First Principles Calculation of pK_a Values for 5-Substituted Uracils

Yun Hee Jang,[†] Lawrence C. Sowers,[‡] Tahir Çağin,[†] and William A. Goddard III*,[†]

Materials and Process Simulation Center, Beckman Institute (139-74), California Institute of Technology, Pasadena, California 91125, and Divisions of Pediatrics and Molecular Medicine, City of Hope National Medical Center, 1500 E. Duarte Road, Duarte, California 91010

Received: December 17, 1999; In Final Form: September 7, 2000

Oxidation of uracil (U) and thymine (5-Me-U) are believed to play a role in genetic instability because of the changes these oxidations cause in the ionization constants (pK_a values), which in turn affects the base pairing and hence coding. However, interpretation of the experimental evidence for the changes of pK_a with substitution at U has been complicated by the presence of two sites (N1 and N3) for ionization. We show that a procedure using first principles quantum mechanics (density functional theory with generalized gradient approximation, B3LYP, in combination with the Poisson−Boltzmann continuum-solvation model) predicts such pK_a values for a series of 5-substituted uracil derivatives in excellent correlation with experiment. In particular, this successfully resolves which cases prefer ionization at N1 and N3. Such first principles predictions of ionization constant should be useful for predicting and interpreting pK_a for other systems.

1. Introduction

The 5-substituted pyrimidines (see Figure 1) comprise a biologically important class of base analogues. In particular, uracil and thymine (with a hydrogen atom or a methyl group in the 5-position, respectively) are major constituents of RNA and DNA, respectively. The thymine 5-methyl group in DNA is a frequent site of oxidation,[1−4] generating the 5-hydroxymethyl, formyl, and carboxyl derivatives. The oxidation of DNA is an established source of genomic instability,[5−7] possibly because of oxidation of the thymine. Indeed, the 5-halo uracil derivatives have demonstrated antitumor and antiviral properties.[8−11]

Substitution at the 5-position of uracil can substantially alter the electronic properties of the pyrimidine as indicated experimentally by changes in the UV spectra and ionization constants (pK_a's).[12−16] In aqueous solution this can induce significant changes in the physical and biological properties of the pyrimidine. In particular, ionization of the pyrimidine moiety in DNA could change the coding properties of the base during polymerase-mediated replication, resulting in a base substitution mutation.[3]

Previously,[3] it was demonstrated that electron-withdrawing substituents in the 5-position of uracil reduces measured pK_a values while electron-donating substituents have the opposite effect. Within the uracil series, a good correlation was obtained between the inductive properties of the 5-substituent (Hammett constant) and the pK_a value measured in aqueous solution.[3] However, interpretation of this correlation is ambiguous because ionization could occur at either the N1 or N3 sites, but only the lower value can be observed experimentally. The 5-substituent can influence the N1 and N3 positions differently, obviating simple correlations between inductive properties and pK_a values.

In order to interpret the experimentally measured changes in pK_a for a series of 5-substituent uracil derivatives and to develop a methodology for predicting the changes in pK_a for other possible oxidation products, we initiated a project to calculate the pK_a values for pyrimidine derivatives from first principles quantum mechanics using the methodology discussed in section 2. As established in section 3, there is a strong correlation between the calculated first pK_a value and the measured one, validating the computational method. We are now using this method to predict other systems for which experimental data are not available.

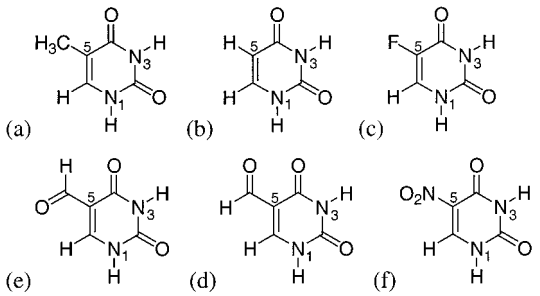


Figure 1. 5-Substituted uracils. (a) thymine, (b) uracil, (c) 5-fluorouracil, (d) *trans*-5-formyluracil, (e) *cis*-5-formyluracil, and (f) 5-nitrouracil.

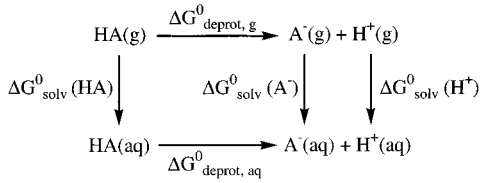


Figure 2. Thermodynamic cycle used in the calculation of pK_a.[17,18]

2. Calculation Details

2.1. pK_a Calculations. The pK_a is estimated theoretically as follows.[17,18] In Figure 2, $\Delta G^0_{\text{deprot,g}}$ and $\Delta G^0_{\text{deprot,aq}}$ are the gas-phase and solution-phase standard free energy of deprotonation, respectively. $\Delta G^0_{\text{solv}}(\text{HA})$, $\Delta G^0_{\text{solv}}(\text{A}^-)$ and $\Delta G^0_{\text{solv}}(\text{H}^+)$

* To whom correspondence should be addressed. E-mail: wag@wag.caltech.edu.
[†] California Institute of Technology.
[‡] City of Hope National Medical Center.

