JCTC Journal of Chemical Theory and Computation

Article

Cite This: J. Chem. Theory Comput. 2019, 15, 6025−6035

pubs.acs.org/JCTC

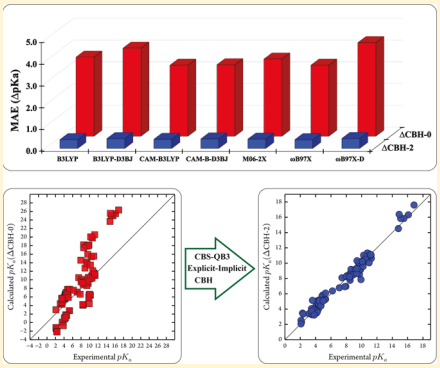
Accurate pK_a Evaluations for Complex Bio-Organic Molecules in Aqueous Media

Bishnu Thapa and Krishnan Raghavachari*

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

Supporting Information

ABSTRACT: Despite the numerous computational efforts on estimating acid dissociation constant (pK_a's), an accurate estimation of pK_a's of bio-organic molecules in the aqueous medium is still a challenge. The major difficulty lies in the accurate description of the aqueous environment and the cost and accuracy of quantum mechanical (QM) methods. Herein, we report a well-defined quantum chemical protocol for accurately calculating pK_a's of a wide range of bio-organic molecules in aqueous media. The performance of our method has been assessed using test sets containing molecules with a range of sizes and a variety of functional groups, including alcohols, phenols, amines, and carboxylic acids, and obtained an impressive mean absolute accuracy of 0.5 pK_a units. For the smaller molecules, we use a high-level QM method (CBS-QB3) and a calibrated explicit−implicit solvation model that yields accurate pK_a values for a range of functional groups. For the larger molecules, we combine this approach with an efficient error-cancellation scheme that eliminates the systematic errors in different density functional methods to yield accurate pK_a values for simple to complex molecular systems. Our protocol is efficient, applicable to large molecules, covers all the common functional groups present in bio-organic molecules, and should find widespread applications in diverse research areas including drug−protein binding, catalysis, and chemical synthesis.



1. INTRODUCTION

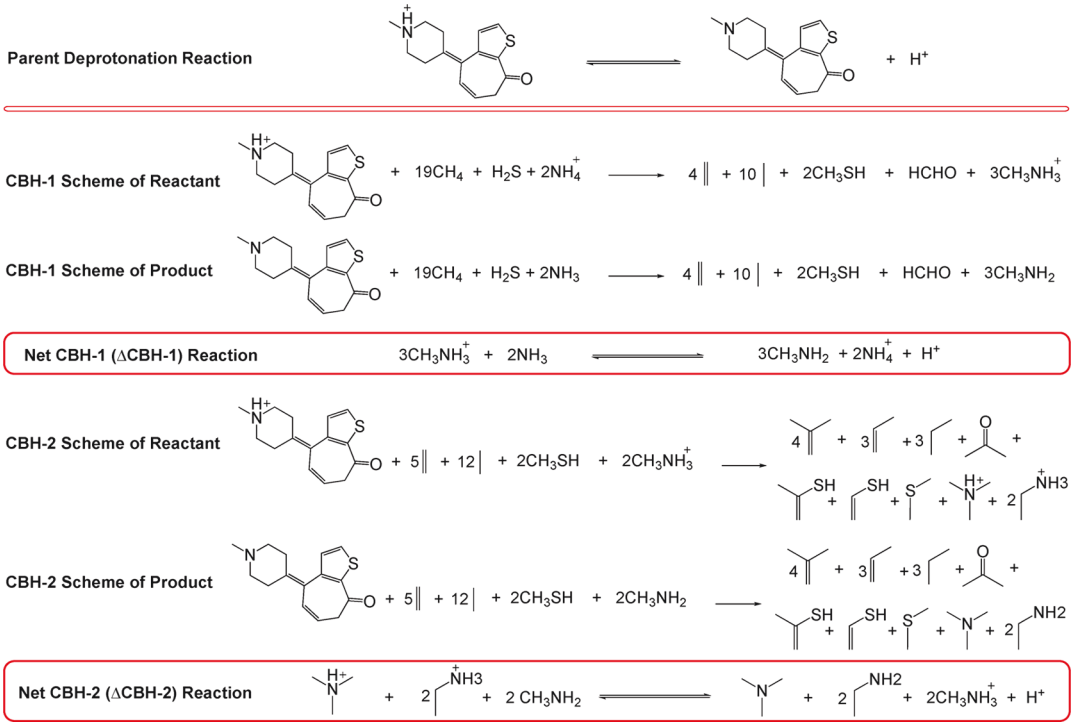
Knowledge of the acid dissociation constant (pK_a value) of a molecule is essential to understand its stability as well as reactivity. In chemical processes such as drug−protein binding, catalysis, and chemical synthesis, the pK_a can determine the fate of a chemical reaction by altering the overall reaction mechanism.[1−5] Performing accurate experimental pK_a measurements in aqueous solutions is not always feasible due to the problems associated with synthetic difficulties, isolation and purification, interference from other solutes and the solvent environment, etc.[6] Additionally, it may be advantageous to screen the physicochemical properties of chemical compounds for potential applications without having to synthesize the actual molecules. In this regard, quantum mechanical (QM) calculations can be useful to obtain the desired property with sufficiently high accuracy.[7]

A significant number of efforts have been made for a quantum mechanical calculation of the pK_a values of various bio-organic and organometallic molecules.[8−13] In general, solute−solvent interactions are described by using continuum solvation models to make the QM calculations tractable.[6,14−17] The accuracy of the calculated pK_a value primarily depends on the quality of the QM method used and the solvation model employed to describe the solute−solvent interaction.[10] It has been shown that high-quality pK_a values can be obtained if a correct level of theory along with an appropriate solvation model is selected for the particular functional group of interest.[18−24] However, it becomes a real challenge, even for the experts, to select a method to use from a plethora of available QM methods. Moreover, the insufficient accuracy of empirically fitted implicit solvation models further adds to the source of errors in the pK_a calculations.[6,11,12] Some of these issues have been addressed, to a limited extent, by several research groups through the extensive screening of QM methods for a particular functional group,[20,23] correcting systematic errors via a linear regression fit approach,[25−28] using reference acids,[12,29−32] reparameterizing the implicit solvent parameters,[33−36] or using explicit−implicit solvation.[27,37−44] Since there are a large number of small molecules with experimentally measured pK_a values, and the errors in QM-calculated deprotonation free energies are somewhat systematic for a specific functional group, the free energy changes can be calculated with a cheaper QM method and fitted separately for different functional groups to obtain a reasonable overall accuracy. A similar approach has been used in Jaguar pK_a,[45] a popular pK_a prediction software, that uses experimental pK_a values of 1134 small molecules to obtain linear fit equations for approximately 100 specific functional groups and uses that information to estimate the pK_a of a new molecule. However, there is no single QM method-based computational protocol that works uniformly for most of the commonly studied bio-organic molecules without such corrections.

Received: June 18, 2019
Published: October 9, 2019

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

Scheme 1. Illustrative Example Showing the Derivation of ΔCBH-1 and ΔCBH-2 Reaction Schemes for a Deprotonation Reaction



In this study, we present a novel and computationally efficient protocol for calculating accurate pK_a values of small (and simple) to larger (more complex) bio-organic molecules containing a wide variety of functional groups including alcohols, amines, thiols, and carboxylic acids. In particular, we address some of the major questions in the current literature regarding pK_a calculations:

(1) Is it possible to have a single theoretical protocol that works accurately for a wide range of functional groups?
(2) Can we obtain accurate pK_a values for a functional group of interest using a range of QM methods, particularly with a variety of density functionals?
(3) If the protocol is accurate, is it computationally affordable?

