The Structure of an Intermolecular Complex between Cytosine and 5-Fluorouracil

Donald Voet and Alexander Rich

Contribution from the Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.
Received October 15, 1968

Abstract: Crystals of an intermolecular complex containing cytosine and 5-fluorouracil in 1:1 molar ratio have been isolated from an aqueous solution containing these components. The structure of the crystal has been determined using single-crystal X-ray diffraction techniques. The crystals belong to the space group $P\bar{1}$ with one molecule each of cytosine and 5-fluorouracil in the asymmetric unit. The two molecules are joined together by two hydrogen bonds to form a cyclic dimer. These cyclic dimers are in turn joined by hydrogen bonds across centers of symmetry to other such cyclic dimers to form infinite sheets so that the crystal is a layered structure. In each unit cell of this structure there is a centrosymmetric cavity that contains two disordered water molecules.

The properties of purines and pyrimidines are determined to a great extent by their ability to form hydrogen bonds. Hydrogen bonds between purines and pyrimidines are responsible for the specificity that is believed to be the basis of intracellular information transfer in biological systems. In the double helical form of DNA, adenine forms hydrogen bonds with uracil and guanine with cytosine. This specificity of hydrogen bonding can be seen both in the solid state and in solution. A large number of intermolecular complexes have been formed between purines and pyrimidines in which specific interactions are found that are related to those occurring in the double helical form of DNA. Several adenine derivatives have been crystallized with uracil or thymine derivatives,[1-5] and likewise, derivatives of guanine and cytosine have been crystallized together.[6,7] Infrared[8,9] and proton nmr[10] studies in solution have shown the existence of selective hydrogen bonding between the naturally occurring purines and pyrimidines that have substituents on the glycosidic nitrogen. Thus, adenine derivatives have a high association constant for hydrogen bonding with uracil or thymine derivatives but do not form hydrogen bonds with guanine or cytosine derivatives. The latter pair, however, associate strongly with each other. This selective affinity has been described as being due to an electronic complementarity which is a property of the individual purine and pyrimidine molecules. However, it is of considerable interest to look for other kinds of interactions between the purines and pyrimidines. In the present work, we have discovered that it is possible to make a hydrogen-bonded one-to-one intermolecular complex between two pyrimidines, cytosine and 5-fluorouracil (see Figure 1). The structure of this complex has been solved by X-ray diffraction analysis. The molecules crystallize in a layer structure involving a complex system of hydrogen bonding. The lattice is somewhat unusual in that it contains a cavity that is partially filled with water molecules in a disordered state.

Experimental Section

Materials. 5-Fluorouracil was obtained as a gift of the Hoffman La Roche Co. The cytosine was purchased from the Cyclo Chemical Co. Crystals of the intermolecular complex were obtained by preparing an aqueous solution of the two materials in 1:1 molar ratio and allowing this solution to evaporate to dryness at room temperature. Only one class of needle-shaped crystals appeared in the beaker, suggesting that the two materials had cocrystallized. Single crystals were isolated from the beaker and dissolved in water. The ultraviolet absorption spectrum of the resulting solution suggested that both cytosine and 5-fluorouracil were present in 1:1 molar ratio. The solution was also analyzed by paper chromatography using an aqueous phosphate–ammonium sulfate solvent, pH 6.8.[11] This analysis showed that both materials were in the solution in approximately 1:1 molar ratio.

A chemical analysis of the crystals, carried out by Dr. S. Nagy, showed that the best empirical formula that could be written for the crystals involved cytosine and 5-fluorouracil in 1:1 molar ratio plus an added molecule of water.

Anal. Calcd for $C_8N_5O_4FH_{10}$: C, 37.07; N, 27.02; F, 7.33; H, 3.89; O, 24.69. Found: C, 37.05; N, 27.40; F, 7.15; H, 3.77; O (by difference), 24.63.

Crystallographic Analysis. A suitable crystal, which measured 0.5 mm along its needle axis and 0.2 mm in cross-section, was mounted along the needle axis (*a* axis). The absence of any diffraction symmetry other than Friedel symmetry implied that the Bravais lattice of the crystal was triclinic. The unit cell parameters were determined from an *0kl* Weissenberg photograph that had been calibrated with superimposed Al powder lines and from *hk*0 and *h*0*l* precession photographs. The unit cell parameters were found to be $a = 4.29 \pm 0.02$, $b = 9.59 \pm 0.01$, and $c = 15.14 \pm 0.02$ Å; $\alpha = 111.8 \pm 0.2$, $\beta = 98.0 \pm 0.2$, and $\gamma = 101.3 \pm 0.2°$.

The bouyant density of the crystals in a CCl_4–$CHCl_3$ solution was found to be 1.594 g/cm³. The calculated density for the crystal, if it contains the complex cytosine–5-fluorouracil–water, with two such complexes in the unit cell, is 1.560 g/cm³.

