Acta Cryst. (1965). **19**, 797

The Crystal and Molecular Structure of Isocytosine*

BY BRAHAMA D. SHARMA† AND J. F. MCCONNELL‡

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California 91109, *U.S.A.*

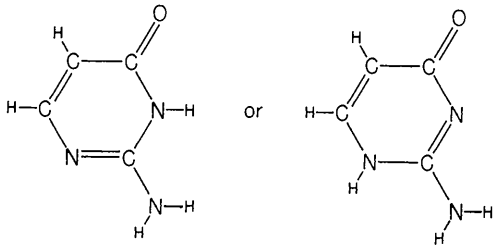
(*Received* 23 *February* 1965)

The crystal structure of the pyrimidine isocytosine, $C_4H_5N_3O$, has been determined and refined by the analysis of complete three-dimensional intensity data, collected about all three crystallographic axes with the use of Cu X-radiation. The crystals are monoclinic with space group $P2_1/n$; the unit-cell dimensions are $a=8{\cdot}745$, $b=11{\cdot}412$, $c=10{\cdot}414$ Å, $\beta=94{\cdot}79°$. There are eight molecules in the unit cell and hence two in the asymmetric unit. The refinement of the positional parameters of all 26 independent atoms, of the individual anisotropic temperature factors for the 16 heavy atoms, and of the individual isotropic temperature factors for the 10 hydrogen atoms was by the method of full-matrix least-squares. For 1826 observed reflections of non-zero weight, the final R index is 0·061.

The crystals of isocytosine consist of two distinct chemical entities, which are tautomers of each other, in an exact 1:1 ratio. These tautomers are hydrogen bonded to one another in a manner analogous to that proposed for guanine and cytosine pairing in deoxyribonucleic acid; the three hydrogen bonds, N(7*B*)–H · · · O(8*A*), N(3*B*)–H · · · N(3*A*), O(8*B*) · · · H–N(7*A*) are 2·861, 2·908, and 2·904 Å in length. The base pairs of isocytosine are further linked to one another *via* hydrogen bonds of the type N–H · · · N, 2·980 Å in length, around crystallographic centers to give discrete, nearly planar, tetramers of molecules of isocytosine. Each base pair of isocytosine is parallel to, and separated by a distance of 3·36 Å from, another base pair, so as to give packs of two π-interacting base pairs of isocytosine.

Introduction

The crystal structure of isocytosine (I)



(I)

was undertaken as a part of a program of research on the structure of compounds related to nucleic acids. Isocytosine can exist in two tautomeric forms shown in (I) and has the same arrangement of heavy atoms as does the six-membered ring of guanine; this six-membered ring plays an important role in the proposed pairing, through hydrogen bonds, of guanine and cytosine in deoxyribonucleic acid (Watson & Crick, 1953*a*,*b*; Pauling & Corey, 1956). In the absence of structural information on the neutral molecule of guanine (which is extremely difficult to crystallize), it seemed important to examine the crystals of isocytosine.

A preliminary report (McConnell, Sharma & Marsh, 1964) of this investigation has already been presented. The present report describes the details of the structure analysis, including further refinement based on additional data, and discusses the unusual features of the existence of two tautomers of isocytosine in an exact 1:1 ratio.

Experimental

Pure isocytosine was obtained in crystalline form from Mann Research Laboratories (their catalog No. 2514). The crystals are acicular, elongated along the *a* direction. The space group was determined from oscillation and zero- and first-layer Weissenberg photographs about the *a*, *b*, and *c* axes. The unit-cell dimensions, given in Table 1, were determined from least-squares analysis of Straumanis-type zero-layer Weissenberg photographs about the *a* and *b* axes. The density was measured by the flotation method in a mixture of ethyl bromide and benzene.

Table 1. *Crystal data for isocytosine*

Crystal system: monoclinic. Systematic absences: $h0l$ with $h+l=2n+1$ and $0k0$ with $k=2n+1$. Space group: $P2_1/n$.

$a=8{\cdot}745\pm1$ Å,	$b=11{\cdot}412\pm2$ Å, (λ Cu $K\alpha=1{\cdot}5418$ Å)	$c=10{\cdot}441\pm1$ Å
$\beta=94{\cdot}79\pm1°$	$V=1038$ Å^3	$Z=8$
$D_m=1{\cdot}403$ g.cm^{-3}	$D_x=1{\cdot}421$ g.cm^{-3}	

Complete intensity data for Cu $K\alpha$ radiation were collected by the multiple film (three packs of three films each) equi-inclination Weissenberg technique. Layer

* Contribution No. 2921 from the Gates and Crellin Laboratories of Chemistry. This investigation was supported, in part, by Research Grant H-2143 from the National Heart Institute, Public Health Service to the California Institute of Technology.

† Present address: Department of Chemistry, Oregon State University, Corvallis, Oregon, U.S.A.

‡ Present address: School of Physics, University of New South Wales, Sydney, Australia.

lines 0–7 around the *a* axis were collected from a needle-shaped crystal about $0{\cdot}2\times0{\cdot}2$ mm in cross-section. This crystal was then demounted and cut under a drop of amyl acetate to form a rhomb with dimensions $0{\cdot}2\times0{\cdot}2\times0{\cdot}2$ mm from which layer lines 0–8 around the *c* axis and 0–6 around the *b* axis were collected. By this means the entire effective copper diffraction sphere was surveyed; of a total of 2218 reflections within the sphere, 1848 were of measurable intensity. All intensities were measured visually.

The standard deviation, $\sigma(I)$, for each intensity reading was estimated according to the following expression:

$$\sigma(I)=[I_{\min}/3+BI+0{\cdot}1I^2/(I_{\max}-I)^2]\times [1+0{\cdot}25\exp\{-50{\cdot}0(0{\cdot}5-\sin^2\theta)^2\}]\times 1/W_{\text{ext}}$$

(Duchamp & Marsh, 1964). Here, $I_{\min}$ and $I_{\max}$ represent the limits of the reliable portion of the intensity strip; the values chosen were 3·3 and 40·0. The constant *B*, determined empirically for each axis from a statistical analysis of the intensity readings of all reflections measured on more than one film, had the values 0·062, 0·068, and 0·112 for the *a*-, *c*-, and *b*-axis data. The exponential term reflects the higher uncertainties of measurements made in the region where $\alpha_1-\alpha_2$ splitting begins to occur.

The external weight, W_{ext}, was taken as unity except for those intensity readings which were subjectively judged to be poor. The intensities and their standard deviations were corrected for film factor, pack factor, Lorentz-polarization factor, and inter-layer scaling factor (Duchamp, Gramaccioli & Marsh, 1964) to yield values of F^2 and $\sigma(F^2)$, on an arbitrary scale, for all reflections. No absorption correction was applied.

The preliminary report (McConnell, Sharma & Marsh, 1964) was based on intensity data measured on the *a*- and *c*-axis photographs only.

