Complexes between Nucleotide Base and Amino Acid. IV. Crystal and Molecular Structure of Cytosine : *N,N*-Phthaloyl-DL-glutamic Acid Complex Dihydrate

Akio TAKENAKA,* Minoru OHKI, and Yoshio SASADA

Laboratory of Chemistry for Natural Products, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227

(Received February 19, 1980)

Crystals of cytosine: *N,N*-phthaloyl-DL-glutamic acid complex dihydrate were obtained from an aqueous solution of the two components. The space group is $P2_1/c$ with unit-cell dimensions of $a=5.243(1)$, $b=39.133(6)$, $c=9.192(1)$ Å, $\beta=93.66(1)°$, and $Z=4$. The structure was solved by the direct method and its parameters were refined by a block-diagonal least-squares method. The complex is formed by the two hydrogen bonds between the oxygen atoms of the ionized α-carboxyl group of glutamic acid and the nitrogen atoms, N(4) and protonated N(3), of cytosine. This binding mode is the same as that found in cytidine: *N*-benzyloxycarbonyl-L-glutamic acid complex, but quite different from that of 5-bromocytosine: *N*-acylglutamic acid complexes. The substituted bromine atom at C(5) of cytosine is thought to prescribe the binding mode in complex formation. The binding mode between cytosine and glutamic acid, which might be a possible elementary pattern of protein: nucleic acid interactions, is discussed in terms of dissociation constant.

In the studies of the elementary interaction patterns between amino acid and purine-pyrimidine base, we have found that *N*-tosyl-L-glutamic acid and *N,N*-phthaloyl-DL-glutamic acid are hydrogen-bonded to 5-bromocytosine with their γ-carboxyl group,[1,2] while Hata, Yoshikawa, Sato, and Tamura reported that in cytidine: *N*-benzyloxycarbonyl-L-glutamic acid complex dihydrate, cytosine base binds to α-carboxyl group of the amino acid.[3] It is interesting to examine what causes such different mode of cytosine: glutamic acid interactions. An X-ray analysis of the crystal structure of cytosine: *N,N*-phthaloyl-DL-glutamic acid complex has been performed, whose structure has revealed the effect of bromine atom on the complex formation.

Experimental and Structure Determination

Single crystals suitable for X-ray studies were obtained by slow evaporation of an aqueous solution containing equimolar amounts (0.2 mmol) of cytosine and *N,N*-phthaloyl-DL-glutamic acid, both of which were purchased from Tokyo Kasei Kogyo Co. The complex formation was confirmed by a preliminary X-ray experiment, an elemental analysis and IR spectrum of these crystals. Found: C, 48.0; H, 4.62; N, 13.2%. Calcd for $C_4H_5N_3O \cdot C_{13}H_{11}NO_6 \cdot 2H_2O$: C, 48.1; H, 4.75; N, 13.2%. Density was measured by the flotation method. Weissenberg photographs showed the systematic absences, $h0l$ for $l=2n+1$ and $0k0$ for $k=2n+1$, indicating the space group to be $P2_1/c$. A crystal $0.8\times0.8\times0.9$ mm was used for data collection on a Rigaku four-circle diffractometer with graphite-monochromated Mo $K\alpha$ radiation ($\lambda=0.71069$ Å). The unit-cell dimensions were calculated with 41 high-angle reflexions. The crystal data are given in Table 1.

TABLE 1. CRYSTAL DATA

Cytosine: *N,N*-Phthaloyl-DL-glutamic Acid Dihydrate	
$C_4H_5N_3O: C_{13}H_{11}NO_6: 2H_2O$	
Space group : $P2_1/c$	
$a=5.243(1)$ Å	$U=1882.2(4)$ Å^3
$b=39.133(6)$	$Z=4$
$c=9.192(1)$	$D_x=1.497$ g cm^{-3}
$\beta=93.66(1)°$	$D_m=1.49_7$

Intensities were measured in the ω scan mode with a scan width of 1° (in ω) plus α_1-α_2 divergence and a scan speed of 4° (in 2θ) min^{-1}. Five reference reflexions showed no significant intensity deterioration throughout the data collection. The intensities were corrected for Lorentz and polarization factors but not for absorption effects. Of 4306 independent reflexions in the range $2<2\theta<55°$, 1090 weak reflexions below background were considered as zero-reflexions; the observed threshold value, F_{lim}, was 2.55. The standard deviations were estimated by the equation of $\sigma^2(F_o)=\sigma_P^2(F_o)+qF_o^2$, where $\sigma_P(F_o)$ is evaluated by counting statistics and q (3.39×10^{-5}) was derived from the variations of the monitored reflexions.[4]

The structure was solved by the direct method and its parameters were refined by block-diagonal least-squares techniques. All the hydrogen atoms, found on a difference map, were included. The quantity minimized was $\sum w(|F_o|-|F_c|)^2$, with $w=1/\sigma^2(F_o)$. The zero-reflexions, for which $|F_c|>F_{lim}$, were included in the least-squares calculation by assuming $F_o=F_{lim}$ and $w=w(F_{lim})$. The final R was 0.55 for 2949 reflexions with $F_o>3\sigma(F_o)$; the maximum shift of parameters in the last cycle was 0.03σ for C, 0.45σ for H, 0.07σ for N, and 0.17σ for O atoms. Atomic scattering factors used were taken from "International Tables for X-Ray Crystallography."[5]