10.1021/jp994432b CCC: $20.00
Published on Web 12/08/2000

are the standard free energies of solvation of HA, A^-, and H^+, respectively. The pK_a of HA(aq) is given by

$$pK_a = \frac{1}{2.303RT}\Delta G^0_{\text{deprot,aq}} \quad (1)$$

where

$$\Delta G^0_{\text{deprot,aq}} = \Delta G^0(A^-(aq)) + \Delta G^0(H^+(aq)) - \Delta G^0(HA(aq))$$

$$= \{\Delta G^0(A^-(g)) + \Delta G^0_{\text{solv}}(A^-)\} + \{\Delta G^0(H^+(q)) + \Delta G^0_{\text{solv}}(H^+)\} - \{\Delta G^0(HA(g)) + \Delta G^0_{\text{solv}}(HA)\}$$

$$= \Delta G^0_{\text{deprot,g}} + \{\Delta G^0_{\text{solv}}(A^-) + \Delta G^0_{\text{solv}}(H^+) - \Delta G^0_{\text{solv}}(HA)\} \quad (2)$$

2.2. Free Energies. The standard Gibbs free energy of each state in gas phase (ΔG^0[HA](g)], ΔG^0[A^-(g)], and ΔG^0[H^+](g)]) is obtained by

$$\Delta G^0 = E_{0K} + \text{ZPE} + \Delta\Delta G_{0\rightarrow 298K} \quad (3)$$

The total energy of the molecule at 0 K (E_{0K}) is calculated at the optimum geometry from quantum mechanics (QM). The zero-point energy (ZPE) and the Gibbs free energy change from 0 to 298 K ($\Delta\Delta G_{0\rightarrow 298K}$) are calculated from the vibrational frequencies calculated using QM. The translational and rotational free energy contribution is also calculated in the ideal gas approximation. We used $\Delta G^0[H^+(g)] = 2.5RT - T\Delta S^\circ = 1.48 - 7.76 = -6.28$ kcal/mol from the literature.[17,18]

2.3. QM Calculations. All QM calculations used the Jaguar v3.5 quantum chemistry software.[19,20] To calculate the geometries and energies of the various molecules, we used the B3LYP flavor of density functional theory (DFT) which includes the generalized gradient approximation and a component of the exact Hartree−Fock (HF) exchange.[21−25] These calculations used the cc-pVTZ(-f)++ basis set and started from the geometries optimized using the HF method with the 6-31G* basis set. The "cc-pVTZ(-f)++" basis set used in this paper (and included in Jaguar) [also denoted as "aug-cc-pVTZ" (augmented correlation-consistent basis set with polarized valence triple-ζ)] is the cc-pVTZ++ basis set of Dunning et al.,[26,27] but with the outmost polarization and diffuse functions (d-functions of H and f-functions of other first-shell elements) deleted. Thus this basis set uses the Huzinaga (6s3p)/[4s3p] for H (instead of (6s3p2d)/[4s3p2d]) and the Huzinaga (11s6p3d)/[5s4p3d] for other first-row elements (instead of (11s6p3d2f)/[5s4p3d2f]).

To determine whether it is sufficient to use HF frequencies for the ZPE and free energy, we calculated frequencies at both HF/6-31G* and B3LYP/cc-pVTZ(-f) for two cases (Uracil and Thymine, neutral and anionic) and multiplied them by the appropriate scaling factors.[28] As shown in Table 1, the HF/6-31G* values are in excellent agreement with the B3LYP/cc-pVTZ(-f) values. Thus, for the other cases, we calculated frequencies at HF/6-31G*, starting from the geometries optimized at this level. The calculated ZPE's were scaled down by 0.9135, and the enthalpy and entropy contributions to the thermodynamic corrections ($\Delta\Delta G_{0\rightarrow 298K}$) were scaled down by 0.8905 and 0.8978, respectively.[28]

2.4. Solvation Energies. The standard free energy of solvation in water [ΔG^0_{solv}(HA) and $\Delta G^0_{\text{solv}}(A^-)$] was calculated using the continuum-solvation approach[29−31] by solving the Poisson−Boltzmann (PB) equation numerically.[32] In this approach, the solute is described as a low-dielectric cavity ($\epsilon_{QM} = 1$) immersed in a high-dielectric continuum of solvent ($\epsilon_{H_2O} = 80$ for water[33]). The solute/solvent boundary is described by the surface of closest approach as a sphere of radius 1.4 Å (probe radius for water) is rolled over the van der Waals (vdW) envelope of the solute. The charge distribution of the solute is represented by a set of atom-centered point charges, which are determined by fitting to the electrostatic potential calculated from the wave function.

The procedure is as follows. A gas-phase calculation is carried out first to obtain the electrostatic-potential fitted (ESP) charges by the CHELPG method.[34−36] On the basis of these charges, the PB equation is solved to obtain the reaction field of the solvent (as a set of polarization charges located on the solute/solvent boundary surface). The Fock Hamiltonian for the HF calculation is then modified to include the solute−solvent interaction due to the reaction field. This is solved to obtain a new wave function and a new set of atom-centered ESP charges. This process is repeated self-consistently until convergence is reached (to 0.1 kcal/mol in the solvation energy). This constitutes the electrostatic or "polar" contribution to the solvation energy.

An additional "nonpolar" contribution due to creation of a solute cavity in the solvent is accounted for by a term proportional to the solvent-accessible surface area of the solute.[31]

The solvation free energy calculation was done at the B3LYP/cc-pVTZ(-f) level, and the geometry was relaxed (i.e., reoptimized) in solution.

Based on earlier studies,[31] the following atomic radii were used to build the vdW envelope of the solute: 2.0 Å for sp^2-hybridized carbon, 1.9 Å for sp^3-hybridized carbon, 1.55 Å for sp^2-hybridized oxygen, 1.5 Å for sp^2-hybridized nitrogen, 2.0 Å for nitrogen and oxygen in $-NO_2$ group, 1.25 Å for hydrogen attached to sp^2-hybridized carbon, and 1.15 Å for hydrogen attached to sp^3-hybridized carbon. For oxygen and nitrogen, we used the same radii corresponding to sp^2-hybridization before and after deprotonation, because even after deprotonation there is strong resonance and electron-delocalization and thus the nitrogen and oxygen still have sp^2-hybridized characteristics.

