ACTA CHEMICA SCANDINAVICA 4 (1950) 751–761

An X-Ray Study of the Stereochemistry of the Nucleosides

S. FURBERG

Birckbeck College Research Laboratories, London, England

The nucleosides are compounds of great biological interest, and have therefore been subjected to comprehensive chemical investigations, as a result of which their chemical formulae are now well established. The *d*-ribose exists in the furanose condition [1], and the N-glycosidic linkage is almost certainly of the β-type [2], although direct confirmation of the latter point would appear very desirable. It still remains, however, to determine the actual shape of these molecules, especially the mutual orientation of the ribose ring and the pyrimidine or purine ring. The solution of this problem is not only of general stereochemical interest, but should also provide a clue to the structure of the nucleic acids. Only few physical investigations related to the structure of the nucleosides have been carried out. On basis of X-ray studies of fibres of Na-thymonucleate Astbury [3] has suggested that the base is parallel to the sugar, the bond joining the two rings making an angle with both ring planes. Hendricks [4] investigated the ions of adenosine and guanosine by measuring the interplanar cleavage spacings, d(001), of their salts with the clay mineral montmorillonite. He found that the atoms in these two compounds lie in or near two parallel planes 1.5 Å apart; one plane contains the purine radical, the other the ribose ring, and it is held that the hydroxyl groups as well as the primary alcohol group must be approximately in the plane of the purine. These investigations do not, however, give any detailed information about the structure of the molecules, and can hardly be regarded as conclusive even as far as their general shape is concerned. X-ray analysis of single crystals would appear to be essential. No such X-ray work on nucleosides is reported in the literature, but crystal structure determinations have been completed of some of their components, or related compounds. Valuable information relevant to the present problem is given by the work of Pitt [5] on hydroxypyrimidines, Clews and Cochran [6] on aminopyrimidines, Broomhead [7] on adenine

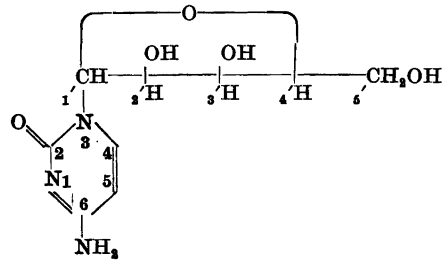
hydrochloride, and Beevers and Cochran [8] on sucrose sodium bromide dihydrate.

In the following crystal data will be given for each of the four nucleosides obtained from yeast nucleic acid, and their structures discussed.

CYTIDINE

Crystallographical. The crystals of cytidine are orthorhombic, elongated along the *c*-axis, bounded by faces (110) and (1$\bar{1}$0), with unit cell dimensions $a = 13.93$ Å, $b = 14.75$ Å and $c = 5.10$ Å. Density 1.53_2 g/cm^3. Space group $P\ 2_12_12_1$; four molecules (calc. 3.98) per unit cell.

Atomic co-ordinates were postulated by a careful consideration of available chemical, physical and crystallographic evidence, and the structure refined by two-dimensional Fourier syntheses. The structure determination is described in detail elsewhere [9], and here only points of stereochemical interest will be dealt with. The main results of the structure analysis of cytidine are evident from Fig. 1, which shows the molecule in the *c*-projection, with bond lengths



Cytidine

and bond angles. No great accuracy can be claimed, as only two-dimensional methods are employed and as rather bad molecular overlap occurs in the *a*- and *b*-projections. The agreement between $F_{obs.}$ and $F_{calc.}$ is, however, satisfactory, and the maximum possible error in the bond lengths is thought to be about 0.1 Å. The ribose ring is imperfectly resolved also in the important *c*-projection, and the greatest errors are therefore likely to occur in this part of the molecule.

Stereochemical. The investigation proves directly that the glycosidic linkage is of the β-type, as the bonds C_1'—N_3 and C_2'—O_2' are in *trans* position. The sugar is in the furanose condition, and cytidine can thus be described as cytosine-3-β-*d*-ribofuranoside, in accordance with the chemists findings.

