Accurate Proton Affinity and Gas-Phase Basicity Values for Molecules Important in Biocatalysis

Adam Moser,[†] Kevin Range,[‡] and Darrin M. York*,[†,§]

Department of Chemistry, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431, United States, Department of Chemistry, Lock Haven University of Pennsylvania, Lock Haven, Pennsylvania 17745, United States, and BioMaPS Institute and Department of Chemistry and Chemical Biology, Rutgers University, 610 Taylor Road, Piscataway, New Jersey 08854-8087, United States

Benchmark quantum calculations of proton affinities and gas-phase basicities of molecules relevant to biochemical processes, particularly acid/base catalysis, are presented and compared for a variety of multilevel and density functional quantum models. Included are nucleic acid bases in both keto and enol tautomeric forms, ribose in B-form and A-form sugar pucker conformations, amino acid side chains and backbone molecules, and various phosphates and phosphoranes, including thio substitutions. This work presents a high-level thermodynamic characterization of biologically relevant protonation states and provides a benchmark database for development of next-generation semiempirical and approximate density functional quantum models and parametrization of methods to predict pK_a values and relative solvation energies.

I. Introduction

The charge states of proteins and nucleic acids play an important role in both structure and reactivity. A key mechanism of controlling charge state and influencing acid−base catalysis is through protonation/deprotonation events of ionizable residues.[1] There has been considerable effort devoted to the prediction of proton affinities (PAs) and pK_a values with quantum chemistry and implicit/explicit solvation models.[2−29] Much of the quantitative work has focused on the prediction of pK_a shifts of ionizable residues with respect to a closely related reference state (for which reliable absolute pK_a values have been determined) rather than on the prediction of the absolute pK_a values themselves.

When carefully tested, this type of indirect approach often leads to cancellation of systematic errors that ultimately result in more quantitative accuracy. Examples of commonly used reference states for calculation of pK_a shifts include a closely related molecule or residue or the same molecule or residue in a different environment. The absolute pK_a values can be recovered from the calculated pK_a shift and the experimentally known absolute pK_a value of the reference state. Nonetheless, this area remains a challenge due to the small differences in free energy that give rise to pK_a shifts (a pK_a unit corresponds to 1.364 kcal/mol of energy at 298.15 K).

Reliable prediction of pK_a shifts using known experimental data may not always be possible since the determination of the experimental pK_a value of an appropriate reference state might not be available, for example, for metaphosphates and phosphoranes that form reactive intermediates in phosphoryl transfer reactions. For these important intermediates, experiment can only provide a rough estimate of the pK_a values. Consequently, methods that improve the prediction of absolute pK_a values are
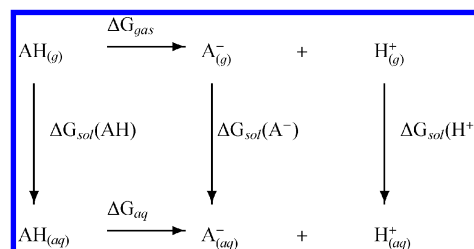

Figure 1. A commonly employed thermodynamic cycle for calculating pK_a values.

also of importance and ultimately may lead to the determination of pK_a shifts with increased reliability.

The accurate prediction of absolute pK_a values is often made using a thermodynamic cycle such as the one shown in Figure 1, although many variations have also been tested.[17,20,30] A key quantity in the cycle is the gas-phase basicity (GPB) that involves deprotonation of the species of interest in the gas-phase (ΔG_{gas} in Figure 1). Related to the GPB is the PA, which is the enthalpic component of the same process.[31] Given that the active site of most enzymes or the center of large biomolecules is an environment that is neither that of bulk solution nor the gas phase, benchmarking these two end point environments is key to making pK_a predictions.

The purpose of the present work is to deliver a consistent, benchmark database of quantum calculations for PAs and GPBs of residues involved in biocatalysis. Compounds were considered that represent titratable amino acid backbone and side-chain residues, nucleic acid bases, sugar and phosphate moieties, as well as biological and chemically modified phosphates and phosphoranes important in phosphoryl transfer reactions and RNA catalysis. The accuracy of a series of high-level quantum model chemistries was assessed against known experimental PA and GPB values. This comparison was then used to derive a set of empirical bond enthalpy and free-energy corrections and a linear regression correction that improves the accuracy and, for closely related molecules, the predictive capability of

* To whom correspondence should be addressed. E-mail: york@biomaps.rutgers.edu.
[†] University of Minnesota.
[‡] Lock Haven University of Pennsylvania.
[§] Rutgers University.

the methods. The results presented herein provide valuable data for the development of next-generation approximate quantum models that can be used in combined QM/MM simulations of biocatalysis,[32−34] for the determination of relative solvation free energies of conjugate acid/base pairs from pK_a data in solution,[35−38] and for the development of improved models for the theoretical determination of pK_a shifts in biomolecules.[29,39−44]

II. Methods

A. Electronic Structure Calculations. All electronic structure and thermochemical analyses were performed using the Gaussian03 suite of programs.[45] Three "multilevel" methods were studied, CBS-QB3,[46,47] G3B3,[48] and G3MP2B3.[48] CBS-QB3 (abbreviated CBS in this article) is a multilevel model chemistry that combines the results of several electronic structure calculations and empirical terms to predict molecular energies to around 1 kcal/mol accuracy.[49] The required electronic structure calculations are outlined below; see ref 46 for details.

CBS:

- B3LYP/6-311G(2d,d,p) geometry optimization and frequencies
- MP2/6-311+G(3df,2df,2p) energy and CBS extrapolation
- MP4(SDQ)/6-31+G(d(f),p) energy
- CCSD(T)/6-31+ G† energy

The G3B3 and the related G3MP2B3 (abbreviated G3MP2 in this article and not to be confused with the usual meaning referring to the method that uses Hartree−Fock geometry and frequency calculations) methods are both modifications of the Gaussian-3 multilevel theory for the calculation of molecular energies.[50] Like CBS-QB3, G3B3 and G3MP2 use density functional theory with the B3LYP functional for geometries and frequencies and combine the results of several electronic structure calculations and empirical terms to predict molecular energies to around 1 kcal/mol accuracy. The required electronic structure calculations for the G3B3 method are outlined below; see ref 48 for details.

G3B3:

- B3LYP/6-31G(d) geometry optimization and frequencies
- MP2/G3Large energy
- MP4/6-31G(d) energy
- MP4/6-31+G(d) energy
- MP4/6-31G(2df,p) energy
- QCISD(T)/6-31G(d) energy

The G3MP2 method eliminates all of the MP4 calculations above, trading some accuracy for speed.

All of the multilevel methods studied here formally scale as $O(N^7)$ due to the CCSD(T) step in CBS, the QCISD(T) step of the G3 methods, and the MP4 steps in G3B3 (for the molecules in the present study, the large basis set MP4 step was usually the computational bottleneck with the G3B3 method). For the systems studied in this paper, the scaling of the multilevel methods was not prohibitive, but for larger systems of biological interest, the $O(N^7)$ will eventually dominate and make the calculation unfeasible. Therefore, several model chemistries based solely on hybrid density functionals were investigated that have more favorable scaling properties, specifically PBE0/6-311++G(3df,2p) (designated PBE0), B1B95/6-311++G(3df,2p) (designated B1B95), and B3LYP/6-311++G(3df,2p) (designated B3LYP). Additionally, the B3LYP/6-311++G(3df,2p)//B3LYP/6-31++G(d,p) model chemistry

(designated QCRNA) that has been extensively applied to model biological phosphorus compounds[7,51−57] was also included.

B3LYP is a three-parameter hybrid functional[58] using the B88 exchange functional of Becke[59] and the Lee, Yang, and Parr correlation functional.[60] PBE0 is a zero-parameter hybrid functional[61] using the Perdew, Burke, Ernzerhof exchange and correlation functional.[62] B1B95 is a one-parameter hybrid functional[63] using the B88 exchange functional of Becke[59] and B95 correlation functional.[63]

In an effort to give the hybrid density functional methods the best possible accuracy for benchmark purposes and avoid problems in the frequency calculations,[64] all calculations (except the QCRNA and multilevel theories) were run with the so-called "ultrafine" numerical integration grid (a pruned grid based on 99 radial shells and 590 angular points per shell) and tight convergence criteria for the geometry optimizations. The use of large basis sets, ultrafine numerical integration meshes, and tight convergence criteria serve to make these calculations benchmark-quality. The QCRNA model is significantly less expensive than the related B3LYP model since it avoids the geometry optimization and frequency calculation steps with the large 6-311++G(3df,2p) basis and uses the default integration grid (a pruned grid based on 75 radial shells and 302 angular points per shell) and geometry convergence criteria. As demonstrated below, QCRNA gives results in very close agreement to those of B3LYP.

B. Calculation of Proton Affinities and Gas-Phase Basicities. The PA and GPB of a species A$^−$ are related to the gas-phase reaction

$$A^-_{(g)} + H^+_{(g)} \rightarrow AH_{(g)} \tag{1}$$
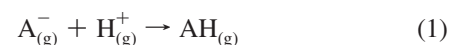
The PA of A$^−$ is defined as the negative of the enthalpy change (ΔH) of the process in eq 1, and the GPB of A$^−$ is defined as the negative of the corresponding Gibbs free-energy change (ΔG).[31] Here, the A$^−$ is the conjugate base associated with the neutral acid AH.

The required thermodynamic properties were obtained from the electronic structure calculations using standard statistical mechanical expressions for separable vibrational, rotational, and translational contributions within the harmonic oscillator, rigid rotor, and ideal gas/particle-in-a-box models in the canonical ensemble.[65] The standard state in the gas phase was for a mole of particles at 298.15 K and 1 atm pressure.