Atomic parameters are listed in Table 2, observed and calculated structure factors in Table 3.[6] Figure 1 shows a stereoscopic view of the cytosine: *N,N*-phthaloyl-DL-glutamic acid complex dihydrate. Bond distances and angles are shown in Fig. 2 together with the atomic numbering system. Bond angles involving hydrogen atoms are given in Table 4, least-squares planes together with deviations of atoms in Table 5, and hydrogen bond distances and angles in Table 6.

Results and Discussion

Molecular Structure. *N,N*-Phthaloyl-DL-glutamic acid is dissociated at the α-carboxyl group but not at the γ-carboxyl group, as confirmed by a difference synthesis. The bond distances and angles of the α- and γ-carboxyl groups are typical as those of the ionized and neutral carboxyl groups, respectively. The two identical C–O lengths of the α-carboxyl group suggest that the negative charges are equally localized on the two oxygen atoms.

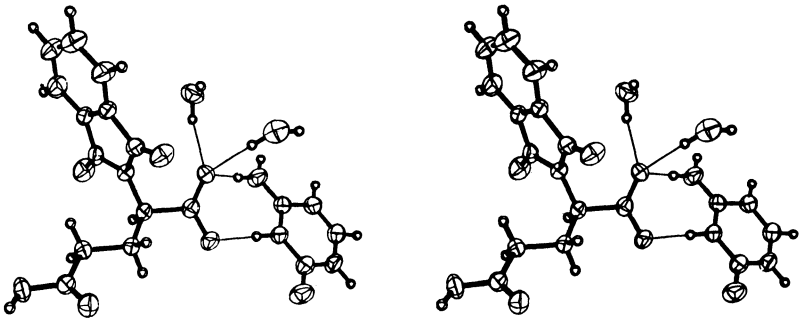


Fig. 1. Stereoview of the cytosine: *N,N*-phthaloyl-DL-glutamic acid complex dihydrate with 50% probability ellipsoids for the non-hydrogen atoms.

TABLE 2. FINAL POSITIONAL AND THERMAL PARAMETERS OF CYTOSINE: *N,N*-PHTHALOYL-DL-GLUTAMIC ACID COMPLEX DIHYDRATE

Standard deviations are given in parentheses. The anisotropic thermal factors has the form $\exp[-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+\beta_{12}hk+\beta_{13}hl+\beta_{23}kl)]$.