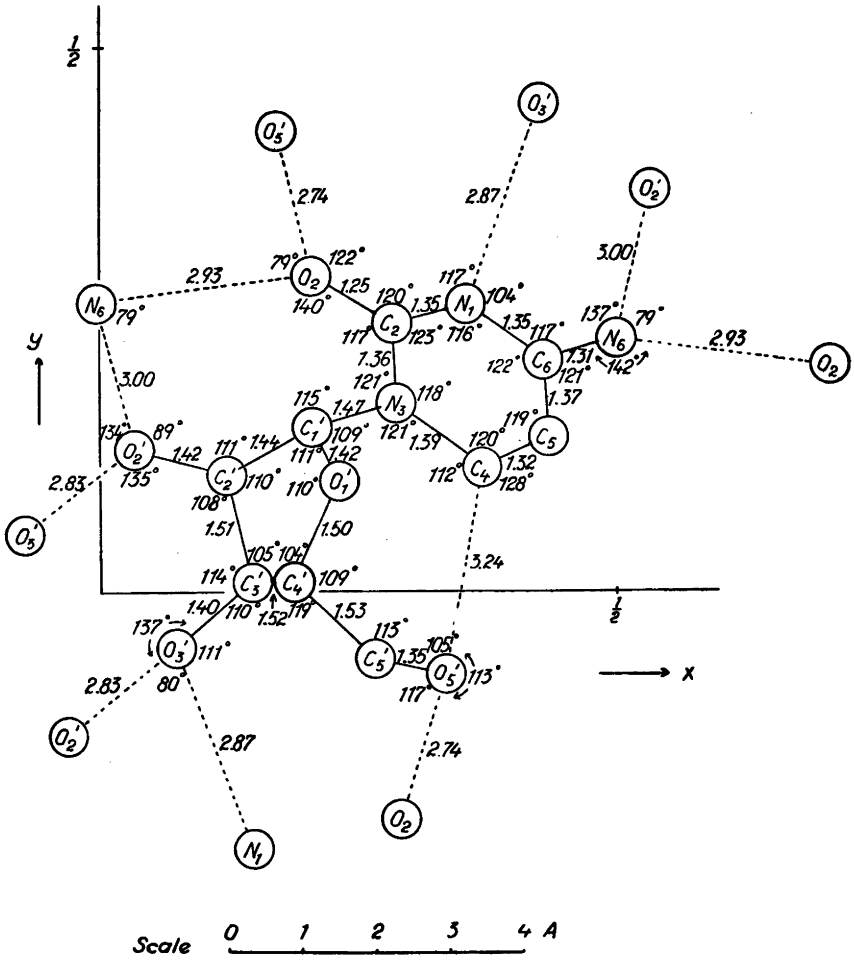


Fig. 1. The c-projection of a molecule of cytidine and its nearby contacts, with bond lengths (in Å units) and bond angles. Hydrogen bonds are dotted.

Consider first the pyrimidine ring and its substituents. The six ring atoms as well as O_2, N_6 and C_1' are all found to lie in the same plane within the experimental error. The bond lengths in the ring vary from 1.32 Å to 1.39 Å, and the bond angles are all near 120°. Although the rather wide limits of error make it impossible to discuss these results in detail, it would appear that we are confronted with a *resonating* system, in which the ring itself and the bonds to the keto (1.25 Å) and the amino group (1.31 Å) take part. The bond to the

carbohydrate does not seem to be involved, as it is found to have the length of a single bond (1.47 Å), but it lies in the plane of the pyrimidine ring in accordance with the double bond character of the ring bonds to N_3. Besides the "normal" structure with the groups $\diagdown C—NH_2$ and $\diagdown C = O$, structures with formal charges containing the groups $\diagdown C = \overset{+}{N}H_2$ and/or $\diagdown C—O^-$ must be supposed to take part, in order to account for the bond lengths. No estimation can be given from the present two-dimensional analysis as to which extent each canonical form contributes to the molecule, but the shortness of the bonds $C_2—O_2$ and $N_6—C_6$ indicate that important contributions may come from forms containing the groups $\diagdown C = O$ and $\diagdown C = \overset{+}{N}H_2$. It is in this connection interesting that both the external bonds from the latter group lie in the plane of the pyrimidine ring within the experimental error.

