*Acta Cryst.* (1973). B29, 697

# Refinement of the Crystal Structure of Tubercidin*

BY J. ABOLA AND M. SUNDARALINGAM†

*Crystallography Laboratory, Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin, Madison, Wisconsin* 53706, *U.S.A.*

(*Received* 10 *November* 1970; *accepted* 28 *November* 1972)

The crystal structure of tubercidin, a C(7) analog of adenosine, has been refined using three-dimensional Cu $K\alpha$ ($\lambda = 1{\cdot}5418$ Å) diffractometer data. The coordinates for the refinement of the nonhydrogen atoms were obtained from the preceding paper.

Full-matrix least-squares refinement using isotropic temperature factors for hydrogen atoms and anisotropic temperature factors for nonhydrogen atoms gave a final $R$ value of 0·027. The bond distances and bond angles for the pyrimidine half of the base are close to those of neutral adenosine compounds, but those for the pyrrole ring are significantly different from the imidazole ring of adenosine. The conformation of the ribose ring belongs to the C(2′)-*endo* ($^2E$) class. However, relative to the plane through the three atoms C(3′), C(4′), O(1′), the puckering is C(2′)-*endo*–C(1′)-*exo* ($^2T_1$). The conformation about the C(4′)–C(5′) bond is *gauche–trans* which is one of the possible conformations for nucleosides. The observed glycosyl torsion angle ($\chi_{CN} = 73{\cdot}0°$) is in the *anti* range and the glycosyl bond distance ($1{\cdot}438 \pm 0{\cdot}004$ Å) is considerably shorter than the value normally found in nucleosides and nucleotides. It appears that the glycosyl bond distance depends on the glycosyl torsion angle and gets shorter with increasing glycosyl angle in the 0–90° range of $\chi$. Further it is noticed that this distance is a function of the sugar puckering. These are attributed to the differences in the nature of the steric interactions between base and sugar which arise as a consequence of the variation of sugar puckering and glycosyl angle. All potential donor and acceptor sites in the molecule are involved in hydrogen bonding. The pyrrole and pyrimidine parts of the bases are stacked with their respective counterparts of adjacent screw-axis related molecules.

## Introduction

We report here a refinement of the crystal structure of tubercidin (Fig. 1), a C(7) analogue of adenosine possessing anti-tumor activity (Bloch & Nichol, 1964). The crystal structure of this compound was determined by Stroud (1973, preceding paper) using the data collected on multiple-film equi-inclination Weissenberg photographs. While we were in the process of carrying out the structure analysis using data collected on a diffractometer, we learnt that Stroud (1973) had already determined its crystal structure. The present refinement was therefore initiated using his coordinates in order to obtain a more precise structure of tubercidin for comparison with adenosine derivatives.

## Experimental

A sample of tubercidin was obtained from Dr R. L. Tolman of the ICN Nucleic Acid Research Institute, Irvine, California and was recrystallized by slow evaporation of an aqeous solution at room temperature. Preliminary oscillation and Weissenberg photographs showed the crystals to be monoclinic with the space group $P2_1$. Accurate unit-cell dimensions and the crystal density measured by flotation techniques are given in Table 1. These values are compared with those of Stroud (1973).

Table 1. *Crystal data for tubercidin*

$C_{11}H_{14}N_4O_4$, M.W. 266·26

| | This work | Stroud (1973) |
|---|---|---|
| $a$ | 9·675 ± 0·002 Å | 9·6752 (3) Å |
| $b$ | 9·303 ± 0·002 | 9·3038 (2) |
| $c$ | 6·720 ± 0·001 | 6·7166 (1) |
| $\beta$ | 94·60 ± 0·02° | 94·5536 (1)° |
| $Z$ | 2 | |
| Volume | 602·908 Å$^3$ | 602·667 Å$^3$ |
| $D_{obs}$ | 1·455 g cm$^{-3}$ Flotation in chloroform and ether | 1·449 g cm$^{-3}$ Flotation in carbon tetrachloride and benzene |
| $D_{calc}$ | 1·466 g cm$^{-3}$ | 1·462 g cm$^{-3}$ |

X-ray intensities were collected on the Picker FACS1 automated diffractometer using a crystal of approximate dimensions 0·10 × 0·15 × 0·30 mm mounted about the $c$ axis and parallel to the $\varphi$ axis of the goniostat. Data over a hemisphere in reciprocal space were recorded using nickel-filtered copper radiation. Equivalent reflections showed a mean difference of only about 1%, and these were averaged and corrected for Lorentz and polarization effects. Altogether 1050 independent reflections representing 78·5% of the total number (1339) possible were collected. Of these, five reflections had intensities less than $1{\cdot}5\sigma$ [where $\sigma$ is the error in the intensities (Stout & Jensen, 1968)] and were con-

* This paper is part XX of a series on 'Stereochemistry of Nucleic Acids and Their Constituents'.

† To whom to address correspondence.

sidered unobserved. Absorption was not considered to be serious and no corrections were applied.