The protocol presented here corrects for the systematic errors in calculated pK_a values associated with various DFT (or WFT) methods and does not require any input from experimental data. The protocol is based on a modified version of our connectivity-based hierarchy (CBH)[46,47] approach, where the systematic error in the calculated pK_a (for a given molecule with a specified QM scheme) is identified as the intrinsic error in the free energy change of an elementary reaction that represents the fundamental chemical change during deprotonation (see refs 46 and 47 for details about the CBH scheme). This enables accurate pK_a calculations on a large number of relatively complex molecules with a mean absolute deviation of 0.5 pK_a units from the experiment.

Our proposed scheme is accomplished in two major steps. We first develop a standard and accurate pK_a calculation protocol for small molecules using a single reliable QM method (CBS-QB3) in conjunction with an explicit−implicit (SMD) solvation model that is calibrated and applicable for small-to-medium-sized molecules containing various common functional groups. The explicit−implicit solvation approach adopted here has been shown to represent the solute−solvent interactions much more accurately compared to the implicit-only solvation approach.[48] We then formulate a well-defined CBH-based systematic error-correction scheme that enables accurate pK_a's of significantly larger bio-organic molecules in the aqueous medium.

The following sequence represents our overall proposed scheme.

(1) Calculate the pK_a of the parent deprotonation reaction (PR) using a selected low-level (DFT or WFT) method.

$$pK_a^{low\_qm,implicit}(PR) = \frac{\Delta G^*_{deprot}}{2.303RT} \quad (1)$$

where ΔG^*_{deprot} is the free energy change for a deprotonation reaction in the aqueous medium calculated at the low-level QM (DFT or WFT) method, R is the gas constant, and T is the temperature (see Section 2 for details). Note that these calculations only involve implicit solvation models and are computationally highly efficient.

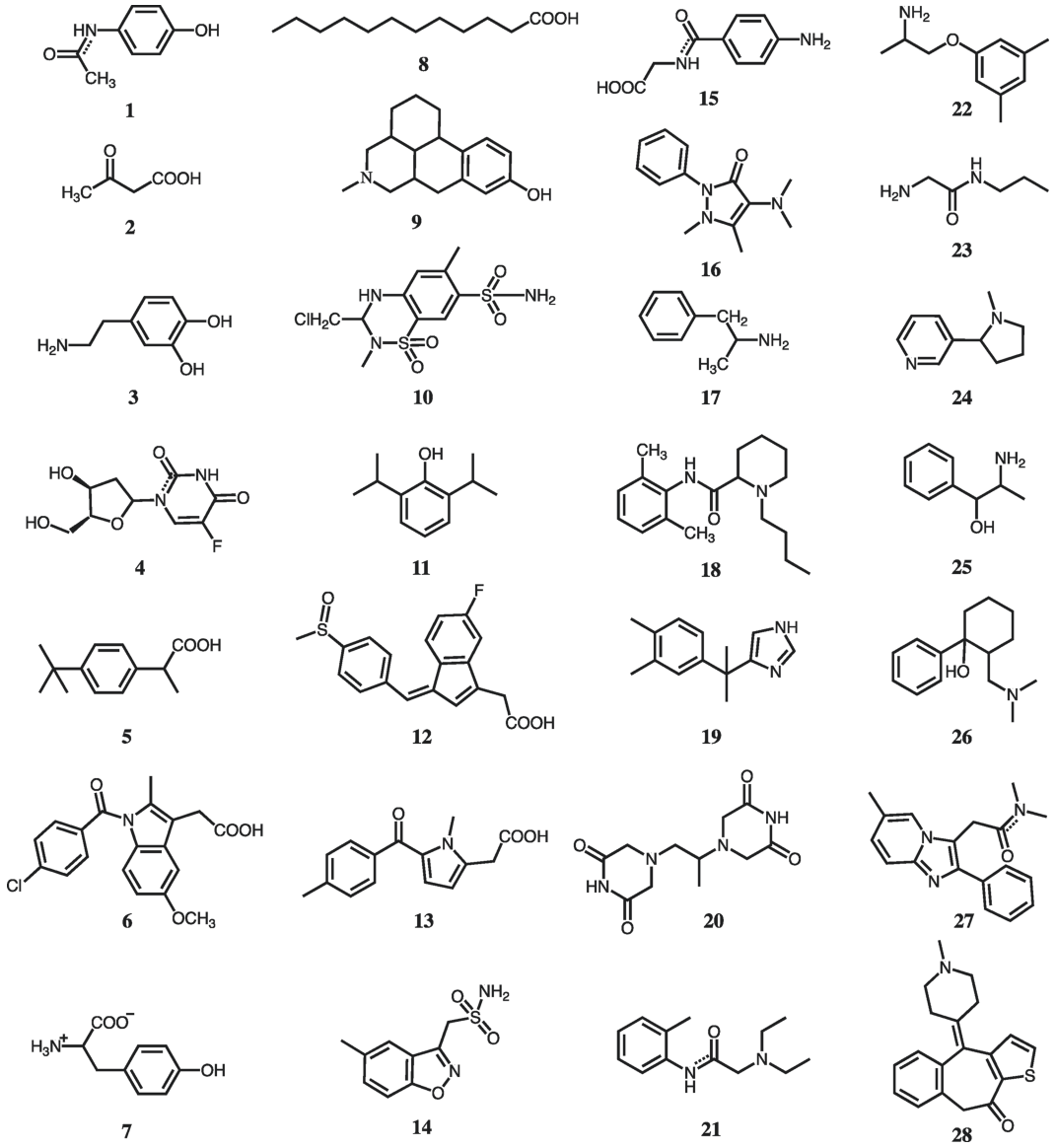
(2) Compare the calculated pK_a with the experimentally measured pK_a for the same reaction to obtain the raw error for the given method. This error is denoted as "ΔpK_a-0".

$$\Delta pK_a\text{-}1 = pK_a^{low\_qm,implicit}(PR) - pK_a^{experiment}(PR) \quad (2)$$

(3) To correct the error associated with the given low-level method, construct the CBH reaction scheme at the desired CBH-n rung (here, only n = 1 and 2 are considered) to find out the elementary reaction involved in deprotonation (Scheme 1 and Scheme S1). CBH rungs are constructed for both the protonated and deprotonated species based on the local bonding

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

Scheme 2. Molecules Used in Set-III



structure and connectivity of the molecule. Each increasing CBH rung preserves a larger local environment in the CBH fragments and yields an increasing level of sophistication. The simplest rung is the isogyric scheme (i.e., CBH-0; not considered in this study) where each heavy (nonhydrogen) atom is extracted in its saturated valance state (e.g., CH_4, NH_3, and H_2O). The next rung, CBH-1, is the same as the popular Pople's isodesmic scheme.[49] In this rung, fragments are constructed by extracting all the heavy atom−heavy atom bonds in the molecule and saturated to satisfy their valence states (e.g., CH_3−OH, CH_2=CH_2, and CH_3−SH). In the CBH-2 scheme, the immediate connectivity of each heavy atom is included to preserve the local bonding environment (Scheme 1 and Scheme S1). Similarly, higher rungs of CBH such as CBH-3 and CBH-4 can be constructed, preserving successively larger local bonded environments.

(4) The net reactions derived from the desired CBH-*n* rung are denoted as ΔCBH-*n* (e.g., ΔCBH-1 for rung 1, and ΔCBH-2 for rung 2) (Scheme 1 and Scheme S1). Since the structural change during deprotonation is small, a significant part of the reactant and product molecules remains identical and gets canceled (vide infra). Thus, the sizes of the molecules involved in these net reactions are quite small.