The shortness of the bond $C_6—N_6$ might also be explained by assuming that the molecule exists in its iminoform $—NH\diagdown C = NH$ in the crystal. This is, however, considered unlikely for chemical reasons, and also crystallographic evidence in favour of the aminoform is obtained from a study of the external bonding scheme in the crystal structure.

The *d*-ribose ring is found to be non-planar. Four of the atoms of the ring lie nearly in one plane, but the fifth member of the ring, C_3', is out of this plane by about 0.5 Å. It is very interesting that Beevers and Cochran [8] found the same feature in the case of fructo-furanoside, as it indicates that deviation from planarity may be a characteristic property of all saturated furanose rings. Some of their other results concerning the five-membered ring could, however, not be confirmed. They found all the bond distances in the ring to be short, 1.42 Å—1.45 Å, and a rather small bond angle of 104°. In addition the angles between the bonds to the hydroxyl groups and the adjacent ring bonds were definitely high at 113°—118°. The bond lengths found in the furanose ring in cytidine vary from 1.42 Å to 1.52 Å, and having in mind the rather wide limits of error these values do not differ significantly from normal single bonds. The mean bond angle in the ring is 108°, and the angles between the bonds to $(OH)_2'$ and $(OH)_3'$ and the ring bonds are not far from the tetrahedral angle, having the values 111°, 108°, 114° and 110°. *D*-ribose is thus found to be more "normal" than fructofuranoside, with bond angles nearer the tetrahedral angle, and bond lengths nearer to a single bond.

The central bond $N_3—C_1'$ lies in the plane of the pyrimidine ring and makes nearly tetrahedral angles with the adjacent ring bonds in the *d*-ribose. Con-

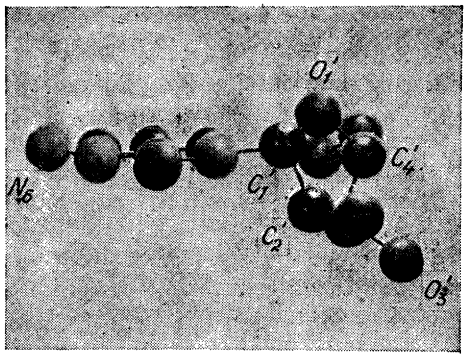


Fig. 2. A model of the cytidine molecule.

trary to Astbury's suggestion of the two rings being parallel, they are oriented in such a way that they are nearly *perpendicular* to each other as illustrated by the model in Fig. 2. Van der Waals forces would appear to play an important role in determining this mutual orientation. A study of a model shows that these forces will tend to impose restrictions on the freedom of rotation about the central bond, as the atoms in the *d*-ribose ring in many positions will come unfavourably near the keto group in the pyrimidine ring. With the *d*-ribose in *cis* position to the atom O_2 (referred to the bond N_3—C_1') the distances from O_2 to some of the atoms in this ring are only 2.4 Å—2.7 Å, and a *trans* arrangement, as actually found in the crystal structure, is therefore energetically more favourable.

The distance between the atom C_4 in the pyrimidine ring and O_5' in the primary alcohol group of the *d*-ribose is found to be as short as 3.24 Å, which would appear to be significantly less than the normal van der Waals approach of 3.4—3.5 Å. This may possibly be due to some kind of attraction between these two atoms, presumably of the hydrogen bond type. The relative position of C_4 and O_5' is also what we should expect in case of bond formation, as O_5' does not lie far away from the plane of the pyrimidine ring. The reason for the formation of the proposed bond may be sought in the possible polarisation of the group at C_4 in the pyrimidine ring to C^-—H^+, due to the inductive effect of the electronegative substituents in the pyrimidine ring (the keto and amino group).

Chemical evidence shows that in cytidine (and uridine) the carbohydrate is more firmly linked to the rest of the molecule than in other glycosides (inclusive of adenosine and guanosine). This abnormal behaviour may partly be related to the suggested intra-molecular bond, but the bond, if it exists, appears to be much too weak to offer a full explanation of the conduct of

cytidine. Another interesting fact in this connection is that 4,5-dihydrocytidine behaves chemically like a normal glycoside, quite unlike cytidine, and in this compound we should not expect such a bond to be formed, as the (CH_2)-groups can hardly be polarised to any extent.