Description of the structure

Immediately after the intensity data around the *a* and *c* axes had been collected and scaled, sharpened and unsharpened three-dimensional Patterson functions were calculated. The vector peaks around the origin in the sharpened Patterson map defined a six-membered ring which provided the orientation of the pyrimidine nucleus and also indicated that the two molecules of the asymmetric unit are parallel to one another. Pronounced peaks in the unmodified and modified Patterson functions corresponding to inter-ring vectors were easily located. One conspicuous vector of length 3·6 Å and oriented almost normal to the plane of the pyrimidine ring suggested a stacking of the molecules; another vector, of length 5·8 Å, lying in the plane of the pyrimidine ring was taken to arise from coplanar molecules hydrogen bonded together. Thus a model of the structure was constructed. A superposition map based on the six bond vectors of the pyrimidine ring was in agreement with the proposed model, which was improved upon by structure-factor and electron-density calculations based on the *h0l* data. Attention was then turned to the three-dimensional intensity data.

Refinement of the parameters

Initial least-squares refinement was carried out on a Burroughs 220 computer. The proposed model, with individual anisotropic temperature factors for the 16 heavy atoms and a fixed isotropic temperature factor and assumed coordinates for the hydrogen atoms, refined steadily to an *R* index of 0·068.

At this stage the *b*-axis photographs were measured and the entire set of intensity data were processed in the manner described in the experimental section; these and all subsequent computations were carried out with

Table 2. *Positional and thermal parameters for the heavy atoms*

The temperature factors are in the form $T_i=\exp(-B_{11}h^2-B_{22}k^2-B_{33}l^2-B_{12}hk-B_{13}hl-B_{23}kl)$. All parameters have been multiplied by 10^4.

Atom	$x(\sigma_x)$	$y(\sigma_y)$	$z(\sigma_z)$	$B_{11}(\sigma_{B_{11}})$	$B_{22}(\sigma_{B_{22}})$	$B_{33}(\sigma_{B_{33}})$	$B_{12}(\sigma_{B_{12}})$	$B_{13}(\sigma_{B_{13}})$	$B_{23}(\sigma_{B_{23}})$
Molecule *A*									
N(1*A*)	7724(2)	8326(2)	3948(2)	111(2)	96(2)	114(2)	−56(3)	77(3)	34(3)
C(2*A*)	6861(2)	7350(2)	3709(2)	99(3)	81(2)	90(2)	−3(3)	61(4)	31(3)
N(3*A*)	5667(2)	7102(1)	4376(1)	101(2)	71(1)	85(2)	−23(2)	65(3)	1(2)
C(4*A*)	5274(2)	7852(2)	5309(2)	101(3)	74(1)	84(2)	−8(3)	44(3)	5(3)
C(5*A*)	6204(2)	8875(2)	5582(2)	145(3)	79(2)	111(2)	−48(3)	55(4)	−18(3)
C(6*A*)	7379(3)	9086(2)	4883(2)	151(3)	86(2)	121(3)	−75(3)	46(5)	7(3)
N(7*A*)	7217(2)	6630(2)	2787(2)	159(3)	106(2)	120(2)	−46(4)	172(4)	−24(3)
O(8*A*)	4120(2)	7636(1)	5895(1)	122(2)	97(1)	114(2)	−54(2)	115(3)	−52(2)
Molecule *B*									
N(1*B*)	1166(2)	4282(2)	3863(2)	104(2)	88(1)	107(2)	−35(3)	50(3)	−28(3)
C(2*B*)	2168(2)	5132(2)	4181(2)	94(2)	74(1)	87(2)	−13(3)	50(3)	−4(3)
N(3*B*)	3525(2)	5228(1)	3623(2)	99(2)	68(1)	90(2)	−16(2)	58(3)	−12(2)
C(4*B*)	3948(2)	4478(2)	2684(2)	116(3)	71(1)	90(2)	27(3)	58(4)	−4(3)
C(5*B*)	2881(2)	3566(2)	2349(2)	140(3)	80(2)	104(2)	5(3)	52(4)	−43(3)
C(6*B*)	1578(2)	3519(2)	2966(2)	129(3)	80(2)	113(2)	−22(3)	12(4)	−25(3)
N(7*B*)	1867(2)	5924(2)	5050(2)	129(3)	104(2)	129(2)	−80(3)	131(4)	−83(3)
O(8*B*)	5185(2)	4654(1)	2205(2)	136(2)	88(1)	131(2)	7(2)	133(3)	−27(2)

the IBM 7094–7040 crystallographic system CRYRM (Duchamp, 1964). The quantity minimized in the least-squares calculations was $\Sigma w(F_o^2-F_c^2/k^2)^2$, with weights w assigned to each observation taken as the inverse of the square of $\sigma(F^2)$. The form factors were adapted from *International Tables for X-ray Crystallography* (1962). Six full-matrix least-squares cycles, which included positional parameters for all 26 atoms, anisotropic temperature factors for the 16 heavy atoms, individual isotropic temperature factors for the 10 hydrogen atoms, and an overall scale factor as refinable parameters, were carried out. In the final cycle, the indicated shift in none of the 185 parameters was as large as 0·01 of its standard deviation. The final R index for 1826 observed reflections of non-zero weight is 0·061. The overall 'goodness of fit' $[\Sigma w(F_o^2-F_c^2/k^2)^2/m-s]^{\frac{1}{2}}$ (Peterson & Levy, 1957) is 1·5. This value is close to the expected value of unity; hence, we feel confident that convergence has been achieved and that the standard deviations, calculated from the residuals and the diagonal elements of the inverse matrix, are indeed reliable.

At the conclusion of the refinement, electron density and difference maps were evaluated in the planes of the two molecules *A* and *B*. These maps are shown in Figs. 1 and 2.

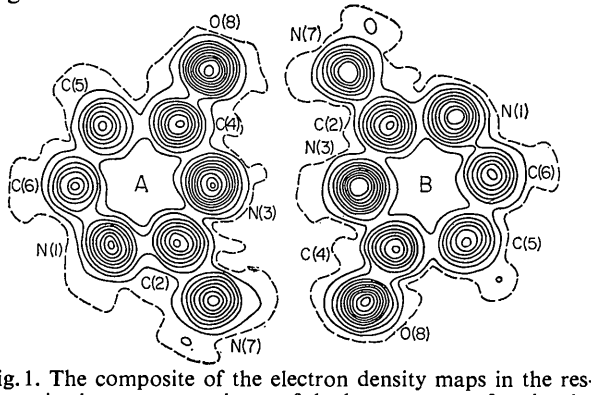


Fig. 1. The composite of the electron density maps in the respective least-squares planes of the heavy atoms of molecules *A* and *B*. Contours are at intervals of 1 e.Å^{-3} beginning with zero electron Å^{-3} (dashed).

