pKa Values of Guanine in Water: Density Functional Theory Calculations Combined with Poisson−Boltzmann Continuum−Solvation Model

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · DECEMBER 2002

Impact Factor: 3.3 · DOI: 10.1021/jp020774x

6 AUTHORS, INCLUDING:

William A. Goddard
California Institute of Technology
1,333 PUBLICATIONS **68,233** CITATIONS
SEE PROFILE

Lawrence C Sowers
Loma Linda University
53 PUBLICATIONS **1,915** CITATIONS
SEE PROFILE

pK_a Values of Guanine in Water: Density Functional Theory Calculations Combined with Poisson−Boltzmann Continuum−Solvation Model

Yun Hee Jang,[†,‡] William A. Goddard III,[*,†] Katherine T. Noyes,[§] Lawrence C. Sowers,[§] Sungu Hwang,[‡,||] and Doo Soo Chung[*,‡]

Materials and Process Simulation Center (139-74), California Institute of Technology, Pasadena, California 91125, School of Chemistry, Seoul National University, Seoul 151-747, Korea, and Department of Biochemistry and Microbiology, Loma Linda University School of Medicine, Loma Linda, California 92350

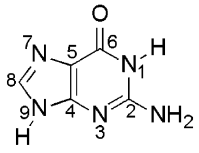
Received: March 20, 2002; In Final Form: September 27, 2002

It has long been postulated that rare tautomeric or ionized forms of DNA bases may play a role in mispair formation. To investigate the role this phenomenon plays in the mispairing of guanine and to develop a calculation methodology that can be extended to mutagenic DNA damage products, we used first principles quantum mechanics (density functional theory (B3LYP) with the Poisson−Boltzmann continuum−solvation model) to calculate the relative stabilities of tautomers of guanine in various environments and their pK_a values in aqueous solution. This method allows us to calculate site specific pK_a values − information that is experimentally inaccessible − as well as overall pK_a values for each stage, wherein our numbers are in agreement with experimental values. We find that neutral guanine exists in aqueous phase as a mixture of two major keto tautomers, the N_9H form (**1**) and a N_7H form (**3**). These keto forms are also major species present in the gas phase, as well as the O_6H enol tautomer (**7a**). These results show that tautomeric configurations can be drastically different depending on the environment. Here, we discuss the reasons for this environmental variability and suggest some possible implications. Finally, we estimate that the relative population of deprotonated guanine is 0.2−2% in the range of pH 7−8, a significant enough population to potentially play a role in mispair formation.

1. Introduction

Guanine (**1**) is a purine base found in the nucleic acids of all living organisms. Under physiological conditions, guanine exists predominantly in the neutral, keto tautomeric form. It has long been postulated that the presence of unpreferred or rare tautomeric forms might be involved in base mispair formation during polymerase-mediated DNA replication, resulting in genetic mutations. However, it has also been estimated that these unpreferred tautomeric forms might be present, under physiological conditions, at a low frequency of 10^{-6} to 10^{-5}.[1] Alternatively, hydrogen-bonding interactions and base pair formation could be perturbed by base ionization.[2] The ionization constants (pK_a values) of the normal DNA bases are roughly 2−3 pH units from physiological pH, predicting that the ionized forms might exist several orders of magnitude more frequently than the rare tautomeric forms. Despite the vast amount of work that has been done on these bases, the mechanisms of mispair formation during polymerase-mediated DNA replication are still unresolved.

Various theoretical studies have been conducted on the tautomerism of neutral guanine,[3−7] although little has been reported for charged species − protonated or deprotonated guanine − in the aqueous phase. Previous quantum mechanics (QM) calculations [B3LYP/6-31+G(d,p) level] for charged guanine have limited their focus to the monohydrated complex in the gas phase.[7] While protonation at N7 has been studied in the aqueous phase, the pK_a value was not calculated.[3]



1 (guanine)

There has been a great deal of theoretical work done to calculate acidic properties of numerous organic compounds, but most efforts have focused on gas phase basicities, proton affinities, or relative pK_a values in aqueous solution.[8,9] It was not until recently that the effort has been made to calculate absolute pK_a values in aqueous solution accurately.[10−15] Various levels of theory have been employed in these studies. Recently, Liptak and Shields[14] used the CPCM continuum−solvation approach with the complete basis set and Gaussian-*n* model to calculate the absolute pK_a values of six carboxylic acids within half a pK_a unit of experimental values. Nascimento and co-workers[11,13] used the PCM continuum−solvation model at the level of HF/6-31+G** and calculated the absolute pK_a values of carboxylic acids, aliphatic alcohols, thiols, and halogenated carboxylic aliphatic acids. Topol and co-workers[10,12] used the SCRF solvation model at the level of B3LYP/6-311+G(d,p), and the calculated absolute pK_a values were within 0.8 pK_a units of error for substituted imidazoles and within 1−2 pK_a units of error for weak organic acids such as hydrocarbons. Kallies and

* To whom correspondence should be addressed. E-mail: wag@wag.caltech.edu (W.A.G.). E-mail: dschung@snu.ac.kr (D.S.C.).
† California Institute of Technology.
‡ Seoul National University.
§ Loma Linda University School of Medicine.
|| Current address: Division of Food Science and Environmental Engineering, Miryang National University, Miryang, Gyeongnam 627-702, Korea.

10.1021/jp020774x CCC: $25.00
Published on Web 12/05/2002

Mitzner[16] used the SCI-PCM continuum−solvation approach at the levels of B3LYP/6-31G* and B3LYP/aug-cc-pVDZ//B3LYP/6-31G* and calculated the absolute pK_a values of aliphatic, alicyclic, and aromatic amines within 0.7 pK_a units of experimental values.

Recently, we have used first principles QM methods (density functional theory (DFT), B3LYP, in combination with the Poisson−Boltzmann (PB) continuum−solvation model) to calculate pK_a values for DNA bases in aqueous solution.[15,17] The agreement between calculated and experimental values for a series of pyrimidine derivatives is excellent (within 0.7 pK_a units). We further demonstrated that these methods can be used to predict the site of ionization where multiple, potential sites exist. These studies suggested to us that ionization and tautomerization might be linked, in that ionization might provoke a shift to an alternative tautomeric form.

As a first attempt to investigate how the proton configuration of a DNA base might be altered by ionization in aqueous solution, we have conducted a computational study of the ionization sites of guanine as well as the potential tautomers of the various ionized forms. In aqueous solution, the preferred keto tautomeric form of guanine strongly predominates. However, in the gas phase, some unusual tautomeric forms are indicated, consistent with experimental data that have demonstrated the existence of rare tautomers in the gas phase[18−20] but not in aqueous solution. Our studies indicate that the preferred site of protonation of guanine is N7 and that the primary site of deprotonation is N1. The calculated pK_a values are 3−4 and 9−10, respectively, consistent with reported experimental data.[21−27] The experimentally determined pK_a values represent a composite pK_a, representing a weighted average of the pK_a values of the various ionization sites and tautomers. We propose a scheme in which such composite pK_a values can be calculated.

This method can extend to various damaged DNA bases such as 8-oxoguanine (8-oxoG),[28−30] potentially explaining why certain DNA modifications increase base mispair formation and mutagenesis[31−34] as well as giving insight into possible repair mechanisms.[35−38] One difficulty in the study of many bases such as 8-oxoG is that they are often extremely insoluble in water.[21,39,40] Our approach could be powerful in that it can be applied to obtain pK_a values and tautomeric behaviors in solution even when the experimental determination is extremely difficult or almost impossible. Another important application of our theory would be to predict the impact of various postulated modifications on DNA bases or new designs prior to synthesis.

2. Calculation Details

2.1. pK_a Calculations. The pK_a of an acid HA is given by[10,14,41]

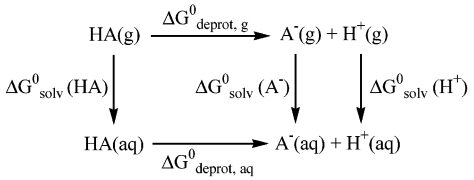
$$\mathrm{p}K_a = \frac{1}{2.303RT}\Delta G^\circ_{\mathrm{deprot,aq}} \quad (1)$$

where R is the gas constant and T is the temperature. The standard free energy of deprotonation of HA in water, $\Delta G^\circ_{\mathrm{deprot,aq}}$, is defined as (Scheme 1)

$$\Delta G^\circ_{\mathrm{deprot,aq}} = \Delta G^\circ(\mathrm{A^-(aq)}) + \Delta G^\circ(\mathrm{H^+(aq)}) - \Delta G^\circ(\mathrm{HA(aq)}) \quad (2)$$

The standard free energy of each species (HA, A^-, and H^+) in water (ΔG°_{aq}; with respect to the reference state where all of the nuclei and electrons are completely separated from one another at 0 K in the gas phase) can be written by the sum of the gas phase standard free energy ΔG°_g and the standard free energy of solvation in water ΔG°_{solv} (Scheme 1):

SCHEME 1: Thermodynamic Cycle[a] Used in the Calculation of pK_a[10,14,16,41]



[a] The proton (H^+) is not an isolated H^+ but a simple representation of H_3O^+ or $[H(H_2O)_n]^+$. An alternative thermodynamic cycle employing an H_2O explicitly on each side (that is, H_2O on the left and H_3O^+ in place of H^+ on the right) has also been used by Da Silva[11,13] and Schüürmann.[8]

$$\Delta G^\circ(\mathrm{HA(aq)}) = \Delta G^\circ(\mathrm{HA(g)}) + \Delta G^\circ_{\mathrm{solv}}(\mathrm{HA}) \quad (3a)$$

$$\Delta G^\circ(\mathrm{A^-(aq)}) = \Delta G^\circ(\mathrm{A^-(g)}) + \Delta G^\circ_{\mathrm{solv}}(\mathrm{A^-}) \quad (3b)$$

$$\Delta G^\circ(\mathrm{H^+(aq)}) = \Delta G^\circ(\mathrm{H^+(g)}) + \Delta G^\circ_{\mathrm{solv}}(\mathrm{H^+}) \quad (3c)$$

2.2. Gas Phase Free Energies. The standard Gibbs free energy of each species in the gas phase at its standard state (ideal gas at 1 atm and 298 K)[42,43] (ΔG° [HA(g)], ΔG° [A^-(g)], and ΔG° [H^+(g)]) is obtained by

$$\Delta G^\circ_g = E_{0\mathrm{K}} + \mathrm{ZPE} + \Delta\Delta G_{0\to 298\mathrm{K}} \quad (4)$$

The total energy of the molecule at 0 K (E_{0K}) is calculated at the optimum geometry from QM. The zero-point energy (ZPE) and the Gibbs free energy change from 0 to 298 K at 1 atm ($\Delta\Delta G_{0\to 298K}$) and are calculated from the vibrational frequencies calculated using QM. Translational and rotational free energy contribution is also calculated in the ideal gas approximation. We use ΔG° [H^+(g)] = $2.5RT - T\Delta S^\circ = 1.48 - 7.76 = -6.28$ kcal/mol at 298 K and 1 atm, which was taken from the literature.[10,12,41]