2.5. $\Delta G^0_{\text{solv}}(H^+)$. Calculating the p$K_a$ also requires the experimental standard free energy of solvation of a proton in water [$\Delta G^0_{\text{solv}}(H^+)$]. Unfortunately, this value remains uncertain.[17,37,38] "The precision of $\Delta G^0_{\text{solv}}(H^+)$ is limited by the fact that the standard hydrogen potential cannot be obtained by measurement alone; some independent quantity is needed to determine an absolute half-cell potential."[17,37] The $\Delta G^0_{\text{solv}}(H^+)$ from the measurements of the standard hydrogen potential range from −254 to −261 kcal/mol.[17,37] From a set of cluster-ion solvation data, ΔG^0_{solv} (H^+) has been estimated to be −263.98 ± 0.07 kcal/mol.[39]

Because of these uncertainties, we chose ΔG^0_{solv} (H^+) to minimize the rms deviation between the calculated and experimental pK_a values for the 5-substituted uracils. This leads to a final value of 258.32 kcal/mol (vide infra, section 2.6), which falls in the middle of the range of experimental $\Delta G^0_{\text{solv}}(H^+)$. The combination with −6.28 kcal/mol of $\Delta G^0[H^+(g)]$ leads to −264.60 kcal/mol of $\Delta G^0[H^+(aq)]$. [For an alternative approach to determine $\Delta G^0[H^+(aq)]$, see Appendix.]

TABLE 1: Calculated Energies Used To Estimate the pK_a's of 5-R-Uracil's[a]

R	energy terms[b]	neutral (kcal/mol)	anion N3(−) (kcal/mol)	pK_a	anion N1(−) (kcal/mol)	pK_a
CH_3	E_{0K}(g)[c]	−285077.96	−284723.65		−284735.20	
	ZPE(g)	71.33	62.53		62.91	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−20.37	−20.42		−20.24	
	total ΔG^0(g)	−285027.00	−284681.54		−284692.53	
	ΔG^0_{solv}	−19.32	−86.51		−73.89	
	total ΔG^0(aq)	−285046.33	−284768.04	10.04	−284766.42	11.23
H	ZPE(g)	−260396.85	−260043.00		−260055.80	
	E_{0K}(g)[c]	54.23	45.45		45.90	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−19.09	−19.17		−18.90	
	total ΔG^0(g)	−260361.71	−260016.72		−260028.81	
	ΔG^0_{solv}	−20.29	−87.94	9.34	−74.31	
	total ΔG^0(aq)	−260382.00	−260104.67		−260103.12	10.47
F	E_{0K}(g)[c]	−322684.07	−322337.03		−322348.30	
	ZPE(g)	49.26	40.52		40.97	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−19.95	−20.15		−19.75	
	total ΔG^0(g)	−322654.76	−322316.65		−322327.08	
	ΔG^0_{solv}	−20.51	−84.12		−71.25	
	total ΔG^0(aq)	−322675.27	−322400.78	7.26	−322398.34	9.05
CHO	E_{0K}(g)[c]	−331533.21	−331191.30		−331206.41	
trans[d]	ZPE(g)	60.42	51.88		52.30	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−20.95	−20.88		−20.76	
	total ΔG^0(g)	−331493.74	−331160.30		−331174.86	
	ΔG^0_{solv}	−22.74	−80.72	7.96	−67.58	
	total ΔG^0(aq)	−331516.48	−331241.03		−331242.44	6.93
CHO	E_{0K}(g)[c]	−331528.27	−331184.85		−331202.33	
cis[e]	ZPE(g)	60.24	51.53		52.13	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−21.22	−21.22		−20.90	
	total ΔG^0(g)	−331489.26	−331154.54		−331171.10	
	ΔG^0_{solv}	−27.38	−87.57		−71.46	
	total ΔG^0(aq)	−331516.63	−331242.11	7.28	−331242.56	6.95
NO_2	E_{0K}(g)[c]	−388759.07	−388422.55		−388439.33	
	ZPE(g)	56.35	47.74		48.18	
	$\Delta\Delta G_{0\rightarrow 298K}$(g)	−22.06	−21.80		−21.68	
	total ΔG^0(g)	−388724.79	−388396.62		−388412.83	
	ΔG^0_{solv}	−23.24	−77.40		−62.88	
	total ΔG^0(aq)	−388748.03	−388474.01	6.91	−388475.71	5.66

[a] $\Delta G^0(H^+(g)) = -6.28$ kcal/mol, $\Delta G^0_{solv} = -258.32$ kcal/mol, $\epsilon_{QM} = 0.92$. [b] Total $\Delta G^0(g) = E_{0K}(g) + ZPE(g) + \Delta\Delta G_{0\rightarrow 298K}(g)$, total $\Delta G^0(aq)$ = total $\Delta G^0(g) + \Delta G^0_{solv}$. [c] Converted from hartree using 1 hartree = 627.5095 kcal/mol. [d] Figure 1d. [e] Figure 1e.

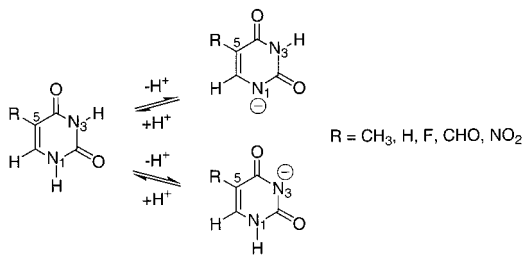


Figure 3. Two possible deprotonation sites of 5-substituted uracils: N1 and N3.