Weissenberg intensity photographs were made using the multiple film technique. The levels *hkl* were collected for $0 \leqslant h \leqslant 3$ using Cu Kα ($\lambda = 1.5418$ Å) X-rays. Data for correlating the individual levels were recorded with timed exposures of precession photographs employing Mo Kα radiation ($\lambda = 0.7107$ Å). For this latter set of data the levels *hkl* for $0 \leqslant k \leqslant 2$ were collected. The intensities of the diffracted spots were estimated visually using a calibrated film strip that was made using the same crystal from which the intensity data were collected. Lorentz-polarization corrections were applied and the intensities from individual films within each level were placed on the same scale using the program

(1) K. Hoogsteen, *Acta Cryst.*, **16**, 907 (1963).
(2) L. Katz, K. Tomita, and A. Rich, *ibid.*, **21**, 754 (1966).
(3) K. Tomita, L. Katz, and A. Rich, *J. Mol. Biol.*, **30**, 545 (1967).
(4) H. M. Sobell, *ibid.*, **18**, 1 (1966).
(5) L. L. Labana and H. M. Sobell, *Proc. Natl. Acad. Sci. U. S.*, **57**, 459 (1967).
(6) H. M. Sobell, K. Tomita, and A. Rich, *ibid.*, **49**, 885 (1963).
(7) E. J. O'Brien, *Acta Cryst.*, **23**, 92 (1967).
(8) Y. Kyogoku, R. C. Lord, and A. Rich, *Science* **154**, 518 (1966).
(9) E. Küchler and J. Derkosch, *Z. Naturforsch.*, **21B**, 209 (1966).
(10) L. Katz and S. Penman, *J. Mol. Biol.*, **15**, 220 (1966).
(11) Pabst Laboratory Circular OR-10, Pabst Brewing Co., Milwaukee, Wis., 1956.

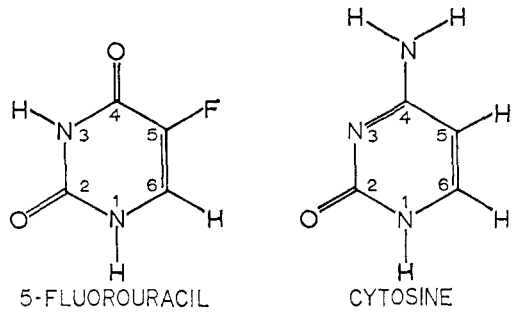

Figure 1.

XRDP1.[12] The intensities determined from the Weissenberg films were corrected for spot-shape distortion by the method of Phillips.[13] All the intensities were then placed on a common scale using the method of Simpson.[12] All computer calculations were performed using an IBM 7094 computer.

Structure Determination and Refinement

A total of 1730 unique reflections were collected. Of these, 1218 were considered intense enough to have been accurately estimated. The absolute scale factor and mean isotropic temperature factor were found by Wilson's method[14] to be $k = 1.080$ and $B = 1.84$, respectively. The statistics of the normalized structure factors,[15] normalized to $\langle E^2 \rangle = 1.00$, were found to be as shown in Table I. The results indicated that the crystal structure is centric; therefore the space group of the crystal is $P\bar{1}$. This result was subsequently confirmed by the successful refinement of the structure.

Table I

	Exptl	Centric	Acentric
$\langle E^2 - 1 \rangle$	1.050	0.968	0.736
$\langle \|E\| \rangle$	0.718	0.798	0.886
$\|E\| \geq 3.0$, %	0.99	0.30	0.01
$\|E\| \geq 2.0$, %	4.70	5.00	1.80
$\|E\| \geq 1.0$, %	27.54	32.00	37.00

Initial attempts were made to solve the structure by direct methods using the Σ_2 relationship.[16] This attempt was unsuccessful, probably due to the effect of the lack of symmetry in the unit cell in limiting the number of vectors with known phases. Accordingly, Patterson methods were employed to solve the crystal structure. The $1\bar{2}2$ reflection was by far the most intense reflection in the diffraction pattern. Its corresponding interplanar spacing of 3.45 Å suggested that the complex between cytosine and 5-fluorouracil crystallized in a layer structure because this spacing is very close to the thickness of the pyrimidine ring. The sharpened Patterson map with the origin partially removed was calculated. The $1\bar{2}2$ section of this map showed an extremely well-resolved and extensive hexagonal pattern that served to confirm our initial assignment of the $1\bar{2}2$ plane as the molecular plane.

The structure was solved by a consideration of the packing properties of the pyrimidines. If the pyrimidines crystallize with a center of symmetry, they often do so by forming infinite chains of hydrogen-bonded molecules that are organized into cyclic dimers across a center of symmetry. It also seemed reasonable to assume that the cytosine and 5-fluorouracil molecules were forming a hydrogen-bonded pair. There are eight possible ways of forming a hydrogen-bonded cyclic dimer between cytosine and 5-fluorouracil that incorporate the necessary centers of symmetry. From the Patterson map the orientation of the six-membered rings was evident. Accordingly, there are six different orientations of the structures which could be considered, making a total of 48 possible packing configurations. Only one of these possible structures was found to have the translational symmetry which is required for proper packing into the unit cell. From the coordinates of the atoms in this model the structure factors of a trial structure were calculated using the 12 ring atoms of the cytosine and 5-fluorouracil rings. In this initial calculation all atoms were assumed to be carbon atoms. The residual, $R = \Sigma||F_o| - |F_c||/\Sigma|F_o|$, of this initial structure factor calculation was 0.58. Two more iterations of Fourier map calculations brought out the remaining five cytosine and 5-fluorouracil atoms. When all 17 of these atoms were used in the structure factor calculation the value of the residual became $R = 0.44$.