The external bonding scheme in the crystal structure of cytidine consists of a very extensive system of hydrogen bonds, in which all its active groups are involved. Each molecule is linked to its neighbours by ten hydrogen bonds of lengths 2.74 Å—3.00 Å. The hydroxyl groups $(OH)_2'$ and $(OH)_3'$, as well as the amino group, are engaged in two external bonds, whereas the hydroxyl group $(OH)_5'$ makes only one external bond, possibly because of its participation in the intra-molecular hydrogen bond proposed above. For a fuller description of the external bonds, the reader is referred to a contemporary paper [9].

URIDINE

Crystallographical. The crystals of uridine are monoclinic, elongated along the *c*-axis, and bounded by faces $(1\bar{1}0)$ and (110). The cell dimensions are: $a = 13.88$ Å, b = 14.62 Å, $c = 5.00$ Å, $\beta = 95°$ (approx.). Cell volume 1010 Å^3; density 1.59 g/cm^3. Four molecules per unit cell (calc. 3.97).

The only systematic absences occur in the $0k0$-reflexions for odd values of k, and as uridine is optically active, the space group is $P\,2_1$. The asymmetric unit in the crystal consists of two molecules of uridine. Although uridine belongs to the space group $P\,2_1$, it has some of the features of cytidine's space group $P\,2_12_12_1$, as the reflexions 001, 100 and 300 are very weak. At the same time, there is a very great similarity in the cell dimensions of the two compounds, and the molecules of uridine are therefore placed roughly in the same positions in the unit cell as those of cytidine.

Stereochemical. Chemically, uridine differs from cytidine only in having a hydroxyl group in the 6-position instead of the amino group. The resemblance in the cell dimensions indicates that also the *shape* of the molecule is likely to be nearly the same in both cases, and our conclusions regarding the stereochemistry of cytidine can therefore probably be applied directly to uridine.

ADENOSINE

This purine nucleoside is of special importance on account of its participation in coenzyme molecules, and attempts to elucidate its structure would appear to be of great interest.

Crystallographical. The needle-shaped crystals are monoclinic, elongated along the c-axis, and bounded by faces (110) and $(1\bar{1}0)$. Even the best speci-

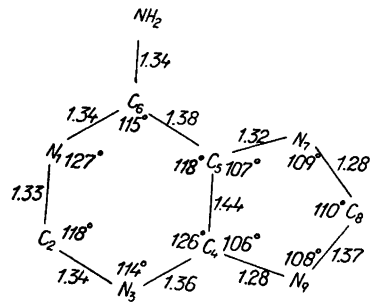


Fig. 3. Bond lengths and bond angles in adenine (J. M. Broomhead).

mens available were poorly developed and very small, with a cross-section not exceeding 0.04 × 0.04 mm. The unit cell dimensions are: $a = 11.94$ Å, $b = 10.22$ Å, $c = 4.82$ Å, $\beta = 103°$ (approx.). Cell volume 572 Å^3; molecular weight 267. The density was found to be about 1.57 g/cm^3, but an accurate determination was difficult because of the small, imperfect crystals. There are two (calc. 2.02) molecules in the unit cell.

The only systematic absences occur in the $0k0$-reflexions for odd values of k, and the space group is $P\ 2_1$. There is no centre of symmetry in the important c-projection, and no attempt was therefore made at this stage to work out the structure completely. In view of the very short c-axis, however, even only an estimate of the extension of the molecule in the c-projection may be of value. This was done by studying the low order reflexions, supported by a Patterson projection. The $h00$-reflexions show interesting intensity variations; 100 is absent, and 300 and 500 weak, whilst 400 is stronger than 200. This means that there must be atoms near $x = 0$, $x = 0.25$ and $x = 0.50$. A consideration of other reflexions indicates that the molecule must extend at least half a cell edge also in the y-direction, probably considerably more.

