Computational Determination of Aqueous pK_a Values of Protonated Benzimidazoles (Part 1)

Trevor N. Brown and Nelaine Mora-Diez*

Department of Chemistry, Thompson Rivers University, Kamloops, BC, V2C 5N3 Canada

Received: September 9, 2005; In Final Form: March 17, 2006

Benzimidazoles are the organic compounds investigated in this work. The experimental determination of the pK_a values of protonated benzimidazoles in water is a challenge because of their low solubility. In addition, some derivatives are involved in tautomeric equilibria which increase the complexity of the theoretical pK_a determinations. In the present study, different approaches are considered to develop a methodology for the accurate prediction of aqueous pK_a values of protonated benzimidazoles at 298.15 K. We have considered different reaction schemes for approximating the acid dissociation equilibrium; two distinct equations are used for the calculation of pK_a values, and a number of levels of theory and empirical corrections are applied in the process of working toward this aim. The best correlations between the experimental and calculated data are obtained at the B3LYP/6-31+G(d,p)-PCM(opt) level of theory. The predictive capabilities of the methodologies attempted are tested with two compounds that were not included in the set of benzimidazoles initially investigated. The direct calculations differ significantly from the expected values, but the pK_a values calculated using the correlation equations are very similar and in reasonable agreement with the expected pK_a values.

1. Introduction

Acid equilibrium constants (K_a, pK_a = −log K_a) are an important property of organic compounds, with extensive effects on many biological and chemical systems. Aqueous pK_a values are especially useful because of their environmental and pharmacological applications.[1,2] pK_a values are related to a number of properties of drugs, such as solubility, extent of binding, and rate of absorption. In addition, the determination of dosage forms and the regimes of drugs are also related to their pK_a values.[2] A number of methods, both experimental and theoretical, have been employed to calculate pK_a values.[3−11] The correlation of theoretical and experimental data can allow the development of predictive models to determine the pK_a of compounds for which no experimental data are yet available.[12]

The theoretical determination of aqueous pK_a values remains a challenging problem for computational chemists. The ab initio calculation of pK_a values in solvents other than water is easily achieved using continuum solvation models.[13] Water is a very challenging solvent to model due to the large amount of hydrogen bonding that is not considered in continuum solvation models.[14]

A number of different approaches have been developed to deal with the problem of calculating aqueous pK_a values. A common practice is to incorporate experimentally determined values for the free energies of the solvated proton into pK_a calculations. This practice generally seems to increase the accuracy of pK_a calculations and has been employed along with high-level ab initio calculations (CBS-QB3) and a continuum solvation model to calculate pK_a values within 0.5 pK_a units.[4] Another approach is to include explicit water molecules in addition to a continuum solvation model. This approach has produced better results than using only a continuum model at the same level of theory (MP2).[5] A third approach for increasing the accuracy of calculated aqueous pK_a values is the development of a qualitative structure−property relationship (QSPR).[6]

* Corresponding author. E-mail: nmora@tru.ca.

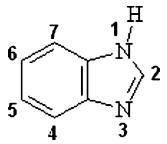

Figure 1. Benzimidazole and its substitution positions.

Using a multiple linear regression, a recent study reproduced experimental pK_a values within a mean absolute deviation of 0.09 pK_a units.[7]

Benzimidazoles are an interesting group of compounds that offer a number of challenges to the determination of their aqueous pK_a values. They are a class of compounds related to imidazole, which forms the base of the amino acid histidine. Benzimidazoles have potential and realized applications in pharmacology as bactericides,[15] antihistamines, analgesics, antiviral compounds and antiulcer agents, among others.[16] Aside from the listed pharmacological applications, benzimidazoles have also found applications as fungicides[15] and as corrosion inhibitors.[17] Figure 1 shows the structure of benzimidazole, the parent compound, and its substitution positions.

Benzimidazoles are soluble in organic solvents but have very low water solubility, making the experimental determination of their aqueous acid equilibrium constant difficult. In addition, some derivatives are involved in tautomeric equilibria which increase the complexity of theoretically determining their pK_a values.[18] An experimental method for determining pK_a values of benzimidazoles involves dissolving the compound in a very small amount of glacial acetic acid and subsequently diluting it with water. UV−visible spectra at different pH values and ionic strengths allow the determination of the absolute pK_a values of these compounds.[3] Potentiometric titrations have also been used to determine the pK_a values of benzimidazoles.[8]

Due to the relatively high pK_a of neutral benzimidazole (12.75),[19] most experimentally determined aqueous pK_a values refer to the protonated forms instead of the neutral ones. There are few aqueous pK_a values of protonated benzimidazoles

10.1021/jp055084i CCC: $33.50
Published on Web 04/20/2006

TABLE 1: Experimental Aqueous pK_a Values of Protonated Benzimidazoles at 298.15 K

molecule	experimental pK_a
5-nitrobenzimidazole	4.17[a]
2-chlorobenzimidazole	4.68[b]
benzimidazole	5.41[a]
1-methylbenzimidazole	5.57[c]
1-ethylbenzimidazole	5.62[c]
4-methylbenzimidazole	5.67[c]
5-methylbenzimidazole	5.81[c]
5,6-dimethylbenzimidazole	5.89[c]
2-methylbenzimidazole	6.10[a]
5-aminobenzimidazole	6.11[c]
2-ethylbenzimidazole	6.20[c]
2-isopropylbenzimidazole	6.23[c]
2-aminobenzimidazole	7.18[b]

[a] Reference 3. [b] Reference 20. [c] Reference 21.

TABLE 2: Summary of Equilibria and Associated pK_a Definitions

	specific equilibrium	absolute pK_a of HBz^+ [a]
Scheme 1	$HBz^+ \leftrightarrows Bz + H^+$	$pK_a = pK_1$
Scheme 2	$HBz^+ + H_2O \leftrightarrows Bz + H_3O^+$	$pK_a = pK_2 + pK_a(H_3O^+) = pK_2 - \log[H_2O]$
Scheme 3	$HBz^+ + Bz_1 \leftrightarrows Bz + HBz_1{}^+$ [b]	$pK_a = pK_3 + pK_a(HBz_1{}^+)$
Scheme 4	$HBz^+ + OH^- \leftrightarrows Bz + H_2O$	$pK_a = pK_4 + pK_a(H_2O) = pK_4 + pK_w + \log[H_2O]$

[a] HBz^+ represents a protonated benzimidazole derivative. pK_n refers to the pK of the reaction scheme (n) indicated. $pK_a(H_3O^+) = -1.74$; $pK_a(H_2O) = 15.74$.[22] [b] $HBz_1{}^+$ represents the protonated benzimidazole used as a reference acid.

available in the literature, and to the best of our knowledge, no previous theoretical studies have been reported on this topic. The development of a computational procedure for the prediction of unknown pK_a values of protonated benzimidazoles is of practical interest, as previously noted. The experimental aqueous pK_a values at 298.15 K of the protonated benzimidazoles considered in this study are shown in Table 1.

In this paper, a number of different approaches are considered with the aim of developing a methodology for the accurate prediction of the aqueous pK_a values of protonated benzimidazole derivatives at 298.15 K. Different reaction schemes are considered in order to approximate the acid dissociation equilibrium; two distinct equations are used for the calculation of pK_a values, and direct calculations are performed at four levels of theory for the full set of compounds previously presented, followed by correlations between the experimental and calculated data.

2. Methodology

2.1. Equilibria Used for Calculating the Aqueous pK_a Values of the Protonated Benzimidazoles. A summary of the reaction schemes considered for the study of the acid dissociation of the protonated benzimidazoles, and their related pK_a definitions, is shown in Table 2. The information in this table is discussed below.

The equilibrium used to calculate absolute pK_a values is the dissociation of an acid into its conjugate base and a free proton. Scheme 1 shows the acid dissociation of a protonated benzimidazole (HBz^+) into its conjugate base, neutral benzimidazole (Bz), and a free proton (H^+). The absolute pK_a of HBz^+ is the pK of Scheme 1 (pK_1).

$$pK_a = pK_1 \tag{1}$$

The exact structure of the proton in aqueous solution is difficult to define, though it is commonly assumed to be a hydronium ion complexed by a number of water molecules. A recent paper used ab initio calculations to elucidate a possible structure for this complex.[23] This difficulty in defining the exact structure of the proton in water has led to the use of other equilibria in which the proton does not appear, presumably increasing the accuracy of the calculated pK_a.