Atom	*x**	*y***	*z**	β_{11}*	β_{22}*	β_{33}*	β_{12}*	β_{13}*	β_{23}**
N (1)	4281 (3)	28608 (4)	3234 (2)	376 (9)	38 (1)	130 (3)	3 (2)	93 (8)	−42 (10)
C (2)	5852 (4)	26463 (5)	2533 (2)	282 (9)	49 (2)	103 (3)	2 (2)	11 (9)	0 (12)
N (3)	5507 (3)	23031 (4)	2776 (2)	304 (8)	40 (1)	96 (2)	15 (2)	97 (7)	2 (9)
C (4)	3758 (4)	21745 (5)	3660 (2)	275 (9)	48 (1)	85 (3)	10 (2)	54 (8)	23 (11)
C (5)	2196 (4)	24105 (5)	4380 (2)	303 (10)	48 (2)	103 (3)	14 (2)	120 (9)	16 (12)
C (6)	2513 (4)	27426 (5)	4145 (2)	337 (10)	53 (2)	115 (3)	17 (2)	70 (9)	−23 (12)
O (2)	7453 (3)	27482 (4)	1722 (2)	401 (8)	61 (1)	167 (3)	−6 (2)	201 (7)	115 (9)
N (4)	3567 (4)	18453 (4)	3820 (2)	444 (9)	44 (1)	147 (3)	18 (2)	246 (9)	45 (10)
C (7)	8001 (4)	15518 (5)	1519 (2)	209 (8)	41 (1)	97 (3)	5 (2)	56 (8)	−30 (10)
C (8)	8501 (4)	12757 (4)	381 (2)	231 (8)	35 (1)	73 (3)	6 (2)	16 (7)	3 (10)
C (9)	10535 (4)	13813 (5)	−653 (2)	287 (9)	36 (1)	81 (3)	0 (2)	62 (8)	−33 (10)
C (10)	10959 (4)	11072 (5)	−1774 (2)	309 (9)	44 (1)	77 (3)	−3 (2)	73 (8)	−44 (10)
C (11)	13054 (4)	11847 (5)	−2762 (2)	290 (9)	42 (1)	78 (3)	15 (2)	24 (8)	19 (11)
N (5)	9046 (3)	9433 (4)	1043 (2)	215 (6)	33 (1)	82 (2)	−2 (1)	18 (6)	7 (8)
O (3)	7526 (3)	14584 (3)	2773 (1)	356 (7)	48 (1)	86 (2)	6 (1)	129 (6)	−2 (8)
O (4)	7987 (3)	18540 (3)	1076 (2)	609 (9)	36 (1)	129 (2)	26 (2)	250 (7)	31 (8)
O (5)	13102 (3)	9550 (4)	−3823 (2)	488 (8)	58 (1)	114 (2)	−22 (2)	239 (7)	−204 (8)
O (6)	14583 (2)	14161 (3)	−2605 (2)	274 (6)	51 (1)	118 (2)	−10 (1)	75 (6)	−77 (8)
C (12)	7591 (4)	6499 (5)	726 (2)	277 (9)	44 (1)	93 (3)	−13 (2)	70 (8)	−103 (11)
C (13)	11130 (4)	8741 (5)	2032 (2)	238 (9)	44 (1)	87 (3)	2 (2)	55 (8)	9 (11)
C (14)	8846 (4)	3718 (5)	1601 (2)	320 (9)	35 (1)	93 (3)	−2 (2)	112 (8)	−36 (10)
C (15)	8182 (5)	308 (5)	1733 (3)	476 (12)	46 (2)	137 (4)	−20 (2)	90 (11)	−42 (13)
C (16)	9723 (5)	−1681 (5)	2684 (3)	756 (16)	36 (2)	160 (4)	−3 (3)	147 (13)	53 (14)
C (17)	11847 (5)	−332 (6)	3443 (3)	633 (15)	56 (2)	159 (4)	21 (3)	−8 (13)	134 (15)
C (18)	12515 (4)	3079 (5)	3303 (3)	439 (12)	53 (2)	135 (4)	15 (2)	16 (11)	106 (13)
C (19)	10955 (4)	5045 (5)	2376 (2)	280 (9)	36 (1)	89 (3)	5 (2)	81 (8)	13 (10)
O (7)	5721 (3)	6403 (4)	−111 (2)	361 (7)	62 (1)	150 (3)	−22 (2)	−112 (7)	−66 (9)
O (8)	12674 (3)	10845 (4)	2475 (2)	292 (7)	52 (1)	151 (3)	−17 (1)	−100 (7)	32 (9)
O (W1)	7221 (3)	9177 (4)	4659 (2)	440 (8)	79 (1)	96 (2)	8 (2)	107 (7)	89 (9)
O (W2)	9564 (3)	16577 (4)	5581 (2)	354 (7)	74 (1)	174 (3)	11 (2)	90 (7)	97 (10)

Atom	*x****	*y**	*z****	*B*(/Å²)*****	Atom	*x****	*y**	*z****	*B*(/Å²)*****
H (1)	448 (4)	3090 (5)	308 (2)	21 (5)	H (10B)	1142 (3)	890 (4)	−132 (2)	11 (4)
H (3)	643 (4)	2147 (5)	222 (2)	15 (5)	H (11)	1447 (5)	983 (6)	−432 (3)	64 (8)
H (5)	97 (4)	2322 (5)	503 (2)	18 (5)	H (15)	666 (4)	−54 (5)	120 (2)	18 (5)
H (6)	147 (4)	2920 (5)	462 (2)	25 (5)	H (16)	918 (4)	−407 (5)	283 (2)	31 (5)
H (4A)	462 (4)	1703 (5)	337 (2)	29 (5)	H (17)	1288 (4)	−184 (5)	414 (2)	41 (6)
H (4B)	237 (4)	1767 (6)	443 (2)	40 (6)	H (18)	1400 (4)	409 (5)	387 (2)	31 (5)
H (8)	686 (3)	1247 (4)	−21 (2)	2 (4)	H (W1A)	709 (4)	1063 (6)	395 (3)	42 (6)
H (9A)	1213 (3)	1444 (4)	−9 (2)	3 (4)	H (W1B)	857 (5)	934 (6)	505 (3)	68 (8)
H (9B)	992 (3)	1591 (4)	−116 (2)	0 (4)	H (W2A)	868 (5)	1807 (7)	612 (3)	72 (8)
H (10A)	942 (3)	1073 (4)	−241 (2)	9 (4)	H (W2B)	851 (6)	1584 (8)	466 (3)	111 (11)

*$\times 10^4$, **$\times 10^5$, ***$\times 10^3$, ****$\times 10$.

