

CHAPTER 8

Theoretical Calculations of Acid Dissociation Constants: A Review Article

Kristin S. Alongi¹ and George C. Shields²

| | | |
|-----------------|--|-----|
| Contents | | |
| | 1. Introduction | 114 |
| | 2. Background | 115 |
| | 2.1 Thermodynamic cycles | 115 |
| | 2.2 Gas-phase free energy calculations | 121 |
| | 2.3 Solvation free energy calculations | 121 |
| | 3. Calculating Changes in Free Energy in the Gas Phase | 123 |
| | 4. Calculating Changes in Free Energy in Solution | 125 |
| | 5. Thermodynamic Cycles | 130 |
| | 6. Concluding Remarks | 133 |
| | References | 134 |

Abstract

Acid dissociation constants, or pKa values, are essential for understanding many fundamental reactions in chemistry. These values reveal the deprotonation state of a molecule in a particular solvent. There is great interest in using theoretical methods to calculate the pKa values for many different types of molecules. These include molecules that have not been synthesized, those for which experimental pKa determinations are difficult, and for larger molecules where the local environment changes the usual pKa values, such as for certain amino acids that are part of a larger polypeptide chain. Chemical accuracy in pKa calculations is difficult to achieve, because an error of 1.36 kcal/mol in the change of free energy for deprotonation in solvent results in an error of 1 pKa unit. In this review the most valuable methods for determining accurate pKa values in aqueous solution are presented for educators interested in explaining or using these methods for their students.

¹ Dean's Office and Department of Chemistry & Physics, College of Science & Technology, Armstrong Atlantic State University, Savannah, GA, USA

² Dean's Office & Department of Chemistry, College of Arts & Sciences, Bucknell University, Lewisburg, PA, USA

Keywords: pK_a; acid dissociation constants; theory; free energy; gas phase; solution; thermodynamic cycles; deprotonation

1. INTRODUCTION

Acid dissociation constants, also known as pK_a values, are essential for understanding many fundamental reactions in chemistry and biochemistry. For the acid dissociation reaction



pK_a is defined as

$$\text{pK}_a = -\log K_a \quad (2)$$

$$K_a = \frac{[\text{A}^{-}_{(\text{aq})}][\text{H}^{+}_{(\text{aq})}]}{[\text{HA}_{(\text{aq})}]} \quad (3)$$

Oftentimes pK_a values can be measured quite easily experimentally; however, many times chemists are interested in the pK_a values of molecules that have not been synthesized or for which experiments are not straightforward. For instance, amino acids that are part of a polypeptide chain have pK_a values that vary based on their local environment, which are difficult to determine. Therefore, the ability to computationally calculate these pK_a values accurately is important for scientific advancements in biochemistry and other fields. Chemical accuracy, though, is hard to achieve. Computationally calculating acid dissociation constants is a demanding and arduous process because an error of 1.36 kcal/mol in the change of free energy of [reaction 1](#) results in an error of 1 pK_a unit [1,2]. There are numerous studies that use a variety of methods in an attempt to obtain this chemical accuracy. In recent years, there have been new developments, but many discrepancies still exist. The aim of this present work is to compare all of the significant methods for accurate pK_a calculations for educators interested in explaining or using these methods with their students. We will focus on three areas of development: thermodynamic cycles, gas-phase free energy calculations, and the change in free energy of solvation calculations. An excellent review of

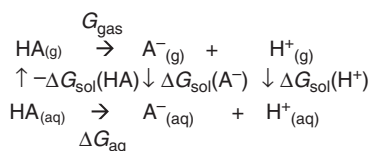


Figure 1 Proton-based thermodynamic cycle.

thermodynamic cycles and the most common solvation models used for pK_a calculations has recently been published by Ho and Coote [3].

2. BACKGROUND

2.1 Thermodynamic cycles

Numerous thermodynamic cycles have been used to calculate pK_a values. One of the most common methods is depicted in Figure 1, based on reaction 1 [1,2,4]:

In Figure 1, ΔG_{aq} represents the overall change in free energy of this reaction in solution, ΔG_{gas} is the change in the gas-phase free energy, and ΔG_{sol} is the change in free energy of solvation. Based on the diagram, pK_a is calculated using the following equations:

$$pK_a = \frac{\Delta G_{aq}}{RT \ln(10)} \quad (4)$$

$$\Delta G_{aq} = \Delta G_{gas} + \Delta \Delta G_{sol} \quad (5)$$

where

$$\Delta G_{gas} = G_{gas}(H^+) + G_{gas}(A^-) - G_{gas}(HA) \quad (6)$$

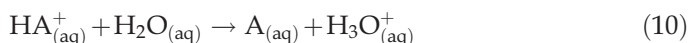
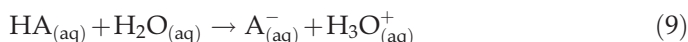
and

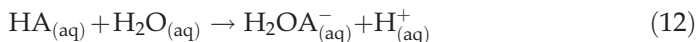
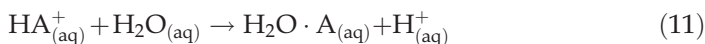
$$\Delta \Delta G_{sol} = \Delta G_{sol}(H^+) + \Delta G_{sol}(A^-) - \Delta G_{sol}(HA) \quad (7)$$

All of these free energy values can be calculated using quantum chemistry, except $\Delta G_{sol}(H^+)$ and $G_{gas}(H^+)$, which must be determined experimentally or using thermodynamic theory. These values will be discussed in the corresponding sections. A similar thermodynamic cycle that is often used is based on the acid dissociation equation of a protonated acid [5]:



In this case, reaction 8 leads to a thermodynamic cycle similar to Figure 1, which was based on reaction 1. Now, Eq. (6) becomes $\Delta G_{gas} = G_{gas}(H^+) + G_{gas}(A) - G_{gas}(HA^+)$ and Eq. (7) becomes $\Delta \Delta G_{sol} = \Delta G_{sol}(H^+) + \Delta G_{sol}(A) - \Delta G_{sol}(HA^+)$. Other thermodynamic cycles are based on the acid's reaction with a water molecule, as depicted in the following reactions [3,5–9]:

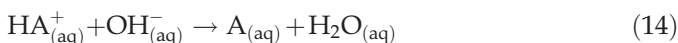
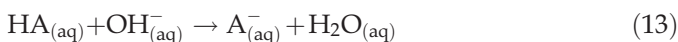




Reactions 9–12 lead to equivalent thermodynamic cycles similar to Figure 1. In these cycles, however, the number of waters included in the cycle may vary [8]. A key point is that the concentrations of all species in a given thermodynamic cycle must have the same standard state, including the solute and solvent, and when water is introduced in the cycle it should have a standard state of 1M [6,9]. One limitation of the cycles derived from reactions 9–12 is that the hydronium ion's change in free energy of solvation is difficult to calculate because of the high charge. However, one can use the accepted experimental $\Delta G_{\text{sol}} \text{H}_3\text{O}^+$ value of -110.3 kcal/mol in the 1M standard state and the accepted experimental value for $\Delta G_{\text{sol}} \text{H}_2\text{O}$ of -6.32 kcal/mol [10–12]. This reduces the number of computations and ensures similar accuracy as using the proton-based thermodynamic cycle and the most recently accepted experimental value for $\Delta G_{\text{sol}} \text{H}^+$.

The addition of another water to the aqueous reaction, such as in reaction 11 or 12, also results in more calculations. This increases computational error in determining $\text{p}K_{\text{a}}$, as the total number of manipulated numbers increases from reaction 1 to reaction 12. For a thorough understanding of the standard state issues that arise when adding water to thermodynamic cycles, study the recent Goddard group paper [9].