(5) Compute the pK_a of the elementary deprotonation reaction obtained from the CBH-*n* scheme with both the current DFT (or WFT) method and an accurate reference method (CBS-QB3 in this work) to obtain the systematic error correction (vide infra). All the CBH fragments (both protonated and deprotonated species) involved in the net CBH-*n* reaction are geometry-optimized independently at the respective level of theory. Note that the reference CBS-QB3 calculations are done with an explicit−implicit solvation model but are computationally feasible due to the small sizes of the molecules involved in the elementary reactions. The difference in the calculated pK_a values between the current low-level (DFT or WFT) method and the reference method for the elementary reaction is the CBH-*n* correction at the given CBH rung, which is then

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

subtracted from the raw low-level pK_a value ($pK_a^{low\_qm,\ implicit}$(PR)) to obtain the CBH-corrected pK_a value for the parent reaction. A fully worked example of CBH scheme including the raw calculated as well as corrected pK_a values is given in the Supporting Information. The overall pK_a and corrections at various rungs are calculated as follows:

For the CBH-1 scheme:

$$\Delta\text{CBH-1 correction} = pK_a^{low\_qm,implicit}(\Delta\text{CBH-1}) - pK_a^{CBS-QB3,explicit-implicit}(\Delta\text{CBH-1}) \quad (3)$$

Calculated pK_a of the parent reaction with CBH-1 correction:

$$pK_a\text{-1} = pK_a^{low\_qm,implicit}(\text{PR}) - \Delta\text{CBH-1 correction} \quad (4)$$

Error in the calculated pK_a of parent reaction with CBH-1 correction:

$$\Delta pK_a\text{-1} = pK_a\text{-1} - pK_a^{experiment}(\text{PR}) \quad (5)$$

Similarly, for the CBH-2 scheme:

$$\Delta\text{CBH-2 correction} = pK_a^{low\_qm,implicit}(\Delta\text{CBH-2}) - pK_a^{CBS-QB3,explicit-implicit}(\Delta\text{CBH-2}) \quad (6)$$

Calculated pK_a of the parent reaction with CBH-2 correction:

$$pK_a\text{-2} = pK_a^{low\_qm,implicit}(\text{PR}) - \Delta\text{CBH-2 correction} \quad (7)$$

Error in calculated pK_a of the parent reaction with CBH-2 correction:

$$\Delta pK_a\text{-2} = pK_a\text{-2} - pK_a^{experiment}(\text{PR}) \quad (8)$$

An alternative possibility that we have explored is to use the experimental pK_a as a reference in the elementary CBH reaction if it is available for all of the species involved in the net pK_a reaction, but we show (vide infra) that our protocol works even more effectively without using any experimental data for error correction.

For most of the molecules considered in this study, the fragments involved in each rung of the CBH scheme can be uniquely identified. However, for molecules containing functional groups that are conjugated with the aromatic rings such as phenols, anilines, benzoic acids, etc., we propose a modified version of the conventional CBH scheme for a more effective treatment of the aromatic environment (Scheme S1). If the group undergoing deprotonation is directly bonded to an aromatic ring, the full aromatic ring is considered as a single group. This helps to preserve the electron delocalization within a fragment. On the other hand, rings that are far from the deprotonation site are not affected since their chemical environment is identical in the protonated and deprotonated states and get canceled in the net CBH reaction.

2. METHODS

2.1. Test Sets. To develop a reliable protocol and demonstrate the applicability of our approach, we constructed three test sets of molecules containing a wide range of functional groups such as amine, alcohol, phenol, thiol, carboxylic acid, etc. (see Scheme 2 and Tables S1−S11), which have the experimentally measured pK_a values in the range of −1 to 20. The first set (**Set-I**) was assembled to establish a standard protocol for calculating accurate pK_a's using a reliable computational model (CBS-QB3) in conjunction with an explicit−implicit model for solvation (SMD[50] solvation model for implicit solvation). It comprises a total of 224 small bio-organic molecules containing 2−12 heavy atoms, with 22 nitrogen-containing aromatic heterocycles (e.g., imidazole, pyridine, and pyridazine), 15 alcohols, 14 thiols, 26 carboxylic acids, 10 primary amines, 8 secondary amines, 5 tertiary amines, 39 phenols, 13 thiophenols, 34 anilines, 10 carbon acids (acidity of α-CH of aldehydes, ketones, and amides), and 28 benzoic acid molecules (Tables S1−S11). In the second test set (**Set-II**), a total of 72 molecules were selected to benchmark the application of the CBH scheme for calculating pK_a's of larger molecules (Table S11). Molecules in this set were selected such that most of the fragments involved in the elementary reactions of the CBH-2 scheme have experimentally measured pK_a's. This helps to calibrate the performance of error-correction schemes based on experiments versus those obtained without recourse to experiments. The third dataset (**Set-III**) includes a total of 28 drug molecules with experimentally measured pK_a's for the parent molecule but not for all the CBH fragments (Scheme 2) and serves as a target dataset to assess the performance of the protocol developed in this study. Some of the molecules in this set have multiple ionizable functional groups. For those molecules, separate microscopic pK_a values were calculated for each of the ionizable groups (see Figures S5−S11), and the value that is closest to the experiment was used for the comparison. While our assignments are unambiguous, a better agreement between the calculated and experimental pK_a values could potentially be achieved using the contributions from multiple ionization states.

2.2. pK_a Calculation. For a deprotonation reaction, BH⇌ B^- + H^+, the acid dissociation constant (pK_a) is calculated as

$$pK_a = \frac{\Delta G^*_{aq}}{2.303RT} \quad (9)$$

where R is the gas constant, T is the temperature (298.15 K), and ΔG^*_{aq} is the aqueous phase free energy change for the deprotonation reaction. ΔG^*_{aq} is calculated as

$$\Delta G^*_{aq} = G^*_{B^-,aq} + G^*_{H^+,aq} - G^*_{BH,aq} \quad (10)$$

where $G^*_{B^-,aq}$, $G^*_{H^+,aq}$, and $G^*_{BH,aq}$ are, respectively, the free energy of deprotonated species (B), free energy of proton (H^+), and the free energy of acid (BH) in aqueous medium. $G^*_{B^-,aq}$ and $G^*_{BH,aq}$ are calculated directly in aqueous medium using the SMD implicit solvation and the QM methods discussed below. No thermodynamic cycle was used to calculate the free energies in solution since it has been recommended in many studies.[16,20,40,42,51] The aqueous phase free energy of proton ($G^*_{H^+,aq}$) is calculated as

$$G^*_{H^+,aq} = G^o_{H^+,gas} + \Delta G^*_{H^+,solv} + \Delta G^{1atm\rightarrow 1M} \quad (11)$$

where $\Delta G^*_{H^+,solv}$ = −265.9 kcal/mol is the solvation energy of a proton, taken from the literature,[52−55] $G^o_{H^+,gas} = H^o_{gas} - TS^o_{gas}$ is the free energy of a proton in the gas phase with $H^o_{gas} = 5/2RT$ = 1.48 kcal/mol and S^o_{gas} = 26.05 cal/(mol·K), and $\Delta G^{1atm\rightarrow 1M}$ = 1.89 kcal/mol is the free energy corresponding to the change

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

in standard state of 1 atm in the gas phase to 1 M in the aqueous phase.

2.3. Computational Details. The pK_a calculations were performed using seven popular density functional methods [four hybrid methods (B3LYP,[56−58] ωB97X,[59] M06-2X, and CAM-B3LYP) and three of their empirical dispersion-corrected variants (B3LYP-D3BJ, ωB97X-D, and CAM-B3LYP-D3BJ)[60,61]] with the 6-311++G(d,p)[62−66] basis set, as well as a reliable compound model chemistry, CBS-QB3.[67] Molecular geometry optimization was carried out in the aqueous medium with the SMD[50] implicit solvation model. For the CBS-QB3 method, geometry optimizations and frequency calculations were performed using the B3LYP/6-311++G(d,p) level, followed by a single-point energy calculation at the CBS-QB3 level. Harmonic frequencies were examined to confirm that the calculated structures are local minima on the potential energy surface. Entropies and thermal corrections were calculated under the standard protocol from statistical thermodynamics of an ideal gas with rigid rotor/harmonic oscillator approximation at 298.15 K. Vibrational frequencies were scaled by 0.99 to match with the CBS-QB3 method for the calculation of zero-point energies, entropies, and thermal corrections. The effect of low-frequency vibration in calculated pK_a was also investigated using the quasi-harmonic oscillator approximation (QHA)[68] (see frequency scaling section of the Supporting Information for details). Comparing the result obtained for the set of 224 molecules (Figures S1 and S2 of the Supporting Information), it is observed that the cutoff value of 100 cm^{-1} improved the performance on average by 0.3 pK_a units compared to QHA-corrected frequencies. Therefore, for the final pK_a calculations, the vibrational modes with lower frequencies were scaled to 100 cm^{-1} before computing their contribution to the Gibbs free energy. For the molecules with multiple rotatable bonds, a conformational search was performed in both the protonated and deprotonated forms using the MMFF94 force field[69] with a 7.0 kcal/mol energy window using the GMMX conformation search module implemented in GaussView.[70] The low-energy conformers were further refined by optimizing the structures at B3LYP/6-31+G(d,p) in aqueous solution including SMD implicit solvation (without explicit solvent molecules) to obtain the lowest-energy conformer. Only the lowest-energy conformers in the protonated and deprotonated forms were used in the pK_a calculations. The effect of including higher-energy conformers in the calculated pK_a was not explored; however, we do realize that including contribution from some of the higher energy conformers can be important in some cases. All the calculations were performed using the Gaussian 16 program suite.[71]

