Theoretical Estimations of the 298 K Gas-Phase Acidities of the Pyrimidine-Based Nucleobases Uracil, Thymine, and Cytosine

Yiqun Huang and Hilkka Kenttämaa*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1393

Received: December 5, 2002; In Final Form: March 17, 2003

Theoretical estimations at the B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level for the 298 K gas-phase acidities (GPAs) of all the possible deprotonation sites of the most stable tautomers of uracil, thymine, and cytosine are reported. The gas-phase acidities of the nucleobases estimated by direct calculations and derived from isodesmic reactions are very similar. For uracil, the GPAs of the N1, N3, C5, and C6 sites are 332.8 ± 2.2, 345.1 ± 2.2, 380.2 ± 0.5, and 366.2 ± 0.5 kcal/mol, respectively. For thymine, the GPAs of the N1, N3, C5−CH_3, and C6 sites are 334.5 ± 2.2, 345.7 ± 2.2, 376.1 ± 2.1, and 367.5 ± 0.5 kcal/mol, respectively. For keto cytosine, the GPAs of N1, N4−H_b, N4−H_a, C5, and C6 are 345.6 ± 2.2, 348.7 ± 2.1, 353.5 ± 2.1, 379.3 ± 0.5, and 371.2 ± 0.5 kcal/mol, respectively. For *trans*-enol cytosine, the GPAs of the hydroxyl group, N4−H_b, N4−H_a, C5, and C6 are 347.2 ± 2.1, 351.9 ± 2.1, 355.4 ± 2.1, 386.4 ± 0.5, and 393.8 ± 0.5 kcal/mol, respectively. For uracil, thymine, and keto cytosine, N1 is the most acidic site while C5 is the least acidic one. For enol cytosine, the most acidic site is the hydroxyl group and the least acidic site is C6. On the basis of these results, uracil and thymine are more acidic than cytosine. The acidities of the keto and *trans*-enol cytosine are approximately equal.

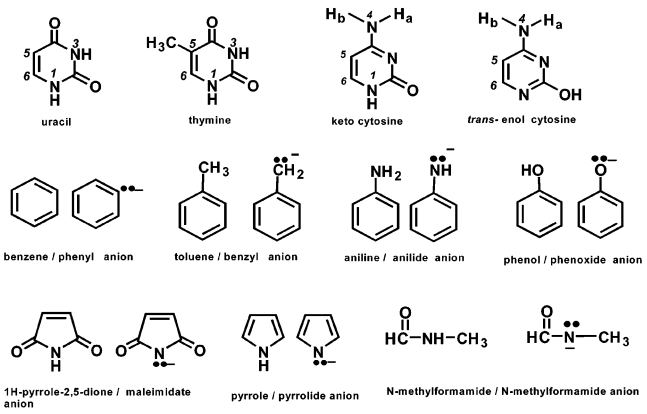
Introduction

The pyrimidine-based nucleobases uracil, thymine, and cytosine are important components of nucleic acids. The protonation and deprotonation of these nucleobases are of importance in many biological processes and are controlled by both the intrinsic acid−base equilibrium properties of the nucleobases and their environment. The intrinsic acid−base properties of nucleobases can be characterized by their gas-phase acidity and basicity (or proton affinity) because, in the gas phase, any interference caused by the presence of solvent is eliminated.

Similar to the situation of aqueous solution, the dioxo form is the major tautomeric form of thymine and uracil in the gas phase. However, both keto and *trans*-enol forms are important for cytosine in the gas phase, while the keto tautomer is predominant in aqueous solution.[1−4] For all these nucleobases, there are more than one site in the molecule that can be deprotonated. For uracil and thymine, the possible deprotonation sites are N1, N3, C6, and C5 (for uracil) and C5−CH_3 (for thymine). For keto cytosine, the possible deprotonation sites are N1, N4−H_a, N4−H_b, C5, and C6, and for *trans*-enol cytosine, they are OH, N4−H_a, N4−H_b, C5, and C6. The structures and numbering schemes of possible deprotonation sites of these nucleobases are shown in Chart 1.