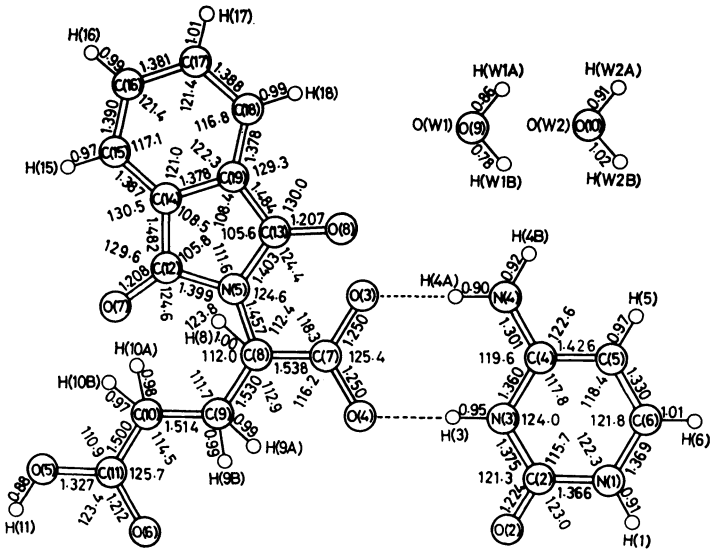


Fig. 2. Atomic numbering of the cytosine: *N,N*-phthaloyl-DL-glutamic acid complex dihydrate with bond lengths (l/Å) and angles (ϕ/°). E. s. d. 's are 0.003 Å and 0.2°, but 0.03 Å for those involving H atom. Bond angles involving H atoms are given in Table 4.

TABLE 4. BOND ANGLES INVOLVING HYDROGEN ATOMS OF CYTOSINE: *N,N*-PHTHALOYL-DL-GLUTAMIC ACID COMPLEX DIHYDRATE

	Angle (ϕ/°)		Angle (ϕ/°)
C(2)–N(1)–H(1)	117(2)	C(7)–C(8)–H(8)	106(1)
C(6)–N(1)–H(1)	121(2)	C(9)–C(8)–H(8)	108(1)
C(2)–N(3)–H(3)	118(2)	N(5)–C(8)–H(8)	106(1)
C(4)–N(3)–H(3)	118(2)	C(8)–C(9)–H(9A)	110(1)
C(4)–N(4)–H(4A)	120(2)	C(8)–C(9)–H(9B)	107(1)
C(4)–N(4)–H(4B)	117(2)	C(10)–C(9)–H(9A)	112(1)
H(4A)–N(4)–H(4B)	122(2)	C(10)–C(9)–H(9B)	109(1)
C(4)–C(5)–H(5)	119(2)	H(9A)–C(9)–H(9B)	106(2)
C(6)–C(5)–H(5)	123(2)	C(9)–C(10)–H(10A)	111(1)
C(5)–C(6)–H(6)	122(2)	C(9)–C(10)–H(10B)	112(1)
N(1)–C(6)–H(6)	116(2)	C(11)–C(10)–H(10A)	106(1)
		C(11)–C(10)–H(10B)	105(1)
H(W1A)–O(W1)–H(W1B)	109(3)	H(10A)–C(10)–H(10B)	108(2)
H(W2A)–O(W2)–H(W2B)	111(3)	C(11)–O(5)–H(11)	110(2)
		C(14)–C(15)–H(15)	119(2)
		C(16)–C(15)–H(15)	124(2)
		C(15)–C(16)–H(16)	117(2)
		C(17)–C(16)–H(16)	121(2)
		C(16)–C(17)–H(17)	119(2)
		C(18)–C(17)–H(17)	120(2)
		C(17)–C(18)–H(18)	122(2)
		C(19)–C(18)–H(18)	121(2)

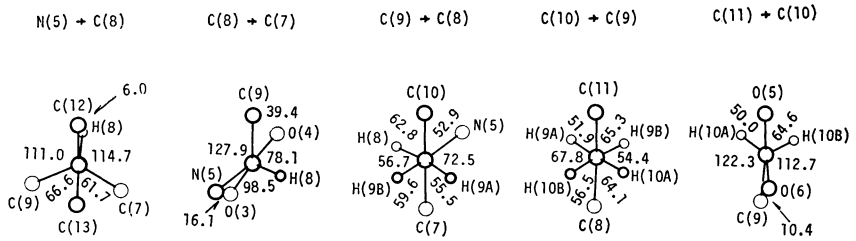


Fig. 3. Torsion angles (ϕ/°) of *N,N*-phthaloyl-DL-glutamic acid component.

TABLE 5. LEAST-SQUARES PLANES AND DEVIATIONS OF ATOMS FROM THE PLANES

X, *Y*, and *Z* are in Å along ***a****, ***b***, and ***c***, respectively. Asterisks denote atoms defining the plane. Standard deviations are given in parentheses.