In the case that a molecule has more than one conformational state accessible at a given temperature, these conformations should be appropriately sampled with Boltzmann-weighted averaging to obtain the most accurate PA and GPB. In this work, the free energies were used to determine the Boltzmann average, and only conformations that changed the PA or GPB by more than 0.1 kcal/mol were considered. This occurred for propanol and propanethiol, where the trans conformation is higher in energy than the gauche conformation but by less than 0.5 kcal/mol in free energy.

The enthalpy of the proton was calculated from the ideal gas expression

$$H(H^+) = U + pV = \frac{5}{2}RT \tag{2}$$

where U is the internal energy, p and V are the pressure and volume, respectively, R is the universal gas constant, and T is

Proton Affinity and Gas-Phase Basicity and Biocatalysis

J. Phys. Chem. B, Vol. 114, No. 43, 2010 **13913**

the absolute temperature. The entropy of the proton was calculated from the Sackur−Tetrode equation[66]

$$S(H^+) = R \ln\left(\frac{e^{5/2}k_B T}{p\Lambda^3}\right) \quad (3)$$

where k_B is the Boltzmann constant and Λ is the thermal de Broglie wavelength $[\Lambda = (h^2/2\pi m k_B T)^{1/2} Z$, where h is Planck's constant and m is the mass of the proton]. Under the standard state conditions, the values of the enthalpy and entropy of the gas-phase proton are $H(H^+) = 1.48$ kcal/mol and $S(H^+) = 26.02$ cal/(mol·K), respectively, and lead to a gas-phase Gibbs free-energy value of $G(H^+) = H(H^+) - TS(H^+) = -6.28$ kcal/mol.

It is sometimes the case that a molecule and/or its conjugate base has more than one indistinguishable microscopic protonation state. All of the GPB values in this paper are microscopic GPBs since that is what naturally comes out of electronic structure calculations of a single protonation state. The conversion between microscopic and macroscopic GPB values was performed as follows.

The equilibrium constant, K^M, (where the M indicates macroscopic) for the reverse process to that of eq 1, assuming unit activity coefficients, is given by

$$K^M = \frac{[A^-]_M[H^+]}{[AH]_M} \quad (4)$$

If A^- has m indistinguishable microscopic protonation states and AH has n indistinguishable microscopic protonation states, then

$$K^M = \frac{m[A^-]_\mu[H^+]}{n[AH]_\mu} = K^\mu\left(\frac{m}{n}\right) \quad (5)$$

where μ indicates microscopic quantities and $K^\mu = [A^-]_\mu[H^+]/[AH]_\mu$. The free-energy change (ΔG) for a process is related to the equilibrium constant by

$$\Delta G = -RT \ln K \quad (6)$$

where R is the ideal gas constant and T is the temperature of interest. Substitution of eq 6 into eq 5 yields the following equation for interconversion of microscopic and macroscopic free-energy changes

$$\Delta G^\mu = \Delta G^M + RT \ln\left(\frac{m}{n}\right) \quad (7)$$

For example, H_3PO_4 has four indistinguishable microscopic protonation states (distribution of three protons among four sites gives $_4C_3 = 4$, where $_nC_k = n!/(n-k)!k!$ is a binomial coefficient), and its conjugate base, $H_2PO_4^-$, has six indistinguishable microscopic protonation states ($_4C_2 = 6$). The experimental (macroscopic) GPB of H_3PO_4 is 323.0 kcal/mol.[67] Application of eq 7 yields a microscopic GPB of 323.2 kcal/mol at 298.15 K.

III. Results and Discussion

In this section, we present calculated PA and GPB results for a wide range of biologically relevant molecules. The first section presents results for a series of O−H, N−H, and S−H molecules that have experimental PA and GPB values for comparison, from which a set of empirical corrections to the calculated values are derived. The second section presents results for molecules that represent amino acid backbone and side chain residues. The third section presents results for nucleic acid nucleobases, including tautomerization energies. The fourth section presents results for nucleic acid ribose sugar models, including different sugar pucker conformations. The fifth section presents results for native and thio-substituted phosphates and phosphoranes of interest in the study of phosphoryl transfer reactions. Throughout this section, we have endeavored, where possible, to consider model compounds that have limited flexibility in order to avoid complications due to multiple minima that can significantly influence PA and GPB values. For example, side chain conformation in methionine can tune the PA and GPB by ∼10 kcal/mol from the value for glycinium.[68] The complete geometrical and thermodynamic data are available in the Supporting Information and via the QCRNA database.[57,69]

A. Experimental Comparison. Table 1 summarizes the PA and GPB errors (calculated − experimental value) for O−H, S−H, and N−H bond-containing molecules. All experimental values are taken from the NIST online database,[67] unless otherwise noted. Included are the maximum error (MAXE), root-mean-squared error (RMSE), mean unsigned error (MUE), and mean signed error (MUE). As expected, the multilevel methods perform better than the density functional methods for both PA and GPB, with PBE0 and B1B95 generally overestimating and B3LYP and QCRNA underestimating the values. As QCRNA consists only of B3LYP calculations, it very closely resembles the B3LYP results, except for propanoic acid GPB, where they differ by 1.8 kcal/mol. The largest outliers for all of the multilevel models are dimethylphosphate and phosphoric acid. For PBE0 and B1B95, water and 4-methylimidazolium have the largest deviation, while B3LYP and QCRNA have maximum error on *p*-nitrophenol. For all models, the PA and GPB generally have the same trends and error metrics. Also provided in Supporting Information for comparison are the dipole moments of the neutral molecules in Table 3 with available experimental data.

To investigate the relative energy lowering between the acid and conjugate base with increasing basis set and the effect on the PA and GPB, the values in Table 1 were obtained using five increasingly larger Dunning basis sets (cc-pVTZ, cc-pVQZ, cc-pV5Z, aug-cc-pVTZ, aug-cc-pVQZ)[70] with the B3LYP functional and compared to the experimental values. Table 2 provides the error metrics for these calculations, while the full table of individual values is provided in Supporting Information. From the error metrics, two general observations can be made. First, basis sets lacking diffuse functions overestimate both PA and GPB, while those with diffuse functions underestimate them slightly. This is most likely associated with the importance of diffuse functions for the anions. Second, the cc-pV5Z, aug-cc-pVTZ, and aug-cc-pVQZ have similar errors, showing that we are approaching the complete basis set limit, with the augmented triple-ζ basis set offering significant computational advantage.

In previous work for the PA and GPB of O−H-containing compounds,[8,71] it was found that for the DFT methods, |MSE| ≈ MUE, indicating a systematic error.[71] This observation motivated a simple additive correction for the bond enthalpy

TABLE 1: PA (left column) and GPB (right column) Error Analysis[a]

