RESEARCH PAPER

PCCP

www.rsc.org/pccp

Benchmark calculations of proton affinities and gas-phase basicities of molecules important in the study of biological phosphoryl transfer

Kevin Range,[a] Demian Riccardi,[b] Qiang Cui,[b] Marcus Elstner[c] and Darrin M. York*[a]

[a] *Department of Chemistry, University of Minnesota, 207 Pleasant St. SE, Minneapolis, MN 55455-0431, USA*

[b] *Theoretical Chemistry Institute, University of Wisconsin, 1101 University Avenue, Madison, Wisconsin 53706, USA*

[c] *Department of Theoretical Physics, University of Paderborn, D-33098 Paderborn, Germany, German Cancer Research Center, Department of Molecular Biophysics, D-60120 Heidelberg, Germany*

Received 14th April 2005, Accepted 20th June 2005
First published as an Advance Article on the web 7th July 2005

Benchmark calculations of proton affinities and gas-phase basicities of molecules most relevant to biological phosphoryl transfer reactions are presented and compared with available experimental results. The accuracy of proton affinity and gas-phase basicity results obtained from several multi-level model chemistries (CBS-QB3, G3B3, and G3MP2B3) and density-functional quantum models (PBE0, B1B95, and B3LYP) are assessed and compared. From these data, a set of empirical bond enthalpy, entropy, and free energy corrections are introduced that considerably improve the accuracy and predictive capability of the methods. These corrections are applied to the prediction of proton affinity and gas-phase basicity values of important biological phosphates and phosphoranes for which experimental data does not currently exist. Comparison is made with results from semiempirical quantum models that are commonly employed in hybrid quantum mechanical/molecular mechanical simulations. Data suggest that the design of improved semiempirical quantum models with increased accuracy for relative proton affinity values is necessary to obtain quantitative accuracy for phosphoryl transfer reactions in solution, enzymes, and ribozymes.

1. Introduction

Biological phosphoryl transfer reactions[1,2] are of considerable importance in processes of cell signaling,[3,4] energy transfer, and RNA catalysis.[2] In particular, the mechanisms whereby RNA is hydrolyzed (Scheme 1) in solution and by enzymes and ribozymes is an area of significant current interest and the focus of a tremendous amount of experimental and theoretical work.[5,6] Consequently, the elucidation of the key factors that govern the reactivity of biological phosphates is of fundamental importance.

One factor that strongly influences biological phosphate stability and reaction mechanisms is the protonation state of the nucleophile/leaving group and of the phosphoryl oxygens at various steps along the reaction.[6] The lifetime of phosphorane intermediates, for example, depend critically on protonation state.[5] Dianionic phosphoranes are transient reactive species, whereas monoanionic and neutral phosphoranes may have sufficient lifetimes to undergo pseudorotation.[7,8] In solution, pseudorotation can be inferred by the uptake of ^{18}O of ethylene phosphate in isotopically labeled water. In RNA systems, pseudorotation and subsequent hydrolysis gives rise to RNA migration.[5]

Enzymes and ribozymes often display complex pH-dependent rate profiles that are suggestive of certain general mechanisms. The hydrolysis of biological phosphates such as RNA is difficult to study experimentally due to the inherent stability of the phosphates under physiological conditions. Consequently, much of the experimental data on non-enzymatic phosphate hydrolysis are derived from phosphate analogs with enhanced leaving groups. Brønsted and Hammett correlation parameters[9] are often used as indexes to characterize phosphoryl transfer reactions. Of particular interest in phosphoryl transfer reactions is the correlation parameter, β_{lg} (lg = "leaving group")[10]

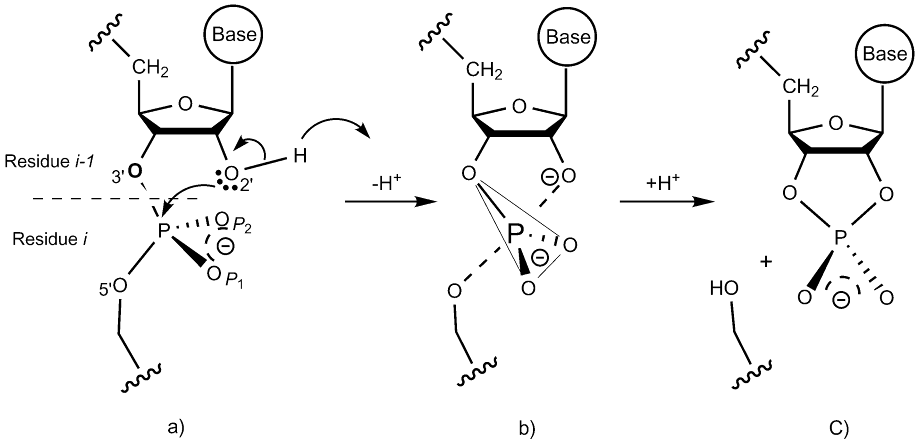
$$\beta_{\mathrm{lg}} = \partial \log k / \partial \mathrm{p}K_\mathrm{a} \qquad (1)$$

where "log k" is the logarithm of the rate constant k for hydrolysis and pK_a is the leaving group pK_a value. Values of β_{lg} derived from the slope of log k *versus* pK_a for a series of leaving groups provide information about the degree of P–O bond cleavage and formation in the transition state.[5]

It is clear that protonation state and its relation to pK_a values and pH-dependent rate profiles is a central consideration in the study of biological phosphoryl transfer in both non-enzymatic and enzymatic systems. However, to date, the direct experimental measurement of pK_a values for reactive intermediates in biological phosphoryl transfer reactions, such as metaphosphates and phosphoranes, has been elusive, due to the transient nature of these species in solution. Estimates of pK_a values for phosphoranes have been recently made based on theory[11–13] and experiment.[13]

The theoretical prediction of pK_a values from quantum chemistry presents tremendous challenges due to the subtle differences in free energy values that give rise to pK_a shifts (a pK_a unit corresponds to 1.364 kcal mol^{-1} of energy at 298.15 K). Several recent studies have addressed the prediction of pK_a values from electronic structure and implicit/explicit solvation methods.[11,12,14–34] The accurate prediction of absolute pK_a values is made using a thermodynamic cycle such as the one shown in Scheme 2. A key quantity in the cycle is the gas-phase

DOI: 10.1039/b504941e



Scheme 1 RNA transesterification. In "a", the 2′ ribose hydroxyl group is deprotonated and attacks the adjacent phosphate, forming a phosphorane transition state/intermediate ("b"), which goes on to form the exocyclic cleavage product ("c") resulting in cleavage of the RNA phosphate backbone.

basicity that involves deprotonation of the species of interest in the gas phase (ΔG_{gas} in Scheme 2). Related to the gas-phase basicity is the proton affinity, which is the enthalpic component of the same process.[35]

The purpose of the present work is to benchmark quantum electronic structure methods against experiment for calculation of proton affinities and gas-phase basicities of molecules most relevant to phosphoryl transfer reactions, and use these results to provide accurate proton affinity and gas-phase basicity values for biological phosphates and phosphoranes for which experimental values do not exist. A test set of molecules that represent important nucleophiles, leaving groups, and phosphorus compounds was assembled for which the experimental proton affinities and gas-phase basicities have been determined. A series of high-level quantum model chemistries was then applied to the test set in order to benchmark the accuracy of the calculations. From these data, a set of empirical bond enthalpy, entropy, and free energy corrections were determined that improve the accuracy and predictive capability of the methods. These corrections were applied to the prediction of proton affinities and gas-phase basicities of important biological phosphates and phosphoranes for which experimental data does not exist. Finally, comparison is made with semiempirical quantum models that are commonly employed in hybrid quantum mechanical/molecular mechanical (QM/MM) simulations of phosphoryl transfer reactions in enzymes and ribozymes.

The results of the present work are instrumental for: (1) the development of new semiempirical quantum models for combined QM/MM calculations of biological phosphoryl transfer reactions, (2) the determination of solvation free energies of phosphates with known pK_a values, and (3) the prediction of pK_a values of biological phosphates and phosphoranes in solution and in enzymes and ribozymes for which experimental data is not available.

