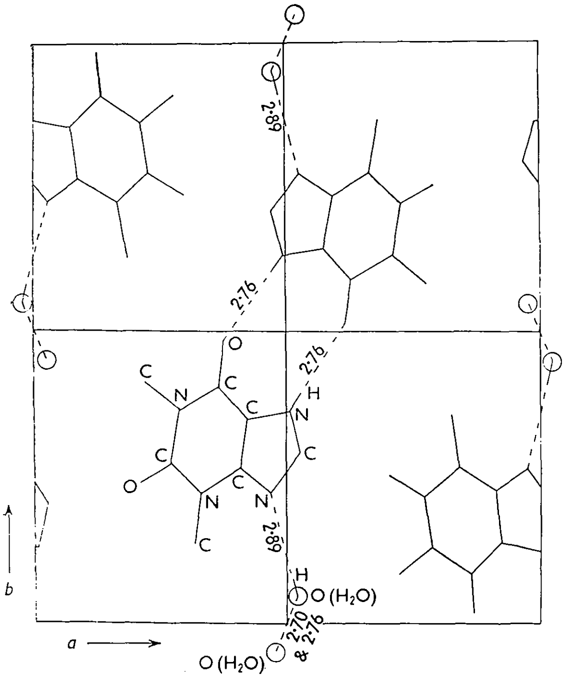

Fig. 3. The hydrogen bond system in the crystal structure of theophylline, projected on to the 001 plane. The origin is at the centre.

the water molecule above, and below, these molecules being related across the assumed centres of symmetry.

In this network of hydrogen bonds between water molecules the assumed centres of symmetry direct the hydrogen atoms towards each other and every alternate bond has no hydrogen atom participating in it. Since these bonds are too long for the hydrogen atom to be situated mid-way between the oxygens, the hydrogen atoms must be directed upwards or downwards in each chain of water molecules. Throughout the crystal structure, independent chains of water molecules may have their hydrogen atoms pointing up or down, and this arrangement is in accordance with the space group $P2_1$. If, however, the hydrogen atoms in independent chains are all directed upwards or all downwards the space group would be $P1$. There is no evidence for this, and it can be assumed that these chains are in agreement with the space group $P2_1$. This arrangement of hydrogen atoms may not represent the only departure from the centrosymmetric space group since the ellipticity of some or all the other atoms may be due to the assumed centre of symmetry; however, any error in atomic coordinates cannot be very great since the departure from $P2_1/a$ is small.

I wish to thank Dr W. Cochran for helpful discussions and advice. The calculations were carried out on the EDSAC and I should like to thank both Dr M. V. Wilkes for his permission to use the computer and the staff of the Mathematical Laboratory for their cooperation. I am grateful to Mrs S. Osmond for preparing the diagrams and for assistance with the intensity measurements and with many of the tedious calculations. Finally I am indebted to the University of New Zealand for the award of a Post-graduate Scholarship in Science, to Newnham College for a Research Studentship and to Girton College for a Research Fellowship.

References

Ahmed, F. R. & Cruickshank, D. W. J. (1953). *Acta Cryst.* **6**, 385.
Broomhead, J. M. (1948). *Acta Cryst.* **1**, 324.
Broomhead, J. M. (1951). *Acta Cryst.* **4**, 92.
Bullen, G. J. (1953). *Acta Cryst.* **6**, 825.
Cochran, W. (1951). *Acta Cryst.* **4**, 81.
Cochran, W. & Dyer, H. B. (1952). *Acta Cryst.* **5**, 634.
Cruickshank, D. W. J. (1949). *Acta Cryst.* **2**, 65.
Cruickshank, D. W. J. & Robertson, A. P. (1953). *Acta Cryst.* **6**, 698.
Donohue, J. & Trueblood, K. N. (1952). *Acta Cryst.* **4**, 419.
Hoerni, J. A. & Ibers, J. A. (1954). *Acta Cryst.* **7**, 744.
Hughes, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
Nowacki, W. & Burki, H. (1955). *Z. Kristallogr.* **106**, 339.
Parry, G. S. (1954). *Acta Cryst.* **4**, 313.
Penfold, B. R. (1953). *Acta Cryst.* **6**, 591.
Phillips, D. C. (1954). *Acta Cryst.* **7**, 746.
Shoemaker, D. P., Barieau, R. E., Donohue, J. & Lu, C.-S. (1953). *Acta Cryst.* **6**, 241.
Sutor, D. J. (1956). *Acta Cryst.* **9**, 969.
Vaughan, P. & Donohue, J. (1952). *Acta Cryst.* **4**, 324.