### Refinement of the structure

The refinement of the structure was initiated using the nonhydrogen atom coordinates from the previous analysis (Stroud, 1973). These were subjected to two cycles of full-matrix least-squares refinement (Busing, Martin & Levy, 1962) using an overall isotropic temperature factor and scale factor that were obtained from a Wilson (1941) plot. The residual index $R = \sum(|F_o| - |F_c|) / \sum|F_o|$ was lowered from 0·166 to 0·095. One cycle of refinement using anisotropic temperature

Table 2. *Observed and calculated structure factors* ( × 10) *for tubercidin*

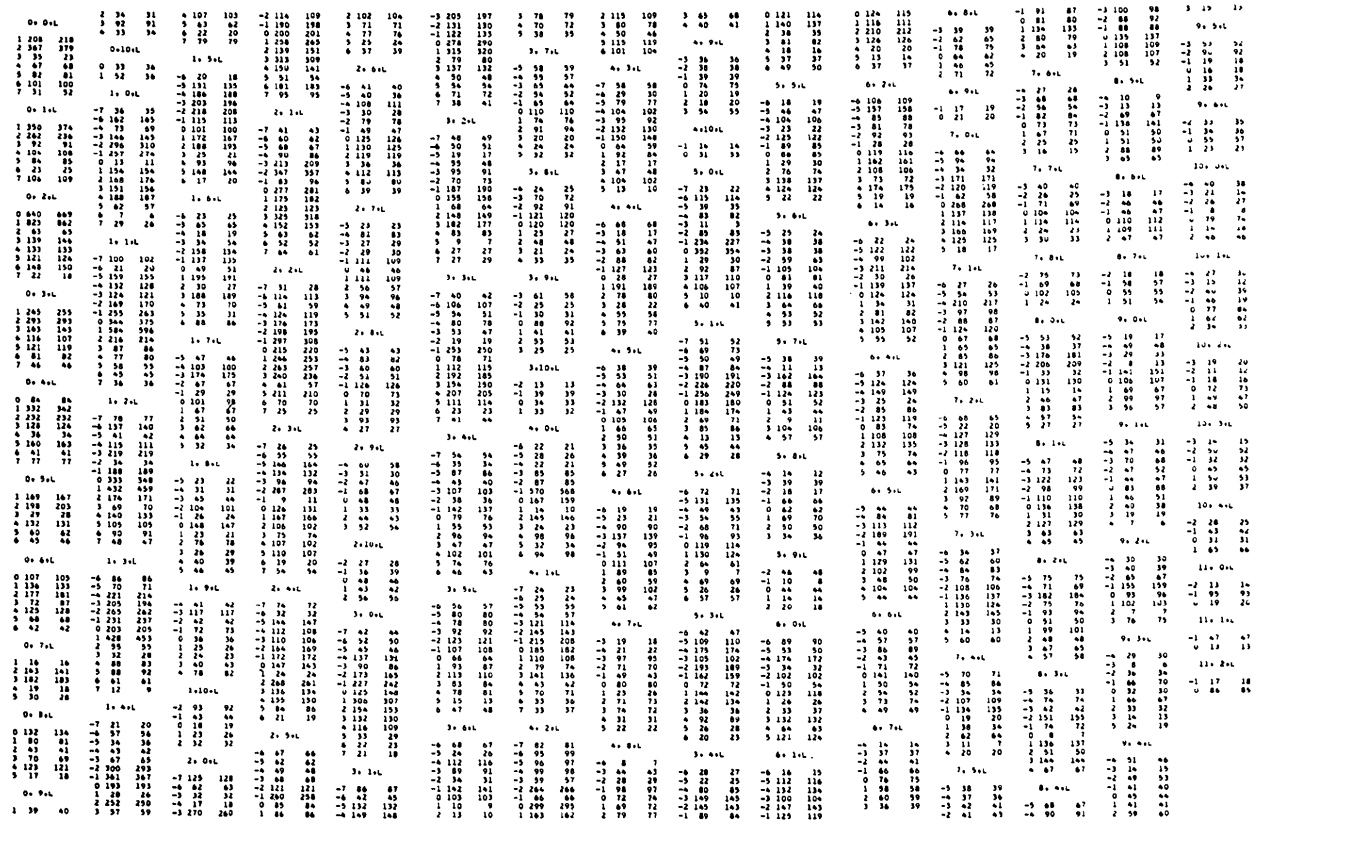

Table 3. *The atomic coordinates and their estimated standard deviations*

Thermal parameters are multiplied by $10^4$.