2.6. Dielectric Constant. For the dielectric constant of water, we used the experimental value of $\epsilon_{H_2O} = 80$ (based on $T \sim 300$ K).[33]

In principle we could take $\epsilon_{QM} = 1$ for the dielectric constant of the region being described in the QM. However, in this solvation model there are several limitations: (1) The setting $\epsilon_{QM} = 1$ assumes that the polarizability of the QM part is exact, but with the level of wave function and basis set considered here it is likely that this is not the case. (2) The experimental pK_a was determined at a constant ionic strength of 0.1 M NaCl at room temperature.[14] The dielectric constant of this solution, especially at the interface with solute, might be different from that of pure water. (3) It might be appropriate to use different radii for each atom after deprotonation due to the change of electron distribution, but we used the same radii for each atom before and after deprotonation.

Consequently, we varied ϵ_{QM} to find the value that gives the same range of pK_a values as the experimental one (Figure A1). We decided to use the slightly lower value of $\epsilon_{QM} = 0.92$. With this choice, the $\Delta G^0_{solv}(H^+)$ was determined to be −258.32 kcal/mol, as mentioned in section 2.5.

We also considered the effect of making ϵ_{H_2O} larger than 80. However, as indicated in Table A2, the range of pK_a is insensitive to this parameter.

The solvation model is sensitive to the atomic vdW radii used to estimate the boundary between QM and continuum space. Keeping $\epsilon_{QM} = 1.0$ but scaling down the atomic radii uniformly by 11% from the initial values, we also obtained the same range of pK_a values as experiment [Figure A1(d)].

3. Results

The detailed energy values used to calculate pK_a are listed in Table 1.

For 5-formyluracil there are two conformational isomers shown in Figure 1(d) and 1(e). The calculations found that the trans conformation (Figure 1(d)) is preferred by 4.5 kcal/mol in gas-phase but has a 4.6 kcal/mol lower solvation energy, leading to similar energies for both conformers in solution (the difference is less than 0.1 kcal/mol). Thus, to calculate the pK_a value for 5-formyluracil we include both conformers with the appropriate Boltzmann-average.

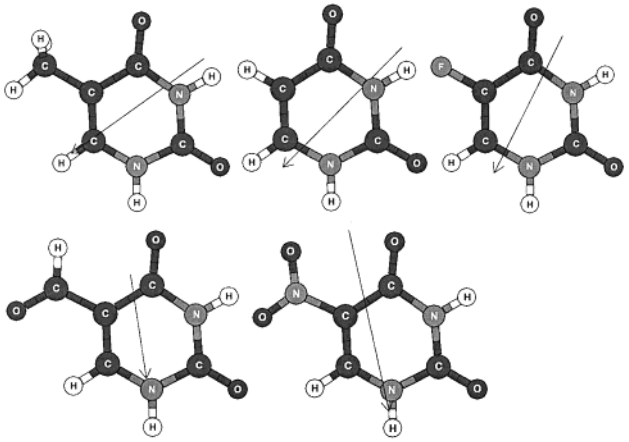
3.1. Site of Protonation. For 5-substituted uracils, there are two possible deprotonation sites: N_1H and N_3H (Figure 3).

TABLE 2: Dipole Moments (in debyes[a]) of 5-R-Uracil's and Their Anions in (a) Gas-Phase [B3LYP/aug-cc-pVTZ(-f)] and (b) Aqueous-Phase [B3LYP/cc-pVTZ(-f)]

	(a) gas phase: μ(calc;g)			(b) aqueous phase: μ(calc;aq)		
5-R	neutral	anion $N_3(-)$	anion $N_1(-)$	neutral	anion $N_3(-)$	anion $N_1(-)$
CH_3	4.47[d]	7.95	3.78	6.68	13.02	6.25
H	4.49[d,e]	7.36	2.15	6.78	12.18	3.84
F	4.12[d]	7.52	2.59	6.37	12.15	3.95
trans-CHO[b]	2.85	5.64	1.39	5.10	9.61	2.67
cis-CHO[c]	5.85	8.01	1.29	9.62	13.88	2.25
NO_2	5.01[d]	6.68	0.10	7.76	11.59	1.19

[a] 1 D = 3.336 × 10^{-30} C m = 0.393 au (the dipole moment of two charges ±e that are 0.2082 Å apart) [b] Figure 1d. [c] Figure 1e. [d] μ(exptl; dioxane solution)[43] = 4.13 ± 0.03 (thymine), 4.16 ± 0.04 (uracil), 4.11 ± 0.05 (5-fluorouracil), 5.47 ± 0.02 (5-nitrouracil). [e] μ(exptl; gas-phase) = 3.87 ± 0.4.[44] μ(calc; g) = 4.12 (MP2/6-311G*),[45] 4.37 (MP2/aug-cc-pVDZ),[45] 4.21 (MP2/6-31G*),[46] 4.43 (DFT(Perdew86)/DZP),[47] 4.44 (DFT(Becke88-Perdew86)/DZP).[47] The dipole moment vectors of neutral 5-R-uracil's in gas-phase are shown in Chart 1.

CHART 1



For thymine, uracil, and 5-fluorouracil, the pK_a values of the N3 site are lower than for the N1 site. In the gas phase, deprotonation from N_1H of these species is 10 to 12 kcal/mol *more* favorable than from N_3H (calculated from gas-phase energy components in Table 1), as pointed out in earlier calculations.[40,41] However, the solvation of the $N_1(-)$ species is 13 to 14 kcal/mol *less* favorable than for the $N_3(-)$ species (Table l). This is because $N_1(-)$ has greater charge delocalization than $N_3(-)$. For example (see Table 2), the gas-phase dipole moment of the $N_1(-)$ species of uracil is 2.15 D, whereas it is 7.36 D for the $N_3(-)$ species. Thus, the experimental pK_a values of thymine, uracil, and 5-fluorouracil correspond to the deprotonation from N_3H.