2.3. Gas Phase QM Calculations. All QM calculations used the *Jaguar v4.0* quantum chemistry software.[44,45] To calculate the geometries and energies of the various molecules, we used DFT, B3LYP, which includes the generalized gradient approximation and a component of the exact Hartree−Fock (HF) exchange,[46−50] as used extensively in other pK_a calculation studies.[10,12,15,16] In a very extensive study of Colominas and coworkers[3] on the level dependence of the tautomerism of neutral and protonated guanine, this B3LYP level of theory combined with the 6-311++G(d,p) or 6-31G(d) basis set has given quite similar results as the highest level of theory they tested [MP4/6-31++G(d,p)]. Because the calculation of vibration frequencies is generally quite time-consuming, the calculation was carried out in two steps. The 6-31G** basis set was first used to calculate the optimized geometry and vibration frequencies, and then, the 6-31++G** basis set, which includes diffuse functions, was used for the final geometry optimization started from the 6-31G** geometry (Appendix A shows how the results depend on the choice of basis set). The gas phase free energy is calculated as

$$\Delta G^\circ_g = \mathrm{ZPE}^{6-31\mathrm{G}^{**}} + \Delta\Delta G^{6-31\mathrm{G}^{**}}_{0\to 298\mathrm{K}} + E^{6-31++\mathrm{G}^{**}}_{0\mathrm{K,g}} \quad (5)$$

2.4. Solvation Free Energies. The standard free energy of solvation of HA and A^- in water [ΔG°_{solv} (HA) and ΔG°_{solv} (A^-)] at their standard states (1 M ideal solution at 298 K)[42,43] was calculated using the continuum−solvation approach by

numerically solving the PB equation.[51−55] 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[56]). The solute/solvent boundary is described by the surface of closest approach as a sphere of radius 1.4 Å (probe radius for water) and 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 (ESP) calculated from the wavefunction, that is, the ESP-fitted charges. In this formulation, the entire solute charge density is placed inside the solute cavity in the form of ESP charges. While this is only an approximate description without a rigorous justification, it is remarkably accurate, and it avoids some problems due to the penetration of the solute electronic cloud beyond the solute cavity.[10]

The solvation process, in which a solute is transferred from a vacuum into a solvent, was depicted hypothetically as two successive steps: (i) the creation of a cavity of the size of the solute in solvent and then (ii) the charging of the solute to turn on the electrostatic interaction with the solvent. Therefore, the free energy of solvation (ΔG_{solv}) was divided into two contributions, nonelectrostatic (nonpolar) term (ΔG_{np}) and electrostatic (polar) term (ΔG_{elec}):

$$\Delta G_{solv} = \Delta G_{np} + \Delta G_{elec} \tag{6}$$

The nonpolar contribution ΔG_{np} includes all of the nonelectrostatic contributions such as the energy cost of the cavity creation as well as the entropy change accompanied by the transfer of the solute from a vacuum into the solvent. It is simply treated to depend linearly on the contact area between the solute and the solvent, that is, the solvent accessible surface area (SA) of the solute in the solution phase:

$$\Delta G_{np} = \gamma \cdot (\mathrm{SA}) + b \tag{7}$$

The relationship had been determined for an aqueous solution by Tannor and co-workers,[52] and we used that relationship in our study without modification.

The electrostatic contribution to the solvation energy (ΔG_{elec}) is determined as follows. A gas phase calculation is carried out first to obtain the ESP charges (CHELP method).[57−59] 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 Hamiltonian 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 (to 0.1 kcal/mol in the solvation energy).

This solvation free energy calculation used *Jaguar v4.0*[44,45] at the B3LYP/6-31++G** level (Appendix B shows how the results depend on the choice of basis set), and the geometry was reoptimized in solution. The standard free energy of each species in water is calculated as a sum

$$\Delta G^\circ_{aq} = \Delta G^\circ_g + \Delta G^{\circ\,6-31++G^{**}}_{solv} \tag{8}$$

2.5. Correction to the Solvation Free Energy Due to the Change in the Reference State. The standard state of a solute in the gas phase is defined as an ideal gas at 1 atm, while the standard state of a solute in solution is defined as an ideal 1 M solution.[42,43,60] Our QM calculation of a solute in the gas phase (ΔG°_g; Sections 2.2 and 2.3) assumes an ideal gas at 1 atm. In order for eqs 3 and 8 to give the solution phase free energy (ΔG°_{aq}) consistently, the free energy of solvation (ΔG°_{solv}) should be defined as the free energy change during the standard solvation process, that is, a transfer of a solute from its gas phase standard state (ideal gas at 1 atm) into its solution phase standard state (1 M ideal solution). However, most of the tabulations of experimental solvation free energies follow a different convention, where the solvation process is defined as a transfer of a solute from its 1 M gas phase state into its 1 M solution phase state.[61−66] Because many continuum−solvation models have been developed to reproduce those experimental values, it is most likely that the solvation free energy calculated from there (ΔG^*_{solv}) should be defined in the same way. This is also the case for the PB solvation model used in this study,[52] where the dependence of the nonpolar solvation free energy (ΔG_{np}) on the SA of solute (eq 7) was determined based on the experimental solvation energies of alkanes.[61] Thus, the estimation of the solution phase free energy (ΔG°_{aq}) by combining the calculated gas and solution phase values ($\Delta G^\circ_g + \Delta G^*_{solv}$) needs a correction term corresponding to the free energy change accompanied by the reversible state change of 1 mol gas from 1 atm (24.47 L mol^{-1}) to 1 M (1 mol L^{-1}):[43]

$$\begin{aligned}\Delta G^\circ_{aq,1M} &= \Delta G^\circ_{g,1atm} + \Delta G^\circ_{solv,1atm(g)\rightarrow 1M(aq)} \\ &= \Delta G^\circ_{g,1atm} + \Delta G^*_{solv,1M(g)\rightarrow 1M(aq)} + \Delta G^{corr}_{1atm(g)\rightarrow 1M(g)} \\ &= \Delta G^\circ_{g,1atm} + \Delta G^*_{solv,1M(g)\rightarrow 1M(aq)} + RT\ln(1/22.47) \\ &= \Delta G^\circ_{g,1atm} + \Delta G^*_{solv,1M(g)\rightarrow 1M(aq)} - 1.89\ (\mathrm{kcal/mol})\end{aligned} \tag{9}$$

That is, a correction of −1.89 kcal/mol should be made to the calculated solvation free energy (ΔG^*_{solv}) to give the standard free energy of solvation (ΔG°_{solv}), as also discussed thoroughly by Abraham.[67] The correction is constant for all solutes.

However, in our current scheme of pK_a calculation (Scheme 1 and eqs 2−5), where the solution phase free energies are calculated only for HA and A^- and that of the proton is determined parametrically from the fitting to the experimental values (Section 2.6), the corrections for HA and for A^- would cancel each other and thus would not alter any final results, neither the final pK_a values nor the final parameters of the solvation free energy of proton, $\Delta G^\circ_{solv}(H^+)$.

2.6. Parameters for Solvation Free Energy Calculation. Several parameters are used in the solvation free energy calculation. These parameters were first taken from the literature and then slightly modified in order to reproduce experimental pK_a values of guanine (see Appendix B for details).

(i) The atomic radii used to build the vdW envelope of the solute were taken from Marten and co-workers[54] and then reduced by 6%. The final radii are (the original literature radii are in parentheses): 1.88 (2.00) Å for sp^2-hybridized carbon, 1.45_7 (1.55) Å for sp^2-hybridized oxygen, 1.41 (1.5) Å for sp^2-hybridized nitrogen, 1.17_5 (1.25) Å for hydrogen attached to sp^2-hybridized carbon, and 1.08_1 (1.15) Å for other types of hydrogen.

(ii) To obtain the pK_a requires the standard free energy of solvation of a proton in water, $\Delta G^\circ_{solv}(H^+)$. Despite considerable experimental and theoretical research, this value remains uncertain.[14,68,69] Experimental measurements of the standard hydrogen potential led to a wide range of $\Delta G^\circ_{solv}(H^+)$ from −254 to −261 kcal/mol.[41,70] Recent studies on the solvation of various ion−water clusters led to $\Delta G^\circ_{solv}(H^+) = -262 \pm 1$ kcal/mol (theoretical)[69] and $\Delta G^\circ_{solv}(H^+) = -263.98 \pm 0.07$

kcal/mol (experimental).[71] As shown in the previous section, the theoretical value (-262 ± 1 kcal/mol)[69] should be corrected by -1.89 kcal/mol to -264 ± 1 kcal/mol to be comparable to the experimental values, and indeed, this is very close to the experimental value of -263.98 ± 0.07 kcal/mol.[71] In our current study, the proton solvation free energy $\Delta G^\circ_{solv}(H^+)$ was treated as a parameter chosen to give the best match between the calculated and experimental guanine pK_a values, and the selected value was -263.47 kcal/mol, in good agreement with the literature values shown above (-264 ± 1 (after correction)[69] and -263.98 ± 0.07 kcal/mol[71]), illustrating the quality of our pK_a calculation scheme. Liptak and Shields have also concluded that the correct value of $\Delta G^\circ_{solv}(H^+)$ must be in the range of -264 kcal/mol, from their pK_a calculations on six carboxylic acids using the complete basis set and Gaussian-*n* models combined with the CPCM continuum−solvation method.[14]

2.7. Summary: Calculation Steps. Our procedure involved three steps: (i) B3LYP/6-31G**(g) preliminary geometry optimization and frequency calculation, (ii) B3LYP/6-31++G**(g) for final geometry optimization, and (iii) B3LYP/6-31++G**(aq) for solution phase geometry optimization.

The final standard free energy of each species in water is expressed as

$$\Delta G^\circ_{aq} = ZPE^{6-31G^{**}} + \Delta\Delta G^{6-31G^{**}}_{0\rightarrow 298K} + E^{6-31++G^{**}}_{0K,g} + \Delta G^{\circ\,6-31++G^{**}}_{solv} \quad (10)$$

This final calculation scheme is different from those used in our previous study on substituted uracils.[15] A larger basis set [Dunning cc-pVTZ(-f)++//cc-pVTZ(-f)] had been used for these smaller pyrimidines, leading to slightly different parameters for solvation calculations [atomic radii scale factor = 0.89; $\Delta G^\circ_{solv}(H^+) = -262.65$ kcal/mol].

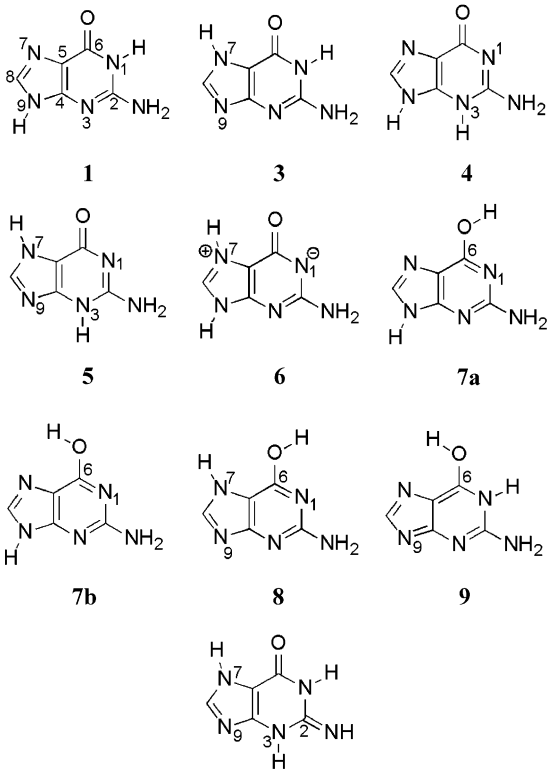
In addition to guanine, this modified method discussed above has now been applied to several systems (cytosine, 9-methylisoguanine, and isoguanine) leading to results in excellent agreement with experiment (within 1 p*K* unit), as summarized in Appendix D. This suggests that these methods can be valuable in prediction of the pK_a values for unknown systems.

