2234 © 1983 *The Chemical Society of Japan* *Bull. Chem. Soc. Jpn.*, ***56***, *2234—2237 (1983)* [Vol. 56, No. 8

Crystal Structure of Adenine-1-(2-Carboxyethyl)uracil (1 : 1) Complex. A Model for Interactions of Amino Acid Side Chains with Nucleic Acid Base Pair

Shigeo Fujita, Akio Takenaka, and Yoshio Sasada*
Laboratory of Chemistry for Natural Products, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 227

(Received March 15, 1983)

As a model for protein–nucleic acid interactions, crystals of the title complex were prepared from a water–ethanol (1 : 1) solution containing equimolar amounts of adenine and 1-(2-carboxyethyl)uracil. The space group is $P2_1/c$ with the unit cell dimensions of $a=12.567(1)$, $b=13.358(1)$, $c=8.032(1)$ Å and $\beta=93.28(1)°$. The crystal structure was determined by X-ray analysis. The final R value was 0.053. The adenine and uracil moieties form a reversed Hoogsteen base pair in which N(7) and N(6) of adenine are hydrogen-bonded with N(3) and O(2) of uracil, respectively. The carboxyl group is dissociated and hydrogen-bonded to the protonated N(1) of adenine in the adjacent base pair at a distance of 2.542(3) Å. It is suggested that the protonation of adenine–uracil reversed Hoogsteen base pair could occur to make the base pair more stable, and the dissociated carboxyl group may be hydrogen-bonded to it in the protein–nucleic acid interactions.

Protein–nucleic acid interactions are essential for gene expression and its regulation.[1] Recently, the three-dimensional structures of cro repressor,[2] λ repressor[3] and catabolite gene activator protein[4] were determined by the X-ray method, and a binding model of cro repressor to its operator site on DNA was constructed by adjusting the structure so as to make hydrogen bonds, assuming the two-fold symmetry.[5] This model suggests that the hydrogen bonds between the nucleotide bases and the amino acid side chains must be the leading part of the specific protein–nucleic acid interactions. For understanding of the molecular basis of recognition between the two kinds of macromolecules, it is important to know the hydrogen bonds between their components. There should be some elementary patterns of hydrogen bonding with different degree of preference or stability, and their suitable combination would give a specificity of the interactions.

Protein crystallography has shown that the atoms located at the protein–protein or protein–substrate interface are packed as close as those in crystals of small molecules.[6] It may be expected that the recognition between different biomolecules are performed in a circumstance similar to crystalline field and the suitable model crystals give us information at atomic level for interpreting the specificity. By X-ray analyses of model crystals in which both the components coexist, we have been studying the elementary binding patterns between nucleotide bases and amino acid side chains.[7–14] We have found the three different binding patterns between adenine and carboxyl group: a carboxyl group is (i) hydrogen-bonded to N(3),[10] (ii) doubly hydrogen-bonded to N(6) and N(7),[15] and (iii) dissociated and doubly hydrogen-bonded to N(6) and protonated N(1).[11] In combination with uracil, a carboxyl group is hydrogen-bonded to O(2)[12] or doubly hydrogen-bonded to N(3) and O(4).[16] Thus, a carboxyl group interacts with both adenine and uracil in the binary model systems. What preference is displayed by a carboxyl group if both adenine and uracil bases are present? We have prepared model crystals that contain these components and revealed the binding geometry by the X-ray method.

Experimental and Structure Determination

1-(2-Carboxyethyl)uracil was prepared from uracil and β-propiolactone, according to Kondo *et al.*[17] The titration curve of 1-(2-carboxyethyl)uracil (0.01 mol dm^{-3}) with NaOH (0.008 mol dm^{-3}) was recorded using Toshiba-Beckman pH meter, and the value of pK_a was evaluated to be 4.1 from the curve. Colorless plate crystals were obtained by slow evaporation of a water–ethanol (1 : 1) solution containing equimolar amounts of adenine and 1-(2-carboxyethyl)uracil. The syste-

Table 1. Fractional coordinates and isotropic temperature factors

The B values accompanied with < > are the equivalent isotropic temperature factors calculated from anisotropic thermal parameters using the equation $B=8\pi^2(U_1+U_2+U_3)/3$, where U_1, U_2, and U_3 are principal components of mean square displacement matrix U. Values in < > are anisotropicity defined by$(\sum(B-8\pi^2U_i)^2/3)^{1/2}$ and those in ( ) are e.s.d.'s; they refer to last decimal places.