SCHEME 1

$$HBz^+ \leftrightarrows Bz + H^+$$

SCHEME 2

$$HBz^+ + H_2O \leftrightarrows Bz + H_3O^+$$

SCHEME 3

$$HBz^+ + Bz_1 \leftrightarrows Bz + HBz_1{}^+$$

SCHEME 4

$$HBz^+ + OH^- \leftrightarrows Bz + H_2O$$

Scheme 1 is a satisfactory representation of the gas-phase deprotonation of the acid HBz^+; in the aqueous phase, water will be involved in the acid−base equilibrium, as shown in Scheme 2. The relation between the calculated pK of Scheme 2 and the absolute pK_a of HBz^+ is shown in eq 2.

$$pK_a = pK_2 - \log[H_2O] = pK_2 + pK_a(H_3O^+) \tag{2}$$

In Scheme 3, the pK_a of HBz^+ is measured relative to the pK_a of a reference acid: in this case, benzimidazole ($HBz_1{}^+$). The relation between the calculated pK of Scheme 3 and the absolute pK_a of HBz^+ is shown in eq 3.

$$pK_a(HBz^+) = pK_3 + pK_a(HBz_1{}^+) \tag{3}$$

Scheme 4 is similar to Scheme 3 except that the reference acid is water instead of HBz^+. The relation between the pK of Scheme 4 and the absolute pK_a of HBz^+ is shown in eq 4.

$$pK_a = pK_4 + pK_w + \log[H_2O] = pK_4 + pK_a(H_2O) \tag{4}$$

2.2. Equations Used for Calculating Aqueous pK_a Values. The theoretical calculation of equilibrium constants (K, $pK = -\log K$) can be accomplished using two primary equations that are related to each other. Equation 5 is used to calculate the pK of an equilibrium that involves i species from the difference in molar energies of the ground states of the products and reactants at 0 K (ΔE_o), the molecular partition functions (q_i), and the stoichiometric coefficient (υ_i) of each species. N_A is Avogadro's number, R is the gas constant, and T is the absolute temperature.[24]

$$pK = -\log\left(\Pi\left(\frac{q_i}{N_A}\right)^{\upsilon_i} e^{-\Delta E_o/RT}\right) \tag{5}$$

The reference state of eq 5 is 1 mol m^{-3}. ΔE_o and q_i must be calculated with respect to the same energy state ($\upsilon = 0$ in our case). The molecular partition functions in eq 5 assume a volume of 1 m^3 and are approximated as a product of electronic, vibrational, rotational, and translational partition functions of noninteracting species.[24] To use eq 5, the molecular partition functions calculated using the molar volume ($V_m = 24.46$ L) of an ideal gas at the standard pressure ($p^\circ = 1$ atm) and 298.15 K must be converted using eq 6.

$$q_i(1\ m^3) = q_i(1\ atm, 298\ K) \times \frac{1\ mol}{V_m} \times \frac{1000\ L}{1\ m^3} \tag{6}$$

When trying to calculate aqueous pK values using eq 5, solvent effects should be included in the geometry optimizations and frequency calculations of the species involved in the equilibrium, so that the ΔE_o and q_i values incorporate the solvent effects. If solvent effects are only included in a single-point energy calculation, the use of eq 5 would imply that the ratio of molecular partition functions is virtually the same regardless of whether it is calculated in the gas phase or in solution. Applying the solvent model in a single-point energy calculation also implies that the relaxation energy (the energy difference between the gas-phase and aqueous-phase geometries) is negligible.

The application of eq 5 to Scheme 1 is impractical due to the uncertainty regarding the structure of the free proton in solution and the fact that its energy cannot be directly calculated. The application of eq 5 to Scheme 2 is shown in eq 7, and the calculation of ΔE_o is shown in eq 8. E_i^{0K} represents the energy at 0 K of species i, which is the total energy calculated plus the zero-point energy correction (ZPC). Since Schemes 3 and 4 are similar in form to Scheme 2 because they have the same number of reactant and product species, the application of eq 5 is very similar to that of eq 7. In the cases of Schemes 2, 3, and 4, the equilibrium constants at the reference states of 1 mol m^{-3} and 1 atm coincide.

$$pK_a = -\log\left(\left(\frac{q_{H_3O^+}q_{Bz}}{q_{H_2O}q_{HBz^+}}\right)e^{-\Delta E_o/RT}\right) \quad (7)$$

$$\Delta E_o = E^{0K}_{H_3O^+} + E^{0K}_{Bz} - E^{0K}_{H_2O} - E^{0K}_{HBz^+} \quad (8)$$

Another approach for calculating aqueous equilibrium constants more frequently found in the literature makes use of eq 9 where ΔG_{aq} is the aqueous Gibbs free energy change of the reaction.[24] If the geometry optimizations and frequency calculations of the species involved in the equilibrium are performed including solvent effects, the values of the aqueous Gibbs free energy (G_{aq}) of each reactant and product are directly obtained by adding the thermal correction to the Gibbs free energy (TCG_{aq}) and the total energy calculated without temperature correction (E^{Total}_{aq}), as indicated in eq 10. In this case, ΔG_{aq} can be calculated from the difference in G_{aq} of the products and reactants, as shown in eq 11. A similar treatment is employed by Klamt et al. in a recent paper.[25]

$$pK = \frac{\Delta G_{aq}}{RT\ln 10} \quad (9)$$

$$G_{aq} = E^{Total}_{aq} + TCG_{aq} \quad (10)$$

$$\Delta G_{aq} = \Sigma G_{aq}(\text{products}) - \Sigma G_{aq}(\text{reactants}) \quad (11)$$

If solvent effects are considered through a single-point (sp) energy calculation instead, a thermodynamic cycle should be used to calculate ΔG_{aq}. In this case, the value of ΔG_{aq} is calculated using eq 12, where ΔG_{gas} is the Gibbs free energy change of the reaction in the gas phase and $\Delta\Delta G_{solv}$ is the difference in solvation free energies (ΔG_{solv}) between products and reactants. The G_{gas} of each species is calculated using eq 13, where E^{Total}_{gas} represents the total gas-phase energy calculated without temperature correction, and TCG_{gas} is the thermal correction to the gas-phase Gibbs free energy. When using the PCM or COSMO solvent models, Gaussian 03[26] can include the ΔG_{solv} in the output file by including the "scfvac" keyword in the input file. Previously referenced thermal corrections refer to the temperature at which the equilibrium constant is to be calculated.

SCHEME 5

$$\begin{array}{ccccc}
G_{gas}(HBz^+) & G_{gas}(H_2O) & \Delta G_{gas} & G_{gas}(Bz) & G_{gas}(H_3O^+) \\
HBz^+_{(gas)} & + \quad H_2O_{(gas)} & \rightarrow & Bz_{(gas)} & + \quad H_3O^+_{(gas)} \\
\downarrow & \downarrow & & \downarrow & \downarrow \\
\Delta G_{solv}(HBz^+) & \Delta G_{solv}(H_2O) & \Delta\Delta G_{solv} & \Delta G_{solv}(Bz) & \Delta G_{solv}(H_3O^+) \\
\downarrow & \downarrow & & \downarrow & \downarrow \\
HBz^+_{(aq)} & + \quad H_2O_{(aq)} & \rightarrow & Bz_{(aq)} & + \quad H_3O^+_{(aq)} \\
G_{aq}(HBz^+) & G_{aq}(H_2O) & \Delta G_{aq} & G_{aq}(Bz) & G_{aq}(H_3O^+)
\end{array}$$

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{solv} \quad (12)$$

$$G_{gas} = E^{Total}_{gas} + TCG_{gas} \quad (13)$$

For aqueous systems, the reference state for eq 9 is 1 mol L^{-1}.[27] The conversion of the Gibbs free energy from a reference state of 1 atm to a reference state of 1 mol L^{-1} is shown in eq 14. In Schemes 2, 3, and 4, these corrections cancel out, and in Scheme 1, an experimental value for the Gibbs free energy of the aqueous proton is employed, meaning that the Gibbs free energy corrections for the acid and the base cancel each other out, so these corrections are not included in our calculations.

$$G(1\text{ mol L}^{-1}, 298.15\text{ K}) = G(1\text{ atm}, 298.15\text{ K}) + RT\ln V_m \quad (14)$$

To overcome the difficulty associated with the structure of the aqueous free proton when applying eq 9 to Scheme 1, the experimentally determined values of $G_{gas}(H^+) = -26.28$ kJ/mol and $\Delta G_{solv}(H^+) = -1105$ kJ mol^{-1} (ref 28) are used when calculating $G_{aq}(H^+)$.