3. RESULTS AND DISCUSSION

3.1. pK_a Calculation Using Explicit−Implicit Solvation. As outlined above, to obtain accurate pK_a's using the CBH scheme, it is necessary to have accurate pK_a values for the elementary reactions. While a large number of small organic molecules have experimentally measured pK_a values, it is not always possible to find experimental pK_a values for all necessary CBH fragments. Despite a significant number of theoretical methods that have been developed and tested over three decades, surprisingly, no pK_a protocol based on a single QM method works satisfactorily across the full range of common functional groups. In particular, it has been demonstrated in recent years that implicit solvent models are not adequate to calculate pK_a's reliably, and at least a few explicit water molecules need to be included.[6,18,20,40−42,50] Therefore, we have benchmarked a standard protocol for calculating accurate pK_a with SMD implicit solvation in conjunction with a few explicit solvent molecules using a single reliable QM method. More specifically, free energy calculations were performed at the CBS-QB3 level of theory with the SMD implicit solvation model (see Section 2 for computational details).

To explore the effect of explicit water molecules on the calculated pK_a's and to obtain the optimum number of explicit water molecules required, we computed the pK_a's by including zero, one, two, and three explicit water molecules placed near the functional group being protonated/deprotonated (e.g., $-OH/-O^-$, $-NH_3^+/-NH_2$) along with the SMD implicit solvation model (Figure 1). For the deprotonation of carbon acids (α-H of aldehydes, ketones, and amides), explicit waters were placed only near the carbonyl oxygen since the deprotonated species (i.e., carbanion) is expected to be in resonance with its more stable enolate anion. Our study is

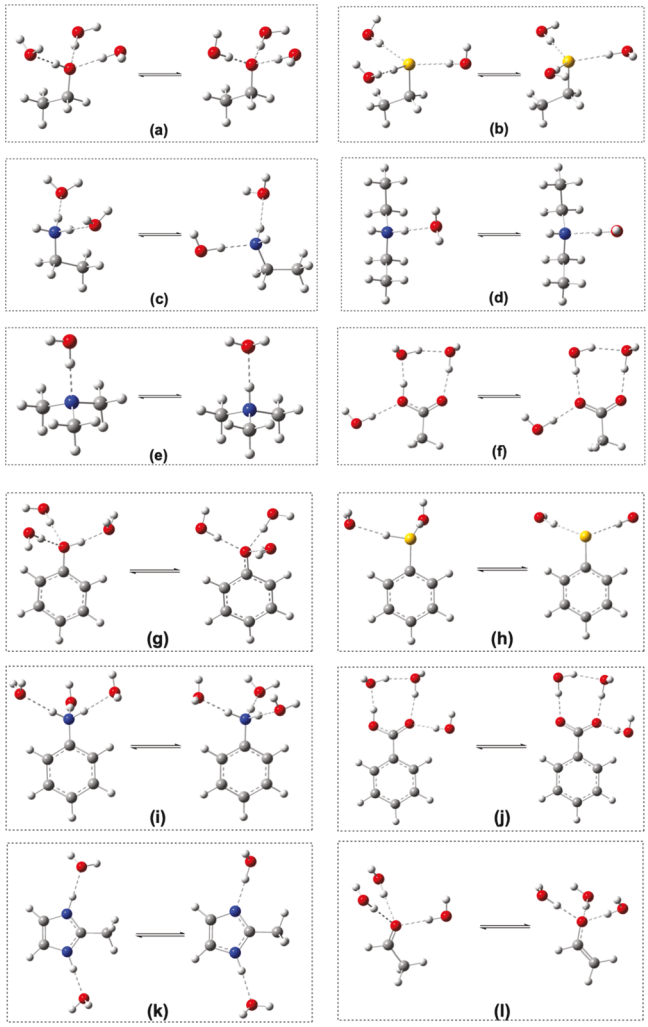


Figure 1. Arrangement of hydrogen-bonded explicit water molecules with the protonated and deprotonated states of (a) aliphatic alcohol, (b) thiol, (c) primary amine, (d) secondary amine, (e) tertiary amine, (f) carboxylic acid, (g) phenol, (h) thiophenol, (i) aniline, (j) benzoic acid, (k) nitrogen-containing aromatic ring (imidazole in this case), and (l) carbon acid.

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

limited to including the explicit water molecules in the first solvation shell based on the assumption that the bulk solvation properties are captured with implicit solvation. Arrangement of water molecules around the various functional groups is shown in Figure 1. Both the protonated and deprotonated states were solvated with an equal number of explicit solvent molecules to maintain the same, balanced solvent environment. For the initial guess of explicit−implicit solvation, water molecules were placed along the (O/N/S)−H bonds and/or in a tetrahedral arrangement with each water, forming only one hydrogen bond with the solute as shown the Figure 1, and geometry optimization was carried out at the B3LYP/6-311++G(d,p) level of theory. For the explicit solvation of nitrogen-containing aromatic compounds, only one water molecule is placed per N/O/S atom in the same plane of the ring, forming only one hydrogen bond. The arrangement of explicit water molecules near the protonated and deprotonated species was kept as consistent as possible (i.e., same number of hydrogen bonds and similar conformations) to avoid any bias that may result in unsystematic contributions to the calculated pK_a values. Maintaining the same explicit−implicit solvation environment on both protonated and deprotonated species also helps treat the solvent hydrogen bonding environment in a more balanced way and cancels some of the systematic errors introduced by the inclusion of explicit waters (e.g., explicit water conformations and contributions from additional low-frequency modes from explicit waters). This approach is slightly different from the commonly used explicit−implicit solvation approach where explicit solvents are added only to the charged species. Errors in pK_a values calculated with zero to three explicit water molecules with respect to the experiment are shown in Figure 1 and Table 1 in the paper and Tables S1−S11 in the Supporting Information. The results show that the required number of water molecules is slightly different for different functional groups to achieve the desired accuracy of <1 pK_a unit.

Table 1. Mean Absolute Error (MAE) in Calculated pK_a's Using the Optimum Number of Explicit Water Molecules and SMD Implicit Solvation Model at the CBS-QB3 Level of Theory for the Molecules in Set-I

functional group	total number of molecules	MAE ($\lvert\Delta pK_a\rvert$)	maximum absolute error	number of explicit water molecules
nitrogen-containing aromatic ring	22	0.78	1.17	1[a]
alcohols	15	0.40	0.99	3
thiols	14	0.50	1.06	3
primary amines	10	0.55 (0.33)	2.05 (0.77)	0 (2)
secondary amines	8	0.29 (0.30)	0.82 (0.66)	0 (1)
tertiary amines	5	0.25 (0.60)	0.74(1.00)	0 (1)
carboxylic acid	26	0.45	1.47	3
phenols	39	0.46	1.41	3
thiophenols	13	0.62	1.69	2
anilines	34	0.34	1.65	3
carbon acids	10	0.52	1.72	3
benzoic acids	28	0.35	1.37	3
total	224	0.45		

[a]One explicit water per nitrogen or oxygen atom present in the ring. ΔpK_a values in parentheses correspond to the number of explicit water molecules given in parentheses.