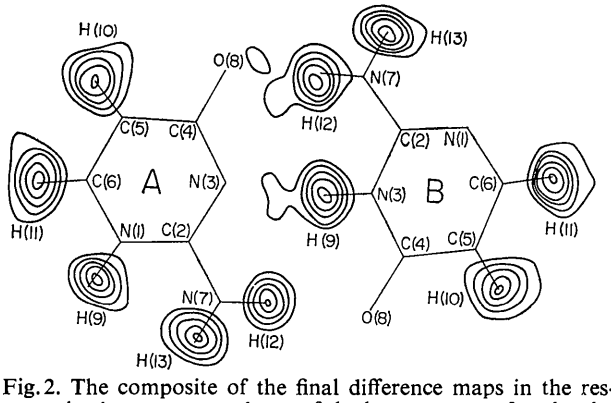


Fig. 2. The composite of the final difference maps in the respective least-squares planes of the heavy atoms of molecules *A* and *B*, in which the hydrogen contributions were omitted from the F_c's. Contours are at intervals of 0·1 e.Å^{-3} beginning with 0·1 e.Å^{-3}.

The final positional and temperature factor parameters of the heavy atoms are listed in Table 2, the parameters of the hydrogen atoms in Table 3, and the observed and calculated structure factors in Table 4.

Table 3. *Positional and thermal parameters for the hydrogen atoms*

All positional parameters have been multiplied by 10^3.

Atom	$x(\sigma)$	$y(\sigma)$	$z(\sigma)$	$B(\sigma)$
Molecule *A*				
H(9*A*)	853(3)	859(2)	344(2)	2·9(5) Å^2
H(10*A*)	595(3)	942(2)	619(2)	3·2(5)
H(11*A*)	811(2)	977(2)	500(2)	2·3(4)
H(12*A*)	660(3)	595(2)	262(3)	3·9(6)
H(13*A*)	797(3)	679(2)	230(2)	3·8(6)
Molecule *B*				
H(9*B*)	420(3)	586(2)	387(2)	2·4(5)
H(10*B*)	308(3)	303(2)	173(2)	2·7(5)
H(11*B*)	91(2)	285(2)	279(2)	2·0(4)
H(12*B*)	260(3)	644(2)	531(2)	2·8(5)
H(13*B*)	86(3)	582(2)	547(3)	4·1(6)

Discussion

The molecular structure of isocytosine

An outstanding and unexpected feature of the crystal structure of isocytosine is that it consists of two distinct, tautomeric, chemical entities, 1,4-dihydro-2-amino-4-oxo-pyrimidine (*A*) and 3,4-dihydro-2-amino-4-oxo-pyrimidine (*B*), in an exact 1:1 ratio.

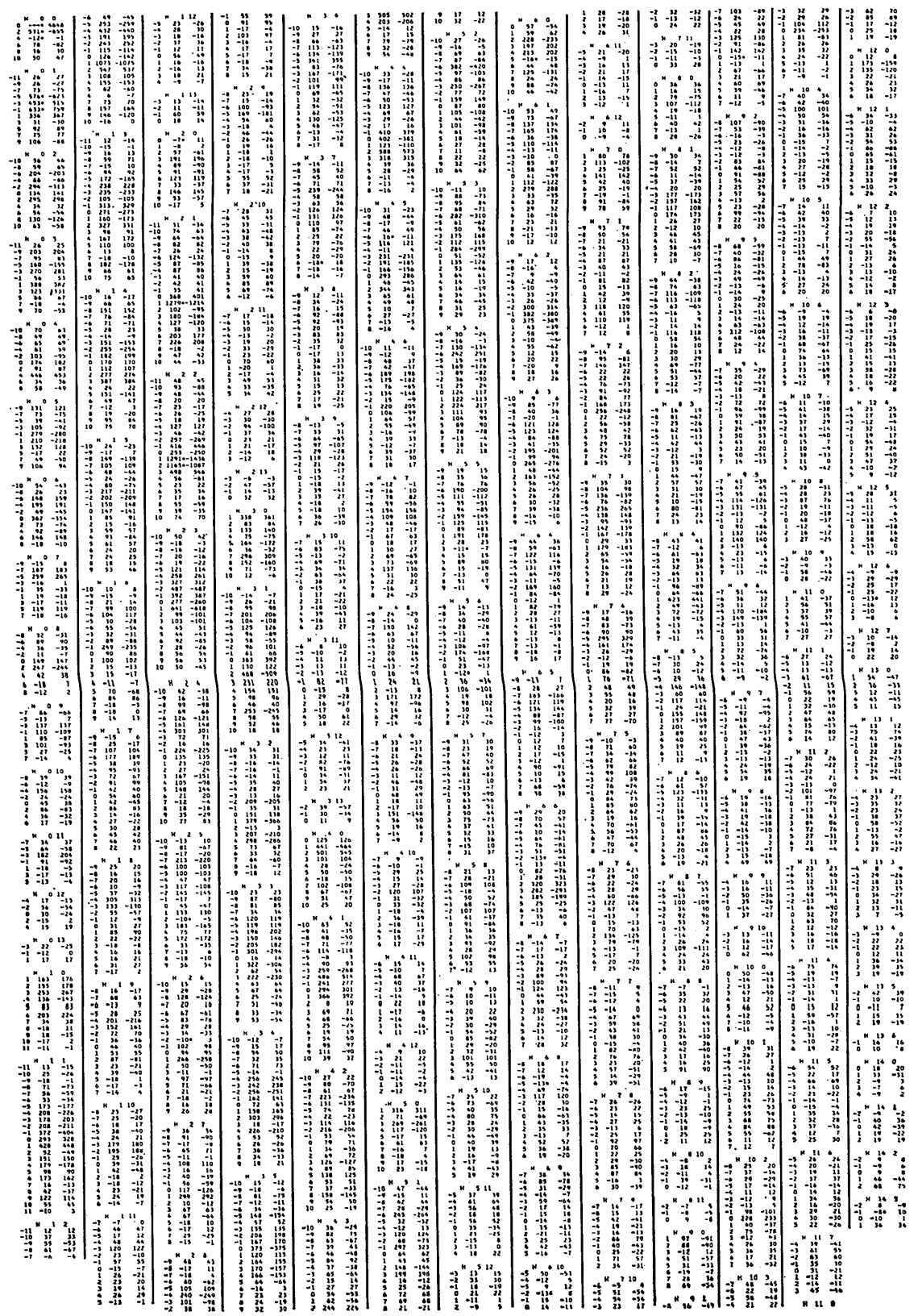
The bond distances and angles calculated from the parameters of Tables 2 and 3 are given in Fig. 3. The estimated standard deviations for the bond distances and angles involving heavy atoms are about 0·003 Å and 0·1°; for those involving hydrogen atoms, about 0·05 Å and 1·0°. In Fig. 4 are shown various canonical structures for the two tautomers; the observed distances can be explained satisfactorily by the indicated relative contributions of these canonical structures. The calculated distances shown were derived from empirical bond number–bond distance curves discussed by Marsh, Bierstedt & Eichhorn (1962).