Atom	x	y	z	B/Å^2
N(1)	0.3663(2)	-0.0302(2)	0.0976(3)	2.9<13>
C(2)	0.3117(2)	0.0517(2)	0.0369(4)	2.9<9>
O(2)	0.2340(2)	0.0444(1)	-0.0633(3)	4.0<25>
N(3)	0.3488(2)	0.1420(2)	0.0918(3)	3.1<16>
C(4)	0.4331(2)	0.1604(2)	0.2060(4)	3.2<10>
O(4)	0.4569(2)	0.2466(1)	0.2434(3)	4.6<27>
C(5)	0.4839(2)	0.0710(2)	0.2681(4)	3.6<17>
C(6)	0.4505(2)	-0.0186(2)	0.2143(4)	3.4<14>
C(9)	0.3341(2)	-0.1313(2)	0.0370(4)	3.1<10>
C(10)	0.2438(3)	-0.1711(2)	0.1329(5)	3.8<11>
C(11)	0.2111(2)	-0.2748(2)	0.0707(4)	3.0<8>
O(12)	0.1264(2)	-0.2855(1)	-0.0178(3)	3.8<16>
O(13)	0.2702(2)	-0.3462(1)	0.1102(3)	4.8<34>
N(1')	-0.0061(2)	0.3312(2)	-0.3065(3)	2.9<9>
C(2')	-0.0152(2)	0.4314(2)	-0.2944(4)	3.5<12>
N(3')	0.0464(2)	0.4920(2)	-0.2056(3)	3.4<17>
C(4')	0.1267(2)	0.4408(2)	-0.1218(4)	4.9<10>
C(5')	0.1440(2)	0.3380(2)	-0.1260(3)	2.5<8>
C(6')	0.0716(2)	0.2790(2)	-0.2186(4)	2.6<3>
N(6')	0.0737(2)	0.1803(2)	-0.2231(4)	3.2<13>
N(7')	0.2326(2)	0.3129(2)	-0.0255(3)	2.9<12>
C(8')	0.2659(2)	0.3990(2)	0.0358(4)	3.2<13>
N(9')	0.2065(2)	0.4794(2)	-0.0161(3)	3.17<151>
H(3)	0.318(2)	0.193(2)	0.054(4)	0.9(7)
H(5)	0.531(2)	0.079(2)	0.342(4)	1.8(8)
H(6)	0.479(2)	-0.087(2)	0.257(3)	1.2(6)
H(9A)	0.309(2)	-0.127(2)	-0.087(4)	1.4(7)
H(9B)	0.393(3)	-0.184(3)	0.054(4)	3.2(8)
H(10A)	0.274(3)	-0.178(3)	0.252(5)	3.6(9)
H(10B)	0.182(3)	-0.128(3)	0.129(5)	4.0(10)
H(1')	-0.065(2)	0.289(2)	-0.388(4)	2.9(8)
H(2')	-0.073(2)	0.461(2)	-0.355(3)	1.1(6)
H(6A')	0.029(3)	0.145(3)	-0.277(4)	2.9(9)
H(6B')	0.128(3)	0.148(2)	-0.157(4)	2.6(8)
H(8')	0.325(2)	0.406(2)	0.120(3)	1.1(6)
H(9')	0.221(3)	0.558(3)	0.023(4)	3.9(9)

matic absence of reflections showed the space group to be $P2_1/c$. The density was measured by the flotation method in hexane and carbon tetrachloride mixture.

A crystal, $0.5\times0.4\times0.1$ mm^3 in size, was used for data collection on a Rigaku four-circle diffractometer with Ni-filtered Cu $K\alpha$ radiation (λ=1.5418 Å). Accurate unit cell dimensions were derived from 2θ values of 80 high-angle reflections. The crystal data are as follows: $C_5H_5N_5\cdot C_7H_8N_2O_4$ F.W. 319.3, monoclinic, space group $P2_1/c$, a=12.567(1), b=13.358(1), c=8.032(1) Å, β=93.28(1)°, Z=4, D_m=1.57 g cm^{-3}, and D_c=1.575 g cm^{-3}.