Another thermodynamic cycle often used is derived from the acid dissociation reaction with a hydroxide ion to produce water [5,7]:



Limitations of these corresponding cycles are similar to those used for reactions 9–12, where the additional reactant increases the number of calculations and, consequently, the computational error. The change in free energy of solvation for OH^- is also difficult to determine computationally because it is a diffuse anion, but the experimental value may be used instead. The accepted experimental $\Delta G_{\text{sol}}(\text{OH}^-)$ value is -104.7 kcal/mol in the 1M standard state [12,13]. Again, as long as the correct experimental values are used for the free energy of solvation of H^+ , H_3O^+ , H_2O , and OH^- , these different thermodynamic cycles should yield similar results. Increasing the number of species primarily affects the calculation of ΔG_{gas} , as Eq. (6) now contains another reactant and product. Reactions 13 and 14 can be modified by hydrating the hydroxide ion with $n+m$ waters, resulting in n waters being attached to $\text{A}_{(\text{aq})}$ or $\text{A}_{(\text{aq})}^-$ and m waters returning to the solvent $\text{H}_2\text{O}_{(\text{aq})}$ after reaction, in the cluster continuum approach first outlined by Pliego and Riveros [3,9,14]. The

difficulty in computing accurate free energy of solvation values for anions is discussed later in Section 2.3.

The thermodynamic cycles derived from reactions 9–14 are depicted in a manner similar to Figure 1. The pK_a values are then calculated using Eqs. (4) and (5); Eqs. (6) and (7) are modified to correlate to the specific thermodynamic cycle.

2.1.1 Limiting experimental values

One of the main sources of error in pK_a calculations is the value used for the free energy of solvation for H^+ , which is explicitly needed in certain thermodynamic cycles.

A proton contains no electrons and its free energy cannot be calculated quantum mechanically. Calculation of this energy using the standard equations of thermodynamics and the Sackur–Tetrode equation [15] yields the same value as can be deduced experimentally from the NIST database. The translational energy of $1.5RT$ combined with $PV = RT$ and $H = E + PV$ yields a value of H° (H^+) equal to $5/2(RT)$ or 1.48 kcal/mol. Use of the Sackur–Tetrode equation yields the entropy, $TS(H^+) = 7.76$ kcal/mol at 298 K and 1 atm pressure. Finally, since $G = H - TS$, $G^\circ(H^+) = -6.28$ kcal/mol.

In contrast, over the past 15 years, the $\Delta G_{\text{sol}} H^+$ value has changed considerably, from an accepted range of -254 to -261 kcal/mol in 1991 [7] to the now accepted -265.9 kcal/mol [12]. The basic uncertainty in any determination of the free energy of solvation of an ion is that ions are never isolated. Determination of, say, ΔG_{sol} for an anion, A^- , can be made if the corresponding value for a cation, C^+ , is already known. That way when the ΔG_{sol} is measured for the salt CA , the value for ΔG_{sol} of A^- can be determined by difference. In practice, all ionic solvation values are referenced against the value for H^+ ; every time the value for $G_{\text{sol}} H^+$ changes, the values for all of the other ions in the various databases change as well. The older value for $G_{\text{sol}} H^+$ of -261.4 kcal/mol was estimated from an average of five independent measurements of the hydrogen ion electrode [12].

When we began our work on pK_a calculations, using carboxylic acids and phenols as test cases, we tried to use the old value of -261.4 kcal/mol. Like previous researchers, it was not possible to calculate accurate absolute pK_a values [16–19]. Karplus and coworkers reported that the continuum dielectric methods were not accurate enough to yield accurate absolute solvation free energies. By 2001, new solvation methods had been developed, and relative pK_a values were easily determined, leading to a suspicion that the value for $\Delta G_{\text{sol}} H^+$ was erroneous [20]. We used an experimental thermodynamic cycle for acetic acid dissociating to H^+ and the acetate anion to derive an experimental value for $\Delta G_{\text{sol}} H^+$ of -264.61 kcal/mol [1,2]. At that time, this was by far the most negative value used for these types of calculations, and it was a bit shocking to be using a value that was more than 3 kcal/mol lower than that obtained from the hydrogen ion electrode. However, two groups had used a combined explicit-implicit theoretical approach to obtain $\Delta G_{\text{sol}} H^+$ values of -264.4 and -264.3 kcal/mol for a standard state of 1M [21,22]. In addition, Coe and coworkers had used experimental ion–water clustering data to derive a

value for $\Delta G_{\text{sol}} \text{H}^+$ of -264.0 kcal/mol [10]. At the time, much of the computational chemistry community thought that this value was for a standard state of 1M [23–25], yet after a few years of confusion it was determined that the correct standard state was 1 atm, so that changing to a standard state of 1M changed the value of $\Delta G_{\text{sol}} \text{H}^+$ to -265.9 kcal/mol [11,12]. In addition, Goddard and coworkers have recently shown that by including concentration corrections to Zhan and Dixon's high-level ab initio calculations of the hydration free energy of the proton [22], their value is corrected to $-265.63 \pm 0.22 \text{ kcal/mol}$ [9]. We owe the Camaioni, Goddard, and Cramer/Truhlar research groups a debt of gratitude for bringing clarity to this issue. Discussion of the standard state conversion can be found below.

The newest value was originally determined by Tissandier et al. in 1998 using correlations between ΔG_{sol} of neutral ion pairs and experimental ion–water clustering data, which is known as the cluster pair approximation method. Kelly et al. [12] confirmed this value in 2006 using a similar method but larger data set. Experimental uncertainty, however, does still exist, which introduces uncertainty in all of these procedures. The estimated error in $\Delta G_{\text{sol}} \text{H}^+$ is thought to be 2 kcal/mol [12]. We note that the most accurate values of a wide range of unclustered cations and anions, based on the accepted value for the free energy of solvation of H^+ , are given in Reference [12]. For clustered ions, standard state corrections for concentration of water clustered to the ions must be included [9], and the corrected values for clustered ions have been published in the supplemental information of a recent paper on Minnesota solvent models [26].

The absolute solvation free energy of a proton can also be calculated using high-level gas-phase calculations with a supermolecule-continuum approach, involving a self-consistent reaction field model. The change in free energy of solvation is calculated by adding waters to H^+ until a converged value is reached. The solvent is approximated by a dielectric continuum medium that is based on its location around the solute, and the number of quantum mechanically treated solvent molecules is increased to improve the calculations. Using this approach, Zhan and Dixon calculated $\Delta G_{\text{sol}} \text{H}^+ = -264.3 \text{ kcal/mol}$ [22]. Correction for a missing standard state term for water concentration changes this value to $-265.63 \pm 0.22 \text{ kcal/mol}$ [9]. Thus the most recent experimental and theoretical determinations of $\Delta G_{\text{sol}} \text{H}^+$ are now less than 0.3 kcal/mol from each other, lending confidence in using either the -265.63 or -265.9 value, and lowering the uncertainty in this value to be closer to 1 kcal/mol . Zhan and Dixon's corrected value is the standard against which all future explicit quantum mechanical calculations of $\Delta G_{\text{sol}} \text{H}^+$ will be evaluated.

