Appendix A

Stability of Oxidized Base and its Mispair in DNA: Quantum Mechanics Calculation and Molecular Dynamics Simulation

Abstract

5-formyluracil (FoU) is a potentially mutagenic lesion of thymine (T) produced in DNA by ionizing radiation and various chemical oxidants. The quantum mechanics (QM) calculation to compute pairing energies of FoU with a purine base was performed at the B3LYP/6-31G**//B3LYP/6-31G**++ level, considering various possible tautomeric, rotameric and ionized form of FoU. The pairing energies of FoU in keto form with either adenine (A) or guanine (G) are comparable to those of T. Although the tautomerism to enol provides triple hydrogen bonds with G, the energy penalty is not fully compensated by the extra hydrogen bond energy. These QM results lead to the conclusion that the ionization at N3 position of FoU would mainly account for the increased mispairing rate of FoU since the deprotonated FoU preferentially form H-bonds with G rather than A and therefore FoU has one more extra possibility of base pairing. The following molecular dynamics (MD) simulations for DNA dodecamers with normal A:T base pair, A:FoU base pair and G:FoU base mispair showed that hydrogen bonds in FoU paired with adenine remained stable in the duplex during the whole simulation, while G:FoU dodecamer showed slightly larger structural fluctuation since it contains non Watson-Crick pairs in the middle. The formyl group of FoU in the anti conformation affects the hydration pattern around the DNA structure. A water molecule that makes a bridge of H-bond between O7 of FoU and

O2P of phosphate seems to be responsible for the well-ordered solvent structure. The interesting result is that, even though the formyl group is located on the major groove side, its presence actually results in severe narrowing of minor grooves. No significant change in helical and backbone parameters is shown for A:FoU and G:FoU dodecamer except for the large shear in G-FoU pairs, which is obvious in Wobble-type geometry.

A.1 Introduction

The modification on DNA bases induces the formation of base mispairing during replication, which is fatal in keeping the genetic integrity in living organism. 5-formyluracil (FoU), one of well-known DNA base lesions is the oxidation product of thymine (T) by ionizing γ-radiation, Fenton-type reactions, and quinone-mediated UV-A photosensitization[1-3]. Privat and Sowers proposed that the electron-withdrawing formyl group increases the stability of the deprotonated form of FoU, which could exist in a non-negligible amount since FoU has a lower pK_a value close to physiological pH than T[4]. The deprotonated form of FoU would be mispaired with guanine (G) in a canonical Watson-Crick geometry. In the following replication process, G might form a correct pair with C and this leads to miscoding (i.e. starting with T, it ends up with C). Masaoka and co-workers also observed that the miscorporation ratio with deoxyguanosine monophosphate (dGMP) increased when FoU on the DNA template was substituted for T and this ratio also increased with increasing pH[5]. It supports the idea proposed by Privat and Sowers that the deprotonated form of FoU plays a key role in mispair mechanism.

The general repair steps carried out by DNA repair enzymes are detection, recognition and removal of mutagenic lesions from DNA. The pathway most commonly employed to remove incorrect bases (like uracil) or damaged bases (like 3-methyladenine) is called base excision repair (BER)[6]. Initially individual DNA glycosylases are targeted to distinct base lesions, which are flipped and cleaved out by the enzymes (damage-specific step) and then a damage-general step restores correct DNA base sequences. Several DNA glycosylases responsible for repair of FoU have been suggested, but the repair mechanism on the molecular level is not well understood yet. In *Escherichia coli*, FoU is reported to be removed from a DNA by the AlkA enzyme with efficiency comparable to that of 7-methylguanine, a good substrate for AlkA[7]. It was proposed that the electron deficient bases flip out from the DNA duplex to form strong π-donor/acceptor interaction with electron-rich aromatic amino acids present in the active site of AlkA. The MutS

proteins involved in methyl-directed mismatch repair also recognize FoU paired with G, but do not recognize it with A. The MutS complex with FoU:G inhibited the activity of AlkA to FoU and thus two independent repair pathways might exist[8]. Zhang *et al.* performed the trapping assay with $NaBH_4$ that is the clue for formation of Schiff base intermediate with NH_2 group in the enzyme at the abasic site[9]. They observed the trapped complex for the Nth, Nei and MutM protein in *E. coli* and also the cleaved bases for these enzymes. With the AlkA protein, no trapping was observed, but the repair mechanism by it cannot be excluded since other pathways without forming the Schiff base intermediate are plausible.

Even though the repair processes help maintain the genetic integrity, the cells are always vulnerable to having base lesions and the following mispairing. The presence of a non-natural base such as FoU would cause the structural changes in DNA double helix. When neutral FoU is mispaired with G and forms the non-canonical hydrogen bond; i.e. in this case, the Wobble type, this local change in H-bonding geometry can cause the overall changes in DNA double helix structure. One of the well-known examples showing the sequence-dependent conformational characteristic is the narrowing of minor grooves in the middle AT-tracts of DNA double helix. It is suggested that the N3 of A and O2 of T in AT base stacks form the electronegative pocket and then the counter cations, e.g., Na^+, are bound to that site and pull two bases closer together. Several studies have been carried out to show the correlation between the width of minor grooves and the location of counter ions in the simulation[10, 11]. One of the reasons why the width of the minor groove matters is that some drugs actually bind to the minor groove. For example, the antitumor antibiotic *netropsin* binds to the B-DNA double helix, especially at the AT base pair regions, without intercalating[12]. The hydration pattern is also critical for the stability of DNA structure and this pattern strongly depends on the sequence of DNA. The hydration spine in the AT-tracts is a good example[13]. Sometimes the hydration pattern also plays a crucial role in protein-DNA interaction. The similar hydration patterns of the protein-DNA interface in the trp

repressor-DNA complex and the naked DNA target were seen and it is proposed that both protein and DNA specially recognize each other's hydration pattern[14].

In this report, we examine the stability of FoU in free base pair system and when incorporated in DNA double helix. We compute pairing energies of various free DNA base pairs with the density functional theory, focusing mainly on mispairing of FoU with G. We also consider pairings of deprotonated form and enol tautomer of FoU with G in all possible hydrogen-bonding geometries. In order to see how stable the oxidized base and the following mispairs are in DNA double helix and how they affect the overall DNA conformation, the molecular dynamics (MD) simulations for DNA dodecamers with normal A:T base pair, A:FoU base pair and G:FoU base mispair are then carried out.

A.2 Computational Methods

A.2.1 Quantum Mechanics (QM) calculation of pairing energies in free DNA base systems

All QM calculations were performed using the Jaguar v4.1 quantum chemistry software[15]. The geometries for 1-methyl pyrimidine, 9-methyl purine bases and all pyrimidine-purine base pairs were first optimized in the gas phase at the B3LYP/6-31G** level. The vibration frequencies for thermodynamic quantities were also calculated at the same level. Then the 6-31G**++ basis set was used for the final geometry optimization starting from the 6-31G**-optimized geometry. Since the calculation of vibration frequencies is a quite time consuming, the diffuse function was not included in the first step. To validate the exclusion of the diffuse function in the calculation of vibration frequencies, we have considered the following combinations of basis sets and compared the calculated enthalpies of base pairing with experimental ones:

(1) 6-31G**/6-31G** (No diffuse function is included in both steps.)

(2) 6-31G**/6-31G**++ (The preliminary geometry optimization and the calculation of frequencies are done with 6-31G** basis set and then the geometry is re-optimized with 6-31G**++ basis set.)

(3) 6-31G**++/6-31G**++ (The diffuse function is included in both steps.)