| | $x/a$ | $y/b$ | $z/c$ | $\beta_{11}$ | $\beta_{22}$ | $\beta_{33}$ | $\beta_{12}$ | $\beta_{13}$ | $\beta_{23}$ |
|---|---|---|---|---|---|---|---|---|---|
| N(1) | 0·6005 (2) | −0·0404 (0) | 0·4522 (4) | 57 (2) | 103 (3) | 171 (6) | −2 (2) | 28 (3) | 44 (4) |
| C(2) | 0·4606 (3) | −0·0350 (5) | 0·4395 (5) | 53 (3) | 117 (4) | 187 (7) | 5 (3) | 23 (4) | 58 (5) |
| N(3) | 0·3763 (2) | −0·0866 (4) | 0·5690 (4) | 49 (2) | 91 (3) | 164 (6) | 0 (2) | 18 (3) | 39 (4) |
| C(4) | 0·4446 (3) | −0·1531 (4) | 0·7253 (4) | 51 (3) | 68 (3) | 111 (6) | −6 (3) | 22 (3) | 1 (4) |
| C(5) | 0·5891 (3) | −0·1674 (4) | 0·7568 (4) | 49 (2) | 72 (3) | 121 (6) | −6 (2) | 15 (3) | − 1 (4) |
| C(6) | 0·6669 (3) | −0·1081 (5) | 0·6092 (4) | 54 (3) | 64 (3) | 130 (6) | −1 (2) | 18 (3) | − 3 (4) |
| N(6) | 0·8058 (2) | −0·1145 (4) | 0·6198 (4) | 51 (3) | 101 (4) | 160 (6) | −0 (2) | −27 (3) | 25 (4) |
| C(7) | 0·6170 (3) | −0·2438 (5) | 0·9407 (4) | 56 (3) | 109 (4) | 135 (6) | − 10 (3) | − 3 (3) | 24 (4) |
| C(8) | 0·4923 (4) | −0·2725 (5) | 1·0117 (5) | 65 (3) | 125 (5) | 113 (6) | − 13 (3) | − 1 (3) | 31 (4) |
| N(9) | 0·3854 (2) | −0·2179 (4) | 0·8807 (3) | 47 (2) | 91 (3) | 116 (5) | −9 (2) | 16 (3) | 18 (3) |
| C(1′) | 0·2389 (3) | −0·2362 (5) | 0·8947 (4) | 53 (3) | 73 (3) | 117 (6) | −2 (3) | 23 (3) | 7 (4) |
| C(2′) | 0·1953 (3) | −0·3878 (4) | 0·9460 (4) | 46 (2) | 65 (2) | 108 (5) | 4 (2) | 14 (3) | − 5 (3) |
| O(2′) | 0·1917 (2) | −0·4855 (4) | 0·7872 (3) | 56 (2) | 99 (3) | 190 (5) | −8 (2) | 38 (3) | −62 (3) |
| O(1′) | 0·1950 (2) | −0·1507 (4) | 1·0533 (3) | 69 (2) | 69 (2) | 207 (5) | − 23 (2) | 75 (3) | −28 (3) |
| C(3′) | 0·0504 (3) | −0·3591 (4) | 1·0113 (4) | 45 (3) | 56 (3) | 117 (6) | 2 (2) | 20 (3) | 12 (4) |
| O(3′) | −0·0470 (2) | −0·3406 (4) | 0·8430 (3) | 47 (2) | 70 (3) | 167 (5) | 3 (2) | − 8 (2) | −10 (3) |
| C(4′) | 0·0692 (3) | −0·2145 (4) | 1·1174 (4) | 43 (3) | 58 (3) | 137 (6) | −0 (3) | 26 (3) | 4 (4) |
| C(5′) | 0·0815 (3) | −0·2233 (5) | 1·3418 (4) | 83 (3) | 71 (4) | 140 (6) | 0 (3) | 21 (4) | −10 (4) |
| O(5′) | 0·0947 (2) | −0·0842 (4) | 1·4288 (3) | 84 (3) | 84 (3) | 147 (4) | 9 (2) | 1 (2) | −28 (3) |

factors lowered the $R$ value to 0·069. A three-dimensional difference-Fourier map revealed the approximate positions of 11 of the 14 hydrogen atoms of the molecule. The 11 hydrogen atoms were subjected to two cycles of least-squares refinement using isotropic temperature factors and keeping the nonhydrogen atoms fixed. A second difference-Fourier map gave the positions of the remaining hydrogen atoms which on refinement dropped the $R$ value to 0·058. It was found at this point that the three low-angle reflections 020, 021, and 111 suffered from secondary extinction. For these reflections the calculated structure amplitudes were substituted as observed values. Two more cycles of refinement of both the hydrogen and nonhydrogen atom coordinates resulted in convergence. The average ratio of the shifts in the coordinates to their estimated standard deviations was 0·068. The final $R$ value is 0·027. The average estimated standard deviations in the positional coordinates are 0·0004 and 0·005 Å for the nonhydrogen and hydrogen atoms respectively. A Hughes (1941) type weighting scheme was used, where $1/\sqrt{w}=8{\cdot}8$ for $|F_o|\leq 153$ and $1/\sqrt{w}=4{\cdot}5+0{\cdot}028\,|F_o|$ for $|F_o|>153$. The quantity minimized in the least-squares refinement was $\sum w(|F_o|-K|F_c|)^2$. The $y$ coordinate of the atom N(1) was held constant during the refinement. The scattering factors used were those of Cromer & Waber (1965) for C, N, and O atoms and that of Stewart, Davidson & Simpson (1965) for hydrogen.

The observed and calculated structure factors are listed in Table 2. The positional coordinates and thermal parameters of the nonhydrogen and hydrogen atoms together with their estimated standard devia-

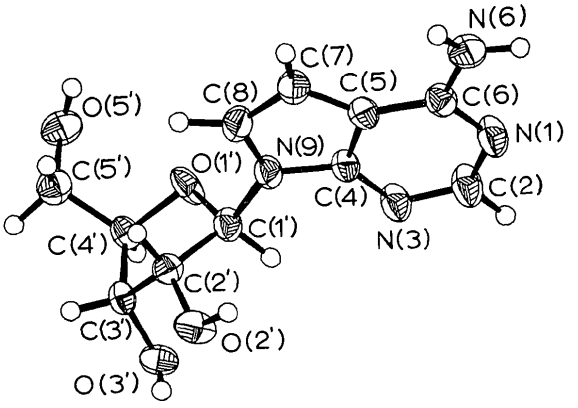


Fig. 1. The atom numbering scheme and the thermal ellipsoids for the atoms in tubercidin.

