Complexes between Nucleotide Base and Amino Acid. I. Crystal Structure of Cytosine: *N*-Formylglycine

Minoru Ohki, Akio Takenaka, Hirotaka Shimanouchi, and Yoshio Sasada

Laboratory of Chemistry for Natural Products, Tokyo Institute of Technology, Ookayama, Meguro-ku, Tokyo 152

(Received September 25, 1974)

Cytosine: *N*-formylglycine complex crystallizes in the space group $Pna2_1$, with cell dimensions of $a=8.444(2)$, $b=21.169(5)$, and $c=5.120(1)$ Å. The structure was solved by the trial-and-error method on the basis of a Patterson map. A full-matrix least-squares refinement led to an R value of 0.035. The difference map revealed the protonation of cytosine by the carboxyl group of *N*-acylamino acid and the involvement of ionic interaction in complex formation. This mode is the same as proposed for cytosine: *N*-benzoylglycine complex in spite of different molecular environment. Such arrangement of cytosine and ionized carboxyl group of amino acid may suggest a mutual recognition scheme between nucleic acid and protein.

Interactions between nucleic acids and proteins take an important part in several processes of protein synthesis, its regulation, the repair of nucleic acids, and so on. In nucleic acid-protein interactions, several kinds of weak bonds must be formed between individual nucleotide and amino acid residues. It is, therefore, interesting to find if there are some specific binding patterns between components of nucleic acids and proteins. They should provide a stereochemical basis for the molecular mechanism of mutual recognition.

Complex formation between nucleotide base and amino acid has been investigated extensively.[1] Tamura, Hata, Sato, and Sakurai[2] showed by an X-ray analysis of cytosine: *N*-benzoylglycine monohydrate (abbreviated hereafter as CBG complex) that the complex is formed by the two hydrogen bonds between the carboxyl group of amino acid and the two nitrogen atoms (amino group and N(3) atom) of cytosine. Recently, we have found a different type of complex formation between cytosine and undissociated acidic side group of glutamic acid.[3] In the former complex, however, the proton migration was deduced from the examination of bond lengths and angles without the direct proof by a difference map. It is well known that cytosine is easily protonated at N(3) by strong acids[4] and the high proton affinity has been justified by the theoretical calculation of π-electron distribution.[5] In view of the nucleic acid-protein interactions, it is important to confirm the protonation of cytosine by amino acid. We have done this by an accurate analysis of the crystal structure of the title complex which is an analogue of the CBG complex.

Experimental

Square-prism crystals were obtained from an aqueous solution containing equimolar cytosine and *N*-formylglycine. The complex formation was confirmed by an elementary analysis and IR spectrum of the crystals. Found: C, 39.1; H, 4.7; N, 25.5%. Calcd for $C_4H_5N_3O \cdot C_3H_5NO_3$: C, 39.3; H, 4.7; N, 26.2%. Laue symmetry and systematic absences of reflexions indicated the space group to be $Pna2_1$ or Pnam. The unti-cell dimensions were determined by least-squares calculation from the spacings of 64 reflexions, $0kl$ and $hk0$, on Weissenberg photographs, calibration being made with superposed silicon lines ($a=5.43075$ Å). The density was measured by flotation in a chloroform-carbon tetrachloride mixture. Crystal data are given in Table 1.

Table 1. Crystal data for cytosine: *N*-formylglycine complex

Crystal system: orthorhombic		
Systematic absences: $0kl$ with $k+l=2n+1$ and $h0l$ with $h=2n+1$.		
Space group: $Pna2_1$.		
$a=8.444(2)$ Å	$b=21.169(5)$ Å	$c=5.120(1)$ Å
$U=915.1(3)$ Å^3	$Z=4$	
$D_x=1.554$ g cm^{-3}	$D_m=1.55_7$ g cm^{-3}	
μ(CuKα)$=11.4$ cm^{-1}		

Intensity data were collected on integrated, equi-inclination Weissenberg photographs ($hk0$-$hk4$ and $0kl$-$6kl$) using Ni-filtered CuKα radiation and the intensities were measured by a TV densitometer.[6] The sizes of crystals used for the a- and c-axis rotations were $0.10\times0.15\times0.40$ mm and $0.15\times0.20\times0.60$ mm, respectively. Corrections were made for Lorentz and polarization factors, but not for absorption. Intensities of both of extended and contracted reflexions on upper layer photographs were corrected by a method proposed by Takenaka and Sasada,[7] and were averaged; a discrepancy index, $R'=\sum||F_{ext}|-|F_{cont}||/\sum|F_{mean}|$, was 0.040, where subscripts ext and cont stand for the extended and contracted spots, respectively. Two sets of intensity data thus obtained were cross-correlated and then placed on an absolute scale by the Wilson's method. A total of 1090 independent reflexions was obtained, of which zero-reflexions numbered 103.