However, for 5-formyluracil and 5-nitrouracil, the solution-phase deprotonation from N_1H is more favorable than from N_3H. The solvation of the $N_1(-)$ species is still 13−16 kcal/mol *less* favorable than for the $N_3(-)$ species. However, in these cases the gas-phase deprotonation from N_1H is 15 to 17 kcal/mol *more* favorable than from N_3H. This is plausible since these 5-formyl and 5-nitro substituents can stabilize the $N_1(-)$ species by delocalizing negative charge more extensively[3] as indicated in Figure 4. This effect can be seen from the dipole moment and geometry change. [In Table 2, the dipole moments of more charge-delocalized $N_1(-)$ species are much less than those of more charge-localized $N_3(-)$ species, and this is especially prominent for 5-formyluracil and 5-nitrouracil. In Figure 5b, the torsion angle of C_4-C_5-N-O of $N_1(-)$ of 5-nitrouracil

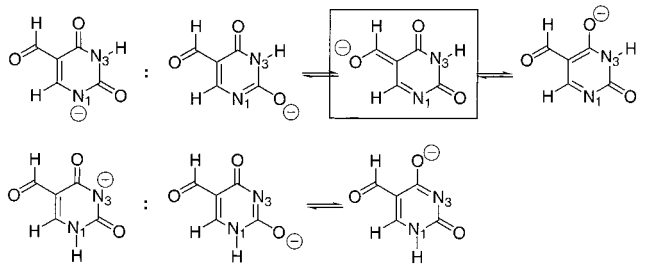


Figure 4. Comparison of delocalization upon ionization at N1 and N3.

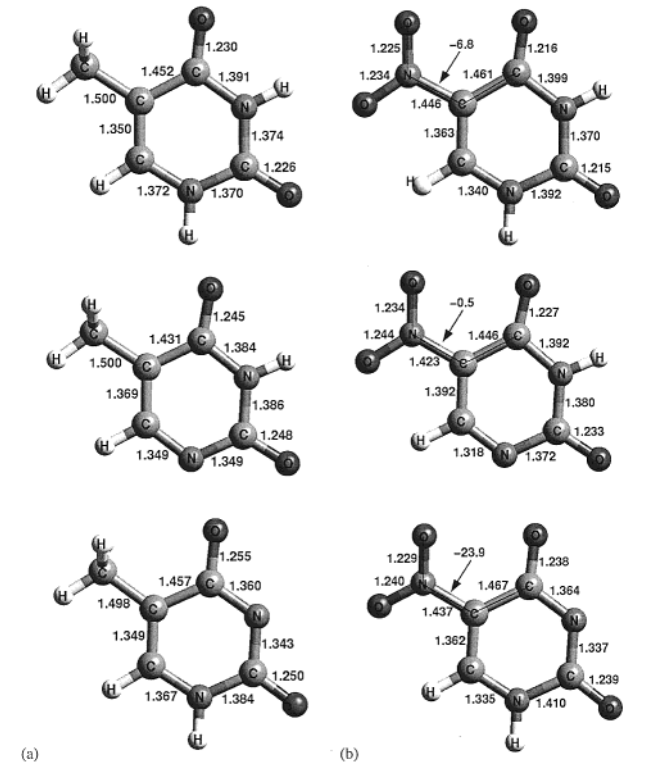


Figure 5. Geometry change during deprotonation of 5-R-uracil's (solution-phase; B3LYP/cc-pVTZ(-f); top) from N_1H (middle) and from N_3H (bottom): (a) thymine and (b) 5-nitrouracil. Bond lengths are in Å and torsion angles given with arrows are in degrees.

(0.5°) is much less than that of its $N_3(-)$ counterpart (23.9°), that is, $N_1(-)$ is more planar than $N_3(-)$, and the C_5-N distance of $N_1(-)$ (1.423 Å) is shorter than that of $N_3(-)$ (1.437 Å). This shows the enhanced resonance of $N_1(-)$ species of 5-formyluracil and 5-nitrouracil as described in Figure 4. This difference is not significant in other 5-substituted Uracils as shown in Figure 5a.] Thus, the experimental pK_a values of 5-formyluracil and 5-nitrouracil correspond to deprotonation from N_1H.

This result is consistent with the experimental observation made on deoxyuridine where the N_1 is bonded to sugar ring so that deprotonation can occur only from the N3 site.[3] The pK_a values of thymine, uracil, and 5-fluorouracil are very similar to those of the corresponding deoxyuridines (9.75 vs 9.69, 9.42 vs 9.26, and 7.93 vs 7.67), but the pK_a of 5-formyluracil is significantly lower than that of 5-formyldeoxyuridine (6.84 vs 8.12). (There is no experimental observation on 5-nitrodeoxyuridine.)

3.2. Comparison with Experiment. Table 1 reports the calculated pK_a values for the N1 and N3 position for this series

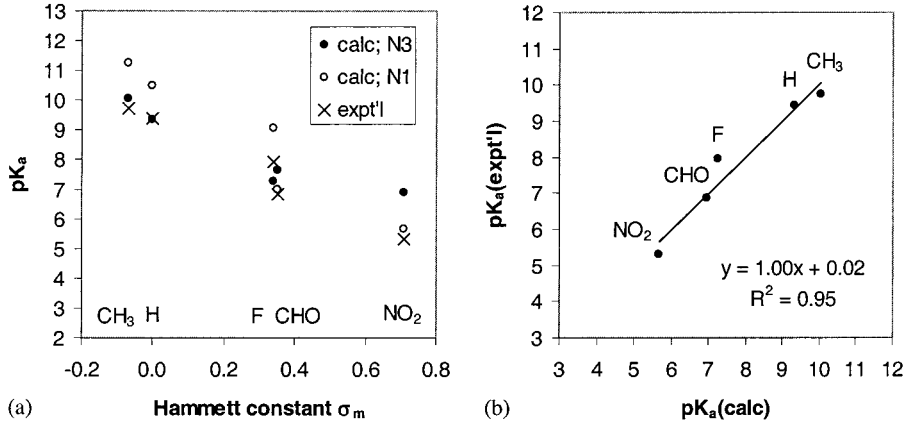


Figure 6. pK_a values of 5-R-uracils (R = CH_3, H, F, CHO, and NO_2): (a) calculated and experimental pK_a's plotted as a function of Hammett constant σ_m, and (b) calculated pK_a versus experimental pK_a.