Equation

Cytosine

$0.6791(6)X+0.0418(8)Y+0.7329(6)Z=4.069(8)$

Phthalimido group

$0.5646(6)X-0.2230(5)Y-0.7947(4)Z=1.339(4)$

Deviations(l/Å) of atoms from the plane

Cytosine		Phthalimido group	
N(1)*	−0.006	N(5)*	−0.012
C(2)*	0.006	C(12)*	0.007
N(3)*	0.000	C(13)*	−0.003
C(4)*	−0.004	C(14)*	0.015
C(5)*	0.003	C(15)*	0.002
C(6)*	0.004	C(16)*	−0.023
O(2)	0.007	C(17)*	−0.010
N(4)	−0.014	C(18)*	0.009
H(1)	−0.01	C(19)*	0.013
H(3)	−0.09	O(7)	0.025
H(4A)	0.01	O(8)	−0.012
H(4B)	−0.01	C(8)	0.007
H(5)	0.02	H(15)	−0.02
H(6)	0.01	H(16)	−0.09
		H(17)	−0.06
		H(18)	−0.01

The phthalimido group is almost planar, though the deviations of atoms from the mean plane indicate the slight bending of the group along its long axis (see Table 3). The N(5) and C(8) atoms also lie on this plane, suggesting an sp^2 hybridization of N(5). The bond distances and angles of the phthaloyl group are in good agreement with those in 5-bromocytosine: *N,N*-phthaloyl-DL-glutamic acid complex[2] and in the other related compounds.[7]

The torsion angles of the glutamic acid are shown in Fig. 3. The γ-carboxyl group has a synplanar C–C–C=O conformation as a common feature of carboxylic acids.[8] The extended conformation of the C–C chain is different from the folded form found in 5-bromocytosine: *N,N*-phthaloyl-DL-glutamic acid complex, where the C(9)–C(10) bond is in gauche conformation.[2] The conformation of the phthaloyl group is similar to that in 5-bromocytosine: *N,N*-phthaloyl-DL-glutamic acid.[2]

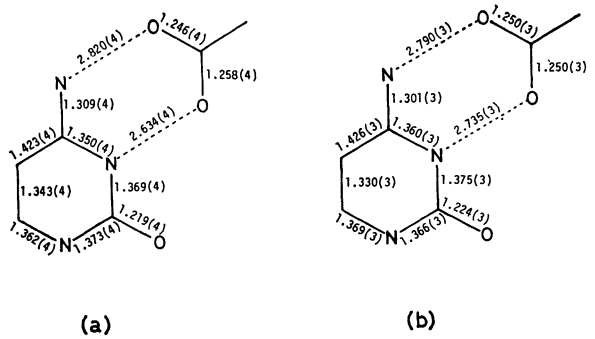
The hydrogen atom released from the α-carboxyl group is transferred to N(3) of cytosine molecule. As compared with the neutral cytosines, the protonation gives rise to a change of cytosine geometry similar to that discussed in cytosine: *N*-formylglycine complex.[9] Figure 4 compares the bond distances of cytosinium cations found in different environments. In cytosine: *N*-formylglycine complex (a), the positive and negative charges are mainly localized on the N(3) atom of

TABLE 6. HYDROGEN BOND DISTANCES AND ANGLES OF CYTOSINE: *N,N*-PHTHALOYL-DL-GLUTAMIC ACID COMPLEX DIHYDRATE

Standard deviations are given in parentheses.

Distances (l/Å)		Angles (ϕ/°)	
N(3)···O(4)	2.735(2)	O(3)···O(W2)···O(2)[f]	115.7(1)
H(3)···O(4)	1.79(2)	H(4B)[e]···O(W2)–H(W2A)	126(2)
N(4)···O(3)	2.790(2)	H(4B)[e]···O(W2)–H(W2B)	90(2)
H(4A)···O(3)	1.91(2)	O(W2)–H(W2B)···O(3)	163(3)
N(4)···O(W2)[a]	2.828(2)	O(W2)–H(W2A)···O(2)[f]	156(3)
H(4B)···O(W2)[a]	1.92(2)	O(W1)–H(W1A)···O(3)	162(3)
O(3)···O(W2)	2.838(2)	O(3)···O(W1)···O(5)[h]	111.9(1)
O(3)···H(W2B)	1.84(3)	H(W1A)–O(W1)···H(11)[h]	106(2)
O(3)···O(W1)	2.746(2)	H(W1B)–O(W1)···H(11)[h]	119(2)
O(3)···H(W1A)	1.91(2)	C(7)–O(3)···N(4)	110.9(2)
O(2)···O(W2)[d]	2.807(2)	C(7)–O(3)···O(W1)	146.2(2)
O(2)···H(W2A)[d]	1.95(3)	C(7)–O(3)···O(W2)	132.3(2)
N(1)···O(6)[c]	2.940(2)	N(4)···O(3)···O(W1)	97.2(1)
H(1)···O(6)[c]	2.04(2)	N(4)···O(3)···O(W2)	77.4(1)
O(5)···O(W1)[b]	2.649(2)	O(W1)···O(3)···O(W2)	70.8(1)
H(11)···O(W1)[b]	1.79(3)	C(7)–O(3)···H(4A)	109(1)
		C(7)–O(3)···H(W1A)	143(1)
Angles (ϕ/°)		C(7)–O(3)···H(W2B)	137(1)
N(3)–H(3)···O(4)	176(2)	H(4A)···O(3)···H(W1A)	97(1)
C(2)–N(3)···O(4)	117.5(2)	H(4A)···O(3)···H(W2B)	77(2)
C(4)–N(3)···O(4)	117.8(2)	H(W1A)···O(3)···H(W2B)	74(2)
N(3)···O(4)–C(7)	114.5(2)	O(5)–H(11)···O(W1)[b]	165(3)
H(3)···O(4)–C(7)	114(1)	C(11)–O(5)···O(W1)[b]	118.6(2)
N(4)–H(4A)···O(3)	165(2)	C(11)–O(6)···N(1)[g]	134.8(2)
N(4)–H(4B)···O(W2)[a]	172(2)	C(11)–O(6)···H(1)[g]	136(1)
C(4)–N(4)···O(3)	115.7(2)	N(1)–H(1)···O(6)[c]	170(2)
C(4)–N(4)···O(W2)[a]	112.8(2)	C(2)–N(1)···O(6)[c]	115.1(2)
O(3)···N(4)···O(W2)[a]	130.7(1)	C(6)–N(1)···O(6)[c]	122.4(2)
N(4)[e]···O(W2)···O(3)	78.5(1)	C(2)–O(2)···O(W2)[d]	143.1(2)
N(4)[e]···O(W2)···O(2)[f]	108.9(1)	C(2)–O(2)···H(W2A)[d]	136(1)