Structure Determination and Refinement

Although $N(z)$ test[8] was not effective, a three-dimensional Patterson map indicated that the space group is $Pna2_1$. The Patterson map also gave the approximate position of the cytosine moiety. The remaining non-hydrogen atoms were found by successive Fourier syntheses. Atomic coordinates and temperature factors, first isotropic and then anisotropic, were refined by the block-diagonal least-squares method; the minimized function was $\sum w\Delta F^2$, where $\Delta F=|F_o|-|F_c|$. After the conventional R factor reached to 0.074, a difference synthesis revealed all the hydrogen atoms. Further refinement including these hydrogen atoms reduced the R factor to 0.037. In the final stage, the full-matrix least-squares method was applied, where the following weight function was used:

$w=\exp(-as^2-bt^2-cst-ds-et-f)$ for $|F_o|\neq 0$, where $s=|F_o|$ and $t=\sin\theta/\lambda$; $w=1/\langle\Delta F^2\rangle$ for $|F_o|=0$. The coefficients, a, b, c, d, e, and f, were evaluated by least-squares at each cycle of the structure refinement so that $\langle w\Delta F^2\rangle=1$; at the final stage, they were 0.001778, 41.00, 0.4658, −0.1752, −42.77, and 8.107, respectively. The refinement was terminated when the largest shift of parameters was less than 0.3σ. The final values of R and R_w, where $R_w=(\sum w\Delta F^2/\sum w|F_o|^2)^{1/2}$, were 0.035 and 0.043, respectively. There were no residual peaks greater than 0.2 e Å^{-3} in the final difference map.

Atomic scattering factors used were taken from "International Tables for X-ray Crystallography".[9]

The computations were done on a HITAC 8800 computer at the University of Tokyo and on a HITAC 8700 computer at Tokyo Institute of Technology. The programs, TLSU for the determination of cell dimensions, TDRW for the data reduction, TFLS for the full-matrix least-squares refinement, and DCMS for the stereo-pair drawing of the crystal and molecular structure, all written by two of the authors (A.T. and Y.S.), and DAPH in the Universal Crystallographic Computation Program System,[10] were used.

The final positional and thermal parameters are given in Table 2. Observed and calculated structure factors are listed in Table 3. Numbering of atoms and thermal ellipsoids are shown in Fig. 1, the crystal structure in Fig. 2, and the bond lengths and angles in Fig. 3. The standard deviations for bond lengths and angles involving C, N, and O are 0.004 Å and 0.4°, respectively, and those involving hydrogen are about ten times of the above values. The hydrogen bonds are listed in Table 4. The least-squares planes of six-membered ring of cytosine, the carboxyl group and the amide group of amino acid are given in Table 5, together with the displacements of atoms from the planes.

Discussion

As shown in Fig. 1, the complex is formed by the two hydrogen bonds between the ionized carboxyl group of amino acid and the two nitrogen atoms, N(4) and the protonated N(3), of cytosine. The present study has proved definitely the protonation of N(3) and the involvement of ionic interaction in complex formation. In this "double hydrogen-bonded system", the two N···O distances are not equal to each other (Fig. 2), contrary to those in CBG complex. The N(4)H···O(3) distance is comparable to the normal distance of NH···O type hydrogen bond,[11] while the