No scaling factor was adopted in the frequency calculation. All thermodynamic quantities were computed at 300 K, based on standard ideal-gas statistical mechanics and the rigid-rotor harmonic oscillator approximation. The enthalpy (or free energy) for each species is defined as:

$$H_{300K}\,(or\ G_{300K}) = E_{0K} + ZPE + \Delta H_{0\to 300K}\,(or\ \Delta G_{0\to 300K}),$$

where the E_{0K} is the total energy of the molecules at 0 K calculated from QM, ZPE is the zero-point energy and, $\Delta H_{0\to 300K}$ (or $\Delta G_{0\to 300K}$) is the change of enthalpy (or free energy) from 0 K to 300 K.

The single point energy calculation was carried out for the free energy of solvation in water, G_{solv}, for the final optimized structure at the B3LYP/6-31G**++ level. The solvation free energies are computed with a self-consistent reaction field method by solving the Poisson-Boltzmann equation. For the dielectric constant of water, we used $\varepsilon_{H2O} = 80.37$ which is at 20 °C[16]. The probe radius was set to 1.40 Å. We used the default values for the van der Waals radii of atoms[17]. The free energy of the system in aqueous solution is given by

$$G_{aq} = G_{300K} + G_{solv}\,.$$

The calculations of pairing free energies were performed for various DNA base pairs, focusing on mispairing of FoU with G. Pairing of deprotonated form or enol tautomer of FoU with G was also considered in all possible hydrogen-bonding geometries. In this calculation, the basis set superposition error (BSSE), which is the artificial lowering in the complex energy relative to that of the separated monomers since the complex basis set is larger than that of each

monomer, should be taken into account. Since the free bases undergo the conformational change upon pairing, their relaxation energy terms were also incorporated into the estimation of the BSSE correction[18]. Therefore, the following BSSE-correction energy, E_{BSSE}, should be added to the "raw" pairing energy, ΔE_{0K}:

$$E_{BSSE} = \left[E_{AB}^{a}(A) - E_{AB}^{a\cup b}(A)\right] + \left[E_{AB}^{b}(B) - E_{AB}^{a\cup b}(B)\right]$$

$$\Delta E_{0K} = E_{AB}^{a\cup b}(AB) - \left[E_{A}^{a}(A) + E_{B}^{b}(B)\right]$$

where *a* and *b* are the basis sets for corresponding bases, *A* and *B*, and the *A*, *B* and *AB* on the subscript represent the geometries where the energies for the species inside the parenthesis were computed. The final equation form for BSSE-corrected free energies in gas and in aqueous solution is followed as:

$$\Delta G_{300K}(g) = \left\{G_{300K}(AB) - \left[G_{300K}(A) + G_{300K}(B)\right]\right\} + E_{BSSE}$$

$$\Delta G_{300K}(aq) = \left\{G_{aq}(AB) - \left[G_{aq}(A) + G_{aq}(B)\right]\right\} + E_{BSSE}.$$

A.2.2 Molecular dynamics (MD) simulation of DNA dodecamer system containing the FoU

The DNA dodecamer containing FoU has been crystallized recently[19]. It was a Dickerson-type dodecamer with the sequence d(CGCGAAT(FoU)CGCG) where one of the middle thymines was replaced with 5-formyluracil. The starting structure for our MD simulation was taken from one of these crystal structures (PDB ID: 1G8V). Three sets of simulations were carried out; one with normal Dickerson sequence, another with the FoU crystal structure and the other where A paired with FoU in the crystal structure was replaced by G. The formyl group in FoU could have a syn or an anti conformation to C4 atom. In the crystallographic study, the formyl group of one FoU adopts a syn conformation, but the other is distorted between the syn and anti conformation with almost equal occupancies. For our dodecamers, the formyl group of the FoU at each strand was assigned to be in the different conformation. The AMBER6 program

package was used for the simulations[20]. However, the FoU is a non-natural DNA base and the AMBER6 does not provide the charges and force-field (FF) parameters for it. Therefore we generated the charges and FF parameters with the consistent way used in the development of the PARM94 in AMBER.

Determination of charges and FF parameters for FoU

We took the thymine nucleoside structure (DTN) in AMBER6 as the initial structure and then changed the methyl group at the C5 position to formyl group. For the enol tautomeric form, the carbonyl group at C4 was converted to the hydroxyl group. We built the syn and the anti conformation of the formyl group separately. With those structures, the geometry optimization was performed at the HF/6-31G* level using Jaguar v4.1. The electrostatic potential (ESP) was calculated for the final geometry and was used as an input for the RESP module in AMBER6 to obtain the charges[21]. In AMBER6, sugar atoms have intermolecularly equivalent charges with the exception of C1' and H1' atoms. Those atoms were constrained to have the same charges given in AMBER6 during charge fitting. The sum of charges for hydrogen and oxygen in the hydroxyl group at the 3' and 5' terminal of nucleoside was constrained. The force-field atom types of modified part were assigned using the Antechamber module in AMBER7[22]. The consistent atoms with thymine kept the same force-field atomic types as in thymine. The common force-field parameters with thymine was taken from the Cornell *et al.* force field[23] given in AMBER6 and non-available parameters there were from the "general amber force field" (gaff.dat). Table A.3 summarizes the FF atomic types and the charges used in this simulation.

MD simulation procedures

The DNA dodecamer was embedded in a rectangular box of TIP3P water molecules extended by 10 Å in each direction of a DNA solute where there were approximately 4000 water molecules. The sodium cations were added to neutralize the system at the electronegative points determined by the electrostatic potential that was calculated at the crude grid points. Some water

molecules clashed with cations were replaced with those ions. Most cations were located near the negatively charged phosphate groups. First, the minimization was performed with the DNA under the harmonic constraint (500 kcal/molÅ^2) while only waters and sodium ions movable to relieve the bad contact between DNA solute and waters or cations. In the next, the constant pressure MD was carried out with isotropic position scaling during 25 ps while the system was gradually heated from 0 K to 277 K under 1 bar with the DNA still constrained (500 kcal/ molÅ^2). Near the end of the simulation the density of the system reached ~1 g/cc. One more 25 ps constant pressure MD was done at constant temperature of 277 K. While releasing the constraint of the solute, the whole system was minimized. Then without any constraint, the whole system was gradually heated up from 0 K to 277 K under the constant pressure of 1 bar. After the system was fully equilibrated in this way, the long-term constant volume MD simulation was performed. All the MD simulations were done with 2 fs integration step. The particle mesh ewald (PME) method was used for the long-range electrostatic interaction. The cutoff distance of 9 Å was used for van der Waals interaction of Lennard–Jones type.

The helical parameter analysis was done using the Curves 5.2 program[24]. The O7 of formyluracil was removed during analysis because the presence of O7 alters the definition of base axis system on the pyrimidine and affects the helical parameters especially related to bases.