3. Results and Discussion

Several tautomers of guanine might be present simultaneously at each ionization state, and the presence of these multiple tautomers complicates the calculation of the pK_a and gas phase proton affinity (PA). Thus, we calculated the free energies of all of the plausible tautomers for neutral, anionic, and cationic guanines in gas and aqueous phases. Then, we estimated the proportions of tautomers at each ionization state assuming a Boltzmann distribution at 298 K (Sections 3.1−3.4), prior to the calculations of gas phase PAs and basicities (Section 3.5) and aqueous phase pK_a values (Sections 3.6−3.8). Biological implications are discussed in Sections 3.9 and 3.10.

As seen below, there are several low-energy tautomers whose free energies were found to lie within 1 kcal/mol of each other: **1**, **3**, and **7a** (neutral; gas); **1** and **3** (neutral; aqueous); **12**, **13**, and **15** (deprotonated; gas); and **11**−**13** (deprotonated; aqueous) (Tables 1 and 6 below). Note that we do not claim to calculate relative energies within 1 kcal/mol. For example, we find that in the gas phase **1** is 0.3 kcal/mol higher than **3** and 0.9 kcal/mol lower than **7a** while in the solution phase **7a** is 8.7 and 7.7 kcal/mol higher than **1** and **3**, respectively. We do not imply that in the gas phase **1** is less stable than **3** and more stable than **7a**. We emphasize that all three tautomers (**1**, **3**, and **7a**) will be present in the gas phase but in the aqueous phase **7a** will not be present. Such estimations are helpful in showing that tautomeric states may shift in different solvents/environments and may have implications for biological microenvironments (such as the active site of an enzyme), where the surroundings are not strictly aqueous.

SCHEME 2: Tautomers of Neutral Guanine Considered in This Study



3.1. Tautomers of Neutral Guanine. *3.1.1. Relative Energies of Neutral Guanine Tautomers.* Tautomers of neutral guanine considered in this study (**1** and **3**−**10**) are shown in Scheme 2, and their relative free energies in gas and aqueous phases are given in Table 1. The relative population of each tautomer in equilibrium was estimated from the Boltzmann distribution at 298 K.

In the gas phase, we predict that the 6-keto tautomer **3** (with protons on N1 and N7) is the most stable, but it is only 0.3 kcal/mol free energy below 6-keto **1** (with protons on N1 and N9). In addition, the 6-enol tautomer **7** (with protons on N9 and O6) is just 0.9 kcal/mol higher. Thus, gas phase guanine at room temperature would be 53% of **3**, 32% of **1**, and 15% of **7**. Other tautomers would have populations below 0.1%.

In the aqueous phase, we predict that the 6-keto tautomer **1** is the most stable, but **3** is only 1.0 kcal/mol higher. However, the best 6-enol tautomer (**7**) is 8.7 kcal/mol higher than **1**. Thus, in the aqueous phase, guanine would be 85% of **1** and 15% of **3** with 0.6% of 6-keto **5** (with protons at N3 and N7).

3.1.2. Solvation Effects. Solvation leads to a significant change in the relative tautomeric stability. To explain this dramatic change, several energetic effects must be considered. One possibility is that differences in the dipole moment between the keto and the enol forms could cause the keto form to be greatly stabilized in a solvent with a high dielectric constant. For example, **1** is probably stabilized to a greater extent than **3** in aqueous solution (by 1.3 kcal/mol in ΔG_{solv}; Table 2), because of the higher dipole moment of **1** (6.81 D) than that of **3** (1.88

TABLE 1: Relative Free Energies (kcal/mol) of Various Tautomers of Neutral Guanine and Their Relative Populations Based on Boltzmann Distribution: (a) Gas Phase and (b) Aqueous Phase

	1 keto amine	**3** keto amine	**4** keto amine	**5** keto amine	**6** keto amine	**7a** enol amine	**7b** enol amine	**8** enol amine	**9** enol amine	**10** keto imine
(a) Gas										
$\Delta G^{\circ}_{g,rel}$[a]	0.3	0.0	18.9	5.9	18.9	0.9	1.5	3.6	21.1	5.7
population	0.32	0.53	7×10^{-15}	3×10^{-5}	8×10^{-15}	0.11	0.04	1×10^{-3}	2×10^{-16}	3×10^{-5}
(b) Aqueous										
$\Delta G^{\circ}_{aq,rel}$[b]	0.0	1.0	5.2	3.0	5.1	8.7	9.7	9.9	17.9	9.8
population	0.85	0.15	1×10^{-4}	6×10^{-3}	2×10^{-4}	4×10^{-7}	6×10^{-8}	5×10^{-8}	4×10^{-14}	6×10^{-8}

[a] Relative free energies with respect to ΔG°_{g} (**3**). [b] Relative free energies with respect to ΔG°_{aq} (**1**).

TABLE 2: Free Energies of Solvation (ΔG_{solv}, kcal/mol) of Tautomers of Neutral Guanine

	1 keto amine	**3** keto amine	**4** keto amine	**5** keto amine	**6** keto amine	**7a** enol amine	**8** enol amine	**9** enol amine	**10** keto amine
ΔG_{solv}	−35.4	−34.1	−48.8	−38.0	−48.9	−27.3	−28.9	−38.3	−31.1
μ_g[a]	6.81	1.88	11.16	4.94	9.08	3.25	3.96	7.07	2.75
$\angle(NH_2)_{g/aq}$[b]	36/23	39/27	41/19	36/21	22/24	24/23	26/26	44/26	
$\angle(NH)_{g/aq}$[c]	3/0 (1)	4/0 (1)	22/0 (3)	8/0 (3)				6/1 (1)	0/0 (1,3)
RMS Δ[d]	0.046	0.052	0.118	0.064	0.040	0.023	0.030	0.084	0.038

[a] Gas-phase dipole moment (Debye or D). [b] Inversion angle of the exocyclic 2-amino (NH_2) group (°) defined by the angle of the N2−H_1 bond from the C2−N2−H_2 plane (gas-phase value/solution-phase value). [c] Angle of the N1−H or N3−H bond from the ring plane (°) (gas-phase value/solution-phase value). Whether the bond is N1−H or N3−H is indicated in parentheses. [d] RMS deviations between atomic coordinates (Å) between the gas-phase structure and the aqueous-phase structure.

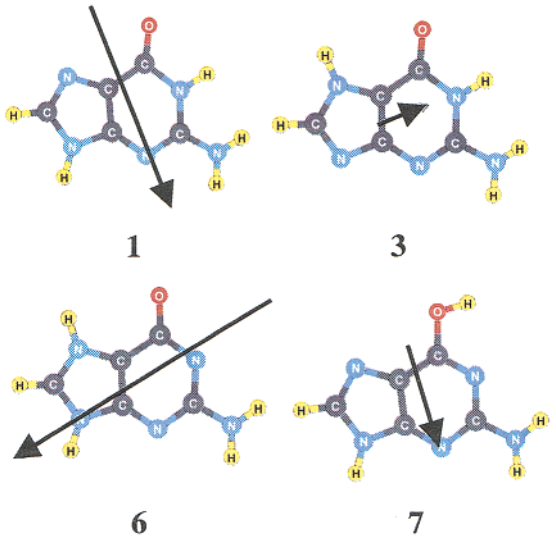


Figure 1. Dipole moment vectors of several tautomers of neutral guanine.

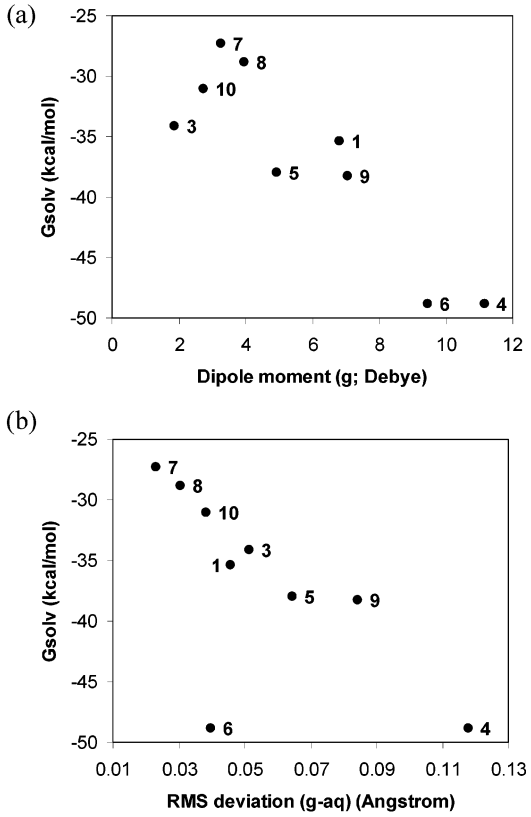


Figure 2. Correlations between free energies of solvation (ΔG_{solv}, kcal/mol) of various tautomers of neutral guanine and (a) their gas phase dipole moments (μ, Debye) or (b) RMS deviations between atomic coordinates (RMS Δ, Å) between gas phase structures and aqueous phase structures.

D) (Table 2 and Figure 1). However, the dipole moment does not completely explain different extents of stabilization of various tautomers by solvation, as can be seen in the poor correlation between ΔG_{solv} and dipole moment (Figure 2a). For example, the enol tautomer **7** has a higher dipole moment (3.25 D; Table 2) than **3**, but the stabilization of **7** by solvation (ΔG_{solv} = −27.3 kcal/mol) is much smaller than that of **3** (ΔG_{solv} = −34.1 kcal/mol). This poor correlation implies that additional factors must account for the solvation effect, in addition to dipole moment differences.

Another possibility involves destabilization of certain tautomers in the gas phase due to unfavorable electrostatic interactions that can be shielded in the aqueous phase by solvent water. In keto tautomers such as **1** and **3**, one of the partially positive hydrogen atoms of the exocyclic 2-amino group is electrostatically repelled by the partially positive N1−H atom. As a result, the exocyclic 2-amino group (−NH_2) is not planar with respect to the aromatic base. The inversion angle of the 2-amino group (defined as the angle between the N2−H_1 bond and the C2−N2−H_2 plane) of **1** and **3** is as large as 36 and 39°, respectively. The corresponding angles for other tautomers in which the N_1H or N_3H protons are in close contact with the

TABLE 3: Properties of Two Representative Tautomers of Neutral Hypoxanthines

	1′ (keto)	**7a′** (enol)
$\Delta\Delta G_g$[a]	0	3.7
$\Delta\Delta G_{aq}$[b]	0	9.8
ΔG_{solv}[c]	−28.7	−22.6
μ_g[d]	5.4	2.6
$\angle(N_1H)_{g/aq}$[e]	0/0	
RMS Δ[f]	0.031	0.017

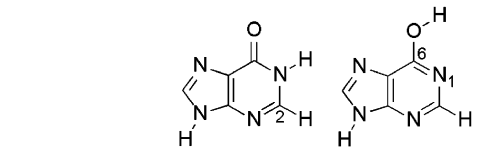
[a,b] Relative free energy (kcal/mol) in the gas phase and in aqueous solution with respect to the keto form **1′**. [c] Free energy of solvation (kcal/mol). [d] Gas-phase dipole moments (Debye or D). [e] Angle of the N1−H bond from the ring plane (°) (gas-phase value/solution-phase value). [f] RMS deviation between atomic coordinates (Å) between the gas-phase structure and the aqueous-phase structure.