Table 2. Final positional and thermal parameters

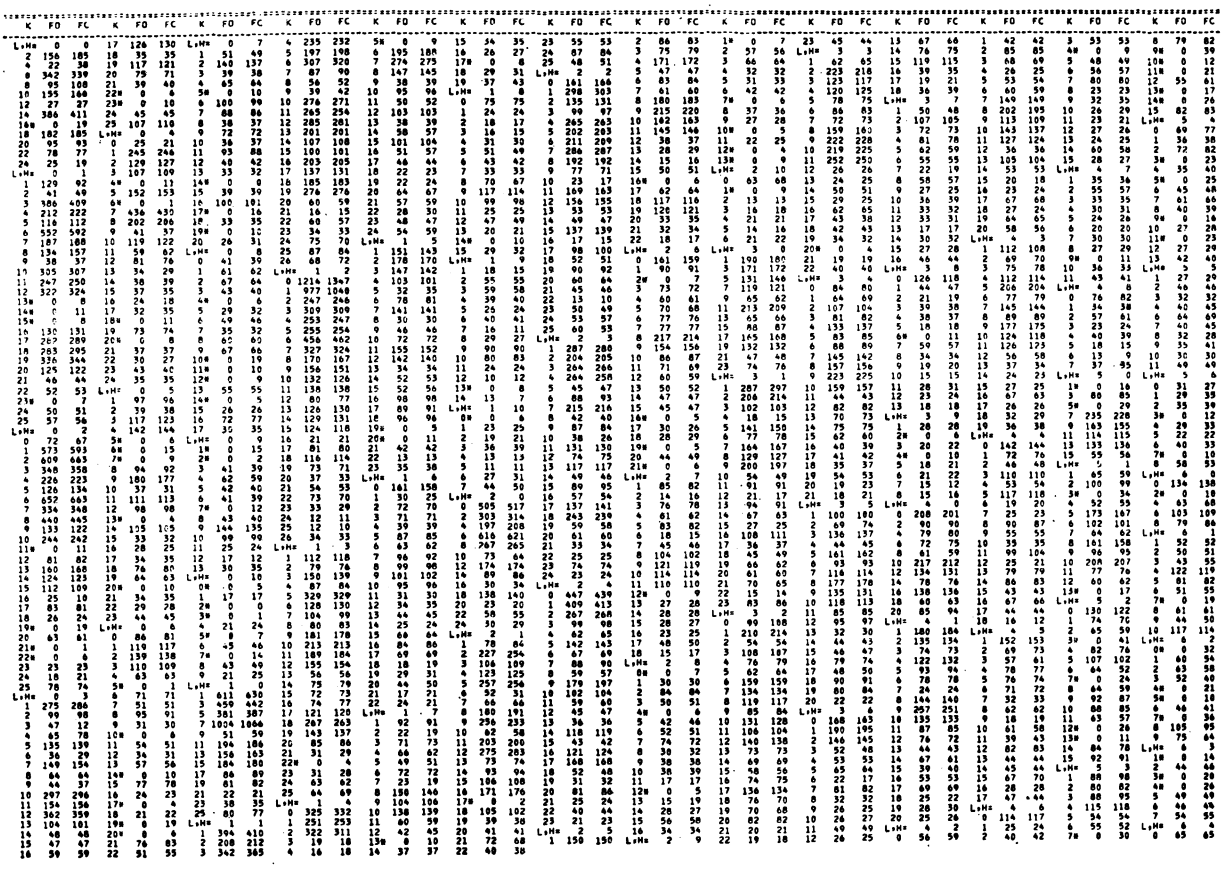
Standard deviations are given in parentheses. The anisotropic temperature factor has the form $\exp[-(B_{11}h^2+B_{22}k^2+B_{33}l^2+B_{12}hk+B_{23}kl+B_{31}lh)]$