Intensities were measured in the $\omega/2\theta$ scan mode with a scan width of 2.6°(2θ) plus α_1—α_2 divergence and a scan speed of 4°(2θ) min^{-1}. Five reference reflections monitored showed no significant intensity deterioration. Correction was not made for absorption effect. Of 2087 independent reflections within the range $2°<2\theta<125°$, 154 weak reflections below background were considered zero-reflections. The structure was solved by the direct method and its parameters were refined by the full-matrix least-squares method. The weighting scheme and the treatment of the zero-reflections were the same as those in Ref. 11. The value of q and F_{lim} were taken as 2.99×10^{-5} and 1.108, respectively. All the hydrogen atoms were found on a difference map and included in the refinement. The R value was 0.053 for 1914 reflections with $F_o>3\sigma$; the maximum shift in atomic parameters was 0.3σ in the last cycle. The final atomic parameters are listed in Table 1.[18] The atomic scattering factors were taken from Ref. 19.

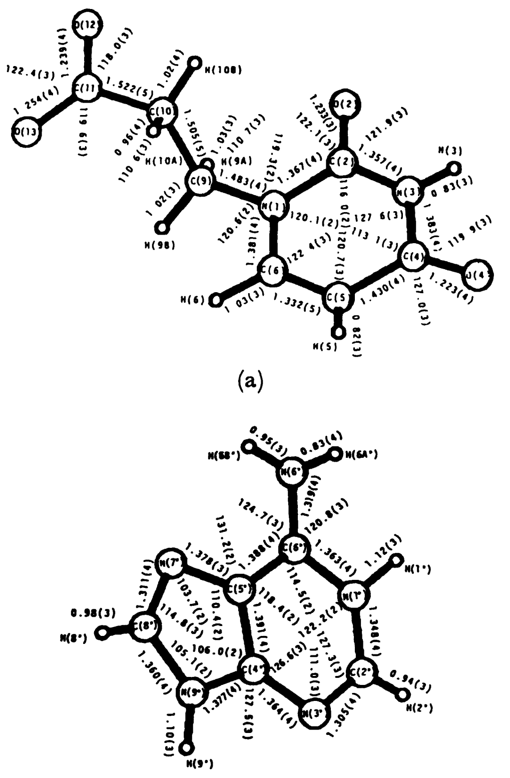
Results and Discussion

Molecular Structure. The carboxyl group in 1-(2-carboxyethyl)uracil is dissociated and a proton acceptor is N(1′) of adenine, as confirmed by the difference synthesis.

Bond lengths and angles are shown in Fig. 1. The dimensions of the carboxyl group are typical of carboxylate ion. The pyrimidine ring is planar within 0.015 Å. The bond lengths and angles are similar to those found in 1-(2-carboxyethyl)uracil.[12] Closer examination shows, however, that the C(2)–N(3) bond is slightly shortened and the C(2)–O(2) bond is lengthened. These changes indicate the increase of the contribution from the quinonoidal form, accompanied by the polarization at H(3)–N(3)–C(2)–O(2). The polarization must be induced by the hydrogen bonds with the adenine monocation. The methylene chain is extended, contrary to the folded form found in 1-(2-carboxyethyl)uracil.[12]

As compared with the neutral adenine, the N(1′)–C(6′) bond is lengthened, the C(6′)–N(1′)–C(2′) bond angle is expanded and the C(5′)–C(6′)–N(1′) and N(1′)–C(2′)–N(3′) bond angles are contracted. Such changes are typical of adenine monocation protonated at N(1′).[20–24] The six-membered ring is ruffled, as usually observed in the purine ring.[25]

Hydrogen Bond in Crystal. As shown in Figs. 2 and 3, the uracil moiety and the adeninium cation form a reversed Hoogsteen base pair[26] through the two hydrogen bonds N(3)–H···N(7′) (2.842(3) Å for N···N) and N(6′)–H···O(2) (2.951(3) Å for N···O). The geometries of both the hydrogen bonds are similar to those of the reversed Hoogsteen base pair without protonation, found in the complexes of 9-ethyladenine: 1-methyl-5-bromouracil[27] and adenosine: 5-bromouridine.[28] The two ring systems are nearly coplanar with a dihedral angle of 5.0(1)° between their respective least-squares planes; similar values are reported for the other reversed Hoogsteen base pairs.[27–28]



(a)

(b)

Fig. 1. Bond lengths (l/Å) and angles (ϕ/°) with e.s.d.'s in the parentheses. (a) 1-(2-Carboxyethyl)uracil and (b) adenine. Bond angles involving hydrogen atoms are deposited.