The least-squares refinement of the structure was carried out employing the program ORFLS[17] using the Hughes weighting scheme.[18] The atomic scattering factors were taken from the International Tables for X-Ray Crystallography.[19] The 16 most intense reflections in the diffraction pattern were not used in the refinement because it was considered that their intensities could not be accurately estimated. In later stages of refinement those reflections with $\sigma(F_c)/F_o > 10$, 13 reflections in total, were deleted so that a total of 1189 reflections were used in the final refinement.[20] Two cycles of isotropic temperature refinement reduced the residual to $R = 0.34$. A difference Fourier map that was subsequently calculated showed a large diffuse peak, elongated in the direction parallel to the a axis, located between the molecular planes in an apparent cavity in the structure. This peak appeared to be caused by the presence of a water molecule. Accordingly, an oxygen atom, atom A, was added to the model of the structure being refined at the position of the maximum of this peak. Two more cycles of isotropic temperature refinement reduced the residual to $R = 0.24$. Two cycles of anisotropic temperature refinement then reduced the residual to $R = 0.17$. However, during the anisotropic temperature refinement the temperature factor, β_{11}, for oxygen atom A appeared to increase without limit. The difference Fourier map at this stage of refinement showed two poorly resolved peaks, one

(12) P. G. Simpson, Ph.D. Thesis, Harvard University, 1963.
(13) D. C. Phillips, *Acta Cryst.*, 7, 746 (1954).
(14) A. J. C. Wilson, *Nature*, 150, 152 (1942).
(15) J. Karle and H. Hauptman, *Acta Cryst.*, 9, 635 (1956).
(16) I. L. Karle and J. Karle, *ibid.*, 16, 969 (1963).
(17) W. R. Busing, K. O. Martin, and H. A. Levy, "A FORTRAN Crystallographic Least Squares Program" ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1962.
(18) E. W. Hughes, *J. Amer. Chem. Soc.*, 63, 1737 (1941).
(19) "International Tables for X-Ray Crystallography," Vol. III, The Kynoch Press, Birmingham,England, 1952.
(20) Observed and calculated structure factors have been deposited as Document No. NAPS-00312 with the ASIS National Auxiliary Publications Service, % CCM Information Sciences, Inc., 22 West 34th St. New York, N. Y. 10001. A copy may be secured by citing the document number and by remitting $1.00 for microfiche or $3.00 for photocopies. Advance payment is required. Make checks or money orders payable to: ASIS-NAPS.

of which was near the position of oxygen atom A. A second oxygen atom, atom B, was therefore added to the model of the structure at the position of the second peak and the refinement was continued with the occupancy factors of both oxygen atoms A and B allowed to vary. Two cycles of anisotropic temperature refinement brought the residual to its final value of $R = 0.162$. In the final cycle of refinement the shifts of the atomic position parameters, except for those of atoms O–A and O–B, were all less than one-fourth of their standard deviations. The same situation was true of the thermal parameters except for those of 5-fluorouracil atoms C(4) and C(5), cytosine atom N(6), and water–oxygen atoms O–A and O–B. The probable reason for these exceptions will be discussed below. The final occupancy factors for partial water–oxygen atoms A and B were 0.76 ± 0.11 and 0.26 ± 0.08, respectively. The refinement program was not constrained in terms of determining the total occupancy of the two sites of the water molecules. However, the fact that the sum of these two partial occupancies is close to unity suggests, in addition to the data previously mentioned, that the crystal is a monohydrate with one water of crystallization for each cytosine–5-fluorouracil pair. It is quite clear that the water molecules are present in the structure since refining the structure without them leads to a final residual of $R = 0.26$.

Results

The final fractional coordinates for all the atoms in the asymmetric unit are given in Table II together with the maximum value of the standard deviation, σ, for each atom type. Table III contains the final anisotropic temperature factors for all unique atoms together with the maximum value of their standard deviations for each atom type. The final anisotropic temperature factors for oxygen atom A can be seen to be very large. These values correspond to root-mean-square thermal displacements of oxygen atom A from its atomic center of 0.81 Å parallel to the a axis, 0.40 Å parallel to the b axis, and 0.41 Å parallel to the c axis.[21] The temperature factors of oxygen atom B are also rather large. Such displacements are much too large to be satisfactorily accounted for by thermal vibrations. It should also be noted that the final water–oxygen atom refinement parameters have very large standard deviations (Tables II and III), whereas the final standard deviations of the parameters describing the atoms of the cytosine and the 5-fluorouracil molecules are all small. This result must be taken as evidence of extreme disorder of the water molecules within the cavity in the crystal structure. This disorder was readily evident in the difference map since the electron density of the water molecule appeared to be very large and diffuse compared with that of a normal oxygen atom. Thus, the existence of water–oxygen atom sites A and B does not imply that the water molecule occupies two discrete sites in the crystal structure but rather, suggests that there is a continuum of sites, some of which have a relatively high probability of occupancy. This continuum is only poorly approximated by the model of discrete atoms held in a harmonic potential field implicit in the functional form of the thermal parameters (see footnote a of Table III).

The values of the anisotropic temperature factors of the atoms of the cytosine and 5-fluorouracil molecules, although somewhat scattered, are mostly within the normal range for these parameters. However, the

Table II. The Atomic Positions in Fractions of a Unit Cell Edge and the Deviation from the Mean Molecular Plane in Ångström Units

Atom	x	y	z	Δ
C–N(1)	−1.1933	0.0082	0.6011	−0.04
C–C(2)	−0.9705	0.1577	0.6394	−0.02
C–N(3)	−0.9599	0.2530	0.7349	0.00
C–C(4)	−1.1384	0.2049	0.7887	0.07
C–C(5)	−1.3742	0.0520	0.7515	0.04
C–C(6)	−1.3802	−0.0435	0.6545	0.02
C–O(2)	−0.8173	0.2000	0.5898	−0.11
C–N(4)	−1.1059	0.3070	0.8796	0.09
U–N(1)	0.4652	0.4634	0.1990	−0.05
U–C(2)	0.4639	0.3513	0.1087	0.09
U–N(3)	0.2360	0.2043	0.0794	0.10
U–C(4)	0.0287	0.1664	0.1353	0.02
U–C(5)	0.0640	0.2932	0.2304	−0.08
U–C(6)	0.2773	0.4325	0.2606	−0.10
U–O(2)	0.6308	0.3762	0.0566	0.15
U–O(4)	−0.1713	0.0370	0.1030	−0.01
U–F(5)	−0.1326	0.2613	0.2859	−0.17
O–A	0.1131	0.4208	0.5147	
O–B	0.3973	0.4162	0.5017	

The Maximum Value of σ for Atomic Positions in Fractions of a Unit Cell Edge

Atom type	Δx	Δy	Δz
C	0.0028	0.0011	0.0008
N	0.0022	0.0010	0.0007
O[a]	0.0020	0.0008	0.0006
F	0.0017	0.0007	0.0005
O[b]	0.0231	0.0043	0.0025

[a] Ring substituent. [b] Water.