TABLE 3: Calculated and Experimental pK_a's and Hammett Meta Constants (σ_m) of 5-Substituted Uracil Derivatives[a]

5-R	ó$_m$[b]	pK_a(calc;N3)	pK_a(calc;N1)	pK_a(exptl)[c]
CH_3	−0.07	**10.04**	11.23	9.75
H	0.00	**9.34**	10.47	9.42
F	0.34	**7.26**	9.05	7.93
CHO	0.35	7.58	**6.94**	6.84
NO_2	0.71	6.91	**5.66**	5.30

[a] The lower value of two calculated pK_a's for each compound (shown in boldface) should be compared with the experimental one. [b] References 3 and 48. [c] Reference 3.

of uracil derivatives. Here we see that the preferred site of ionization (lower pK_a value) is N3 for thymine, uracil, and 5-fluorouracil, whereas the 5-formyl and 5-nitro substituents prefer N1 ionization. In aqueous solution, the experimental pK_a corresponds to this preferred site of ionization.

Table 3 and Figure 6 compare the calculated pK_a values with experiment. We see a strong correlation between the experimental pK_a and that calculated for the preferred site of ionization.

$$pK_a^{expr} = 1.00pK_a^{calc} + 0.02; \quad r^2 = 0.95 \qquad (4)$$

Thus, eq 4 leads to predicted pK_a values in good agreement with experiment.

3.3. Dipole Moment Orientation. Table 2 shows the dipole moment vectors superimposed with the structures of neutral 5-R-uracils. The orientation of dipole moments changes little when the 5-CH_3 (thymine) or 5-H (uracil) were substituted with 5-F, but it changes substantially when substituted with 5-CHO or 5-NO_2. This change in the dipole moment orientation might change the proper orientation for the best electrostatic interaction between base pairs, which might in turn alter the base−base interactions in a DNA helix and the coding properties of those substituted uracils. This issue would be interesting to pursue in the future.

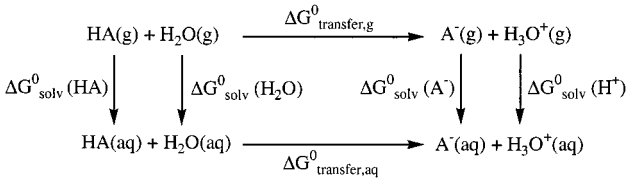
4. Conclusions

We used first principles density functional theory with the Poisson−Boltzmann continuum-solvation model to calculate the pK_a for both the N1 and N3 positions of a series of 5-substituted uracil derivatives. This series of 5-substituted uracil derivatives was chosen to have a wide range of pK_a values, extending both above and below physiological pH. This provides a good test of the theory since the 5-substituent can change the preference for ionization of the N1 and N3 positions. In aqueous solution, the more acidic proton will be lost first with increasing solution pH, so that the experimental pK_a value corresponds to the preferred ionization site. Thus, the experimental pK_a values must be compared with the pK_a value calculated to be the preferred site of ionization. This validates the computational method for predicting both the preferred site of ionization and the pK_a. We can now use this methodology to examine the effects of additional modifications and derivatives including deoxynucleosides, deoxynucleotides, and oligodeoxynucleotides.

Acknowledgment. This work was supported in part by the National Institutes of Health [HD36385 (W.A.G.), GM 41336 (L.C.W.), and CA 33572 (L.C.W.)]. In addition, the facilities of the MSC are also supported by DOE-ASCI, ARO-MURI, ARO-DURIP, National Science Foundation (CHE-95-22179), Exxon Corp., Dow Chemical, 3M, Beckman Institute, Avery-Dennison, Chevron Corp., Seiko Epson, Asahi Chemical, and BP Amoco.

Appendix

The thermodynamic cycle used in the pK_a calculation can also be written as follows:[42]



The equilibrium constant for the proton-transfer process in aqueous phase is given by

$$K' = \frac{[A^-]_{aq}[H_3O^+]_{aq}}{[AH]_{aq}[H_2O]_{aq}} = \frac{[A^-]_{aq}[H^+]_{aq}}{[AH]_{aq}[H_2O]_{aq}} = \frac{K}{[H_2O]_{aq}}$$

where K is the equilibrium constant for the deprotonation process $AH(aq) \rightarrow A^-(aq) + H^+(aq)$. From the relationship

$$\Delta G^0_{transfer,aq} = -2.303RT \log K' = -2.303RT \log \frac{K}{[H_2O]_{aq}} = 2.303RT\, pK_a + 2.36$$

the pK_a of HA(aq) is given by

$$pK_a \left(= \frac{1}{2.303RT}\Delta G^0_{deprot,aq}\right) = \frac{1}{2.303RT}[\Delta G^0_{transfer,aq} - 2.36]$$