The base pair

The unusual feature of the presence of two distinct, tautomeric, molecular species is further enhanced by the fact that these molecules are hydrogen bonded to one another in a manner analogous to that proposed by Watson & Crick (1953*a*, *b*), and modified by Pauling & Corey (1956), for the guanine–cytosine pair in deoxyribonucleic acid (DNA). This feature is apparent from Figs. 5 and 6, which show the crystal structure of isocytosine viewed along the *b* and *c* axes. The details of the hydrogen-bonded pair are given in Fig. 7. Unlike other base pairs (Hoogsteen, 1963*a*; O'Brien, 1963; Sobell, Tomita & Rich, 1963; Haschemeyer & Sobell, 1963; Mathews & Rich, 1964; Haschemeyer & Sobell, 1964) *the base pair of isocytosine was not obtained by deliberate co-crystallization of the two components*; presumably the two tautomers of isocytosine are of approx-

Table 4. *Observed and calculated structure factors*

Reading from left to right, the columns contain values of h, $10k|F_o|$, $10F_c$. A negative sign preceding $10|F_o|$ should be read 'less than'; an asterisk following $10k|F_o|$ signifies that the reflection was given zero weight in the final least-squares cycle.



imately equal stability and are present in appreciable amounts in solution at all times. The base pair of isocytosine can, in a sense, be considered a naturally occurring base pair.

The three hydrogen bonds within this base pair (see Fig. 7), N(7*B*)–H · · · O(8*A*), N(3*B*)–H · · · N(3*A*), and O(8*B*) · · · H–N(7*A*), are of lengths 2·861, 2·908, and 2·904 Å (±0·009 Å) and are approximately linear. In Table 5 these hydrogen bond lengths are compared with other experimental and predicted values of similar hydrogen bonds in 1:1 complexes of guanine and cytosine derivatives; this comparison gives more credence to the predicted values of Pauling & Corey (1956) than to those of Spencer (1959).

The geometry of the two tautomeric molecules and the base pair

The equations of the least-squares planes for the two individual molecules and for the base pair are given in Table 6. The molecules *A* and *B* are significantly non-planar, the non-ring atoms in both molecules being displaced to one side of the least-squares planes of the ring atoms. This displacement is more pronounced for ring *B* than for ring *A*.

The base pair is also non-planar, the two molecules being inclined at an angle of 9° to one another. The angles between the planes of the individual molecules *A* and *B* and the plane of the base pair are 6·5° and 3°; relative to the plane of molecule *B*, molecule *A* is rotated such that atoms N(7*A*) and N(7*B*) lie on one side of the plane of the base pair and the atoms O(8*A*) and O(8*B*) on the other.

The packing of the base pair in the crystal

An outstanding feature of the packing is the existence of packs of double base pairs; the least-squares planes of the base pairs within each pack are parallel to each other and separated by a distance of 3·36 Å. This

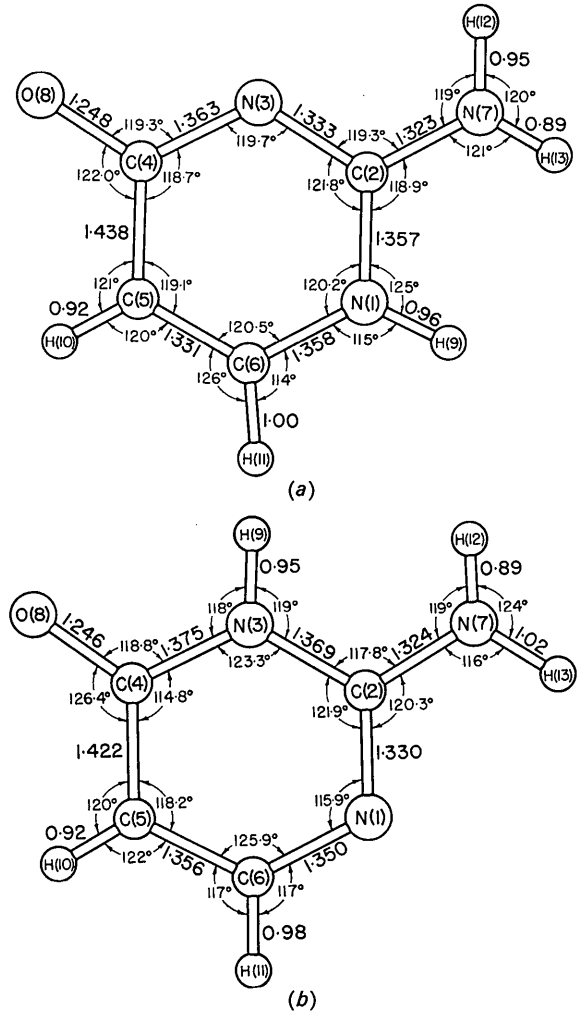


Fig. 3. Bond distances and angles for the tautomeric molecules (*a*) molecule *A* and (*b*) molecule *B* of isocytosine.

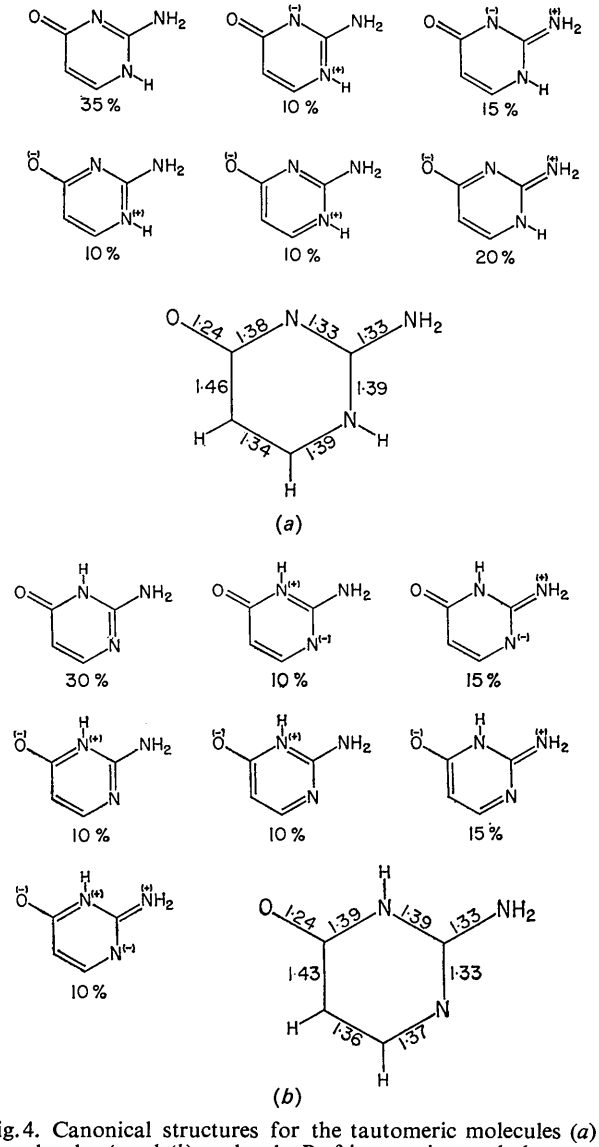
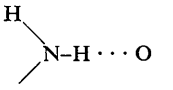
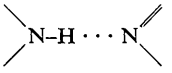
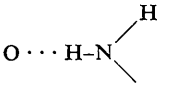


Fig. 4. Canonical structures for the tautomeric molecules (*a*) molecule *A* and (*b*) molecule *B* of isocytosine and the predicted bond distances.