Among the various functional groups considered, aliphatic amines are the only functional groups for which consistently good performance is obtained with implicit solvation models (MAE = 0.25−0.55 pK_a units). Thus, there is a very little room for improvement upon inclusion of explicit waters, and we discuss them only briefly. For primary amines, best results were obtained with two explicit water molecules (MAE = 0.33). For secondary amines, the inclusion of one explicit water molecule gave similarly accurate results (MAE = 0.30 pK_a units), while the results got slightly worse for tertiary amines (MAE = 0.60), possibly due to the steric repulsion from amine nitrogen substituents. As seen below, substantially improved results are obtained for all other functional groups with a small number of explicit solvent molecules.

Compared to the aliphatic amines, the errors in the calculated pK_a's are substantially larger (2−10 pK_a units) for all other functional groups using implicit solvation calculations. These errors can vary depending on the type of implicit solvation used. Inclusion of explicit water molecules was found to be necessary to evaluate the pK_a's accurately. By including one to three explicit water molecules (Figure 1) depending on the functional groups, an impressive accuracy (overall mean absolute error of 0.45 pK_a unit and the maximum mean absolute error of 0.78 pK_a unit) has been achieved for all the functional groups considered in this study (Table 1 and Tables S1−S11). Compared to the SMD-only calculated pK_a's, inclusion of explicit waters most dramatically improved the error in the calculated pK_a values of aliphatic alcohols (MAE lowered from 9.41 to 0.40), aliphatic thiols (MAE improved from 8.42 to 0.50), phenols (MAE lowered from 4.90 to 0.46), and thiophenols (MAE lowered from 5.57 to 0.62). Similarly, even for the more challenging case of carbon acids (α-CH of aldehydes, ketones, and amides), a remarkable improvement in the calculated pK_a is seen with the inclusion of three explicit water molecules placed around the carbonyl oxygen that stabilizes the conjugate base (i.e., enolate anion).

To further assess if the performance of our pK_a calculation protocol is systematic, we plotted the pK_a's calculated using CBS-QB3 and SMD, with and without the explicit solvent molecules, as a function of experimentally measured pK_a values (Figure 2). When the overall results for all functional groups calculated with SMD/CBS-QB3 and no explicit waters are linearly correlated with the experiment, a moderate correlation (R^2 = 0.85) with a significant slope (1.669) and a large intercept (−2.458) is obtained (Figure 3a). However, a better correlation can be obtained for some functional groups if only the pK_a's of those specific functional groups are correlated with the experimental values (Figures S3 and S4). On the other hand, the results obtained with the CBS-QB3 and SMD solvation with the optimum number of explicit waters listed in Table 1 are rather impressive. An excellent correlation (R^2 = 0.98) is found between the calculated and available experimental pK_a values. Furthermore, both the slope (1.005) and intercept (0.00) of the linear fit line improve quite dramatically, close to their ideal values.

Overall, the pK_a results calculated for various functional groups using the CBS-QB3 and SMD implicit solvation with zero to three explicit water molecules are excellent. We note that our study is restricted to the study of functional groups with a measured range of pK_a values from −1 to 20. At sufficiently lower (<−1) and higher (>20) pH ranges where

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

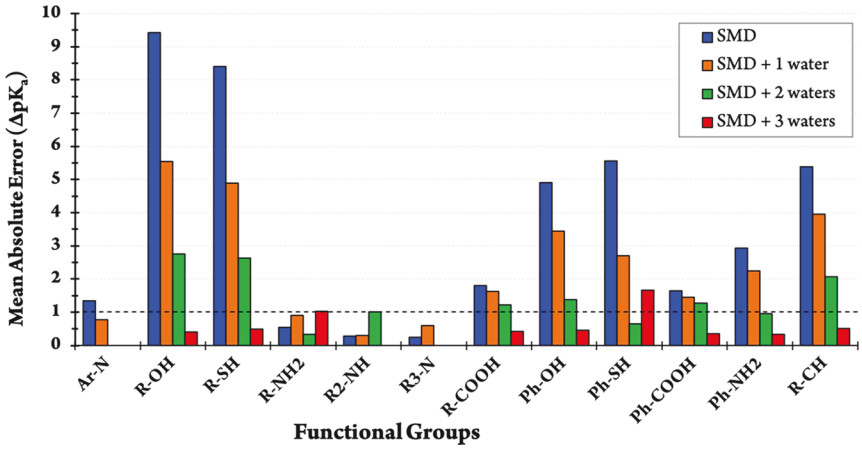


Figure 2. Effect of including explicit water molecules on the calculated pK_a values. "Ar-N", "R-", and "Ph-" in the *x*-axis labels represent the nitrogen-containing aromatic ring, aliphatic group, and phenyl group attached to the functional group, respectively. The dotted black line represents a target accuracy of 1 pK_a unit.

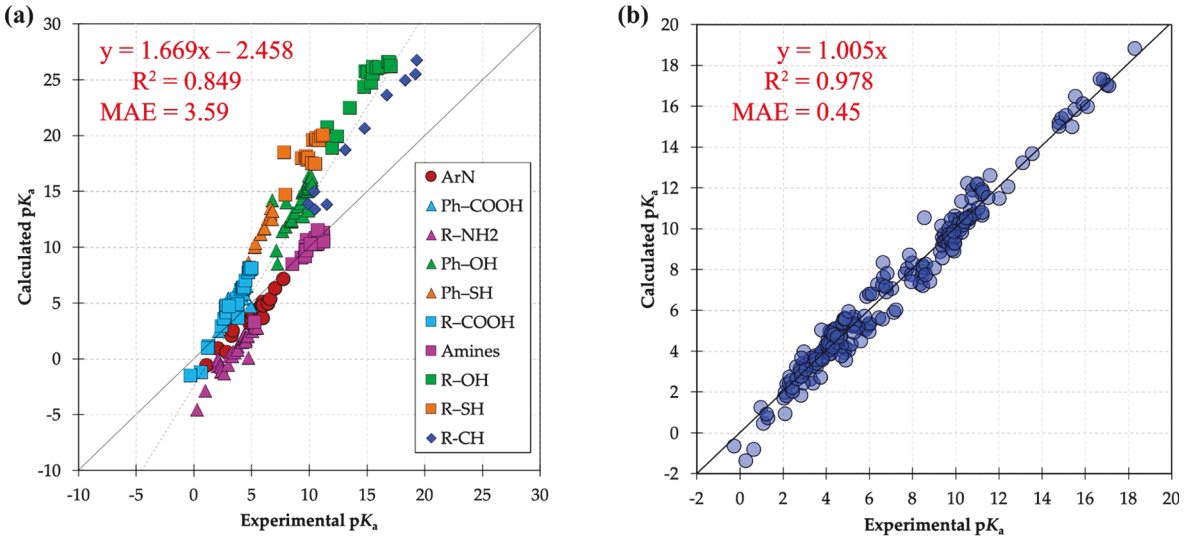


Figure 3. Correlation between the experimental pK_a values and pK_a values calculated with (a) CBS-QB3 and SMD implicit solvation and (b) CBS-QB3 and SMD implicit solvation with the optimum number of explicit waters listed in Table 1. For primary, secondary, and tertiary amines, two, one, and zero explicit waters were used, respectively.

water itself may be in a charged state (either protonated or deprotonated), correcting implicit solvation with neutral water might not be an appropriate model, and further calibrations of computational models may be needed. For the functional groups considered in this study, no other adjustment of the solvation model is required to accurately calculate the pK_a's. To the best of our knowledge, this is the first single QM method-based computational protocol that works well for a variety of functional groups present in many of the common bio-organic molecules.