2-amino group (such as **4**, **5**, and **9**) range from 36 to 44°. In tautomers lacking N1 or N3 protons, such as **7** and **8**, the corresponding angle is much smaller (24 and 26°, respectively). These observations indicate that the nonplanarity of the 2-amino group of guanine comes at least partly from electrostatic repulsion between the NH_2 protons and the neighboring protons (N_1H or N_3H), as also suggested by a previous study.[5]

This repulsion also causes significant displacement of the N_1H or N_3H protons out of the ring plane. Tautomer **4**, which is very unstable in the gas phase, has an N_1H inversion angle (defined as the angle of the N1−H bond with respect to the ring plane) of 23° in the gas phase but is planar within 1° in the aqueous phase, indicating that the intramolecular repulsion is lessened by a polar solvent such as water. This stabilization by the solvation effect would be larger for keto tautomers or others that have more intramolecular repulsion in the gas phase (such as **4**, **5**, and **9**) than enol tautomers (**7** and **8**), for which the repulsion is not significant in the gas phase (Figure 2b).

To confirm this argument on the relationship between the "intramolecular" repulsion caused by the 2-amino group and the solvation energy, we conducted similar calculations on the derivative hypoxanthine, which is essentially guanine without the 2-amino group. We calculated the solvation free energies of two representative tautomers of hypoxanthine: **1′** for a keto form corresponding to **1** of guanine and **7a′** for an enol form corresponding to **7a** of guanine (Table 3). Contrary to guanine, the keto form of hypoxanthine **1′** is perfectly planar. Its N1−H bond does not deviate from the base plane at all in either the gas phase the aqueous solution, and its RMS deviation between the gas and the solution phase structures is as small as 0.031 Å. Its solvation free energy was calculated to be −28.7 kcal/mol, which lies near the upper end in the range of the guanine solvation free energies. This solvation energy value indicates that solvation confers less of a stabilizing effect on hypoxanthine than it does for guanine, as expected from the plot in Figure 2b. Another expectation from the lack of the intramolecular repulsion between the N_1H and the exocyclic amino group in **1′** is that the keto form of hypoxanthine would be quite stable in the gas phase as compared to the enol form. Indeed, **1′** is 3.7 kcal/mol more stable than **7a′** in the gas phase, contrary to the guanine case where the free energies of **1** and **7a** were quite similar to each other (only 0.6 kcal/mol different from each other) in the gas phase. The results reported here are consistent with previous experimental and computational studies, which indicate that the keto form of hypoxanthine is more stable than the 6-enol form.[72−76]

The comparison of guanine and hypoxanthine allows us to determine what portion of the water solvation effect can be attributed to intramolecular N1 imino/N2 amino repulsion. The energy difference between the enol and the keto tautomeric



forms of guanine in the gas phase and in water is 8.1 kcal/mol. The corresponding value for hypoxanthine, which does not have a 2-amino group, is only 6.1 kcal/mol. The difference, 2.0 kcal/mol, can therefore be attributed to the intramolecular repulsion. The remaining 6.1 kcal/mol difference between the keto and the enol forms when comparing guanine in the gas phase and aqueous solution results primarily from the purine dipole−water interaction. The dipole moment for the 6-keto form of guanine is twice that of the enol tautomer as previously discussed.

We have shown here that some tautomeric forms of neutral guanine have dramatically different stabilities depending on the surrounding medium. These differences can be attributed to both dipole moment effects (which only play a role in molecular stability in the presence of a high-dielectric constant solvent) and intramolecular electrostatic repulsion effects (which only play a role in the gas phase, as they can be effectively shielded by a polar solvent). Our observations suggest that varying tautomeric populations of DNA bases could be stabilized in nonaqueous media, which could be similar to the environment provided by a hydrophobic or other nonsolvated pocket of an enzyme. This possibility would hold tremendous implications for the specific mechanisms of various DNA repair enzymes and would be an interesting avenue to pursue in future studies.

3.2. Deprotonated Guanine. Tautomers of guanine after the first deprotonation (**11**$^-$−**16**$^-$) and the second deprotonation (**17**$^{2-}$) considered in this study are shown in Scheme 3, and their relative free energies and populations in gas and aqueous phases are given in Table 4.

In the aqueous phase, deprotonation from N_1H or N_9H of **1** and **3** leads to three tautomers, **11**$^-$, **12**$^-$, and **13**$^-$, with free energies within 0.5 kcal/mol, while deprotonation from the exocyclic NH_2 group leads to an energy 7.6 kcal/mol higher (**14**$^-$). Deprotonated guanine would exist as a mixture of 52% **11**$^-$, 26% **12**$^-$, and 22% **13**$^-$. As in the neutral guanine, the enol tautomer **15**$^-$, which has a significant population in the gas phase (∼12%), is 9−10 kcal/mol higher in free energy than the keto tautomers (**11**$^-$, **12**$^-$, and **13**$^-$) in aqueous solution, indicating that the deprotonation of guanine does not shift the keto−enol tautomeric equilibrium to enol. The second deprotonation from these three tautomers results predominantly in one tautomer, **17**$^{2-}$.

SCHEME 3: Tautomers of Deprotonated Guanine (11$^-$−16$^-$) Considered in This Study and an Exclusive Tautomer after the Second Deprotonation (17^{2-})

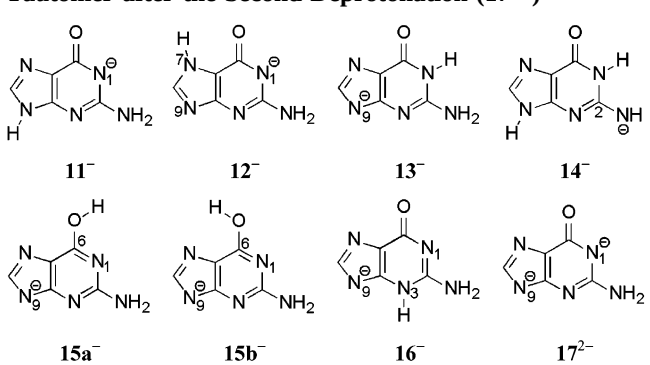
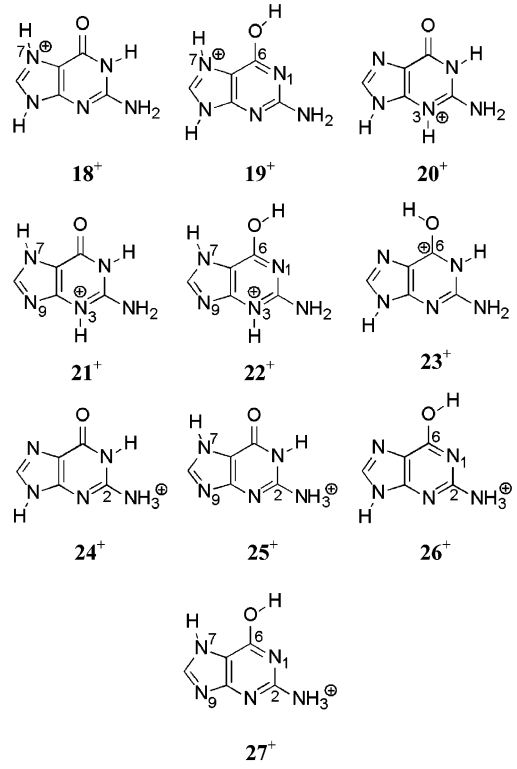


TABLE 4: Relative Free Energies (kcal/mol) of Tautomers of Deprotonated Guanine and Their Relative Populations Based on Boltzmann Distribution: (a) Gas Phase and (b) Aqueous Phase

	11$^-$ keto amine	**12**$^-$ keto amine	**13**$^-$ keto amine	**14**$^-$ keto amine	**15a**$^-$ enol amine	**15b**$^-$ enol amine	**16**$^-$ keto amine
(a) Gas							
$\Delta G^\circ_{g,rel}$[a]	2.6	0.0	1.0	2.4	2.7	1.0	7.1
population	0.01	0.72	0.12	0.01	7×10^{-3}	0.12	5×10^{-6}
(b) Aqueous							
$\Delta G^\circ_{aq,rel}$[b]	0.0	0.4	0.5	7.6	9.6	10.5	2.9
population	0.52	0.26	0.22	1×10^{-6}	5×10^{-8}	1×10^{-8}	4×10^{-3}

[a] Relative free energies with respect to ΔG°_g (**12**$^-$). [b] Relative free energies with respect to ΔG°_{aq} (**11**$^-$).

SCHEME 4: Tautomers of Protonated Guanine Considered in This Study



3.3. Protonated Guanine. Tautomers of protonated guanine considered in this study (**18**$^+$−**27**$^+$) are shown in Scheme 4, and their relative free energies and populations in gas and aqueous phases are given in Table 5. Both in the gas phase and in the aqueous phase, the tautomer protonated at N7 (**18**$^+$) is far more stable than any other tautomer. The enol tautomers **19**$^+$ and **22**$^+$ were 2.9 and 1.6 kcal/mol, respectively, higher in free energy than **18**$^+$ in the gas phase, but the energy difference became much higher (9.4 and 9.6 kcal/mol, respectively) in aqueous solution. This calculation indicates that like ionization, the protonation of guanine does not shift the keto−enol tautomeric equilibrium to enol.

TABLE 6: Relative Free Energies (kcal/mol) of Tautomers of Neutral and Protonated Guanine in the Gas Phase; Our Results vs the Results of Colominas and Coworkers[3]

$\Delta G^\circ_{g,rel}$	(a) neutral[a]					(b) protonated[b]				
this work	**1**	**3**	**7a**	**7b**	**8**	**18**$^+$	**19**$^+$	**21**$^+$	**22**$^+$	**23**$^+$
Colominas	19	17	96c	96t	76c	179	796c	137	376c	196t
this work	0.0	−0.3	0.6	1.2	3.3	0.0	2.9	4.8	1.4	5.6
Colominas: level D	0.0	0.2	1.1	1.8	4.4	0.0	3.6	3.5	1.8	5.1
Colominas: level F	0.0	−0.4	1.1	1.8	3.7	0.0	3.6	4.8	1.9	5.9

[a] Relative free energies with respect to ΔG°_g(**1**). [b] Relative free energies with respect to ΔG°_g (**18**$^+$).

3.4. Agreement with Other Studies. Our results for the tautomerism of neutral and protonated guanine are in agreement with both experimental and theoretical results. Experimentally it has been suggested that a matrix-isolated guanine in Ar or N_2 exists as a mixture of keto-amine (such as **1** and **3**) and enol-amine (such as **7**) tautomers, although the proposed relative population of the enol tautomer varies as follows: 78% for guanine and 86% for 9-methylguanine[18] or ∼50% for 9-methylguanine.[19,20] Among the keto-amine tautomers, **3** has been shown to be more stable than **1** for isolated guanine.[77] However, **1** is known to be the only tautomeric form found in polar solvents[78,79] or in the crystalline state,[80] which is understood in terms of the higher dipole moment of **1** than those of **3** and **7**.[18] Both in the isolated phase and in the polar media (such as dimethyl sulfoxide solution), it has been suggested that the amine tautomer would predominate in the amine−imine equilibrium.[20,81]

An extensive theoretical study by Colominas and co-workers[3] on the tautomers of neutral guanine [G19 (**1**), G17 (**3**), G96c (**7a**), G96t (**7b**), and G76c (**8**)] and on the level dependence of the calculation results in the gas phase [HF/6-31G(d)//HF/6-31G(d) (level A); HF/6-311++G(d,p)//HF/6-31G(d) (level B); MP2/6-311++G(d,p)//HF/6-31G(d) (level C); MP4/6-311++G(d,p)//MP2/6-31G(d) (the highest level D); B3LYP/6-31G(d)//MP2/6-31G(d) (level E); and B3LYP/6-311++G(d,p)//MP2/6-31G(d) (level F)]. Our results on tautomers of neutral guanine are in close agreement with those obtained with the highest level (level D) and with the B3LYP level (level F) in their work (Table 6a).