The gas-phase acidities (ΔH) of nucleobases have not been thoroughly investigated. Chen *et al*. reported the gas-phase acidity of 14.1 eV (325.2 kcal/mol) for adenine, guanine, and thymine, 14.2 eV (327.5 kcal/mol) for uracil, and 14.3 eV (329.8 kcal/mol) for cytosine, as obtained by using negative chemical ionization mass spectrometry.[5] These values correspond to the most acidic sites of the corresponding nucleobases. Beauchamp *et al*. performed AM1 semiempirical molecular orbital calculations on the gas-phase acidities of all nucleobases except for uracil.[6] For thymine, their calculations indicated that the most acidic site is N1, followed by N3, with the gas-phase acidities of 328.0 and 344.1 kcal/mol, respectively. For cytosine (keto form), the most acidic site was reported to be N1, followed by N4−H_b and N4−H_a, with gas-phase acidities of 338.9, 345.6, and 351.5 kcal/mol, respectively. *N*-Methylformamide (Chart 1) was used to estimate the accuracy of the AM1 calculations via comparison to experimental results. In the case of *N*-methylformamide, the gas-phase acidity calculated at the AM1 level is 366.3 kcal/mol, about 5−6 kcal/mol higher than the experimental value of 360.5 ± 2.1 kcal/mol.[6] Beauchamp *et al*. therefore assumed that the gas-phase acidities of keto cytosine and thymine predicted by AM1 calculations were also about 5 kcal/mol higher than the real values.[6] The corrected GPA(N1) of thymine and cytosine would be 323 (lower than Chen *et al*.'s result by about 2 kcal/mol [5]) and 334 kcal/mol (higher than Chen *et al*.'s result by about 4 kcal/mol [5]), respectively.

CHART 1: Structures of Uracil, Thymine, Cytosine and Reference Systems and the Numbering Schemes for Possible Deprotonation Sites



* Corresponding author.

10.1021/jp022530i CCC: $25.00
Published on Web 05/23/2003

Zeegers-Huyskens *et al.*[7] calculated the deprotonation energies of the two NH bonds of uracil by using density functional theory at the B3LYP/6-31G(d,p) and B3LYP/6-31++G(d,p) levels of theory. At the B3LYP/6-31++G(d,p) level of theory, the deprotonation energies of the N1 and N3 sites are 332.8 and 346.2 kcal/mol, respectively. Recently, Lee *et al.*[8] reported experimentally bracketed gas-phase acidities of the N1 (333 ± 4 kcal/mol) and N3 (347 ± 4 kcal/mol) sites of uracil, obtained by using Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry. They also theoretically estimated the gas-phase acidity of the N1 (329 kcal/mol) and N3 (343 kcal/mol) sites of uracil at the B3LYP/6-31+G(d) level of theory. All these values are in reasonably good agreement. However, they are higher than those obtained by Chen *et al.*[5] Wu *et al.*[9] generated the $C6^-$ anion of 1,3-dimethyluracil in a quadrupole ion trap mass spectrometer via collision-activated dissociation (CAD) of precursor carboxylate anions and measured the proton affinity of that anion (369.9 ± 3.1 kcal/mol), a value that is equivalent to the gas-phase acidity of the C6 site of 1,3-dimethyluracil. This value is consistent with their theoretical calculations carried out at the MP2/6-31+G(d,p)//HF/6-31+G(d) (367.6 kcal/mol) and B3LYP/6-31+G(d,p)//HF/6-31+G(d) (366.0 kcal/ mol) levels of theory. It can be assumed that the GPA(C6) of 1,3-dimethyluracil is similar to that of uracil.