The standard state of $\Delta G_{\text{sol}} \text{H}^+$ must also be taken into account to produce reliable results. Free energies may be calculated using an ideal gas at 1 atm as a reference for gas-phase calculations or with an ideal gas at 1 mol/L, which is used with free energy of solvation calculations. The value for ΔG_{sol} and ΔG_{gas} depends on which standard state is used in their determination. Furthermore, a homogeneous equilibrium, where all of the species are in the same standard state, is necessary to obtain reliable pK results. The conversion of the 1 atm standard state to the 1 mol/L standard state can be derived from the relationship between the

equilibrium constant concerning concentration, K_c in 1M, and the equilibrium constant expressed in terms of pressure, K_p in the 1 atm standard state. The relationship between the two constants is derived for the following general reaction:



The corresponding equilibrium constants are

$$K_c = \frac{[B]^b}{[A]^a} \quad (16)$$

$$K_p = \frac{P_B^b}{P_A^a} \quad (17)$$

Using the ideal gas law, $PV = nRT$, where R is given by 0.8206 L atm/K mol, we rewrite Eq. (17):

$$K_p = \frac{(n_B RT/V)^b}{(n_A RT/V)^a} = \frac{(n_B/V)^b}{(n_A/V)^a} \cdots (RT)^{b-a} \quad (18)$$

Since n_B/V and n_A/V now have the units mol/L, they can be replaced by the concentrations of A and B and simplified in terms of K_c :

$$K_p = \left\{ \frac{[B]^b}{[A]^a} \right\} (RT)^{b-a} = K_c (RT)^{\Delta n} \quad (19)$$

where Δn is the change in the number of moles, $b - a$. Equation (19) can be used to show the relationship between the Gibbs free energies in different standard states. The relationship between the 1M state and the standard state of 1 atm is [12]:

$$G^* = G^\circ + \Delta G^{\circ \rightarrow *} \quad (20)$$

$$\Delta G^* = \Delta G^\circ - \Delta G^{\circ \rightarrow *} \quad (21)$$

We can determine $\Delta G^{\circ \rightarrow *}$ by using Eq. (19) and the relations between the equilibrium constants and the free energies at the two different standard states:

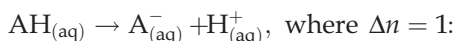
$$\Delta G^* = -RT \ln K_c \quad (22)$$

$$\Delta G^\circ = -RT \ln K_p \quad (23)$$

Using Eqs. (19), (22), and (23), we can find the relationship between the two free energies at 298.15 K:

$$\Delta G^* = \Delta G^\circ - RT \ln(RT)^{\Delta n} = \Delta G^\circ - RT \ln(24.4654)^{\Delta n} \quad (24)$$

In relation to Eq. (2), this illustrates that for the dissociation reaction 1,

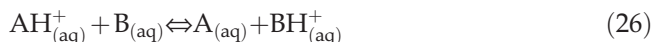


$$\Delta G^{\circ \rightarrow *} = RT \ln(24.4654) \quad (25)$$

At 298.15 K this conversion of standard states, $\Delta G^{\circ \rightarrow *}$, equals 1.89 kcal/mol. When $\Delta G_{\text{sol}} \text{H}^+ = -265.9$, it is in the gas-phase 1M standard state. If reported with the 1 atm standard it is -264.0 kcal/mol. In calculating accurate pK_a values, one must be aware of the standard state because a difference of 1.89 kcal/mol can cause significant error and unreliable values [1,2,12].

The $G_{\text{gas}} \text{H}^+$ value also cannot be determined quantum mechanically. Its value, however, has less uncertainty, and is the same whether determined from experimental values available in the NIST website or from the Sackur-Tetrode equation [15], and is consistently accepted as -6.28 kcal/mol for a standard state of 1 atm [1,2].

The reliance on experimental H^+ values is a significant challenge in producing accurate pK_a calculations using some thermodynamic cycles. Relative pK_a calculations, which solve for the pK_a of one acid in terms of another, are often used [5,20,27–32] to cancel out this error. The calculation is based on the two species in equilibria, where AH^+ and BH^+ are two acids:



This method is limited to reactions that have a particular acid standard available for Reference [20]. For this review, we will focus on absolute pK_a calculations and the most accurate method concerning any molecule of interest.

2.1.2 Alternative methods for calculating pK_a values

Methods other than thermodynamic cycles are often used to calculate acid dissociation constants. Previous publications implement the theoretical relationship between pK_a and structural property [6], bond valence methods and bond lengths [33], pK_a correlations with highest occupied molecular orbital (HOMO) energies and frontier molecular orbitals [34], and artificial neural networks [35] to predict pK_a values. In addition much work has been done using physical properties as quantitative structure-activity relationship (QSAR) descriptors, and regression equations with such descriptors to yield accurate pK_a values for specific classes of molecules [36–47]. The correlation of pK_a s to various molecular properties, however, is often restricted to specific classes of compounds, and it is

usually unwise to apply these relations to other molecules outside of the data set [8,23]. Therefore, in this review we will focus on the use of various thermodynamic cycles in the calculation of acid dissociation constants.

Many other variables such as electronic, thermal, and solvation energy and polarizable force fields may also affect the accuracy of pK_a calculations but will not be discussed in detail in this review [28,48].

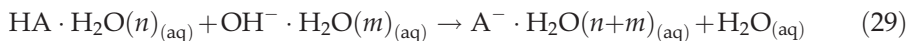
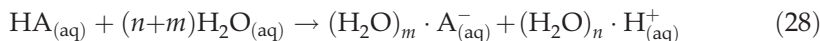
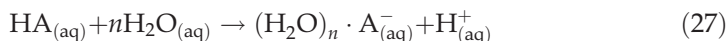
2.2 Gas-phase free energy calculations

The gas-phase free energy calculation is the lowest source of error in pK_a calculations. High levels of theory, such as CBS-QB3 [49] and CBS-APNO [50], produce reliable ΔG_{gas} values with root-mean-square deviations of 1.1–1.6 kcal/mol from the free energy of gas-phase deprotonation reactions compiled in the NIST online database [51–53]. With today's computers and focusing on small molecules, CCSD(T) calculations extrapolated to the complete basis set limit can give gas-phase free energies as accurate as the experiment. Details will be discussed in Section 3. The problem, however, is producing accurate results without using such computationally expensive levels of theory [54]. Combinations of different methods, such as model chemistries, density functional theories (DFTs), and ab initio theories, and different basis sets have been used in an attempt to achieve an accurate but less computationally demanding method.

2.3 Solvation free energy calculations

The largest source of error in pK_a calculations is the change in free energy of solvation calculation for the reaction, which is based on the type of solvation model used and the specific level of theory [1,2,8,12]. The basic problem is that experimental free energies of solvation for ions have error bars of roughly 2–5 kcal/mol, and so models that have been developed to reproduce experimental values have the same inherent uncertainty. It is not possible to improve a particular solvation model by simply increasing the basis set, as one can when calculating ab initio quantum mechanical gas-phase values.

Explicit solvation methods include the addition of solvent molecules directly in the calculation. This method is advantageous because specific solute–solvent interactions are taken into account. These multiple interactions, however, make it more difficult to find a global minimum for the complex [23,25]. The number of necessary solvent molecules included in the reaction also comes into question, leading to the problem of balancing accuracy with computational expense. In addition, conformational effects can be daunting; it is difficult to know how many different ion–water configurations are necessary to get a conformationally averaged result. Reactions 9–12 use only one water molecule but explicit solvation methods can be used to examine the effects of adding additional waters to the reaction:



Reactions 27–29 depict some examples of explicit solvation effects, where n , or $n + m$, is the number of water molecules used in the reaction. Because of the daunting task of computing enough different configurations with a large number of water molecules, complete with frequency calculations to determine the entropic values necessary to obtain free energies for each configuration, quantum chemistry is not yet used routinely for completely explicit solvent models. Recent evidence suggests that if the standard states for water are included correctly, the use of a thermodynamic cycle based on reaction 28 will yield good values for pK_a s if a cluster cycle (and not a monomer cycle) is used for the waters [9].