Atom	x[a]	y[b]	z[a]	B_{11}[a]	B_{22}[b]	B_{33}[a]	B_{12}[a]	B_{23}[a]	B_{31}[a]
N(1)	958(3)	12419(9)	2119(7)	142(3)	115(4)	245(8)	−5(2)	−1(3)	−79(8)
C(2)	1515(3)	18025(10)	3108(7)	127(3)	127(4)	198(9)	−5(2)	6(3)	−76(9)
N(3)	951(3)	23396(9)	1945(6)	129(3)	123(4)	186(7)	−1(2)	7(3)	−88(7)
C(4)	−52(3)	23403(10)	−107(7)	112(3)	151(5)	145(8)	2(2)	8(3)	−29(9)
C(5)	−601(3)	17505(11)	−1090[c]	131(3)	178(5)	192(8)	−12(2)	−11(4)	−84(9)
C(6)	−79(3)	12224(12)	82(7)	143(4)	155(5)	206(9)	−12(2)	−18(3)	−43(9)
O(2)	2465(3)	18192(8)	4900(7)	198(3)	157(4)	350(8)	−2(2)	12(3)	−300(10)
N(4)	−475(3)	28869(11)	−1078(7)	161(3)	177(5)	265(9)	1(2)	24(4)	−156(10)
O(3)	561(3)	40172(8)	1284(7)	184(3)	138(3)	297(8)	10(2)	2(3)	−203(8)
O(4)	1924(2)	34084(7)	4033(7)	191(3)	115(3)	262(7)	3(2)	14(3)	−166(8)
O(5)	3165(2)	52186(8)	347(7)	179(3)	152(3)	273(8)	2(2)	10(3)	128(8)
N(5)	1519(3)	51074(8)	3845(7)	146(3)	111(4)	218(7)	11(2)	−7(3)	52(9)
C(7)	1462(3)	39342(10)	3181(7)	128(3)	115(4)	166(8)	6(2)	10(3)	−12(9)
C(8)	2114(3)	44982(10)	4680(7)	144(4)	131(4)	172(10)	5(2)	1(3)	−15(9)
C(9)	2089(3)	54083(10)	1798(8)	145(3)	106(4)	247(9)	−2(2)	−3(3)	2(10)

Atom	x[d]	y[d]	z[d]	B/Å^2 [e]
H(1)	133(4)	88(2)	297(8)	13(6)
H(3)	133(5)	272(2)	261(9)	25(7)
H(41)	−10(4)	326(2)	−36(10)	25(7)
H(42)	−114(6)	288(2)	−243(12)	36(9)
H(5)	−137(5)	176(2)	−250(10)	25(8)
H(6)	−42(4)	78(2)	−39(8)	18(6)
H(81)	326(4)	450(2)	443(7)	15(6)
H(82)	180(5)	445(2)	649(10)	25(8)
H(9)	163(5)	580(2)	133(10)	28(8)
H(10)	76(5)	528(2)	474(9)	22(7)

a) $\times 10^4$. b) $\times 10^5$. c) The z coordinate of C(5) was fixed from the requirement of the polar space group.
d) $\times 10^3$. e) $\times 10$

TABLE 3. OBSERVED AND CALCULATED STRUCTURE FACTORS ($\times 10$)



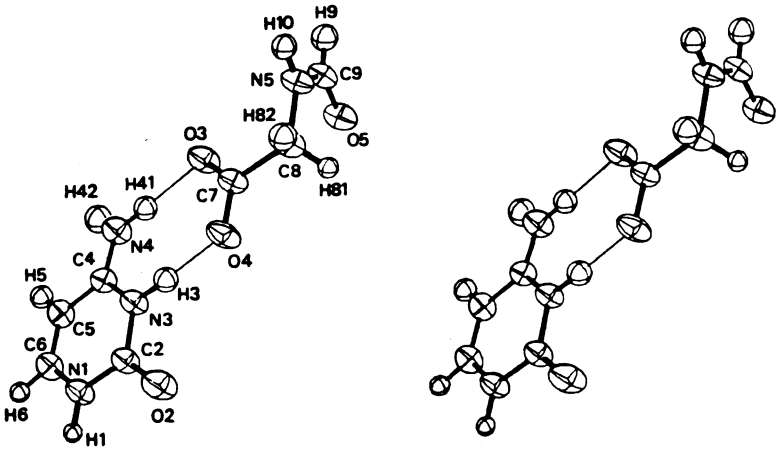


Fig. 1. A stereoscopic view of cytosine: *N*-formylglycine complex. Thermal ellipsoids are drawn at the 50% probability level.

N(3)H···O(4) distance is very close to that between carboxylate oxygen and protonated imidazole nitrogen atoms found in *N*-acetyl-L-histidine monohydrate.[12] This implies that the positive and negative charges are mostly located at N(3) and O(4) atoms, respectively. The differences between the C(7)–O(3) and C(7)–O(4) bond lengths and between the C(8)–C(7)–O(3) and C(8)–C(7)–O(4) angles (Fig. 3) also support this. Other hydrogen bonds are found between the carbonyl O(5) atom of amino acid and the N(1) atom of cytosine, between the N(4) and O(2) atoms of cytosines, and between the O(3) and N(5) atoms of the amino acids. These hydrogen bonds, except for the last one, are schematically the same as those in the CBG complex. The approach of the two amino acids in the CBG complex is impeded by the bulky phenyl groups and they are connected through the hydrogen bonds with the water molecule between them. In the present complex, however, they are linked directly by the N(5)H···O(3) hydrogen bond. The charge distribution in the "double hydrogen-bonded system" seems to be affected by such a different environment.