molecule[b]	CBS		G3B3		G3MP2		PBE0		B1B95		B3LYP		QCRNA		experiment	
water	1.7	1.7	1.2	1.2	1.3	1.3	3.2	3.2	3.3	3.3	0.1	0.1	0.1	0.1	390.3 (0.2)	383.7 (0.2)
hydronium	−1.0	−1.1	−0.3	−0.4	−0.4	−0.5	0.6	0.5	0.5	0.3	−1.1	−1.2	−1.1	−1.2	165.0 (1.0)	157.7 (0.1)
methanol	1.1	1.3	2.2	2.4	2.2	2.3	−0.5	−0.3	−0.1	0.1	−2.3	−2.0	−2.3	−2.1	381.5 (1.0)	374.8 (0.7)
ethanol	0.7	0.4	1.5	1.3	1.7	1.4	−0.3	−0.5	0.3	−0.0	−2.2	−2.5	−2.2	−2.4	378.2 (0.8)	371.3 (1.0)
propanol	1.1	0.7	2.0	1.6	2.2	1.8	0.4	−0.1	0.9	0.4	−1.2	−1.8	−1.2	−1.7	376.0 (1.1)	369.4 (1.1)
2-propanol	0.8	0.3	1.4	1.0	1.7	1.3	0.7	0.3	1.2	0.8	−1.4	−1.8	−1.3	−1.7	375.7 (0.8)	368.8 (1.0)
DMPH[c]	−2.4	−1.7	−1.8	−1.2	−1.3	−0.7	−0.6	0.2	−0.7	0.3	−1.6	−0.8	−1.5	−0.9	331.6 (4.1)	324.6 (4.0)
phosphoric acid	−2.6	−2.5	−2.2	−2.2	−1.8	−1.8	−0.5	−0.3	−0.6	−0.4	−2.4	−2.2	−2.3	−2.2	330.5 (5.0)	323.2 (4.9)
formic acid	−0.4	−1.3	0.4	−0.6	0.9	−0.1	−0.2	−1.1	0.1	−0.8	−2.0	−2.9	−2.0	−2.9	344.0 (1.6)	337.9 (1.2)
acetic acid	0.2	−2.0	1.0	−1.2	1.4	−0.8	0.4	−0.1	0.7	0.1	−1.4	−1.9	−0.8	−3.1	347.2 (1.1)	341.4 (1.2)
propanoic acid	−1.2	−0.3	−0.4	0.6	−0.0	1.0	0.5	−1.2	0.5	−0.5	−1.3	−2.8	−1.9	−1.0	347.4 (1.8)	340.4 (1.4)
glycine	1.5	−0.4	2.3	0.5	2.7	0.8	2.7	0.7	2.9	0.9	1.2	−0.9	1.2	−0.8	340.3 (1.1)	335.1 (1.4)
proline	−0.2	−0.9	0.7	−0.0	0.9	0.2	1.8	1.0	1.8	1.0	0.1	−0.8	0.1	−0.8	340.3 (3.1)	333.4 (3.0)
phenol	−0.8	−0.6	−0.6	−0.5	−0.6	−0.5	−1.4	−1.3	−0.7	−0.7	−2.4	−2.3	−2.5	−2.4	350.1 (1.1)	342.9 (1.4)
o-chlorophenol	0.7	−0.7	1.0	−0.4	1.0	−0.4	0.4	−1.1	1.0	−0.5	−1.0	−2.4	−1.0	−2.4	343.4 (2.3)	337.1 (2.0)
m-chlorophenol	−0.5	−1.2	−0.2	−0.9	−0.2	−0.9	−1.2	−1.8	−0.6	−1.2	−2.4	−3.0	−2.3	−2.9	342.1 (3.1)	335.2 (1.4)
p-chlorophenol	−0.5	−0.9	−0.2	−0.5	−0.2	−0.5	−1.0	−1.3	−0.5	−0.7	−2.2	−2.4	−2.3	−2.6	343.4 (1.6)	336.5 (1.4)
p-methylphenol	−0.3	−0.6	0.0	−0.4	0.0	−0.4	−0.7	−1.0	−0.1	−0.6	−1.8	−2.0	−1.8	−2.2	350.7 (1.3)	343.8 (1.2)
p-nitrophenol	−0.2	−0.1	0.2	0.2	0.7	0.7	−2.3	−2.3	−1.7	−1.8	−4.1	−4.1	−4.1	−4.2	327.8 (2.1)	320.9 (2.0)
hydrogen sulfide	−0.5	−0.5	−0.1	0.0	0.2	0.2	0.5	0.5	0.8	0.8	−0.3	−0.3	−0.4	−0.4	351.3 (0.8)	344.9 (2.0)
methanethiol	−0.5	−0.2	0.1	0.2	0.4	0.7	0.9	1.2	1.4	1.7	0.0	0.4	−0.0	0.3	357.3 (1.2)	350.6 (2.0)
ethanethiol	−1.2	−1.5	−0.5	−0.0	−0.2	−0.6	0.3	−0.1	1.0	0.7	−0.4	−0.9	−0.4	−0.8	355.4 (1.5)	348.9 (2.0)
propanethiol	−1.0	−1.0	−0.4	0.4	−0.1	−0.2	0.6	0.5	1.2	1.0	0.2	−0.2	0.2	−0.2	354.2 (2.2)	347.9 (2.0)
2-propanethiol	−0.5	−1.2	0.1	−0.9	0.4	−0.3	1.4	0.6	2.0	1.2	0.6	−0.1	0.6	−0.1	353.4 (2.2)	347.1 (2.0)
ammonium	0.2	0.3	0.5	−0.5	0.3	0.3	0.6	0.7	0.2	0.3	−1.1	−1.0	−1.1	−1.0	204.0 (0.5)	195.7 (0.5)
methylammonium	0.1	0.2	0.6	−0.6	0.4	0.6	0.5	0.5	0.2	0.3	−0.6	−0.6	−0.6	−0.6	214.9 (0.5)	206.6 (0.5)
ethylammonium	0.1	0.6	0.6	0.6	0.5	1.0	0.9	1.4	0.5	1.0	−0.1	0.3	−0.1	0.3	218.0 (5.0)	210.0 (5.0)
propanammonium	−0.1	0.5	0.4	0.7	0.2	0.9	0.9	1.5	0.7	1.3	−0.0	0.5	−0.1	0.6	219.4 (0.5)	211.3 (0.5)
2-propanammonium	−0.6	0.2	−0.1	1.1	−0.2	0.6	1.0	1.8	0.7	1.5	−0.1	0.8	−0.1	0.7	220.8 (0.5)	212.5 (0.5)
1H-3H-imidazolium	−0.0	−0.5	0.7	1.0	0.4	−0.1	1.9	1.4	1.8	1.2	1.3	0.8	1.3	0.8	225.3 (0.5)	217.7 (0.5)
4-methyl-imidazolium	1.3	1.1	1.9	0.7	1.7	1.5	3.6	3.5	3.5	3.4	3.0	2.9	3.0	2.9	227.7 (2.0)	220.1 (2.0)
pyridinium	−0.4	−0.8	0.3	0.1	0.0	−0.3	2.0	1.7	1.8	1.5	1.8	1.5	1.9	1.5	222.0 (5.0)	214.7 (0.5)
pyrrolidinium	0.1	0.4	0.8	1.7	0.7	1.2	1.0	1.3	0.8	1.1	0.6	0.9	0.6	1.0	226.6 (0.5)	218.8 (0.5)
glycinium	−0.0	0.4	0.4	−0.0	0.3	0.7	1.7	1.7	1.3	1.5	0.1	0.3	0.1	0.5	211.9 (0.5)	203.7 (0.5)
MAXE	−2.6	−2.5	2.3	2.4	2.7	2.3	3.6	3.5	3.5	3.4	−4.1	−4.1	−4.1	−4.2		
RMSE	1.0	1.0	1.1	1.0	1.1	0.9	1.4	1.3	1.4	1.2	1.6	1.8	1.6	1.8		
MUE	0.7	0.8	0.8	0.8	0.8	0.8	1.1	1.1	1.1	1.0	1.2	1.5	1.2	1.5		
MSE	−0.2	−0.4	0.4	0.2	0.5	0.3	0.6	0.3	0.8	0.5	−0.7	−0.9	−0.7	−0.9		

[a] All quantities are in kcal/mol. Experimental values are from ref 67, with error estimates given in parentheses. The error is given as quantum model chemistry minus the experimental value (error = calculated − experiment). Error metrics of maximum error (MAXE), root-mean-squared error (RMSE), mean unsigned error (MUE), and mean signed error (MSE) are given. [b] "Molecule" refers to AH in eq 1. [c] Hydrogen dimethyl phosphate.

TABLE 2: PA (left column) and GPB (right column) Error Analysis for a B3LYP Functional with Various Dunning Basis Sets[a]

	cc-pVTZ		cc-pVQZ		cc-pV5Z		aug-cc-pVTZ		aug-cc-pVQZ	
MAXE	22.7	22.7	13.6	13.7	5.1	5.1	−3.7	−3.7	−3.5	−3.5
RMSE	5.6	5.4	3.1	2.9	1.7	1.8	1.5	1.7	1.4	1.6
MUE	4.2	4.0	2.0	1.9	1.2	1.3	1.3	1.5	1.1	1.3
MSE	4.2	4.0	1.9	1.6	0.6	0.3	−0.6	−0.9	−0.4	−0.6

[a] All quantities are in kcal/mol. Error metrics are maximum error (MAXE), root-mean-squared error (RMSE), mean unsigned error (MUE), and mean signed error (MSE).

and entropy that brought the error metrics of the density function models closer to those of the multilevel models for both the PA and GPB. This PA correction (ΔH^C) is applied as

$$\Delta H'_X = \Delta H_X + \Delta H^C_X \qquad (8)$$

where X is either O−H, S−H, or N−H bonds. This correction is similarly added for the GPB. ΔH^C_X and ΔG^C_X are assigned to enforce the MSE to be 0. This bond energy correction (BEC) model is now compared to a more flexible model based on a linear regression using

$$PA_{pred} = b_X \cdot PA_{model} + a_X \qquad (9)$$

where PA_{model} is the PA predicted by a computational model, a_X and b_X are parameters optimized to best fit the experimental data, X is either the O−H, S−H, or N−H bond, and PA_{pred} is the linear regression model (LRM) prediction. The equation for the GPB is analogous to eq 9. Parameters for both the BEC and LRM models are provided in Supporting Information.

Table 3 presents PA error metrics for the uncorrected (Raw), BEC, and LRM models. Shown are the MUE and MSE with the standard deviation following in parentheses and the total residual, S (i.e., the sum of the deviation squares). Also calculated was the linear correlation coefficients, but for all models, this value was near unity. Similar trends are found for the GPB, and the table is provided in Supporting Information.

For the uncorrected (Raw) values, the CBS model outperforms all other models for O−H and N−H compounds, as shown by a lower residual. Further, CBS shows a tighter distribution around the experimental values, shown in the standard deviation of the MUE and MSE. The BEC model significantly lowers the residual and MUE for the DFT methods as well as the CBS model for the S−H compounds; as discussed above, this is an indication of a systematic error in the methods.