In contrast to explicit solvation, implicit solvent effects, where actual solvent molecules are not included in the thermodynamic cycle, are easily implemented for pK_a calculations. Various methods used to calculate the change in free energy of solvation, such as the Dielectric Polarizable Continuum Model (DPCM) and Conductor-Like Polarizable Continuum Model (CPCM), use implicit solvation by constructing a solvation cavity around the molecule of interest. These methods have been shown to compute the free energy of solvation for neutral molecules within 1 kcal/mol [23]. The implicit models directly approximate a homogenous dielectric continuum for the neutral compounds, which represents the response of a bulk solvent. This is computationally less demanding than explicit solvation methods, but it is not particularly accurate for ionic species. Strong electrostatic effects make solvent modeling using implicit solvation more challenging [8]. The method yields less accurate values for these species [25] and may also impart a false partial positive charge on the system if wave functions penetrate beyond cavity walls [14]. Furthermore, ionic species also have larger free energies of solvation, due to solute–solvent interactions. Consequently, a smaller percentage of error is necessary for a charged species to produce the same level of chemical accuracy as a neutral molecule [23]. Aside from problems with ionic species, an additional limitation of implicit solvation is that the accuracy depends on the selection of proper boundary techniques, such as the type of solvation cavity [25].

While developing an implicit solvation model, Solvation Model 6 (SM6), Kelly, Cramer, and Truhlar found that when calculating the free energies of solvation for molecules with concentrated regions of charge densities, more accurate values were obtained by adding explicit waters in addition to the implicit effects of the model. They concluded that this occurred because of significant local solute–solvent interactions, which their implicit model did not take into account [25]. This method of including explicit solvation effects while using an implicit model is referred to as a cluster continuum model [55] or implicit–explicit model [25]. One limitation of this method is that one

must determine the number of explicit molecules that yield the most accurate results, which varies based on the type of molecules in the data set [55,56].

Along with deciding whether to use implicit or implicit-explicit solvent models, a specific level of theory and basis set must be used for the calculation of the change in free energy of solvation. Similar to the gas-phase free energy, there are a variety of methods and it can be difficult to determine what combination is the most accurate. Further discussion can be found in Section 4.

3. CALCULATING CHANGES IN FREE ENERGY IN THE GAS PHASE

The calculation of ΔG_{gas} should be the lowest source of error in $\text{p}K_{\text{a}}$ calculations because many high levels of theory can accurately predict these values. In 2006, we showed that CCSD(T) [57–60] is a highly effective method for calculating the change in gas-phase free energies for deprotonation [53]. CCSD(T) stands for coupled cluster with all single and double substitutions along with a quasi-perturbative treatment of connected triple excitations, and as of this writing, it is considered the gold standard in ab initio quantum chemistry. CCSD(T) is one of the most effective ways to include electron correlation, which results from the fact that when an electron moves, all other electrons tend to move to avoid the moving electron. Hartree-Fock theory solves the Schrödinger equation for an average electronic potential, including electron correlation is essential for obtaining meaningful energetic values and different ways of doing so consume much of the field of computational chemistry. In this case, the coupled cluster calculations included triple excitations for both the complete fourth-order Møller-Plesset (MP4) and CCSD(T) energies (for instance, by using the E4T keyword in Gaussian). The single-point CCSD(T) energy calculations used the augmented correlation consistent polarized n -tuple zeta basis sets (aug-cc-pV n Z, $n = \text{D, T, Q, 5}$) of Dunning and coworkers [61]. These calculations were performed upon geometries obtained using fourth-order Møller-Plesset perturbation theory [62] with single, double, and quadruple substitutions (MP4 (SDQ)). These optimizations, and their corresponding frequency calculations, employed the aug-cc-pVTZ basis set. The frequency calculations ensured that all structures were optimized to a true energetic minimum on the potential energy surface, and the unscaled thermochemical corrections were used to obtain the zero point energies, enthalpies, and Gibbs free energies. Furthermore, to estimate the energy at the complete basis set limit, a series of two-point extrapolations on the correlation energy were undertaken [53]. In this scheme (Eqs. 30–32), an extrapolated value for the correlation contribution to the total energy is obtained using two consecutive correlation energies, $x - 1$ and x , and is then added to a nonextrapolated Hartree-Fock energy [63–65]:

$$E_x^{\text{corr}} = E_x^{\text{CCSD(T)}} - E_x^{\text{HF}} \quad (30)$$

$$E_{x-1,x}^{\text{corr}} = \frac{x^3 E_x^{\text{corr}} - (x-1)^3 E_{x-1}^{\text{corr}}}{x^3 - (x-1)^3} \quad (31)$$

$$E_{x-1,x} = E_x^{\text{HF}} + E_{x-1,x}^{\text{corr}} \quad (32)$$

The CCSD(T)//MP4(SDQ)/aug-cc-pVTZ method, with the extrapolation to the complete basis set limit using the aug-cc-pVTZ and aug-cc-pVQZ basis sets, yielded a standard deviation of 0.58 kcal/mol when compared to a select set of experimental values of gas-phase deprotonation reactions compiled in the NIST online database, a data set with uncertainty of <1 kcal/mol [51,53]. The low error of the selected NIST data set makes these values extremely useful in determining accurate pK_a calculations and will be referenced throughout the review.

Using model chemistry methods, we also reported ΔG_{gas} calculations with slightly less accuracy, within 1.1–1.6 kcal/mol of experimental values. The model chemistries G3 [66], CBS-QB3 [49], CBS-APNO [50], and W1 [67] produced mean absolute deviations of 1.16, 1.43, 1.06, and 0.95 kcal/mol, respectively [52,53]. This data was confirmed in 2005 where G2, G3, and CBS-APNO predict accurate values of ΔG_{gas} for formation of ion–water clusters when compared to experimental results [68–72]. Contrary to the previous publications, however, CBS-QB3 was less accurate for these clustered ions than the other model chemistries. Although effective methods of gas-phase free energy calculations do exist, it would be useful to find computationally less demanding methods that produce a similar accuracy [54].

DFT may offer a more cost-efficient approach, although it must be remembered that each particular DFT functional and a given basis set represents its own parameterized method. DFT includes some of the correlation energy, although the exact solution to recover all of it is still unknown and is the subject of much theoretical research. For example, fairly accurate results were obtained for PBE1PBE/aug-cc-pVTZ and B3P86/aug-cc-pVTZ [54]. The select NIST data test set included the deprotonation reactions of the following compounds: ammonia, methylamine, dimethylamine, ethylamine, methane, methanol, water, acetylene, ethylene, formaldehyde, hydrogen chloride, propene, nitrous acid, nitric acid, isocyanic acid, furan, and benzene. When compared to experimental results within the NIST database [51], the mean square ΔG_{gas} deviations for both PBE1PBE/aug-cc-pVTZ and B3P86/aug-cc-pVTZ were 1.6 kcal/mol, exhibiting somewhat less accuracy compared to more computationally expensive methods.

In another study, G3MP2, G2, G3, G2MP2, G3B3, G3MP2B3, QCISD(T), CBS-4, CBS-Q, CBS-QB3, and CBS-APNO produced ΔG_{gas} values for nitrous acid within 0–1.6 kcal/mol [73] of the experimental value of 333.7 kcal/mol [74]. They also found that the less expensive density functional B3LYP produced values within 2.72 kcal/mol of experiment. The commonly used Hartree–Fock level of theory, which does not include correlation energy, produced inaccurate results with a large 4.66 kcal/mol discrepancy [73].