Table 4. *Hydrogen atom coordinates and standard deviations*

| | $x/a$ | $y/b$ | $z/c$ | $B$ (Å²) |
|---|---|---|---|---|
| H(2) | 0·408 (4) | 0·012 (4) | 0·319 (5) | 3·6 (0·8) |
| H(61) | 0·850 (4) | −0·172 (5) | 0·701 (5) | 2·5 (0·7) |
| H(62) | 0·849 (4) | −0·076 (4) | 0·512 (6) | 3·5 (0·8) |
| H(8) | 0·464 (5) | −0·325 (6) | 1·139 (7) | 5·9 (1·1) |
| H(7) | 0·705 (4) | −0·277 (4) | 1·002 (6) | 4·0 (0·9) |
| H(1') | 0·201 (4) | −0·208 (5) | 0·755 (6) | 4·9 (1·1) |
| H(2') | 0·259 (3) | −0·422 (4) | 1·061 (5) | 2·6 (0·7) |
| H(02') | 0·273 (5) | −0·494 (5) | 0·733 (6) | 5·1 (1·0) |
| H(3') | 0·019 (3) | −0·432 (5) | 1·099 (5) | 2·2 (0·6) |
| H(03') | −0·050 (4) | −0·409 (5) | 0·783 (6) | 3·8 (0·6) |
| H(4') | −0·006 (3) | −0·158 (4) | 1·072 (4) | 1·9 (0·6) |
| H(5') | 0·166 (4) | −0·282 (4) | 1·380 (6) | 4·9 (0·8) |
| H'(5') | −0·012 (4) | −0·269 (4) | 1·371 (5) | 3·6 (0·9) |
| H(05') | 0·181 (5) | −0·072 (6) | 1·463 (7) | 6·2 (1·2) |

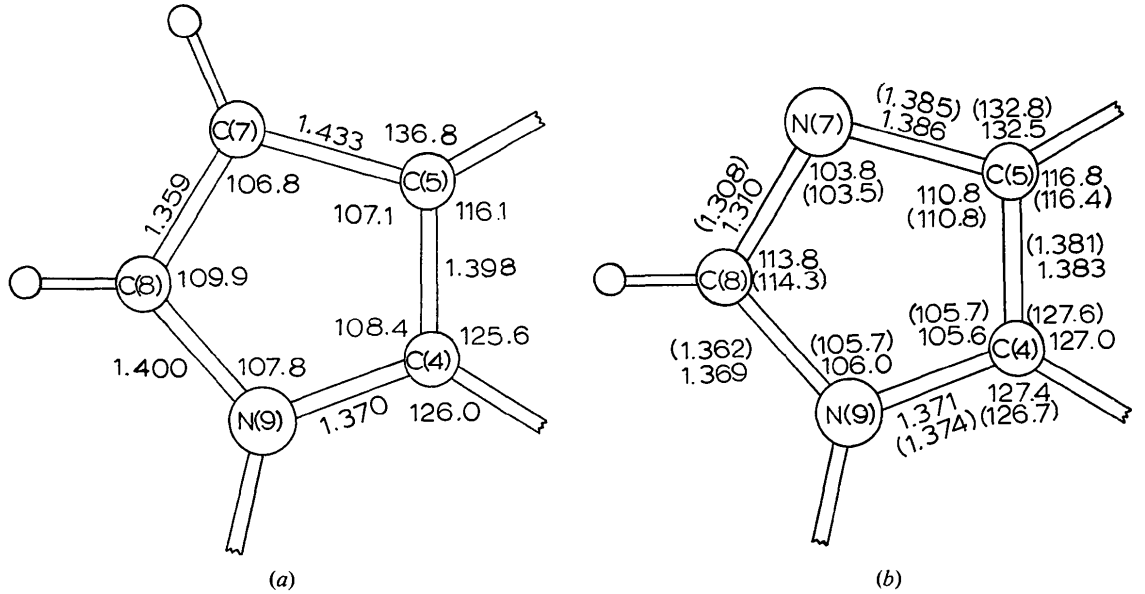


Fig. 2. A comparison of the molecular dimensions of the pyrrole ring in tubercidin (*a*) and the imidazole ring of deoxyadenosine monohydrate and adenosine (*b*) (values in parentheses are for adenosine, from Lai & Marsh).

tions are listed in Table 3 and Table 4 respectively. The thermal vibration ellipsoids are shown in Fig. 1.

## Discussion of the structure

The present refinement has given a more precise structure for tubercidin. The average estimated standard deviations in the bond distances (0·004 Å) and bond angles (0·2°) are much better than the values 0·013 Å and 0·4° obtained in the previous analysis (Stroud, 1973). In this work the hydrogen atoms have been located unequivocally and this has permitted a detailed description of the hydrogen bonding and crystal packing schemes.

### *Bond distances and bond angles*

The bond distances and bond angles in the tubercidin molecule are shown in Table 5. In general, our values agree with those of Stroud (1973) within the errors of the two structure analyses. As expected, the major differences between tubercidin and the neutral adenosine compounds, for example adenosine (Lai & Marsh, 1972) and deoxyadenosine monohydrate (Watson, Sutor & Tollin, 1965; Lin & Sundaralingam, 1972, unpublished), are in the five-membered pyrrole ring of tubercidin and the five-membered imidazole ring of adenosine and deoxyadenosine as seen in Fig. 2. [The bond distances and bond angles shown in Fig. 2 for deoxyadenosine are from a recent refinement (Lin & Sundaralingam, 1972, unpublished) of the structure reported by Watson *et al.* (1965)]. As observed by Stroud (1973), the substitution of C(7)–H for N(7) causes major differences in the bond distances involving the atom C(7) in tubercidin when compared to adenosine and deoxyadenosine. Both the bond distances C(5)–C(7) and C(7)–C(8) in tubercidin are approximately 0·05 Å longer than the C(5)–N(7) and N(7)–C(8) bond distances of adenosine. The more distal bonds to C(7), *viz.* C(8)–N(9) and C(4)–C(5), also show significant lengthening in comparison to the corresponding bonds of adenosine. The C(4)–N(9) bond opposite to C(7) is hardly affected. Similarly, the bond angles in the five-membered rings show large differences. The endocyclic angles at C(5), C(7), C(8) and C(4) of the pyrrole ring differ by about 3° from those of adenosine, while the angle at N(9) shows the smallest difference to adenosine.