2. Methods

2.1 Electronic structure calculations

All electronic structure and thermochemical analyses were performed using the Gaussian03 suite of programs.[36] Three

$$\begin{tikzcd}
AH_{(g)} \arrow[r, "\Delta G_{gas}"] \arrow[d, "\Delta G_{sol}(AH)"] & A^-_{(g)} \arrow[d, "\Delta G_{sol}(A^-)"] & + & H^+_{(g)} \arrow[d, "\Delta G_{sol}(H^+)"] \\
AH_{(aq)} \arrow[r, "\Delta G_{aq}"] & A^-_{(aq)} & + & H^+_{(aq)}
\end{tikzcd}$$

Scheme 2 A thermodynamic cycle for calculating pK_a values.[26]

"multi-level" methods were studied: CBS-QB3,[37,38] G3B3,[39] and G3MP2B3.[39] CBS-QB3 is a multi-level model chemistry that combines the results of several electronic structure calculations and empirical terms to predict molecular energies to around 1 kcal mol^{-1} accuracy.[40] The required electronic structure calculations are outlined below, see ref. 37 for details:

CBS-QB3:

- 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 methods are both modifications of the Gaussian-3 multi-level theory for the calculation of molecular energies.[41] Like CBS-QB3, G3B3 and G3MP2B3 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^{-1} accuracy. The required electronic structure calculations for the G3B3 method are outlined below, see ref. 39 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 G3MP2B3 method eliminates all of the MP4 calculations above, trading some accuracy for speed.

All of the multi-level methods studied here formally scale as $O(N^7)$ due to the CCSD(T) step in CBS-QB3, 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 multi-level 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) (PBE0), B1B95/6-311++G(3df,2p) (B1B95), and B3LYP/6-311++G(3df,2p) (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 phosphorous compounds[7,8,11,42] was also included.

B3LYP is a three parameter hybrid functional[43] using the B88 exchange functional of Becke[44] and the Lee, Yang, and Parr correlation functional.[45] PBE0 is a zero parameter hybrid functional[46] using the Perdew, Burke, Ernzerhof exchange and correlation functional.[47] B1B95 is a one parameter hybrid functional[48] using the B88 exchange functional of Becke[44] and B95 correlation functional.[48] In an effort to give the hybrid density functional methods the best possible accuracy for benchmark purposes and avoid problems in the frequency calculations[49] all calculations (except the QCRNA and multi-level 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, fine 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 that of B3LYP.

2.2 Calculation of proton affinities and gas-phase basicities

The proton affinity (PA) and gas-phase basicity (GPB) of a species A^- are related to the gas phase reaction:

$$A^-_{(g)} + H^+_{(g)} \rightarrow AH_{(g)} \tag{2}$$

The proton affinity of A^- is defined as the negative of the enthalpy change (ΔH) of the process in eqn. (2), and the gas-phase basicity of A^- is defined as the negative of the corresponding Gibbs free energy change (ΔG).[35] 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, ideal gas/particle-in-a-box models in the canonical ensemble.[50] The standard state in the gas phase was for a mole of particles at 298.15 K and 1 atm pressure.

Explicit inclusion of zero-point energy corrections is important for reliable thermodynamic results. Although other studies have made the assumption that the difference in zero point energy between the neutral acid AH and anion A^- is generally small,[20] for the systems studied in the present work, this is not the case. An O–H bond stretch vibration typically falls in the range of 2500–3600 cm^{-1}, and corresponds to a zero-point energy difference of around 7–10 kcal mol^{-1} (this value is consistent with the zero-point energy range in the present work, *e.g.*, 7.22–10.3 kcal mol^{-1} for the B3LYP method). Neglect of the zero-point energy, if applied to the calculation of pK_a values *via* the thermodynamic cycle in Scheme 1 (or similar cycles), would lead to errors in absolute pK_a values of 5–7 pK_a units, and errors in pK_a shifts of up to 2 pK_a units. This would significantly limit the overall reliability and predictive capability in applications to biological systems, and hence it is recommended that zero-point energies be included explicitly in such calculations.

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

$$H(H^+) = U + PV = \frac{5}{2}RT \tag{3}$$

where U is the internal energy, P and V are the pressure and volume, respectively, R is the universal gas constant, and T is the absolute temperature. The entropy of the proton was calculated from the Sackur–Tetrode equation,[51]

$$S(H^+) = R\ln\left(\frac{e^{5/2}k_B T}{p\Lambda^3}\right) \tag{4}$$

where k_B is the Boltzmann constant, p is the pressure, and Λ is the thermal De Broglie wavelength [$\Lambda = (h^2/2\pi m k_B T)^{1/2}$, 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^{-1} and $S(H^+) = 26.02$ cal mol^{-1} K^{-1}, respectively, and lead to a gas-phase Gibbs free energy value of $G(H^+) = H(H^+) - TS(H^+) = -6.28$ kcal mol^{-1}.

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* gas-phase basicities, 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 eqn. (2), assuming unit activity coefficients, is given by:

$$K^M = \frac{[A^-]_M[H^+]}{[AH]_M} \tag{5}$$

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

$$K^M = \frac{m[A^-]_\mu[H^+]}{n[AH]_\mu} = K^\mu\left(\frac{m}{n}\right) \tag{6}$$

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:

$$\Delta G = -RT\ln K \tag{7}$$

where R is the ideal gas constant and T is the temperature of interest. Substitution of eqn. (7) into eqn. (6) yields the following equation for interconversion of microscopic and macroscopic free energy changes:

$$\Delta G^\mu = \Delta G^M + RT\ln\left(\frac{m}{n}\right) \tag{8}$$

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^{-1}.[52] Application of eqn. (8) yields a microscopic GPB of 323.2 kcal mol^{-1}.

3. Results and discussion

In the present work, calculated proton affinity and gas-phase basicity values are compared with experimental values using several error metrics:

$$\text{MSE} = \langle \text{err} \rangle \tag{9}$$

$$\text{MUE} = \langle |\text{err}| \rangle \tag{10}$$

$$\text{RMSE} = \langle |\text{err}|^2 \rangle^{1/2} \tag{11}$$

$$\text{MAXE} = \text{signed\_max(err)} \tag{12}$$

where "err" is the error for each data point defined as the calculated minus experimental value, and signed_max(err) in eqn. (12) is meant to indicate the signed error value with the largest (maximum) magnitude. The mean signed error (MSE) indicates the mean error value (where the distribution of errors is centered). The mean unsigned error (MUE) is an average measure of the magnitude of the errors. The root-mean-square error (RMSE) measures the second moment of the error distribution and is related to the standard deviation σ by

Table 1 Proton affinity error analysis[a]