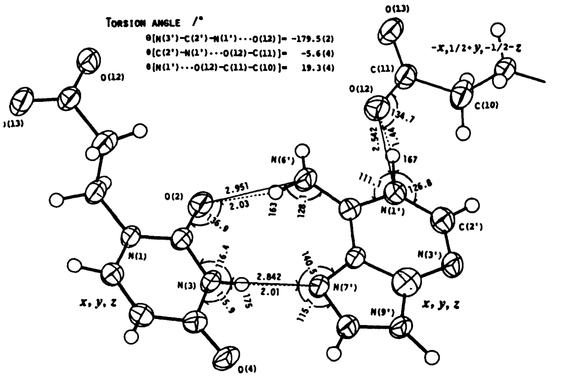


Fig. 2. A projective view of binding mode of the carboxyl group to the adenine–uracil base pair with hydrogen bond distances (l/Å) and angles (ϕ/°). E.s.d.'s are 0.003 Å and 0.2°, respectively, for distances and angles involving nonhydrogen atoms. The corresponding values involving hydrogen atoms are 0.03 Å and 2°, respectively.

The carboxyl group interacts with the protonated N(1′) of adenine through the strong N(1′)–H···O(12) hydrogen bond; the N···O distance is 2.542(3) Å and the torsion angles, θ[N(3′)–C(2′)–N(1′)···O(12)], θ[C(2′)–N(1′)···O(12)–C(11)], and θ[N(1′)···O(12)–C(11)

Table 2. Hydrogen bond distances and angles[a)]
Standard deviations are given in parentheses.

Hydrogen bond distance	l/Å		l/Å
N(3′)···N(6′) (i)[b)]	2.972(4)	N(9′)···O(13) (ii)	2.646(3)
N(3′)···H(6A′) (i)	2.26(4)	H(9′)···O(13) (ii)	1.57(3)
Hydrogen bond angle	ϕ/°		ϕ/°
N(3′)···H(6A′)–N(6′) (i)	145(3)	N(9′)–H(9′)···O(13) (ii)	163(3)
C(2′)–N(3′)···N(6′) (i)	98.1(2)	C(4′)–N(9′)···O(13) (ii)	115.4(2)
C(4)′–N(3′)···N(6′) (i)	150.8(2)	C(8′)–(9′)···O(13) (ii)	139.5(2)
N(3′)···N(6′)–C(6′) (i)	147.1(2)	N(9′)···O(13)–C(11) (ii)	114.5(2)
N(3′)···N(6′) (i)···O(2) (i)	83.7(1)		

a) The other hydrogen bonds found in the present crystal are shown in Fig.2. b) Symmetry codes: (i) $-x$, $1/2+y$, $-1/2-z$; (ii) x, $1+y$, z.

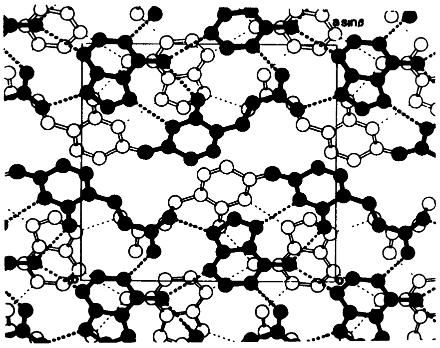


Fig. 3. Crystal structure of adenine: 1-(2-carboxyethyl)-uracil (1 : 1) complex, projected along the c axis. The broken lines indicate the hydrogen bonds.

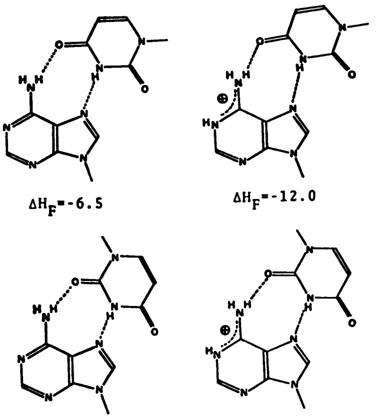


Fig. 4. Formation energies (ΔH_F/kcal mol^{-1}) of four types of adenine–uracil base pairs calculated with the CNDO/2 method.