The thermodynamic cycle employed for eq 12 and Scheme 2 is shown in Scheme 5. Equations 11 and 12 take the form of eq 15 and 16, respectively. The thermodynamic cycle and the calculation of ΔG_{aq} for Schemes 2, 3, and 4 are very similar.

$$\Delta G_{aq} = G_{aq}(Bz) + G_{aq}(H_3O^+) - G_{aq}(HBz^+) - G_{aq}(H_2O) \quad (15)$$

$$\Delta G_{aq} = (G_{gas}(Bz) + G_{gas}(H_3O^+) - G_{gas}(HBz^+) - G_{gas}(H_2O)) + (\Delta G_{solv}(Bz) + \Delta G_{solv}(H_3O^+) - \Delta G_{solv}(HBz^+) - \Delta G_{solv}(H_2O)) \quad (16)$$

Since eq 5 and 9 are related to each other, when the geometry optimizations and harmonic frequency calculations are performed in the aqueous phase, they should yield identical results. When optimizations are performed in the gas phase, followed by single-point energy calculations including solvent effects, the values produced by the two equations might be slightly different, probably due to differences in the calculation of partition functions in the gas and aqueous phases and the fact that calculations with eq 9 employ a thermodynamic cycle.

2.3. Treatment of Tautomeric Equilibria. Benzimidazole derivatives that are not substituted at positions 1 or 3 (see Figure 1) exhibit tautomeric equilibria when in their neutral forms. This type of tautomerism is specific to heterocyclic systems and is called annular tautomerism.[18] In this kind of tautomerism, solution equilibrium is so fast that individual tautomers are impossible to isolate.[18] An example of a tautomeric equilibrium in benzimidazoles is shown in Figure 2.

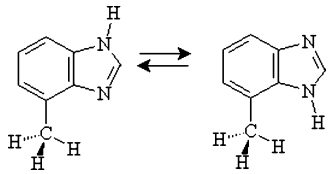

Figure 2. 4-Methylbenzimidazole and its tautomer, 7-methylbenzimidazole.

The tautomers of benzimidazoles substituted at position 2 are indistinguishable and presumably are equally populated. For derivatives with nonidentical tautomers, it must be assumed that one tautomer will be more stable and will therefore be more populated than the other. Equation 17 defines the relative population of the named benzimidazole tautomer (f_1) and the unnamed tautomer (f_2). In cases where the tautomers are identical ($f_1 = f_2 = 0.5$), the acid−base equilibrium of such a derivative will be affected, as would a 2-fold degeneracy of the neutral benzimidazole.[29] If the benzimidazole is substituted at position 1 ($f_1 = 1$, $f_2 = 0$), the acidity of the derivative is due exclusively to the proton at position 3.

$$f_1 + f_2 = 1, \quad 0 \le f_1 \le 1, \quad 0 \le f_2 \le 1 \tag{17}$$

Tautomerism can be defined as an equilibrium between two molecular states. The equilibrium constant for tautomerism (K_T) can be defined in terms of concentrations or relative populations of the two tautomers, as shown in eq 18, where T_1 and T_2 represent the two tautomeric forms of the acid A. The relative population of the named tautomer can be calculated from the tautomeric equilibrium constant (calculated using eqs 5 or 9) as shown in eq 19.

$$K_T = \frac{f_2}{f_1} = \frac{[T_2]}{[T_1]} \tag{18}$$

$$f_1 = \frac{1}{K_T + 1} \tag{19}$$

In any aqueous solution, the concentration of a benzimidazole derivative will be the sum of the concentrations of each tautomer. The concentration of each tautomer, as a portion of the total concentration, is the relative population of the tautomer multiplied by the total concentration. The total K_a is calculated from site-specific acid equilibrium constants using eq 20.

$$K_a = \frac{[H^+][B]}{[A]} = \frac{[H^+][B]}{[A]} f_1 + \frac{[H^+][B]}{[A]} f_2 = K_a^1 + K_a^2 \tag{20}$$

The overall pK_a of benzimidazoles can be viewed as the combination of the site-specific pK_a values of the protons at positions 1 and 3, defined as pK_a^1 and pK_a^2. Equation 20 and the definition of the relative populations in eq 17 can be combined to derive the relation between the total pK_a and the site-specific pK_a values shown in eq 21. The acid dissociation constants of each of the tautomeric forms are pK_a^1 and pK_a^2. Equation 21 is only valid for species with one acid form and two tautomeric basic forms.

$$\mathrm{p}K_a = -\log(K_a^1 + K_a^2) = \mathrm{p}K_a^1 + \log f_1 = \mathrm{p}K_a^2 + \log f_2 \tag{21}$$

2.4. Computational Details. Electronic structure calculations have been performed using the Gaussian 03[26] series of programs on a PQS Quantum Cube with 2.8 GHz Xeon processors running the Mandrake 9.2 Linux operating system. A total of 13 different combinations of methods (AM1,[30] HF, MP2,[31] and B3LYP[32]), basis sets (6-31+G(d,p), 6-311++G(d,p), 6-311++G(2df,2p),[33] and aug-cc-pvdz[34]), and solvent models (PCM[35] and COSMO[36]) are applied in trial calculations to compare their relative accuracies and to select the best combination of methods to calculate the pK_a values of the set of benzimidazoles under study. All of the calculations using the PCM solvent model employ the UAHF atomic radii when constructing the solvent cavity, as recommended in the Gaussian 03 User's Reference when the "scfvac" keyword is used to obtain the free energy of solvation, as is the case in this study.[37] All the geometries are fully optimized, and the character of the stationary points found is confirmed by a harmonic frequency calculation at the same level of theory to ensure a minimum is located. When optimizations are performed with a solvent model, the solvent model is also included in the harmonic frequency calculation.

3. Results and Discussion

3.1. Preliminary Calculations on Benzimidazole. To select the computational methods, basis sets and solvent models to be used in the pK_a calculations of the set of benzimidazoles under study (see Table 1), a series of trial calculations are performed on the smallest molecule studied: benzimidazole. The results of the trial calculations are shown in Table 3, with calculated pK_a values that are within ±0.30 units of the experimental value (5.41) indicated. The raw data for these calculations can be found in Table S1 of the Supporting Information. It was not possible to optimize the hydroxide ion at the MP2 and B3LYP levels of theory (basis set: 6-31+G-(d,p)) while including solvent effects (PCM).

The application of the two equations for pK_a calculations (eqs 5 and 9) produces similar results. When geometries are optimized in solvent, eqs 5 and 9 produce virtually identical results. Only when gas-phase geometries are used is there any difference between the values obtained using both equations. These results were expected and briefly discussed in section 2.2. Mean absolute deviations (AD, the mean of the unsigned deviations between pairs of corresponding values) between the calculated values of 0.25, 0.20, and 1.04 pK_a units are obtained using Schemes 2, 3, and 4, respectively.

Scheme 1 consistently produces values within approximately 3 pK_a units of the literature value. Schemes 2 and 4 produce good results that are very similar when experimental values are used for the free energies of water and the hydroxide and hydronium ions. When Schemes 2 and 4 are applied using only ab initio data, Scheme 2 performs better than Scheme 4, with the exception of the levels of theory employing the COSMO solvation model. Scheme 3 is considered for 9 of the 13 trial calculations; it produces the most consistent and accurate results of the four schemes applied with a mean absolute deviation from the experimental value of 0.64 pK_a units. When using this reaction scheme in the trial calculations, protonated 2-methylbenzimidazole is used as the reference. As can be seen in Table 3, the use of gas- or aqueous-phase geometries with Scheme 3 leads to basically the same pK_a values using both equations.

Four calculations are performed with the MP2//AM1, HF, MP2, and B3LYP methods using gas-phase geometries, the 6-31+G(d,p) basis set and the PCM solvent model with UAHF atomic radii. Similarly, three calculations are performed with the MP2//AM1, MP2, and B3LYP methods using the same type of geometry and basis set but applying the COSMO solvent model. In all cases where experimental free energy values are used (Schemes 1, 2, and 4), the two solvent models produce similar results. When Schemes 2 and 4 are applied using only ab initio data, the PCM solvent model is much more accurate.