To the best of our knowledge, no higher level theoretical investigations beyond the AM1 level[6] have been reported for the gas-phase acidities of the different deprotonation sites of pyrimidine bases except for uracil and 1,3-dimethyluracil.[7−9] The presence of more than one possible deprotonation site in a nucleobase molecule makes the experimental measurement of the gas-phase acidities of all possible sites difficult.[8] Only the gas-phase acidities of the most acidic site have been measured for pyrimidine bases, except for uracil,[8] and such measurements cannot identify which site of the nucleobase is the most acidic.[5] The main purpose of this work was to carry out a comprehensive theoretical investigation on the gas-phase acidities of all acidic sites of the pyrimidine-based nucleobases uracil, thymine, and cytosine. Such a theoretical investigation may help the experimental determination of the gas-phase acidities of the nucleobases. Another purpose of this work was to compare the results obtained from direct and isodesmic reaction-based calculations in order to determine whether the utilization of isodesmic reaction-based calculations is necessary for this type of compounds. Generally speaking, the isodesmic reaction-based methodology helps to cancel the errors caused by limited basis set effects in reaction enthalpy calculations. However, the accuracy of isodesmic reaction-based calculations depends on the accuracy of the known property (gas-phase acidity in our case) of the reference system, and in many cases, suitable reference systems are not available. However, for pyrimidine-based nucleobases, suitable reference systems can be found. Therefore, these bases can be used as a testing set. If direct and isodesmic reaction-based calculations give rise to similar results for uracil, thymine, and cytosine, then it might be reasonable to use the direct calculation methodology to estimate the gas-phase acidities of other nucleobases and their derivatives for which no suitable reference systems are available.

Computational Methods

In this work, the 298 K gas-phase acidities of all possible deprotonation sites of uracil, thymine, and cytosine were evaluated theoretically by using both direct calculations and isodesmic reaction methodologies.

TABLE 1: Gas-Phase Acidities at 298 K of All Possible Deprotonation Sites of Uracil Calculated at the B3LYP/6-31+G(d)//B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) Levels of Theory

	298 K GPA(kcal/mol)			
	B3LYP/6-31+G(d)//B3LYP/6-31+G(d)		B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d)	
deprotonation site	*a*	*b*	*a*	*b*
N1	332.1 ± 2.2	330.5	332.8 ± 2.2	334.2
N3	345.8 ± 2.2[c]	344.1	345.1 ± 2.2[c]	346.6
	329.7 ± 4.6[d]		329.1 ± 4.6[d]	
C5	379.4 ± 0.5	377.6	380.2 ± 0.5	379.8
C6	364.8 ± 0.5	363.0	366.2 ± 0.5	365.9

[a] GPA values were derived from isodesmic reactions. [b] GPA values were calculated directly from the enthalpies of neutral uracil, its conjugate base, and a proton at 298 K. [c] Pyrrole was used as the reference molecule. [d] 1*H*-Pyrrole-2,5-dione was used as the reference molecule.

The gas-phase acidity (GPA) of molecule HA is defined as the 298 K enthalpy change of the reaction:

$$HA \rightarrow H^+ + A^- \quad (1)$$

The calculated 298 K enthalpies of H^+, A^-, and HA can be used to directly obtain the GPA via (1).

Another way to estimate GPA is via an isodesmic reaction 2:

$$HA + R^- \rightarrow RH + A^- \quad (2)$$

where in RH is a reference molecule whose 298 K GPA is known and R^- is its conjugate base. The GPA of HA can be derived using the formula

$$GPA(HA) = GPA(RH) + \Delta H_{298} \quad (3)$$

Here ΔH_{298} is the directly calculated 298 K enthalpy change of reaction 2.

Aniline, benzene, toluene, pyrrole, and phenol, as well as 1*H*-pyrrole-2,5-dione, were chosen as the reference molecules. The structures of these reference molecules and their conjugate bases are shown in Chart 1. All calculations reported in this work were performed using the Guassian98 program.[10] The geometries of the isolated dioxo forms of uracil and thymine, keto and *trans*-enol forms of cytosine, and all reference molecules, as well as the corresponding conjugate bases, were fully optimized by using density functional theory at the B3LYP/6-31+G(d) level. The single point energy calculations were carried out at the B3LYP/aug-cc-pVTZ level of theory on all B3LYP/6-31+G(d) geometries. B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) was the highest affordable level of theory. In both direct and isodesmic reaction-based calculations at the B3LYP/6-31+G(d) level of theory, 298 K enthalpies were estimated as the electronic energy modified by the 298 K enthalpy correction. In B3LYP/aug-cc-pVTZ single point calculations, 298 K enthalpies were estimated as the B3LYP/aug-cc-pVTZ level electronic energy corrected by the B3LYP/6-31+G(d) level 298 K enthalpy corrections. In direct GPA calculations, the enthalpy of the proton at 298 K was chosen to be 0.00236 hartree (5/2*RT*).