Molecule[b]	CBS-QB3		G3B3		G3MP2B3		PBE0		B1B95		B3LYP		QCRNA		Exp
Water	1.7	*1.8*	1.2	*0.9*	1.3	*0.7*	3.2	*3.4*	3.3	*3.1*	0.1	*1.9*	0.1	*1.9*	390.3(0.2)
Methanol	1.1	*1.3*	2.2	*1.9*	2.1	*1.6*	−0.5	*−0.3*	−0.1	*−0.2*	−2.3	*−0.4*	−2.3	*−0.5*	381.5(1.0)
Ethanol	0.7	*0.8*	1.5	*1.2*	1.7	*1.1*	−0.3	*−0.1*	0.3	*0.1*	−2.2	*−0.4*	−2.2	*−0.4*	378.2(0.8)
Propanol	0.8	*0.9*	1.7	*1.3*	1.9	*1.4*	0.3	*0.5*	0.7	*0.6*	−1.5	*0.3*	−1.4	*0.4*	376.0(1.1)
2-Propanol	0.8	*0.9*	1.4	*1.1*	1.7	*1.2*	0.7	*0.9*	1.0	*0.9*	−1.4	*0.4*	−1.3	*0.5*	375.7(0.8)
DMPH[c]	−2.4	*−2.2*	−1.8	*−2.1*	−1.3	*−1.9*	−0.6	*−0.4*	−0.7	*−0.8*	−1.6	*0.3*	−1.5	*0.3*	331.6(4.1)
Phosphoric acid	−2.6	*−2.4*	−2.2	*−2.6*	−1.8	*−2.4*	−0.5	*−0.3*	−0.6	*−0.8*	−2.4	*−0.6*	−2.3	*−0.5*	330.5(5.0)
Formic acid	−0.4	*−0.3*	0.4	*−0.0*	0.9	*0.3*	−0.2	*0.0*	0.1	*−0.0*	−2.0	*−0.1*	−2.0	*−0.1*	344.0(1.6)
Acetic acid	0.2	*0.3*	1.0	*0.6*	1.4	*0.8*	0.4	*0.6*	0.7	*0.5*	−0.8	*1.0*	−0.8	*1.0*	347.2(1.1)
Propanoic acid	−0.5	*−0.3*	0.3	*−0.1*	0.7	*0.1*	0.4	*0.6*	0.5	*0.4*	−1.3	*0.5*	−1.3	*0.5*	347.4(1.8)
Phenol	−0.8	*−0.6*	−0.6	*−1.0*	−0.6	*−1.2*	−1.4	*−1.2*	−0.7	*−0.9*	−2.4	*−0.6*	−2.5	*−0.6*	350.1(1.1)
o-Chlorophenol	0.7	*0.8*	1.0	*0.7*	1.0	*0.4*	0.4	*0.6*	1.0	*0.8*	−1.0	*0.9*	−1.0	*0.8*	343.4(2.3)
m-Chlorophenol	−0.6	*−0.4*	−0.3	*−0.6*	−0.3	*−0.8*	−1.2	*−1.0*	−0.6	*−0.8*	−2.5	*−0.6*	−2.4	*−0.5*	342.2(3.1)
p-Chlorophenol	−0.5	*−0.4*	−0.2	*−0.6*	−0.2	*−0.8*	−1.0	*−0.8*	−0.4	*−0.6*	−2.2	*−0.4*	−2.3	*−0.5*	343.4(1.6)
p-Methylphenol	−0.3	*−0.1*	0.0	*−0.3*	0.0	*−0.6*	−0.8	*−0.5*	−0.1	*−0.3*	−1.8	*0.0*	−1.8	*−0.0*	350.7(1.3)
p-Nitrophenol	−0.2	*−0.0*	0.2	*−0.1*	0.7	*0.1*	−2.3	*−2.1*	−1.7	*−1.9*	−4.1	*−2.3*	−4.2	*−2.3*	327.8(2.1)
MAXE	−2.6	*−2.4*	−2.2	*−2.6*	2.1	*−2.4*	3.2	*3.4*	3.3	*3.1*	−4.1	*−2.3*	−4.2	*−2.3*	
RMSE	1.1	*1.1*	1.2	*1.2*	1.3	*1.1*	1.2	*1.2*	1.1	*1.1*	2.0	*0.9*	2.0	*0.9*	
MUE	0.9	*0.9*	1.0	*0.9*	1.1	*1.0*	0.9	*0.8*	0.8	*0.8*	1.8	*0.7*	1.8	*0.7*	
MSE	−0.1	*0.0*	0.4	*0.0*	0.6	*0.0*	−0.2	*0.0*	0.2	*0.0*	−1.8	*0.0*	−1.8	*0.0*	

[a] All quantities are in kcal mol^{-1}. Experimental values are the unweighted average of all values available from ref. 52. Estimates of the experimental error are shown in parenthesis immediately to the right of the experimental data and were calculated from the individual error estimates using standard propagation of errors.[53] The error metrics (error = calculated − experimental value) shown are the maximum error (MAXE), root-mean-square error (RMSE), mean unsigned error (MUE) and mean signed error (MSE). The proton affinity errors that include the empirical bond enthalpy correction ΔH^c for O–H bonds (see text) are shown in italics immediately to the right of the uncorrected data. [b] "Molecule" refers to the neutral molecule AH in eqn. (2). [c] Hydrogen dimethyl phosphate

$\sigma^2 = \text{RMSE}^2 - \text{MSE}^2$. The MUE and RMSE are equal if all of the errors are equal to one another (*i.e.*, the error distribution has zero standard deviation). Alternately, if the RMSE is found to be considerably larger than the MUE, this suggests there is a broad distribution or errors, or indicates the existence of a few outlier data points that dominate the RMSE. The maximum error (MAXE) is simply the single data point error with the largest magnitude. These error metrics obey the inequality:

$$|\text{MSE}| \leq \text{MUE} \leq \text{RMSE} \leq |\text{MAXE}| \qquad (13)$$

Together, these statistical quantities provide useful information that describes the distribution of errors, and in some cases may provide insight into simple correction terms that capitalize on the presence of systematic errors.

Tables 1 and 2 compare the errors with respect to experiment for proton affinity and gas-phase basicity values calculated with the various quantum model chemistries. The 16 molecules in these tables were selected from available experimental data to represent those most important in the study of biological phosphoryl transfer, with emphasis on the most common types of phosphorus compounds, nucleophiles, and leaving groups that involve formation and cleavage of P–O bonds.

3.1 Performance of multi-level methods

In general, the multi-level methods are fairly comparable to one another, with the CBS-QB3 method having best overall performance for the proton affinities and gas-phase basicities. The CBS-QB3 method predicts 2/3 data points with PA/GPB values that lie outside of the experimental error bars, whereas the G3B3 and G3MP2B3 methods predict 5/4 and 6/5 data point values, respectively, that lie outside the experimental error bars.

The PA MUE/MSE values are 0.9/−0.1 kcal mol^{-1} for CBS-QB3, 1.0/0.4 kcal mol^{-1} for G3B3 and 1.1/0.6 kcal mol^{-1} for G3MP2B3. Examination of the data indicates that the PA errors are more systematic within sets of molecules with the same functional group. For example in the CBS-QB3 method, the errors for alkyl alcohols, acids, and phenols (excluding *o*-chlorophenol that appears to be an outlier for all of the methods) have ranges of 0.8 to 1.1, −0.5 to 0.2, and −0.8 to −0.2, respectively. This suggests that these methods are likely more accurate for determination of relative PA values between similar classes of compounds than for absolute PA values. The PA MAXE values range from −2.1 to −2.6 kcal mol^{-1}. For CBS-QB3, the largest PA error is for phosphoric acid (−2.6 kcal mol^{-1}), and for G3B3 and G3MP3B3 the largest PA error is for methanol (2.2 and 2.1 kcal mol^{-1}, respectively).

The GPB error results are similar overall to the PA error results, although the error trends within the same functional group are less pronounced. For GPB values, G3B3 has the best MSE of the multi-level methods, although the range of MUE and RMSE values are similar (0.9 to 1.0 kcal mol^{-1} and 1.1 to 1.2 kcal mol^{-1}, respectively). The CBS-QB3 method performs better for alkyl alcohols than G3B3 and G3MP2B3, but worse for acids, and slightly worse for phenols.

Among the largest errors for the multi-level methods involve the phosphorus compounds (hydrogen dimethyl phosphate and phosphoric acid). With only two phosphorus data points with fairly large experimental error bars (greater than 4 kcal mol^{-1}), it is not statistically significant to generalize about the performance of the methods for calculation of PA and GPB of biological phosphorus compounds. This underscores the importance of performing robust benchmark quantum calculations for proton affinities and gas-phase basicities against available experimental values in order to assess accuracy and sources of systematic errors and using this information to derive more accurate values for molecules involved in biological phosphoryl transfer reactions. This information can then be used in the prediction of pK_a values or linear free energy relations that provide insight into mechanism.