Figure 2 displays a histographic comparison of the CBS (the most accurate on average) and QCRNA (the most computationally inexpensive) methods for the uncorrected, BEC, and LRM. From this comparison, it is apparent how the uncorrected QCRNA systematically underestimates the PA values. After the BEC or LRM corrections, the QCRNA has a more evenly

Proton Affinity and Gas-Phase Basicity and Biocatalysis

J. Phys. Chem. B, Vol. 114, No. 43, 2010 **13915**

TABLE 3: PA Error Analysis for Each Model Broken down by Bond Type and Fitting Method[a]

model		CBS	G3B3	G3MP2	PBE0	B1B95	B3LYP	QCRNA
Raw O−H	MUE	0.9 (0.7)	1.1 (0.8)	1.2 (0.8)	1.0 (0.9)	1.0 (0.9)	1.7 (0.9)	1.7 (1.0)
	MSE	−0.1 (1.2)	0.5 (1.3)	0.7 (1.2)	0.1 (1.4)	0.4 (1.3)	−1.6 (1.2)	−1.6 (1.2)
	S	24.381	31.889	34.482	33.285	31.089	67.804	68.687
BEC O−H	MUE	0.9 (0.7)	1.0 (0.7)	1.0 (0.7)	1.1 (0.9)	1.0 (0.8)	0.8 (0.8)	0.8 (0.8)
	MSE	−0.0 (1.2)	−0.0 (1.3)	−0.0 (1.2)	0.0 (1.4)	0.0 (1.3)	0.0 (1.2)	0.0 (1.2)
	S	24.274	27.729	25.808	33.196	27.920	23.501	23.501
LRM O−H	MUE	0.6 (0.6)	0.7 (0.6)	0.8 (0.5)	0.9 (0.8)	0.8 (0.7)	0.8 (0.8)	0.9 (0.8)
	MSE	−0.0 (0.8)	0.0 (0.9)	−0.0 (1.0)	−0.0 (1.3)	0.0 (1.1)	0.0 (1.1)	0.0 (1.2)
	S	11.656	14.891	15.529	26.839	20.982	21.951	23.206
Raw S−H	MUE	0.7 (0.3)	0.2 (0.2)	0.3 (0.1)	0.7 (0.4)	1.3 (0.5)	0.3 (0.2)	0.3 (0.2)
	MSE	−0.7 (0.3)	−0.2 (0.3)	0.1 (0.3)	0.7 (0.4)	1.3 (0.5)	0.0 (0.4)	−0.0 (0.4)
	S	3.074	0.447	0.435	3.357	8.983	0.736	0.746
BEC S−H	MUE	0.3 (0.1)	0.2 (0.1)	0.2 (0.1)	0.3 (0.2)	0.3 (0.3)	0.3 (0.2)	0.3 (0.2)
	MSE	−0.0 (0.3)	−0.0 (0.3)	−0.0 (0.3)	0.0 (0.4)	0.0 (0.5)	−0.0 (0.4)	0.0 (0.4)
	S	0.388	0.317	0.332	0.626	0.812	0.734	0.745
LRM S−H	MUE	0.3 (0.1)	0.2 (0.1)	0.2 (0.1)	0.3 (0.3)	0.3 (0.3)	0.3 (0.2)	0.3 (0.2)
	MSE	0.0 (0.3)	0.0 (0.3)	−0.0 (0.3)	0.0 (0.4)	−0.0 (0.4)	−0.0 (0.4)	−0.0 (0.4)
	S	0.386	0.316	0.331	0.608	0.685	0.709	0.722
Raw N−H	MUE	0.3 (0.4)	0.6 (0.5)	0.5 (0.5)	1.4 (0.9)	1.2 (1.0)	0.9 (1.0)	0.9 (1.0)
	MSE	0.1 (0.5)	0.6 (0.5)	0.4 (0.5)	1.4 (0.9)	1.2 (1.0)	0.5 (1.2)	0.5 (1.2)
	S	2.280	6.047	4.266	27.750	22.455	16.292	16.149
BEC N−H	MUE	0.3 (0.4)	0.3 (0.4)	0.3 (0.4)	0.7 (0.6)	0.7 (0.6)	1.0 (0.7)	1.0 (0.7)
	MSE	0.0 (0.5)	0.0 (0.5)	0.0 (0.5)	−0.0 (0.9)	0.0 (1.0)	−0.0 (1.2)	0.0 (1.2)
	S	2.249	2.491	2.439	7.869	8.989	13.840	13.888
LRM N−H	MUE	0.3 (0.3)	0.3 (0.3)	0.3 (0.3)	0.6 (0.4)	0.6 (0.4)	0.6 (0.3)	0.6 (0.3)
	MSE	0.0 (0.5)	−0.0 (0.5)	0.0 (0.5)	−0.0 (0.7)	−0.0 (0.7)	−0.0 (0.7)	−0.0 (0.7)
	S	2.095	2.000	1.876	4.655	4.687	4.191	4.374

[a] Raw indicates that no correction has been made. BEC and LRM are bond energy correction and linear regression model, respectively (see text for details). Error metrics given are mean unsigned error (MUE) and mean signed error (MSE), with their standard deviations in parentheses, and the sum of the squares of the deviations (S).

distributed error distribution but is still is not as tightly distributed as the CBS model.

Summarizing the model comparison, all models, with corrections, give comparable error metrics and have errors within 1−3 kcal/mol. The CBS model generally is the most reliable model, except for the S−H compounds, which are easily compensated by the BEC or LRM correction. The main purpose of presenting the BEC and LRM correction models is to analyze the nature of systematic errors in the various model chemistries. For molecules that are similar to those for which the corrections have been derived, it is expected that application of the corrections may lead to improved quantitative accuracy (coefficients in eqs 8 and 9 for the models are available in the Supporting Information). Although the model corrections lead to considerable improvement for the set of data from which they were derived, we do not advocate their use as general corrections. In the following sections, we provide the uncorrected (i.e., Raw) PA and GPB values in tables and discussion.

B. Amino Acids. Amino acids contain two protonation sites along the backbone (an amine and a carboxylic acid) and possibly one or two more on the side chain. In the gas phase, the amino acid backbone does not exist in isolation as a zwitterion as it does in solution.[72,73] It is known that the backbone titratable sites, isolated in the gas phase, are highly sensitive to the side chain conformation.[5,74] Here, we report model compounds for the amino acid backbone and side chains separately to reduce the conformational complexity[75] and focus more directly on the intrinsic affinity of each site for protonation. These model compounds are based on the amino acid side chain capped by a methyl group. Table 4 reports the PA and GPB for the amino acid model compounds as well as both titratable sites of glycine and proline to provide reference values for the backbone sites.

Overall, the quantum models are internally consistent and reliably predict the experimental values for the model compounds. The largest errors are for histidine, though most models

are still within the experimental error. For the arginine side chain model (*n*-methylguanidine), no experimental value was found. Hunter and Lias report a PA of 235.7 kcal/mol and a GPB of 226.9 kcal/mol for guanidine,[78] which is close to our calculated value (e.g., the calculated PA values ranged from 232.3 to 236.8 kcal/mol). Norberg et al. report a computational prediction for *n*-methylguanidine PA of 242−248 kcal/mol (MP2/6-31G*) but note that this structure is basis-set-dependent and that correlated methods provide lower values. Our quantum models diverge for this compound, with the DFT models predicting a significantly larger (∼3 kcal/mol) PA and GPB compared to those of the multilevel models, consistent with Norberg et al.'s suggestion.

Histidine values are provided both for the π (near backbone) and τ (far from backbone) nitrogen positions.[79] All of the models show a preference for deprotonation at the π position but only by ∼0.5 kcal/mol, which indicates that the system environment and solvation will dominate the site preference.

Figure 3 shows the GPB values plotted as a function the experimental pK_a value in solution. It is apparent that the GPB values do not correlate directly with the pK_a values. It is, nonetheless, noteworthy that while certain residues have very disparate pK_a values, they can have similar GPBs, and alternatively, similar pK_a values can have significantly different GPBs. For example, tyrosine and glutamic acid only differ by 3.4 kcal/mol (equivalent to 2.5 pK_a units) in the gas phase but are 6.4 pK_a units apart in solution, indicating that most of the difference is the preferential solvation of the propanate ion relative to the *p*-methylphenolate ion. Alternatively, glycinium's GPB is 12.3 kcal/mol (9 pK_a units) greater than prolinium, but the solution pK_a values differ by only 0.5 pK_a units. Moreover, in several instances, the basicity trend is reversed in the gas phase relative to that in solution. Examples include histidine and lysine, the latter of which is more basic in solution by 4.5 units but less basic in the gas phase. In this case, although the His side chain has a greater GPB value than that of Lys, the smaller primary ammonium of the protonated Lys side chain is much smaller
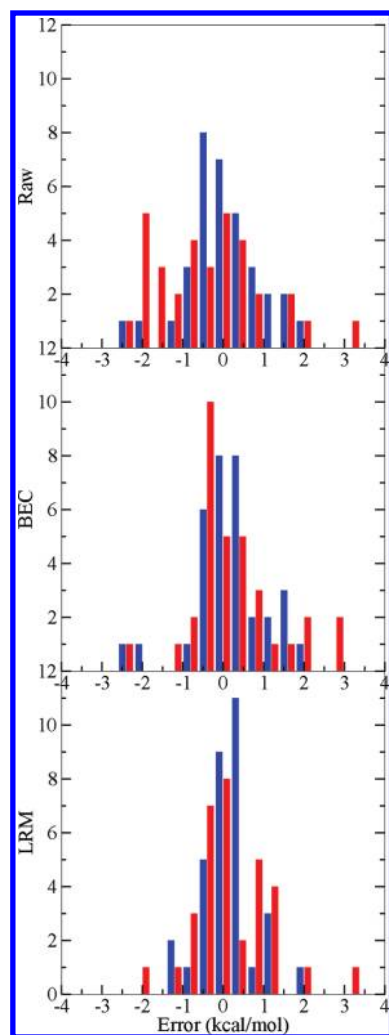
Figure 2. PA error histogram for the CBS (blue, left bar) and QCRNA (red, right bar) against experiment (top), with bond energy correction (middle) and linear regression model (bottom). See text for bond energy correction (BEC) and linear regression model (LRM) information. Histogram bins are 0.5 kcal/mol wide.

and better solvated than the imidazolium of the protonated His side chain. Analogously, Asp is more acidic than Cys in solution by 4.5 pK_a units but less acidic in the gas phase. The soft sulfur of Cys is able to more effectively accommodate the negative charge on the deprotonated side chain than the hard oxyacid group of Asp, but in solution, this larger size of the thiolate anion of Cys makes it less solvated than the carboxylate anion of Asp. There is great interest in the study of the protonation state of amino acids in proteins, the local dielectric of which may be considerably less than that of aqueous solution. Consequently, a quantitative understanding of the balance between intrinsic electronic effects and the effect of generalized solvation is critical.