Results and Discussions

1. Uracil. The theoretically evaluated 298 K gas-phase acidities of different deprotonation sites of uracil are shown in Table 1. For the N1 site, pyrrole was chosen as the reference

molecule with 298 K GPA of 358.6 ± 2.2 kcal/mol to form its conjugate base pyrrolide anion.[11] At the B3LYP/6-31+G(d) level of theory, direct and isodesmic reaction-based calculations predict GPA(N1) to be 330.5 kcal/mol and 332.1 ± 2.2 kcal/mol, respectively. The former is equivalent to the 329 kcal/mol calculated by Lee *et al*.[8] at the same level of theory at 0 K. At the B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory, the GPA values predicted by our direct and isodesmic-reaction-based calculations are 334.2 and 332.8 ± 2.2 kcal/mol, respectively. The uracil results obtained at this highest level of theory are also very close to the experimental values (333 ± 4 kcal/mol) bracketed by Lee *et al*.[8] and the theoretical predictions (332.8 kcal/mol) obtained by Zeegers-Huyskens *et al*.[7] at the B3LYP/6-31++G(d,p) level of theory at 0 K. The GPA(N1) of 345 kcal/mol estimated by Zeegers-Huyskens *et al*.[7] at the B3LYP/6-31G(d,p) level of theory is obviously too high and contradicts the experimental results (333 ± 4 kcal/mol).[8] The reason for this overestimation of the gas-phase acidity might be the lack of including diffusion functions into the basis set. It is commonly recognized that adding diffusion functions into the basis set is essential for the calculation of anions. On the basis of our calculations, it is reasonable to assume that 332.8 ± 2.2 kcal/mol is a good estimation for the 298 K gas-phase acidity of the N1 site of uracil. Therefore, the gas-phase acidity of uracil of 327.5 kcal/mol (14.2 eV) experimentally estimated by Chen *et al*.[5] might be slightly low.

In general, for almost all cases studied in this paper, the results obtained by direct and isodesmic reaction-based calculations are very similar. Therefore, from now on, only the GPA values predicted by isodesmic reaction-based calculations are discussed unless there is a significant discrepancy between the results obtained from direct and isodesmic reaction-based calculations.

For the N3 sites of uracil, the pyrrole/pyrrolide anion pair, and the 1*H*-pyrrole-2,5-dione/maleimidate anion pair (with 298 K GPA of 325.1 ± 4.6 kcal/mol)[11] were chosen as the reference systems. For pyrrole as the reference molecule, the derived values of GPA(N3) are 345.8 ± 2.2 and 345.1 ± 2.2 kcal/mol at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory, respectively. These data are close to the experimental result (347 ± 4 kcal/mol) reported by Lee *et al*.[8] and theoretical estimation (346.2 kcal/mol, B3LYP/6-31++G(d,p) level of theory) made by Zeegers-Huyskens *et al*.[7] When 1*H*-pyrrole-2,5-dione was used as the reference molecule, the GPA values derived from the isodesmic reaction-based calculations are significantly lower (329.7 ± 4.6 and 329.1 ± 4.6 kcal/mol for B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) calculations, respectively) and deviate from the value experimentally bracketed by Lee *et al*.[8]

For the C5 site, benzene with the 298 K gas-phase acidity of 401.7 ± 0.5 kcal/mol was used as the reference molecule.[11] GPA(C5) is estimated to be 379.4 ± 0.5 and 380.2 ± 0.5 kcal/mol, respectively, at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory. Benzene was also utilized as the reference molecule for GPA(C6) calculations. The result obtained at the B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level (366.2 ± 0.5 kcal/mol) is close to the recently calculated values (367.6 kcal/mol at the MP2/6-31+G(d,p)//HF/6-31+G(d) level and 366.0 kcal/mol at the B3LYP/6-31+G(d,p)//HF/6-31+G(d) level of theory) and the experimentally bracketed value (369.9 ± 3.1 kcal/mol) of GPA(C6) of 1,3-dimethyluracil.[9] In this case, the results appear not to be sensitive to the choice of the basis set.