Table III. The Anisotropic Thermal Parameters[a] (× 10^4)

Atom	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
C–N(1)	198	58	50	−22	23	18
C–C(2)	154	37	32	−18	34	8
C–N(3)	297	53	30	−16	60	8
C–C(4)	137	37	24	38	41	7
C–C(5)	154	68	30	−12	28	7
C–C(6)	109	78	38	−8	50	11
C–O(2)	490	61	35	−9	113	5
C–N(4)[b]	224	105	16	−24	68	9
U–N(1)	253	48	33	23	42	4
U–C(2)	63	60	28	−7	33	13
U–N(3)	179	60	40	−11	13	14
U–C(4)[b]	−3	68	33	−26	26	14
U–C(5)[b]	107	92	26	14	72	12
U–C(6)	336	59	38	52	22	6
U–O(2)	333	90	39	−2	75	8
U–O(4)	419	56	21	−50	51	−5
U–F(5)	445	100	39	−58	114	2
O–A	5195	352	139	75	55	122
O–B	139	153	86	33	97	70

The Maximum Value of σ for Thermal Parameters (× 10^4)

Atom type						
C	83	13	7	24	17	7
N	64	11	6	19	15	7
O[c]	60	10	4	18	13	5
F	52	9	4	16	11	5
O[d]	1439	71	31	219	127	38

[a] The thermal parameters are in the form $\exp[-(h^2\beta_{11} + k^2\beta_{22} + l^2\beta_{33} + 2hk\beta_{12} + 2hl\beta_{13} + 2kl\beta_{23})]$. [b] The matrix of this quantity is not positive definite. [c] Ring substituent. [d] Water.

(21) D. W. J. Cruickshank, *Acta Cryst.*, 9, 747 (1956).

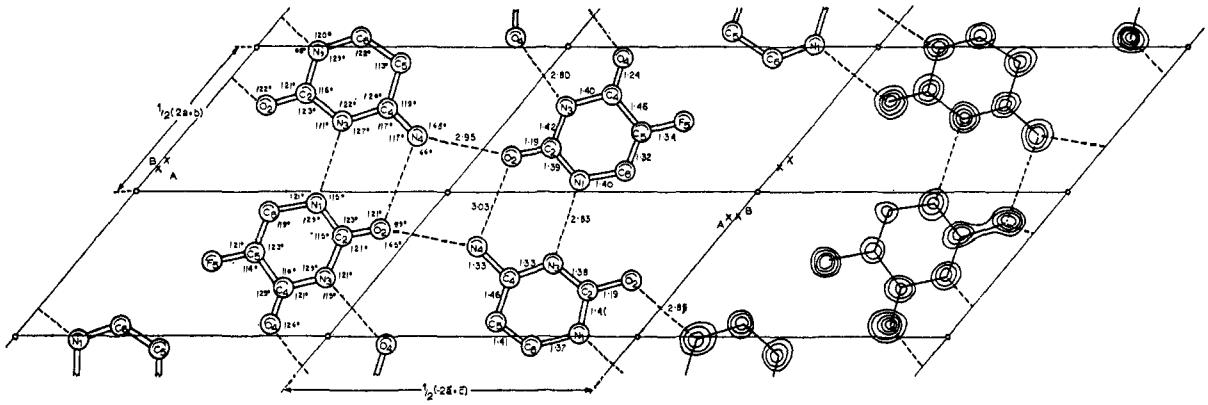

Figure 2. The 1 $\bar{2}$ 2 section of the electron-density map together with a schematic drawing of the molecular structure indicating the atomic numbering scheme, the bond lengths (in ångström units) and the bond angles. Hydrogen bonds are indicated by dashed lines. The positions of the partial water oxygen atoms, A and B, as projected along the *a* axis onto the 1 $\bar{2}$ 2 plane, are indicated by ×'s.

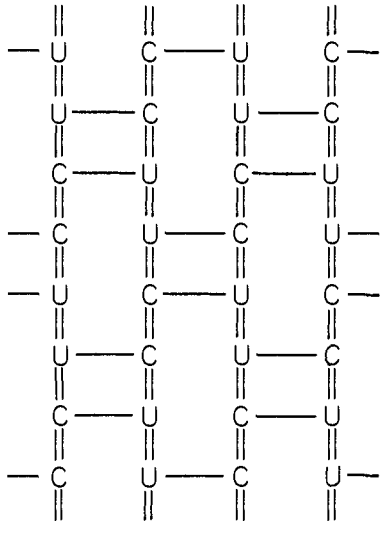

Figure 3. A schematic representation of the hydrogen bonding in the crystal. Here C represents cytosine, U represents 5-fluorouracil, and a solid line represents a single hydrogen bond.

matrices of the anisotropic temperature factors for 5-fluorouracil atoms C(4) and C(5) and cytosine atom N(4) are not positive definite quantities, a situation that is not physically meaningful. It is likely that this situation, together with the large shifts in the thermal parameters for these atoms, is due to the disorder of the water molecule which has a distribution that does not fit the model of discrete atoms assumed in the refinement procedure. This disorder is undoubtedly also responsible for the slightly high value of the residual after the final refinement.