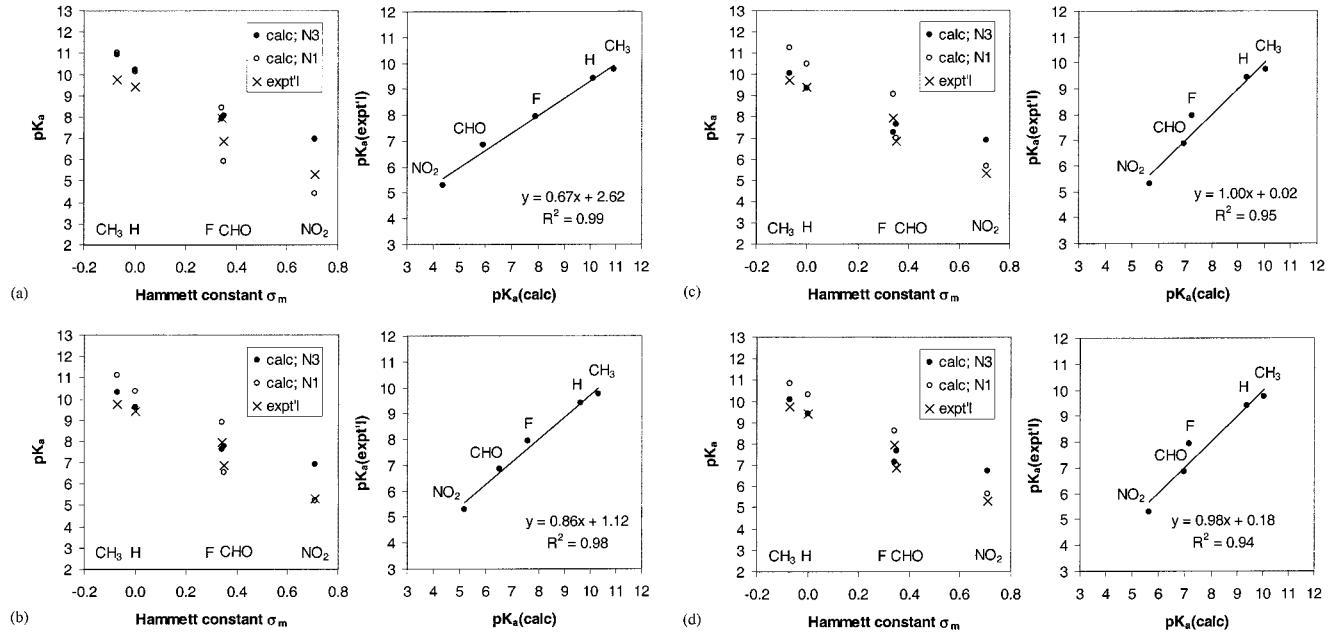


Figure A1. Dependence of calculated pK_a on the dielectric constant ϵ_{QM} (a, b, and c) and on the vdW radii (a and d): (a) $\epsilon_{QM} = 1.00$ [$\Delta G^0_{solv}(H^+)$] $= -263.01$ kcal/mol], (b) $\epsilon_{QM} = 0.95$ [$\Delta G^0_{solv}(H^+)$] $= -260.11$ kcal/mol], (c) $\epsilon_{QM} = 0.92$ [$\Delta G^0_{solv}(H^+)$] $= -258.32$ kcal/mol], and (d) $\epsilon_{QM} = 1.00$, the vdW radii scaled down by 11% [$\Delta G^0_{solv}(H^+)$] $= -262.65$ kcal/mol].

TABLE A1: Zero-Point Energies and Free Energy Corrections Calculated in Two Different Levels and Multiplied by the Appropriate Scaling Factors:[28] (a) HF/6-31G* and (b) B3LYP/cc-pVTZ(-f)[a]

	(a) HF/6-31G*		(b) B3LYP/cc-pVTZ(-f)	
	ZPE[b]	$\Delta\Delta G_{0\rightarrow 298K}$[c]	ZPE[d]	$\Delta\Delta G_{0\rightarrow 298K}$[e]
uracil				
neutral	54.23	−19.09	54.51 (54.11)[d]	−19.13
anion: $N_1(-)$	45.90	−18.90	45.93 (45.60)	−18.96
anion: $N_3(-)$	45.45	−19.17	45.43 (45.10)	−19.25
thymine				
neutral	71.33	−20.37	71.81 (71.29)	−20.41
anion: $N_1(-)$	62.91	−20.24	63.16 (62.70)	−20.31
anion: $N_3(-)$	62.53	−20.42	62.76 (62.30)	−20.49

[a] We conclude that the HF level is adequate. [b] Scaled down by 0.9135. [c] The enthalpy and entropy contributions were scaled down by 0.8905 and 0.8978, respectively. [d] Not scaled down. [e] Since ref 28 does not list the scaling factor for B3LYP/cc-pVTZ(-f), we estimated it as 0.9927 from those for B3LYP/6-31G*, HF/6-31G*, and HF/6-311G** by 0.9806[B3LYP/6 − 31G*] × {0.9248[HF/6 − 311G**]/0.9135[HF/6 − 31G*]} = 0.9927[B3LYP/cc − pVTZ(-f)] assuming the same scaling factor for both triple-ζ basis sets, cc-pVTZ(-f) and 6-311G**.

Since

$$\Delta G^0_{transfer,aq} = \Delta G^0(A^-(aq)) + \underline{\Delta G^0(H_3O^+(aq)) - \Delta G^0(H_2O(aq))} - \Delta G^0(HA(aq))$$

and

$$\Delta G^0_{deprot,aq} = \Delta G^0(A^-(aq)) + \underline{\Delta G^0(H^+(aq))} - \Delta G^0(HA(aq))$$

$\Delta G^0(H^+(aq))$ is expected to be equal to [$\Delta G^0(H_3O^+(aq)) - \Delta G^0(H_2O(aq)) - 2.36$]. From the standard free energies of H_2O(aq) and H_3O^+(aq) calculated in the same way as done for 5-R-uracil's and their anions, $\Delta G^0(H^+(aq))$ is determined to be −265.66 kcal/mol. This alternative approach gives a very similar value to the one (−264.60 kcal/mol) determined from the best match between the experimental pK_a's and the calculated ones in sections 2.5 and 2.6.