A.3 Results and discussion

A.3.1 QM calculations of base pairing energies

Figure A.1 shows the hydrogen bonding patterns of FoU and deprotonated FoU with A or G. They are final QM-optimized structures obtained by the method described in the previous section. While the keto tautomer of FoU forms the hydrogen bonds of Watson-Crick type with A, the enol tautomer forms the Wobble type. The enol tautomer can form the Watson-Crick type of

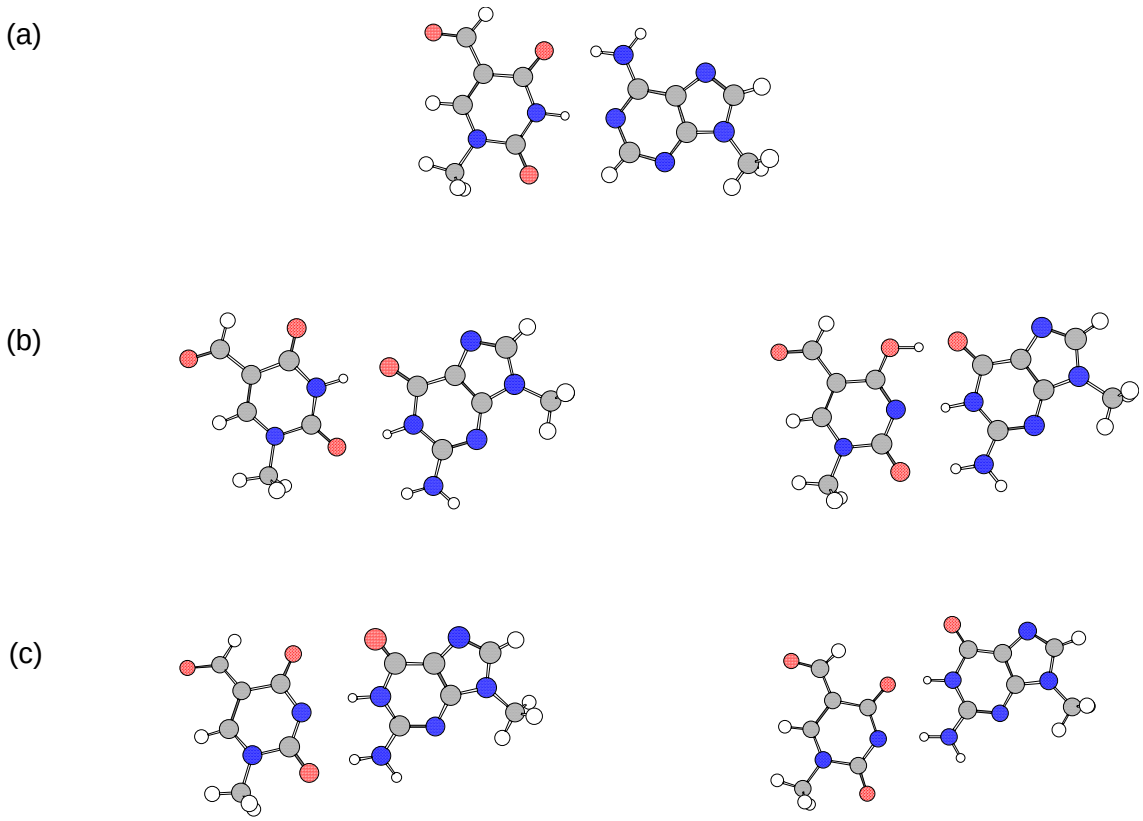


Figure A.1 Hydrogen bonding patterns of FoU with purine bases; QM-optimized structures, (a) A-FoU(keto) Watson-Crick, (b) G-FoU(keto) Wobble (*left*) and G-FoU(enol) Watson-Crick (*right*), (c) G-FoU(deprotonated) Watson-Crick (*left*) and Wobble (*right*).

mispairing with G through three hydrogen bonds. For the deprotonated FoU, both types of hydrogen bonding are possible with G.

The enthalpies of base pairing in gas phase for the canonical AT and GC pairs are shown in Table A.1. It can be seen that the exclusion of diffuse function on the frequency calculation does not make any difference on calculation of enthalpies. In all three cases, the calculated values agree fairly well with experimental ones, even though the slight improvement on the pairing enthalpy of GC is shown when the 6-31G**++ basis set is used for the final geometry optimization. In the case of the anion species like the deprotonated FoU, the diffuse function

Table A.1 Base pairing enthalpies in gas phase at 300 K for GC and AT calculated using B3LYP DFT method with three combinations of basis set[§]

	6-31G**/ 6-31G**	6-31G**/ 6-31G**++	6-31G**++/ 6-31G**++	Exptl[a]
AT (Watson-Crick)	-10.9	-10.5	-10.6	
AT (Hoogsteen)	-11.5	-11.1	-11.1	-13.0
GC (Watson-Crick)	-24.4	-23.3	-23.4	-21.0

[a] Reference [25], [§] unit: kcal/mol

should be included and the 6-31G**/6-31G**++ combination has been chosen for all other calculations in this study.

When the FoU is paired with A, the pairing free energies slightly increase in solution phase as well as in gas phase, compared with those of A-T Watson-Crick pair. Since the formyl group is an electron-withdrawing group, the inductive effect makes the charges on the pyrimidine ring deficient, and the hydrogen bond would become stronger if the FoU plays a role as a hydrogen donor. However, the FoU forms two hydrogen bonds with A both as a donor and as an acceptor. Therefore such an enhancement would be nullified and the pairing energy of A-FoU would become similar to that of AT. From the fact of the slight stabilization in A-FoU, it can be said that the hydrogen bond between H3 in FoU and N1 in A plays a more important role in the A-FoU pairing as previously shown by Kawahara et al.[26].

The FoU and G can form two hydrogen bonds in the Wobble geometry and their pairing free energy is comparable to that of FoU-A pair. The T-G pair also has the similar strength of hydrogen bond to T-A pair. It shows that the FoU and T can be paired with G as frequently as with A when they exist as the free bases.

Tautomerism of FoU

Table A.2 Pairing free energies in gas phase and in aqueous solution calculated using B3LYP DFT method with 6-31G**//6-31G**++ basis sets[a]

	$\Delta E(BSSE)$	ΔH_{300K}	$\Delta G_{300K}(g)$	G_{solv}	$\Delta G_{300K}(aq)$	$\Delta\Delta G_{300K}(aq)^{keto\rightarrow enol}$
TA [WC][e]	-11.9	-10.5	1.0	10.0	11.0	
TA [H][e]	-12.5	-11.1	0.5	10.3	10.9	
CG	-25.1	-23.3	-9.1	19.8	10.7	
FoUA	-12.4	-11.1	0.34	10.0	10.3	
FoUG	-12.9	-11.5	-0.05	10.3	10.2	
FoU´G[b]	-25.9 (-14.8)[c]	-25.2 (-14.4)[c]	-12.4 (-1.4)[c]	18.9	6.5 (15.9)[c]	9.4
TG	-12.8	-11.3	0.5	10.6	11.1	
T´G[b]	-27.1 (-15.3)[c]	-26.1 (-14.7)[c]	-13.6 (-1.5)[c]	19.5	5.9 (15.4)[c]	9.5
FoU⁻G [WC][e]	-23.4	-21.9	-10.1	19.2	9.1 (10.4)[c]	1.3*[d]
FoU⁻G [W][e]	-29.0	-27.5	-16.2	25.3	9.1 (10.4)[c]	1.3*[d]

[a] Unit: kcal/mol, [b] FoU´ and T´: enol tautomers of FoU and T, [c] ( ): considering energy penalty relative to the keto form of neutral FoU, [d] *: from *J. Phys. Chem. A* **105** 274 (5-formyluracil, T = 298 K, pH = 7.00), [e] WC: Watson-Crick; H: Hoogsteen; W: Wobble

The FoU and T can have enol tautomeric forms. In both FoU and T cases, the keto form is energetically more favorable than the enol form as shown in Table A.2 and the calculated equilibrium constants of tautomerism, which are defined as the concentration ratio of enol form to keto form, are 1.2 x 10^{-7} and 1.3 x 10^{-7} at 300 K in aqueous solution for FoU and T, respectively. However, one of enol tautomers could form three hydrogen bonds with G as shown in Figure A.1 and the barrier of tautomerism would be compensated by one extra hydrogen bonding. Actually the calculation results show that the pairing of enol form with G in gas phase is slightly favorable even after considering energy penalties (11.1 kcal/mol in E_{0K} for FoU and 11.8 kcal/mol in E_{0K} for T with respect to the keto form). On the other hand, it becomes unfavorable in aqueous solution because of large cost of solvation energy on pairing. If we assume that the DNA bases would be in the lower dielectric environment in oligonucleotides than in water, the pairing of enol form with G would be energetically plausible in the biological system.

