Cytosine-Carboxylate Interactions: Crystal Structure of Cytosinium Hydrogen Maleate

Thailampillai Balasubramanian,[#] **Packianathan Thomas Muthiah,*** **and Ward Thomas Robinson**[†]

Department of Chemistry, Bharathidasan University, Tiruchirapalli-620024, India

†Department of Chemistry, University of Canterbury, New Zealand

(Received November 13, 1995)

Interactions of carboxylate groups of proteins with nucleobases are some of the important stereochemical patterns involved in protein-nucleic acid recognition. The title compound serves as a model for such interactions, involving constrained carboxyl groups. Space group: $P\bar{1}$, $a=7.424(1)$, $b=7.952(2)$, $c=8.248(2)$ Å, $\alpha=106.61(3)°$, $\beta=94.97(3)°$, and $\gamma=91.26(3)°$. The structure was solved by direct methods. Cytosine is protonated at N(3). The usual fork-like interaction of the carboxylate groups with cytosine is absent and a network of hydrogen bonding involving the cytosine moiety and hydrogen maleate is present. The interaction of carboxylate group and the amino group of cytosine observed here is reminiscent of the carboxylate–GC pair interaction observed in a protein–DNA complex (DNA repressor of phage 434 complex). There is also a C(6)–H····O (of maleate) interaction.

Studies of interactions between nucleobases and functional side chains of amino acids are helpful to understand the elementary stereochemical patterns involved in protein-nucleic acid recognition. Among the four nucleobases (A, T, G, and C), G can be recognized by its carboxylate group.[1] Adenine-carboxylate interactions[2] and cytosine-carboxylate interactions[3—5] have been extensively studied by X-ray crystallography. Protonated cytosine forms a fork-like hydrogen bonding pattern with carboxylate groups as shown in Fig. 1. In this study, we have a pair of constrained carboxylate groups (maleic acid) to interact with cytosine. Hydrogen maleate ion is a structural motif having carboxyl-carboxylate interactions which are known to occur in many proteins.[6] Hence it is interesting to note the nature of the interactions of this structural motif with cytosine.

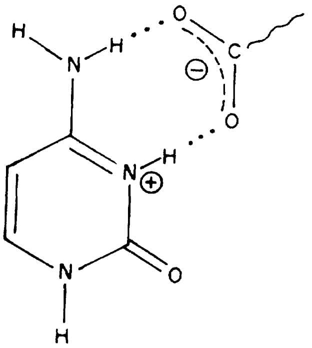

Fig. 1. Common mode of cytosine-carboxylate interaction.

Experimental

Cytosine (560 mg) dissolved in 80 ml of water was mixed with maleic acid (585 mg) dissolved in 50 ml of water. The mixture was warmed for about 30 min over a water bath and on cooling, thin plate-shaped crystals appeared within three days.

X-Ray Structure Analysis. X-Ray data were collected in a four-circle diffractometer. The crystallographic data are summarized in Table 1. Lorentz and polarization corrections were applied. The structure was solved by direct methods using the package SHELXS86[7] and refined using the package SHELXL93.[8] All the hydrogens were located by difference Fourier mapping and were included in the refinement with isotropic temperature factors, but the non-hydrogen atoms were refined anisotropically. The refinement was completed at $R=0.03$. Final atomic coordinates of the non-hydrogen atoms and equivalent isotropic temperature factors are listed in Table 2. The tables of the atomic coordinates and the isotropic temperature factors for hydrogen atoms, and F_o-F_c are deposited as Document No. 69043 at the Office of the Editor of Bull. Chem. Soc. Jpn.