3.2 Performance of density-functional methods

It is generally accepted that multi-level schemes afford a high general level of accuracy for most problems of

Table 2 Gas-phase basicity error analysis[a]

Molecule[b]	CBS-QB3		G3B3		G3MP2B3		PBEO		B1B95		B3LYP		QCRNA		Exp
Water	1.7	*2.2*	1.2	*1.3*	1.3	*1.1*	3.2	*3.8*	3.3	*3.4*	0.1	*2.4*	0.1	*2.5*	383.7(0.2)
Methanol	1.3	*1.9*	2.4	*2.4*	2.3	*2.2*	−0.3	*0.3*	0.1	*0.2*	−2.0	*0.3*	−2.1	*0.2*	374.8(0.7)
Ethanol	0.4	*0.9*	1.3	*1.3*	1.4	*1.3*	−0.6	*0.0*	0.0	*0.1*	−2.5	*−0.2*	−2.4	*−0.1*	371.3(1.0)
Propanol	0.2	*0.7*	1.2	*1.2*	1.4	*1.2*	−0.3	*0.3*	0.2	*0.2*	−2.0	*0.3*	−2.0	*0.3*	369.4(1.1)
2-propanol	0.3	*0.8*	1.0	*1.0*	1.3	*1.1*	0.3	*0.8*	0.7	*0.8*	−1.8	*0.5*	−1.7	*0.6*	368.8(1.0)
DMPH[c]	−1.7	*−1.1*	−1.2	*−1.1*	−0.7	*−0.8*	0.2	*0.8*	0.3	*0.3*	−0.8	*1.5*	−0.9	*1.5*	324.6(4.0)
Phosphoric acid	−2.5	*−2.0*	−2.2	*−2.2*	−1.8	*−1.9*	−0.3	*0.3*	−0.4	*−0.3*	−2.2	*0.1*	−2.2	*0.1*	323.2(4.9)
Formic acid	−1.3	*−0.8*	−0.6	*−0.6*	−0.1	*−0.2*	−1.1	*−0.6*	−0.8	*−0.8*	−2.9	*−0.6*	−2.9	*−0.6*	337.9(1.2)
Acetic acid	−2.0	*−1.4*	−1.2	*−1.1*	−0.8	*−0.9*	−0.1	*0.5*	0.1	*0.2*	−3.6	*−1.3*	−3.1	*−0.8*	341.4(1.2)
Propanoic acid	−0.9	*−0.3*	−0.1	*−0.1*	0.3	*0.1*	−1.2	*−0.6*	−0.5	*−0.4*	−2.8	*−0.5*	−3.6	*−1.3*	340.4(1.4)
Phenol	−0.6	*−0.1*	−0.5	*−0.5*	−0.5	*−0.7*	−1.3	*−0.7*	−0.6	*−0.6*	−2.3	*−0.0*	−2.4	*−0.0*	342.9(1.4)
o-Chlorophenol	−0.8	*−0.2*	−0.4	*−0.3*	−0.4	*−0.6*	−1.1	*−0.5*	−0.5	*−0.4*	−2.4	*−0.1*	−2.4	*−0.0*	337.1(2.0)
m-Chlorophenol	−1.2	*−0.6*	−0.9	*−0.8*	−0.9	*−1.0*	−1.8	*−1.3*	−1.2	*−1.2*	−3.1	*−0.7*	−2.9	*−0.6*	335.2(1.4)
p-Chlorophenol	−0.9	*−0.3*	−0.5	*−0.5*	−0.5	*−0.7*	−1.3	*−0.7*	−0.7	*−0.6*	−2.4	*−0.1*	−2.6	*−0.3*	336.5(1.4)
p-Methylphenol	−0.6	*−0.0*	−0.4	*−0.3*	−0.4	*−0.5*	−1.0	*−0.5*	0.7	*0.8*	−2.0	*0.3*	−2.2	*0.2*	343.8(1.2)
p-Nitrophenol	−0.1	*0.4*	0.2	*0.2*	0.7	*0.5*	−2.3	*−1.8*	−1.8	*−1.7*	−4.1	*−1.8*	−4.2	*−1.8*	320.9(2.0)
MAX	−2.5	*2.2*	2.4	*2.4*	2.3	*2.2*	3.2	*3.8*	3.3	*3.4*	−4.1	*2.4*	−4.2	*2.5*	
RMSE	1.2	*1.1*	1.1	*1.1*	1.1	*1.1*	1.3	*1.2*	1.1	*1.1*	2.5	*1.0*	2.5	*1.0*	
MUE	1.0	*0.9*	0.9	*0.9*	0.9	*0.9*	1.0	*0.8*	0.7	*0.7*	2.3	*0.7*	2.4	*0.7*	
MSE	−0.5	*0.0*	−0.0	*0.0*	0.2	*0.0*	−0.6	*0.0*	−0.1	*0.0*	−2.3	*0.0*	−2.4	*0.0*	

[a] All quantities are in kcal mol^{-1}. Experimental values are the unweighted average of all values available from ref. 52, corrected for degenerate protonation sites where appropriate to produce microscopic gas-phase basicities. Estimates of the experimental error are shown in parenthesis immediately to the right of the experimental data and were calculated from the individual error estimates using standard propagation of errors.[53] The error metrics (error = calculated − experimental value) shown are the maximum error (MAXE), root-mean-square error (RMSE), mean unsigned error (MUE) and mean signed error (MSE). The gas-phase basicity errors that include the empirical bond free energy correction ΔG^{C} for O–H bonds (see text) are shown in italics immediately to the right of the uncorrected data. [b] "Molecule" refers to the neutral molecule AH in eqn. (2). [c] Hydrogen dimethyl phosphate.

thermochemistry, whereas the robustness of density-functional methods is typically less well established. In the present case of the calculation of proton affinities and gas-phase basicities for phosphoryl transfer, of primary concern is the accurate modeling of the O–H bond energy in the AH molecule and of the stability of the A^- anion that results from deprotonation. For these properties, density-functional theory provides a quite reasonable description, and although in some cases there may be errors in the absolute values, the errors are sufficiently systematic that simple corrections may be introduced. The advantage in exploring the use of density-functional theory as an alternative to the prediction of proton affinities and gas-phase basicities centers on the tremendous reduction in computational cost, particularly in applications to larger molecules of biological interest. Toward this end, a set of benchmark density-functional calculations have been performed using several established hybrid density functionals: PBE0, B1B95, and B3LYP (see the Methods section for details).

It is encouraging that the PA MSE for the PBE0 (−0.2 kcal mol^{-1}) and B1B95 (0.2 kcal mol^{-1}) functionals are impressively small, and the PA MUE and RMSE values (0.9 to 0.8 and 1.2 to 1.1 kcal mol^{-1}, respectively) are comparable to those of the multi-level methods. The maximum error for both of these functionals (3.2 to 3.3 kcal mol^{-1}) occurs for water, arguably the most important of all the nucleophiles. The multi-level methods predict errors in the proton affinity of water in the range 1.2–1.7 kcal mol^{-1}, and have MAXE values across the entire data set that are around 1 kcal mol^{-1} less in magnitude than the corresponding MAXE values for PBE0 and B1B95.

The B3LYP functional, on the other hand, has a fairly large MSE of −1.8 kcal mol^{-1} that indicates a systematic underprediction of the proton affinity values relative to experiment. This large MSE carries over to the MUE, RMSE and MAXE values causing them to similarly be significantly larger than any of the other methods. It is of mild interest to note that the B3LYP predicts the PA of water nearly exactly (0.1 kcal mol^{-1} error), although this is likely due to a fortuitous cancellation of errors.

As with the multi-level methods, the GPB results are similar to the PA results. The B3LYP functional shows a large, systematic error for the GPB values (MSE = −2.3 kcal mol^{-1}) that is 0.5 kcal mol^{-1} larger than for PA. B1B95 slightly outperforms PBE0, both of which compare favorably to the multi-level methods. The maximum GPB error for both of these functionals is for water (3.2 to 3.3 kcal mol^{-1}). This is potentially significant since water and hydroxide are both important nucleophiles in phosphate hydrolysis. The GPB value for water (383.7 kcal mol^{-1}[52,54]) also forms an important component of the thermodynamic cycle in Scheme 2 and those used in calculating the solvation free energies of H^+, H_3O^+, and OH^-, accurate values for which have been recently reported.[55]

It is of further interest to note than the geometry optimization and frequency calculation with the large 6-311++G(3df,2p) basis set, tight geometry convergence criteria, and finer integration mesh contributes only very mildly to the proton affinity values. The much less computationally demanding QCRNA level which uses the smaller 6-31++G(d,p) basis set for geometries and frequencies and the default integration mesh and geometry convergence criteria produces almost identical results to B3LYP. As is discussed below, the errors in the absolute proton affinity values for the B3LYP and QCRNA methods are systematic and simply correctable, and the relative proton affinity values are actually among the most accurate of all the methods.