Figure 2 shows the 1$\bar{2}$2 section of the electron density map together with a schematic drawing that illustrates the hydrogen bonds between the 5-fluorouracil and the cytosine molecules. The atomic numbering scheme and the bond angles and distances are also shown in this figure together with the relationship of the molecular structure to the unit cell. The standard deviations of the cytosine and the 5-fluorouracil bond lengths and bond angles as determined from the errors in the unit cell parameters and from the variance–covariance matrix of the final refinement are all very close to 0.01 Å and 1°, respectively.

The 122 section of the electron density map is very close to the molecular plane but not exactly through it. The equation of the mean molecular plane through all the atoms of the cytosine and the 5-fluorouracil residues as determined by the method of least squares is given by the equation $x - 1.955y + 1.924z = -0.041$, where x, y, and z are the fractional coordinates of the unit cell. The deviations of the atoms of the cytosine and the 5-fluorouracil molecules from this plane are given in Table II. The root-mean-square deviation of these 17 atoms out of the mean molecular plane is 0.027 Å. The maximum deviation out of this plane, which is that of 5-fluorouracil atom F(5), is 0.17 Å. The interplanar spacing between these mean molecular planes is 3.50 Å, a figure which is very close to the value cited above of 3.45 Å for the 1$\bar{2}$2 plane. In Figure 2 it was convenient to draw the structure in the 1$\bar{2}$2 plane. However, this action resulted in a slight deviation of a portion of the 5-fluorouracil molecule from the plane of the drawing. For example, the electron density peaks of atoms C(6), C(5), and F(5) of the 5-fluorouracil molecule are slightly out of the plane of Figure 2 which causes these peaks to appear somewhat smaller in the electron density map of this figure than they actually are at their maxima.

The 5-fluorouracil and cytosine molecules are organized into a hydrogen-bonded net as is shown in Figure 2. In the vertical direction of this figure there are infinite chains of molecules held together by two hydrogen bonds between each pyrimidine molecule. These chains consist of the pair 5-fluorouracil–cytosine followed by the pair cytosine–5-fluorouracil. The former pair is related to the latter pair by a center of symmetry. This hydrogen-bonding scheme forms an alternating double repetition of cytosine and 5-fluorouracil residues. These infinite chains are cross linked by single hydrogen bonds between cytosine and 5-fluorouracil residues. This scheme is illustrated schematically in Figure 3 in which the lines between the symbols for cytosine (C) and 5-fluorouracil (U) stand for the number of hydrogen bonds connecting them. It can be seen that adjacent vertical columns have the alternating sequence of two cytosine and two 5-fluorouracil residues out of phase in such a way that they are held together by single hydrogen bonds between the cytosine and the 5-fluorouracil residues in the horizontal

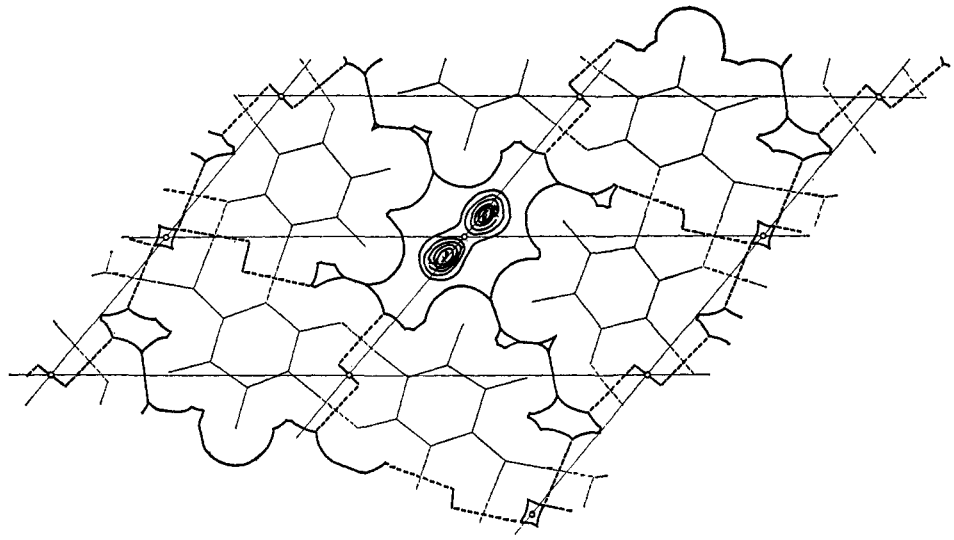

Figure 4. The van der Waals packing diagram in the 122 plane shown in Figure 2. The contours represent the electron-density distribution of the water molecules projected along the a axis onto the $1\bar{2}2$ plane. The positions of the partial water–oxygen atoms, A and B, are indicated by ×'s.

direction. The open areas without hydrogen bonds in Figure 3 are schematic representations of the cavity that contains the disordered water molecules. This elongated cavity is also illustrated in Figure 2 which has ×'s over the positions of partial water–oxygen atoms A and B as projected along the a axis.

The bond lengths and angles listed in Figure 2 are similar to those which have been observed in crystal structure analyses of other cytosine or uracil derivatives. The hydrogen bond distances shown in Figure 2 are all within the normal range,[22] the only exception being a slightly long hydrogen bond of length 3.03 Å between the amino group, N(4), of cytosine and the carbonyl oxygen atom, O(2), of 5-fluorouracil.