Table 5. *Comparison of hydrogen bond lengths within the isocytosine base pair with other observed and predicted values of similar hydrogen bonds in* 1:1 *complexes of guanine and cytosine derivatives*

For isocytosine the hydrogen bonds from left to right are N(7*B*)–H · · · O(8*A*), N(3*B*)–H · · · N(3*A*), and O(8*B*) · · · H–N(7*A*). The molecule (*B*) of isocytosine is considered, from a tautomeric point of view, analogous to the generally assumed tautomeric structure of guanine.

Experimental values			
*Isocytosine (this work)	2·861 Å	2·908 Å	2·904 Å
*Cytosine-5-acetic Acid (Marsh *et al.*, 1962)	2·790	2·823	2·790
†O'Brien (1963)	2·82	2·91	2·93
†Sobell, Tomita, & Rich (1963)	2·91	2·95	2·86
†Haschemeyer & Sobell (1964)	2·78	2·91	2·82
Predicted values			
Pauling and Corey (1956)	2·93	2·96	2·93
Spencer (1959)	3·03	2·95	2·98

* The uncertainties in these lengths are ±0·01 Å.
† The uncertainties in these lengths were not quoted by the respective authors.

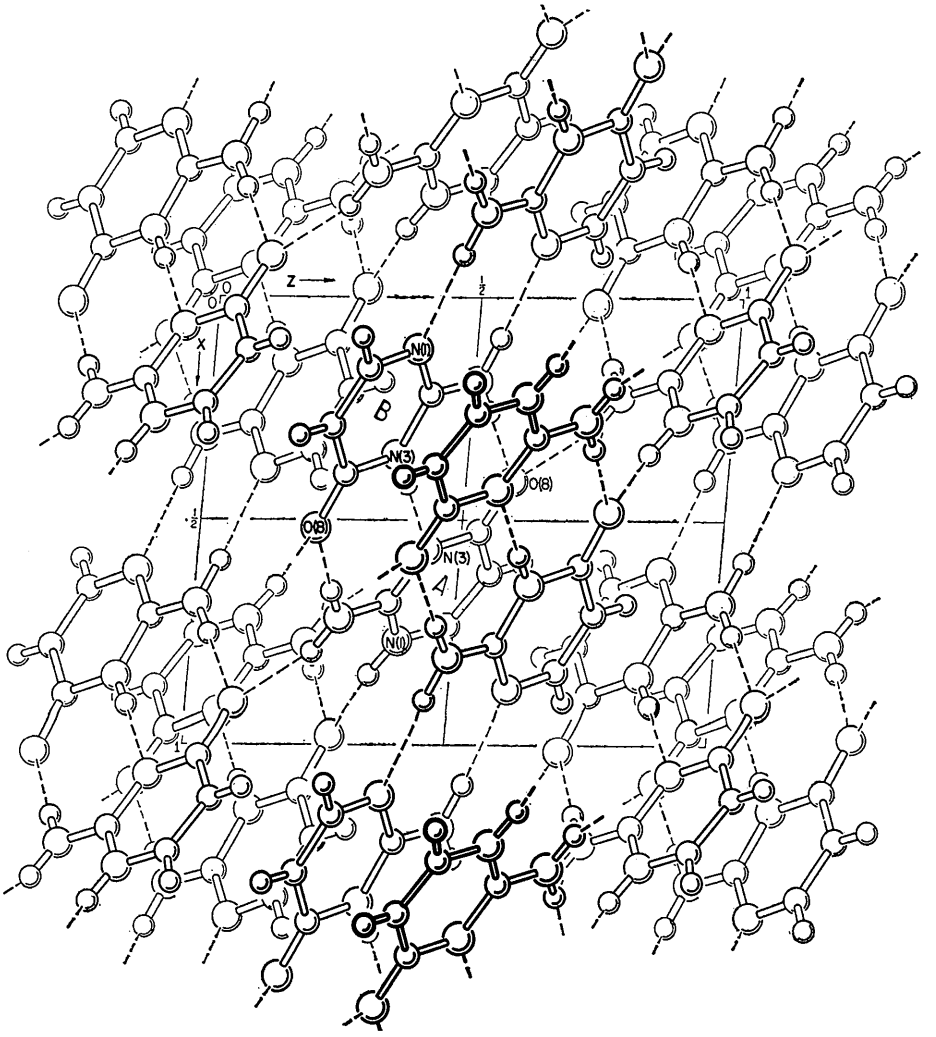


Fig. 5. The structure viewed along the *b* axis. The dashed lines represent hydrogen bonds.