Results and Discussions

The bond lengths, bond angles, and mean-square planes were calculated using the program PARST.[9] An ORTEP[10] view of the complex is shown in Fig. 2. The bond lengths and the bond angles of the cytosine moiety are listed in Table 3 and compared with the compilation based on the Cambridge Structural Database.[11] Cytosine is protonated at N(3) leading to widening of the corresponding internal angle as compared with neutral cytosine.[11] The bond lengths and bond angles of the hydrogen maleate ion are in agreement with those reported for the crystal structure of (+)-*N*,*N*-dimethyl-3-(4-chlorophenyl)-3-(2-pyridyl)propylammonium hydrogen maleate.[12] In the hydrogen maleate ion, the intra-

---

[#] Present address: Department of Physics, Nehru Memorial College, Puthanampatti, India 621007.

Table 1. Crystal Data and Structure Refinement

Empirical formula	$C_8H_9N_3O_5$
Formula weight	227.18
Temperature/K	130(2)
Wavelength/Å	0.71073
Crystal system	Triclinic
Space group	$P\bar{1}$
Unit cell dimensions a/Å	7.4240(10)
b/Å	7.952(2)
c/Å	8.248(2)
α/°	106.61(3)
β/°	94.97(3)
γ/°	91.26(3)
Volume/Å^3	464.3(2)
Z	2
Density (calcd)/Mg m^{-3}	1.625
Absorption coefficient/mm^{-1}	0.137
F(000)	236
Theta range for data collection	2.59° to 25°
Index ranges	$0 \leq h \leq 5$, $-9 \leq k \leq 9$, $-9 \leq l \leq 9$
Reflections collected	1451
Independent reflections	1314 [R(int) = 0.0150]
Refinement method	Full-matrix least-squares on F^2
Data/restraints/parameters	1314/0/180
Goodness-of-fit on F^2	1.094
Final R indices [$I > 2\sigma(I)$]	$R_1 = 0.0312$, $wR_2 = 0.0752$
R indices (all data)	$R_1 = 0.0384$, $wR_2 = 0.0805$
Largest diff. peak and hole	0.174 and -0.273 e A^{-3}

Table 2. Atomic Coordinates ($\times 10^4$) and Equivalent Isotropic Displacement Parameters (Å$^2 \times 10^3$)
U(eq) is defined as one third of the trace of the orthogonalized Uij tensor.

Atom	x	y	z	U(eq)
N1	3395(2)	3340(2)	3736(2)	17(1)
C2	3830(3)	2712(2)	5096(2)	15(1)
O2	4856(2)	3515(1)	6323(1)	20(1)
N3	3034(2)	1103(2)	4980(2)	16(1)
C4	1933(3)	135(2)	3626(2)	16(1)
N4	1260(2)	−1384(2)	3622(2)	19(1)
C5	1559(3)	853(2)	2246(2)	19(1)
C6	2299(3)	2428(2)	2351(2)	19(1)
C21	6925(3)	1746(2)	12376(2)	16(1)
O21	6269(2)	287(2)	12332(2)	23(1)
C21′	7885(2)	2731(2)	13666(1)	22(1)
C22	6527(3)	2363(2)	10836(2)	18(1)
C23	7133(3)	3807(2)	10505(2)	18(1)
C24	8393(3)	5273(2)	11522(2)	16(1)
O24	8742(2)	6495(2)	10956(1)	22(1)
O24′	9151(2)	5254(2)	13006(1)	20(1)

Table 3. Bond Lengths (Å) and Bond Angles (°) for Cytosine Moiety Compared with a Previous Compilation[11)]

	Present work	A previous compilation	
		Protonated	Neutral
Bond lengths:			
N1–C2	1.370(2)	1.381(2)	1.399(4)
C2–N3	1.373(2)	1.387(2)	1.356(3)
N3–C4	1.353(2)	1.352(2)	1.334(2)
C4–C5	1.421(2)	1.413(3)	1.426(4)
C5–C6	1.334(3)	1.341(2)	1.337(2)
N1–C6	1.353(2)	1.362(3)	1.364(2)
C2–O2	1.219(2)	1.211(2)	1.237(2)
C4–N4	1.297(2)	1.313(3)	1.337(4)
Bond angles:			
C6–N1–C2	122.9(2)	121.5(1)	120.6(1)
N1–C2–N3	115.2(2)	114.9(2)	118.9(2)
C4–N3–C2	124.3(1)	125.1(2)	120.0(2)
N3–C4–C5	117.6(2)	117.5(2)	121.8(2)
C6–C5–C4	118.8(2)	118.5(1)	117.6(2)
C5–C6–N1	121.2(2)	122.5(1)	121.0(2)
O2–C2–N1	122.8(2)	123.5(2)	119.2(2)
O2–C2–N3	122.0(2)	121.6(1)	121.9(2)
N4–C4–N3	119.9(2)	119.5(2)	117.9(3)
N4–C4–C5	122.5(2)	123.0(3)	120.3(2)