3.3 Bond enthalpy, entropy and free energy corrections

Examination of the error statistics for the quantum data reveals that a main source of error, at least for some of the methods, arises from a simple shift of the average PA and GPB values relative to experiment. This behavior is manifested when the magnitude of the MSE is approximately equal to the MUE (|MSE| ≈ MUE). If further, the |MSE| is comparable to the RMSE (*i.e.*, the standard deviation of errors is very small), then a shift of the data by a constant would offer considerable

Table 3 Thermodynamic O–H bond corrections for several quantum models[a]

	Multi-level methods			DFT methods			
Correction	CBS-QB3	G3B3	G3MP2B3	PBE0	B1B95	B3LYP	QCRNA
ΔH^{C}	0.1	−0.4	−0.6	0.2	−0.2	1.8	1.8
ΔS^{C}	−1.3	−1.3	−1.3	−1.3	−1.0	−1.7	−1.7
ΔG^{C}	0.5	0.0	−0.2	0.6	0.1	2.3	2.3

[a] ΔH^{C} and ΔG^{C} are in kcal mol^{-1} and correspond to 298.15 K. ΔS^{C} is in cal/(mol K).

improvement. In the present case, two methods stand out as particularly fitting this scenario: the B3LYP and related QCRNA models (Tables 1 and 2).

This observation motivates the adoption of simple constant enthalpic and entropic correction term to the process in eqn. (2) to afford a means of improving the quantitative results and making accurate prediction of PA and GPB values for larger systems with more favorably scaling and affordable quantum models such as QCRNA. Since all of the molecules in the present work involve O–H type bonds, a simple O–H bond enthalpy correction (ΔH^{C}) and bond entropy correction (ΔS^{C}) is proposed to correct the calculated PA and GPB values (although it should be noted that the process in eqn. (2) is not formally correspond to a "bond energy" but rather a deprotonation that involves cleavage of the O–H bond followed by transfer of an electron from hydrogen). The corrected change in enthalpy ($\Delta H'$), entropy ($\Delta S'$) and free energy ($\Delta G'$) of the process in eqn. (2) take the form:

$$\Delta H' = \Delta H + \Delta H^{C} \tag{14}$$

$$\Delta S' = \Delta S + \Delta S^{C} \tag{15}$$

$$\Delta G' = \Delta G + \Delta G^{C} = \Delta H' - T\Delta S' \tag{16}$$

where $\Delta G^{C} = \Delta H^{C} - T\Delta S^{C}$.

The thermodynamic O–H correction terms, based on the set of 16 phosphoryl transfer molecules with available experimental PA and GPB values, are listed in Table 3 for each of the multi-level and density-functional methods.

Examination of the corrected PA and GPB values in Tables 1 and 2 indicate the thermodynamic O–H corrections improve the agreement with the experimental values and bring the error metrics of the different methods into a tighter range. The most dramatic effect occurs for the corrected QCRNA method that now has the lowest MUE and RMSE values for both PA and GPB values. Figs. 1 and 2 illustrate a regression of the uncorrected and corrected CBS-QB3 and QCRNA PA and GPB values. Only 2 calculated PA values and 1 calculated GPB value fall outside the experimental error bars for the corrected QCRNA method, whereas 3 calculated PA values and 2 calculated GPB values fall outside the experimental error bars for the corrected CBS-QB3 method. This is of tremendous importance since the QCRNA method is the most affordable of all of the multi-level and DFT methods compared, and can be extended to large phosphoryl transfer systems. It is yet to be explored as to whether similar types of corrections might offer further improvement when applied with other DFT functionals and different basis sets.

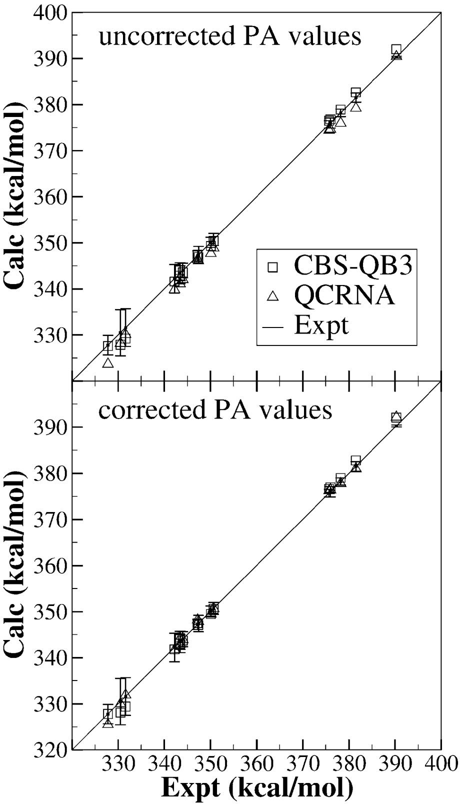


Fig. 1 Regression plots of the CBS-QB3 and QCRNA test set proton affinities. The top panel shows the uncorrected data, and the bottom panel shows the bond enthalpy corrected data. The linear correlation coefficient (R_C), linear best-fit slope (m) and intercept (b) for the corrected values are: $R_C = 0.9995$, $m = 1.046$, $b = -16.2$ for corrected CBS-QB3, $R_C = 0.9992$, $m = 1.021$, $b = -7.5$ for corrected QCRNA.

3.4 Comparison of semiempirical methods

Table 4 compares the proton affinity values calculated with several standard semiempirical methods that are commonly used in QM/MM simulations[56–62] with experimental values. All of the semiempirical proton affinities were calculated with respect to water, and are shown with and without thermodynamic O–H correction to the enthalpy (ΔH^{C}, defined here as minus the MSE for each method).

Of all of the semiempirical methods, PM3 performs the best for the uncorrected proton affinity calculations, with an RMSE of 13.8 kcal mol^{-1}, almost seven times higher than the uncorrected QCRNA. Interestingly MNDO/d has larger errors than regular MNDO. MNDO/d reduces to MNDO when only first and second-row atoms are present, so this difference must be in the proton affinities of the phosphates and the chlorophenols. The addition of d orbitals to the chlorine in the chlorophenols has very little effect on the calculated proton affinities; the RMSE for MNDO/d is 0.3 kcal mol^{-1} higher than for MNDO for the chlorophenols. The phosphates on the other hand are heavily influenced by the addition of d orbitals which radically lower the energy of the anions, increasing the PA errors by a factor of two for MNDO/d compared to MNDO. For dianionic phosphoryl transfer reaction that do not involve protonated phosphoryl oxygens, nucleophiles or leaving groups, the MNDO/d method has recently been shown to be quite reliable.[67]

The semiempirical PA values have a large systematic error component that is reflected by the MSE values which range from −12.0 kcal mol^{-1} (PM3) to −30.0 kcal mol^{-1} (MNDO/d). An O–H bond enthalpy correction reduces the semiempirical PA RMSE by a factor of 2–5. With O–H bond enthalpy correction, the MNDO, AM1, and SCC-DFTB methods perform the best of the semiempirical models with RMSE values in the range of 5–6 kcal mol^{-1} and MUE values in the range of 4–5 kcal mol^{-1}. These errors are still several times the average experimental error and larger than corrected QCRNA errors by a factor of 4–6. Most importantly, the relative proton affinity values are often of the wrong sign for the semiempirical methods, and could be problematic in the prediction of linear free energy relationships, pK_a values, or in QM/MM simulations of reactions in enzymes