The closest nonbonding intermolecular contacts are between O(4) of amino acid and C(5) of cytosine at $1/2+x$, $1/2-y$, $1+z$ and between C(9) of amino acid and O(2) of cytosine at $1/2-x$, $1/2+y$, $-1/2+z$.

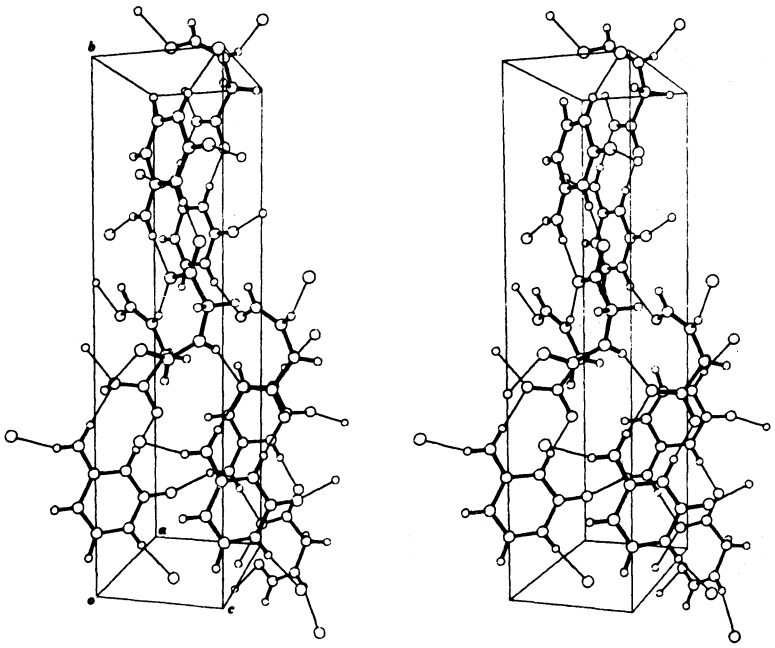


Fig. 2. A stereoscopic view of the crystal structure.

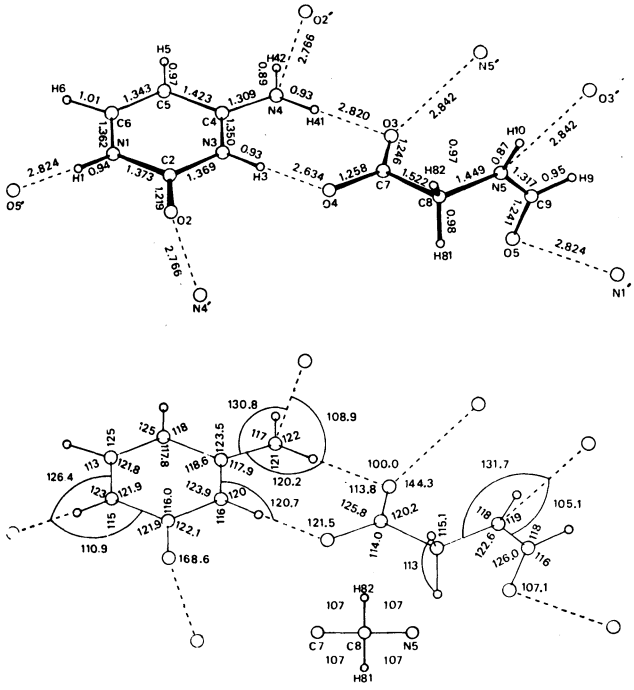


Fig. 3. Distances and angles.
The standard deviations for bond distances and angles are 0.004 Å and 0.4 °, respectively, while those involving hydrogen are about ten times larger.

These distances, 3.274 and 3.164 Å, respectively, are slightly larger than the minimum van der Waals contacts.[13] If the hydrogen atoms are taken into account, however, the H···O distances are somewhat shorter than the sum of these van der Waals radii;[13] H(5)···O(4) 2.32 Å and H(9)···O(2) 2.40 Å. Relevant angle is 170° for C(5)–H(5)–O(4) and 137° for C(9)–H(9)–O(2). It seems that there are some C–H···O interactions, at least in the former.