C. Nucleic Acid Bases. The DNA and RNA bases each contain multiple protonation sites, but the most biologically relevant (titratable) sites are shown in Figure 4[80,81] along with their pK_a values.[82–84] Table 5 gives the PA and GPB values for protonation (+) and deprotonation (−) at selected sites in their standard tautomeric forms, along with available experimental values.[78,85]

The deviations between the calculated and experimental values for the nucleic acids are generally larger for the nucleobases than those for the amino acid side chains considered above, although there remains a high degree of correlation. In

particular, thymine and uracil have some of the largest deviations for the multilevel models, even after proton orientation and conformational averaging are considered. Additionally, CBS seems to underestimate the experimental values. Alternatively, if we compare our computation to experimental values derived from fast atom bombardment mass spectrometry provided by Greco et al.,[86] we see different levels of deviation depending on the models used to interpret the data and derive PA values. Wolken and Tureek suggest a PA of 205.1 kcal/mol for uracil,[87] which lies within the range of the calculations using multilevel models (204.4−205.5 kcal/mol) but is somewhat outside of the range of the DFT methods (206.4−207.4 kcal/mol). The large variation in experimentally determined PA and GPB values underscore the need for a systematic high-level quantum chemical characterization. Overall, the PA and GPB follow the trend

$$GUA > CYT > ADE > THY > URA$$

which is consistent with previous studies.[88,89]

Another important aspect in the consideration of protonation states of nucleobases is their tautomeric equilibria. The nucleobases can exist in different tautomeric forms (Figure 5), which are sometimes accessible in biological systems.[82] The enthalpy and free-energy values for the tautomerization reactions shown in Figure 5 are listed in Table 6. In the gas phase, the enol/imino forms for cytosine and guanine are the most accessible relative to the standard keto/amino forms. The calculated free-energy difference between keto and enol forms for guanine ranges from −0.4 to 1.1 kcal/mol, and the free-energy difference between amino and imino forms of cytosine ranges from 0.8 to 2.1 kcal/mol. The protonation state for these nucleobases may have implications in RNA enzymes as guanine and cytosine have been implicated as acting as general base and acid catalysts, respectively.[90] The remainder of the tautomerization free-energy values are more than 10 kcal/mol in the gas phase. A complete list of the PAs and GPBs of all protonation sites for DNA/RNA bases in both their keto and enol form as well as the tautomerization energies is provided in Supporting Information.

D. RNA Ribose Models. One of the defining differences between DNA and RNA is the RNA 2′ hydroxyl group. This group is a very weak acid but plays an important role in DNA and RNA cleavage and ligation[91] and has been implicated as a general acid catalyst in the hammerhead ribozyme[92] possibly activated by a divalent metal ion.[93] On the basis of experimental and computational estimates, the O2′ pK_a ranges between approximately 12.5 and 14.9.[19,94–96] This range is likely due to the local chemical environment (including a mild dependence on the attached nucleobase[95]), sugar pucker, and solution ionic strength.

To provide the most biologically relevant sugar pucker conformations, we have calculated the PA and GPB using C2′-endo ($P = 163.797$, $\tau_m = 34.495$) and C3′-endo ($P = 13.506$, $\tau_m = 37.747$) sugar puckers,[97] where P is the pseudorotation phase angle and τ_m is the degree of puckering. These conformations are based on X-ray diffraction data[98,99] and are consistent with canonical B and A form DNA and RNA.[82,100] Additionally, the dihedral synonymous to the ε dihedral (C4′−C3′−O2′−P) for the nucleic acid phosphodiester backbone was in the biological trans conformation.

Similar to the amino acids, use of the full ribose structure created conformational obstacles that obfuscate the intrinsic acidity of the 2′ oxygen. In particular, a hydroxyl group at the 3′ position created a significant intramolecular hydrogen bond

Proton Affinity and Gas-Phase Basicity and Biocatalysis

J. Phys. Chem. B, Vol. 114, No. 43, 2010 **13917**

TABLE 4: PAs and GPBs for Amino Acids Model Compounds with Experimental Values[b] (experimental error in parentheses) and pK_a Values[c] in Solution as a Reference[a]

model (amino acid)	CBS-QB3	G3B3	G3MP2B3	PBE0	B1B95	B3LYP	QCRNA	experiment	pK_a
				PA					
prolinium (Pro)	224.3	224.8	225.0	225.4	225.3	225.0	225.0	224.0 (2.0)	1.9
glycinium (Gly)	211.9	212.3	212.2	213.6	213.2	212.0	212.0	211.9 (0.5)	2.4
acetic acid (Asp)	347.4	348.2	348.6	347.6	347.9	345.8	346.4	347.2 (1.1)	3.9
propanic acid (Glu)	346.2	347.0	347.4	347.8	347.9	346.1	345.5	347.4 (1.8)	4.1
4-methylimidazolium (His π)	229.0	229.6	229.4	231.3	231.2	230.7	230.7	227.7 (2.0)	6.0
4-methylimidazolium (His τ)	229.6	230.3	230.1	231.8	231.7	231.3	231.3		6.0
methanethiol (Cys)	356.8	357.4	357.7	358.2	358.7	357.3	357.3	357.3 (1.2)	8.4
glycine (Gly)	341.8	342.6	343.0	343.0	343.2	341.5	341.5	340.3 (1.1)	9.8
p-methylphenol (Tyr)	350.4	350.7	350.7	350.0	350.6	348.9	348.9	350.7 (1.3)	10.5
propanammonium (Lys)	219.3	219.8	219.6	220.3	220.1	219.4	219.3	219.4 (0.5)	10.5
proline (Pro)	340.1	341.0	341.2	342.1	342.1	340.4	340.4	340.3 (3.1)	10.6
n-methylguanidine (Arg)	239.5	239.6	239.0	243.1	242.9	241.9	241.8		12.5
				GPB					
prolinium (Pro)	216.8	217.3	217.4	217.8	217.6	217.4	217.4	216.0 (2.0)	1.9
glycinium (Gly)	204.1	204.5	204.4	205.4	205.2	204.0	204.2	203.7 (0.5)	2.4
acetic acid (Asp)	339.4	340.2	340.6	341.3	341.5	339.5	338.3	341.4 (1.2)	3.9
propanic acid (Glu)	340.1	341.0	341.2	339.2	339.9	337.6	339.4	340.4 (1.4)	4.1
4-methylimidazolium (His π)	221.2	221.8	221.6	223.6	223.5	223.0	223.0	220.1 (2.0)	6.0
4-methylimidazolium (His τ)	221.9	222.6	222.4	224.2	224.0	223.7	223.6		6.0
methanethiol (Cys)	350.4	351.0	351.3	351.8	352.3	351.0	350.9	350.6 (2.0)	8.4
glycine (Gly)	334.7	335.6	335.9	335.8	336.0	334.2	334.3	335.1 (1.4)	9.8
p-methylphenol (Tyr)	343.2	343.4	343.4	342.8	343.2	341.8	341.6	343.8 (1.2)	10.5
propanammonium (Lys)	211.8	212.3	212.2	212.8	212.6	211.8	211.9	211.3 (0.5)	10.5
proline (Pro)	332.5	333.4	333.6	334.4	334.4	332.6	332.6	333.4 (3.0)	10.6
n-methylguanidine (Arg)	232.7	232.9	232.3	236.8	237.3	235.5	235.0		12.5

[a] All quantities are in kcal/mol except pK_a. [b] All experimental values are from NIST (ref 67), except for those for prolinium, which can be found in ref 76. [c] pK_a values are taken from ref 77.
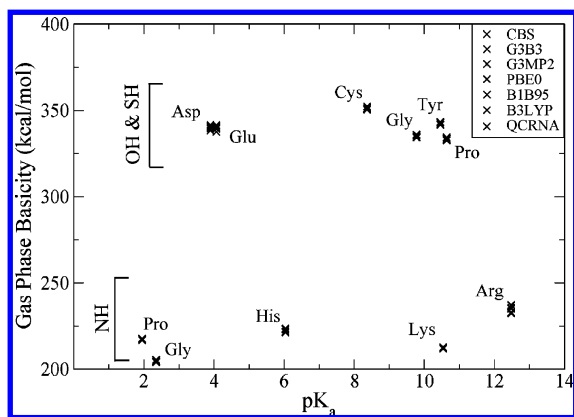

Figure 3. Calculated GPBs of amino acid model compounds versus pK_a values.
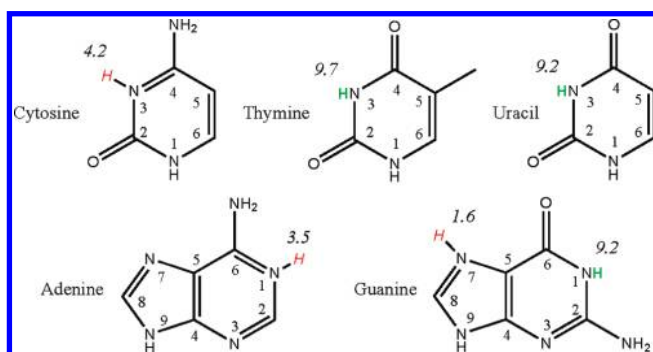

Figure 4. DNA/RNA base pair protonation points with pK_a values. Hydrogens in green are points of deprotonation. Hydrogens in red are points of protonation.

that strongly shifted the O2′ values. We consider here three model compounds for the RNA ribose (Figure 6), 2-hydroxytetrahydrofuran (THF), 2-hydroxy-1,3,4-trimethyltetrahydrofurn (methyl), and 2-hydroxy-1,4-dimethyl-3-methoxytetrahydrofuran (ribose). These successively more complicated models provide insight into how the ribose structure tunes the intrinsic PA and GPB of the 2′-hydroxyl group, with the ribose model being the most similar to the sugar ring of RNA.