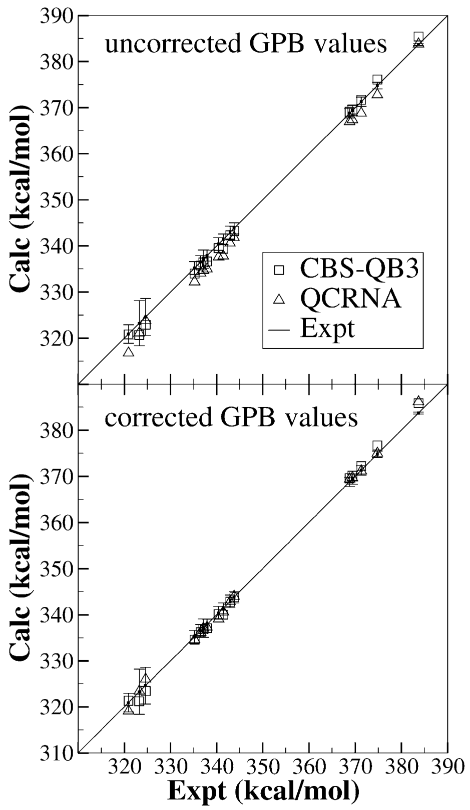


Fig. 2 Regression plots of the CBS-QB3 and QCRNA test set gas-phase basicities. The top panel shows the uncorrected data, and the bottom panel shows the bond free energy corrected data. The linear correlation coefficient (R_C), linear best-fit slope (m) and intercept (b) for the corrected values are: $R_C = 0.9996$, $m = 1.047$, $b = -16.4$ for corrected CBS-QB3, $R_C = 0.9991$, $m = 1.026$, $b = -9.0$ for corrected QCRNA.

and ribozymes. For example the proton affinity of the alkyl alcohols is experimentally ordered methanol > ethanol > propanol > 2-propanol. This trend is predicted by all of the multi-level and DFT methods, but AM1 orders the alcohols methanol > propanol > 2-propanol > ethanol. Such errors could have serious consequences for QM/MM simulations of processes that are sensitive to protonation state such as phosphoryl transfer. The reliable prediction of pK_a values for biological systems, for example, is beyond the scope of current semiempirical methods without taking recourse into corrections at much higher and more expensive levels of electronic structure theory. These data underscore the need for design of improved semiempirical quantum models[58,68–72] for biological reactions.

Table 4 Performance of standard semiempirical methods on this set for proton affinities[a]

	MNDO		AM1		PM3		MNDO/d		SCC-DFTB	
MAXE	−34.8	*8.9*	−28.1	*−9.7*	−23.3	*15.5*	−46.9	*−16.8*	−26.9	*13.1*
RMSE	27.4	*5.5*	19.1	*5.2*	13.8	*6.8*	31.0	*7.7*	17.4	*5.9*
MUE	26.8	*5.0*	18.3	*4.8*	12.4	*5.4*	30.0	*6.2*	15.1	*4.1*
MSE	−26.8	*0.0*	−18.3	*0.0*	−12.0	*0.0*	−30.0	*0.0*	−16.4	*0.0*

[a] MNDO, AM1, PM3, and MNDO/d calculations were performed with a modified version of MNDO97.[63] SCC-DFTB calculations were performed using a modified version of CHARMM and standard parameters for C, H, O, N,[64] P,[65] and S[66] and do not include the chlorophenols since chlorine parameters are not yet available. All calculations are relative to water. The error metrics (error = calculated − experimental value) shown are the maximum error (MAXE), root-mean-square error (RMSE), mean unsigned error (MUE) and mean signed error (MSE). The proton affinity errors that include an empirical bond enthalpy correction are shown in italics immediately to the right of the uncorrected data.

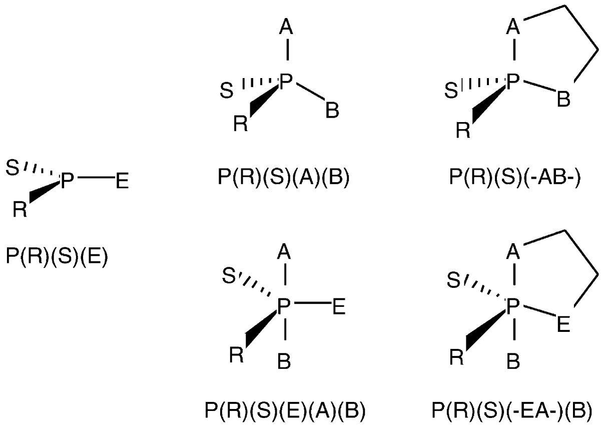
3.5 Predicted PA and GPB values for biological phosphorus compounds

Tables 5 and 6 list the predicted proton affinity and microscopic gas-phase basicity values, respectively, for several key model compounds involved in biological phosphoryl transfer reactions for which experimental data is unavailable. All predicted quantities include the bond enthalpy, entropy, and free energy corrections discussed above. Related compounds for which experimental data is available (*e.g.*, H_3PO_4) are included for completeness and comparison. The nomenclature convention for ligand designations in metaphosphate, acyclic and cyclic phosphate and phosphorane compounds is illustrated in Scheme 3 and is consistent with that used in previous work.[8,11]

There is remarkable agreement of predicted PA and GPB values between all of the quantum models tested with variations between methods rarely exceeding 2 kcal mol^{-1}. A key quantity to reproduce accurately is the relative PA and GPB values between different ionic species. In the case of phosphoric acid (H_3PO_4), the PA and GPB values are predicted to be between 129.0–130.6 and 129.7–130.9 kcal mol^{-1} lower than those of the singly deprotonated species ($H_2PO_4^-$), respectively. The corresponding PA and GPB differences between singly and doubly deprotonated species range from 119.8–123.0 and 121.5–125.1 kcal mol^{-1}, respectively.

Analysis of Tables 5 and 6 reveals several important trends:

- For neutral molecules of the general formula PO_nH_{2n-5}, PA and GPB values increase with increasing valence n (n = 3,4,5) of phosphorus.
- For neutral phosphates of the form $H_{3-m}PO_{4-m}(OCH_3)_m$, PA and GPB values only slightly increase with increasing methyl substitution m (m = 0,1,2) of the phosphoryl oxygens.
- For monoanionic phosphates of the form $H_{2-m}PO_{4-m}(OCH_3)_m^-$, PA and GPB values decrease with increasing methyl substitution (m) (m = 0,1) of the phosphoryl oxygens.
- The GPB value for dimethyl phosphate is slightly greater than that of ethylene phosphate (the cyclic analog of dimethyl phosphate), whereas their PA values are very similar.
- For neutral phosphoranes, PA and GPB values of phosphoryl oxygens in the axial position are considerably greater than in the equatorial position, and this difference decreases with increasing degree of alkylation.
- For neutral cyclic phosphoranes, PA and GPB values of equatorial phosphoryl oxygens are not significantly affected by



Scheme 3 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.[8,11]

Table 5 Predicted proton affinities for phosphorus compounds of biological interest[a]

Molecule[b]	CBS-QB3	G3B3	G3MP2B3	PBE0	B1B95	B3LYP	QCRNA
P(O)(OH)(OH)(OH)	328.1	327.9	328.1	330.2	329.7	329.9	330.0
P(O)(O)(OH)(OH)$^-$	458.9	458.2	458.8	460.7	460.6	459.6	459.6
P(O)(O)(O)(OH)$^{2-}$	581.1	578.0	580.3	583.3	583.6	581.2	581.3
P(O)(O)(OH)	310.6	310.7	311.2	312.6	312.3	312.8	312.9
P(O)(OH)(–O–CH$_2$CH$_2$–O–)	329.5	329.5	329.9	330.5	330.3	331.2	331.3
P(OH)(OH)(–O–CH$_2$CH$_2$–O–)(OH*)	351.9	352.0	352.1	352.9	352.8	352.8	352.9
P(OH*)(OH)(–O–CH$_2$CH$_2$–O–)(OH)	343.2	343.1	343.4	344.5	344.3	343.9	344.0
P(OH*)(OH)(–O–CH$_2$CH$_2$–O–)(OCH$_3$)	343.5	343.6	343.9	344.6	344.4	344.2	344.4
P(OH)(OCH$_3$)(–O–CH$_2$CH$_2$–O–)(OH*)	352.0	352.4	352.4	352.8	352.7	353.0	353.2
P(OH*)(OCH$_3$)(–O–CH$_2$CH$_2$–O–)(OH)	345.2	345.5	345.7	346.4	346.3	346.3	346.5
P(OH)(OH)(OH)(OH*)(OH)	351.0	350.8	350.9	352.6	352.4	352.4	352.3
P(OH*)(OH)(OH)(OH)(OH)	341.0	340.3	340.7	342.0	341.6	341.0	341.1
P(O)(OH)(OCH$_3$)(OCH$_3$)	329.4	329.5	329.8	331.2	330.8	331.9	331.9
P(O)(OH)(OH)(OCH$_3$)	329.3	329.3	329.6	331.2	330.7	331.5	331.5
P(O)(O)(OH)(OCH3)$^-$	454.9	454.3	454.9	456.2	456.2	455.3	455.4

[a] The nomenclature convention for is shown in Scheme 3. All quantities are in kcal mol^{-1} and include the bond enthalpy correction from Table 3. For molecules with more than one chemically distinct deprotonation state, the hydrogen position for deprotonation is superscripted with an asterisk "*". For molecules with two axial or two/three equatorial protons only the results from the lowest energy anion configuration are shown. Deprotonation at the other site(s) yield nearly identical proton affinity values (*i.e.*, far less than the deviation between different methods). [b] "Molecule" refers to the neutral molecule AH in eqn. (2).