Deprotonated form of FoU

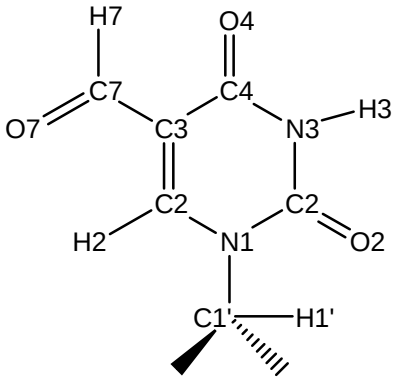
The smaller pKa of 5-formyl deoxyuridine predicts the existence of N3-deprotonated, negative species on a larger amount at the physiological pH[4]. At pH = 7 and T = 300 K, the 7.6% of 5-formyl deoxyuridine would dissociate into negative deoxyuridinium ion and proton while only 0.2% dissociates for deoxythymidine. The deprotonated FoU plays a role only as a hydrogen acceptor and therefore the pairing with A is expected to be extremely weak. Two possible geometries of pairing between FoU^- and G are shown in Figure A.1. The first one (geometry I) corresponds to Watson-Crick geometry, which could be optimal since it does not distort the overall backbone geometry in the normal DNA. The interesting thing is that both geometries have a big stabilization on pairing in gas phase and their pairing free energies are comparable to that of the triple hydrogen-bonding pair such as GC. In the geometry I the repulsion between two electronegative oxygens destabilizes the hydrogen bonding, and actually the purine and pyrimidine rings are no longer co-planar in this structure. The extra stability in the pair of deprotonated FoU with G could come from the ion and ion-induced dipole interaction since the permanent dipole for the isolated guanine does not point toward the negatively charged FoU. In aqueous solution, the solvation energy for the isolated FoU^- is quite huge and the final free energy in solution becomes comparable to those of neutral G-FoU Wobble pair and A-FoU Watson-Crick pair. Considering that the base pair is not fully exposed to water in the oligonucletide, these results support the mechanism that the ionization could allow formation of mispair with G during DNA replication and it would induce the transition mutation at the oxidized T site.

A.3.2 MD simulations of dodecamers

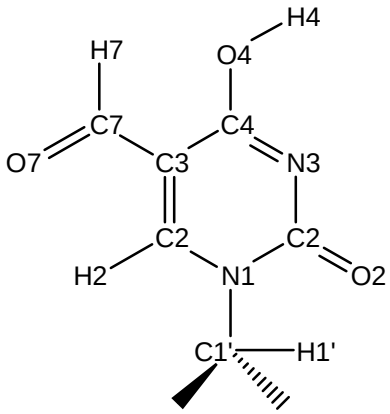
In the present MD simulation, we consider the most dominant keto form of FoU. The deprotonated FoU that might play a role in mispairing during the DNA replication step would turn into the thermodynamically most stable keto species.

AMBER force field parameters of FoU

Table A.3 The AMBER type force field parameters of 5-formyluracil – keto form (*top*) and enol form (*bottom*)



atom label	*FF atomic type*	*charges*	
		(anti)	*(syn)*
C1'	CT	0.166	0.142
H1'	H2	0.137	0.155
N1	N*	-0.010	-0.001
C6	CM	-0.196	-0.280
H6	H4	0.298	0.286
C5	CM	-0.055	-0.041
C7	C	0.396	0.396
O7	O	-0.518	-0.455
H7	HA	0.067	0.001
C4	C	0.412	0.522
O4	O	-0.527	-0.512
N3	NA	-0.295	-0.393
H3	H	0.305	0.319
C2	C	0.496	0.555
O2	O	-0.554	-0.573



atom label	*FF atomic type*	*charge*	
		(anti)	*(syn)*
C1'	CT	0.205	0.182
H1'	H2	0.104	0.121
N1	N*	-0.110	-0.083
C6	CM	-0.072	-0.178
H6	H4	0.243	0.237
C5	CM	-0.106	-0.094
C7	C	0.343	0.348
O7	O	-0.491	-0.457
H7	HA	0.074	0.038
C4	CA	0.620	0.715
O4	OH	-0.614	-0.587
H4	HO	0.469	0.454
N3	NC	-0.738	-0.775
C2	C	0.787	0.800
O2	O	-0.592	-0.600

Table A.4 The base pairing energies and the distances between H-bond donors and acceptors for the base pairs involved in the dodecamers of this work (FUA : anti conformer, FUS : syn conformer)

(Unit: kcal/mol)

	AMBER FF[a]	QM[b]
A:T	- 15.1	-11.9
G:C	-29.0	-25.1
A:FUA	-14.3	-12.4
A:FUS	-14.3	-12.5
G:FUA	-16.3	-12.9
G:FUS	-16.4	-12.7

(Unit: Å)

	AMBER FF[a]	QM[b]	X-ray[c]
A:T Watson-Crick			
N1-N3	2.85	2.91	2.82
N6-O4	2.79	3.01	2.95
G:C Watson-Crick			
N1-N3	2.85	2.98	2.95
N2-O2	2.74	2.94	2.86
O6-N4	2.78	2.83	2.91
A:FUA Watson-Crick			
N1-N3	2.86	2.89	
N6-O4	2.80	2.90	
A:FUS Watson-Crick			
N1-N3	2.86	2.88	
N6-O4	2.81	3.08	
G:FUA Wobble			
N1-O2	2.75	2.95	
O6-N3	2.80	2.77	
G:FUS Wobble			
N1-O2	2.74	2.83	
O6-N3	2.80	2.98	

[a] dielectric constant = 1; scaling of 1-4 vdW interaction = 0.5; scaling of 1-4 electrostatic interaction = 0.83; for deoxynucleosides [b] gas phase calculation; BSSE corrected; for 1-methylpyrimidines and 9-methylpurines [c] From experimental X-ray crystallographic data [26].

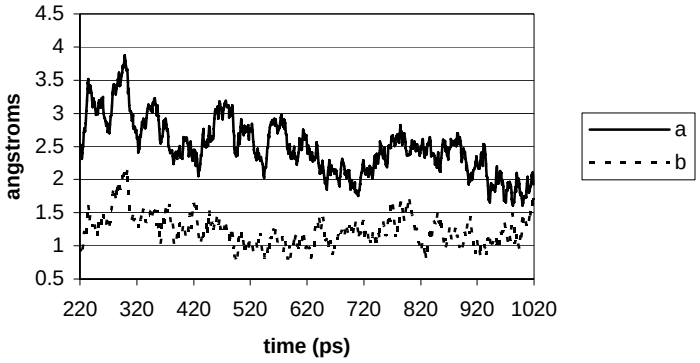


Figure A.2 Fluctuation in the root mean square deviation of coordinates (CRMSD) of DNA dodecamer containing FoUs (1G8V.pdb) during 1 ns simulation after equilibration; a : CRMSD with respect to the minimized DNA structure, b : CRMSD with respect to the mean structure over 220–1020 ps.

The force field (FF) atomic types and charges for FoU developed for this study are tabulated in Table A.3. To validate these parameters the base paring energies and geometries obtained with AMBER FF are compared with those from QM calculations as shown in Table A.4. The overall pairing energies are a little overestimated even in the cases of the canonical AT and GC pair, although this might result from the extra non-bond interaction between sugar rings in nucleosides. However, the extent of the overestimation for the pairs with formyluracil is comparable to the GC and AT cases. The hydrogen bond distances agree well with each other and the differences are within 0.3 Å. To check the stability of the DNA conformation during the MD simulation with newly implemented charges and FF parameters for FoU, the time evolution of the root mean square deviation of coordinates (CRMSD) was calculated for the dodecamer X-ray crystallographic structure containing FoU (1G8V) in Figure A.2. The CRMSD value with respect to the mean structure is 1.24 ± 0.24 Å and it is comparable to the one calculated for the DNA system with normal base sequence.