Table 7 provides the PA and GPB for THF, methyl, and ribose models illustrated in Figure 6. Unlike the nucleic acid bases, the multilevel and DFT calculations have a more narrow distribution of values (typically around 2 kcal/mol) for a given model system. C3′-endo has a greater PA and GPB for all models compared to the C2′-endo sugar pucker conformation, and this gap increases with model complexity. The CBS GPB is 3.3 kcal/mol larger and more acidic for the ribose model in the C3′-endo conformation than that in the C2′-endo conformation. The addition of the 3′ methyl and then the 3′ methoxy group successively raises the PA and GPB of the 2′ hydroxyl group, making it less likely to release its proton. This is mainly due to the stabilization of the protonated structure via intramolecular hydrogen bonding, as shown by the difference with the THF model and the larger effect for C2′-endo conformation.

E. Phosphates. Metaphosphates, phosphates, and phosphoranes represent important states in phosphoryl transfer reactions.[91,101−105] Metaphosphates, for example, are presumed intermediates in dissociative phosphoryl transfer mechanisms for phosphate monoesters in solution.[101,106−110] Acyclic and cyclic phosphates and cyclic phosphoranes are models for different states along the reaction path of RNA transesterification, hydrolysis, and ligation.[91,103,105] Thio substitutions at native phosphoryl oxygen positions are useful experimental probes that give insight into catalytic mechanism in ribozymes.[111,112] The interpretation of thio effect experiments can be aided by computational techniques,[55,113] and hence, their high-level quantum characterization is of considerable interest.

TABLE 5: PA (top) and GPB (bottom) Values for DNA and RNA Bases[a]

base[b]	CBS	G3B3	G3MP2	PBE0	B1B95	B3LYP	QCRNA	expt.
				PA				
A:N 1(+)	223.6	224.9	224.7	227.1	227.1	226.7	226.8	225.3
C:N 3(+)	226.9	227.6	227.4	229.5	229.3	228.9	228.8	227.0
G:N 7(+)	227.6	228.2	227.9	231.0	230.8	230.5	230.5	229.3
G:N 1(−)	337.7	338.2	338.0	340.9	341.0	339.9	339.9	
T:O 4(+)	206.1	207.1	207.3	209.2	209.3	208.1	208.2	210.5
T:N 3(−)	346.3	347.0	346.9	348.2	348.1	347.1	346.9	
U:O 4(+)	204.4	205.4	205.5	207.3	207.4	206.4	206.4	208.6
U:N 3(−)	346.3	347.0	346.9	348.2	348.1	347.1	346.9	
				GPB				
A:N 1(+)	216.3	217.1	217.0	218.3	218.4	218.2	218.3	218.1
C:N 3(+)	218.9	219.8	219.7	220.9	220.2	220.4	219.9	219.0
G:N 7(+)	220.2	220.9	220.6	223.6	223.3	223.0	223.0	221.7
G:N 1(−)	330.2	330.6	330.4	333.3	333.5	332.4	332.3	
T:O 4(+)	198.4	199.3	199.5	201.4	201.5	200.4	200.5	203.2
T:N 3(−)	338.5	339.1	339.1	340.4	340.4	339.3	339.0	
U:O 4(+)	196.7	197.6	197.8	199.6	199.6	198.6	198.7	201.2
U:N 3(−)	338.1	338.6	338.5	339.9	340.0	338.7	338.4	

[a] The uncertainty of the experimental values is ~2.0 kcal/mol.[78] All quantities are in kcal/mol. [b] Base identifies the nucleic acid base (A, C, G, T, or U) and the position that the proton is being removed (−) or added (+).

Figure 5. DNA/RNA base pair tautomers.

TABLE 6: Enthalpy (top) and Free Energy (bottom) of Tautomerization for DNA and RNA Base Pairs in the Gas Phase[a]

	CBS	G3B3	G3MP2	PBE0	B1B95	B3LYP	QCRNA
				Enthalpy			
adenine	12.5	11.7	11.4	12.3	12.2	11.9	11.9
cytosine	1.2	1.0	0.8	2.4	2.4	2.1	2.1
guanine	−0.4	−0.2	−0.3	0.7	0.5	1.1	1.1
thymine	12.0	11.8	11.6	12.3	11.9	12.5	12.5
uracil	11.1	10.9	10.7	11.3	10.9	11.6	11.5
				Free Energy			
adenine	12.3	12.0	11.7	13.6	13.4	13.0	12.8
cytosine	1.5	1.2	1.0	3.4	3.8	3.0	3.4
guanine	−0.3	0.0	−0.1	0.8	0.6	1.2	1.2
thymine	12.2	12.0	11.8	12.5	12.0	12.7	12.6
uracil	11.2	11.1	10.8	11.4	11.1	11.7	11.6

[a] All quantities are in kcal/mol.

methods. PA and GPB show similar trends, and values for all of the quantum models can be found in the Supporting Information. The quantitative discussion and analysis here will

Figure 6. Neutral and deprotonated model compounds for D-ribofuranose sugar; THF (2-hydroxytetrahydrofuran), methyl (2-hydroxy-1,3,4-trimethyltetrahydrofuran), and ribose (2-hydroxy-1,4-dimethyl-3-methoxytetrahydrofuran).

To fully investigate the protonation states of the phosphate moiety, we provide benchmark values for metaphosphates, acyclic and cyclic phosphates, and phosphoranes, each with thio substitutions (see Supporting Information for full tables). Figure 7 provides the nomenclature for the different compounds, and Table 8 reports the PA values using the CBS and QCRNA

Proton Affinity and Gas-Phase Basicity and Biocatalysis

J. Phys. Chem. B, Vol. 114, No. 43, 2010 **13919**

TABLE 7: PA (top) and GPB (bottom) Values for RNA-like Sugar Molecules THF, Methyl, and Ribose[a]

pucker	model	CBS	G3B3	G3MP2	PBE0	B1B95	B3LYP	QCRNA
				PA				
C2'[b]	THF	369.2	369.1	370.1	370.2	370.6	368.2	368.3
	methyl	366.4	367.0	367.4	367.3	367.4	365.6	365.7
	ribose	360.4	361.2	361.6	362.9	363.2	360.8	361.5
C3'[c]	THF	370.2	370.9	371.2	370.8	371.3	368.7	368.8
	methyl	367.7	368.3	368.7	368.2	368.6	366.4	366.5
	ribose	364.9	365.7	366.1	365.6	365.8	363.9	363.9
				GPB				
C2'[b]	THF	361.2	360.9	362.0	362.1	362.4	360.1	360.1
	methyl	358.0	358.6	359.0	358.7	358.8	357.1	357.2
	ribose	354.5	355.4	355.8	354.9	355.1	355.1	353.5
C3'[c]	THF	362.9	363.5	363.9	363.6	364.1	361.4	361.5
	methyl	360.3	360.9	361.3	360.8	361.2	359.0	359.2
	ribose	357.8	358.6	359.0	358.5	358.7	356.8	356.8

[a] Figures in Supporting Information. All molcules are given in conformations consistent with B-form and A-form DNA/RNA. All quantities are in kcal/mol. [b] $P = 163.797$; $\tau_m 4 = 34.495$. [c] $P = 13.506$; $\tau_m = 37.747$.
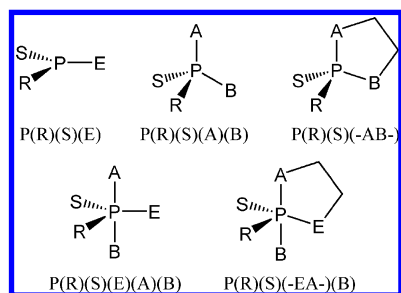
Figure 7. Nomenclature convention for ligand designations in metaphosphate, acyclic and cyclic phosphate, and phosphorane compounds of biological interest. This nomenclature is consistent with the naming convention used for similar compounds in previous work.[7,53,71]

be largely restricted to the CBS PA data, although the qualitative insights thus provided are more generally applicable.

We report a PA for metaphosphate of 310.5 kcal/mol using the CBS method, in excellent agreement with the experimental value of 310.8 ± 2.0 kcal/mol.[67] For the phosphates, recall from Table 1 that the errors in the PA calculation for phosphoric acid and dimethyl phosphate are -2.6 and -2.4 kcal/mol, respectively, which is well within the 5.0 and 4.1 kcal/mol error bars from the experimental data. These results specifically, along with the general successes noted above, indicate that the predictions presented in Table 8 and the Supporting Information are of similar quality to those that might be obtained by further experiments. Additionally, these calculations provide values for systems that may be very difficult to probe directly via experiment.

Single sulfur substitution (not at the protonation site) lowers the PA of metaphosphates by 3.0 kcal/mol, whereas substitution at the protonation site lowers the PA by 6.0 kcal/mol. The corresponding double sulfur substitutions lower the PA values further by 3.3 and 7.2 kcal/mol, respectively.

The PA values for the cyclic phosphates are quite similar to those of the acyclic analogues; for example, the PA values for dimethyl phosphate and ethylene phosphate differ by only 0.2 kcal/mol. Sulfur substitution at the site of protonation has the largest effect on the PA, lowering it from the corresponding unsubstituted phosphate by 8.4 to 10.9 kcal/mol. Sulfur substitution at an unprotonated nonbridge position has the smallest effect, lowering the PA relative to the unsubstituted phosphate by 4.1 to 6.5 kcal/mol. It should be noted that this large effect on the PA value is likely to have considerably less effect on the corresponding pK_a values in solution due to the reduced solvation of the larger sulfur atom.