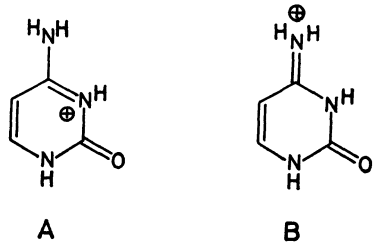
Symmetry codes (a) $-1+x, y, z$, (b) $1+x, y, -1+z$, (c) $-1+x, 1/2-y, 1/2+z$, (d) $x, 1/2-y, -1/2+z$, (e) $1+x, y, z$, (f) $x, 1/2-y, 1/2+z$, (g) $1+x, 1/2-y, -1/2+z$, (h) $-1+x, y, 1+z$.



(a) (b)

Fig. 4. A comparison of bond distances between (a) cytosine: *N*-formylglycine complex and (b) the present complex, cytosine: *N,N*-phthaloyl-DL-glutamic acid.

cytosine and its hydrogen-bonded oxygen atom of carboxyl group, respectively, and in consequence their distance becomes considerably shorter than that of N(4)···O. In the present complex (b), however, the two C–O distances of α-carboxyl group are identical and in addition these two oxygen atoms are hydrogen-bonded to N(3) and N(4) of cytosine at almost the same distances. In such a situation, the cytosinium cation is perturbed, so that the contribution from the canonical formula B increases, as compared with (a) in which the canonical formula A is predominant.



A B

Slight differences in bond lengths seem to be consistent with this interpretation. Similar trend is also observed in cytidine: salicylic acid complex.[10]

The planarity of the present cytosine is excellent; only the N(4) atom deviates slightly from the mean plane (Table 5).

Crystal Structure. As shown in Figs. 1 and 5, the complex is formed by the two hydrogen bonds between the oxygen atoms of the dissociated α-carboxyl group of glutamic acid and the nitrogen atoms, N(4) and protonated N(3), of cytosine. This double hydrogen-bonded system is the same as that found in cytidine: *N*-benzyloxycarbonyl-L-glutamic acid complex.[3] The two N···O distances, N(3)···O(3) and N(4)···O(4), are nearly equal to each other, although the planar α-carboxyl group makes an angle of 35.3(7)° with the cytosine plane. This feature of the system is contrary to that of the above complex.[3]

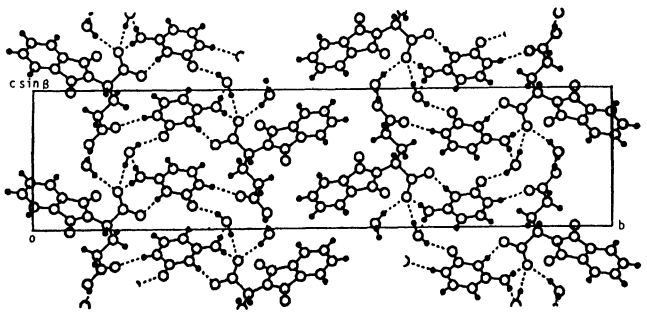
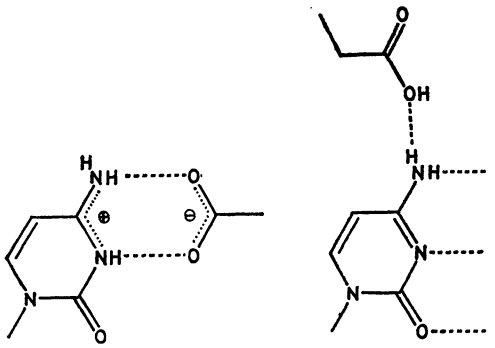


Fig. 5. Crystal structure of the cytosine: *N,N*-phthaloyl-DL-glutamic acid dihydrate, projected along the c axis.