TABLE A2: Dependence of Solvation Energies of 5-R-Uracils and Their Anions on ϵ_{QM} and ϵ_{H_2O}

ΔG_{solv}		$\epsilon_{H_2O} = 80$[a] $\epsilon_{QM} = 1.00$[a]	$\epsilon_{H_2O} = 96$ $\epsilon_{QM} = 1.00$	$\epsilon_{H_2O} = 80$[b] $\epsilon_{QM} = 0.92$[b]
CHO	N3(−)	−73.166	−73.370	−80.723
	N1(−)	−61.545	−61.704	−67.583
	neutral	−20.054	−20.130	−22.741
F	N3(−)	−76.054	−76.268	−84.123
	N1(−)	−64.914	−65.073	−71.254
	neutral	−17.994	−18.059	−20.507
H	N3(−)	−79.658	−79.883	−87.941
	N1(−)	−67.699	−67.937	−74.310
	neutral	−17.877	−17.938	−20.289
CH_3	N3(−)	−78.256	−78.481	−86.507
	N1(−)	−67.132	−67.307	−73.889
	neutral	−16.976	−17.042	−19.322
NO_2	N3(−)	−69.910	−70.110	−77.398
	N1(−)	−57.212	−57.359	−62.881
	neutral	−20.532	−20.612	−23.237

[a] The original value. [b] The value used in this work.

References and Notes

(1) Myers, L. S.; Ward, J. F.; Tsukomato, W. T.; Holmes, D. E.; Julca, J. R. *Science* **1965**, *148*, 1234−1235.
(2) Teebor, G. W.; Frenkel, K.; Goldstein, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 318−321.
(3) Privat, E.; Sowers, L. *Mutat. Res.* **1996**, *354*, 151−156.
(4) Privat, E.; Sowers, L. *Chem. Res. Toxicol.* **1996**, *9*, 745−750.
(5) Imlay, J. A.; Linn, S. *Science* **1988**, *240*, 1302−1309.
(6) Loeb, L. A. *Cancer Res.* **1989**, *49*, 5489−5496.
(7) Ames, B. N.; Gold, L. S.; Willett, W. C. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 5258−5265.
(8) Heidelberger, C. *Prog. Nucleic Acid Res. Mol. Biol.* **1965**, *4*, 1−50.
(9) Rowe, W. P.; Lowy, D. R.; Teich, N.; Hartley, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 1033−1035.
(10) Sternglanz, H.; Bugg, C. E. *Biochim. Biophys. Acta* **1975**, *378*, 1−11.
(11) Morris, S. M. *Mutat. Res.* **1993**, *297*, 39−51.
(12) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1983.
(13) Lawley, P. D.; Brooks, P. D. *J. Mol. Biol.* **1962**, *4*, 216−219.
(14) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. *Mutat. Res.* **1987**, *177*, 201−218.
(15) Sowers, L. C.; Goodman, M. F.; Eritja, R.; Kaplan, B.; Fazakerley, G. V. *J. Mol. Biol.* **1989**, *205*, 437−447.

(16) Yu, H.; Eritja, R.; Bloom, L. B.; Goodman, M. F. *J. Biol. Chem.* **1993**, *268*, 15935−15943.
(17) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610−5620.
(18) Topol, I. A.; Tawa, G. J.; Burt, S. K. *J. Phys. Chem. A* **1997**, *101*, 10075−10081.
(19) *Jaguar 3.5*; Schrodinger Inc.: Portland, OR, 1998.
(20) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; III, W. A. G.; Donnelly, R. E.; Ringalda, M. N. *J. Chem. Phys.* **1994**, *101*, 4028−4041.
(21) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(22) Slater, J. C. *Quantum Theory of Molecules and Solids. Vol. 4. The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974.
(23) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(24) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(25) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789.
(26) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(27) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(28) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502−16513.
(29) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M. N.; Goddard, W. A., III; Honig, B. A. *J. Am. Chem. Soc.* **1994**, *116*, 11875−11882.
(30) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144−1149.
(31) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775−11788.
(32) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(33) Archer, D. G.; Wang, P. *J. Phys. Chem. Ref. Data* **1990**, *19*, 371−411.
(34) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1990**, *8*, 894−907.
(35) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H., Jr. *J. Comput. Chem.* **1990**, *11*, 297−310.
(36) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(37) Reiss, H.; Heller, A. *J. Phys. Chem.* **1985**, *89*, 4207−4213.
(38) Marcus, Y. *Ion Solvation*; John Wiley and Sons: Ltd., 1985; pp 105−109.
(39) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys. Chem. A* **1998**, *102*, 7787−7794.
(40) Nguyen, M.; Chandra, A.; Zeegers-Huyskens, T. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1277−1280.
(41) Chandra, A.; Nguyen, M.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1998**, *102*, 6010−6016.
(42) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194−11199.
(43) Kulakowska, I.; Geller, M.; Lesyng, B.; Wierzchowski, K. L.; Bolewski, K. *Biochim. Biophys. Acta* **1974**, *361*, 119−130.
(44) Brown, R. B.; Godfrey, P. D.; McNaughton, D.; Pierlot, A. P. *J. Am. Chem. Soc.* **1988**, *110*, 2329−2330.
(45) van Mourik, T.; Price, S.; Clary, D. *J. Phys. Chem. A* **1999**, *103*, 1611−1618.
(46) Leszczynski, J. *J. Phys. Chem.* **1992**, *96*, 1649−1653.
(47) Estrin, D. A.; Paglieri, L.; Corongiu, G. *J. Phys. Chem.* **1994**, *98*, 5653−5660.
(48) Perrin, D. D.; Dempsey, B.; Serjeant, E. P. *pK_a prediction for organic acids and bases*; Chapman and Hall: London and New York, 1981.