The van der Waals packing of the cytosine–5-fluorouracil complex in the $1\bar{2}2$ plane is shown in Figure 4. As can be seen in this figure, the pyrimidine rings fit together quite well with no abnormally short intermolecular contacts in forming the hydrogen-bonded net but they leave empty a flattened and somewhat elongated cavity which is shown in the middle of the figure. In Figure 4 the electron-density distribution of the water molecules is projected along the a axis onto the molecular plane producing the contours shown. The position of partial water oxygen atoms A and B are indicated by ×'s in Figure 4. Figure 5 shows the electron-density map between the molecular planes. This section, which is part of the plane $y = 0.433$, is taken through the water peak. It shows the diffuseness of the water peak and the position of the two partial water–oxygen atoms A and B. It can be seen that peak B is located just below the midline between the two molecular planes, whereas peak A is far above it. An identical electron density peak, related to the peak shown in Figure 5 by a center of symmetry, lies 1.15 Å away in the plane $y = 0.567$.

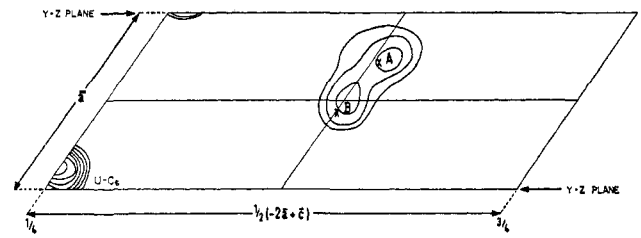

Figure 5. A portion of the section $y = 0.433$ of the electron-density map in the vicinity of the water molecules. The positions of the partial water–oxygen atoms, A and B, are indicated by ×'s.

Discussion

The cavity in the cytosine–5-fluorouracil crystal is quite narrow in the molecular plane; if an oxygen atom were located at the level of the plane it would be quite crowded. Accordingly, the water molecules adopt positions in between the two molecular levels. The walls of the cavity are lined with atoms with which it is unlikely that a water molecule would have any significant interaction. The only exceptions to this statement are the four symmetry-related cytosine O(2) atoms which are located at the ends of the flattened crevice. However, the approach to the only available position on each of these oxygen atoms which is most likely to take part in strong hydrogen bonding is completely blocked by the hydrogen atom which is covalently bonded to 5-fluorouracil atom C(6) adjacent to the atom under consideration. The closest approach of any of the symmetry-related cytosine O(2) atoms surrounding a particular cavity to the positions of partial oxygen atom A is 2.77 Å, about the distance one would expect for a strong O–H···O hydrogen bond.[22] However, it is also about the distance one would expect for an O···O van der Waals contact.[22] It could be argued that the relative orientations of cytosine carbonyl oxygen atom O(2) and the closest of the two water molecules centered at the two symmetry related positions, O–A, are unfavorable for hydrogen bonding but this argument is, at best, tenuous.[23] However, it is to be expected that if the water molecule were hydrogen bonded to cytosine atom O(2) it would be strongly

(22) L. Pauling, "The Nature of the Chemical Bond," Cornell University Press, Ithaca, N. Y., 1960.

(23) J. Donahue in "Structural Chemistry and Molecular Biology," A. Rich and N. Davidson, Ed., W. H. Freeman and Co., San Francisco, Calif., 1968.

localized near that atom but, as mentioned above, the water molecules are extremely disordered within the cavity. Thus it appears unlikely that a water molecule at position O–A, and similarly, one at position O–B, hydrogen bonds to any of the symmetry-related cytosine O(2) atoms. Therefore it seems reasonable to argue that there are only very small, if any, attractive forces between the water molecules and the atoms lining the cavity wall and that the observed stoichiometry of the crystal is simply a reflection of the fact that the total volume of the cavity is such that it will accommodate only two water molecules.

The cytosine–5-fluorouracil structure has an interesting four-membered hydrogen-bonded ring. As can be seen in Figure 2, this ring involves two amino groups and two carbonyl groups so that, in effect, two sets of cytosine and 5-fluorouracil molecules each held together with two hydrogen bonds are then joined with two additional symmetry related hydrogen bonds to make a cyclic tetramer. Arrangements of this type have been found in other crystal structures involving purines and pyrimidines. In the intermolecular complex of 1-methyl-5-bromocytosine and 9-ethylguanine[4] the two base pairs are organized into a cyclic tetramer with a set of hydrogen bonds similar to that observed in the present structure. A similar ring is also found in the structure of cytosine-5-acetic acid[24] in which the cytosine derivative itself occupies all four sides of the hydrogen-bonded ring. A third example is found in the closely related structure of 1-methylcytosine–5-fluorouracil.[25] The occurrence of this type of hydrogen-bonded tetramer in which pairs of bases are brought together is of some interest because of the frequent occurrence of hydrogen-bonded pairs of bases in the double-helical form of the nucleic acids. These structures provide a kind of structural rationale for believing that cyclic tetramers of bases in polynucleotide structures may form in configurations which are transiently stabilized.[26] This phenomenon might be of direct relevance in the phenomenon of genetic crossing over which involves the participation of two double-helical molecules of DNA.

The molecular structure of the complex between 1-methylcytosine and 5-fluorouracil[25] is very closely related to present structure. In both crystals the network of hydrogen-bonded cytosine and uracil derivatives are identical, including the cyclic tetramer described above, except that the structure containing 1-methylcytosine is blocked by the methyl group from forming a cyclic dimer between the two cytosine derivatives. The similarity between the two structures is remarkable considering that the crystal structures of the two complexes are completely different. The space group of the crystal containing 1-methylcytosine is Pbca which causes the rings to pack into a herring-bond pattern, whereas the rings in the present structure are packed into a layered array. We must therefore conclude that there is some type of specificity between the bases which is common to the two structures.