Table 6 Predicted gas-phase basicities for phosphorous compounds of biological interest[a]

Molecule[b]	CBS-QB3	G3B3	G3MP2B3	PBE0	B1B95	B3LYP	QCRNA
P(O)(OH)(OH)(OH)	321.2	321.1	321.3	323.5	322.9	323.4	323.4
P(O)(O)(OH)(OH)$^-$	451.7	451.3	451.9	453.4	453.2	452.4	452.4
P(O)(O)(O)(OH)$^{2-}$	576.0	572.8	575.1	578.1	578.3	576.2	576.4
P(O)(O)(OH)	304.8	304.9	305.4	306.7	306.2	307.0	307.2
P(O)(OH)(–O–CH$_2$CH$_2$–O–)	321.8	322.4	322.7	324.9	324.6	323.7	324.1
P(OH)(OH)(–O–CH$_2$CH$_2$–O–)(OH*)	345.0	345.3	345.4	346.0	345.8	346.0	346.2
P(OH*)(OH)(–O–CH$_2$CH$_2$–O–)(OH)	335.7	335.9	336.2	337.2	337.0	336.6	336.8
P(OH*)(OH)(–O–CH$_2$CH$_2$–O–)(OCH$_3$)	335.9	336.2	336.5	337.1	336.7	336.7	336.9
P(OH)(OCH$_3$)(–O–CH$_2$CH$_2$–O–)(OH*)	344.7	345.5	345.5	345.5	345.5	345.8	345.4
P(OH*)(OCH$_3$)(–O–CH$_2$CH$_2$–O–)(OH)	338.1	338.6	338.9	339.4	339.1	339.4	339.6
P(OH)(OH)(OH)(OH*)(OH)	343.9	343.8	343.9	345.5	345.1	345.3	344.9
P(OH*)(OH)(OH)(OH)(OH)	323.1	323.1	323.4	325.0	324.5	325.4	325.4
P(O)(OH)(OCH$_3$)(OCH$_3$)	323.4	323.4	323.7	325.3	324.9	326.1	326.0
P(O)(OH)(OH)(OCH$_3$)	323.1	323.1	323.4	325.0	324.5	325.4	325.4
P(O)(O)(OH)(OCH$_3$)$^-$	447.7	447.1	447.6	449.1	449.0	448.4	448.4

[a] The nomenclature convention for is shown in Scheme 3. All quantities are in kcal mol^{-1} and include the bond free energy correction from Table 3. For molecules with more than one chemically distinct deprotonation state, the hydrogen position for deprotonation is superscripted with an asterisk "*". For molecules with two axial or two/three equatorial protons only the results from the lowest energy anion configuration are shown. Deprotonation at the other site(s) yield nearly identical gas-phase basicity values (*i.e.*, far less than the deviation between different methods). [b] "Molecule" refers to the neutral molecule AH in eqn. (2).

methylation in the axial position, but are increased by more than 2 kcal mol^{-1} upon methylation in the equatorial position. These trends are important for quantum models to reproduce when applied in hybrid QM/MM simulations of phosphoryl transfer in solution, enzymes and ribozymes.

It is interesting to note that the PA values for the phosphates (CBS-QB3 values range from 328.1 to 329.5 kcal mol^{-1}) are quite similar to the PA value for *p*-nitrophenolate (CBS-QB3 and experimental value of 327.0 kcal mol^{-1}), a common enhanced leaving group in phosphoryl transfer experiments.[5] Moreover, the PA values for equatorial phosphorane oxygens are similar to those of the acids and chlorophenols, the former representing models for acidic side chains that mediate phosphoryl transfer in kinases and phosphatases, and the latter another experimentally common leaving group.[2] The quantitative prediction of the PA and GPB values in Tables 5 and 6 is a first step toward the accurate modeling of pK_a values and linear free energy relations in biological phosphoryl transfer reactions.

4. Conclusion

Benchmark calculations for proton affinities and gas-phase basicities have been performed for a series of molecules with available experimental values. It has been shown that multi-level methods, in particular CBS-QB3, can reliably calculate proton affinities and gas-phase basicities to within experimental errors in almost all cases. Fairly recent density functionals such as PBE0 and B1B95 are competitive with much more expensive multi-level methods for these properties.

From the benchmark data, a set of empirical bond enthalpy, entropy, and free energy corrections are introduced. These corrections improved the accuracy of the methods considerably. One of the most widely used density functionals, B3LYP, has large but systematic errors that are easily corrected. With the bond enthalpy, entropy, and free energy corrections, an inexpensive model chemistry, QCRNA (based on B3LYP), was tested and shown to predict PA and GPB values with accuracy rivaling or exceeding the much more expensive multi-level methods.

The corrected methods were applied to a set of biological phosphorus compounds involved in phosphoryl transfer reactions for which there is no experimental PA and GPB data available. The results of these calculations reveal several interesting trends for PA and GPB values for metaphosphate, cyclic and acyclic phosphates and phosphoranes. The benchmark PA and GPB data for biological phosphorus compounds provide a first step toward the accurate calculation of pK_a values of reactive intermediates in phosphoryl transfer reactions.

Semiempirical methods commonly employed to study biological systems (*e.g.*, MNDO, AM1, PM3, MNDO/d, and SCC-DFTB) with QM/MM simulations are found to do be unreliable for prediction of proton affinities, and cannot be significantly improved with bond enthalpy, entropy, and free energy corrections. This may have important consequences for QM/MM simulations of phosphoryl transfer. The results of the present work identify an important challenge in the design of new semiempirical quantum models for linear-scaling and hybrid QM/MM simulations of large-scale biological phosphoryl transfer reactions: the accurate prediction of PA and GPB values. The benchmark data presented here can be used to facilitate development of more accurate and robust quantum models for biological reactions, a topic of intense effort by the authors. The successful design of such new semiempirical quantum models would represent a major advance in the arsenal of computational simulation tools for the prediction of pK_a shifts and linear free energy relations of phosphoryl transfer reactions in solution, enzymes and ribozymes.

Acknowledgements

D.Y. is grateful for financial support provided by the National Institutes of Health (Grant GM62248), and the Army High Performance Computing Research Center (AHPCRC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under Cooperative Agreement number DAAD19-01-2-0014. This work was partially supported by grants from the National Science Foundation (MCB-0314327, CHE-CAREER-0348649) and the National Institutes of Health (R01-GM071428) and the Deutsche Forschungsgemeinschaft (Forschergruppe 490). Q.C. also acknowledges a Research Fellowship from the Alfred P. Sloan Foundation. Computational resources were provided by the Minnesota Supercomputing Institute. K.R. acknowledges Adam Moser for illuminating discussions of the data, and Patton Fast for technical support.