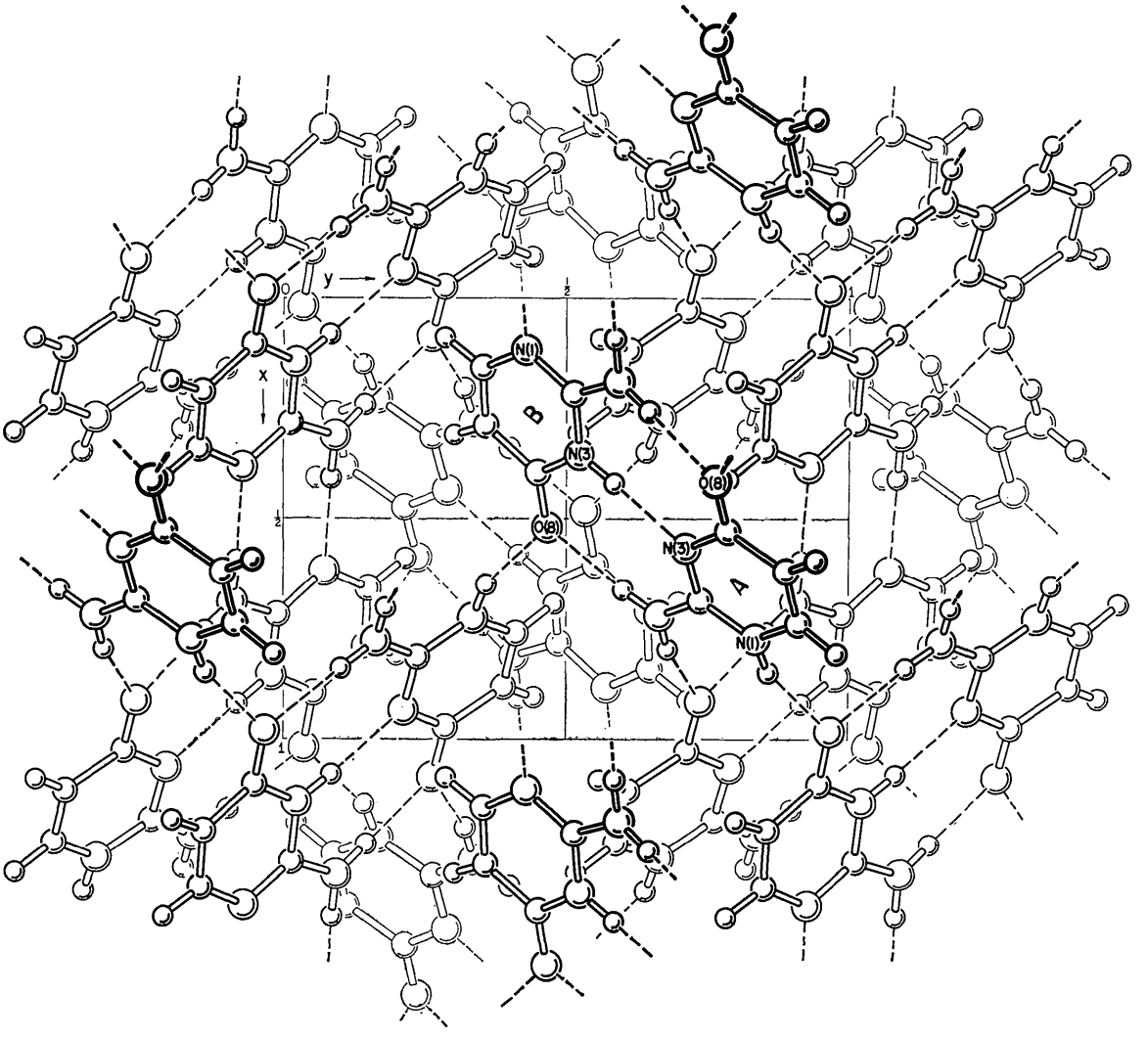


Fig. 6. The structure viewed along the *c* axis. The dashed line represents hydrogen bonds.

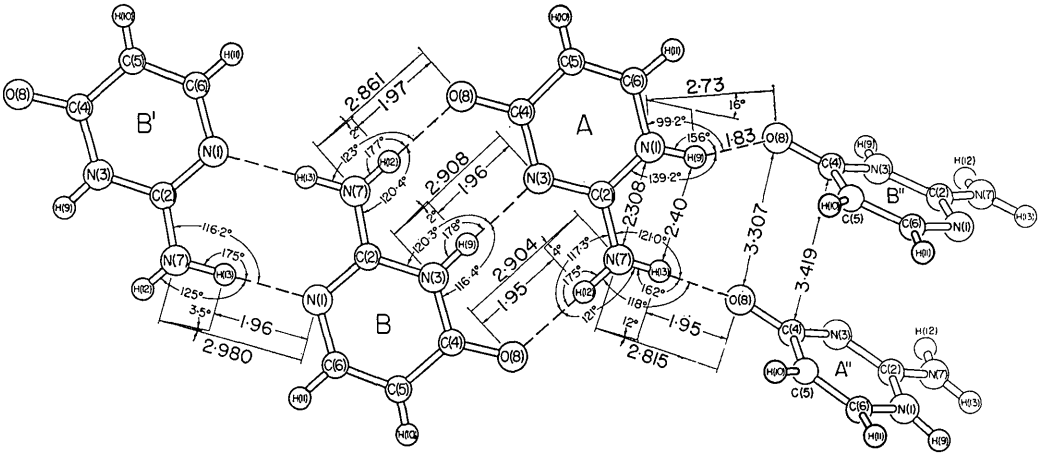


Fig. 7. Details of the hydrogen bonds within the base pair and those linking the base pairs to one another.

Table 6. *Least-squares planes and deviations for molecules A and B and the base pair comprising the two tautomeric molecules*

The orthogonal coordinate set for all the planes is (a, b, c^*)

Molecule *A*

Plane 1*A*, ring atoms (1–6):
$0{\cdot}5440X-0{\cdot}5094Y+0{\cdot}6666Z=1{\cdot}3873$ Å
Plane 2*A*, all heavy atoms (1–8):
$0{\cdot}5351X-0{\cdot}5141Y+0{\cdot}6703Z=1{\cdot}3087$ Å

Atom	Deviations from plane 1*A*	Deviations from plane 2*A*
N(1*A*)	−0·002 Å	−0·010 Å
C(2*A*)	0·000	0·003
N(3*A*)	0·007	0·024
C(4*A*)	−0·013	0·007
C(5*A*)	0·011	0·020
C(6*A*)	−0·004	−0·010
N(7*A*)	−0·007	−0·008
O(8*A*)	−0·058	−0·025
H(9*A*)	−0·10	−0·12
H(10*A*)	−0·04	−0·03
H(11*A*)	0·02	0·01
H(12*A*)	−0·02	−0·01
H(13*A*)	−0·06	−0·07

Molecule *B*

Plane 1*B*, ring atoms (1–6):
$0{\cdot}3954X-0{\cdot}5829Y+0{\cdot}7097Z=0{\cdot}2813$ Å
Plane 2B, all heavy atoms (1–8):
$0{\cdot}4004X-0{\cdot}5783Y+0{\cdot}7107Z=0{\cdot}3132$ Å

Atom	Deviations from plane 1*B*	Deviations from plane 2*B*
N(1*B*)	−0·008 Å	−0·010 Å
C(2*B*)	−0·003	0·005
N(3*B*)	0·010	0·023
C(4*B*)	−0·006	0·004
C(5*B*)	−0·004	−0·003
C(6*B*)	0·011	0·007
N(7*B*)	−0·022	−0·011
O(8*B*)	−0·032	−0·015
H(9*B*)	0·00	0·02
H(10*B*)	−0·02	−0·02
H(11*B*)	0·10	0·09
H(12*B*)	0·07	0·09
H(13*B*)	−0·01	−0·00

The hydrogen bonded base pair

Plane *P*, all heavy atoms within the pair:
$0{\cdot}4464X-0{\cdot}5768Y+0{\cdot}6840Z=0{\cdot}3425$ Å

(The plane of the base pair at $[(1-x), (1-y), (1-z)]$ is 3·7073 Å from the origin).

Atom	Deviations from plane *P*	Atom	Deviations from plane *P*
N(1*A*)	−0·151 Å	N(1*B*)	−0·107 Å
C(2*A*)	−0·007	C(2*B*)	−0·061
N(3*A*)	0·139	N(3*B*)	0·030
C(4*A*)	0·120	C(4*B*)	0·056
C(5*A*)	−0·007	C(5*B*)	0·015
C(6*A*)	−0·158	C(6*B*)	−0·047
N(7*A*)	−0·014	N(7*B*)	−0·115
O(8*A*)	0·205	O(8*B*)	0·102
H(9*A*)	−0·35	H(9*B*)	0·05
H(10*A*)	−0·06	H(10*B*)	0·02
H(11*A*)	−0·24	H(11*B*)	0·02
H(12*A*)	0·08	H(12*B*)	0·01
H(13*A*)	−0·16	H(13*B*)	−0·16

Table 6 (*cont.*)

The tetramer

Plane *T*, all heavy atoms within the tetramer; the atoms (1*A*–8*A*), (1*B*–8*B*), (1*A*′–8*A*′), and 1*B*′–8*B*′). The molecules *A*′ and *B*′ are related to *A* and *B* by a center at (0, ½, ½).
$0{\cdot}3967X-0{\cdot}5714Y+0{\cdot}7183Z=0{\cdot}3031$ Å

The deviations are given only for the atoms within the asymmetric part of the tetramer.