TABLE 3: Aqueous pK_a Values of Protonated Benzimidazole at 298.15 K, Calculated Using Different Methods, Basis Sets, and Solvent Models[h]

level of theory	solvent model	geometry[a]	S1[c] eq 9	S2[b] eq 5	S2[b] eq 9	S2[c] eq 9	S3[b,d] eq 5	S3[b,d] eq 9	S4[b] eq 5	S4[b] eq 9	S4[c] eq 9	AD[e]
MP2/6-31+G(d,p)//AM1	PCM	gas-phase	3.53	3.26	3.16	4.91	**5.42**	6.11	−0.65	2.76	4.89	1.86
MP2/6-31+G(d,p)//AM1	COSMO	gas-phase	2.25	16.75	17.12	3.63	4.76	**5.31**	−9.60	−7.62	3.61	6.51
HF/6-31+G(d,p)	PCM	aqueous-phase	7.23	**5.67**	**5.68**	8.62	4.41	4.41	3.14	3.14	8.59	1.70
HF/6-31+G(d,p)	PCM	gas-phase	6.80	8.52	7.98	8.18	4.42	4.47	0.57	0.42	8.16	2.71
MP2/6-31+G(d,p)	PCM	aqueous-phase	3.05	1.37	1.37	4.43	**5.22**	**5.22**			4.41	1.83
MP2/6-31+G(d,p)	PCM	gas-phase	2.73	**5.18**	4.85	4.12	**5.22**	**5.63**	0.90	3.90	4.10	1.39
MP2/6-31+G(d,p)	COSMO	gas-phase	2.53	17.72	18.04	3.91			−8.65	−6.72	3.89	8.15
B3LYP/6-31+G(d,p)	PCM	aqueous-phase	4.64	3.47	3.47	6.02	4.46	4.46			6.00	0.97
B3LYP/6-31+G(d,p)	PCM	gas-phase	4.12	7.28	6.82	**5.50**	4.42	4.46	2.22	1.99	**5.48**	1.48
B3LYP/6-31+G(d,p)	COSMO	gas-phase	4.04	19.85	19.77	**5.43**	4.66	4.71	−6.96	−7.11	**5.40**	6.28
B3LYP/6-311++G(d,p)	PCM	gas-phase	3.79	7.31	6.84	**5.18**			1.71	1.50	**5.16**	1.86
B3LYP/6-311++G(2df,2p)	COSMO	gas-phase	4.55	20.50	20.43	5.93			−7.02	−7.18	5.91	8.15
B3LYP/aug-cc-pvdz	PCM	gas-phase	3.77	7.80	7.33	**5.16**			0.91	0.69	**5.14**	2.24
MD/AD of eq 5 from eq 9[f]				0.14/0.25			−0.18/0.20		−0.75/1.04			
AD of gas-phase pK_a from aqueous phase pK_a[g]			0.42	3.49	3.04	0.42	0.02	0.15	2.57	2.72	0.42	0.65

[a] A gas-phase geometry exists when the solvent model is applied in a single-point energy calculation (sp). An aqueous-phase geometry exists when the solvent model is applied during optimization and frequency calculations (opt). [b] The pK_a values are calculated using only ab initio data. [c] The pK_a values were calculated using experimental data for the free energy values of OH^-, H_2O, H_3O^+, and H^+. [d] The application of Scheme 3 (S3) in this case uses 2-methylbenzimidazole as the reference molecule. [e] Mean absolute deviation of the calculated pK_a values from the experimental pK_a of benzimidazole (5.41). [f] Mean deviation (MD) and mean absolute deviation (AD) of the pK_a values calculated using eqs 5 and 9. [g] Mean absolute deviation between pK_a values calculated using aqueous- and gas-phase geometries at the HF, MP2, and B3LYP levels of theory using the 6-31+G(d,p) basis set and the PCM solvent model. [h] The experimental pK_a value is 5.41.[3] Sx is an abbreviation for Scheme x.

Using the following combinations of methods and basis set, HF/6-31+G(d,p)-PCM, MP2/6-31+G(d,p)-PCM, and B3LYP/6-31+G(d,p)-PCM, pK_a values are calculated by applying the solvent model first during the optimization and frequency calculation (aqueous-phase geometry) and then only during a single-point energy calculation (gas-phase geometry), as described in section 2.2. When experimental free energies are used (Schemes 1, 2, and 4), aqueous-phase geometries produce pK_a values higher by 0.31−0.52 pK_a units, not always in better agreement with the experimental data. When using only ab initio data in Schemes 2 and 4, the deviation between pK_a values calculated using aqueous-phase and gas-phase geometries became much more significant. In this case, the use of gas-phase geometries produces pK_a values higher by a maximum of approximately 3 pK_a units. The results using Scheme 3 with gas- and aqueous-phase geometries are similar, with a mean absolute deviation between the values of 0.02 using eq 5 and 0.15 using eq 9.

B3LYP calculations combined with the PCM solvation model and using gas-phase geometries are performed with three basis sets: 6-31+G(d,p), 6-311++G(d,p), and aug-cc-pvdz. Similarly, B3LYP calculations with the COSMO solvation model and the 6-31+G(d,p) and 6-311++G(2df,2p) basis sets are also performed. From these calculations, we conclude that the increase in size of the basis set appears to have a negligible effect on the accuracy of the calculated pK_a values and significantly increases the computational time.

By far, the most pronounced effects on the calculated pK_a values are produced by using experimental free energy values and by the reaction scheme used to calculate the pK_a values. At the HF level of theory, Scheme 2 produces the best results, while, at the MP2 level of theory, Scheme 3 produces the best results. At the B3LYP level, Schemes 2 and 4 with experimental free energies give the best results. The most accurate results produced using ab initio data alone are at the HF/6-31+G(d,p)-PCM(opt) and MP2/6-31+G(d,p)-PCM(sp) levels of theory, but considering the relatively poor performance of these methods with other schemes, the results are most likely coincidental. The smallest mean absolute deviation (0.97 pK_a units), taking into account all the reaction schemes considered, is obtained at the B3LYP/6-31+G(d,p)-PCM(opt) level of theory (see Table 3).

The main conclusion from the trial calculations is that if the proper reaction scheme is employed, an accurate aqueous pK_a of protonated benzimidazoles at 298.15 K can be calculated using several of the computational methods shown in Table 3. Other conclusions from the trial calculations are that (a) incorporating solvent effects through a single-point (sp) energy calculation using gas-phase geometries produces results comparable to calculations using aqueous-phase geometries when experimental free energy values are used or when Scheme 3 is applied; (b) the smallest basis set used, 6-31+G(d,p), should be adequate; and (c) the COSMO solvent model produces results similar to the PCM model when experimental free energy values are used, with the exception of the MP2//AM1 calculations. When using only ab initio data, the PCM model is much more accurate than COSMO.

On the basis of the accuracy of the trial calculations, the best computational method is not obvious, so our selection is based on the most effective use of the computing resources available while trying to explore a variety of method combinations. Four levels of theory are selected from the trial calculations and used to determine the pK_a values of the set of thirteen benzimidazoles: MP2/6-31+G(d,p)//AM1-COSMO(sp), HF/6-31+G(d,p)-PCM(opt), B3LYP/6-31+G(d,p)-COSMO(sp), and B3LYP/6-31+G(d,p)-PCM(opt). Some calculations performed on benzimidazole derivatives using the MP2 method did not show any significant improvement over the B3LYP calculations and consumed more computing resources.

3.2. Calculations on the Set of Benzimidazoles under Study. The four different equilibria shown in Table 2 are considered during the pK_a calculations at the four levels of theory chosen. For Scheme 1, only eq 9 is applied using the experimental values for the Gibbs free energies of the proton. Equations 5 and 9 are both used to calculate the pK_a values of the other three equilibria (Schemes 2, 3, and 4). For Schemes 2 and 4, eq 9 is applied in two ways, using only ab initio data and using the experimental Gibbs free energies of water and the hydroxide and hydronium ions, as explained in the trial calculations. The raw data of the calculations performed can be found in Tables S2−S5 in the Supporting Information.

The relative populations of the tautomers are calculated for each of the four benzimidazole derivatives with nonidentical tautomers and used to correct the calculated pK_a values. At each level of theory, the equilibrium constant for tautomerism is

TABLE 4: Relative Populations of Nonidentical Tautomers

		MP2/6-31+G(d,p)//AM1-COSMO(sp)		HF/6-31+G(d,p)-PCM(sp)		HF/6-31+G(d,p)-PCM(opt)		B3LYP/6-31+G(d,p)-COSMO(sp)		B3LYP/6-31+G(d,p)-PCM(opt)	
molecule	eq	f_1[a]	f_2[b]	f_1[a]	f_2[b]	f_1[a]	f_2[b]	f_1[a]	f_2[b]	f_1[a]	f_2[b]
5-nitrobenzimidazole	5	0.594	0.406	0.748	0.252	0.737	0.263	0.682	0.318	0.644	0.356
	9	0.630	0.370	0.755	0.245	0.737	0.263	0.686	0.314	0.644	0.356
4-methylbenzimidazole	5	0.577	0.423	0.470	0.530	0.501	0.499	0.634	0.366	0.494	0.506
	9	0.580	0.420	0.531	0.469	0.501	0.499	0.605	0.395	0.494	0.506
5-methylbenzimidazole	5	0.462	0.538	0.413	0.587	0.404	0.596	0.441	0.559	0.413	0.587
	9	0.445	0.555	0.421	0.579	0.404	0.596	0.442	0.558	0.414	0.586
5-aminobenzimidazole	5	0.444	0.556	0.227	0.773	0.230	0.770	0.291	0.709	0.286	0.714
	9	0.355	0.645	0.229	0.771	0.230	0.770	0.289	0.711	0.286	0.714

[a] The value of f_1 is the relative population of tautomer 1 which is the named tautomer; e.g., tautomer 1 for 5-nitrobenzimidazole is substituted at position 5. [b] The value of f_2 is the relative population of tautomer 2 which is the unnamed tautomer; e.g., tautomer 2 for 5-nitrobenzimidazole is substituted at position 6.