C(10)] are −179.5(2), −5.6(4), and 19.3(4)°, respectively. The other oxygen atom O(13) is hydrogen-bonded to N(9′) at a distance of 2.646(3) Å. These two hydrogen bonds are very short compared with normal N–H···O hydrogen bonds. Between adeninium cations related by 2_1 screw, the N(6′)–H···N(3′) hydrogen bond is formed; the distance of 2.972(4) Å is normal, but this hydrogen bond deviates to some extent from linearity, the N(6′)–H···N(3′) being 145(3)°. The O(4) atom of uracil moiety does not participate in hydrogen bonding and surrounded by the hydrophobic groups. The O(4) atom points to the hydrogen atom of C(6) at $1-x$, $1/2+y$, $1/2-z$ at a distance of 3.357(3) Å for O···C. The base pairs related by 2_1 screw around $x=0$ and $z=1/4$ form a ribbon along the b axis. The detailed hydrogen bond distances and angles are given in Fig. 2 or Table 2.

Hydrogen Bonding between Carboxyl Group and Adenine–Uracil Base Pair. The present crystal structure suggests that the proton transfer from the carboxyl group to N(1) of adenine, accompanied by hydrogen bonding between them, makes the reversed Hoogsteen base pair more stable. This is confirmed by a CNDO/2 calculation[29)] of the formation energies; it is obvious from Fig. 4 that the protonation at N(1) stabilizes by about 5 kcal mol^{-1} both the reversed and normal Hoogsteen base pair.** It is plausible that the base pair to which a suitable proton donor is close may be protonated in a hydrophobic conditions. But such stabilization may be unlikely for the Watson-Crick base pair because of the small pK_a value for N(7) or N(3) of adenine (pK_a<1). If the protonation is compelled, it would occur at N(1) of adenine (pK_a=4.25) to lead a break-down of Watson-Crick base pair.[30)]

The dissociation constant, as an indication, of 1-(2-carboxyethyl)uracil (pK_a=4.1) is comparable to those of the acidic side chains of amino-acid derivatives and peptides,[31)] which shows that the chemical behavior of the carboxyl group of model compound is similar to that of the acidic side chains of protein. Therefore, one could expect that the interaction between the carboxylate ion and the charged reversed Hoogsteen base pair found in the present model crystal is also realized in a living system and takes some part in recognition of the special type of base pair by protein side chains.

As mentioned in the previous section, the specificity in protein–nucleic acid interactions would arise from the suitable combinations of preferred hydrogen bonds between their components which are supported by the three-dimensional structure of macromolecule. It is important to examine if the hydrogen bonding found in the present model crystal could be applied to the binding of the amino acid side chain with bases in a particular secondary or tertiary structure of nucleic acid.

The Hoogsteen base pair commonly exists between U8 and A14 residues in t-RNA for which, it is known,

---

** 1 cal=4.186 J.

aminoacylsynthetase,[32] elongation factor Tu,[33] ethidium bromide[34] and Mg^{2+} ion[35] can be accessible. Based on the geometry of the present hydrogen bond, model fitting of the dissociated carboxyl group into the t-RNA^{phe} molecule[36] was tried by interactive computer graphics. The carboxyl group is close to the G20 residue which does not form a pair with other bases. X-Ray studies have shown that this residue, located at the surface of the molecule, is mobile even in crystalline state.[37] If the guanine moiety of G20 is rotated by 60° around its glycosidic bond (the resulting χ^{CN} of 51.2° is within the range reported for purine nucleosides and nucleotides[38]), there is no abnormal contacts between atoms. Therefore, the present binding could be justified as one of possibilities of the extention from the model crystal structure to the protein-nucleic acid interactions.

Recently, it has been reported that the carbamoyl group is also hydrogen-bonded to N(1) of adenine in the normal Hoogsteen base pair between adenine and thymine in the crystalline complex of 9-(2-carbamoylethyl)adenine:1-methylthymine.[14] So, N(1) of adenine may be a preferential binding site in the normal and reversed Hoogsteen pairs.

The present work was supported in part by a Grant-in-Aid for Scientific Research No. 57380026 from the Ministry of Education, Science and Culture, and Kawakami Foundation, to which the authors' thanks are due. A part of calculation was carried out on a HITAC M200 at the Computer Center of the Institute for Molecular Science.

References

1) "The Mechanism of Protein Synthesis and Its Regulation," ed by L. Bosch, North-Holland Publishing Company, Amsterdam, London (1972).

2) W. F. Anderson, D. H. Ohlendorf, Y. Takeda, and B. W. Matthews, *Nature* (*London*), **290**, 754 (1981).

3) C. O. Pabo and M. Lewis, *Nature* (*London*), **298**, 443 (1982).

4) D. B. Mckay and T. A. Steitz, *Nature* (*London*), **290**, 744 (1981).