ion hydrogen bond is of length 2.416(2) Å and is asymmetric as reported earlier.[12)] The two ends of the maleate ion have different environments. The angle O–H····O is 176(2)°, which has been reported earlier to be 159°.[12)] The cytosines are paired about inversion centers through a pair of N(1)–H····O(2) hydrogen bonds.

The usual fork-like interaction (involving the carboxylate group, N3–H, and C4 amino group of cytosine) observed in the earlier cytosine carboxylate complexes is absent in this structure. Instead the carboxylate group is involved in hydrogen bonding with the amino group of cytosine without the involvement of Watson–Crick sites. Hence this type of interaction may represent a model for interaction between

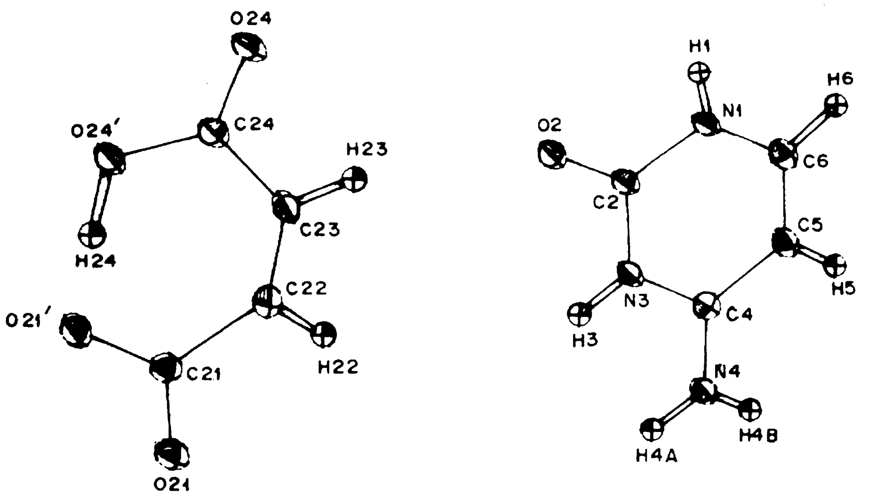

Fig. 2. An ORTEP view of cytosinium hydrogen maleate.

Table 4. Hydrogen Bonding Geometry

D–H (Å)	D···A (Å)	H···A (Å)	D–H···A (°)
N1–H1	N1···O2 (1)	H1···O2 (1)	N1–H1···O2 (1)
0.87(2)	2.809(2)	1.94(2)	175(2)
C6–H6	C6···O24 (1)	H6···O24 (1)	C6–H6···O24 (1)
1.02(2)	3.126(2)	2.14(2)	160(1)
N4–H4B	N4···O24′ (2)	H4B···O24′ (2)	N4–H4B···O24′ (2)
0.88(2)	2.959(2)	2.46(2)	117(2)
N4–H4B	N4···O24 (2)	H4B···O24 (2)	N4–H4B···O24 (2)
0.88(2)	2.864(2)	1.99(2)	174(2)
O24′–H24	O24′···O21′ (0)	H24···O21′ (0)	O24′–H24···O21′ (0)
1.05(3)	2.416(2)	1.37(3)	173(2)

Equivalent positions: (0) x, y, z; (1) $-x+1, -y+1, -z+1$; (2) $+x-1, +y-1, +z-1$.