3.2. CBH Scheme for pK_a Calculation. In the pK_a calculation protocol discussed above, the free energies are computed at the CBS-QB3 level. This method can be readily applied to small molecules containing up to 10−20 heavy atoms within a reasonable amount of CPU time. However, it becomes prohibitively expensive for larger molecules and hence limits the direct application of the abovementioned approach. For such systems, the use of cheaper methods such as DFT or more applicable WFT methods (e.g., MP2) is commonly followed. As shown in the previous studies, the accuracy of these methods is very system-dependent. Therefore, no single QM method can be used to calculate the pK_a values for different functional groups. In contrast, our CBH-based error cancellation scheme can reduce the computational cost while maintaining a similar high accuracy for all functional groups. To demonstrate this definitively, we computed the pK_a of 72 small-to-medium-sized molecules (**Set-II**) using the CBH scheme. These selected molecules have the experimentally measured pK_a's for the full molecule as well as for the elementary reactions involved in the CBH rungs. A representative net chemical reaction involved in CBH-1 and CBH-2 scheme is given in Scheme 1 (also in Scheme S1).

Figure 4 and Table S12 show the mean absolute error in pK_a's calculated using various rungs of the CBH-*n* (*n* = 0, 1, and 2) scheme. The pK_a values calculated using most DFT functionals without any corrections are found to deviate by more than 3 pK_a units from the experimental values. While some of the DFT methods performed relatively well for a subset of molecules, none of them worked satisfactorily for all of the considered functional groups. The errors in the

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

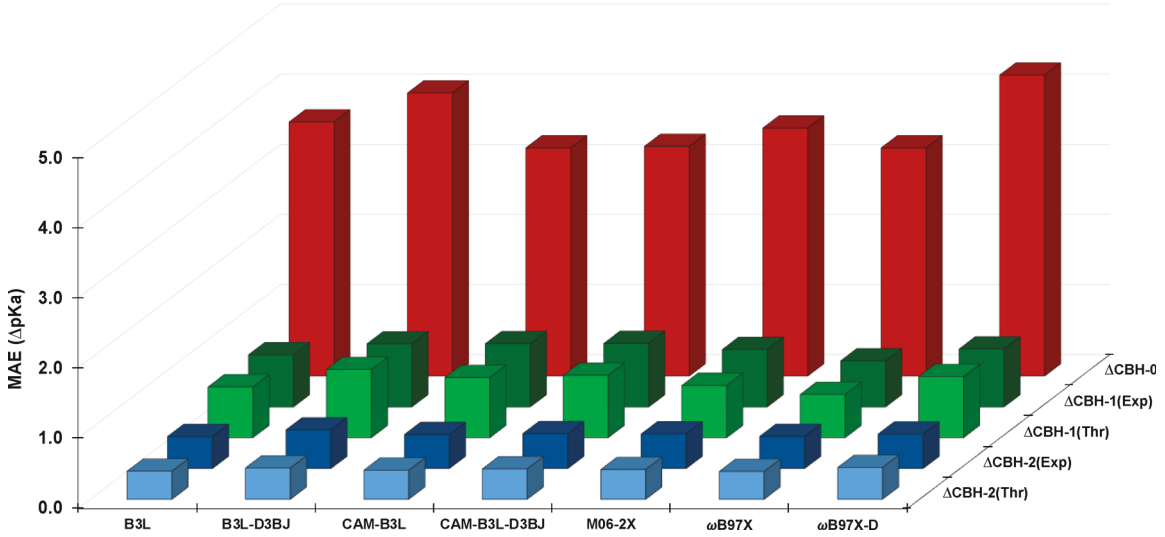


Figure 4. Calculated mean absolute error in pK_a's (ΔpK_a-0) and CBH-corrected pK_a errors (ΔpK_a-1 and ΔpK_a-2) at various DFT methods for the molecules in **Set-II**. ΔCBH-*n*(Exp) and ΔCBH-*n*(Thr) along the *z*-axis respectively denote experimental and theoretically calculated pK_a values used as the reference for the elementary CBH fragments.

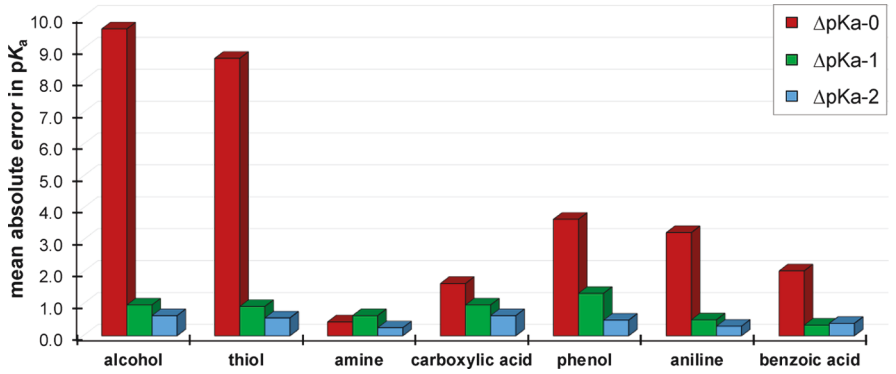


Figure 5. Mean absolute error in pK_a's (ΔpK_a-0) and CBH-corrected pK_a errors (ΔpK_a-1 and ΔpK_a-2, respectively) for various functional groups at the B3LYP/6-311++G(d,p) level of theory for the molecules in **Set-II**. Theoretically calculated pK_a values were used as the reference for the elementary CBH fragments.

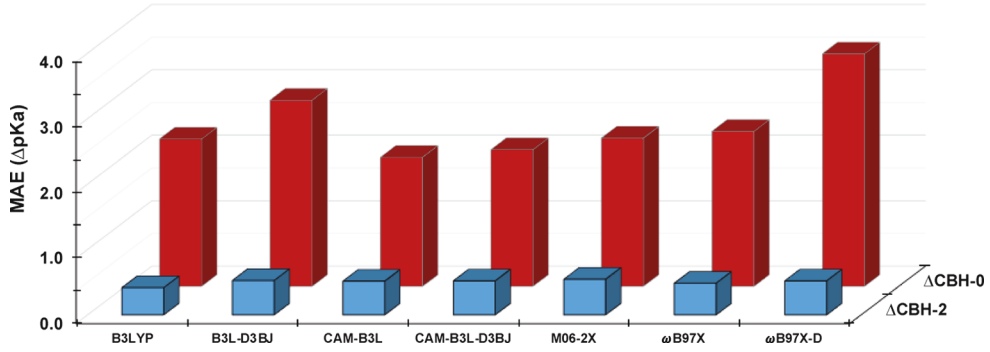


Figure 6. Calculated mean absolute error in pK_a (ΔpK_a-0) and ΔCBH-2-corrected pK_a errors (ΔpK_a-2) for various DFT methods for the molecules in **Set-III**.

calculated pK_a (ΔpK_a-0) drop significantly when the ΔCBH-1 scheme is used (ΔpK_a-1 = 0.8−1.3). While the results with ΔCBH-1 are significantly better than DFT methods alone without any correction, some of the errors are still considerably large (>2 pK_a units) due to the inconsistent treatment of local bonding environments by the CBH-1 scheme (Table S12). This shows that the pK_a correction via the isodesmic reaction scheme (i.e., CBH-1) alone is not sufficient to achieve the target accuracy of <1 pK_a unit. The errors drop dramatically when the ΔCBH-2 scheme is used. The mean unsigned error is in the range of 0.40−0.45 pK_a units for all the DFT functionals used. It is impressive to see that the results are systematically improved for all of the deprotonating functional groups. The best results were obtained for B3LYP with ΔpK_a-2 = 0.40 pK_a units (Figures 4 and 5). Interestingly, the errors in the calculated pK_a using the explicit−implicit approach (ΔCBH-

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

n(Thr)) for the elementary reactions are similar or better than using the experimental pK_a values (ΔCBH-n(Exp)), partly due to a better cancelation of errors in the theoretically calculated values. The overall results show that an impressive accuracy in the calculated pK_a values can be obtained by using ΔCBH-2 scheme without using any experiment-based corrections.