5) D. H. Ohlendorf, W. F. Anderson, R. G. Fisher, Y. Takeda, and B. W. Matthews, *Nature* (*London*), **298**, 718 (1982).

6) D. M. Segal, E. A. Padlan, G. H. Cohen, S. Rudikoff, M. Potter, and D. R. Davies, *Proc. Natl. Acad. Sci. U.S.A.*, **71**, 4298 (1974); H. Frauenfelder, G. A. Petsko, and D. Tsernoglou, *Nature* (*London*), **280**, 558 (1979); J. Baldwin and C. Chothia, *J. Mol. Biol.*, **129**, 175 (1979).

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

8) M. Ohki, A. Takenaka, H. Shimanouchi, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **50**, 2573 (1977).

9) A. Takenaka, M. Ohki, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **53**, 2724 (1980).

10) A. Takenaka and Y. Sasada, 38th National Meeting of the Chemical Socirty of Japan, Nagoya, October 1978, Abstr. 1004.

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

12) S. Fujita, A. Takenaka, and Y. Sasada, *Acta Crystallogr., Sect. B*, **38**, 2936 (1982).

13) M. Takimoto, A. Takenaka, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **54**, 1635 (1981).

14) M. Takimoto, A. Takenaka, and Y. Sasada, *Bull. Chem. Soc. Jpn.*, **55**, 2734 (1982).

15) T. Ishida, M. Inoue, A. Ota, and T. Kurihara, *J. Chem. Soc., Chem. Commun.*, **1980**, 1074.

16) R. Destro and R. E. Marsh, *Acta Crystallogr., Sect. B*, **28**, 2971 (1972).

17) K. Kondo, T. Sato, and K. Takemoto, *Chem. Lett.*, **1973**, 967.

18) Lists of structure factors, anisotropic thermal parameters and least-squares plane equations and deviations of atoms from these planes have been deposited as Document No. 833, at the office of the Chemical Society of Japan.

19) "International Tables for X-ray Crystallography," Kynoch Press, Birmingham (1974), Vol. IV, p. 71.

20) K. Shikata, T. Ueki, and T. Mitsui, *Acta Crystallogr., Sect. B*, **29**, 31 (1973).

21) T. J. Kistenmacher and T. Shigematsu, *Acta Crystallogr., Sect. B*, **30**, 166 (1974).

22) H.-S. Shieh and D. Voet, *Acta Crystallogr., Sect. B*, **31**, 2192 (1975).

23) V. Langer, K. Huml, and L. Lessinger, *Acta Crystallogr., Sect. B*, 34, 2229 (1978).

24) B. E. Hingerty, J. R. Einstein, and C. H. Wei, *Acta Crystallogr., Sect. B*, **37**, 140 (1981).

25) J. Sletten and L. H. Jensen, *Acta Crystallogr., Sect. B*, **25**, 1608 (1969).

26) T. D. Sakore, S. S. Tavale, and H. M. Sobell, *J. Mol. Biol.*, **43**, 361 (1969).

27) L. Katz, K. Tomita, and A. Rich, *Acta Crystallogr.*, **21**, 754 (1966).

28) A. E. Haschemeyer and H. M. Sobell, *Acta Crystallogr.*, **18**, 525 (1965).

29) J. A. Pople and G. A. Segal, *J. Chem. Phys.*, **44**, 3289 (1966).

30) Ph. Hermann and E. Fredericq, *Nucleic Acids Res.*, **4**, 2939 (1977).

31) T. Miyazawa, private communication.

32) S. Yamashiro-Matsumura and M. Kawata, *J. Biol. Chem.*, **256**, 9308 (1981).

33) S. Bertram and R. Wagner, *Biochem. Int.*, **4**, 117 (1982).

34) M. Liebman, J. Rubin, and M. Sundaralingam, *Proc. Natl. Acad. Sci. U. S. A.*, **74**, 4821 (1977).

35) S. R. Holbrook, J. L. Sussman, R. W. Warrant, G. M. Church, and S.-H. Kim, *Nucleic Acids Res.*, **4**, 2811 (1977).

36) F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, J. M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).

37) C. D. Stout, H. Mizuno, S. T. Rao, P. Swaminathan, J. Rubin, T. Brennan, and M. Sundaralingam, *Acta. Crystallogr., Sect. B*, **34**, 1529 (1978).

38) C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, **94**, 8205 (1972).