Hydrogen bond distance

The stability of DNA structure is directly related to the hydrogen bonds (H-bonds) between two base pairs. The bond distances between H-bond donor and acceptor atoms were measured

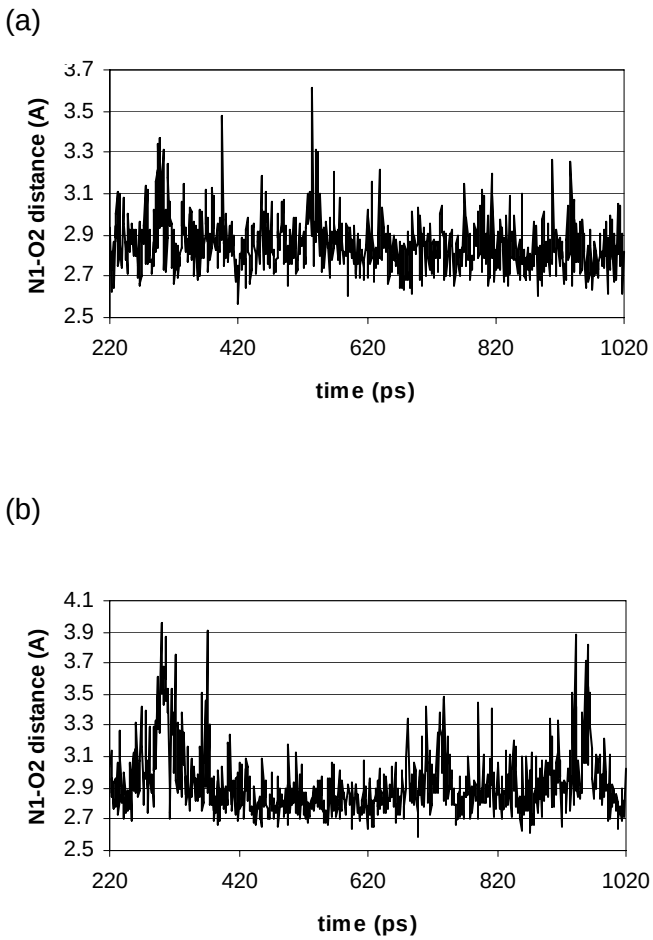


Figure A.3 The time profile of H-bond distance between N1 from G and O2 from FoU; (a) for the syn conformer of FoU at the 5th position and (b) for the anti conformer of FoU at the 8th position.

every 1 ps after 220 ps during the production period. For the DNA with the normal Dickerson sequence, the distances are within 3.1 Å in most times except for the bases at the terminal. The base pairs at the 5' terminal started unraveling around 420 ps and formed the H-bonds back in 50 ps later. The H-bonds at the 3' terminal broke around 620 ps and stayed unraveled until the end of the simulation. The floppiness of the bases at the terminal is usual since they have only the one-side stacking interaction. In the case of the dodecamer with A:FoU pairs, a similar phenomena were observed. The FoU in the middle of the dodecamer does not cause any instability in H-bonds and the DNA kept the stable conformation during the simulation. However, when the FoU

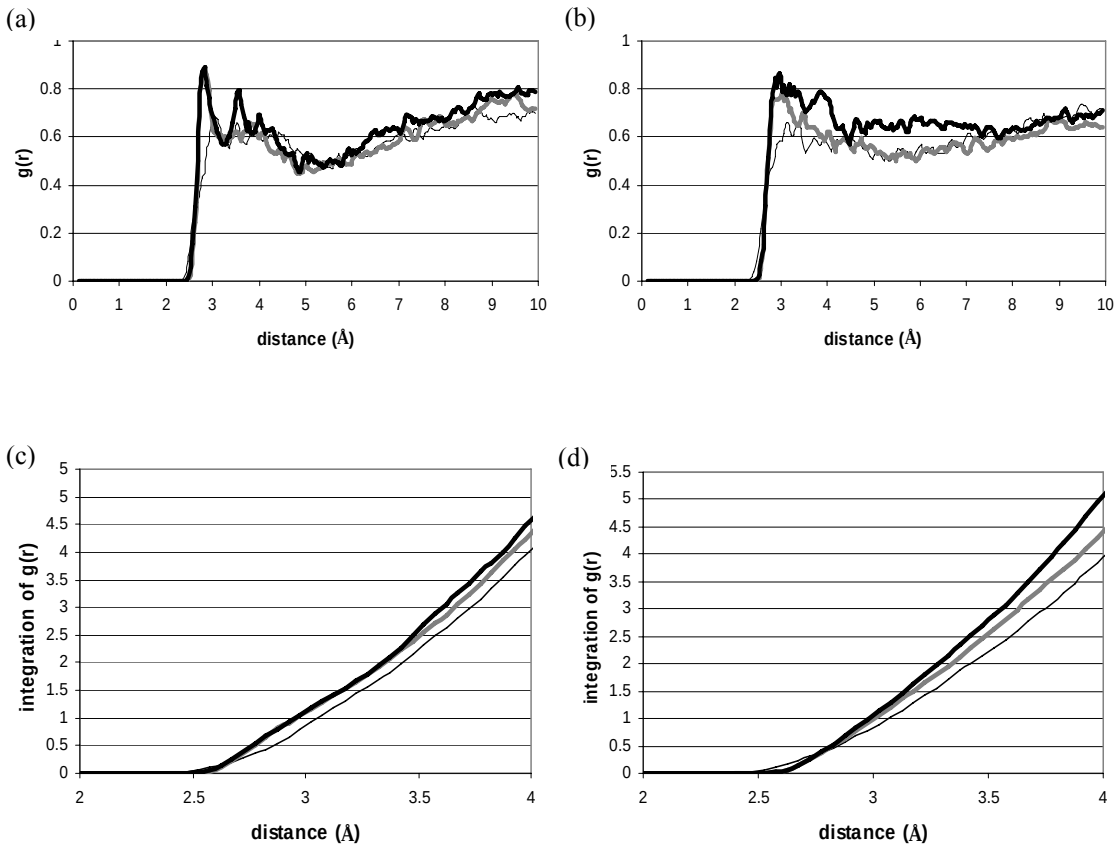


Figure A.4 [(a), (b)] Normalized radial distribution functions g(r) of water-oxygen [(c), (d)] and the number of waters in the solvation shell obtained by integrating g(r). The thick black line, the thick gray line and the thin black line show g(r) of the target atoms, O7 of FoU in the G:FoU case, O7 of FoU in the A:FoU case and H7 of T in the A:T case respectively. (a) and (c) : the formyl group of FoU is anti at the 8th base pair position; (b) and (d) : the formyl group is syn at the 5th base pair position. g(r) was normalized by the water of 1 g/cm^3. $n = 4\pi\rho\int g(r)r^2 dr$, where ρ is 0.033 molecules/Å^3 for water of 1 g/cm^3.

is paired with G, the large fluctuation in H-bonds for the G:FoU pair was observed, especially in the case of FoU with the formyl group in the anti conformation. For the syn conformer, the H-bonds were pretty steady.

DNA hydration

When the methyl group in thymine is substituted with the formyl group, this extra oxygen (O7) can play a role as a hydrogen bond acceptor. Figure A.4 shows the radial distribution function, g(r) of oxygens in water solvent.