The bond lengths and angles of amino acid are in good agreement with those found in other amino acids and oligopeptides. The torsion angles of C(7)–C(8)–N(5)–C(9) and of O(3)–C(7)–C(8)–N(5) are −81.1 and 4.2°, respectively, which are close to those found in many dipeptides.[14] The O(5) and C(8) atoms are *cis* around the C(9)–N(5) bond.

Table 4. Distances and angles of the hydrogen bonds

N	H	O	Distance (*l*/Å) N···O	Distance (*l*/Å) H···O	Angle (*φ*/°) N–H···O
N(3)	H(3)	O(4) (a)	2.634	1.71	176
N(4)	H(41)	O(3) (a)	2.820	1.89	176
N(4)	H(42)	O(2) (b)	2.766	1.91	160
N(1)	H(1)	O(5) (c)	2.824	1.90	167
N(5)	H(10)	O(3) (d)	2.842	2.03	157
	(a) at		x,	y,	z
	(b) at		$-1/2+x$,	$1/2-y$,	$-1+z$
	(c) at		$1/2-x$,	$-1/2+y$,	$1/2+z$
	(d) at		$-x$,	$1-y$,	$1/2+z$

The protonation to cytosine results in the lengthening of the C(2)–N(3) and N(3)–C(4) bonds, the shortening of the C(2)–O(2) and C(4)–N(4) bonds, and the widening of the C(2)–N(3)–C(4) angle, these features being similar to those in other protonated cytosines as shown in Fig. 4 where the average bond distances and angles are given for the neutral and protonated species.[15–19] The slightly different dimensions between the present and other protonated cytosines are caused by the effects of alkylation at N(1)[20] and hydrogen bond with O(2),[15] since the O(2) atoms of other protonated species are not hydrogen bonded.

The N(3) atom deviates by 0.011 Å from the mean plane of the six-membered ring. The exocyclic atoms, O(2) and N(4), displace in the direction opposite to N(3) by 0.040 and 0.015 Å, respectively. Such a tendency is the same as found in 1-(β-D-arabinofuranosyl)cytosine hydrochloride,[19] but is different from

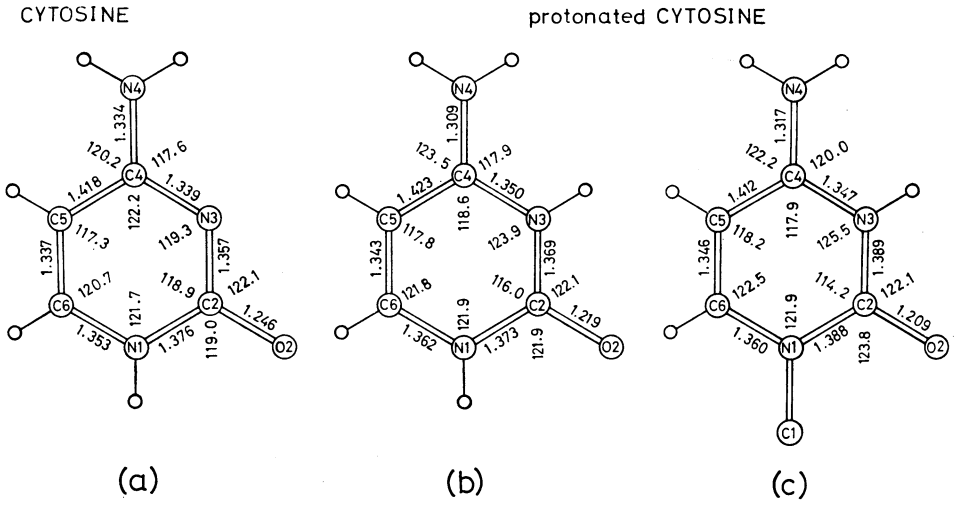


Fig. 4. A comparison of the dimensions of neutral and protonated cytosines.
(a): Average values of cytosine[15] and cytosine monohydrate.[15]
(b): This investigation.
(c): Average values of two forms of cytidylic acid,[16,17] 1-methylcytosine hydrochloride,[18] and 1-(β-D-arabinofuranosyl)-cytosine hydrochloride.[19]

Table 5. Least-squares planes and deviations of the atoms

X, *Y*, *Z* are in Å along the crystal axes, respectively.