The bond distances and bond angles of the pyrimidine portion of tubercidin are in good agreement with those of adenosine and deoxyadenosine.

### *Glycosyl bond distance*

It is interesting to note that tubercidin exhibits a high value (73°) for the glycosyl torsion angle $\chi_{CN}$. It may be noted that the magnitude of the glycosyl bond distance is correlated with the glycosyl angle (Sundaralingam, 1966; Lin, Sundaralingam & Arora, 1970). From the available data on the glycosyl torsion angle and bond distance on various purine and pyridine nucleosides and nucleotides, it is noticed that for both purine and pyrimidine systems, the glycosyl bond distance shows a tendency to shorten with increasing $\chi$ in the 0–90° range (Sundaralingam, 1972). Further it is noticed that in general pyrimidines tend to have slightly higher ($\simeq$0·03 Å) values for C(1′)–N than the purine systems. Another noteworthy point of interest is that the glycosyl bond distance is closely connected with the mode of sugar puckering. C(2′)-*endo* systems possess comparatively low C(1′)–N distance than the C(3′)-*endo* systems. These may be

Table 5. *Bond distances, bond angles and their estimated standard deviations*

| | | | |
|---|---|---|---|
| Base | | | |
| N(1)–C(2) | 1·350 (4) Å | C(2)–N(1)–C(6) | 118·7 (2)° |
| N(1)–C(6) | 1·347 (4) | N(1)–C(2)–N(3) | 127·6 (2) |
| C(2)–N(3) | 1·329 (4) | C(2)–N(3)–C(4) | 112·9 (2) |
| N(3)–C(4) | 1·346 (4) | N(3)–C(4)–C(5) | 125·6 (2) |
| C(4)–C(5) | 1·403 (4) | C(4)–C(5)–C(6) | 116·1 (2) |
| C(5)–C(6) | 1·406 (4) | C(5)–C(6)–N(1) | 119·1 (2) |
| C(6)–N(6) | 1·341 (4) | N(1)–C(6)–N(6) | 118·4 (2) |
| C(5)–C(7) | 1·433 (4) | C(5)–C(6)–N(6) | 122·4 (2) |
| C(7)–C(8) | 1·359 (5) | C(4)–C(5)–C(7) | 107·1 (2) |
| C(8)–N(9) | 1·400 (4) | C(5)–C(7)–C(8) | 106·8 (2) |
| N(9)–C(4) | 1·370 (4) | C(7)–C(8)–N(9) | 109·9 (2) |
| N(9)–C(1′) | 1·438 (4) | C(8)–N(9)–C(4) | 107·8 (2) |
| | | N(9)–C(4)–C(5) | 108·4 (2) |
| | | C(4)–N(9)–C(1′) | 125·2 (2) |
| | | C(8)–N(9)–C(1′) | 126·8 (2) |
| | | C(6)–C(5)–C(7) | 136·8 (2) |
| | | N(3)–C(4)–N(9) | 126·0 (2) |
| C(7)–H(7) | 0·97 (4) | C(5)–C(7)–H(7) | 129 (1·6) |
| C(8)–H(8) | 1·04 (5) | C(8)–C(7)–H(7) | 124 (1·5) |
| C(2)–H(2) | 1·01 (4) | C(7)–C(8)–H(8) | 133 (1·8) |
| N(6)–H(61) | 0·86 (4) | N(9)–C(8)–H(8) | 117 (1·8) |
| N(6)–H(62) | 0·94 (4) | N(1)–C(2)–H(2) | 120 (1·4) |
| | | N(3)–C(2)–H(2) | 112 (1·4) |
| | | C(6)–N(6)–H(61) | 121 (1·4) |
| | | C(6)–N(6)–H(62) | 117 (1·5) |
| Ribose | | | |
| C(1′)–C(2′) | 1·520 (6) | O(1′)–C(1′)–C(2′) | 104·1 (2) |
| C(2′)–C(3′) | 1·526 (4) | C(1′)–C(2′)–C(3′) | 100·4 (2) |
| C(3′)–C(4′) | 1·527 (5) | C(2′)–C(3′)–C(4′) | 102·1 (2) |
| C(4′)–O(1′) | 1·451 (4) | C(3′)–C(4′)–O(1′) | 107·1 (2) |
| C(1′)–O(1′) | 1·422 (4) | C(4′)–O(1′)–C(1′) | 107·6 (2) |
| C(2′)–O(2′) | 1·401 (4) | C(1′)–C(2′)–O(2′) | 114·8 (2) |
| C(3′)–O(3′) | 1·423 (3) | O(2′)–C(2′)–C(3′) | 111·5 (2) |
| C(4′)–C(5′) | 1·506 (4) | C(2′)–C(3′)–O(3′) | 110·9 (2) |
| C(5′)–O(5′) | 1·421 (5) | C(4′)–C(3′)–O(3′) | 108·1 (2) |
| | | O(1′)–C(4′)–C(5′) | 108·0 (2) |
| | | C(3′)–C(4′)–C(5′) | 114·6 (2) |
| | | C(4′)–C(5′)–O(5′) | 111·1 (2) |
| | | O(1′)–C(1′)–N(9) | 109·8 (2) |
| | | C(2′)–C(1′)–N(9) | 114·7 (2) |