The accuracy of B3LYP has been examined by numerous researchers. In 2003, Fu and colleagues reported that the MP2/6-311++G(d,p) and B3LYP/6-311++ G(2df, p) methods yielded gas-phase acidities, or the change in free energy of the reaction:



within 2.2 and 2.3 kcal/mol of experimental values of various organic acids reported in the NIST online database [27,51]. Two years later, Range also reported that B3LYP with the 6-311++G(3df,2p) basis produced a root-mean-square error of 2.5 kcal/mol for reaction 33, when compared to experimental values from the NIST online database [51]. The article also reported that previously discussed high levels of theory, CBS-QB3, G3B3, G3MP2B3, PBE0, and B1B98, have a RMSE all within 1.3 kcal/mol of experimental values [75]. Reaction 33 represents the gas-phase dissociation of an acid, which is the top line of Figure 1.

Other publications, however, report more accurate values of B3LYP gas-phase free energy calculations on aliphatic amines, diamines, and aminoamines. In 2007, Bryantsev et al. reported that B3LYP calculations with the basis set 6-31++G** had a mean absolute error of 0.78 kcal/mol from experimental values of the gas-phase basicity (ΔG_{gas}) of the reverse reaction of Eq. (1) reported in the NIST database [51]. This accuracy is comparable to that of expensive, high-level model chemistries but, because the experimental values have uncertainties of ± 2 kcal/mol, it is difficult to discern exactly how accurate the calculations are in comparison to values in the other publications [76]. The take-home message remains the same: always benchmark DFT calculations for the systems you are interested in computing.

4. CALCULATING CHANGES IN FREE ENERGY IN SOLUTION

The change in free energy of solvation calculation for the reaction is the largest source of error in $\text{p}K_{\text{a}}$ calculations. To determine the most accurate method we must look both at the type of solvation model used, implicit, explicit, or cluster continuum method (likewise described as implicit-explicit), and the specific level of theory. As previously mentioned, ionic species, in particular, are extremely difficult to calculate because of their strong electrostatic effects and large free energy of solvation values [8,14,23,25].

Implicit solvation models developed for condensed phases represent the solvent by a continuous electric field, and are based on the Poisson equation, which is valid when a surrounding dielectric medium responds linearly to the charge distribution of the solute. The Poisson equation is actually a special case of the Poisson-Boltzmann (PB) equation: PB electrostatics applies when electrolytes are present in solution while the Poisson equation applies when no ions are present. Solving the Poisson equation for an arbitrary equation requires numerical methods, and many researchers have developed an alternative way to approximate the Poisson equation that can be solved analytically, known as the

Generalized Born (GB) approach. The most common implicit models used for small molecules are the Conductor-Like Screening Model (COSMO) [77,78], the DPCM [79], the Conductor-Like Modification to the Polarized Continuum Model (CPCM) [80,81], the Integral Equation Formalism Implementation of PCM (IEF-PCM) [82] PB models, and the GB SMx models of Cramer and Truhlar [23,83–86]. The newest Minnesota solvation models are the SMD universal Solvation Model based on solute electron density [26] and the SMLVE method, which combines the surface and volume polarization for electrostatic interactions model (SVPE) [87–89] with semiempirical terms that account for local electrostatics [90]. Further details on these methods can be found in Chapter 11 of Reference [23].

Kelly et al. [8] used the cluster continuum model in their study of aqueous acid dissociation constants. They compared the correlation between experimental pK_a values and the calculated acid dissociation free energies of anions with and without an additional explicit water molecule using SM6. Note that because of the relation between pK_a and ΔG_{aq} as shown in Eq. (4), a plot of pK_a versus ΔG_{aq} should yield a slope of $1/2.303RT$ or $1/RT \ln(10)$. This single water molecule was added only to ions containing three or fewer atoms or ones with oxygen atoms bearing a more negative partial atomic charge than that of the water solute. They reported that when only implicit effects were included, a regression equation with a slope of $0.71/RT \ln(10)$ and correlation $r^2 = 0.76$ was computed. When an explicit water was added to the acids, however, the new regression yielded a slope of $0.87/RT \ln(10)$ with a correlation of $r^2 = 0.86$. From this observation, Kelly, Cramer, and Truhlar concluded that, for some anions, the accuracy of acid dissociation energies greatly increased with the addition of one explicit water molecule. Furthermore, they attested that an implicit model alone cannot produce such accurate results because it does not take into account strong solute–solvent interactions. They argue that previous publications using implicit methods with strong correlations between pK_a values and free energy calculations actually have underlying systematic errors in their methods, as indicated by lower slopes [8].

Related to Kelly, Cramer, and Truhlar's conclusion, Klamt, Eckert, and Diedenhofen's 2003 publication studied the correlation between experimental pK_a values and the free energies of dissociation for 64 organic and inorganic acids for reaction 9. Like the Kelly publication, Klamt, Eckert, and Diedenhofen used an explicit water molecule. Their solvent calculations, however, used Klamt's COSMO-RS level of theory [78]. They reported a correlation of $r^2 = 0.984$ with a standard deviation of only 0.49. The slope of the regression line, however, was fairly low at $0.58/RT \ln(10)$. Klamt, Eckert, and Diedenhofen claim that this discrepancy is not due to the weakness of the calculation method [91]. Another study by Eckert and Klamt in 2006 confirmed these results by reporting that a correlation of experimental pK_a values with free energies of dissociation had an $r^2 = 0.98$ with a deviation of 0.56 pK_a units and again a significantly smaller slope than the accepted $1/RT \ln(10)$ [92].

These values seem to indicate that COSMO-RS, contrary to Kelly, Cramer, and Truhlar's publication, is actually more accurate than SM6. However, Kelly and colleagues might claim that this lower slope indicates underlying systematic error, and we would agree.

Although Kelly, Cramer, and Truhlar found that a single explicit water molecule increased the accuracy of acid dissociation free energies for SM6, this trend was not common to all solvation methods or for the addition of multiple water molecules. They reported the gas-phase binding free energies of $(\text{H}_2\text{O})_n \text{CO}_3^{-2}$ with $n = 0$ to $n = 3$ for SM6, SM5.43R, and DPCM/98 with UAHF atomic radii levels of theory. As the number of water molecules increased from zero to three, the accuracy also increased for SM6. The two other continuum solvation models, SM5.43R and DPCM, however, decreased in the accuracy of gas-phase binding free energies as the number of explicit water molecules increased. With the addition of one water molecule, SM5.43R did become more accurate. It got significantly worse, however, as the number of water molecules increased, as the absolute deviations went from 1 kcal/mol with one water molecule to 10 kcal/mol with three molecules. The most accurate calculation for DPCM was with no explicit water molecules and the calculations continued to become less accurate with increased numbers of waters surrounding CO_3^{-2} . Overall, the most accurate method for the study was SM5.43R with one explicit water molecule, outperforming SM6 with three water molecules, which had an absolute deviation of 3 kcal/mol [8].

To supplement their previous publication, Kelly, Cramer, and Truhlar also conducted an extensive study surrounding the absolute aqueous solvation free energies of ions and ion–water clusters containing a single water molecule. They reported the following mean unsigned errors using values from their recent [12] and previous publication [25].