TABLE 5: Calculated Aqueous pK_a Values for the Protonated Benzimidazoles at 298.15 K, Using Different Reaction Schemes (S1−4) and Equations at the MP2/6-31+G(d,p)//AM1-COSMO(sp) Level of Theory

molecule	lit value	S1 eq 9[a]	S2 eq 5[b]	S2 eq 9[b]	S2 eq 9[a]	S3 eq 5[b]	S3 eq 9[b]	S4 eq 5[b]	S4 eq 9[b]	S4 eq 9[a]
5-nitrobenzimidazole	4.17	0.02	13.21	14.90	1.40	1.87	3.18	−13.15	−9.85	1.38
2-chlorobenzimidazole	4.68	−2.43	12.33	12.45	−1.04	0.99	0.73	−14.02	−12.29	−1.07
benzimidazole	5.41	2.25	16.75	17.12	3.63	5.41[c]	5.41[c]	−9.60	−7.62	3.61
1-methylbenzimidazole	5.57	4.21	18.62	19.09	5.59	7.28	7.37	−7.73	−5.66	5.57
1-ethylbenzimidazole	5.62	3.31	17.60	18.19	4.70	6.26	6.48	−8.75	−6.55	4.68
4-methylbenzimidazole	5.67	2.74	17.23	17.62	4.13	5.89	5.90	−9.13	−7.12	4.10
5-methylbenzimidazole	5.81	2.67	17.20	17.54	4.05	5.85	5.83	−9.16	−7.20	4.03
5,6-dimethylbenzimidazole	5.89	3.30	17.73	18.18	4.69	6.38	6.46	−8.63	−6.56	4.66
2-methylbenzimidazole	6.10	3.03	18.09	17.91	4.42	6.75	6.20	−8.26	−6.83	4.40
5-aminobenzimidazole	6.11	2.10	16.50	16.98	3.49	5.16	5.27	−9.85	−7.76	3.47
2-ethylbenzimidazole	6.20	2.16	17.14	17.03	3.54	5.79	5.32	−9.22	−7.71	3.52
2-isopropylbenzimidazole	6.23	1.82	16.93	16.70	3.21	5.58	4.98	−9.43	−8.05	3.18
2-aminobenzimidazole	7.18	4.11	19.49	18.99	5.50	8.15	7.28	−6.86	−5.75	5.47
MD[d]		−3.49	11.09	11.39	−2.10	**−0.27**	**−0.35**	−15.26	−13.35	−2.13
AD[e]		3.49	11.09	11.39	2.11	**1.06**	**0.97**	15.26	13.35	2.13

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Experimental pK_a value. All other pK_a values are calculated relative to this value, which was not included in the calculation of deviations. [d] Mean deviation of the calculated pK_a values from the literature values. [e] Mean absolute deviation of the calculated pK_a values from the literature values.

calculated using both eqs 5 and 9. The relative population of each tautomer is calculated with both equilibrium constants. When calculating pK_a values with eq 5, the relative population derived using eq 5 is employed, and when calculating pK_a values with eq 9, the relative population derived using eq 9 is employed. At the four selected levels of theory, the relative populations calculated by the two equations are always within 0.06 of each other.

Table 4 shows the relative populations calculated at each level of theory for the four benzimidazole derivatives with nonidentical tautomers, as well as the results obtained at the HF/6-31+G(d,p)-PCM(sp) level of theory, for comparison purposes. It is important to note that the maximum difference in aqueous Gibbs free energies between any pair of tautomers is of 3.04 kJ mol^{-1} (for 5-aminobenzimidazole at the HF/6-31+G(d,p)-PCM(sp) level of theory). For most of the tautomeric pairs, the difference is less than 1.00 kJ mol^{-1}. Equations 5 and 9 give virtually identical values when aqueous-phase geometries are used (with mean absolute deviations of 0.000−0.009) and similar values when gas-phase geometries are used (with mean absolute deviations of 0.020−0.036). Table 4 also shows that the use of gas- or aqueous-phase geometries seems to have no significant influence on the calculated relative populations of the tautomeric benzimidazoles. The mean absolute deviation between relative populations using HF calculations using both types of geometries is 0.015. Almost identical values are also obtained from the two B3LYP combinations employed, even though the solvent models used are not identical. The qualitative prediction of the most stable tautomer is very consistent between the different levels of theory, with the exception of 4-methylbenzimidazole at the HF/6-31+G(d,p)-PCM(sp) level of theory using eq 5 and at the B3LYP/6-31+G(d,p)-PCM(opt) level of theory. We were unable to find experimental determinations of these relative populations.

Tables 5−8 show the results of the calculations performed at the MP2/6-31+G(d,p)//AM1-COSMO(sp), HF/6-31+G(d,p)-PCM(opt), B3LYP/6-31+G(d,p)-COSMO(sp), and B3LYP/6-31+G(d,p)-PCM(opt) levels of theory, respectively. It was not possible to optimize the protonated form of 2-isopropylbenzimidazole at the HF/6-31+G(d,p)-PCM(opt) and B3LYP/6-31+G(d,p)-PCM(opt) levels of theory, so this compound is not included in the data set at these levels of theory. One of the lowest mean and mean absolute deviations at all the levels of theory is obtained when working with Scheme 3 (using both pK_a expressions). Once again, it is shown that pK_a calculations using eq 5 and 9 give very similar results with Schemes 2, 3, and 4 when only ab initio data is used.

Scheme 1 produces pK_a values that are both higher and lower than the experimental values depending on the level of theory and performs best with the two B3LYP levels of theory. Schemes 2 and 4 give results comparable to or better than Scheme 1 when experimental free energy values are used for water and the hydroxide and hydronium ions. When using only ab initio data, Scheme 2 produces a large positive deviation while Scheme 4 produces a large negative deviation from experimental values when the COSMO solvation model is used. When the PCM solvation model is used, the calculated values are much improved and are comparable to the values calculated using experimental free energies. Scheme 3 gives the most accurate results for the MP2/6-31+G(d,p)//AM1-COSMO(sp) and HF/6-31+G(d,p)-PCM(opt) levels of theory, showing a mean absolute deviation of 0.97−1.10 pK_a units and mean

TABLE 6: Calculated Aqueous pK_a Values for the Protonated Benzimidazoles at 298.15 K, Using Different Reaction Schemes (S1−4) and Equations at the HF/6-31+G(d,p)-PCM(opt) Level of Theory

molecule	lit value	S1 eq 9[a]	S2 eq 5[b]	S2 eq 9[b]	S2 eq 9[a]	S3 eq 5[b]	S3 eq 9[b]	S4 eq 5[b]	S4 eq 9[b]	S4 eq 9[a]
5-nitrobenzimidazole	4.17	3.25	1.69	1.69	4.63	1.43	1.43	−0.85	−0.85	4.61
2-chlorobenzimidazole	4.68	1.18	−0.38	−0.38	2.56	−0.64	−0.64	−2.92	−2.92	2.54
benzimidazole	5.41	7.23	5.67	5.68	8.62	5.41[c]	5.41[c]	3.14	3.14	8.59
1-methylbenzimidazole	5.57	7.90	6.34	6.35	9.29	6.08	6.08	3.81	3.81	9.26
1-ethylbenzimidazole	5.62	8.07	6.52	6.52	9.46	6.25	6.25	3.98	3.98	9.44
4-methylbenzimidazole	5.67	7.31	5.76	5.76	8.70	5.49	5.49	3.22	3.22	8.68
5-methylbenzimidazole	5.81	7.68	6.12	6.12	9.07	5.86	5.86	3.59	3.59	9.04
5,6-dimethylbenzimidazole	5.89	8.14	6.58	6.58	9.52	6.32	6.32	4.04	4.04	9.50
2-methylbenzimidazole	6.10	8.92	7.37	7.37	10.31	7.10	7.10	4.83	4.83	10.29
5-aminobenzimidazole	6.11	8.13	6.57	6.58	9.52	6.31	6.31	4.04	4.04	9.49
2-ethylbenzimidazole	6.20	9.07	7.51	7.51	10.45	7.25	7.25	4.97	4.97	10.43
2-aminobenzimidazole	7.18	9.82	8.27	8.27	11.21	8.00	8.00	5.73	5.73	11.18
MD[d]		1.53	**−0.03**	**−0.03**	2.91	**−0.32**	**−0.32**	−2.57	−2.57	2.89
AD[e]		2.26	**1.22**	**1.23**	3.26	**1.10**	**1.10**	2.57	2.57	3.24

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Experimental pK_a value. All other pK_a values are calculated relative to this value, which was not included in the calculation of deviations. [d] Mean deviation of the calculated pK_a values from the literature values. [e] Mean absolute deviation of the calculated pK_a values from the literature values.