To further assess the applicability of our CBH-based pK_a calculation approach in larger and chemically interesting molecules, we computed the pK_a of 28 drug molecules randomly selected from a drug database (Scheme 2). This test set covers a pK_a range of 2−12 pK_a units, typical of the pK_a values of common bio-organic systems. For this data set, only the ΔCBH-2 scheme was used since it gave the best results (MAE = 0.40−0.45 pK_a units) for the DFT methods in **Set-II**. The calculated pK_a values and the errors for **Set-III** molecules are shown in Figure 6 and Table S4. For the molecules with multiple ionizable groups, separate CBH-2 schemes were constructed to calculate the corresponding microscopic pK_a values (Figure S5−S11). With the ΔCBH-2 scheme, calculated errors are significantly lower compared to the uncorrected pK_a's (ΔpK_a-0 = 1.8−3.5 units versus ΔpK_a-2 = 0.44−0.54 units). Interestingly, the functional group R−SO_2−NH_2 present in two of the molecules in **Set-III** (molecules **III.10** and **III.14**) was not included in any of the previous two datasets yet yielded accurate results. This demonstrates the applicability of our pK_a calculation approach in molecules containing functional groups that were not explicitly calibrated in this study.

The pK_a calculation protocol based on CBH scheme also works if a molecule possesses multiple ionizable groups or tautomeric forms (e.g., **III.2**, **III.3**, and **III.9**). For such molecules, separate CBH schemes can be constructed for each of the deprotonating functional groups to derive the separate elementary deprotonation reactions. In such cases, the explicit water molecules are placed only near the deprotonating functional group under consideration. In this way, each of the functional groups can be microsolvated locally and separately without having to include explicit solvent molecules around all of the functional groups at once. This also avoids the complication that may arise while placing explicit water molecules around all of the functional groups at the same time. For example, for the molecule **III.3**, which has two phenolic OH groups and one aliphatic amine group, three separate CBH reactions are constructed for each of the functional groups, and the corresponding pK_a values are then obtained (see Figures S5−S11).

4. CONCLUSIONS

In summary, this study presents a standard protocol for accurately calculating pK_a's of a wide range of bio-organic molecules in the aqueous medium. By including a few (0−3) explicit solvent molecules and a SMD implicit solvation model, an impressive accuracy of MAE = 0.45 pK_a units was achieved for small bio-organic molecules with experimental pK_a values in the range of −1 to 20 at the CBS-QB3 level. For the functional groups benchmarked in this study, our explicit−implicit solvation protocol with the recommended number of explicit water molecules and SMD/CBS-QB3 method can be directly used without further benchmarking. For the larger molecules, where CBS-QB3-based approach is computationally unaffordable, we present an efficient pK_a calculation protocol based on the CBH error-cancelation scheme. Using the CBH-2 scheme, we showed that our protocol eliminates the systematic errors in different DFT methods to yield accurate pK_a values (MAE of 0.40−0.54 pK_a units) for simple to complex molecular systems. Our results showed that, by treating the elementary deprotonation reactions at the CBS-QB3 level with explicit−implicit solvation, the calculated pK_a's are nearly independent of the DFT method used. It is possible that, for some bio-organic molecules, CBH-2 scheme may not be sufficient to capture the local bonding environment in the elementary reactions. In such cases, the accuracy can be further improved by using higher rungs of the CBH schemes (e.g., ΔCBH-3). Nonetheless, we note that the current protocol covers most of the common functional groups present in organic and biomolecular systems. Our unique pK_a calculation protocol should find widespread applications in different areas of research such as exploring reaction mechanisms, synthetic chemistry, and structure-based drug design.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00606.

Discussion of frequency scaling, CBH scheme for molecules with aromatic ring, arrangement of hydrogen-bonded explicit water molecules, molecules used in **Set-I** and **Set-II**, experimental and calculated pK_a values, and pK_a calculation of molecules with multiple functional groups/tautomeric states (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: kraghava@indiana.edu.

ORCID

Bishnu Thapa: 0000-0003-3521-1062

Krishnan Raghavachari: 0000-0003-3275-1426

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support from the National Science Foundation grant CHE-1665427 at Indiana University. B.T. was supported by a Lilly Research Award Program. The Big Red II supercomputing facility at Indiana University was used for most of the calculations in this study.

■ REFERENCES

(1) Manallack, D. T. The pKa Distribution of drugs: Application to drug discovery. *Perspect. Med. Chem.* **2007**, *1*, 25−38.

(2) Charifson, P. S.; Walters, W. P. Acidic and basic drugs in medicinal chemistry: A perspective. *J. Med. Chem.* **2014**, *57*, 9701−9717.

(3) Bell, R. P. *The proton in chemistry*; Springer Science & Business Media: 2013.

(4) Stewart, R. *The proton: Applications to organic chemistry*; Academic Press: New York, 2012.

(5) Huynh, M. H. V.; Meyer, T. J. Proton-coupled electron transfer. *Chem. Rev.* **2007**, *107*, 5004−5064.

(6) Ho, J.; Coote, M. L. A universal approach for continuum solvent pK$_a$ calculations: Are we there yet? *Theor. Chem. Acc.* **2010**, *125*, 3.

(7) Pauling, L.; Wilson, E. B. *Introduction to quantum mechanics with applications to chemistry*; Courier Corporation: 2012.

(8) Alongi, K. S.; Shields, G. C. Theoretical calculations of acid dissociation constants: A review article. *Annu. Rep. Comput. Chem.* **2010**, *6*, 113−138.