Table 1 shows that SM6 outperformed all continuum models, with SM6/MPW25/6-31G(d) producing the lowest mean unsigned error (MUE) of

Table 1 Mean unsigned errors in absolute aqueous solvation free energies of ions and ion–water clusters, with a single water molecule, for various continuum solvent models [12,25]

| Solvent model | Clustered data set ^a | All ions ^b |
|------------------------------|---------------------------------|-----------------------|
| SM6/MPW25/MIDI! | 3.7 | 4.8 |
| SM6/MPW25/6-31G(d) | 3.3 | 4.5 |
| SM6/MPW25/6-31+G(d) | 3.5 | 4.6 |
| SM6/MPW25/6-31+G(d,p) | 3.5 | 4.5 |
| SM6/B3LYP/6-31+G(d,p) | 3.6 | 4.7 |
| SM6/B3PW91/6-31+G(d,p) | 3.5 | 4.6 |
| SM5.43R/MPW25/6-31+G(d,p) | 5.3 | 6.1 |
| DPCM/98/HF/6-31G(d) | 5.8 | 5.7 |
| DPCM/03/HF/6-31G(d) | 13 | 14.3 |
| CPCM/98/HF/6-31G(d) | 6 | 6 |
| CPCM/03/HF/6-31G(d) | 7.3 | 7.1 |
| IEF-PCM/03/HF/6-31G(d) | 7.4 | 7.2 |
| IEF-PCM/03/MPW25/6-31+G(d,p) | 8.6 | 8.4 |

^a Gas-phase optimized geometries at the B97-1/MG3S level of theory.

^b Gas-phase optimized geometries at the MPW25/MIDI! level of theory.

3.3 kcal/mol when used with clustered ions. The data also shows that the clustered ions resulted in lower MUE than the unclustered ions for all SM6 calculations by about 1 kcal/mol, reaffirming their conclusion derived from their prior publication [8]. Other levels of theory, however, do not produce as conclusive results and do not always produce lower MUE when implementing the cluster pair approximation. Overall, Kelly et al. [12] concluded that SM6 with diffuse basis functionals and clustered ions produce the most reliable values for the absolute aqueous solvation free energies ($A^-_{(\text{gas})} \rightarrow A^-_{(\text{aq})}$). This implies that this method would also lead to the most accurate ΔG_{sol} values. Furthermore, as the parameters in SM6 were originally developed based on an older accepted ΔG_{sol} H^+ value of -264.3 kcal/mol reported by Zhan and Dixon [22], this was a significant finding because it showed that the SM6 calculations are also accurate when combined with the currently accepted ΔG_{sol} H^+ value of -265.9 kcal/mol [10,12].

Jia et al. also studied the cluster continuum method using PCM with the HF/6-31+G(d), HF/6-311++G(d,p), and B3LYP/6-311++G(d,p) levels of theory for Eq. (13). For a data set of five organic acids, they found that the accuracy of the pK_a calculations increases as the number of explicit water molecules increases from 0 to 3 [93]. In this study, relative pK_a values were computed, so that lack of electron correlation in the gas-phase calculation apparently cancelled.

Focusing on implicit calculations, da Silva et al. compared the success of DPCM and IEF-PCM in pK_a calculations at the HF/6-31G(d) level using these polarizable continuum solvent models with UAHF radii and 15 different levels of free energy gas-phase calculations, HF, MP2, QCISD(T), B3LYP, G1, G2, G2MP2, G3, G3B3, G3MP2, CBS-4, CBS-4M, CBS-Q, CBS-QB3, CBS-APNO, for the ΔG_{gas} calculation. The HF, MP2, and QCISD(T) theories use the 6-311++G(3df,3pd) basis set. Overall, they found that DPCM was more successful than IEF-PCM at calculating the pK_a value of nitrous acid. The most successful method was DPCM with the gas-phase calculation at B3LYP/6-311++G(3df,3pd), which produced an error of only 0.3 pK_a units. da Silva et al. however, concluded that this was probably due to a cancellation of errors. The other accurate values were calculated using high-level theories: CBS-APNO, CBS-QB3, and G2 [73].

Using the DPCM method, da Silva et al. also examined the effect of different basis sets with the HF level of theory using gas-phase geometries from G2, CBS-Q, CBS-QB3, and CBS-APNO calculations. They reported that as the basis set size increased, excluding aug-cc-pVTZ, the accuracy of the pK_a calculation also increased. The most accurate basis set paired with HF was aug-cc-pVQZ, which produced an absolute average error of 0.39 pK_a units. da Silva also studied DPCM with DFT methods instead of HF. The average pK_a values were calculated using G2, CBS-Q, CBS-QB3, and CBS-APNO for ΔG_{gas} values. The free energy of solvation was calculated using B3LYP, TPSS, PBE0, B1B95, VSXC, B98, and O3LYP with 6-31G(d), 6-311++G(3df,3pd), aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ basis sets. Results indicated that the use of DFT methods produce much more accurate results than HF, with all theories within 0.3 pK_a units of experimental values. The most accurate methods were VSXC, TPSS, B98, and B1B95, all with absolute average errors of less than 0.15 pK_a units. Unlike the HF

results, da Silva found no benefit in using larger basis sets with DFT [73]. This observation rings true as DFT methods are semiempirical and each method with a given basis set is its own distinct model chemistry suitable for specific systems [94]. Increasing the basis set size does not systematically improve the result like it does in *ab initio* quantum chemistry [23,54].

In 2007, Sadlej-Sosnowska compared the DPCM, CPCM, and IEF-PCM levels of theory for the free energy of solvation calculations. The three methods were observed at the HF and B3LYP levels of theory with basis sets 6-31+G*, 6-311++G*, pVDZ, and pVTZ. DPCM was used with a UAHF radius and IEF-PCM was paired with UAHF and UA0. Sadlej-Sosnowska found that IEF-PCM with UAHF was more accurate than DPCM with UAHF when applied to neutral molecules, in contrast to the da Silva et al. results [73]. The most accurate level of theory was IEF-PCM with UAHF at the HF/cc-pVTZ level. In comparing radii, IEF-PCM with UAHF was more accurate than UA0 [95].

Takano and Houk studied several computational methods of solvation calculations and various cavity models in their 2005 publication [7]. They found that for the calculation of aqueous solvation free energies the CPCM method at the HF/6-31G(d)//HF/6-31+G(d) and HF/6-31+G(d)//B3LYP/6-31+G(d), with UAKS cavities that have radii optimized with PBE0/6-31G(d), are the most accurate, and produce mean absolute deviations of 2.6 kcal/mol. This mean absolute deviation was based on calculations concerning neutral and charged species. The accuracy of each cavity, UAKS, UAHF(G03), UAHF(G98), Bondi, Pauling, UA0, and UFF, varied based on the type of molecule. CPCM with a UAKS cavity model was also compared to COSMO, SM5.42R, PCM, IPCM [96], and cluster continuum with MP2/6-31+G(2df,2p)//HF//6-31+G(d,p) and IPCM. The CPCM data in Takano's publication was compared with SM5.42R, PCM, IPCM, and cluster continuum data from Pliego and Riveros [97] and COSMO data from various publications [77,98,99]. CPCM was found to have the highest accuracy with a mean absolute deviation from experimental aqueous solvation free energies of 3.04 kcal/mol [7]. The other methods had about a 10 kcal/mol deviation, except IPCM with a MAD of 20 kcal/mol [7].

Yu, Liu, and Wang also studied the effect of cavity models. Their pK_a calculations were based upon a system containing an explicit water molecule. Wang and colleagues studied the effect of UAHF, UAKS, Pauling, and Bondi cavity models on the accuracy of CPCM pK_a calculations at the B3LYP/6-311++G(2df,2p) level of theory on a B3LYP/6-31+G(d) optimized geometry. They reported that the pK_a calculation depends greatly on the choice of solvation cavity. The most accurate methods were CPCM with an UAKS or UAHF cavity, which produced mean absolute deviations of 0.38 and 0.40 pK_a units, respectively, from experimental pK_a values [100]. This correlates well with the 2005 data reported by Takano and Houk [7].