TABLE 7: Calculated Aqueous pK_a Values for the Protonated Benzimidazoles at 298.15 K, Using Different Reaction Schemes (S1−4) and Equations at the B3LYP/6-31+G(d,p)-COSMO(sp) Level of Theory

molecule	lit value	S1 eq 9[a]	S2 eq 5[b]	S2 eq 9[b]	S2 eq 9[a]	S3 eq 5[b]	S3 eq 9[b]	S4 eq 5[b]	S4 eq 9[b]	S4 eq 9[a]
5-nitrobenzimidazole	4.17	0.12	15.91	15.85	1.50	1.47	1.48	−10.90	−11.04	1.48
2-chlorobenzimidazole	4.68	−1.05	14.79	14.68	0.34	0.35	0.32	−12.02	−12.20	0.32
benzimidazole	5.41	4.04	19.85	19.77	5.43	5.41[c]	5.41[c]	−6.96	−7.11	5.40
1-methylbenzimidazole	5.57	4.84	20.63	20.57	6.23	6.19	6.21	−6.17	−6.31	6.20
1-ethylbenzimidazole	5.62	4.98	20.74	20.71	6.36	6.30	6.35	−6.07	−6.18	6.34
4-methylbenzimidazole	5.67	4.19	20.07	19.92	5.57	5.62	5.56	−6.74	−6.96	5.55
5-methylbenzimidazole	5.81	4.47	20.28	20.20	5.86	5.84	5.84	−6.53	−6.68	5.84
5,6-dimethylbenzimidazole	5.89	5.11	20.89	20.84	6.49	6.45	6.48	−5.91	−6.05	6.47
2-methylbenzimidazole	6.10	5.43	21.29	21.16	6.82	6.85	6.80	−5.52	−5.72	6.80
5-aminobenzimidazole	6.11	5.58	21.35	21.31	6.97	6.91	6.95	−5.46	−5.57	6.94
2-ethylbenzimidazole	6.20	4.96	20.58	20.69	6.35	6.14	6.33	−6.22	−6.19	6.32
2-isopropylbenzimidazole	6.23	4.73	20.54	20.45	6.11	6.09	6.09	−6.27	−6.43	6.09
2-aminobenzimidazole	7.18	7.20	23.00	22.93	8.58	8.56	8.57	−3.81	−3.95	8.56
MD[d]		**−1.54**	14.25	14.19	**−0.16**	**−0.21**	**−0.19**	−12.56	−12.70	**−0.18**
AD[e]		**1.54**	14.25	14.19	**0.96**	**1.01**	**1.03**	12.56	12.70	**0.95**

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Experimental pK_a value. All other pK_a values are calculated relative to this value, which was not included in the calculation of deviations. [d] Mean deviation of the calculated pK_a values from the literature values. [e] Mean absolute deviation of the calculated pK_a values from the literature values.

TABLE 8: Calculated Aqueous pK_a Values for the Protonated Benzimidazoles at 298.15 K, Using Different Reaction Schemes (S1−4) and Equations at the B3LYP/6-31+G(d,p)-PCM(opt) Level of Theory

molecule	lit value	S1 eq 9[a]	S2 eq 5[b]	S2 eq 9[b]	S2 eq 9[a]	S3 eq 5[b]	S3 eq 9[b]	S4 eq 9[a]
5-nitrobenzimidazole	4.17	0.34	−0.83	−0.83	1.73	1.12	1.12	1.71
2-chlorobenzimidazole	4.68	−0.50	−1.67	−1.67	0.88	0.27	0.27	0.86
benzimidazole	5.41	4.64	3.47	3.47	6.02	5.41[c]	5.41[c]	6.00
1-methylbenzimidazole	5.57	5.19	4.02	4.02	6.57	5.96	5.96	6.55
1-ethylbenzimidazole	5.62	5.39	4.22	4.22	6.78	6.16	6.16	6.75
4-methylbenzimidazole	5.67	4.86	3.69	3.69	6.24	5.63	5.63	6.22
5-methylbenzimidazole	5.81	5.16	3.99	3.99	6.54	5.93	5.93	6.52
5,6-dimethylbenzimidazole	5.89	5.73	4.56	4.56	7.11	6.50	6.50	7.09
2-methylbenzimidazole	6.10	6.27	5.11	5.11	7.66	7.05	7.05	7.64
5-aminobenzimidazole	6.11	6.22	5.05	5.05	7.60	6.99	6.99	7.58
2-ethylbenzimidazole	6.20	6.37	5.20	5.20	7.75	7.14	7.14	7.73
2-aminobenzimidazole	7.18	8.78	7.61	7.61	10.16	9.55	9.55	10.14
MD[d]		**−0.83**	−2.00	−2.00	**0.55**	**−0.07**	**−0.06**	**0.53**
AD[e]		**1.17**	2.07	2.07	**1.59**	**1.30**	**1.30**	**1.58**

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Experimental pK_a value. All other pK_a values are calculated relative to this value, which was not included in the calculation of deviations. [d] Mean deviation of the calculated pK_a values from the literature values. [e] Mean absolute deviation of the calculated pK_a values from the literature values.

deviations between −0.27 and −0.35 pK_a units. Scheme 2−eq 9 and Scheme 4−eq 9 (when using experimental free energies) produce the most accurate results at the B3LYP/6-31+G(d,p)-COSMO(sp) level of theory, showing a mean absolute deviation of 0.95−0.96 pK_a units and mean deviations between −0.16 and −0.18 pK_a units. Scheme 1 gives the most accurate results for the B3LYP/6-31+G(d,p)-PCM(opt) level of theory, showing a mean absolute deviation of 1.17 pK_a units and a mean deviation (MD) of −0.83 units.

To improve the accuracy of the results, a linear correction is applied. The correlation between the calculated and the experimental pK_a values is determined for each set of data, and then, the individual pK_a values are recalculated by inserting their values into the correlation equation. We refer to this process as

TABLE 9: Correlation between the Experimental and the Calculated Aqueous pK_a Values (at 298.15 K) of the Protonated Benzimidazoles at the Four Levels of Theory Chosen, Using Different Reaction Schemes (S1−4) and Equations

	S1 eq 9[a]	S2 eq 5[b]	S2 eq 9[b]	S2 eq 9[a]	S3 eq 5[b]	S3 eq 9[b]	S4 eq 5[b]	S4 eq 9[b]	S4 eq 9[a]
				MP2/6-31+G(d,p)//AM1-COSMO(sp)					
A[c]	0.287	0.309	0.287	0.287	0.308	0.287	0.309	0.287	0.287
B[c]	5.09	0.539	0.824	4.70	4.07	4.21	8.68	7.93	4.70
R^2	0.473	0.686	0.473	0.473	0.696	0.481	0.686	0.473	0.473
AD[d]	0.44	0.34	0.44	0.44	0.35	0.44	0.34	0.44	0.44
				HF/6-31+G(d,p)-PCM(opt)					
A[c]	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265
B[c]	3.79	4.20	4.20	3.42	4.30	4.30	4.87	4.87	3.43
R^2	0.762	0.761	0.762	0.762	0.773	0.773	0.761	0.762	0.762
AD[d]	0.26	0.26	0.26	0.26	0.27	0.27	0.26	0.26	0.26
				B3LYP/6-31+G(d,p)-COSMO(sp)					
A[c]	0.301	0.302	0.301	0.301	0.301	0.300	0.302	0.301	0.301
B[c]	4.48	−0.30	−0.25	4.06	4.09	4.10	7.80	7.83	4.07
R^2	0.818	0.817	0.818	0.818	0.827	0.828	0.817	0.818	0.818
AD[d]	0.25	0.25	0.25	0.25	0.26	0.25	0.25	0.25	0.25
				B3LYP/6-31+G(d,p)-PCM(opt)					
A[c]	0.281	0.281	0.281	0.281	0.281	0.281			0.281
B[c]	4.33	4.66	4.66	3.94	4.14	4.14			3.95
R^2	0.903	0.903	0.903	0.903	0.910	0.910			0.903
AD[d]	0.17	0.17	0.17	0.17	0.16	0.16			0.17