2. Thymine. The theoretically evaluated 298 K GPAs of different sites of thymine are listed in Table 2. N1 is the most acidic site. Calculations performed at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory (using pyrrole/pyrrolide anion as the reference system) predict GPA(N1) of 333.9 ± 2.2 and 334.5 ± 2.2 kcal/mol, respectively. These values are about 10 kcal/mol higher than that estimated by Beauchamp *et al*. (323 kcal/mol).[6] To compare our results with those calculated by Beauchamp *et al*.,[6] we estimated the 298 K gas-phase acidity of the N site of *N*-methylformamide by using direct calculations at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory. The corresponding results are 359.6 and 361.7 kcal/mol, respectively, in excellent agreement with the experimental value of 360.5 ± 2.1 kcal/mol.[6] Therefore, our recommended 298 K GPA of the N1 site is 334.5 ± 2.2 kcal/mol. This value is almost 10 kcal/mol higher than that measured by Chen *et al*. (325.2 kcal/mol).[5]

TABLE 2: Gas-Phase Acidities at 298 K of All Possible Deprotonation Sites of Thymine Calculated at the B3LYP/6-31+G(d)//B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) Levels of Theory

deprotonation site	298 K GPA (kcal/mol)			
	B3LYP/6-31+G(d)//B3LYP/6-31+G(d)		B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d)	
	a	*b*	*a*	*b*
N1	333.9 ± 2.2	332.2	334.5 ± 2.2	335.9
N3	346.4 ± 2.2[c]	344.7	345.7 ± 2.2[c]	347.1
	330.3 ± 4.6[d]		329.6 ± 4.6[d]	
C5−CH_3	375.0 ± 2.1	376.1	376.1 ± 2.1	375.9
C6	366.3 ± 0.5	364.5	367.5 ± 0.5	367.2

[a] GPA values were derived from isodesmic reactions. [b] GPA values were calculated directly from the enthalpies of neutral thymine, its conjugate base, and a proton at 298 K. [c] Pyrrole was used as the reference molecule. [d] 1*H*-Pyrrole-2,5-dione was used as the reference molecule.

The gas-phase acidity of the N3 site in thymine is predicted to be 346.4 ± 2.2 and 345.7 ± 2.2 kcal/mol, respectively, at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory (using pyrrole/pyrrolide anion as the reference system). However, if 1*H*-pyrrole-2,5-dione is utilized as the reference molecule, the gas-phase acidity derived from the isodesmic reaction is significantly lower than that obtained from direct and the other isodesmic reaction calculations. There are no experimental results for thymine that help to determine which reference system is more suitable. However, comparison of computational results to experimental data in the case of uracil demonstrated that 1*H*-pyrrole-2,5-dione is not a suitable reference system for gas-phase acidity calculations on the N3 site in this pyrimidine base. Therefore, we draw the conclusion that the pyrrole/pyrrolide anion pair is likely to be the better reference system for N3 gas-phase acidity calculations, and the best estimation for the GPA(N3) is 345.7 ± 2.2 kcal/mol.

For the C5−CH_3 site, the toluene/$PhCH_2^-$ anion pair with 298 K GPA of 380.8 ± 2.1 kcal/mol was selected as the reference system.[11] GPA(C5) is estimated to be 375.0 ± 2.1 and 376.1 ± 2.1 kcal/mol, respectively, at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory. For the C6 site, the calculations suggest that reasonable 298 K gas-phase acidities are 366.3 ± 0.5 and 367.5 ± 0.5 kcal/mol, respectively, at the B3LYP/6-31+G(d)//B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory, using benzene as the reference molecule. These values are very close to that of the C6 site of uracil. For both uracil and thymine, C5, C5−CH_3, and C6 sites are less acidic than N1 and N3, while C5 and C5−CH_3 are the least acidic sites.