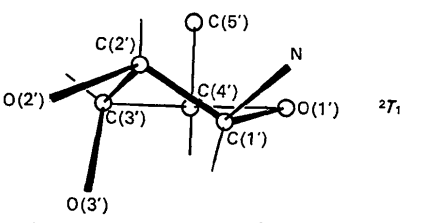


Fig. 3. The ribose conformation ($^2T_1$) in tubercidin; the C(2′)–C(1′) bond is twisted with respect to the plane formed by the remaining ring atoms O(1′), C(4′) and C(3′).

attributed to the differences in the nature of the steric interactions (short-range forces) involving the base-sugar components which arise as a consequence of the variation of the sugar puckering and the glycosyl torsion angle. The so-called anomeric effect may also be partially responsible for these changes.

*Planarity of the base*

The equation of the least-squares plane through the nine ring atoms of the base is $-0{\cdot}009X-0{\cdot}866Y-0{\cdot}501Z=-1{\cdot}233$. The deviations of the atoms from this plane are given in Table 6. The ring atoms all lie in a plane, however, the substituent ribose carbon C(1′) shows a large displacement (0·119 Å) from this plane, in agreement with the observation made by Stroud (1973). The amino group is twisted out of the base plane by 17°.

Table 6. *Deviations of the atoms from various least-squares planes*

| Base | | Ribose | | | |
|---|---|---|---|---|---|
| | (1) | | (2)† | (3)† | (4) |
| N(1) | −0·006* | C(1′) | 0·032* | 0·574 | 0·213 |
| C(2) | 0·005* | C(2′) | −0·632 | −0·066* | −0·465 |
| N(3) | −0·006* | C(3′) | −0·030* | 0·102* | 0·000* |
| C(4) | 0·000* | C(4′) | 0·048* | −0·107* | 0·000* |
| C(5) | −0·002* | O(1′) | −0·051* | 0·071* | 0·000* |
| C(6) | 0·008* | C(5′) | −1·032 | −1·463 | −1·179 |
| C(7) | −0·005* | N(9) | −0·594 | 0·256 | −0·312 |
| C(8) | −0·001* | O(2′) | −0·353 | 0·548 | −0·085 |
| N(9) | 0·007* | O(3′) | 1·296 | 1·431 | 1·336 |
| N(6) | 0·013 | | | | |
| C(1′) | 0·119 | | | | |
| O(1′) | −1·097 | | | | |
| r.m.s. deviation | 0·005 | | 0·041 | 0·088 | |

* Atoms included in the calculation of the plane.
† (2) Best four-atom plane.
(3) Next best four-atom.

*Conformation of the ribose*

Some least-squares planes through the ribose ring atoms are given in Table 6. The primary puckering (relative to the best four-atom least-squares plane) of the ribose is C(2′)–*endo* (Sundaralingam, 1965). The secondary puckering (relative to the next best four-atom least-squares plane) is C(1′)–*exo*. Thus the puckering is described as a twist ($T$) of the C(2′)–C(1′) bond with respect to the three-atom plane containing C(3′), C(4′) and O(1′) atoms and is referred to as C(2′)–*endo*–C(1′)–*exo* ($^2T_1$) (Fig. 3). A similar mode of puckering has been observed for other nucleosides exhibiting large glycosyl angles (Sundaralingam, 1972). The torsion angles of the ribose ring bonds and the H···H dihedral angles of the ribose are shown in Table 7. The ring dihedral angles can be expressed in terms of the pseudorotation parameters, the phase angle of pseudorotation ($P$) and the maximum amplitude of pseudorotation $\tau_m$ (Altona & Sundaralingam, 1972). The latter values for tubercidin along with a number of other torsion angles of interest are also listed in Table 7.

*Hydrogen bonding and base stacking*

The hydrogen bonding scheme and molecular packing are shown in Fig. 4. The bond distances and angles