The O(3) atom is also an acceptor of hydrogen bonds with the two water molecules. The O(W2) atom donates its hydrogen atom to O(2) of neighbouring cytosine and at the same time, accepts from N(4) of another cytosine. The γ-carboxyl group forms two hydrogen bonds; O(5) is a donor for O(W1) and O(6) is an acceptor from N(1) of adjacent cytosine. One of the two hydrogen atoms of water O(W1) is free from hydrogen bonding but directed to O(5) at $x,y,1+z$ with the O(W1)–H(W1B)···O(5) angle of 175(3)° and the distances, 3.306(2) Å for O(W1)···O(5) and 2.53(3) Å for H(W1B)···O(5).

The cytosine molecules are arranged parallel to one another on the c-glide plane. On the other hand, the phthaloyl groups come together around the inversion centres and packed parallel to form a sheet.

Complex Formation. In the present cytosine: *N,N*-phthaloyl-DL-glutamic acid complex and cytidine: *N*-benzyloxycarbonyl-L-glutamic acid complex,[3] the



(a) (b)

Fig. 6. Hydrogen bonding modes between cytosine and carboxyl group; (a) Type I and (b) Type II.

TABLE 7. DISSOCIATION CONSTANTS OF CYTOSINE, ITS DERIVATIVES AND *N*-ACYLAMINO ACIDS

	pK_a	Ref.
Cytosine	4.58	12
Cytidine	4.08	12
5-Bromocytosine	3.04	13
Cytidine 5′-phosphate	4.5	12
Cytidine 2′-deoxy-5′-phosphate	4.44	12
N-Acetylglycine	3.67	14
N-Benzoylglycine	3.65	14
N-Carbamoylglycine	3.88	14
N-Formylglycine	3.43	14
Glycylglycine	3.15	14
Glycylalanine	3.15	14
	pK_a^I	pK_a^{II} Ref.
N-Benzoylglutamic acid	3.49	4.99 14

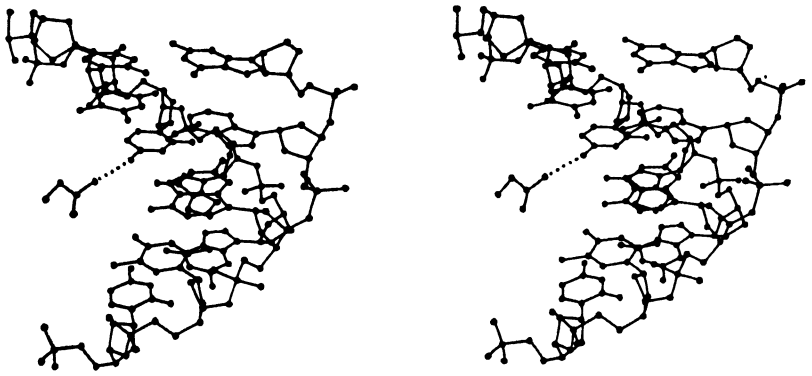


Fig. 7. A fitting feature between acidic side group of protein and cytosine moiety paired with guanine residue in the major groove of a double helical DNA (conformation B), in Type II mode.

binding modes are the same one (hereafter we designate Type I, see Fig. 6(a)) inspite of the different substitutions of glutamic acid and of cytosine, and the different crystal packings. On the other hand, the complexes with 5-bromocytosine are formed in the different mode (hereafter Type II, see Fig. 6(b)) for the two different acyl groups of the amino acids.[1,2] Consequently, it seems that the bromine atom at C(5) of cytosine determines the binding mode. The geometrical change of cytosine skeleton, which is caused by the steric and electronic effects of bromination at C(5), can be related to their dissociation constants in solution.[11] Accordingly, the complex formation modes may be interpreted in terms of pK_a values.

Table 7 lists the dissociation constants of cytosine, its derivatives and some *N*-acylamino acids. The pK_a values of *N*-acylglycines are almost invariant for different acyl groups. So, it is reasonable to assume pK_a^I and pK_a^{II} values of *N*-acylglutamic acids to be close to those of *N*-benzoylglutamic acid. Accordingly, the pK_a value of the α-carboxyl group of *N*-acylglutamic acids may be lower than that of cytosine or cytidine. In a solution containing *N*-acylglutamic acid and cytosine or cytidine, the latter molecule is easily protonated by the amino acid and both interact with each other in the Type I binding mode. On the other hand, the pK_a value of 5-bromocytosine is even less than that of the α-carboxyl group of *N*-acylglutamic acid. Therefore, 5-bromocytosine molecule could scarcely take a proton from *N*-acylglutamic acid. Consequently, the α-carboxyl groups are associated to form a dimer or tetramer[2] in the crystal, whereas 5-bromocytosine molecules are coupled with each other by the two N(4)–H···N(3) hydrogen bonds, the amino group of which are hydrogen-bonded to the γ-carboxyl group.[1,2]

Protein: Nucleic Acid Interaction. In the previous paper, we have proposed that the two binding modes, Type I and Type II, are elementary patterns of protein: nucleic acid interactions.[1,2] The pK_a values of proteins at the C terminal and at the acidic side groups respectively might be near pK_a^I of glycylglycine or glycylalanine and pK_a^{II} of *N*-acylglutamic acid. On the other hand, in nucleic acid cytosine bases would have a pK_a close to that of cytidine 5′-phosphate (CMP),

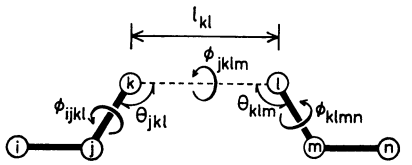


Fig. 8. Definition of the binding parameters in Table 8.