TABLE 5: Relative Free Energies (kcal/mol) of Tautomers of Protonated Guanine and Their Relative Populations Based on Boltzmann Distribution: (a) Gas Phase and (b) Aqueous Phase

	18$^+$	**19**$^+$	**20**$^+$	**21**$^+$	**22**$^+$	**23**$^+$	**24**$^+$	**25**$^+$	**26**$^+$	**27**$^+$
(a) Gas										
$\Delta G^\circ_{g,rel}$	0.0	2.9	16.8	4.8	1.4	5.6	39.3	35.4	23.2	22.4
population	0.91	7×10^{-3}	4×10^{-13}	3×10^{-4}	0.08	7×10^{-5}	1×10^{-29}	1×10^{-26}	9×10^{-18}	4×10^{-17}
(b) Aqueous										
$\Delta G^\circ_{aq,rel}$	0.0	9.4	2.8	1.8	9.6	11.8	10.7	10.7	13.5	13.8
population	0.95	1×10^{-7}	8×10^{-3}	0.05	9×10^{-8}	2×10^{-9}	1×10^{-8}	1×10^{-8}	1×10^{-10}	7×10^{-11}

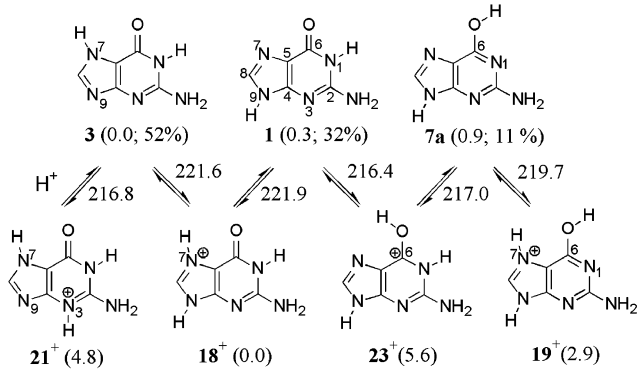
[a] Relative free energies with respect to ΔG°_g (**18**$^+$). [b] Relative free energies with respect to ΔG°_{aq} (**18**$^+$).

TABLE 7: Gas Phase Proton Affinities (PA) and Basicities (GB) of Guanine (kcal/mol)

	calculation	experiment
PA	229.3 (**1** → **18**$^+$; N7), 223.7 (**1** → **23**$^+$; O6)	
	229.0 (**3** → **18**$^+$; N9), 223.9 (**3** → **21**$^+$; N3)	229 ± 2[a]
	227.3 (**7** → **19**$^+$; N7), 224.1 (**7** → **23**$^+$; N1)	
GB	221.9 (**1** → **18**$^+$; N7), 216.4 (**1** → **23**$^+$; O6)	
	221.6 (**3** → **18**$^+$; N9), 216.8 (**3** → **21**$^+$; N3)	222 ± 2[a]
	219.7 (**7** → **19**$^+$; N7), 217.0 (**7** → **23**$^+$; N1)	

[a] Ref 82.

SCHEME 5: Series of Gas Phase Basicities (GB, kcal/mol) of Guanine[a]



[a] Relative free energies of tautomers at each ionization state are shown together.

Colominas and co-workers also examined the tautomers of protonated guanine [pG179 (**18**$^+$), pG796 (**19**$^+$), pG137 (**20**$^+$), pG376c (**22**$^+$), and pG196t (**23**$^+$)]. Again, our results on tautomers of protonated guanine are in close agreement with those obtained with the highest level (level D) and with the B3LYP level (level F) in their work (Table 6b).

3.5. Gas Phase PA and GB. The gas phase proton affinity (PA) and basicity (GB) of an acid HA are defined as enthalpy change and free energy change, respectively, during the protonation process in the gas phase:

$$\mathrm{PA} = \Delta H_g^\circ(\mathrm{A}^-) + \Delta H_g^\circ(\mathrm{H}^+) - \Delta H_g^\circ(\mathrm{HA}) \qquad (11)$$

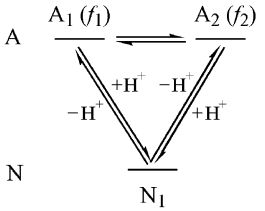
$$\mathrm{GB} = \Delta G_g^\circ(\mathrm{A}^-) + \Delta G_g^\circ(\mathrm{H}^+) - \Delta G_g^\circ(\mathrm{HA}) \qquad (12)$$

Here, $\Delta H_g^\circ(\mathrm{H}^+) = 2.5RT = 1.48$ kcal/mol and $\Delta G_g^\circ(\mathrm{H}^+) = 2.5RT - T\Delta S^\circ = 1.48 - 7.76 = -6.28$ kcal/mol at 1 atm and 298 K.[16]

For the neutral state, we find three tautomers with significant populations: **1** (32%), **3** (52%), and **7a** (11%). We expect that the most probable protonation on **1** is on N7, leading to **18**$^+$ with a GB of 221.9 kcal/mol. We find that the next best protonation is on O6 leading to **23**$^+$ with a much smaller GB of 216.4 kcal/mol. We expect that the most probable protonation on **3** is on N9, leading to the same cationic tautomer **18**$^+$ with a GB of 221.6 kcal/mol. We expect that the most probable protonation on **7a** is on N7 leading to **19**$^+$ with a GB of 219.7 kcal/mol.

These calculated GBs (221.6 for **3** → **18**$^+$ and 221.9 for **1** → **18**$^+$) agree well with experiment (222 ± 2).[82,83] Because several tautomers can be present simultaneously in the neutral and cationic states and the tautomerization between them in the gas phase can be slower than the measurement, it could be possible to examine experimentally the protonation from a specific neutral tautomer to a specific cationic tautomer as in Table 7 and Scheme 5.

SCHEME 6: Simple Model Case for Illustrating the Calculation of pK_a in the Presence of Multiple Tautomers



3.6. Calculation of pK_a in the Presence of Multiple Tautomers. In the aqueous phase, tautomerization by the solvent-mediated proton transfer is expected to be fast as compared to the time scale of the pK_a measurement, leading establishment of a rapid equilibrium between tautomers during the measurement. Assuming this rapid equilibrium, we calculate the pK_a values in the presence of multiple tautomers as follows.

First, we consider the simplest case, with only one tautomer at the protonated state (N) and two tautomers at the deprotonated state (A_1 and A_2; $[A] = [A_1] + [A_2]$) (Scheme 6). The free energy of deprotonation from N to A_1 and to A_2 is given by ΔG_1 and ΔG_2, respectively. The relative populations of the deprotonated tautomers, f_1 and f_2, are given by

$$f_1 = \frac{[\mathrm{A}_1]}{[\mathrm{A}]} = \frac{[\mathrm{A}_1]}{[\mathrm{A}_1] + [\mathrm{A}_2]} \qquad (13a)$$

$$f_2 = \frac{[\mathrm{A}_2]}{[\mathrm{A}]} = \frac{[\mathrm{A}_2]}{[\mathrm{A}_1] + [\mathrm{A}_2]} \qquad (13b)$$

where $f_1 + f_2 = 1$, $0 \le f_1 \le 1$, and $0 \le f_2 \le 1$. These populations are calculated from the Boltzmann distribution based on the relative free energies of those tautomers as done in Sections 3.1−3.3. The overall dissociation constant, K_a, is given as

$$K_a = \frac{[\mathrm{H}^+][\mathrm{A}]}{[\mathrm{N}]} \qquad (14)$$

The site specific dissociation constants, K_a^1 and K_a^2, can be calculated from the deprotonation free energies of the corresponding processes:

$$K_a^{\,1} = \frac{[\mathrm{H}^+][\mathrm{A}_1]}{[\mathrm{N}]} = \exp\left(-\frac{\Delta G_1}{RT}\right) \qquad (15a)$$

$$K_a^{\,2} = \frac{[\mathrm{H}^+][\mathrm{A}_2]}{[\mathrm{N}]} = \exp\left(-\frac{\Delta G_2}{RT}\right) \qquad (15b)$$

Equations 13−15 can be rewritten as

$$K_a^{\,1} = \frac{[\mathrm{H}^+][\mathrm{A}_1]}{[\mathrm{N}]} = \frac{[\mathrm{H}^+][\mathrm{A}]}{[\mathrm{N}]} \cdot f_1 = K_a \cdot f_1 \qquad (16a)$$

$$K_a^{\,2} = \frac{[\mathrm{H}^+][\mathrm{A}_2]}{[\mathrm{N}]} = \frac{[\mathrm{H}^+][\mathrm{A}]}{[\mathrm{N}]} \cdot f_2 = K_a \cdot f_2 \qquad (16b)$$

Thus, the overall K_a is calculated from a site specific K_a^i as

$$K_a = \frac{K_a^{\,1}}{f_1} = \frac{K_a^{\,2}}{f_2} \qquad (17)$$

That is, the overall pK_a value is calculated from a site specific pK_a value (pK_a^i) as

$$pK_a = pK_a^1 + \log f_1 = pK_a^2 + \log f_2 \quad (18)$$

where $\log f_1 \le 0$ and $\log f_2 \le 0$.

For the special case of a deprotonated state consisting of two degenerate tautomers [$\Delta G_1 = \Delta G_2$, $f_1 = f_2 = 0.5$], eq 18 reduces to

$$pK_a = pK_a^1 - \log 2 \quad (19)$$

Multiplying eq 19 by 2.303*RT* leads this to

$$\Delta G = \Delta G_1 - RT \ln 2 \quad (20)$$

as expected from a 2-fold degeneracy.[43]

For the case where both the protonated state and the deprotonated state consist of several tautomers with fractions f_i at the protonated state and f'_j at the deprotonated state (Scheme 7), eq 16 for a site specific dissociation constant becomes

$$K_a^{ij} = \frac{[H^+][A_j]}{[N_i]} = \frac{[H^+][A] \cdot f'_j}{[N] \cdot f_i} = K_a \frac{f'_j}{f_i} \quad (21)$$

corresponding to the deprotonation from an *i*-th tautomer to a *j*′-th tautomer. The overall K_a is written as

$$K_a = K_a^{11} \frac{f_1}{f'_1} = K_a^{12} \frac{f_1}{f'_2} = K_a^{21} \frac{f_2}{f'_1} = \cdots = K_a^{ij} \frac{f_i}{f'_j} \quad (22)$$

with the overall pK_a given by

$$pK_a = pK_a^{ij} - \log f_i + \log f'_j \quad (23)$$

3.7. pK_a of Guanine. The calculated pK_a values of guanine are summarized in Scheme 8 and Table 8. They are in good agreement with experimental values.[21−27]

The experimental values for pK_{a1} ($[G + H]^+ \rightarrow G^0$) are 3.2−3.3, in good agreement with the calculated value of 3.15. This protonation occurs at the N7 position.

The experimental values for pK_{a2} ($G^0 \rightarrow [G - H]^-$) are 9.2−9.6, in excellent agreement with the calculated value of 9.44. This involves the protonation at the N1 position.