Stereochemical. The two components of adenosine have both been examined by X-rays before, adenine by Broomhead [7] and *d*-ribose by the author in his work on cytidine. The result of the former investigation is given in Figure 3, of the latter in Figure 1. It is proved that the glycosidic linkage is between N_9 in the purine and C_1' in the *d*-ribose [10], and that it is of the β-type [2]. One important remaining problem is the direction of the central bond N_9—C_1' connecting the two ring systems. Hendricks [4] and Astbury [3] have both suggested that it forms an angle with the plane of the purine (see above), so as to make the purine and the furanose ring parallel. In cytidine, however, this bond lies *in* or very near the plane of the pyrimidine ring, and the high degree of resonance in adenine, which is apparent from the bond lengths in Fig. 3, would seem to make this likely to be the case also in adenosine. Such a "cyti-

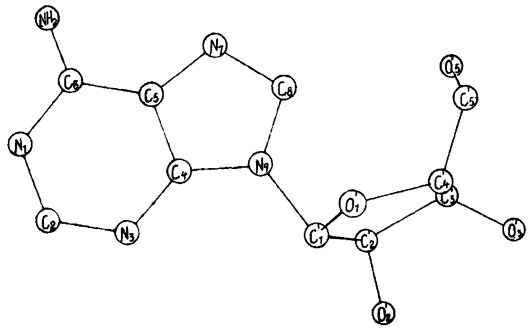


Fig. 4. The structure proposed for adenosine.

dine-like" structure is shown in Fig. 4. Here N_9—C_1' lies in the plane of the purine and forms tetrahedral angles with the adjacent ring bonds in the *d*-ribose; the two rings, far from being parallel, are nearly perpendicular to one another. Although other arrangements can not be definitely ruled out, this structure appears to give the most satisfactory explanation of the available data. Firstly, it is in agreement with the X-ray evidence, which indicates a rather flat ($c = 4.82$ Å) and widely spread-out molecule. The structure in Fig. 4 is about as flat as the molecule can be provided normal bond angles, and sufficiently extended. Secondly, it is correct from a chemical point of view, being adenine-9-β-*d*-ribofuranoside, whereas the sugar in the structure proposed by Hendricks actually is α-lyxose, not *d*-ribose, as pointed out by Gulland and coworkers [11]. Thirdly, Hendricks' results can be reinterpreted in terms of the proposed structure, in which also two parallel planes about 1.5 Å apart can be roughly recognised, the one containing the purine plus the atoms C_1', C_4' and C_5', the other passing through C_2', C_3' and O_2'. The remaining atoms do not lie in any of these planes, but none of them is far out. The structure appears thus to be consistent with the measurements of Hendricks.

In cytidine we found some evidence for a weak intramolecular bond, supposedly due to the polarisation of a (CH) group. There is no reason for assuming such a bond in adenosine, where the only substituent in the purine ring is far away from the group concerned, $(CH)_8$, actually in another ring.

GUANOSINE

Crystallographical. The crystals of guanosine are monoclinic, elongated along the *c*-axis, and bounded by {110}, {010} and {100}, the last form dominating. Their shape indicates monoclinic sphenoidal symmetry (class 2). The cell

dimensions are: $a = 17.45$ Å, $b = 11.44$ Å, $c = 6.65$ Å, $\beta = 98°$ (approx.). Cell volume: 1315 Å^3. Density: 1.60 g/cm^3. Guanosine crystallises with two molecules of water, and there are four (calc. 3.97) molecules $C_{10}H_{13}O_5N_5 \cdot 2H_2O$ in the unit cell.

The $0k0$-reflexions are absent for odd values of k, and the space group is $P\,2_1$. The asymmetric unit in the crystal consists of two molecules of guanosine and four molecules of water.

Stereochemical. No conclusion regarding the shape of the guanosine molecule can be drawn from these preliminary data. Also the guanosine ion was, however, investigated by Hendricks, who found it to be very similar to the adenosine ion. It appears therefore reasonable to assume that our discussion of the stereochemistry of adenosine is valid also for guanosine.

THE DEOXYRIBONUCLEOSIDES

The deoxyribonucleosides differ chemically from the ribonucleosides in that they contain 2-deoxy-*d*-ribose instead of *d*-ribose. No crystals of these compounds have been investigated, but the results from the study of cytidine and the other ribonucleosides can probably be applied to them. In cytidine the hydroxyl group $(OH)_2'$ is at a distance of as much as 3.75 Å from the keto group in the pyrimidine, and is unlikely to play any important role in determining the mutual orientation of the carbohydrate and the pyrimidine. The absence of $(OH)_2'$ should therefore not alter significantly the forces which act between the two ring systems, and the perpendicular relationship between them would appear likely to be maintained also in the case of the deoxyribonucleosides.