Table 8. Binding parameters between cytosine and carboxyl group in fig. 7

Binding parameter			
Interaction bond distance (/Å)			
l_{kl}			2.80
Bond angle(/°)			
θ_{jkl}			158
θ_{klm}			150
Torsion angle (/°)			
ϕ_{ijkl}			183
ϕ_{jklm}			−8
ϕ_{klmn}			−13
Atoms in cytosine		Atoms in carboxyl group	
i	N(3)	l	O
j	C(4)	m	C(1)
k	N(4)	n	C(2)

which is a value between those of carboxyl groups in proteins mentioned above. Therefore, the above discussion in terms of pK_a is also applicable and supports our previous proposal that the Type I is a binding pattern, *e.g.* between the unpaired cytosine in CCA terminus of tRNA and the C-terminal carboxyl group in enzyme, and the Type II is an elementary pattern of interaction between paired cytosine and acidic side group of amino acid residue.

To examine the above-mentioned argument, we have attempted to fit the binding geometry of Type II into a major groove of a double helical DNA (conformation B) using an interactive modelling programme[15] with computer graphics; atomic coordinates were taken from the works of Arnott and Hukins.[16] Trials show that the major groove† of B-DNA is not so wide for binding

† The major groove of B-DNA has been suggested to be a binding site from its surface accessibility[17] and from its methylation reactions.[18]

of carboxyl group to cytosine residue, due to their directional properties of hydrogen bondings. Base sequences also restrict the region. Interacting feature free from any steric hindrance, shown in Fig. 7, was constructed with small adjustment of torsional angles or bond angles. The binding data used are given in Table 8, parameters of which are defined in Fig. 8. The hydrogen bonding patterns described in the present series, together with those in adenine: indole,[19] adenine: carboxyl group,[20] and adenine: amide group[21] systems, will be useful for understanding the protein: nucleic acid interactions.

The present work was partially supported by a Grant-in-Aid for Scientific Research from the Ministry of Education and by the Kawakami Foundation, to which the authors' thanks are due.

References

1) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **49**, 3493 (1976).
2) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **50**, 90 (1977).
3) T. Hata, M. Yoshikawa, S. Sato, and C. Tamura, *Acta Crystallogr., Sect. B*, **31**, 312 (1975).
4) L. E. McCandlish and G. H. Staut, *Acta Crystallogr., Sect. A*, **31**, 245 (1975).
5) "International Tables for X-Ray Crystallography," Kynoch Press, Birmingham (1974), Vol. IV. p. 71.
6) Table 3 has been deposited as Document No. 8034, at the Office of the Bulletin of the Chemical Society of Japan.
7) K. Gerst and C. E. Nordman, *Acta Crystallogr., Sect. B*, **33**, 1588 (1977); M. Sikirica, I. Vicković, and D. Viterbo, *Acta Crystallogr., Sect. B*, **35**, 1735 (1979).
8) L. Leiserowitz, *Acta Crystallogr., Sect. B*, **32**, 775 (1976).
9) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **48**, 848 (1975).
10) C. Tamura, M. Yoshikawa, S. Sato, and T. Hata, *Chem. Lett.* **1973**. 1221 (1973).
11) M. Kato, A. Takenaka, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **52**, 49 (1979).
12) A. Albert, "Synthetic Procedures in Nucleic Acid Chemistry," ed by W. W. Zorbach and R. S. Tipson, John Wiley and Sons, Inc., New York (1973), p. 1.
13) I. Wempen and J. J. Fox, *J. Am. Chem. Soc.*, **86**, 2474 (1964).
14) "Lange's Handbook of Chemistry," 11th ed, ed by J. A. Dean, McGraw-Hill Book Co., New York (1973), pp. 5—15.
15) A. Takenaka and Y. Sasada, *J. Cryst. Soc. Jpn.*, **22**, 214 (1980).
16) S. Arnott and D. W. L. Hukins, *Biochem. Biophys. Res. Commun.*, **47**, 1504 (1972).
17) C. J. Alden and S. H. Kim, *J. Mol. Biol.*, **132**, 411 (1979).
18) L. Johnsrud, *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 5314 (1978).
19) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **50**, 2573 (1977); M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Acta Crystallogr., Sect. B*, **33**, 2954 (1977).
20) A. Takenaka and Y. Sasada, 38th National Meeting of the Chemical Society of Japan, Nagoya, October 1978, Abstr., p. 1004.
21) M. Takimoto, A. Takenaka, and Y. Sasada, Annual Meeting of the Crystallographic Society of Japan, Nagoya, November 1979, Abstr., p. 1B-13.