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Coefficients of the linear correlation equation of the form $pK_{a(experimental)} = ApK_{a(calculated)} + B$ and the correlation factor R^2. [d] Mean absolute deviation of calculated values from experimental aqueous pK_a values after application of the linear correction.

a linear correction to the data. Table 9 shows the coefficients (A and B) and R^2 values of the correlation equation for each set of data at the four levels of theory. The mean absolute deviation (AD) for the calculated values after applying the linear correction is also shown. The pK_a values calculated at the MP2/6-31+G(d,p)//AM1 level of theory show the lowest correlation to the experimental pK_a values. At this level of theory, the pK_a values calculated with eq 9 show a weaker correlation to the experimental values (R^2 of 0.473−0.481) than those calculated with eq 5 (R^2 of 0.686−0.696). At the HF/6-31+G(d,p)-PCM-(opt) level of theory, the correlations using both equations are slightly stronger and much more consistent. At the B3LYP/6-31+G(d,p)-COSMO(sp) and B3LYP/6-31+G(d,p)-PCM(opt) levels of theory, the correlation is much stronger and more consistent regardless of which reaction scheme or equation is used to calculate the pK_a values. The R^2 values obtained are between 0.817 and 0.828 at the B3LYP/6-31+G(d,p)-COSMO-(sp) level of theory, which indicates an adequate correlation. The correlation at the B3LYP/6-31+G(d,p)-PCM(opt) level of theory improves to an R^2 value of 0.903−0.910, which is a good correlation. The AD values for all the sets of data are very similar and are greatly improved with respect to the pK_a values calculated before the linear correction.

The best possible correlation between the calculated and experimental values should have a slope of 1 and a y-intercept of 0. In all of the data sets calculated, the slope of the correlation equation is approximately 0.3 and the y-intercept varies considerably depending on the reaction scheme used in the pK_a calculations.

From the data presented here, it appears that altering the equilibrium used to calculate the pK_a values shifts the correlation equation up or down (changes the y-intercept, B) but has very little effect on the quality of the correlation between the calculated and experimental values (R^2) or on the slope of the correlation equation (A). The pK_a calculations using Scheme 2 give y-intercepts that are the closest to 0 when the COSMO solvation model is used; however, the values calculated have a large positive deviation from the experimental values, so this is most likely coincidental. Figure 3 displays a typical plot of the linear correlation of the experimental aqueous pK_a values (at 298.15 K) of the protonated benzimidazoles under study to their calculated values. This figure shows the application of Scheme 2 (with experimental free energy values for water and the hydronium ion) and eq 9 at the B3LYP/6-31+G(d,p)-COSMO(sp) level of theory. The points have been numbered in ascendant order of experimental aqueous pK_a value.

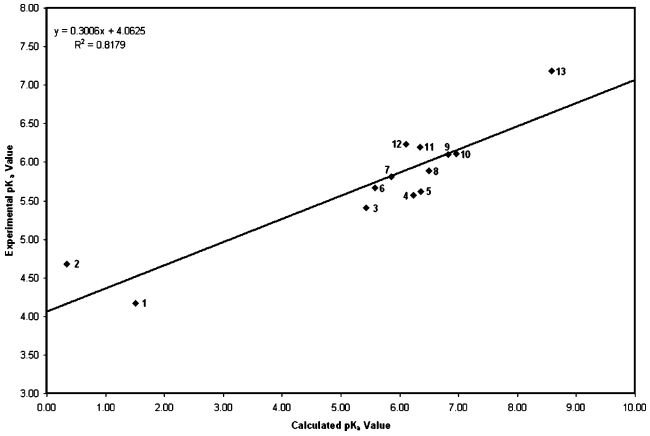

Figure 3. Plot of the correlation between the experimental aqueous pK_a values (at 298.15 K) of the protonated benzimidazoles under study and their calculated values using Scheme 2 and eq 9 at the B3LYP/6-31+G(d,p)-COSMO(sp) level of theory. The points have been numbered in ascendant order of experimental aqueous pK_a value.

The use of a linear correction significantly reduces the mean absolute deviation of the calculated values from the experimental pK_a values with respect to the deviation of the uncorrected values (as shown by comparing the AD values reported in Tables 5−8 with those in Table 9). However, this approach is limited by the accuracy of the method used to calculate the pK_a values (reaction scheme, pK_a expression, and level of theory) and the reliability of the experimental data. A general observation about the correlation plots can be made regarding the pK_a order of 5-nitrobenzimidazole and 2-chlorobenzimidazole. While the experimental results predict the protonated 5-nitrobenzimidazole to be more acidic than the protonated 2-chloro analogue, the

TABLE 10: Predictions for the Aqueous pK_a Values (at 298.15 K) of the Protonated 5-Chlorobenzimidazole and 2-Methoxybenzimidazole Using Various Methodologies and Correlations from Table 9

		5-chlorobenzimidazole		2-methoxybenzimidazole	
level of theory	scheme/eq	direct	correlated	direct	correlated
B3LYP/6-31+G(d,p)-COSMO(sp)	S1 eq 9[a]	2.61	5.26	3.36	5.49
	S2 eq 5[b]	18.41	5.26	19.26	5.52
	S2 eq 9[b]	18.34	5.26	19.09	5.49
	S2 eq 9[a]	4.00	5.26	4.74	5.49
	S3 eq 5[b]	3.96	5.29	4.82	5.55
	S3 eq 9[b]	3.98	5.29	4.73	5.51
B3LYP/6-31+G(d,p)-PCM(opt)	S1 eq 9[a]	3.30	5.26	4.51	5.60
	S2 eq 5[b]	2.13	5.26	3.34	5.60
	S2 eq 9[b]	2.13	5.26	3.34	5.60
	S2 eq 9[a]	4.69	5.26	5.89	5.60
	S3 eq 5[b]	4.07	5.28	5.28	5.62
	S3 eq 9[b]	4.07	5.28	5.28	5.62
experimental		4.86[c]			
predicted range		4.68−5.41		6.10−7.18	

[a] Values of G_{gas} and ΔG_{solv} for OH^-, H_2O, H_3O^+, and H^+ are taken from the literature.[28] [b] The pK_a values are calculated using only ab initio data. [c] Reference 8.

calculated pK_a values in all the cases attempted predict the opposite acidic order. Most likely, theoreticians would send this case back to the experimentalist for further determinations, but another publication by our group on this topic,[38] exploring QSPR approaches for pK_a calculations, shows that when either certain atomic charges or orbital energies are included in the correlation with the experimental pK_a values (together with the calculated aqueous change in the Gibbs free energy for the acid−base process), the expected pK_a order between the protonated 5-nitrobenzimidazole and the 2-chloro analogue is obtained. Substituent effects modify the way molecules interact with each other (e.g., in solute−solvent interactions) from electrostatic and/or orbitalic points of view. This situation might seem to indicate that certain orbitalic and electrostatic factors that influence the thermodynamics of a chemical (in this case, acid−base) equilibrium are not taken into account when performing direct pK calculations. We are currently exploring this topic in more depth.

3.3. Predicting Aqueous pK_a Values of Protonated Benzimidazoles. Calculations are performed for two additional benzimidazole derivatives not shown in Table 1: 5-chlorobenzimidazole and 2-methoxybenzimidazole. Experimental aqueous pK_a values could not be found for these two molecules at the time the calculations were performed; however, an experimental value for the aqueous pK_a of 5-chlorobenzimidazole of 4.86 was subsequently located.[8] Predictions are made for the pK_a values of these two compounds using the B3LYP/6-31+G(d,p)-COSMO(sp) and B3LYP/6-31+G(d,p)-PCM(opt) levels of theory. The results of the direct pK_a calculations using Schemes 1, 2, and 3 (with both pK_a equations), as well as the calculations using the correlation equations (see Table 9), are summarized in Table 10.