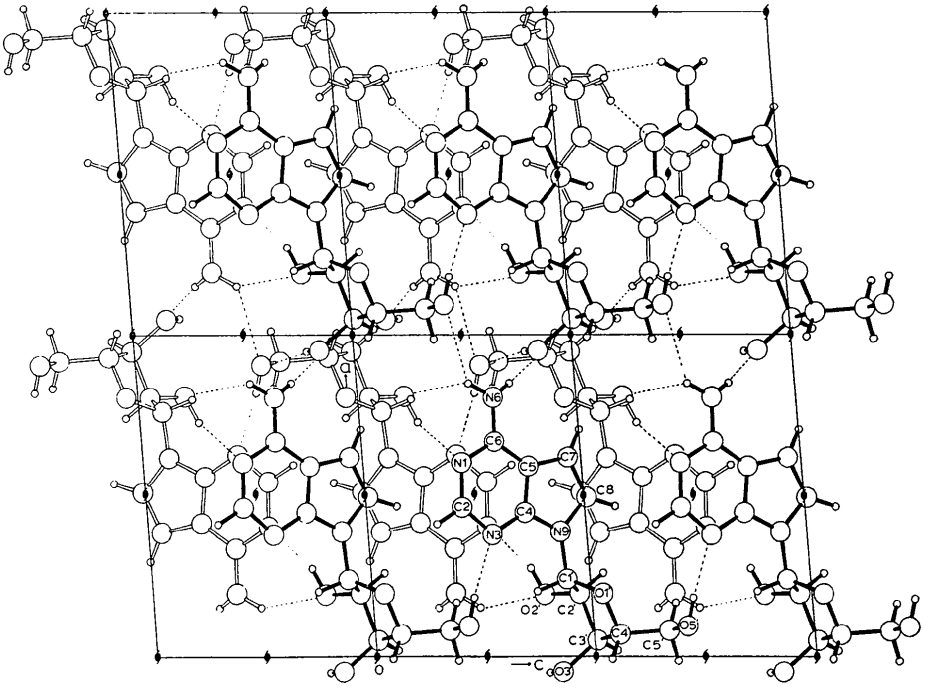


Fig. 4. The projection down the *b* axis showing the hydrogen-bonding scheme and the molecular packing.

Table 7. *Torsion angles in tubercidin**

| | | | | | | |
|---|---|---|---|---|---|---|
| Glycosyl bond | $\chi_{CN}$ | C(8) –N(9) –C(1′)–O(1′) | 73·0° | | | Conformation *anti* |
| Furanose ring bonds | $\tau_0$ | C(4′)–O(1′)–C(1′)–C(2′) | −32·9 | | | |
| | $\tau_1$ | O(1′)–C(1′)–C(2′)–C(3′) | 43·3 | H(1′)–H(2′) | 161° | |
| | $\tau_2$ | C(1′)–C(2′)–C(3′)–C(4′) | −36·3 | H(2′)–H(3′) | −43 | C(2′)-*endo*–C(1′)-*exo* |
| | $\tau_3$ | C(2′)–C(3′)–C(4′)–O(1′) | 18·2 | H(3′)–H(4′) | −104 | ($^2T_1$) |
| | $\tau_4$ | C(3′)–C(4′)–O(1′)–C(1′) | 9·0 | | | |
| Exocyclic bond | $\varphi_{OO}$ | O(1′)–C(4′)–C(5′)–O(5′) | 62·0 | H(4′)–H(5′) | −178 | *gauche–trans* |
| C(4′)–C(5′) | $\varphi_{OC}$ | C(3′) C(4′) C(5′) O(5′) | −178·3 | H(4′) H′(5′) | 60 | |
| Pseudorotation | $P$ | | 149·3 | | | |
| parameters | $\tau_m$ | | −43·8 | | | |
| Hydroxyl groups | | H(2′)–C(2′)–O(2′)–H(O2′) | −65 | | | |
| | | H(3′)–C(3′)–O(3′)–H(O3′) | −64 | | | |
| | | H(5′)–C(5′)–O(5′)–H(O5′) | −18 | | | |
| | | H′(5′)–C(5′)–O(5′)–H(O5′) | −147 | | | |
| Others | | C(4) –N(9) –C(1′)–O(1′) | −112·8 | | | |
| | | N(9) –C(1′)–O(1′)–C(4′) | 89·6 | | | |
| | | C(1′)–O(1′)–C(4′)–C(5′) | 133·3 | | | |
| | | N(9) –C(1′)–C(2′)–C(3′) | 163·3 | | | |
| | | C(1′)–C(2′)–C(3′)–O(3′) | 78·7 | | | |
| | | O(3′)–C(3′)–C(4′)–C(5′) | 140·6 | | | |
| | | O(2′)–C(2′)–C(3′)–O(3′) | −43·3 | | | |

* The estimated standard deviations in torsion angles involving nonhydrogen atoms are about 0·4° while those involving hydrogen atoms are about 3°.

Table 8. *Hydrogen-bond distances and angles**

Estimated standard deviations are given in parentheses and refer to the least significant digit.

| Symbol | Translation | | | | Angle | Length | Length from |
|---|---|---|---|---|---|---|---|
| No.* | $X$ | $Y$ | $Z$ | Atoms | (°) | (Å) | hydrogen (Å) |
| 1 | +1 | 0 | 0 | N(6) –H(61) ···O(3′) | 167 (1) | 2·891 (5) | 2·05 (4) |
| 2 | +1 | 0 | +1 | N(6) –H(62) ···O(2′) | 143 (1) | 2·989 (4) | 2·19 (4) |
| 1 | +1 | 0 | −1 | N(6) –H(62) ···O(5′) | 131 (1) | 3·180 (3) | 2·49 (4) |
| 2 | +1 | −1 | +1 | O(2′)–H(02′)···N(1) | 159 (1) | 2·722 (3) | 1·87 (4) |
| 2 | 0 | −1 | +2 | O(3′)–H(03′)···O(5′) | 168 (1) | 2·924 (4) | 2·18 (4) |
| 1 | 0 | 0 | +1 | O(5′)–H(05′)···N(3) | 167 (1) | 2·811 (3) | 1·97 (4) |

* Symmetry operations: (1) $X$, $Y$, $Z$; (2) $-X$, $\frac{1}{2}+Y$, $-Z$.

involving the hydrogen bonds are given in Table 8. The hydroxyl groups of the ribose appear to be involved in a donor and an acceptor hydrogen bond. O(2′) is involved simultaneously in hydrogen bonding to N(1) and N(6) of the base. The base is hydrogen bonded to the ribose hydroxyls only and there are no interbase hydrogen bonds. N(6) is involved in a strong hydrogen bond to O(2′) and a possible weak hydrogen bond to O(5′).