Plane 1 (cytosine ring)
$0.7703X-0.0281Y-0.6371Z+0.1435=0$

Plane 2 (carboxyl group)
$-0.7884X-0.0335Y+0.6143Z+0.2537=0$

Plane 3 (amide group)
$-0.6740X-0.4542Y-0.5826Z+6.9236=0$

Deviations (*l*/Å)

Plane 1		Plane 2	
N(1)*	0.001	O(3)*	−0.001
C(2)*	0.008	O(4)*	−0.001
N(3)*	−0.011	C(7)*	0.002
C(4)*	0.005	C(8)*	−0.001
C(5)*	0.004	H(3)	−0.00
C(6)*	−0.007	H(41)	−0.02
O(2)	0.040		
N(4)	0.015	Plane 3	
H(1)	−0.01	N(5)*	0.002
H(3)	−0.01	O(5)*	0.001
H(41)	0.00	C(8)*	−0.001
H(42)	0.02	C(9)*	−0.002
H(5)	−0.04	H(9)	0.02
H(6)	−0.05	H(10)	−0.00

* Atoms included in the calculation of the least-squares plane.

that in other protonated species.[16–18] The dihedral angle between pyrimidine and carboxyl planes, 3.9°, is close to 3.0° found in CBG complex.

In the present and CBG complexes, a common feature of the "double hydrogen-bonded system" is observed, in spite of different molecular arrangements in their crystals. So, this seems to be a specific arrangement between cytosine and ionized carboxyl group of amino acid. The "double hydrogen-bonded system" could be a model for mutual recognition between nucleic acid and protein, such as binding mode between the unpaired cytosine in CCA terminus of *t*RNAs and the *C*-terminal carboxyl group in amino-acyl-*t*RNA synthetase or CCA pyrophosphorylase.

References

1) C. Tamura, N. Sakurai, and S. Sato, This Bulletin, **44**, 1473 (1971).
2) C. Tamura, T. Hata, S. Sato, and N. Sakurai, *ibid*, **45**, 3254 (1972).
3) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Chem. Lett.*, **1974**, 551.
4) R. F. Bryan and K. Tomita, *Acta Crystallogr.*, **15**, 1174 (1962).
5) B. Pullman and A. Pullman, "Quantum Biochemistry," John Wiley & Sons, New York (1963), p. 117.
6) T. Izumi, *Japan. J. Apbl. Phys.*, **10**, 1724 (1971).
7) A. Takenaka and Y. Sasada, presented at the Annual Meeting Cryst. Soc. Japan, Osaka, Japan, Abstracts, p. 27 (1973).
8) E. R. Howells, D. C. Phillips, and D. Rogers, *Acta Crystallogr.*, **3**, 210 (1950).
9) "International Tables for X-Ray Crystallography," Vol. 3, Kynoch Press, Birmingham (1962), p. 201.
10) T. Ashida, "The Universal Crystallographic Computing System (I)," ed. by T. Sakurai, The Crystallographic Society of Japan (1967), p. 76.
11) W. Fuller, *J. Amer. Chem. Soc.*, **81**, 1700 (1959).
12) T. J. Kistenmachen, D. J. Hunt, and R. E. Marsh, *Acta Crystallogr.*, **B28**, 3352 (1972).
13) L. Pauling, "The Nature of the Chemical Bond," Cornell University Press, Ithaca (1960), p. 260.
14) R. E. Marsh and J. Donohue, *Advanc. Protein Chem.*, **22**, 235 (1967).
15) R. J. McClure and B. M. Craven, *Acta Crystallogr.*, **B29**, 1234 (1973).
16) M. Sundaralingam and L. H. Jensen, *J. Mol. Biol.*, **13**, 914 (1965).
17) C. E. Bugg and R. E. Marsh, *ibid*, **25**, 67 (1967).
18) B. L. Trus and R. E. March, *Acta Crystallogr.*, **B28**, 1834 (1972).
19) J. S. Sherfinski and R. E. Marsh, *ibid*, **B29**, 192 (1973).
20) K. Hoogsteen, *ibid*, **16**, 28 (1963).