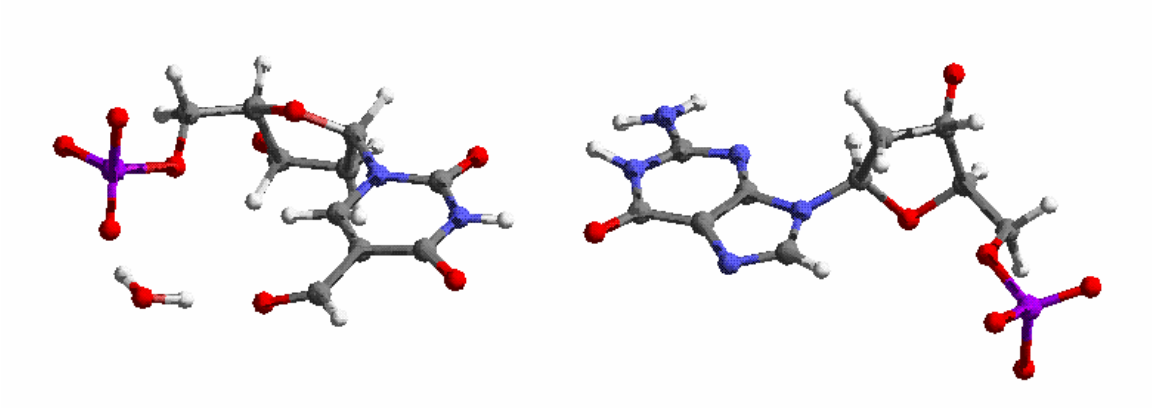


Figure A.5 The snapshot of guanine and formyl uracil with the formyl group in anti conformation (at 301 ps). The water makes a hydrogen bond bridge between O7 of FoU and O2P of the phosphate group. The O-O distance between water and O7 is 2.7 Å and the other O-O distance between water and O2P is 2.5 Å.

The first sharp peak that is not prominent in the thymine case is observed for the FoU case, especially when the formyl group is in the anti conformation. It shows that the waters around the oxygen of formyl group in anti are well ordered. This is because the water can make a bridge between O7 of FoU and O2P or O5' from the backbone when the formyl group is anti as shown in Figure A.5. In the syn conformation, the O7 is located away from these oxygen atoms and the O4 of FoU is too close to O7 atom for a water to make H-bond bridge. When g(r) is integrated over the first coordination shell ($r \sim 3.3$ Å) for FoU at the 8th position of G:FoU and A:FoU case, the number of waters is approximately 1.8 for both cases. We can clearly see that the more water molecules are around the FoU than the thymine.

Groove widths

The widths of the major and minor grooves for Dickerson crystal structure (PDB ID: 1BNA) and FoU crystal structure (PDB ID: 1G8V) were calculated using Curves program. They have the same crystal symmetry. If we ignore that the different experimental condition where the crystals were grown might affect the conformational differences, Figure A.6 shows definitely sequence-dependence of the groove widths. Although the overall shapes in the graphs are similar

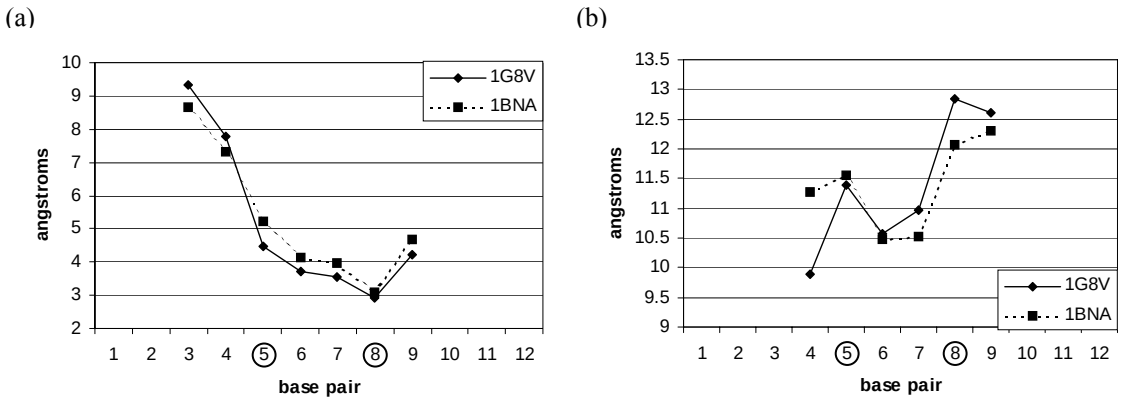


Figure A.6 The widths of minor (a) and major (b) grooves. The solid line is for the crystal structure with 5-formyluracil (1G8V) and the dotted line is for Dickerson crystal structure (1BNA). The positions where the sequence differences are shown are circled.

to each other, the widths of minor grooves for the FoU case become slightly narrower and the major grooves get wider.

The groove widths averaged over the MD simulation from 220 ps to 1020 ps are shown in Figure A.7. Since they are isolated DNA molecules immersed into the water box and less constrained, the absolute values of widths are larger than in the crystal structure. The dodecamer with the normal Dickerson sequence has an asymmetric distribution in minor groove throughout the sequence even though the sequence itself is symmetric. However, the dodecamer with FoU:A has a symmetric pattern, and the different conformation of formyl group does not seem to affect the width of minor groove. The replacement of T with FoU results in the significant decrease in the width of the minor groove. In the FoU:G case, the minor groove becomes narrower around the 5th base pair position where the syn FoU is located and on the other hand the minor groove becomes wider at the 8th base pair position where the anti FoU is located. The major grooves become narrower for the A:FoU DNA and this change is more prominent on the side of anti FoU. There is a huge increase in the width of the major groove near the 5th base pair for the G:FoU DNA.

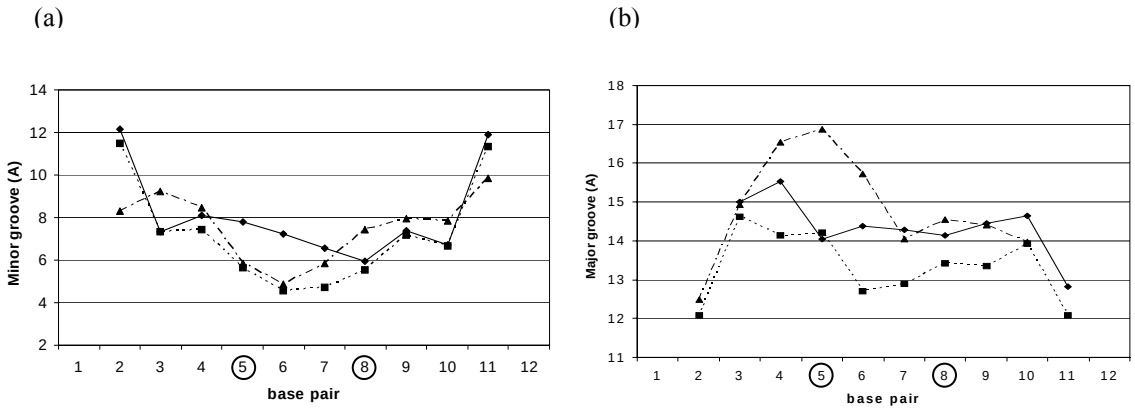


Figure A.7 The minor (a) and major (b) groove widths averaged over the MD simulation from 220 ps and 1020 ps. The diamond, square and triangle symbol are for the normal, A:FoU and G:FoU DNA dodecamer, respectively. The syn FoU is at the 5th position and the anti FoU is at the 8th position as indicated by circles.