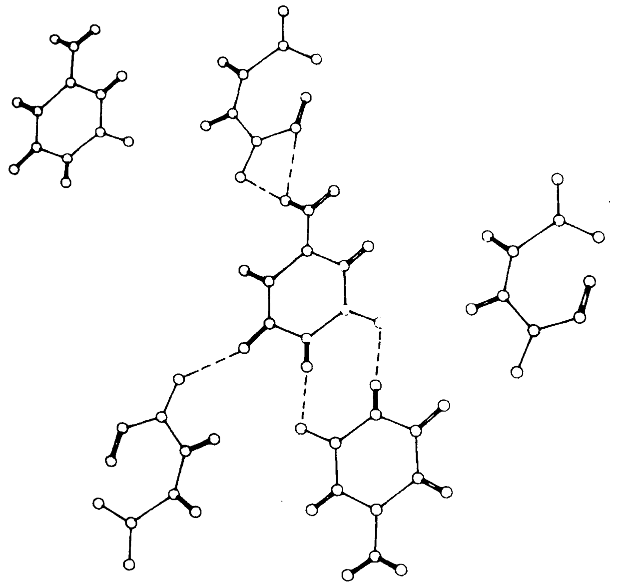

Fig. 3. Hydrogen bonding pattern in cytosinium hydrogen maleate.

the side chains of amino acid residues Asp and Glu in protein and the cytosine group of DNA. Since the usual fork-like interaction of carboxylate with cytosine would not occur in double-stranded DNA, the hydrogen bonding pattern described in this structure is potentially an interesting example of DNA–protein interaction in the major groove. This type of hydrogen bonding pattern has also been observed in the crystal structure of a protein–DNA complex (DNA-repressor of phage 434 complex) where the carboxylate group of a glutamic acid residue of the protein interacts with the C4 amino group of a GC pair of the nucleic acid.[13]

It is also interesting to note that there is C(6)–H····O (of maleate) hydrogen bonding and this tendency of C(6)–H of cytosine to participate in hydrogen bonding has been observed earlier.[14] This type of C–H····O interaction involving nucleobases may also play a subtle role in nucleic acid-ligand recognition. The hydrogen bonding patterns and their parameters are listed Table 4. A view of these interactions is shown in Fig. 3.

TB acknowledges the University Grants Commission (UGC) for a Teacher Fellowship and PTM is a Career Awardee of UGC.

References

1) C. Helene and G. Lancelot, *Prog. Biophys. Mol. Biol.*, **39**, 1 (1982).

2) A. Takenaka and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **55**, 680 (1982).

3) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **48**, 848 (1975).

4) C. Tamura, N. Sakurai, S. Sato, and N. Sakurai, *Bull. Chem. Soc. Jpn.*, **45**, 3254 (1972).

5) M. Gdaniec, B. Brychi, and M. Szafran, *J. Mol. Struct.*, **195**, 57 (1989).

6) L. Sawyer and M. N. G. James, *Nature*, **295**, 79 (1982).

7) G. M. Sheldrick, "SHELXS86, Program for automatic solution of crystal structures," University of Göttingen, Germany (1986).

8) G. M. Sheldrick, "SHELXS93, Program for crystal structure refinement," University of Göttingen, Germany (1993).

9) PARST: A system of computer routines for calculating molecular parameters from results of crystal structure analysis: M. Nardelli, *Comput. Chem.*, **7**, 95 (1983).

10) C. K. Johnson, "ORTEP Program, Technical report ORNL-3794," Oak Ridge National Laboratory (1965).

11) R. Taylor and O. Kennard, *J. Mol. Struct.*, **78**, 1 (1982).

12) M. N. G. James and G. J. B. Williams, *Can. J. Chem.*, **52**, 1872 (1974).

13) A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison, *Science*, **242**, 99 (1988).

14) "Hydrogen Bonding in Biological Structures," ed by G. A. Jefferey and W. Saenger, Spinger-Verlag, New York (1991), p. 156.