Namazian and Halvani studied pK_a calculations with an explicit water using the B3LYP/6-31+G(d,p) level of theory for free energy calculations in the gas phase and PCM/B3LYP/6-31+G(d,p) with the UA0 radius for solvation calculations. Using a data set of 66 acids, they found the method accurate within an average of 0.58 pK_a units. The thermodynamic cycles used an explicit water, as in Eq. (9). Although the method produced pK_a values within 0.6 pK_a units, there is

some uncertainty in the free energy of solvation of H_3O^+ and the B3LYP level of theory is not the most accurate for free energy calculations in the gas phase [31].

Gao and colleagues studied several methods of solvation calculations [29]:

- S1: CPCM/HF/6-311+G(d,p) // HF/6-311+G(d,p)
- S2: CPCM/B3LYP/6-311+G(d,p) // HF/6-311+G(d,p)
- S3: CPCM/HF/6-311+G(d,p) // CPCM/HF/6-311+G(d,p)
- S4: CPCM/B3LYP/6-311+G(d,p) // CPCM/HF/6-311+G(d,p)
- S5: CPCM/HF/6-31G(d) (Radii=UAHF) // HF/6-311+G(d,p)
- S6: CPCM/HF/6-31G(d) (Radii=UAHF) // CPCM/HF/6-311+G(d,p)
- S7: SM5.4/AM1 calculated from Spartan 04
- S8: SM5.4/PM3 calculated from Spartan 04
- S9: SM5.4/AM1 taken from Reference [101] (AMSOL)
- S10: SM5.4/PM3 taken from Reference [101] (AMSOL)
- S11: SM5.43R/mPW1PW91/6-31+G(d) // mPW1PW91/MIDI!
- S12: CPCM/HF/6-31+G(d) (Radii=UAKS) // HF/6-311+G(d,p)
- S13: CPCM/HF/6-31+G(d) (Radii=UAKS) // CPCM/HF/6-311+G(d,p)
- S14: Monte Carlo QM/MM

They found that methods CPCM/HF/6-31G(d) (Radii=UAHF) // CPCM/HF/6-311+G(d,p), SM5.4/PM3 calculated from Spartan 04, SM5.4/AM1 taken from Reference [101] (AMSOL), and Monte Carlo QM/MM produced free energy of hydration values within 1 kcal/mol of experimental values. The free energy of hydration for the anion, $\text{A}^-_{\text{gas}} \rightarrow \text{A}^-_{\text{aq}}$, would be a vertical line on the right side of the thermodynamic cycle shown in Figure 1.

The CPCM/HF/6-31G(d) (Radii=UAHF) // CPCM/HF/6-311+G(d,p) and Monte Carlo QM/MM levels of theory were the most accurate and produced free energy of hydration within 0.4 kcal/mol of experimental values [29].

Recognizing that the continuum solvent calculations are the weakest link in pK_a calculations, Ho and Coote [3] used the CPCM (with UAKS and UAFH radii), SM6, IPCM, and COSMO-RS models to predict pK_a values for a common data set of neutral organic and inorganic acids. They used four different thermodynamic cycles, and in general found that the COSMO-RS, CPCM, and SM6 models worked best depending on the thermodynamic cycle used. We turn to a discussion of thermodynamic cycles in the next section.

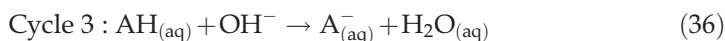
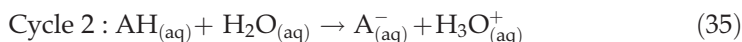
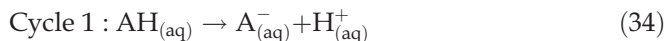
5. THERMODYNAMIC CYCLES

Various thermodynamic cycles are used in pK_a calculations. Although previously a source of confusion in the field, it is now clear that as long as the most accurate experimental values are used, and no explicit water molecules are added, the choice of cycle should just be a matter of convenience. The most common is based on Eq. (1) and is diagrammed in Figure 1, where a molecule is simply deprotonated, yielding a corresponding base and the proton in solution [1,2,4]. This cycle depends on the accuracy of the continuum model used to determine the anion (reaction 1) or cation (reaction 8) solvation energies, calculations that vary in

accuracy depending on the system in question and the parameterization of the solvation model. The most recent Minnesota solvation model, SMVLE [90], appears quite promising in this regard. In this method the SVPE term accounts for bulk electrostatics; the atomic surface tensions account for solvent cavitation, changes in dispersion energy, and any changes in local solvent structure; and a new functional term that accounts explicitly for local electrostatics at the solute-solvent boundary [90]. Because the SMVLE model specifically corrects for surface regions showing high charge concentrations, it gives the most accurate aqueous solvation free energies for ions relative to any other continuum method as of this writing.

Other thermodynamic cycles used in various calculations contain explicit water molecules. The effectiveness of this implicit-explicit method in terms of calculating the free energy of solvation is discussed in the preceding section. Since the free energy of solvation is the largest source of error in determining pK_a values, the accuracy of the calculation often determines the validity of various thermodynamic cycles.

In 2007, Sadlej-Sosnowska studied three different thermodynamic cycles based on the following reactions [95]:



For Cycle 1, the free energies of solvation of AH, A^{-} , and H^{+} are required, along with the gas-phase free energy difference for $\text{AH}(\text{g}) \rightarrow \text{A}^{-}(\text{g}) + \text{H}^{+}(\text{g})$. For Cycle 2, the free energies of solvation for AH, H_2O , A^{-} , and H_3O^{+} are required, along with the gas-phase free energy difference for $\text{AH}(\text{g}) + \text{H}_2\text{O}(\text{g}) \rightarrow \text{A}^{-}(\text{g}) + \text{H}_3\text{O}^{+}(\text{g})$. Finally, for Cycle 3, instead of H_3O^{+} , we need the free energy of solvation for OH^{-} and the gas-phase free energy difference for $\text{AH}(\text{g}) + \text{OH}^{-}(\text{g}) \rightarrow \text{A}^{-}(\text{g}) + \text{H}_2\text{O}(\text{g})$.

Sadlej-Sosnowska found that Cycle 1 was the most accurate because there was a large uncertainty in the calculation of the free energy of solvation of H_3O^{+} and OH^{-} in Cycle 2 and 3, respectively. However, when experimental values were used in place of calculated values for ΔG_{sol} and ΔG_{gas} of OH^{-} and H_3O^{+} , all three cycles produced the same pK_a values. Sadlej-Sosnowska stated that the accuracy of Cycle 1 resulted from the use of the most accurate experimental ΔG_{sol} H^{+} value and if this value is changed then the results will shift accordingly [95]. This illustrates a good point, since water dissociates to H^{+} and OH^{-} , or two waters dissociate to H_3O^{+} and OH^{-} ; if the correct value for ΔG_{sol} H^{+} is used in Cycle 1 for reaction 34, that is equivalent to using the correct values for ΔG_{sol} for H_2O and H_3O^{+} in reaction 35, and to the correct values for ΔG_{sol} for OH^{-} and H_2O in reaction 36. Once the value for ΔG_{sol} H^{+} is set, the rest of the values are easily found from this reference, and the $\Delta \Delta G_{\text{sol}}$ values should be consistent [12].

Many thermodynamic cycles contain the hydrogen proton and anions, which often leads to a large error in the computational calculation of the free energy of solvation of the anion. As a result, cycles with water molecules or additional acids [5,27–31] are often used to try and remove these sources of error. If accurate free energy values are used, pK_a calculations can be fairly accurate, but many papers report pK_a calculations with less accurate free energy values for H^+ . These publications would need to be recalculated with more accepted values to produce reliable and accurate data.