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

(9) Ho, J.; Coote, M. L. First-principles prediction of acidities in the gas and solution phase. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 649−660.
(10) Shields, G. C.; Seybold, P. G. *Computational approaches for the prediction of pKa values*; CRC Press: Boca Raton, London, New York, 2013.
(11) Ho, J. Predicting pK_a in implicit solvents: Current status and future directions. *Aust. J. Chem.* **2014**, *67*, 1441−1460.
(12) Casasnovas, R.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. Theoretical pK_a calculations with continuum model solvents, alternative protocols to thermodynamic cycles. *Int. J. Quantum Chem.* **2014**, *114*, 1350−1363.
(13) Seybold, P. G.; Shields, G. C. Computational estimation of pK_a values. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015**, *5*, 290−297.
(14) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.
(15) Guerard, J. J.; Arey, J. S. Critical evaluation of implicit solvent models for predicting aqueous oxidation potentials of neutral organic compounds. *J. Chem. Theory Comput.* **2013**, *9*, 5046−5058.
(16) Ho, J.; Ertem, M. Z. Calculating free energy changes in continuum solvation models. *J. Phys. Chem. B* **2016**, *120*, 1319−1329.
(17) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 699−709.
(18) Haworth, N. L.; Wang, Q.; Coote, M. L. Modeling flexible molecules in solution: A pK_a case study. *J. Phys. Chem. A* **2017**, *121*, 5217−5225.
(19) Mobley, D. L.; Wymer, K. L.; Lim, N. M.; Guthrie, J. P. Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 135−150.
(20) Thapa, B.; Schlegel, H. B. Density functional theory calculation of pK_a's of thiols in aqueous solution using explicit water molecules and the polarizable continuum model. *J. Phys. Chem. A* **2016**, *120*, 5726−5735.
(21) Klicić, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. *J. Phys. Chem. A* **2002**, *106*, 1327−1335.
(22) Pickard, F. C., IV; König, G.; Tofoleanu, F.; Lee, J.; Simmonett, A. C.; Shao, Y.; Ponder, J. W.; Brooks, B. R. Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pK_a corrections. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1087−1100.
(23) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. Accurate relative pK_a calculations for carboxylic acids using complete basis set and Gaussian-n models combined with continuum solvation methods. *J. Chem. Phys.* **2001**, *114*, 4595−4606.
(24) Liptak, M. D.; Shields, G. C. Accurate pK_a calculations for carboxylic acids using complete basis set and Gaussian-n models combined with cpcm continuum solvation methods. *J. Am. Chem. Soc.* **2001**, *123*, 7314−7319.
(25) Zhang, S. A reliable and efficient first principles-based method for predicting pK_a values. III. Adding explicit water molecules: Can the theoretical slope be reproduced and pK_a values predicted more accurately? *J. Comput. Chem.* **2012**, *33*, 517−526.
(26) Zhang, S.; Baker, J.; Pulay, P. A reliable and efficient first principles-based method for predicting pK_a values. 2. Organic acids. *J. Phys. Chem. A* **2010**, *114*, 432−442.
(27) Eckert, F.; Klamt, A. Accurate prediction of basicity in aqueous solution with COSMO-RS. *J. Comput. Chem.* **2006**, *27*, 11−19.
(28) Muckerman, J. T.; Skone, J. H.; Ning, M.; Wasada-Tsutsui, Y. Toward the accurate calculation of pK_a values in water and acetonitrile. *Biochim. Biophys. Acta* **2013**, *1827*, 882−891.
(29) Namazian, M.; Halvani, S. Calculations of pKa values of carboxylic acids in aqueous solution using density functional theory. *J. Chem. Thermodyn.* **2006**, *38*, 1495−1502.
(30) Namazian, M.; Halvani, S.; Noorbala, M. R. Density functional theory response to the calculations of pK_a values of some carboxylic acids in aqueous solution. *J. Mol. Struct.: THEOCHEM* **2004**, *711*, 13−18.
(31) Zeng, Y.; Chen, X.; Zhao, D.; Li, H.; Zhang, Y.; Xiao, X. Estimation of pK_a values for carboxylic acids, alcohols, phenols and amines using changes in the relative Gibbs free energy. *Fluid Phase Equilib.* **2012**, *313*, 148−155.
(32) Gangarapu, S.; Marcelis, A. T. M.; Zuilhof, H. Accurate pK_a calculation of the conjugate acids of alkanolamines, alkaloids and nucleotide bases by quantum chemical methods. *ChemPhysChem* **2013**, *14*, 990−995.
(33) Lian, P.; Johnston, R. C.; Parks, J. M.; Smith, J. C. Quantum chemical calculation of pK_as of environmentally relevant functional groups: Carboxylic acids, amines, and thiols in aqueous solution. *J. Phys. Chem. A* **2018**, *122*, 4366−4374.
(34) Psciuk, B. T.; Lord, R. L.; Munk, B. H.; Schlegel, H. B. Theoretical determination of one-electron oxidation potentials for nucleic acid bases. *J. Chem. Theory Comput.* **2012**, *8*, 5107−5123.
(35) Provorse Long, M. R.; Isborn, C. M. Combining explicit quantum solvent with a polarizable continuum model. *J. Phys. Chem. B* **2017**, *121*, 10105−10117.
(36) Lee, T. B.; McKee, M. L. Dependence of pK_a on solute cavity for diprotic and triprotic acids. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10258−10269.
(37) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants. *J. Phys. Chem. A* **2006**, *110*, 2493−2499.
(38) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A., III Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *J. Phys. Chem. B* **2008**, *112*, 9709−9719.
(39) da Silva, E. F.; Svendsen, H. F.; Merz, K. M. Explicitly representing the solvation shell in continuum solvent calculations. *J. Phys. Chem. A* **2009**, *113*, 6404−6409.
(40) Thapa, B.; Schlegel, H. B. Theoretical calculation of pK_a's of selenols in aqueous solution using an implicit solvation model and explicit water molecules. *J. Phys. Chem. A* **2016**, *120*, 8916−8922.
(41) Thapa, B.; Schlegel, H. B. Improved pK_a prediction of substituted alcohols, phenols, and hydroperoxides in aqueous medium using density functional theory and a cluster-continuum solvation model. *J. Phys. Chem. A* **2017**, *121*, 4698−4706.
(42) Thapa, B.; Schlegel, H. B. Calculations of pK_a's and redox potentials of nucleobases with explicit waters and polarizable continuum solvation. *J. Phys. Chem. A* **2014**, *119*, 5134−5144.
(43) Pliego, J. R.; Riveros, J. M. The cluster− continuum model for the calculation of the solvation free energy of ionic species. *J. Phys. Chem. A* **2001**, *105*, 7241−7247.
(44) Pliego, J. R.; Riveros, J. M. Theoretical calculation of pK_a using the cluster− continuum model. *J. Phys. Chem. A* **2002**, *106*, 7434−7439.
(45) Bochevarov, A. D.; Watson, M. A.; Greenwood, J. R.; Philipp, D. M. Multiconformation, density functional theory-based pK_a prediction in application to large, flexible organic molecules with diverse functional groups. *J. Chem. Theory Comput.* **2016**, *12*, 6001−6019.
(46) Ramabhadran, R. O.; Raghavachari, K. Connectivity-based hierarchy for theoretical thermochemistry: Assessment using wave function-based methods. *J. Phys. Chem. A* **2012**, *116*, 7531−7537.
(47) Ramabhadran, R. O.; Raghavachari, K. Theoretical thermochemistry for organic molecules: Development of the generalized connectivity-based hierarchy. *J. Chem. Theory Comput.* **2011**, *7*, 2094−2103.
(48) Pliego, J. R., Jr.; Riveros, J. M. Hybrid discrete-continuum solvation methods. *Wiley Interdiscip. Rev.: Comput. Mol. Sci* **2019**, No. e1440.
(49) Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. Molecular orbital theory of the electronic structure of organic compounds V. Molecular theory of bond separation. *J. Am. Chem. Soc.* **1970**, *92*, 4796−4801.
(50) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035

model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.
(51) Ho, J. Are thermodynamic cycles necessary for continuum solvent calculation of pK_as and reduction potentials? *Phys. Chem. Chem. Phys.* **2015**, *17*, 2859−2868.
(52) Camaioni, D. M.; Schwerdtfeger, C. A. Comment on "accurate experimental values for the free energies of hydration of h^+, oh_-, and h_3o^+". *J. Phys. Chem. A* **2005**, *109*, 10795−10797.
(53) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Aqueous solvation free energies of ions and ion− water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. *J. Phys. Chem. B* **2006**, *110*, 16066−16081.
(54) Isse, A. A.; Gennaro, A. Absolute potential of the standard hydrogen electrode and the problem of interconversion of potentials in different solvents. *J. Phys. Chem. B* **2010**, *114*, 7894−7899.
(55) Marenich, A. V.; Ho, J.; Coote, M. L.; Cramer, C. J.; Truhlar, D. G. Computational electrochemistry: Prediction of liquid-phase reduction potentials. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15068−15106.
(56) Becke, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **1992**, *96*, 2155−2160.
(57) Becke, A. D. Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals. *J. Chem. Phys.* **1997**, *107*, 8554−8560.
(58) Lee, C.; Yang, W.; Parr, R. G. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785.
(59) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom−atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(60) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(61) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456−1465.
(62) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+g basis set for first-row elements, Li-F. *J. Comput. Chem.* **1983**, *4*, 294−301.
(63) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(64) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(65) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret. Chim. Acta* **1973**, *28*, 213−222.
(66) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(67) Montgomery, J. A., Jr; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A complete basis set model chemistry Vii. Use of the minimum population localization method. *J. Chem. Phys.* **2000**, *112*, 6532−6542.
(68) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation. *J. Phys. Chem. B* **2011**, *115*, 14556−14562.
(69) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of mmff94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(70) Dennington, R.; Keith, T.; Millam, J.; Eppinnett, K.; Hovell, W.; Gilliland, R. *Gaussview. Version 6*; Semichem. Inc.: Shawnee Mission, KS, USA, 2016.
(71) Frisch, M. J. T.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16, revision-a03;* Gaussian, Inc.: Wallingford, CT, 2016.

DOI: 10.1021/acs.jctc.9b00606
J. Chem. Theory Comput. 2019, 15, 6025−6035