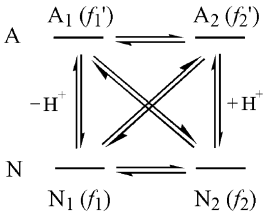
The experimental values for pK_{a3} ($[G - H]^- \rightarrow [G - 2H]^{2-}$) are 12.2−12.4, in good agreement with the calculated value of 12.61.

Several site specific pK_a values from the most stable neutral tautomer **1** are summarized in Table 9. We see that N_1H of neutral guanine is more acidic (pK_a 9.65) than N_9H (pK_a 10.03) although the difference is small. Also, N7 of neutral guanine (pK_a 3.20) accepts a proton much more easily than N3 (pK_a 1.14).

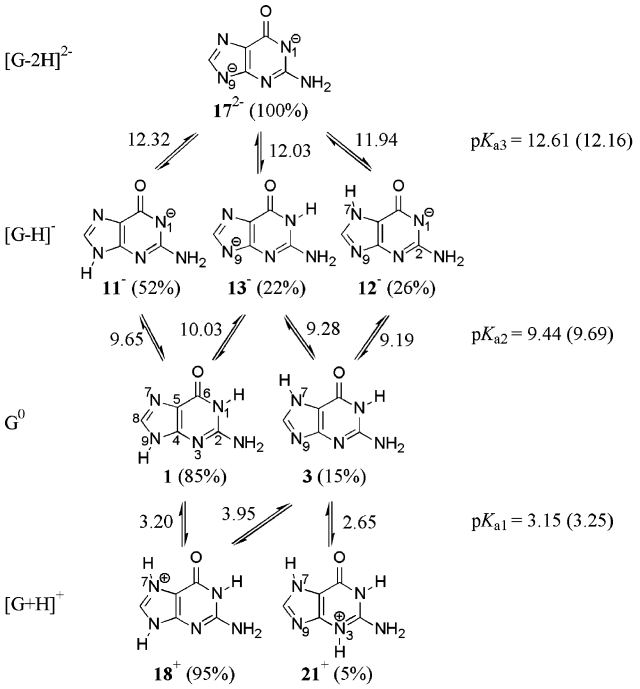
3.8. pK_a with a Proton Kept at N9. In nucleosides, the N9 of guanine is connected to a sugar rather than a proton and cannot participate in the deprotonation. Thus, to discuss more clearly the implication of our model study on the mutagenicity of guanine in DNA, it is necessary to take into account only the tautomers with a proton at N9 (as a crude model of sugar) (Scheme 9). In this case, guanine would exist predominantly as a single tautomer in each state. The neutral species would exist as a keto form **1**. Deprotonation would occur predominantly at N_1H, and protonation would occur predominantly at N7. The resulting pK_a values (9.65 and 3.20) are essentially the same as the major site specific pK_a values (Table 9).

3.9. Enol Tautomers and Base Mispairing. Our calculations indicate that under physiological conditions neutral guanine exists predominantly as 6-keto forms **1** and **3**. Because the 6-enol tautomer of guanine is known to be stable in the gas phase, the possibility that this tautomer might be involved in the formation of noncomplementary base pairs has drawn great attention. Indeed, for the gas phase, we calculate that the 6-enol tautomer **7** is only 0.9 kcal/mol higher in free energy than the keto tautomer **3**, leading to a significant population (12%). However, this enol tautomer **7** is not stable in the aqueous phase. It is 8.7 kcal/mol higher in free energy than **1** leading to a population in the aqueous phase of 4×10^{-7}. These results suggest that under normal physiological conditions, rare enol tautomers are not likely to induce base mispair formation.

SCHEME 7: Extended Model for the Calculation of Overall pK_a from Site Specific Values



SCHEME 8: Calculated pK_a Values in the Aqueous Phase and the Populations of Major Tautomers of Guanine in Each Ionization State[a]



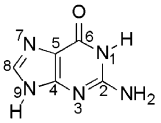
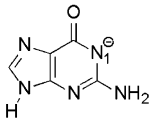
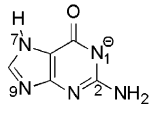
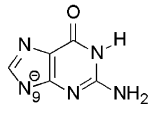
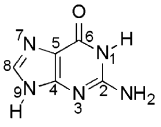
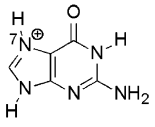
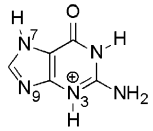
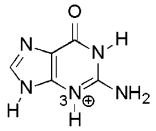
[a] Numbers in parentheses are the pK_a values that would be calculated from the Boltzmann-averaged free energy of each ionization state rather than eq 23.

TABLE 8: Calculated pK_aValues and Major Protonation Sites Corresponding to Each pK_a

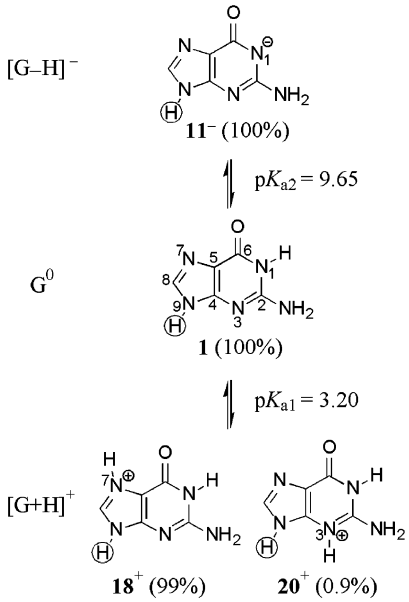
	calculation	experiment[d]
pK_{a3}[a]	12.6 (N9)	12.3−12.4
pK_{a2}[b]	9.4 (N1)	9.2−9.6[e]
pK_{a1}[c]	3.2 (N7)	3.2−3.3[e]

[a] From $\Delta G^\circ_{deprot,aq} = \Delta G^\circ_{aq}([G-2H]^{2-}) + \Delta G^\circ_{aq}(H^+) - \Delta G^\circ_{aq}([G-H]^-)$. [b] From $\Delta G^\circ_{deprot,aq} = \Delta G^\circ_{aq}([G-H]^-) + \Delta G^\circ_{aq}(H^+) - \Delta G^\circ_{aq}(G)$. [c] From $\Delta G^\circ_{deprot,aq} = \Delta G^\circ_{aq}(G) + \Delta G^\circ_{aq}(H^+) - \Delta G^\circ_{aq}([G + H]^+)$. [d] Experimental values for guanine base.[21−27] [e] The values measured at 40 °C are 9.92 (pK_{a2}) and 3.22 (pK_{a1}).[24,27]

TABLE 9: Site Specific pK_a Values: (a) Site Specific pK_{a2} Corresponding to Deprotonation from Each Site of Neutral Guanine 1 and (b) Site Specific pK_{a1} Corresponding to Protonation on Each Site of Neutral Guanine 1

(a)	**1**	From N_1H: **11**$^-$	From N_1H: **12**$^-$	From N_9H: **13**$^-$
				
pK_{a2}	–	9.65	9.94	10.03
(b)	**1**	On N_7: **18**$^+$	On N_3: **21**$^+$	On N_3: **20**$^+$
				
pK_{a1}	–	3.20	1.89	1.14

SCHEME 9: pK_a Values Calculated for a Simple Model of Guanine in DNA[a]



[a] The proton on N9 represents the deoxyribose unit and is not involved in deprotonation.

3.10. Deprotonation and Base Mismatch. Protonation or deprotonation of DNA bases could alter their hydrogen-bonding characteristics.[84−86] Because the guanine N_1H proton participates in a Watson−Crick hydrogen bond with cytosine, deprotonation from this site might lead to a breakdown of this hydrogen-bonding scheme resulting in potential base mispairing.

The population of deprotonated or protonated species at a specific pH can be estimated from the pK_a values by the Henderson−Hasselbalch equation:[87]

$$\mathrm{pH} = \mathrm{p}K_a + \log \frac{[\mathrm{A}^-]}{[\mathrm{HA}]} \tag{24}$$

The relative population of deprotonated form [A$^-$] (x) is given by

$$x = \frac{10^{\Delta}}{1 + 10^{\Delta}} \text{ where } \Delta = \mathrm{pH} - \mathrm{p}K_a \tag{25}$$

From pK_{a1} (3.20) and pK_{a2} (9.65) of guanine in DNA (Scheme 10), we calculated the population of cationic or anionic species under physiological conditions (neutral pH ∼ 7.0 in human body or pH = 6.5−8.5 in experiments). The population of cationic species is quite negligible in this range of pH values, but we estimate that 0.2% of guanine would exist as deprotonated anions at pH 7 and about 2% at pH 8.0. These amounts are quite significant, 5000 times larger than the contribution of enol tautomers. Previously, it has been proposed that base ionization could contribute to base mispair formation.[2] Structural studies with modified bases such as 5-fluorouracil[84] and 5-bromouracil[85] could form base pairs in pseudo Watson−Crick geometry upon ionization of the halouracil. Furthermore, in vitro polymerase assays have demonstrated an increase in mispair formation with guanine by both halouracils[86] with increasing solvent pH, indicating that such ionized base pairs can form in DNA and be recognized by DNA polymerase. The results of the calculations given here are consistent with the proposal that base ionization is more likely to provoke mispair formation than tautomerization.

Summary

Here, we use first principles QM (DFT (B3LYP) in combination with the PB continuum−solvation model) to calculate the relative energies of a number of neutral and ionized tautomers of guanine in both gas and aqueous phases. Using this model, we also present a method whereby the site specific pK_a values for guanine can be determined theoretically. This approach gives us numbers that are in agreement with other theoretical studies as well as experimentally determined values, validating its use for other systems—especially DNA damage products wherein solubility or other factors prevent experimental pK_a determination.

The major deprotonation and protonation sites of guanine are calculated to be N1 and N7, respectively, both in the gas phase and in the aqueous phase. Other minor sites are N9 for deprotonation (14% in the gas phase and 26% in the aqueous phase) and N3 for protonation (8% in the gas phase and 5% in the aqueous phase). The relative population of deprotonated species is estimated as 0.2−2% in the range of pH 7−8, suggesting in accord with previous studies that base ionization plays a significant role in base mispairing during DNA polymerase-mediated replication.

We calculate that neutral guanine exists as a mixture of two major keto tautomers, a N_9H form (**1**) and a N_7H form (**3**). These

TABLE 10: Gas Phase Proton Affinities (PA) and Basicities (GB) of Guanine (kcal/mol) Calculated at the B3LYP Level with Various Basis Sets, Which Shows the Importance of Diffuse Functions in $E_{0K,g}$ Calculations

	(a) 6-31G**	(b) 6-31++G**//6-31G**	(c) 6-31++G**/6-31G**	(d) 6-31++G**	(e) expt
ZPE, $\Delta\Delta G_{0\rightarrow 298K}$	6-31G**	6-31G**	6-31G**	6-31++G**	
optimization	6-31G**	6-31G**	6-31++G**	6-31++G**	
$E_{0K,g}$	6-31G**	6-31++G**	6-31++G**	6-31++G**	
PA	235.9	229.4	229.3	229.3	229 ± 2[a]
GB	228.5	222.0	221.9	221.8	222 ± 2[a]

[a] Ref 82.