Helical parameters

The global base-base parameters are analyzed and shown in Figure A.8. These parameters could be the indication of the stability in hydrogen bonding. The G:FoU DNA dodecamer has the large shearing at the 5th and 8th position where G:FoU mispairs are located. The G and FoU have the Wobble geometry and they are sliding each other compared to the pyrimidine-purine pair in Watson-Crick geometry. Therefore the large values in shear parameter reflect the Wobble geometry of G:FoU pair. Since the measurement is done in the 5'→3' direction, they have the opposite sign even though the sequence is symmetric. The conformation of formyl group does not make any difference in shearing.

Large buckling is detected at the 4th and 9th base pair in the A:FoU DNA dodecamer, compared with the case of normal Dickerson sequence. These pairs that flank the central AATT sequence have respectively positive and negative buckles that bend the center of these base pair away from the central tetramer. This may contribute to severe narrowing of the minor grooves in the A:FoU dodecamer. The similar huge positive buckling is shown only at the 4th base pair

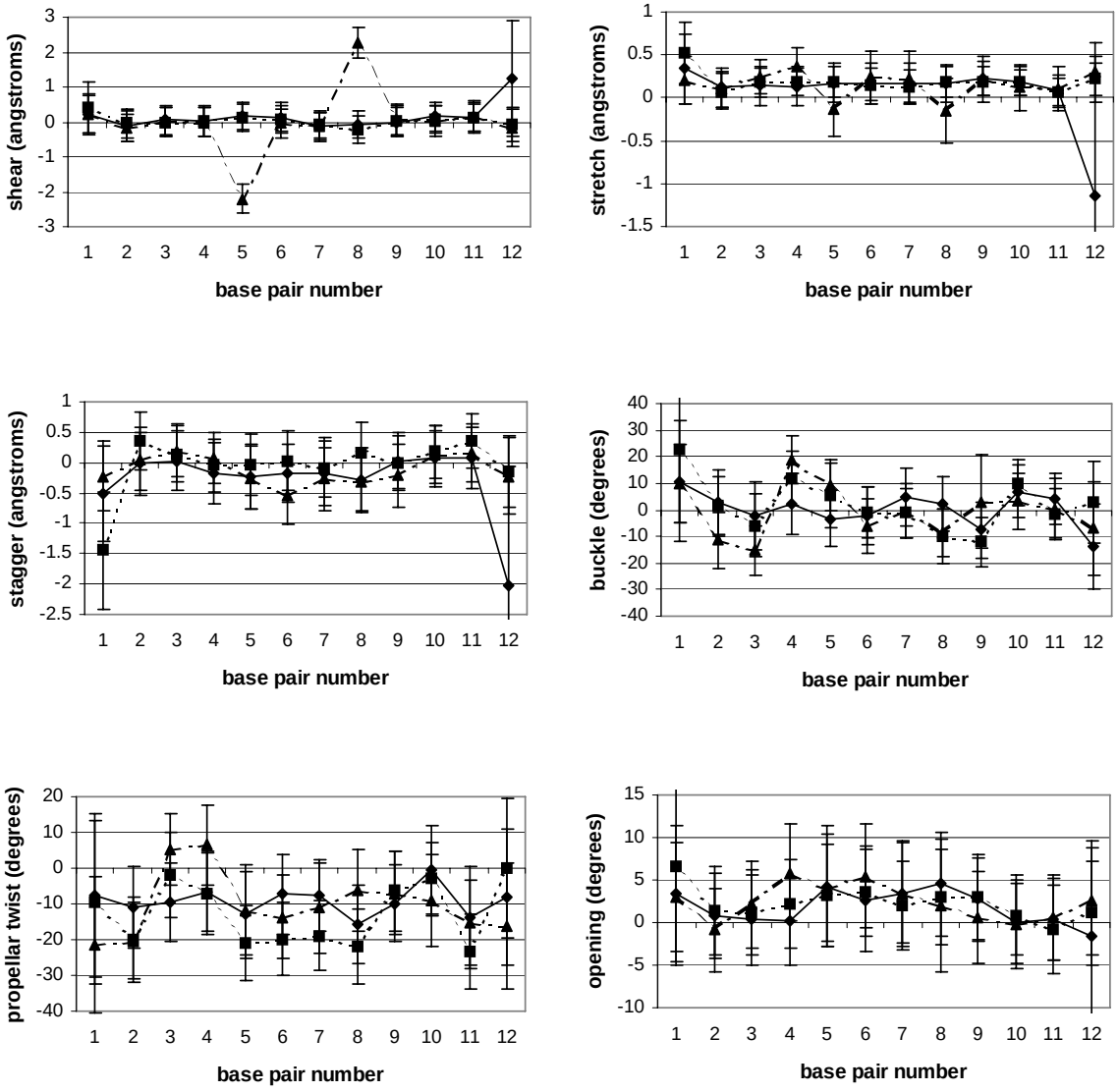


Figure A.8 The global base-base parameters. They are averaged values over the MD simulation from 220 ps and 1020 ps. The diamond, square and triangle symbol are for the normal, A:FoU and G:FoU DNA dodecamer, respectively.

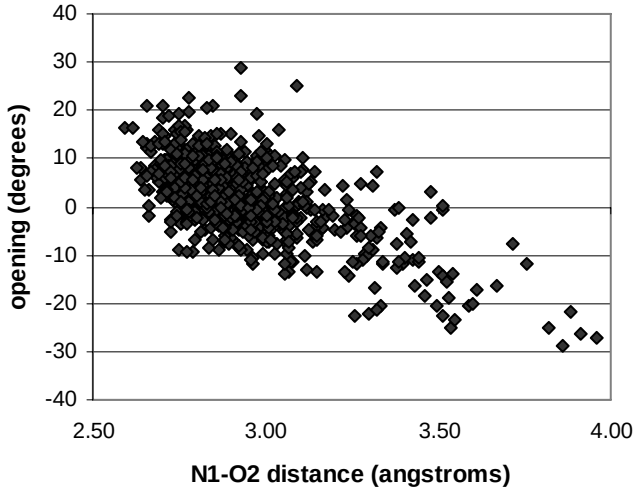


Figure A.9 The plot of N1-O2 distance versus the opening at the 8th G:FoU pair. The snapshot was taken every 1 ps during 220-1020 ps of MD simulation.

position in the G:FoU mispair case and it is reflected as the asymmetric narrowing of minor groove in the central part of the G:FoU dodecamer.

The large positive opening in base pair means opening in the major groove and thus narrowing of the minor groove. The deviation from the normal Dickerson sequence case that might explain the significant narrowing in the dodecamers with FoU present is not small, and in all three systems the width of minor groove shows the negative correlation with opening. Figure A.9 shows the correlation between the H-bond distance of N1-O2 at the 8th base pair in G:FoU dodecamer and the opening. In most times, the N1-O2 distance is near 2.9 Å and the opening fluctuates by 10º around zero. When the N1 and O2 get apart, the opening becomes more negative. This loose H-bond at the 8th base pair contributes to the larger width of minor groove than one at the 5th base pair as shown in Figure A.7.

Backbone parameters

The torsion angles for a polydeoxyribonucleotide chain and the pseudorotation phase angle P of a sugar ring are calculated for three dodecamer systems. The distinct sequence-dependent

aspects are not seen here. The preferred sugar puckering modes are O4'-endo, C1'-exo and C2'-endo that correspond to the "south" conformations. The structures keep B-type DNA conformations over the MD simulation. The phase angle P and the δ torsion (C5'-C4'-C3'-O3') at the 7th base residue and the 16th base residue show a little correlation with the opening of G and FoU base at the 8th base pairs. The 7th and 16th bases are right ahead of this G:FoU pair in the 5'→3' direction. The correlation coefficients for the P and δ torsion of the 7th base are 0.36 and 0.36, respectively. For the 16th base, they are 0.32 and 0.34 for the P and δ torsion.