TABLE 3: Gas-Phase Acidities at 298 K of All Possible Deprotonation Sites of Cytosine Calculated at the B3LYP/6-31+G(d)//B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) Levels of Theory

	298 K GPA(kcal/mol)			
	B3LYP/6-31+G(d)//B3LYP/6-31+G(d)		B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d)	
deprotonation site	*a*	*b*	*a*	*b*
keto cytosine				
N4−H_b	347.7 ± 2.1	346.6	348.7 ± 2.1	349.4
N4−H_a	353.7 ± 2.1	352.6	353.5 ± 2.1	354.2
N1	345.0 ± 2.2	343.3	345.6 ± 2.2	347.1
C5	377.9 ± 0.5	376.2	379.3 ± 0.5	378.9
C6	369.5 ± 0.5	367.7	371.2 ± 0.5	370.8
enol cytosine				
N4−H_b	351.0 ± 2.1	349.8	351.9 ± 2.1	352.6
N4−H_a	355.2 ± 2.1	354.1	355.4 ± 2.1	356.1
OH	347.3 ± 2.1	341.6	347.2 ± 2.1	346.2
C5	385.1 ± 0.5	383.4	386.4 ± 0.5	386.0
C6	392.9 ± 0.5	391.1	393.8 ± 0.5	393.4

[a] GPA values were derived from isodesmic reactions. [b] GPA values were calculated directly from the enthalpies of neutral cytosine, its conjugate base, and a proton at 298 K.

3. Cytosine. For cytosine, the reference systems used in isodesmic reaction-based gas-phase acidity calculations are aniline with 298 K GPA of 366.4 ± 2.1 kcal/mol [11] (for the N4 site), benzene (for the C5 and C6 sites), pyrrole (for the N1 site), and phenol with 298 K GPA of 349.2 ± 2.1 kcal/mol [11] (for the hydroxyl group). The calculated 298 K gas-phase acidities of cytosine are listed in Table 3.

Similar to the situations of uracil and thymine, in the case of cytosine, both direct and isodesmic reaction-based calculations predict very similar gas-phase acidities. The only exception is the gas-phase acidity of the hydroxyl group of enol cytosine predicted at the B3LYP/6-31+G(d) level of theory. For this site, the difference between the results obtained from direct and isodesmic reaction-based calculations is larger than 5 kcal/mol. However, at the higher B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory, the difference is reduced to about 1 kcal/mol.

For the keto tautomer, the most acidic site is N1 (345.6 ± 2.2 kcal/mol, B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory) although GPA(N4−H_b) is only about 3 kcal/mol higher. GPA(N4−H_a) is about 5 kcal/mol higher than that of N4−H_b due to the repulsion between the electron lone pairs centered at N3 and at N4. C6 is about 8 kcal/mol more acidic than C5, and the latter is the most basic site (379.3 ± 0.5 kcal/mol, B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory). All our calculated GPA values for cytosine are higher than those calculated by Beauchamp at the AM1 level of theory.[6] Similar to the situation of thymine, the excellent agreement between the experimental GPA data of *N*-methylformamide and the corresponding values calculated at the B3LYP/6-31+G(d) and B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) levels of theory supports the reliability of our calculations on GPA(N1) of keto cytosine. For the *trans*-enol cytosine, the most acidic site is the hydroxyl group with a GPA of 347.2 ± 2.1 kcal/ mol (B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory), followed by N4−H_b and N4−H_a (351.9 ± 2.1 kcal/mol for N4−H_b and 355.4 ± 2.1 kcal/mol for N4−H_a, B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) calculations). Different from the situation of the keto tautomer, our calculations show that GPA(C6) is larger than that of C5 and that C6 is the least acidic site of the *trans*-enol cytosine with GPA of 393.8 ± 0.5 kcal/mol (B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory). In $C6^-$ anion, the doubly occupied σ orbital centered at C6 can interact with the lone pair orbital of N1. Such a four-electron interaction is repulsive, which leads to the destabilization of the $C6^-$ anion. This type of an interaction is not present for the keto tautomer of cytosine. In addition, it is worthwhile to note that for a given deprotonation site, the gas-phase acidity of the *trans*-enol cytosine is always greater than that of the keto cytosine. The reason for this phenomenon is unclear. In general, on the basis of our calculations, cytosine is less acidic than both uracil and thymine.