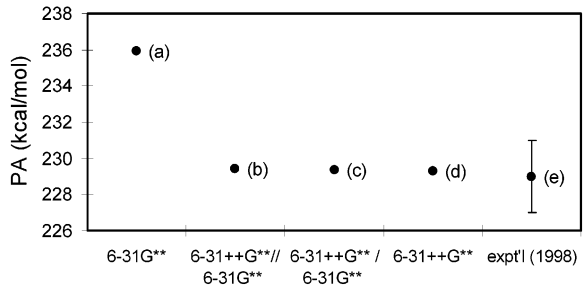


Figure 3. Gas phase proton affinities of guanine calculated at the B3LYP level with various basis sets. This illustrates the importance of including diffuse functions in the calculation of E_{0K}.

tautomers have similar relative populations in the gas phase, along with a significant population of the 6-enol tautomer (**7**) in the gas phase, but **1** becomes dominant (85%) in the aqueous phase. The energetic differences seen between the gas and the aqueous phase tautomers of guanine can be attributed to both dipole moment effects and electrostatic repulsion between the N_1H and the exocyclic amino group (this repulsion is shielded by solvation). Similar calculations on hypoxanthine, which lacks the 2-amino group, tell us to what extent this phenomenon can be attributed to the dipole moment effect, as hypoxanthine is not expected to have an intramolecular repulsion effect. These results suggest that in hydrophobic environments (wherein the surroundings probably behave as some intermediate between the extremes of the gas and aqueous phase), such as the active site of an enzyme, alternative tautomeric forms may play a significant role.

Acknowledgment. This work was supported in part by the National Institutes of Health [HD36385 (W.A.G.), GM 41336 (L.C.S.), and CA 85779 (L.C.S. and W.A.G.)] and the BK21 program and CMC of Korea (Y.H.J., S.H., and D.S.C.). In addition, the facilities of the MSC are also supported by DOE-ASCI, ARO-MURI, ARO-DURIP, National Science Foundation [CHE-99-85574 and 99-77872], Dow Chemical, 3M, Beckman Institute, Avery-Dennison, Chevron Corporation, Seiko Epson, Asahi Chemical, and Kellogg's.

Appendix A. Basis Set Dependence of Gas Phase PA and GB

We considered several basis sets for the calculation of gas phase PA and GB of guanine, to determine the optimum basis set to be used in the gas phase calculations reported above in Section 2.3. The 6-31G** basis includes polarization functions on all atoms, while 6-31++G** also includes diffuse functions on all atoms. Such diffuse functions are expected to be important for negative ions (deprotonation). We considered four cases (Table 10 and Figure 3): (i) No diffusion functions were included (6-31G**, Table 10a). (ii) Geometry optimization and frequency calculations (i.e., ZPE and $\Delta\Delta G_{0\rightarrow 298K}$ calculation) used 6-31G** and a single point energy calculation were done with 6-31++G** to improve E_{0K} (6-31++G**//6-31G**, Table 10b). (iii) Preliminary geometry optimization and frequency calculation were done with 6-31G**, and further geometry optimization was done with 6-31++G** starting from the optimum 6-31G** geometry to improve E_{0K} (6-31++G**/6-31G**, Table 10c). (iv) Diffuse functions were included in every step of calculation, i.e., E_{0K}, ZPE, and $\Delta\Delta G_{0\rightarrow 298K}$ calculation (6-31++G**, Table 10d).

TABLE 11: Basis Set Dependence of Site Specific pK_{a2} Values of Guanine[a]

	$E_{0K,g}$	aq opt[b]	ΔG_{solv}	ZPE/$\Delta\Delta G_{0\rightarrow 298K}$[c]	pK_{a2} ($\mathbf{1} \rightarrow \mathbf{11}^-$)
(a)	6-31G**	no	6-31G**	gas 6-31G**	19.2
(b)	6-31G**	yes	6-31G**	gas 6-31G**	19.1
(c)	6-31++G**	no	6-31G**	gas 6-31G**	9.7
(d)	6-31++G**	yes	6-31G**	gas 6-31G**	9.6
(e)	6-31++G**	no	6-31++G**	gas 6-31G**	10.2
(f)	6-31++G**	yes	6-31++G**	gas 6-31G**	9.8
(g)	6-31++G**	no	6-31++G**	gas 6-31++G**	10.3
(h)	6-31++G**	yes	6-31++G**	gas 6-31++G**	9.9
(i)	exp (pK_{a2})[d]				9.2−9.6 9.92 (40 °C)

[a] This paper uses level f for all calculations. [b] Geometry reoptimization in aqueous phase when calculating ΔG_{solv}. [c] ZPE and $\Delta\Delta G_{0\rightarrow 298K}$ were calculated in the gas phase. [d] Experimental values. Refs 21−27.

Case (i) (using only 6-31G**) gives results very different from the others, indicating that diffuse functions are very important especially for the calculation of $E_{0K,g}$. Frequency calculations are more time-consuming than geometry optimizations, and diffuse functions have only a minor effect on these results [energy difference less than 0.2 kcal/mol between cases (iii) and (iv)]. Thus, the most efficient basis set among the four tested seems to be 6-31++G**//6-31G** and 6-31++G**/6-31G**, which led to values in good agreement with experiment (Table 1e).[82,83] Consequently, we used the 6-31++G**/6-31G** basis set for all calculations reported in this paper.

Appendix B. Basis Set Dependence of pK_a Values for Guanine

The site specific pK_{a2} corresponding to $\mathbf{1} \rightarrow \mathbf{11}^-$ for guanine at the B3LYP level was calculated with the same basis sets used in Appendix A in order to check the importance of diffuse functions for aqueous phase calculations (Table 11 and Figure 4). The effect of geometry relaxation in the aqueous phase was also investigated. Because the frequency calculation is quite expensive, especially in solution, the ZPE and $\Delta\Delta G_{0\rightarrow 298K}$ were calculated only for the gas phase. The vdW radii were taken from the literature.[54]

The most complete calculations (Table 11h) lead to $pK_{a2} = 9.9$ in reasonable agreement with experiment[21] (9.92 and 9.2−9.6). These results show that ignoring diffuse functions in calculating $E_{0K,g}$ leads to a pK_{a2} too high by 10 units. The other factors (diffuse functions in ΔG_{solv}, ZPE, or $\Delta\Delta G_{0\rightarrow 298K}$ calculations and geometry relaxation in aqueous phase) have only minor effects on the result (within 0.4 pK_a units) (Table

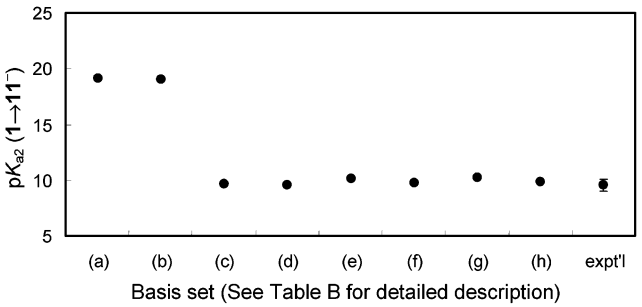


Figure 4. pK_a values of guanine calculated with various basis sets at the B3LYP level. This paper uses level f for all calculations.

TABLE 12: Calculated and Experimental pK_a Values of Guanine

	(a) calculation[a]	(b) calculation[b]	(c) experiment[c]
pK_{a3}[d]	13.41	12.61	12.3−12.4
pK_{a2}[e]	9.45	9.44	9.2−9.6[g]
pK_{a1}[f]	1.79	3.15	3.2−3.3[g]

[a,b] Calculated with parameters given in Table 13a,b, respectively. [c] Refs 21−27. [d] From $\Delta G^\circ_{\text{deprot,aq}} = \Delta G^\circ_{\text{aq}}([\text{G}-2\text{H}]^{2-}) + \Delta G^\circ_{\text{aq}}(\text{H}^+) - \Delta G^\circ_{\text{aq}}([\text{G}-\text{H}]^-)$. [e] From $\Delta G^\circ_{\text{deprot,aq}} = \Delta G^\circ_{\text{aq}}([\text{G}-\text{H}]^-) + \Delta G^\circ_{\text{aq}}(\text{H}^+) - \Delta G^\circ_{\text{aq}}(\text{G})$. [f] From $\Delta G^\circ_{\text{deprot,aq}} = \Delta G^\circ_{\text{aq}}(\text{G}) + \Delta G^\circ_{\text{aq}}(\text{H}^+) - \Delta G^\circ_{\text{aq}}([\text{G}+\text{H}]^+)$. [g] The values measured at 40 °C are 9.92 (pK_{a2}) and 3.22 (pK_{a1}).[24,27]

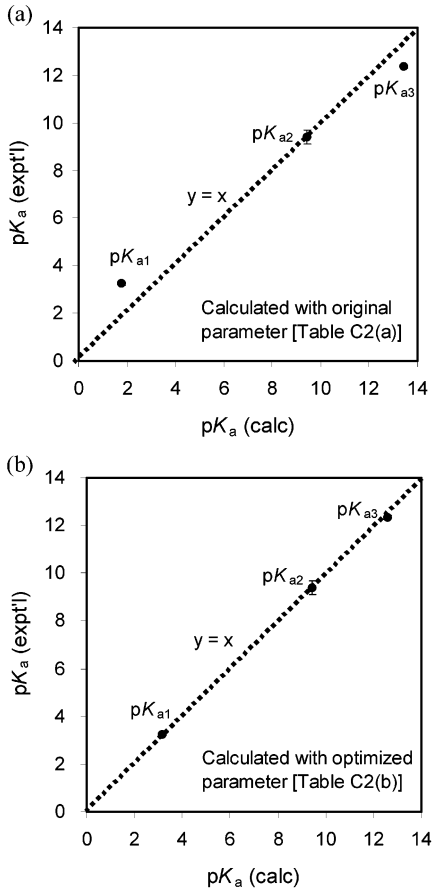


Figure 5. Calculated pK_a values vs experimental pK_a values of guanine. Dependence of calculated pK_a values on vdW radii used in the solvation free energy calculations. (a) Using radii from Marten[54] in Table 13a. (b) Using scaled parameters given in Table 13b. Dashed lines represent "$y = x$", the perfect match between calculation and experiments.

11c−g). On the basis of these results, all calculations in this work obtained ZPE and $\Delta\Delta G_{0\rightarrow 298\text{K}}$ using the gas phase 6-31G**, then calculated $E_{0\text{K,g}}$ with 6-31++G**, and calculated ΔG_{solv} with 6-31++G** after reoptimization in the aqueous phase (Table 11f).

TABLE 13: Parameters for Solvation Free Energy Calculation

atomic vdW radii	(a) Marten[a]	(b) this paper[c]
O sp^2	1.55	1.46
N sp^2	1.50	1.41
C sp^2	2.00	1.88
H attached to C sp^2	1.25	1.18
all other H	1.15	1.08
free energy solvation of proton	−263.98 kcal/mol[b]	−263.47 kcal/mol[d]

[a] Ref 54. [b] Ref 71. Other experimental values range from −261 to −254 kcal/mol,[41,70] and a recent calculation[69] gives −264 kcal/mol after a correction (Sections 2.5 and 2.6). [c] Reduced by 6% from Marten's set (a). [d] From the fit in Figure 5.

TABLE 14: Calculated and Experimental pK_a Values of Several DNA Bases

		calculation	experiment
cytosine	pK_{a1}	4.5	4.45−4.6[a]
	pK_{a2}	13.0	12.2[a]
isoguanine	pK_{a1}	3.9[b]	4.5 ± 0.2[c]
	pK_{a2}	9.6[b]	9.0 ± 0.2[c]
9-methylisoguanine	pK_{a1}	3.4[b]	3.85 ± 0.05[d]
	pK_{a2}	10.8[b]	9.9 ± 0.05[d]
8-oxoG	pK_{a1}	0.22[e]	
	pK_{a2}	8.76[e]	
	pK_{a3}	12.53[e]	

[a] Refs 24, 26, and 88. [b] Jang et al., in preparation. [c] Ref 40. [d] Ref 89. [e] Ref 90.