The bases show a head-to-tail stacking pattern, which is one of the characteristic modes of base stacking in crystal structures of purine derivatives (Bugg, Thomas, Sundaralingam & Rao, 1971). Within the stacked column of bases adjacent pyrimidine rings form a staggered arrangement with overlapping N(1) and N(3) atoms. The only parts of adjacent pyrrole rings that overlap are the C(8)–H atoms of molecules related by the $2_1$ axis. The stacked 'wall' of bases form the hydrophobic region at $a/2$ while the hydrophilic sugar residues are interleaved between the hydrophobic regions.

We gratefully thank Dr M. Stroud for making available his coordinates for this refinement. We also wish to thank Dr R. L. Tolman for supplying the tubercidin and Dr S. T. Rao for many of the computer programs used in this work. This work was supported by a grant GM 17378 from the National Institute of Health of the USPHS.

## References

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

BLOCH, A. & NICHOL, C. A. (1964). *Antimicrobial. Agents and Chemotherapy*, pp. 530–539. American Society for Microbiology.

BUGG, C. E., THOMAS, J. M., SUNDARALINGAM, M. & RAO, S. T. (1971). *Biopolymers*, **10**, 175–219.

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

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

HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737–1751.

LAI, T. F. & MARSH, R. E. (1972). *Acta Cryst.* B**28**, 1982–1989.

LIN, H-Y., SUNDARALINGAM, M. & ARORA, S. K. (1971). *J. Amer. Chem. Soc.* **93**, 1235–1241.

STEWART, R. F., DAVIDSON, E. R. & SIMPSON, W. T. (1965). *J. Chem. Phys.* **42**, 3175–3187.
STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination.* New York: Macmillan.
STROUD, R. M. (1973). *Acta Cryst.* B**29**, 690–696.
SUNDARALINGAM, M. (1965). *J. Amer. Chem. Soc.* **87**, 599–606.
SUNDARALINGAM, M. (1966). *Acta Cryst.* **21**, 495–506.
SUNDARALINGAM, M. (1972). In *The Jerusalem Symposia on Quantum Chemistry and Biochemistry.* Jerusalem: The Israel Academy of Sciences and Humanities. In the press.
WATSON, D. G., SUTOR, D. J. & TOLLIN, P. (1965). *Acta Cryst.* **19**, 111–124.
WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 151–152.

*Acta Cryst.* (1973). B**29**, 703

# Seven Basic Conformations of Nucleic Acid Structural Units

BY S.-H. KIM*

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts* 02139, *U.S.A.*

HELEN M. BERMAN

*The Institute for Cancer Research, Philadelphia, Pennsylvania* 19111, *U.S.A.*

NADRIAN C. SEEMAN†

*Department of Biological Science, Columbia University, New York, New York* 10027, *U.S.A.*

AND MARSHALL D. NEWTON

*Department of Chemistry, Brookhaven National Laboratory, Upton, New York* 11973, *U.S.A.*

(*Received* 20 *November* 1972; *accepted* 25 *November* 1972)

Seven basic conformations for dinucleoside phosphates are described. These conformations, derived by examination of the results of crystal structure determinations of dinucleoside phosphates and mononucleotides, conformation energy calculations, and *ab initio* molecular orbital calculations, can be considered as fundamental structural units of nucleic acids. Some possible nucleic-acid secondary structures are proposed from a consideration of these conformations.

## Introduction

Studies of nucleotides and oligonucleotides can be very useful in describing and laying the groundwork for predicting possible secondary structures in nucleic acids. Structural information essential for nucleic acid model building can be obtained from (*a*) the crystal structures of nucleosides and mononucleotides (*e.g.* Arnott, 1970); (*b*) interpretation of diffraction patterns from nucleic acid fibers (*e.g.* Watson & Crick, 1953; Marvin, Spencer, Wilkins & Hamilton, 1961); (*c*) the crystal structures of dinucleoside phosphates (Seeman, Sussman, Berman & Kim, 1971; Rubin, Brennan & Sundaralingam, 1972) and oligonucleotides and (*d*) calculations of conformation energies (*e.g.* Olson & Flory, 1972). In the past, all model building of nucleic acids has been based on information from (*a*) and (*b*). We have extended this by adding information from (*c*) and (*d*) to allow us to present a comprehensive proposal of a basic set of building blocks for use in nucleic acid model building based on base sequences or low-resolution electron density maps. We describe a set of seven basic building units, each of which is a dinucleoside phosphate of a particular conformation and which represents an overlapping building block. The derivation of the seven basic conformers and their polymers is given below.

## The conformation of the nucleoside portion in nucleotides is the same

From surveys of the structures of 5′-mononucleotides and related compounds (Sundaralingam, 1969; Arnott, 1970), and from the results of the determination of the crystal structure of a dinucleoside phosphate, uridylyl-3′,5′-adenosine phosphate (UpA), it was pointed out by Seeman *et al.* (1971) that all four major nucleosides have the same preferred conformations: the relationship between the base and the ribose sugar is the *anti* conformation (Donohue & Trueblood, 1960); the conformation of the ribose sugars is either C(3′)-*endo* or C(2′)-*endo*; there are CH···O close contacts between O(5′) of the sugar and the hydrogen bonded to either C(6) of the pyrimidine or C(8) of the purine. The latter interaction was also observed in n.m.r. studies of dinucleoside phosphates in aqueous solution (T'so,

* Present address: Dept. of Biochemistry, Duke University School of Medicine, Durham, North Carolina 27706, U.S.A.
† Present address: Dept. of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, U.S.A.