The acidities of 5-chlorobenzimidazole and 2-methoxybenzimidazole can be estimated from substituent effects. Substituting chlorine onto benzimidazole increases its acidity due to the electron-withdrawing effect of chlorine. At position 5, the substituent effect is smaller than at position 2 (beside the acidic nitrogen), so the pK_a of 5-chlorobenzimidazole should be between that of 2-chlorobenzimidazole (4.68) and benzimidazole (5.41). A similar effect, but in the opposite direction, can be seen between 5-aminobenzimidazole (pK_a = 6.11) and 2-aminobenzimidazole (pK_a = 7.18) and between 5-methylbenzimidazole (pK_a = 5.81) and 2-methylbenzimidazole (pK_a = 6.10). Electron donating groups (NH_2 and CH_3) at position 5 reduce the acidity of protonated benzimidazole to a lesser extent than at position 2.

A methoxy group has an electron-donating effect larger than that of a methyl group, which would make the pK_a higher than that of 2-methylbenzimidazole (6.10). The electron-donating effect of the methoxy group should be smaller than that of an amino group, so the pK_a would be lower than the pK_a of 2-aminobenzimidazole (7.18).

In the direct and correlated pK_a calculations, the relative pK_a ordering of the two benzimidazoles is correct. However, the direct predictions fell outside the estimated upper and lower pK_a limits in every case. The correlation equations at the two levels of theory produced a pK_a value for 5-chlorobenzimidazole that fell well within the estimated upper and lower bounds of the aqueous pK_a value. The pK_a value for 2-methoxybenzimidazole is improved by using the linear correction, but the values still fell outside the estimated upper and lower bounds.

4. Conclusions

A variety of methodologies for the accurate theoretical determination of aqueous pK_a values of protonated benzimidazoles have been explored. To the best of our knowledge, this is the first theoretical study on this topic for this family of compounds.

Aqueous pK_a values are calculated for a group of thirteen benzimidazole derivatives using four equilibria and two equations at four levels of theory. The accuracy of the directly calculated pK_a values is good in some cases but probably insufficient for practical applications. It is found that, in general, the accuracy of the direct calculations is virtually independent of the level of theory applied and appears to be determined by the reaction scheme used.

Furthermore, it is shown via several examples that the two pK_a equations used (eq 5 and 9) produce the same results when solvent effects are included in both the geometry optimization and frequency calculations and that very similar results are obtained when the optimizations are carried out in the gas phase followed by a single point including solvent effects. Our results show that these equations could be used interchangeably, even though, to the best of our knowledge, eq 5 has not been used in the literature for pK_a calculations.

Linear corrections produce pK_a values with a lower mean absolute deviation from experimental pK_a values than the noncorrected calculations in all cases. The strongest correlations between experimental and calculated data are obtained at the B3LYP/6-31+G(d,p)-PCM(opt) level of theory.

The predictive capabilities of all the methodologies attempted are tested with two compounds that were not included in the

set of benzimidazoles initially investigated. The direct calculations differ significantly from the expected values, but the pK_a values calculated using the correlation equations are very similar and in reasonable agreement with the expected pK_a values.

Acknowledgment. We gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) and Thompson Rivers University (CUEF program) for financial support. Thanks are also due to Dr. K. Donkor for valuable discussions and for providing us with previous experimental results and bibliography on this topic. We are very appreciative of the referees' thorough comments on our paper and of the important issues they raised.

Supporting Information Available: Tables S1−S5 summarizing the raw data used in the aqueous pK_a calculations performed on the compounds studied. The Appendix section shows five examples of how the pK_a values are calculated in this paper using different equations, reaction schemes, and levels of theory. This information is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Stumm, W.; Morgan, J. J. *Aquatic Chemistry: Chemical Equilibria and Rates in Natural Waters*; Wiley-Interscience: New York, 1996.

(2) Thomas, G. *Medicinal Chemistry: An Introduction*; John Wiley & Sons: West Sussex, 2000.

(3) Donkor, K. K.; Kratochvil, B. *J. Chem. Eng. Data* **1993**, *38*, 569.

(4) Some examples: (a) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314. (b) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421. (c) Murlowska, K.; Sadlej-Sosnowska, N. *J. Phys. Chem. A* **2005**, *109*, 5590.

(5) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.

(6) Some examples: (a) Duchowicz, P. R.; Castro, E. A. *Mendeleev Comm.* **2002**, *5*, 187. (b) Citra, M. J. *Chemosphere* **1999**, *38*, 191. (c) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963. (d) Hennemann, M.; Clark, T. *J. Mol. Model*. **2002**, *8*, 95.

(7) Soriano, E.; Cerdán, S.; Ballesteros, P. *J. Mol. Struct. (THEOCHEM)* **2004**, *684*, 121.

(8) Kapinos, L. E.; Song, B.; Sigel, H. *Chem.—Eur. J.* **1999**, *5*, 1794.

(9) Silva, C. O.; Nascimento, M. A. C. *Adv. Chem. Phys.* **2002**, *123*, 423.

(10) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.

(11) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.

(12) Young, D. C. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*; John Wiley and Sons: New York, 2001.

(13) Some recent examples: (a) Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. *J. Phys. Chem. A* **2004**, *108*, 166. (b) Fu, Y.; Liu, L.; Li, R.; Liu, R.; Guo, Q. *J. Am. Chem. Soc.* **2004**, *126*, 814. (c) Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717.

(14) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(15) Some examples: (a) Özden, S.; Atabey, D.; Yıldız, S.; Göker, H. *Bioorg. Med. Chem.* **2005**, *13*, 1587. (b) Kazimierczuk, Z.; Andrzejewska, M.; Kaustová, J.; Klimešová, V. *Eur. J. Med. Chem*. **2005**, *40*, 203. (c) Özden, S.; Karataş, H.; Yıldız, S.; Göker, H. *Arch. Pharm. Pharm. Med. Chem.* **2004**, *337*, 556. (d) Klimešová, V.; Koči, J.; Pour, M.; Stachel, J.; Waisser, K.; Kaustová, J. *Eur. J. Med. Chem*. **2002**, *37*, 409.

(16) Lednicer, D. *Strategies for Organic Drug Synthesis and Design*; Wiley-Interscience: New York, 1998.

(17) Some examples: (a) Zhang, D.; Gao, L.; Zhou, G. *Corros. Sci.* **2004**, *46*, 3031. (b) Popova, A.; Christov, M.; Raicheva, S.; Sokolova, E. *Corros. Sci.* **2004**, *46*, 1333. (c) López, D. A.; Simison, S. N.; de Sánchez, S. R. *Electrochim. Acta* **2003**, *48*, 845.

(18) Eicher, T.; Hauptmann, S. *The Chemistry of Heterocycles*; Wiley-VCH: Weinheim, 2003.

(19) Taft, R. W.; Catalán, J.; Claramunt, R. M.; Elguero, J.; Laynez, J.; Menéndez, M.; Anvia, F.; Quian, J. H.; Taagepera, M. *J. Am. Chem. Soc.* **1988**, *110*, 4105.

(20) Schenkeveld, S.; Donkor, K. K. *Spectrometric Determination of Aqueous Ionization Constants (pK_a) of Benzimidazoles*; Directed Studies Report, Thompson Rivers University Library, 2002.

(21) Hofman, K. *Imidazole and its Derivatives*; Interscience Publishers: New York, 1953.

(22) (a) Starkey, R.; Norman, J.; Hintze, M. *J. Chem. Educ.* **1986**, *63*, 473. (b) Campbell, M. L.; Waite, B. A. *J. Chem. Educ.* **1990**, *67*, 386. (c) Delange, A. M.; Potgieter, J. H. *J. Chem. Educ.* **1991**, *68*, 304.

(23) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241.

(24) Laidler, K. J.; Meiser, J. H.; Sanctuary, B. C. *Physical Chemistry*; Houghton Mifflin: Boston, 2003.

(25) Klamt, A.; Eckert, F.; Diedenhofen, M. *J. Phys. Chem. A* **2003**, *107*, 9380.

(26) 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.

(27) (a) Pliego, J. R., Jr. *Chem. Phys. Lett.* **2003**, *367*, 145. (b) Pliego, J. R., Jr. *Chem. Phys. Lett.* **2003**, *381*, 246.

(28) Palascak, M. W.; Shields, G. C. *J. Phys. Chem. A* **2004**, *108*, 3692.

(29) 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.

(30) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(31) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(32) (a) B3 functional: Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) LYP functional: Lee, C.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.

(33) Ditchfie, R.; Hehre, W. J.; Pople, J. A. *J. Chem. Phys.* **1971**, *54*, 724.

(34) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(35) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032.

(36) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799.

(37) Frisch, Æ.; Frisch, M.; Trucks, G. W. *Gaussian 03 User's Reference*; Gaussian Inc.: Wallingford, CT, 2003.

(38) Brown, T. N.; Mora-Diez, N. to be submitted for publication.