Atom	Deviations from plane *T*	Atom	Deviations from plane *T*
N(1*A*)	−0·239 Å	N(1*B*)	0·062 Å
C(2*A*)	−0·072	C(2*B*)	0·082
N(3*A*)	0·150	N(3*B*)	0·093
C(4*A*)	0·190	C(4*B*)	0·059
C(5*A*)	0·040	C(5*B*)	0·045
C(6*A*)	−0·188	C(6*B*)	0·064
N(7*A*)	−0·136	N(7*B*)	0·081
O(8*A*)	0·348	O(8*B*)	0·033
H(9*A*)	−0·491	H(9*B*)	0·094
H(10*A*)	0·029	H(10*B*)	0·015
H(11*A*)	−0·298	H(11*B*)	0·144
H(12*A*)	−0·027	H(12*B*)	0·184
H(13*A*)	−0·329	H(13*B*)	0·094

unusual feature is shown in Fig. 8 and is apparent from the views of the crystal structure shown in Figs. 5 and 6. The base pairs within each pack are arranged in such a manner as to bring heavy atoms of the same kind in close proximity; the existence of these packs can be ascribed to the possible π-interactions of the kind envisioned for adenine–thymine and guanine–cytosine base pairs in DNA and to the interaction between atoms N(3*B*) and N(3*A*) which, in view of the tautomeric nature of the molecule, have opposite charges. In Table 7 are given the shortest interatomic distances within such a pack of double base pairs. None of these distances is shorter than the expected van der Waals contact.

Table 7. *Distances within the asymmetric part of the double base pair pack*

From atom in the molecule *B* at x, y, z	To atom in the molecule *A* or *B* at $[(1-x), (1-y), (1-z)]$	Distance
N(1*B*)	C(2*A*)	3·49 Å
N(1*B*)	N(3*A*)	3·56
C(2*B*)	C(2*A*)	3·65
C(2*B*)	N(3*A*)	3·45
C(2*B*)	N(7*A*)	3·75
N(3*B*)	N(3*B*)	3·74
N(3*B*)	N(3*A*)	3·42
C(4*B*)	N(3*A*)	3·55
C(4*B*)	C(4*A*)	3·42
C(4*B*)	O(8*A*)	3·24
C(5*B*)	N(3*A*)	3·63
C(5*B*)	C(4*A*)	3·25
C(5*B*)	C(5*A*)	3·58
C(5*B*)	O(8*A*)	3·37
C(6*B*)	C(2*A*)	3·76
C(6*B*)	N(3*A*)	3·59
C(6*B*)	C(4*A*)	3·53
C(6*B*)	C(5*A*)	3·61
C(6*B*)	C(6*A*)	3·79
N(7*B*)	O(8*B*)	3·75
N(7*B*)	N(7*A*)	3·73
O(8*B*)	N(7*B*)	3·75
O(8*B*)	O(8*A*)	3·31

The base pairs are further linked to each other through hydrogen bonds. Molecule *B* of each base pair is linked to a centrosymmetrically related molecule (*B'*, Fig. 7) by two, N(7*B*)–H · · · N(1*B'*) andN(1*B*) · · · H–N(7*B'*), hydrogen bonds of length 2·980 Å (inadvertently an incorrect value of 3·03 Å was quoted in the preliminary communication, McConnell, Sharma & Marsh, 1964). These hydrogen bonds give rise to discrete tetramers of isocytosine molecules in the solid. The least-squares plane of this tetramer and deviations of the atoms of the asymmetric part are given in Table 6.

Molecule *A* of each base pair acts as a donor for two additional hydrogen bonds, N(1*A*)–H · · · O(8*B''*) and N(7*A*)–H · · · O(8*A''*), of length 2·730 Å and 2·815 Å (±0·009 Å). These strong hydrogen bonds, which are shown in Fig. 7, deviate from linearity by 16° and 12·1° respectively; the non-linearity is dictated by the van der Waals separation of the two acceptor oxygen atoms. There are no unusual C–H · · · O interactions of the kind reported for other purine and pyrimidine crystals (Hoogsteen, 1963*b*; Sutor, 1963).

The hydrogen bonding and packing scheme discussed above reveals certain interesting features which can be useful in the overall study of the secondary structure of DNA. Each pack of double base pairs is a part of two, one left handed and the other right handed, spirals; the spiral axes are the crystallographic twofold screw axes. Molecules *A* and *B* are hydrogen bonded to one another on the surface of these spirals. This spiral-like structure consisting of π-interacting base pairs is reminiscent of the general features of the proposed packing of adenine-thymine and guanine–cytosine base pairs in the DNA molecule; however, the distance between successive base pair layers is smaller here than in DNA. The normal to the plane of the base pairs of isocytosine makes an angle of 60° to the spiral axes, which is larger than any proposed inclination of base pairs in DNA and ribonucleic acid.

The temperature factors

The magnitudes and direction cosines of the principal axes of thermal motion, as derived from the parameters of Table 2, are listed in Table 8. In general the minimum thermal motion is along the *a* direction, nearly the direction of hydrogen bonds responsible for the formation of discrete tetramers, and the maximum motion is in the direction normal to the molecular plane. In view of the three-dimensional hydrogen-bonding scheme a rigid-body analysis of the thermal motion was not carried out.

Table 8. *The magnitudes and the direction cosines of the principal axes of the thermal ellipsoids*

The direction cosines are with respect to *abc** and have been multiplied by 10^4.

Molecule *A*

Atom	Axis, *i*	B_i(Å^2)	g_i^a	g_i^b	g_i^{c*}
N(1*A*)	1	5·84	−1376	7970	5882
	2	5·28	5733	−4201	7035
	3	2·08	8077	4340	−3991
C(2*A*)	1	4·88	2242	6930	6852
	2	3·76	−5312	6764	−5102
	3	2·32	8170	2496	−5198
N(3*A*)	1	4·36	5725	−3802	7264
	2	3·68	−30	8850	4656
	3	2·24	8199	2687	−5056
C(4*A*)	1	3·96	2545	6500	7160
	2	3·88	−4247	7403	−5212
	3	2·68	8688	1714	−4645
C(5*A*)	1	5·72	5901	−5099	6259
	2	4·24	−4035	4853	7757
	3	3·28	6993	7103	−807
C(6*A*)	1	6·04	−7180	6610	−2180
	2	5·28	−419	2716	9615
	3	2·92	6948	6995	−1673
N(7*A*)	1	7·92	6175	−3845	6862
	2	5·12	1876	9192	3463
	3	2·04	−7639	−851	6397
O(8*A*)	1	7·20	−4514	5918	−6679
	2	3·92	2495	8023	5423
	3	2·32	8567	781	−5098