A.4 Summary and Conclusion

We calculated pairing free energies of various free DNA base pairs at the B3LYP/6-31G**//B3LYP/6-31G**++ level, focusing on mispairing of 5-formyluracil which is an oxidative form of thymine. The free energy of keto FoU with G is comparable to that with A in both gas phase and solution phase while the pairing of enol FoU with G in solution phase is most unfavorable due to large cost of solvation free energy on pairing in addition to the barrier on tautomerism. The N3-deprotonated FoU forms strong hydrogen bonding with G in gas phase, which is energetically comparable to the triple hydrogen bonding of GC pair. The calculation in aqueous phase shows that mispairings of both neutral keto and deprotonated FoU with G are as probable as normal base pairings.

Considering that the neutral FoU could be mispaired as frequently as T with G from the aspect of energetics, we conclude that the ionization at N3 position of FoU would mainly account for the increased misparing rate of FoU since the deprotonated FoU preferentially form H-bonds with G rather than A and therefore FoU has one more extra possibility of base pairing.

The 1 ns MD simulations were then carried out for three DNA dodecamer-explicit water systems; one with normal Dickerson-type sequence, another with two thymines replaced by formyluracil in the Dickerson sequence and the other where these formyluracils are paired with

guanines. Even though the formyl group is on the side of the major groove, its presence actually leads to the severe narrowing of the minor grooves. In the case of G:FoU dodecamer, the same kind of narrowing was shown especially around the 5th base pair where the syn FoU makes a pair with G. The slightly wider minor groove at the 8th base pair position where the formyl group of FoU has anti conformation, is correlated with loosening of H-bonds between G:FoU.

The formyl group of FoU in the anti conformation affects the hydration pattern around the DNA structure. A water molecule makes a bridge of H-bond between O7 of FoU and O2P of phosphate and it provides the well-ordered water structure. No significant change in backbone parameters is shown for A:FoU and G:FoU dodecamer.

Overall the incorporation of FoU paired with A does not cause the significant structural change in DNA double helix except for the narrowing of the minor groove. On the other hand, the G:FoU dodecamer shows relatively larger fluctuation since it contains non Watson-Crick pairs. It might be worth studying how this kind of conformational distortion affects the interaction with a DNA-binding protein, for example like the DNA repair enzyme. More detailed molecular level description also should be investigated to explain the effect of the formyl group on the width of minor grooves.

References

1. Ames, B.N., M.K. Shigenaga, and T.M. Hagen, *Oxidants, Antioxidants, and the Degenerative Diseases of Aging.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(17): p. 7915-7922.
2. Kasai, H., et al., *5-Formyldeoxyuridine - a New Type of DNA Damage Induced by Ionizing-Radiation and Its Mutagenicity to Salmonella Strain Ta102.* Mutation Research, 1990. **243**(4): p. 249-253.
3. Bjelland, S., et al., *Cellular effects of 5-formyluracil in DNA.* Mutation Research-DNA Repair, 2001. **486**(2): p. 147-154.
4. Privat, E.J. and L.C. Sowers, *A proposed mechanism for the mutagenicity of 5-formyluracil.* Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis, 1996. **354**(2): p. 151-156.
5. Masaoka, A., et al., *Oxidation of thymine to 5-formyluracil in DNA promotes misincorporation of dGMP and subsequent elongation of a mismatched primer terminus by DNA polymerase.* Journal of Biological Chemistry, 2001. **276**(19): p. 16501-16510.
6. Mol, C.D., et al., *DNA repair mechanisms for the recognition and removal of damaged DNA bases.* Annual Review of Biophysics and Biomolecular Structure, 1999. **28**: p. 101-128.
7. Masaoka, A., et al., *Enzymatic repair of 5-formyluracil I. Excision of 5- formyluracil site-specifically incorporated into oligonucleotide substrates by AlkA protein (Escherichia coli 3- methyladenine DNA glycosylase II).* Journal of Biological Chemistry, 1999. **274**(35): p. 25136-25143.
8. Terato, H., et al., *Enzymatic repair of 5-formyluracil II. Mismatch formation between 5-formyluracil and guanine during DNA replication and its recognition by two proteins*

involved in base excision repair (AlkA) and mismatch repair (MutS). Journal of Biological Chemistry, 1999. **274**(35): p. 25144-25150.

9. Zhang, Q.M., et al., *Identification of repair enzymes for 5-formyluracil in DNA - Nth, Nei, and MutM proteins of Escherichia coli.* Journal of Biological Chemistry, 2000. **275**(45): p. 35471-35477.
10. Hamelberg, D., et al., *Flexible structure of DNA: Ion dependence of minor-groove structure and dynamics.* Journal of the American Chemical Society, 2000. **122**(43): p. 10513-10520.
11. Young, M.A., B. Jayaram, and D.L. Beveridge, *Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: Fractional occupancy of electronegative pockets.* Journal of the American Chemical Society, 1997. **119**(1): p. 59-69.
12. Kopka, M.L., et al., in *Structure and motion : membranes,nucleic acids & proteins*. 1985, Adenine Press. p. 461-483.
13. Shui, X.Q., et al., *The B-DNA dodecamer at high resolution reveals a spine of water on sodium.* Biochemistry, 1998. **37**(23): p. 8341-8355.
14. Schwabe, J.W.R., *The role of water in protein DNA interactions.* Current Opinion in Structural Biology, 1997. **7**(1): p. 126-134.
15. Jaguar. 2000, Schrodinger Inc.: Portland.
16. *CRC Handbook of Chemistry and Physics*. 60th ed, ed. R.C. Weast. 1979, Boca Raton, FL: CRC Press.
17. Tannor, D.J., et al., *Accurate First Principles Calculation of Molecular Charge-Distributions and Solvation Energies from Ab-Initio Quantum- Mechanics and Continuum Dielectric Theory.* Journal of the American Chemical Society, 1994. **116**(26): p. 11875-11882.

18. Xantheas, S.S., *On the importance of the fragment relaxation energy terms in the estimation of the basis set superposition error correction to the intermolecular interaction energy.* Journal of Chemical Physics, 1996. **104**(21): p. 8821-8824.

19. Tsunoda, M., et al., *Crystallization and preliminary X-ray analysis of a DNA dodecamer containing 2 '-deoxy-5-formyluridine; what is the role of magnesium cation in crystallization of Dickerson-type DNA dodecamers?* Acta Crystallographica Section D-Biological Crystallography, 2001. **57**: p. 345-348.

20. Case, D.A., et al., *AMBER6*. 1999, University of California, San Francisco.

21. Cieplak, P., et al., *Application of the Multimolecule and Multiconformational Resp Methodology to Biopolymers - Charge Derivation for DNA, Rna, and Proteins.* Journal of Computational Chemistry, 1995. **16**(11): p. 1357-1377.

22. Case, D.A., et al., *AMBER7*. 2002, University of California, San Francisco.

23. Cornell, W.D., et al., *A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules.* Journal of the American Chemical Society, 1995. **117**(19): p. 5179-5197.

24. Lavery, R. and H. Sklenar, *Curves 5.2*. 1997.

25. Yanson, I.K., A.B. Teplitsky, and L.F. Sukhodub, *Experimental Studies of Molecular-Interactions between Nitrogen Bases of Nucleic-Acids.* Biopolymers, 1979. **18**(5): p. 1149-1170.

26. Kawahara, S., et al., *Ab initio and density functional studies of substituent effects of an A-U base pair on the stability of hydrogen bonding.* Journal of Physical Chemistry A, 1999. **103**(42): p. 8516-8523.