A survey of reported crystal structures containing self-bonding uracil derivatives revealed nine such additional structures.[27–35] In eight of these nine structures, those of uracil,[27] 1-methyluracil,[28] 1-methylthymine,[29] the complex 5-fluorouracil–9-ethylhypoxanthine,[30] 5-ethyl-6-methyluracil,[31] 2,4-dithiouracil,[32] 5-nitrouracil monohydrate,[33] and the photodimer of 1-methylthymine,[34] a cyclic dimer is formed between the uracil residues which is identical with that formed between the uracil residues in the structure of cytosine–5-fluorouracil. Only in the structure of thymine monohydrate[35] does a different cyclic dimer form. In that case atom O(4) is hydrogen bonded to a water molecule. The frequent occurence of the type of cyclic dimer found between the uracil residues in the structure of cytosine–5-fluorouracil suggests that its association constant is far greater than those of the two other possible cyclic dimers which might exist between uracil derivatives. When such a cyclic dimer is formed between two 5-fluorouracil molecules there are only two ways of forming a cyclic dimer with cytosine using the available hydrogen-bonding sites. However, only one of these ways can be used to form a cyclic dimer with 1-methylcytosine. It is interesting that the cyclic dimer which is possible with both cytosine derivatives is the only one which allows the formation of the above-mentioned cyclic tetramer. It is likely that the formation of the cyclic tetramer confers additional stability on the crystal structure.

Only in the present structure has the mode of self-complexing between cytosine molecules shown in Figures 2 and 4 been found. In two of the other three known structures in which a neutral cytosine derivative forms a self-complex, those of cytosine[36] and cytosine monohydrate,[37] an unsymmetrical cyclic dimer is formed utilizing, respectively, as the hydrogen donor and acceptor atoms, atoms N(1) and O(2) on one molecule and atoms N(4) and N(3) on the other molecule. In the remaining reported structure in which a neutral cytosine derivative forms a self-complex, that of 1-methylcytosine[38] atom N(1) is blocked by a methyl group. The molecule forms a hydrogen-bonded pair with itself through a centrosymmetric cyclic dimer utilizing atoms N(3) and N(4) in each molecule as the hydrogen-acceptor and -donor atoms, respectively. The limited data above do not suggest that the cyclic dimer formed between cytosine molecules in the present structure is especially favorable energetically since all three possible types of cyclic dimers which can be formed between two neutral cytosine molecules do occur. Thus it is likely that the strength of the cyclic dimer formed between the 5-fluorouracil molecules in the present structure and in that of the complex 1-methylcytosine–5-fluorouracil[25] and the formation of the cyclic tetramers mentioned above are the energetically controlling factors in these structures.

(24) R. E. Marsh, R. Bierstadt, and E. L. Eichorn, *Acta Cryst.*, **15**, 310 (1962).
(25) S. H. Kim and A. Rich, *J. Mol. Biol.*, in press.
(26) O. Siddiqi, "The Nucleic Acids," Sree Sarawaty Press Ltd., Calcutta, 1965, p 175.
(27) (a) G. S. Parry, *Acta Cryst.*, 7, 313 (1954); (b) R. F. Stewart, and L. H. Jensen, *ibid.*, **23**, 1102 (1967).
(28) D. W. Green, F. S. Mathews, and A. Rich, *J. Biol. Chem.*, **237**, 3573 (1962).
(29) K. Hoogsteen, *Acta Cryst.*, **16**, 28 (1963).
(30) S. H. Kim and A. Rich, *Science*, **158**, 1046 (1967).
(31) G. N. Reeke and R. E. Marsh, *Acta Cryst.*, **20**, 703 (1966).
(32) E. Shefter and H. G. Mautner, *J. Amer. Chem. Soc.*, **89**, 1249 (1967).
(33) B. M. Craven, *Acta Cryst.*, **23**, 376 (1967).
(34) J. R. Einstein, J. L. Hossau, J. W. Longworth, R. O. Rahn, and C. H. Wei, *Chem. Commun.*, 1063 (1967).
(35) R. Gerdil, *Acta Cryst.*, **14**, 333 (1961).
(36) D. L. Barker and R. E. Marsh, *ibid.*, **17**, 1581 (1964).
(37) G. A. Jeffrey, and Y. Kinoshita, *ibid.*, **16**, 20 (1963).
(38) F. S. Mathews, and A. Rich, *Nature*, **201**, 179 (1964).

The present structure is an intermolecular complex involving cytosine and a derivative of uracil. It appears to go against the rule of electronic complimentarity described above, since infrared studies of derivatives of these molecules have shown that there is no selective affinity for hydrogen bonding between uracil and cytosine derivatives.[8] Similarly, 1-methyl-5-fluorouracil forms a crystalline complex with 9-ethyladenine[3] but does not form one with cytosine derivatives.[39] However, it is important to note that the molecules involved in the present structure do not have side chains on their glycosidic nitrogen atoms. Instead, atom N(1) is protonated in both molecules and, of course, plays an important role in developing the hydrogen bonding system which stabilizes the entire crystal. Thus, it is likely that the electronic distributions in the pyrimidines are modified somewhat when they are substituted on the N(1) position so that their hydrogen-bonding properties are altered. Of course, in the present crystal structure, and that of the complex 1-methylcytosine–5-fluorouracil, the hydrogen bonding between the cytosine derivatives and 5-fluorouracil involves the atom N(1) of 5-fluorouracil. Such a mode of hydrogen bonding would be impossible if a substituent were present on this atom.