Conclusions

1. In this work, the 298 K gas-phase acidities of all possible deprotonation sites of the most stable tautomers of uracil, thymine, and cytosine were theoretically estimated by using both direct and isodesmic reaction-based calculation methodologies. The results show that, in these gas-phase acidity calculations, the GPA estimated by direct calculations is similar to that derived from isodesmic reactions. The B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) appears to be a reliable level for the calculation of gas-phase acidities of nucleobases, while the B3LYP/6-31+G(d) level gives similar results in all but one case.

2. For uracil and thymine, the most acidic site is N1, with a GPA of about 12 and 11 kcal/mol lower, respectively (B3LYP/aug-cc-pVTZ//B3LYP/6-31+G(d) level of theory), than the GPA of N3, the second most acidic site. This result is consistent with the experimental results obtained for uracil by Lee *et al*.[8] The C5, C5−CH_3, and C6 sites, on the contrary, are not acidic (C5 and C5−CH_3 are the least acidic site). As pointed out by Lee *et al*.,[8] even if the deprotonation can occur at these two sites, the corresponding anion may readily isomerize via the interaction with another neutral nucleobase molecule. Therefore, it might be difficult to observe the deprotonation of these two sites via experiments.

3. For the keto cytosine, the most acidic site is N1, followed by N4−H_b and N4−H_a. C5 and C6 are less acidic, with the GPAs being about 34 and 26 kcal/mol higher than that of N1, respectively. For the *trans*-enol cytosine, the hydroxyl group is the most acidic site, followed by N4−H_b and N4−H_a. C5 and C6 are less acidic. Different from the situations of uracil, thymine, and the keto cytosine, C6 is the least acidic site for the *trans*-enol cytosine. The possible reason is the repulsive interaction between the doubly occupied C6 σ orbital and the N1 lone pair orbital in the $C6^-$ anion.

4. Among the three pyrimidine-based nucleobases studied, uracil and thymine are the most acidic bases with the minimum GPA of 333−335 kcal/mol. The keto and *trans*-enol cytosine are less acidic with the minimum GPA 346−347 kcal/mol.

Acknowledgment. This work was partially supported by the National Institutes of Health (NIH).

References and Notes

(1) Leś, A.; Adamowicz, L.; Bartlett, R. J. *J. Phys. Chem.* **1989**, *93*, 4001−4005.

(2) Gould, I. R.; Burton, N. A.; Hall, R. J.; Hillier, I. H. *J. Mol. Struct. (THEOCHEM)* **1995**, *331*, 147−154.

(3) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811−6821.

(4) Civcir, P. Ü. *J. Mol. Struct. (THEOCHEM)* **2000**, *532*, 157−169, and references therein.

(5) Chen, E. C. M.; Chen, E. S. *J. Phys. Chem. B* **2000**, *104*, 7835−7844.

(6) Rodgers, M. T.; Campbell, S.; Marzluff, E. M.; Beauchamp, J. L. *Int. J. Mass Spectrom. Ion Processes* **1994**, 121−149, and references therein.

(7) Nguyen, M. T.; Chandra, A. K.; Zeegers-Huyskens, T. *J. Chem. Soc., Faraday Trans*. **1998**, *94* (9), 1277−1280.

(8) Kurinovich, M. A.; Lee, J. K. *J. Am. Chem. Soc*. **2000**, *122*, 6258−6262.

(9) Feng, W. Y.; Austin, T. J.; Chew, F.; Gronert, S.; Wu, W. *Biochemistry* **2000**, *39*, 1778−1783.

(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian98*; Gaussian, Inc.: Pittsburgh, PA, 1998.

(11) Linstrom, P. J., Mallard, W. G., Eds. *NIST Chemistry WebBook;* NIST Standard Reference Database Number 69, July; National Institute of Standards and Technology: Gaithersburg MD, 2001 (http://webbook.nist.gov).