References

1 R. R. Holmes, *Acc. Chem. Res.*, 2004, **37**, 746–753.
2 A. C. Hengge, *Acc. Chem. Res.*, 2002, **35**, 105–112.
3 A. M. Stock, V. L. Robinson and P. N. Goudreau, *Annu. Rev. Biochem.*, 2000, **69**, 183–215.
4 N. Ahn, *Chem. Rev.*, 2001, **101**(8), 2207–2208.
5 D. M. Perreault and E. V. Anslyn, *Angew. Chem. Int. Ed.*, 1997, **36**, 432–450.
6 D.-M. Zhou and K. Taira, *Chem. Rev.*, 1998, **98**, 991–1026.
7 C. S. López, O. N. Faza, B. A. Gregersen, X. Lopez, A. R. de Lera and D. M. York, *Chem. Phys. Chem.*, 2004, **5**, 1045–1049.
8 C. S. López, O. N. Faza, A. R. de Lera and D. M. York, *Chem. Eur. J.*, 2005, **11**, 2081–2093.
9 W. P. Jencks, *Chem. Rev.*, 1985, **85**(6), 511–527.
10 D. Herschlag and W. P. Jencks, *J. Am. Chem. Soc.*, 1990, **112**, 1951–1956.
11 K. Range, M. J. McGrath, X. Lopez and D. M. York, *J. Am. Chem. Soc.*, 2004, **126**, 1654–1665.
12 X. Lopez, M. Schaefer, A. Dejaegere and M. Karplus, *J. Am. Chem. Soc.*, 2002, **124**(18), 5010–5018.
13 J. Davies, N. Doltsinis, A. Kirby, C. Roussev and M. Sprik, *J. Am. Chem. Soc.*, 2002, **124**, 6594–6599.
14 Y. Alexeev, T. Windus, C. -G. Zhan and D. Dixon, *Int. J. Quantum Chem.*, 2005, **102**, 775–784.
15 G. I. Almerindo, D. W. Tondo and J. R. Pliego Jr., *J. Phys. Chem. A*, 2004, **108**, 166–171.
16 Y. Fu, L. Liu, R.-Q. Li, R. Liu and Q.-X. Guo, *J. Am. Chem. Soc.*, 2004, **126**, 814–822.
17 P. Hudáky and A. Perczel, *J. Phys. Chem. A*, 2004, **108**, 6195–6205.
18 A. M. Magill, K. J. Cavell and B. F. Yates, *J. Am. Chem. Soc.*, 2004, **126**, 8717–8724.
19 Y. H. Jang, W. A. Goddard III, K. T. Noyes, L. C. Sowers, S. Hwang and D. S. Chung, *J. Phys. Chem. B*, 2003, **107**, 344–357.
20 A. Klamt, F. Eckert, M. Diedenhofen and M. E. Beck, *J. Phys. Chem. A*, 2003, **107**, 9380–9386.
21 K. R. Adam, *J. Phys. Chem. A*, 2002, **106**(49), 11963–11972.
22 D. M. Chipman, *J. Phys. Chem. A*, 2002, **106**, 7413–7422.
23 J. J. Klicić, R. A. Friesner, S.-Y. Liu and W. C. Guida, *J. Phys. Chem. A*, 2002, **106**, 1327–1335.
24 J. R. Pliego Jr. and J. M. Riveros, *J. Phys. Chem. A*, 2002, **106**, 7434–7439.
25 M. D. Liptak and G. C. Shields, *J. Am. Chem. Soc.*, 2001, **123**(30), 7314–7319.
26 M. D. Liptak and G. C. Shields, *Int. J. Quantum Chem.*, 2001, **85**, 727–741.
27 I.-J. Chen and A. D. MacKerell Jr., *Theor. Chem. Acc.*, 2000, **103**, 483–494.
28 P. D. Lyne and M. Karplus, *J. Am. Chem. Soc.*, 2000, **122**, 166–167.
29 C. O. Silva, E. C. da Silva and M. A. C. Nascimento, *J. Phys. Chem. A*, 2000, **104**(11), 2402–2409.
30 J. E. Yazal, F. G. Prendergast, D. E. Shaw and Y.-P. Pang, *J. Am. Chem. Soc.*, 2000, **122**, 11411–11415.
31 M. Peräkylä, *Phys. Chem. Chem. Phys.*, 1999, **1**, 5643–5647.
32 C. O. da Silva, E. C. da Silva and M. A. C. Nascimento, *J. Phys. Chem. A*, 1999, **103**, 11194–11199.
33 G. Schüürmann, M. Cossi, V. Barone and J. Tomasi, *J. Phys. Chem. A*, 1998, **102**, 6706–6712.
34 W. H. Richardson, C. Peng, D. Bashford, L. Noodleman and D. A. Case, *Int. J. Quantum Chem.*, 1997, **61**, 207–217.
35 A. D. McNaught and A. Wilkinson, *Compendium of Chemical Terminology: IUPAC Recommendations*, Blackwell Science, Inc., Oxford, 2nd edn., 1997.
36 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *GAUSSIAN 03 (Revision B.05)*, Gaussian, Inc., Wallingford, CT, 2003.
37 J. A. Montgomery, Jr., M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 1999, **110**(6), 2822–2827.
38 J. A. Montgomery Jr., M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 2000, **112**, 6532–6542.
39 A. G. Baboul, L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 1999, **110**, 7650–7657.
40 E. K. Pokon, M. D. Liptak, S. Feldgus and G. C. Shields, *J. Phys. Chem. A*, 2001, **105**, 10483–10487.
41 L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 7764–7776.
42 E. Mayaan, K. Range and D. M. York, *J. Biol. Inorg. Chem.*, 2004, **9**(7), 807–817.
43 A. D. Becke, *J. Chem. Phys.*, 1993, **98**(7), 5648–5652.
44 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
45 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
46 C. Adamo and G. E. Scuseria, *J. Chem. Phys.*, 1999, **111**, 2889–2899.
47 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
48 A. D. Becke, *J. Chem. Phys.*, 1996, **104**(3), 1040–1046.

49 J. W. Ochterski, *Vibrational Analysis in Gaussian* http://gaussian.com/g whitepap/vib.htm [accessed March 2005), 1999.

50 C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons, Chichester, England, 2nd edn., 2002.

51 D. A. McQuarrie, *Statistical Mechanics*, University Science Books, Mill Valley, CA, 1973.

52 *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*, ed. P. Linstrom and W. Mallard, National Institute of Standards and Technology, Gaithersburg, MD, 2003.

53 H. C. Harris and D. C. Harris, *Quantitative Chemical Analysis*, W. H. Freeman Company, New York, 6th edn., 2002.

54 J. R. Smith, J. B. Kim and W. C. Lineberger, *Phys. Rev. A*, 1997, **55**(3), 2036–2043.

55 M. Palascak and G. Shields, *J. Phys. Chem. A*, 2004, **108**, 3692–3694.

56 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, *Science*, 2004, **303**, 186–195.

57 M. J. Field, P. A. Bash and M. Karplus, *J. Comput. Chem.*, 1990, **11**, 700–733.

58 W. Thiel, *Adv. Chem. Phys.*, 1996, **93**, 703–757.

59 F. J. Luque, N. Reuter, A. Cartier and M. F. Ruiz-López, *J. Phys. Chem. A*, 2000, **104**(46), 10923–10931.

60 Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim and M. Karplus, *J. Phys. Chem. B*, 2001, **105**(2), 569–585.

61 M. Elstner, T. Frauenheim and S. Suhai, *J. Mol. Struct. (THEOCHEM)*, 2003, **632**, 29–41.

62 B. A. Gregersen, X. Lopez and D. M. York, *J. Am. Chem. Soc.*, 2004, **126**, 7504–7513.

63 W. Thiel, *MNDO97, version 5.0*, University of Zurich, 1998.

64 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**(11), 7260–7268.

65 Q. Cui and M. Elstner, to be published.

66 T. A. Niehaus, M. Elstner, T. Frauenheim and S. Suhai, *J. Mol. Struct. (THEOCHEM)*, 2001, **541**, 185–194.

67 B. A. Gregersen, J. Khandogin, W. Thiel and D. M. York, *J. Phys. Chem. B*, 2005, **109**, 9810–9817.

68 T. Clark, *J. Mol. Struct. (THEOCHEM)*, 2000, **530**, 1–10.

69 P. Winget, C. Selçuki, A. Horn, B. Martin and T. Clark, *Theor. Chem. Acc.*, 2003, **110**(4), 254–266.

70 W. Thiel, in *Handbook of Molecular Physics and Quantum Chemistry*, ed. S. Wilson, John Wiley and Sons, Chichester, 2003, **vol. 2**, pp. 487–502.

71 B. Brauer, G. M. Chabanb and R. B. Gerbe, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2543–2556.

72 T. Bredow and K. Jug, *Theor. Chem. Acc.*, 2005, **113**, 1–14.