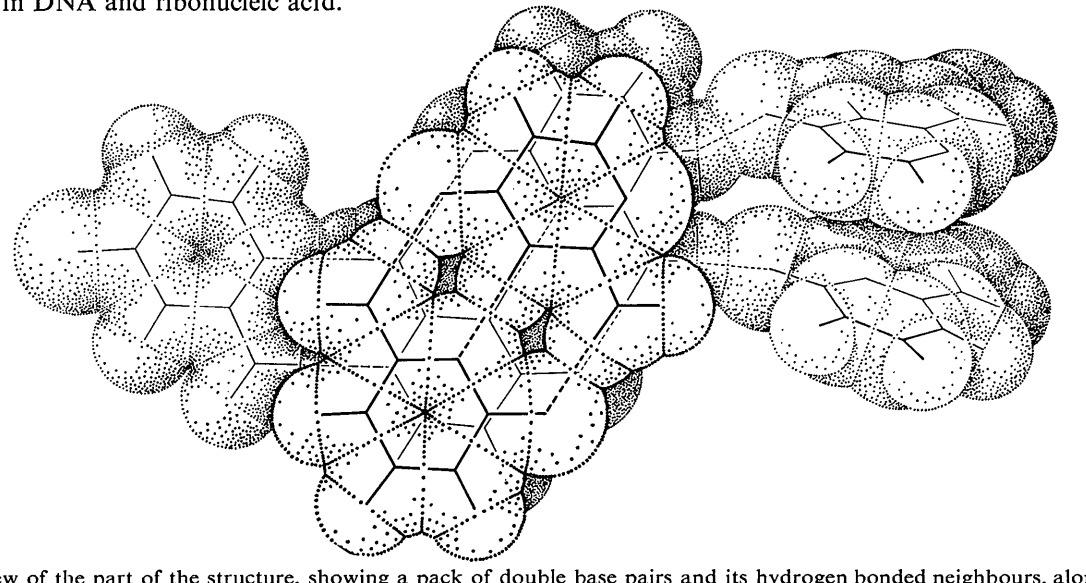


Fig. 8. A view of the part of the structure, showing a pack of double base pairs and its hydrogen bonded neighbours, along the normal to the least-squares plane of the heavy atoms of molecule *B*.

Table 8 (*cont.*)

Molecule *B*

Atom	Axis, *i*	B_i(Å^2)	$g_i{}^a$	$g_i{}^b$	$g_i{}^{c*}$
N(1*B*)	1	5·52	−3137	6679	−6749
	2	3·92	−671	6934	7174
	3	2·76	9472	2703	−1727
C(2*B*)	1	4·16	4044	−5255	7486
	2	3·76	1109	8406	5302
	3	2·48	9079	1314	−3982
N(3*B*)	1	4·44	4488	−3870	8055
	2	3·40	653	9131	4024
	3	2·48	8913	1280	−4350
C(4*B*)	1	4·52	6300	3512	6926
	2	3·88	1310	8311	−5405
	3	2·64	7654	−4313	−4776
C(5*B*)	1	5·48	2512	−5796	7752
	2	4·32	8630	4968	918
	3	3·04	−4383	6460	6250
C(6*B*)	1	5·24	243	−4964	8678
	2	4·32	7556	−5592	−3411
	3	3·48	6546	6640	3615
N(7*B*)	1	8·64	−4293	6084	−6675
	2	3·60	2007	7849	5862
	3	2·40	8806	1177	−4591
O(8*B*)	1	6·96	5045	−1864	8430
	2	4·64	2999	9535	314
	3	2·44	8096	−2370	−5369

Isocytosine and nucleic acids

Earlier physico-chemical studies of isocytosine, such as the spectral studies in the solid (Stimson & O'Donnell, 1952) and dissociation constants studies in solution (Levene, Bass & Simms, 1926), were based on the assumption that one tautomer predominates over the other. Similar assumptions have also been made for other purine and pyrimidine derivatives. The fact that both tautomers of isocytosine exist in an exact 1:1 ratio in the solid not only suggests further studies on isocytosine but also points out the necessity of a careful probe of the phenomenon of tautomerism in other purine and pyrimidine derivatives, and of a consideration of *p*H in formulating structures of nucleic acids.

A discussion of other possible base pairs and the process of replication and mutation of nucleic acids in the light of the crystal structure of isocytosine will be presented elsewhere.

We are grateful to Dr Richard E. Marsh for ever available counsel. We also wish to express our sincere appreciation to Professor Robert B. Corey for his continued interest. Help of (Miss) Lillian Casler in the preparation of the drawings is gratefully acknowledged.

References

DUCHAMP, D. J. (1964). A.C.A. Meeting, Bozeman, Montana, Paper B-14, p. 29.

DUCHAMP, D. J. & MARSH, R. E. (1964). Private communication.

DUCHAMP, D. J., GRAMACCIOLI, C. M. & MARSH, R. E. (1964). Private communication.

HASCHEMEYER, A. E. V. & SOBELL, H. M. (1963). *Proc. Nat. Acad. Sci. Wash.* **50**, 872.

HASCHEMEYER, A. E. V. & SOBELL, H. M. (1964). *Nature, Lond.* **202**, 969.

HOOGSTEEN, K. (1963*a*). *Acta Cryst.* **16**, 907.

HOOGSTEEN, K. (1963*b*). *Acta Cryst.* **16**, 28.

International Tables for X-ray Crystallography (1962). Vol. III, p. 202, Birmingham: Kynoch Press.

LEVENE, P. A., BASS, L. W. & SIMMS, H. S. (1926). *J. Biol. Chem.* **70**, 229.

MCCONNELL, J. F., SHARMA, B. D. & MARSH, R. E. (1964). *Nature, Lond.* **203**, 399.

MARSH, R. E., BIERSTEDT, R. & EICHHORN, E. L. (1962). *Acta Cryst.* **15**, 310.

MATHEWS, F. S. & RICH, A. (1964). *J. Mol. Biol.* **8**, 89.

O'BRIEN, E. J. (1963). *J. Mol. Biol.* **7**, 107.

PAULING, L. & COREY, R. B. (1956). *Arch. Biochem. Biophys.* **65**, 164.

PETERSON, S. W. & LEVY, H. A. (1957). *Acta Cryst.* **10**, 70.

SOBELL, H. M., TOMITA, K. & RICH, A. (1963). *Proc. Nat. Acad. Sci. Wash.* **49**, 885.

SPENCER, M. (1959). *Acta Cryst.* **12**, 66.

STIMSON, M. M. & O'DONNELL, M. J. (1952). *J. Amer. Chem. Soc.* **74**, 1805.

SUTOR, D. J. (1963). *J. Chem. Soc.* p. 1105.

WATSON, J. D. & CRICK, F. H. C. (1953*a*). *Nature, Lond.* **171**, 737.

WATSON, J. D. & CRICK, F. H. C. (1953*b*). *Nature, Lond.* **171**, 964.