The protonation state of phosphoranes is of considerable interest as it affects the lifetime of intermediates in phosphoryl

TABLE 8: Predicted PAs for Metaphosphate, Phosphate, Cyclic Phosphate (left columns), and Acyclic and Cyclic Phosphorance Compounds (right column) of Biological Interest[a]

molecule[b]	CBS	QCRNA	molecule[b]	CBS	QCRNA
P(O)(O)(OH)	310.5	311.1	P(OH*)(OH)(OH)(OH)(OH)	340.8	339.3
P(O)(O)(SH)	304.5	306.2	P(OH)(OH)(OH)(OH)(OH*)	350.8	350.5
P(S)(O)(OH)	307.5	308.5	P(OH*)(OH)(−O−CH₂CH₂−O−)(OH)	343.1	342.6
P(S)(S)(OH)	307.2	308.6	P(OH)(OH*)(−O−CH₂CH₂−O−)(OH)	343.4	342.9
P(S)(O)(SH)	303.3	305.2	P(OH*)(OH)(−O−CH₂CH₂−O−)(OCH₃)	343.8	343.0
P(O)(OH)(OH)(OH)	327.9	328.2	P(OH)(OH*)(−O−CH₂CH₂−O−)(OCH₃)	343.4	342.9
P(O)(O)(OH)(OH)⁻	458.7	457.8	P(OH*)(OH)(−O−CH₂CH₂−S−)(OH)	335.2	335.0
P(O)(O)(O)(OH)²⁻	580.9	579.5	P(OH)(OH*)(−O−CH₂CH₂−S−)(OH)	335.4	335.0
P(S)(OH)(OH)(OH)	322.5	322.9	P(OH*)(OH)(−O−CH₂CH₂−S−)(OCH₃)	331.8	334.5
P(O)(OH)(OH)(SH*)	318.2	319.9	P(OH)(OH*)(−O−CH₂CH₂−S−)(OCH₃)	332.1	334.4
P(O)(OH)(OH*)(SH)	320.9	320.9	P(OH*)(OH)(-S-CH₂CH₂−O−)(OH)	336.1	336.2
P(O)(OCH₃)(OH)(OH)	330.0	330.3	P(OH)(OH*)(−S−CH₂CH₂−O−)(OH)	336.0	336.2
P(O)(OCH₃)(O)(OH)−	453.9	452.9	P(OH*)(OH)(−S−CH₂CH₂−O−)(OCH₃)	339.0	338.6
P(S)(OCH₃)(OH)(OH)	323.5	324.1	P(OH)(OH*)(−S−CH₂CH₂−O−)(OCH₃)	339.6	338.8
P(S)(OCH₃)(O)(OH)−	437.9	437.5	P(OH*)(OH)(−O−CH₂CH₂−O−)(SCH₃)	333.1	333.0
P(O)(OCH₃)(OH)(SH*)	319.1	321.4	P(OH)(OH*)(−O−CH₂CH₂−O−)(SCH₃)	333.1	333.0
P(O)(OCH₃)(OH*)(SH)	321.6	322.1	P(SH*)(OH)(−O−CH₂CH₂−O−)(OH)	333.6	335.0
P(O)(SCH₃)(OH)(OH)	322.2	322.8	P(SH)(OH*)(−O−CH₂CH₂−O−)(OH)	338.4	337.7
P(O)(SCH₃)(O)(OH)−	443.7	441.0	P(SH*)(OH)(−O−CH₂CH₂−O−)(OCH₃)	335.3	336.8
P(O)(OCH₃)(OCH₃)(OH)	329.2	330.1	P(SH)(OH*)(−O−CH₂CH₂−O−)(OCH₃)	339.1	338.2
P(S)(OCH₃)(OCH₃)(OH)	325.1	325.9	P(OH*)(SH)(−O−CH₂CH₂−O−)(OH)	337.3	336.7
P(O)(SCH₃)(OCH₃)(OH)	324.2	326.2	P(OH)(SH*)(−O−CH₂CH₂−O−)(OH)	332.6	334.1
P(O)(OCH₃)(OCH₃)(SH)	320.8	323.2	P(OH*)(SH)(−O−CH₂CH₂−O−)(OCH₃)	341.1	340.1
P(O)(OH)(−O−CH₂CH₂−O−)	329.4	329.5	P(OH)(SH*)(−O−CH₂CH₂−O−)(OCH₃)	335.9	336.8
P(O)(SH)(−O−CH₂CH₂−O−)	320.1	321.9			
P(S)(OH)(−O−CH₂CH₂−O−)	324.0	324.5			
P(O)(OH)(−S−CH₂CH₂−O−)	324.1	324.7			

[a] All values are in kcal/mol. See Figure 7 for the naming scheme. [b] "Molecule" refers to AH in eq 1.

transfer reactions and could lead to formation of alternate reaction products such as in RNA migration.[104] Dianionic phosphoranes are generally believed to exist as high-energy intermediates or else true transition states. Further, it is known from Westheimer's rules that anionic ligands prefer to occupy equatorial positions[114] and serve as pivotal positions in pseudorotations.[51,53] Indeed, in the present work, we find that the differences in PA for deprotonation at the equatorial and apical positions differ by 10 kcal/mol for pentahydroxyphosphorane with the CBS model (Table 8). We have endeavored to provide a fairly comprehensive set of native and thio-substituted phosphoranes in neutral and monoanionic forms in the Supporting Information. Here, we restrict discussion to the PA values in only the equatorial positions for a series of native and thio-substituted cyclic phosphoranes that represent models for reactive intermediates in RNA transesterification and hydrolysis reactions (Table 8). It should be noted that the pro-R and pro-S nonbridging oxygens are not equivalent in many biological systems.[115,116] In the present model systems, the pro-R and pro-S nonbridging positions are not necessarily equivalent for a given conformational minima of the puckered five-membered ring. In Table 8, we list PA values for both the pro-R and pro-S positions for a given conformational minima, the values for which differ by typically only 0.5 kcal/mol or less.

The cyclic oxyphosphoranes have PA values between 343.1 and 343.8 kcal/mol and are 2−3 kcal/mol larger than those for the acyclic pentahydroxyphosphorane reference of 340.8 kcal/mol. Sulfur substitution at the nonbridge position decreases the PA values by around 4.5 kcal/mol relative to the native cyclic phosphorane, and sulfur substitution at the nonbridge deprotonation site decreases the PA value by around 8−9.5 kcal/mol. Substitution at the equatorial and apical positions within the ring decreases the PA values by about 3.8−7.2 and 8.0−11.7 kcal/mol, respectively. Substitution at the acyclic apical position decreases the PA by around 12 kcal/mol.

IV. Conclusion

Benchmark calculations of the PA and GPB for oxygen-, sulfur-, and nitrogen-containing compounds are compared for a variety of multilevel and density functional quantum methods against experimental values. These methods are corrected using a BEC and LRM to evaluate the quantum methods' performance. Here, we report that while all of the tested quantum chemistries provide reliable predictions, the CBS-QB3 method in general is the most reliable.

These models are then used to predict the PA and GPB for approximately 150 different molecules of biological interest comprising more than 2000 quantum calculations. These systems include the amino acid backbone and side chains, nucleic acids in two tautomeric forms, ribose-like structures for RNA sugar's O2′, and various metaphosphates, cyclic and acyclic phosphates, and phosphoranes, all with and without thio substitution.

This work provides a consistent database of PA and GPB to be used in pK_a prediction, biochemical evaluation of protonation/deprotonation events, parametrization of improved semiempirical quantum models used in QM/MM biocatalysis simulation, and comparison to experimental PA and GPB determination.

References and Notes

(1) Silverman, R. B., *The Organic Chemistry of Enzyme-Catalyzed Reactions*; Academic Press: San Diego, CA, 2000.
(2) Alexeev, Y.; Windus, T.; Zhan, C.-G.; Dixon, D. *Int. J. Quantum Chem.* **2005**, *102*, 775.
(3) Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. *J. Phys. Chem. A* **2004**, *108*, 166.
(4) Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126*, 814.
(5) Hudáky, P.; Perczel, A. *J. Phys. Chem. A* **2004**, *108*, 6195.
(6) Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717.
(7) Range, K.; McGrath, M. J.; Lopez, X.; York, D. M. *J. Am. Chem. Soc.* **2004**, *126*, 1654.
(8) Range, K.; López, C. S.; Moser, A.; York, D. M. *J. Phys. Chem. A* **2006**, *110*, 791.
(9) Jang, Y. H.; Goddard, W. A., III; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *J. Phys. Chem. B* **2003**, *107*, 344.
(10) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A* **2003**, *107*, 9380.
(11) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963.
(12) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.
(13) Klicić, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327.
(14) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010.
(15) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.
(16) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.
(17) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727.
(18) Chen, I.-J.; MacKerell, A. D., Jr. *Theor. Chem. Acc.* **2000**, *103*, 483.
(19) Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 166.
(20) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.
(21) Yazal, J. E.; Prendergast, F. G.; Shaw, D. E.; Pang, Y.-P. *J. Am. Chem. Soc.* **2000**, *122*, 11411.
(22) Peräkylä, M. *Phys. Chem. Chem. Phys.* **1999**, *1*, 5643.
(23) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.
(24) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.
(25) Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, L.; Case, D. A. *Int. J. Quantum Chem.* **1997**, *61*, 207.
(26) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2004**, *55*, 689.
(27) Jensen, J. H.; Li, H.; Robertson, A. D.; Molina, P. A. *J. Phys. Chem. A* **2005**, *109*, 6634.
(28) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715.
(29) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253.
(30) Pliego Jr., J. R. *Chem. Phys. Lett.* **2003**, *367*, 145.
(31) McNaught, A. D.; Wilkinson, A. *Compendium of Chemical Terminology: IUPAC Recommendations*, 2nd ed.; Blackwell Science, Inc.: Oxford, U.K., 1997; http://www.iupac.org/publications/compendium/index.html.
(32) Nam, K.; Cui, Q.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2007**, *3*, 486.
(33) Yang, Y.; Yu, H.; York, D. M.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861.
(34) Nam, K.; Gao, J.; York, D. M. In *Multiscale Simulation Methods for Nanomaterials*; Ross, R. B., Sanat, M., Eds.; Wiley: New York, 2008; pp 201−218.
(35) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, *111*, 4190.
(36) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760.
(37) Śmiechowski, M. *J. Mol. Struct.* **2009**, *924*, 170.
(38) Liu, S.; Pedersen, L. G. *J. Phys. Chem. A* **2009**, *113*, 36.