THE NUCLEIC ACIDS

Astbury [3] has on basis of X-ray studies of fibres and powders of nucleic acids suggested that these molecules are presumably stiff columns of nucleosides fitting closely on top of one another, with all the rings flat and parallel. In the present investigation it is established that the purine or pyrimidine is not parallel, but perpendicular to the ribose ring at least in one of the nucleosides, namely cytidine, and probably also in all the others. It appears likely that the nucleosides maintain their characteristic shape when they unite to form nucleic acids, and Astbury's theory is therefore in need of some modification. A full discussion of the problem is postponed pending the completion of the structure of a nucleotide, but one point will be made here. There is

evidence that the purines and the pyrimidines are nearly parallel to one another, and presumably *perpendicular* to the axis of the rod-like molecule, at least in thymonucleic acid. If this is the case, the ribose rings will consequently lie with their planes more or less *parallel* to the axis of the molecule, and it can be shown by models that a dominating spacing of about 3.4 Å, as found by Astbury, can be maintained in the molecule also under these circumstances.

EXPERIMENTAL

Considerable difficulties were encountered in getting single crystals big enough for X-ray work, and several techniques of growing crystals were tried. For cytidine the best results were obtained by slow evaporation of a solution in alcohol-water, and for adenosine by slow cooling of a concentrated solution in water. In the case of uridine and guanosine, crystals from the supplied samples could be used.

Density measurements were made by flotation of the crystals in a mixture of carbon tetrachloride and benzene.

The cell dimensions were derived from oscillation photographs, and are believed to be correct to at least within 1 %. Copper radiation ($\lambda = 1.54$ Å) is used throughout the investigation.

In the case of cytidine and adenosine also Weissenberg photographs were taken. The intensities were estimated visually and corrected for the Lorentz-polarisation factor in the usual way.

SUMMARY

Single crystals of cytidine, uridine, adenosine and guanosine have been studied by X-ray diffraction methods. The complete crystal structure of *cytidine* has been determined by two-dimensional Fourier syntheses, showing directly that cytidine is cytosine-3-β-*d*-ribofuranoside. The planar pyrimidine ring and the non-planar ribose ring are linked in such a way that they are nearly perpendicular to each other; the central bond N_3—C_1' lies in the plane of the pyrimidine ring. Preliminary investigations of the other three ribonucleosides, supported by other physical data, indicate that these molecules have a shape similar to that of cytidine, and it is considered likely that this holds also for the deoxyribonucleosides.

My best thanks are due to Prof. J. D. Bernal, F. R. S., and Dr. C. H. Carlisle for their continued interest and encouragement, to Prof. A. R. Todd, F. R. S., for an interview, and to Dr. D. O. Jordan and Dr. R. J. C. Harris for supplying the specimens. I also wish to thank The British Council for the award of a scholarship.

REFERENCES

1. Levene, P. A., and Tipson, R. S. *J. Biol. Chem.* **94** (1932) 809; **101** (1933) 529.
2. Davoll, J., Lythgoe, B., and Todd, A. R. *J. Chem. Soc.* (1946) 833.
3. Astbury, W. T. *Symp. Soc. Exptl. Biol.* **1** (1947) 66.
4. Hendricks, S. B. *J. Phys. Chem.* **45** (1941) 65.
5. Pitt, G. *Acta Cryst.* **1** (1948) 168.
6. Clews, C. J. B., and Cochran, W. *Acta Cryst.* **2** (1949) 46.
7. Broomhead, June M. *Acta Cryst.* **1** (1948) 324.
8. Beevers, C. A., and Cochran, W. *Proc. Roy. Soc.* **A 190** (1947) 257.
9. Furberg, S. *Acta Cryst.* In press.
10. Gulland, J. M., Holiday, E. R., and Macrae, T. F. *J. Chem. Soc.* (1934) 1639.
11. Gulland, J. M., Barker, G. R., and Jordan, D. O. *Nature* **151** (1943) 109.

Received April 1, 1950.