(39) A. Rich, unpublished observations.

Acknowledgments. We wish to thank Dr. S.-H. Kim for his assistance in pointing out the existence of the complex. This work was supported by grants from the National Institutes of Health, the U. S. Public Health Service, and the National Aeronautics and Space Agency.

Synthesis of Carbocyclic Analogs of Purine Ribonucleosides[1]

Y. Fulmer Shealy and Joe D. Clayton

Contribution from the Kettering-Meyer Laboratory, Southern Research Institute, Birmingham, Alabama 35205. *Received November 8, 1968*

Abstract: The racemic forms of the carbocyclic (cyclopentane) analogs of adenosine, inosine, 6-mercaptopurine ribonucleoside, and 6-(methylthio)purine ribonucleoside were synthesized from (±)-4β-amino-2α,3α-dihydroxy-1β-cyclopentanemethanol. This amine was synthesized by two routes from 2α,3α-diacetoxy-1β,4β-cyclopentanedicarboxylic acid *via* its anhydride and monoamide. The two routes differed in the order in which the amino group was introduced by a Hofmann reaction and the hydroxymethyl group, by metal hydride reduction of an ester or acid chloride group. The stereochemistry of the starting cyclopentane was fixed by the preparation of this compound from *exo-cis*-5-norbornene-2,3-diol, which was prepared by *cis*-dihydroxylation of norbornadiene. The structure of the norbornenediol was confirmed by chemical means and by nmr analysis.

It is now well established that alterations of either the furanose or the base moiety of naturally occurring purine and pyrimidine nucleosides may produce derivatives that exert interesting and powerful biological effects.[2] Replacement of the furanose oxygen atom of nucleosides with a methylene group would produce carbocyclic (cyclopentane) analogs in which the hydroxyl groups occupy the same positions, have the same *cis–trans* relationships, and may be expected to assume similar conformations.[4] The carbocyclic analogs have the potential, therefore, either to mimic or to antagonize the functions of the naturally occurring nucleosides and nucleotides. Unlike the nucleosides, the carbon–nitrogen bond joining the heterocyclic base to the cyclopentane ring should be comparable in stability to that of a simple alkyl derivative and should, therefore, be much less susceptible to enzymatic fission than the analogous bond of nucleosides. For these reasons we have synthesized the carbocyclic analogs (racemic forms) of the naturally occurring nucleosides adenosine and inosine and of two biologically active purine nucleosides. A preliminary account of our synthesis of the adenosine analog (C-Ado, **28**) has appeared;[3] subsequently, Kishi and coworkers[5] reported in a preliminary communication that X-ray analysis revealed the structure of a recently isolated antibiotic[6] to be an optically active form of C-Ado (**28**). Earlier, Murdock

(1) This investigation was supported by the Cancer Chemotherapy National Service Center, National Cancer Institute, National Institutes of Health, Contract No. PH43-64-51 and by the C. F. Kettering Foundation.

(2) References to some of the large number of publications in this field may be found in the following reviews: J. A. Montgomery and H. J. Thomas, *Advan. Carbohydrate Chem.*, **17**, 301 (1962); J. J. Fox, K. A. Watanabe, and A. Bloch, *Progr. Nucl. Acid Res. Mol. Biol.*, **5**, 251 (1966); S. S. Cohen, *ibid.*, **5**, 1 (1966); C. Heidelberger, *Ann. Rev. Pharmacol.*, **7**, 101 (1967); J. A. Montgomery, *Progr. Med. Chem.*, in press; *cf.* references cited in ref 3.

(3) Y. F. Shealy and J. D. Clayton, *J. Amer. Chem. Soc.*, **88**, 3885 (1966).

(4) The preferred conformations of simple substituted cyclopentanes are the puckered envelope and half-chair forms [K. S. Pitzer and W. E. Donath, *ibid.*, **81**, 3213 (1959); F. V. Brutcher, Jr., T. Roberts, S. J. Barr, and N. Pearson, *ibid.*, **81**, 4915 (1959); M. Hanack, "Conformation Theory," Academic Press, New York, N. Y., 1965, pp 72–78; E. L. Eliel, N. L. Allinger, S. J. Angyal, and G. A. Morrison, "Conformational Analysis," John Wiley & Sons, Inc., New York, N. Y., 1965. In crystalline purine and pyrimidine nucleosides and nucleotides either C2′ or C3′ is displaced (either *endo* or *exo* to the base moiety) from the mean plane of the other four furanose-ring atoms by 0.5–0.6 Å; see the following publications and references cited therein: M. Sundaralingam, *J. Amer. Chem. Soc.*, **87**, 599 (1965); M. Sundaralingam and L. H. Jensen, *J. Mol. Biol.*, **13**, 930 (1965); A. E. V. Haschemeyer and A. Rich, *ibid.*, **27**, 369 (1967); P. Tollin, H. R. Wilson, and D. W. Young, *Nature*, **217**, 1148 (1968). There is evidence, however, of some difference in the stabilities of the preferred conformers of comparably substituted cyclopentanes and tetrahydrofurans [H. R. Buys, C. Altona, and E. Havinga, *Tetrahedron*, **24**, 3019 (1968)].

(5) T. Kishi, M. Muroi, T. Kusaka, M. Nishikawa, K. Kamiya, and K. Mizuno, *Chem. Commun.*, 852 (1967).

(6) T. Kusaka, H. Yamamoto, M. Shibata, M. Muroi, T. Kishi, and K. Mizuno, *J. Antibiot.* (Tokyo), **21**, 255 (1968).