Proton Affinity and Gas-Phase Basicity and Biocatalysis

J. Phys. Chem. B, Vol. 114, No. 43, 2010 **13921**

(39) Khandogin, J.; Brooks, C. L., III. *Biochemistry* **2006**, *45*, 9363.

(40) Chen, J.; Brooks, C. L., III; Khandogin, J. *Curr. Opin. Struct. Biol.* **2008**, *18*, 140.

(41) Marti, M. A.; Gonzalez Lebrero, M. C.; Roitberg, A.; Estrin, D. A. *J. Am. Chem. Soc.* **2008**, *130*, 1611.

(42) Click, T. H.; Kaminski, G. A. *J. Phys. Chem. B* **2009**, *113*, 7844.

(43) Machuqueiro, M.; Baptista, A. M. *J. Am. Chem. Soc.* **2009**, *131*, 12586.

(44) Swails, J. M.; Meng, Y.; Walker, F. A.; Marti, M. A.; Estrin, D. A.; Roitberg, A. E. *J. Phys. Chem. B* **2009**, *113*, 1192.

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(46) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(47) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.

(48) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650.

(49) Pokon, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, *105*, 10483.

(50) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(51) López, C. S.; Faza, O. N.; Gregersen, B. A.; Lopez, X.; de Lera, A. R.; York, D. M. *ChemPhysChem* **2004**, *5*, 1045.

(52) Mayaan, E.; Range, K.; York, D. M. *J. Biol. Inorg. Chem.* **2004**, *9*, 807.

(53) López, C. S.; Faza, O. N.; de Lera, A. R.; York, D. M. *Chem.−Eur. J.* **2005**, *11*, 2081.

(54) Liu, Y.; Lopez, X.; York, D. M. *Chem. Commun.* **2005**, *31*, 3909.

(55) Liu, Y.; Gregersen, B. A.; Lopez, X.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 19987.

(56) Liu, Y.; Gregersen, B. A.; Hengge, A.; York, D. M. *Biochemistry* **2006**, *45*, 10043.

(57) Giese, T. J.; Gregersen, B. A.; Liu, Y.; Nam, K.; Mayaan, E.; Moser, A.; Range, K.; Nieto Faza, O.; Silva Lopez, C.; Rodriguez de Lera, A.; Schaftenaar, G.; Lopez, X.; Lee, T.; Karypis, G.; York, D. M. *J. Mol. Graphics Modell.* **2006**, *25*, 423.

(58) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(59) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(60) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(61) Adamo, C.; Scuseria, G. E. *J. Chem. Phys.* **1999**, *111*, 2889.

(62) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(63) Becke, A. D. *J. Chem. Phys.* **1996**, *104*, 1040.

(64) Ochterski, J. W. Vibrational Analysis in Gaussian. http://gaussian.com/g_whitepap/vib.htm (accessed March 2005).

(65) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.; John Wiley & Sons: Chichester, England, 2002.

(66) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Mill Valley, CA, 1973.

(67) Linstrom, P., Mallard, W., Eds. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; National Institute of Standards and Technology: Gaithersburg, MD, 2003; http://webbook.nist.gov.

(68) Lioe, H.; O'Hair, R. A.; Gronert, S.; Austin, A.; Reid, G. E. *Int. J. Mass Spectrom.* **2007**, *267*, 220.

(69) QCRNA. http://theory.rutgers.edu/Database/QCRNA.

(70) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.

(71) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070.

(72) Tortonda, F. R.; Pascual-Ahuir, J. L.; Silla, E.; Tuñón, I. *Chem. Phys. Lett.* **1996**, *260*, 21.

(73) Yang, G.; Zu, Y.; Liu, C.; Fu, Y.; Zhou, L. *J. Phys. Chem. B* **2008**, *112*, 7104.

(74) Jones, C. M.; Bernier, M.; Carson, E.; Colyer, K. E.; Metz, R.; Pawlow, A.; Wischow, E. D.; Webb, I.; Andriole, E. J.; Poutsma, J. C. *Int. J. Mass Spectrom.* **2007**, *267*, 54.

(75) Kovačevic, B.; Rožman, M.; Klasinc, L.; Srzić, D.; Maksić, Z. B.; Yáñez, M. *J. Phys. Chem. A* **2005**, *109*, 8329.

(76) Kuntz, A. F.; Boynton, A. W.; David, G. A.; Colyer, K. E.; Poutsma, J. C. *J. Am. Soc. Mass Spectrom.* **2002**, 72–81.

(77) Voet, D.; Voet, J.; Pratt, C. *Fundamentals of Biochemistry*; John Wiley & Sons, Inc.: New York, 1999.

(78) Hunter, E.; Lias, S. *J. Phys. Chem. Ref. Data* **1998**, *27*, 413.

(79) McNaught, A. D., Wilkinson, A., Eds. *IUPAC Compendium of Chemical Terminology (the Gold Book)*, 2nd ed.; Blackwell Scientific Publications: Oxford, U.K., 1997.

(80) Lee, J. *Int. J. Mass Spectrom.* **2005**, *240*, 261.

(81) Donna, L. D.; Napoli, A.; Sindona, G. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 1080.

(82) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I., Jr. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.

(83) Shapiro, R.; Danzig, M. *Biochemistry* **1972**, *11*, 23.

(84) Lippert, B. Alterations of Nucleobase pK_a Values upon Metal Coordination: Origins and Consequences. *Progress in Inorganic Chemistry*; John Wiley and Sons: New York, 2005; Vol. 54, pp 385−445.

(85) Chen, E. C.; Herder, C.; Chen, E. S. *J. Mol. Struct.* **2006**, *798*, 126.

(86) Greco, F.; Liguori, A.; Sindona, G.; Uccella, N. *J. Am. Chem. Soc.* **1990**, *112*, 9092.

(87) Wolken, J. K.; Tureek, F. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1065.

(88) Russo, N.; Toscano, M.; Grand, A.; Jolibois, F. *J. Comput. Chem.* **1998**, *19*, 989.

(89) Podolyan, Y.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 7346.

(90) Scott, W. G. *Curr. Opin. Struct. Biol.* **2007**, *17*, 280.

(91) Lilley, D. M.; Eckstein, F., Eds. *Ribozymes and RNA Catalysis*; RSC Biomolecular Series; RSC Publishing: Cambridge, U.K., 2008.

(92) Martick, M.; Lee, T.-S.; York, D. M.; Scott, W. G. *Chem. Biol.* **2008**, *15*, 332.

(93) Lee, T.-S.; Silva Lopez, C.; Giambasu, G. M.; Martick, M.; Scott, W. G.; York, D. M. *J. Am. Chem. Soc.* **2008**, *130*, 3053.

(94) Li, Y.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364.

(95) Velikyan, I.; Acharya, S.; Trifonova, A.; Földesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2001**, *123*, 2893.

(96) Acharya, S.; Földesi, A.; Chattopadhyaya, J. *J. Org. Chem.* **2003**, *68*, 1906.

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

(98) Arnott, S.; Hukins, D. *Biochem. J.* **1972**, *130*, 453.

(99) Arnott, S.; Hukins, D. W. L. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 1504.

(100) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(101) Hengge, A. C. *Acc. Chem. Res.* **2002**, *35*, 105.

(102) Hengge, A. C. *Adv. Phys. Org. Chem.* **2005**, *40*, 49.

(103) Swamy, K. C. K.; Kumar, N. S. *Curr. Sci.* **2003**, *85*, 1256.

(104) Perreault, D. M.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **1997**, *36*, 432.

(105) Bevilacqua, P. C. Proton Transfer in Ribozyme Catalysis. *Ribozymes and RNA Catalysis*; RSC Publishing: Cambridge, U.K., 2008; pp 11−36.

(106) Wolfenden, R.; Ridgway, C.; Young, G. *J. Am. Chem. Soc.* **1998**, *120*, 833.

(107) Allen, K. N.; Dunaway-Mariano, D. *Trends Biochem. Sci.* **2004**, *29*, 495.

(108) Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 719.

(109) Khandogin, J.; Gregersen, B. A.; Thiel, W.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 9799.

(110) Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2.

(111) Frey, P. A.; Sammons, R. D. *Science* **1985**, *228*, 541.

(112) Herschlag, D.; Piccirilli, J. A.; Cech, T. R. *Biochemistry* **1991**, *30*, 4844.

(113) Gregersen, B. A.; Khandogin, J.; Thiel, W.; York, D. M. *J. Phys. Chem. B* **2005**, *109*, 9810.

(114) Westheimer, F. H. *Acc. Chem. Res.* **1968**, *1*, 70.

(115) Zhou, D.-M.; He, Q.-C.; Zhou, J.-M.; Taira, K. *FEBS Lett.* **1998**, *431*, 154.

(116) Allin, C.; Gerwert, K. *Biochemistry* **2001**, *40*, 3037.