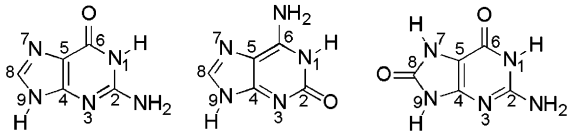
Appendix C. Parameter Optimization for Solvation Free Energy Calculation

The standard parameters from ref 54 lead to a pK_{a2} of 9.45 (Table 12) in excellent agreement with the experimental values of 9.2−9.6. However, pK_{a1} is too low by 1.4 units and pK_{a3} is too high by 1 unit. This means that the cationic guanines were calculated to be too acidic (too easy to lose a proton) and that anionic guanines, especially doubly anionic guanines, were calculated to be too basic (too easy to gain a proton) (Figure 5). This suggested to us that the atomic radii used to mark the separation of the continuum solvent from the explicit charges might be a bit too large. Indeed, we find reducing the radii by 6% (Table 13) leads to all three pK_a values within 0.2 units of experiment. We consider here the solvation free energy of proton to be a variable. However, adjusting it for the best fit to experiment leads to a value (−263.47 kcal/mol) in excellent agreement with the best literature value, −263.98 kcal/mol.[71] Thus, these parameters were used throughout all of the calculations on guanine.

Appendix D. Extension of Methodology to Other DNA Bases

We had calculated the pK_a values of guanine to be in good agreement with experiments. To find out whether this agreement is also expected for other systems, we applied exactly the same scheme (including the parameters determined in this study) to calculate pK_a values of several other DNA bases (cytosine, isoguanine, and 9-methylisoguanine) (Table 14). The agreement with experiments was within one pK_a unit for all of those cases, showing the predictive power of our pK_a calculation scheme. The pK_a values of 8-oxoG were also calculated (Table 14), but

there is no experimental data available on this base, probably because of its extremely low solubility. Only the pK_a values of 8-oxoguanosine (the nucleoside analogue of 8-oxoG) have been reported.[39] This case illustrates that our approach can be applied to obtain pK_a values even when the experimental determination is extremely difficult or almost impossible. The details of these calculations for isoguanine will be covered in a separate paper by the same authors. The 8-oxoG calculations are detailed in ref 90.



References and Notes

(1) Topal, M. D.; Fresco, J. R. *Nature (London)* **1976**, *263*, 285.
(2) Sowers, L. C.; Shaw, B. R.; Veigl, M. L.; Sedwick, W. D. *Mutat. Res.* **1987**, *177*, 201.
(3) Colominas, C.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **1996**, *118*, 6811.
(4) Katritzky, A. R.; Karelson, M. *J. Am. Chem. Soc.* **1991**, *113*, 1561.
(5) Gorb, L.; Leszczynski, J. *J. Am. Chem. Soc.* **1998**, *120*, 5024.
(6) Barsky, D.; Colvin, M. E. *J. Phys. Chem. A* **2000**, *104*, 8570.
(7) Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* **1999**, *103*, 8853.
(8) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.
(9) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114*, 4595.
(10) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Phys. Chem. A* **1997**, *101*, 10075.
(11) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.
(12) Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. *J. Phys. Chem. A* **2000**, *104*, 9619.
(13) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.
(14) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.
(15) Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A., III. *J. Phys. Chem. A* **2001**, *105*, 274.
(16) Kallies, B.; Mitzner, R. *J. Phys. Chem. B* **1997**, *101*, 2959.
(17) La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A., III; Sowers, L. C. *Chem. Res. Toxicol.* **2000**, *13*, 462.
(18) Szczepaniak, K.; Szczesniak, M. *J. Mol. Struct.* **1987**, *156*, 29.
(19) Sheina, G. G.; Stephanian, S. G.; Radchenko, E. D.; Blagoi, Y. P. *J. Mol. Struct.* **1987**, *158*, 275.
(20) Szczepaniak, K.; Szczesniak, M.; Person, W. B. *Chem. Phys. Lett.* **1988**, *153*, 39.
(21) *CRC Handbook of Chemistry and Physics*, 80th ed.; CRC Press: Boca Raton, 1999−2000.
(22) Dawson, R. M. C.; Elliott, D. C.; Elliott, W. H.; Jones, K. M. *Data for Biochemical Research*, 3rd ed.; Oxford University Press: Oxford, 1986.
(23) Fasman, G. D. *CRC Handbook of Biochemistry and Molecular Biology. Nucleic Acids. Volume 1*, 3rd ed.; CRC Press: Cleveland, OH, 1975.
(24) Ts'o, P. O. P. *Basic Principles in Nucleic Acid Chemistry*; Academic Press: New York, 1974.
(25) Jordan, D. O. *The Chemistry of Nucleic Acids*; Butterworth and Co.: Washington, 1960.
(26) Chargaff, E.; Davidson, J. N. *The Nucleic Acids. Chemistry and Biology*; Academic Press Inc.: New York, 1955.
(27) Bundari, S. *The Merck Index*, 12th ed.; Merck and Co., Inc.: Whitehouse Station, NJ, 1996.
(28) Kawanishi, S.; Hiraku, Y.; Oikawa, S. *Mutat. Res.* **2001**, *488*, 65.
(29) Kasai, H. *Mutat. Res.* **1997**, *387*, 147.
(30) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109.
(31) Shibutani, S.; Takeshita, M.; Grollman, A. P. *Nature* **1991**, *349*, 431.
(32) Wood, M. L.; Dizdaroglu, M.; Gajewski, E.; Essigmann, J. M. *Biochemistry* **1990**, *29*, 7024.
(33) Moriya, M.; Ou, C.; Bodepudi, V.; Johnson, F.; Takeshita, M.; Grollman, A. P. *Mutat. Res.* **1991**, *254*, 281.
(34) Cheng, K. C.; Cahill, D. S.; Kasai, H.; Nishimura, S.; Loeb, L. A. *J. Biol. Chem.* **1992**, *267*, 166.
(35) Lindahl, T.; Wood, R. D. *Science* **1999**, *286*, 1897.
(36) Michaels, M. L.; Miller, J. H. *J. Bacteriol.* **1992**, *174*, 6321.
(37) Grollman, A. P.; Moriya, M. *Trends Genet.* **1993**, *9*, 246.
(38) Bruner, S. D.; Norman, D. P. G.; Verdine, G. L. *Nature* **2000**, *403*, 859.
(39) Cho, B. P. *Magn. Reson. Chem.* **1993**, *31*, 1048.
(40) Albert, A.; Brown, D. J. *J. Chem. Soc.* **1954**, 2060.
(41) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.
(42) Pearson, R. G. *J. Am. Chem. Soc.* **1986**, *108*, 6109.
(43) Tinoco, I., Jr.; Sauer, K.; Wang, J. C. *Physical Chemistry. Principles and Applications in Biological Sciences*, 2nd ed.; Prentice-Hall: New Jersey, 1985.
(44) *Jaguar 4.0*; Schrödinger Inc.: Portland, OR, 2000.
(45) Greeley, B. H.; Russo, T. V.; Mainz, D. T.; Friesner, R. A.; Langlois, J.-M.; Goddard, W. A., III; Donnelly, R. E., Jr.; Ringnalda, M. N. *J. Chem. Phys.* **1994**, *101*, 4028.
(46) Slater, J. C. *Quantum Theory of Molecules and Solids. Vol. 4. The Self-Consistent Field for Molecules and Solids*; McGraw-Hill: New York, 1974.
(47) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(48) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.
(49) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(50) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*, 200.
(51) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101.
(52) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, *116*, 11875.
(53) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144.
(54) 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.
(55) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435.
(56) Archer, D. G.; Wang, P. *J. Phys. Chem. Ref. Data* **1990**, *19*, 371.
(57) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.
(58) Woods, R. J.; Khalil, M.; Pell, W.; Moffat, S. H.; Smith, V. H., Jr. *J. Comput. Chem.* **1990**, *11*, 297.
(59) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.
(60) Ben-Naim, A. *J. Phys. Chem.* **1978**, *82*, 792.
(61) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(62) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016.
(63) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147.
(64) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.
(65) Li, J. B.; Zhu, T. H.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
(66) Wang, J. M.; Wang, W.; Huo, S. H.; Lee, M.; Kollman, P. A. *J. Phys. Chem. B* **2001**, *105*, 5055.
(67) Abraham, M. H. *J. Chem. Soc., Faraday Trans. 1* **1984**, *80*, 153.
(68) Marcus, Y. *Ion Solvation*; John Wiley and Sons, Ltd.: New York, 1985; p 105.
(69) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Caldwell, R. A.; Rashin, A. A. *J. Chem. Phys.* **1998**, *109*, 4852.
(70) Reiss, H.; Heller, A. *J. Phys. Chem.* **1985**, *89*, 4207.
(71) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R., Jr. *J. Phys. Chem. A* **1998**, *102*, 7787.
(72) Civcir, P. U. *Struct. Chem.* **2001**, *12*, 15.
(73) Costas, M. E.; Acevedo-Chavez, R. *J. Mol. Struct. (THEOCHEM)* **1999**, *489*, 73.
(74) Costas, M. E.; Acevedo-Chavez, R. *J. Mol. Struct. (THEOCHEM)* **1999**, *468*, 39.
(75) Chenon, M.-T.; Pugmire, R. J.; Grant, D. M.; Panzica, R. P.; Townsend, L. B. *J. Am. Chem. Soc.* **1975**, *97*, 4636.
(76) Hernandez, B.; Luque, F. J.; Orozco, M. *J. Org. Chem.* **1996**, *61*, 5964.
(77) Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L. K.; LeBreton, P. R. *J. Phys. Chem.* **1980**, *84*, 1006.
(78) Shapiro, R. *Prog. Nucleic Acid Res. Mol. Biol.* **1968**, *8*, 73.
(79) Miles, H. T.; Howard, F. B.; Frazier, J. *Science* **1963**, *142*, 1458.
(80) Thewalt, U.; Bugg, C. E.; Marsh, R. E. *Acta Crystallogr. B* **1971**, *27*, 2358.
(81) Gatlin, L.; Davis, J. C., Jr. *J. Am. Chem. Soc.* **1962**, *84*, 4464.
(82) Hunter, E. P. L.; Lias, S. G. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413.
(83) Greco, F.; Liguori, A.; Sindona, G.; Uccella, N. *J. Am. Chem. Soc.* **1990**, *112*, 9092.
(84) Sowers, L. C.; Eritja, R.; Kaplan, B.; Goodman, M. F.; Fazakerly, G. V. *J. Biol. Chem.* **1988**, *263*, 14794.
(85) Sowers, L. C.; Goodman, M. F.; Eritja, R.; Kaplan, B.; Fazakerley, G. V. *J. Mol. Biol.* **1989**, *205*, 437.
(86) Yu, H.; Eritja, R.; Bloom, L. B.; Goodman, M. F. *J. Biol. Chem.* **1993**, *268*, 15935.

(87) Harris, D. C. *Exploring Chemical Analysis*; W. H. Freeman and Company: New York, 1997.
(88) Levene, P. A.; Bass, L. W.; Simms, H. S. *J. Biol. Chem.* **1926**, *70*, 229.
(89) Sepiol, J.; Kazimierczuk, Z.; Shugar, D. *Z. Naturforsch.* **1976**, *31c*, 361.
(90) Jang, Y. H.; Goddard, W. A., III; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *Chem. Res. Toxicol.* **2002**, *15*, 1023.