For example, in 2006, Nino and coworkers studied the efficiency of model chemistries G1 [102], G2 [103], and G3 [66] for pK_a calculations [30]. They reported that G1 is the most accurate of the three models, producing an average difference of 0.51 pK_a units from experimental values when used with the CPCM solvation model in the study of aminopyridines. Results showed that the models decreased in accuracy, $G1 > G2 > G3$. The publication, however, contains the older accepted $\Delta G_{sol} H^+$ value of -264.61 kcal/mol in the 1M standard state, obtained from the acetic acid system [1,2], which as we have discussed in Section 2.1.1 is 1.29 kcal/mol more positive than the currently accepted -265.9 kcal/mol reported by Tissandier et al. in 1998 [10] and confirmed by Kelly et al. in 2006 [12]. This difference creates a discrepancy in those pK_a values calculated in the paper and the actual values produced by the model chemistries. Changing the value for $\Delta G_{sol} H^+$ from -264.61 to -265.9 produces new pK_a values that are approximately 0.95 pK_a units less than those reported by Caballero. This changes the order of accuracy of the methods. After calculating new pK_a values and comparing them to the experimental values reported in the Caballero publication, we find that the previously reported least accurate method, G3, is now the most effective in producing reliable pK_a values. This correction outlines the importance of using correct experimental values and how a difference of merely 1.29 kcal/mol can change the conclusions about the most efficient method of pK_a calculation. G3 has been shown to be superior to G2 (and G2 to G1) for many different gas-phase processes, including deprotonation, so the new ordering makes more sense. Quite simply, Nino and colleagues were led astray by the use of the accepted value for $\Delta G_{sol} H^+$ at the time they started their work.

The significance of the free energy of solvation of a proton is also apparent in the publication by Bryantsev et al. [76]. In this work the $\Delta G_{sol} H^+$ value was treated as a parameter and fitted in order to obtain the most accurate pK_a values. The Goddard group used $\Delta G_{sol} H^+ = -267.9$ kcal/mol and -267.6 kcal/mol for solution-phase and gas-phase optimized calculations, respectively, for the 1M standard state. The values are off from the accepted value of -265.9 kcal/mol, however, still within the 2 kcal/mol error bars assigned by Kelly et al. [12]. Nevertheless, because of this discrepancy, the reported accuracy of less than 0.5 pK_a units for solution-phase optimized geometries and greater than 0.5 pK_a units for calculations on gas-phase optimized geometries might be overstated. These results will change when the accepted $\Delta G_{sol} H^+$ value is used.

Recent work on using explicit waters in cluster continuum or implicit-explicit thermodynamic cycles show much promise, as long as the standard state issues

for water are consistent [3,9]. The key point is that water as a solvent, water as a solute, and all species involved in the thermodynamic cycle must be in a 1M standard state. At this point it is not clear how many explicit waters should be used in a cycle [3], although use of the variational method to determine the number of waters to be used, and putting the waters together as clusters instead of monomers, appears to have much promise [9].

6. CONCLUDING REMARKS

Many variables affect the accuracy of pK_a calculations. With regard to the free energy calculation in the gas phase, extra computational expense might be necessary to achieve values within 1.0 kcal/mol for reaction 33. However, getting this right is straightforward. CCSD(T) single-point energy calculations on MP2 or MP4 geometries are accurate within a half kcal/mol or better. DFT methods should be benchmarked against appropriate experimental or ab initio results to ensure that the DFT method of choice is suitable for the systems of interest. Compound model chemistry methods such as G3, CBS-APNO, and W1 are also highly accurate.

For the free energy of solvation calculation, however, it is difficult to discern the most accurate method. Recently, there have been numerous publications exploring the use of the cluster continuum method with anions. With regard to implicit solvation, there are no definite conclusions to the most accurate method, yet for the PB models the conductor-like models (COSMO; CPCM) appear to be the most robust over the widest range of circumstances [23]. At this writing, the SMVLE method seems to be the most versatile, as it can be used by itself, or with the implicit-explicit model, and the error bars for bare and clustered ions are the smallest of any continuum solvation method. The ability to add explicit water molecules to anions and then use the implicit method (making it an implicit-explicit model) improves the results more often than the other implicit methods that have been used in the literature to date.

Concerning thermodynamic cycles, the most important component is the treatment of the free energy of the hydrogen ions. Even a slight difference in values can produce drastically different trends in pK_a . The most accurate experimental value should be used in the equation. As of this writing the best values for the experimental free energies of solvation, for a standard state of 1M and 298 K, are -265.9 kcal/mol for H^+ , -104.7 kcal/mol for OH^- , -110.3 kcal/mol for H_3O^+ , and -6.32 kcal/mol for H_2O [9-12]. These values are all consistent to each other, as can be seen by using them in thermodynamic cycles to calculate the dissociation of water into its component ions, where ΔG_{gas} is obtained from the NIST website. For the classic thermodynamic cycle displayed in Figure 1, using the accepted value for $\Delta G_{sol} H^+$ and considering the conversion of gas-phase calculations to the 1M standard state (Eq. 25), pK_a values for the reaction in Eq. (1) at 298.15 K can be determined using Eq. (37), with the four calculated energies in kcal/mol:

$$pK_a = [G_{gas}(A^-) - G_{gas}(HA) + \Delta G_{sol}(A^-) - G_{sol}(HA) - 270.28567]/1.36449 \quad (37)$$

Extra caution should be taken when performing pK_a calculations on ionic species, as their strong electrostatic effects and large free energies of solvation make accurate calculations difficult. Cycles involving explicit water molecules have their merit when dealing with these compounds. Interested readers should refer to the recent literature to ensure that they correct for the standard state of water, which should be 1M and not 55.34M in all cycles [3,9]. To further complicate matters, various functional groups or acidic/basic strength of the molecules may also affect the accuracy of methods. If the implicit solvent method used in the calculation of $\Delta G_{\text{sol}}(\text{A}^-)$ and $\Delta G_{\text{sol}}(\text{HA})$ are believed to yield good results for the species in question, then using thermodynamic cycle 1 of Figure 1 and Eq. (37) is the most straightforward way to calculate pK_a values. Investigators are encouraged to use the highest level of theory they can afford to calculate $G_{\text{gas}}(\text{A}^-)$ and $G_{\text{gas}}(\text{HA})$.

Due to the numerous potential cycles using explicit molecules, levels of theory, basis sets, and types of molecules, it is impossible to determine one specific method that produces the most accurate pK_a values. Rather, this review serves to summarize the current literature and illustrate various schemes that have been successful. Accurate attention to detail and the use of benchmark calculations or experimental values to assist in determination of the correct method to use for a particular system is highly recommended. Further research on thermodynamic cycles using explicit cycles, clustered water structures, conformational effects, and advances in continuum solvation calculations will continue to advance this field.

REFERENCES

1. 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(30), 7314–9.
2. Liptak, M.D., Gross, K.C., Seybold, P.G., Feldgus, S., Shields, G.C. Absolute pK_a determinations for substituted phenols. *J. Am. Chem. Soc.* 2002, 124(22), 6421–7.
3. Ho, J.M., Coote, M.L. A universal approach for continuum solvent $pK(a)$ calculations: Are we there yet? *Theor. Chem. Acc.* 2010, 125(1–2), 3–21.
4. Liptak, M.D., Shields, G.C. Experimentation with different thermodynamic cycles used for pK_a calculations on carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *Int. J. Quantum Chem.* 2001, 85, 727–41.
5. Brown, T.N., Mora-Diez, N. Computational determination of aqueous $pK(a)$ values of protonated benzimidazoles (part 1). *J. Phys. Chem. B* 2006, 110(18), 9270–79.
6. Pliego, J.R. Thermodynamic cycles and the calculation of $pK(a)$. *Chem. Phys. Lett.* 2003, 367(1–2), 145–9.
7. Takano, Y., Houk, K.N. Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. *J. Chem. Theory. Comput.* 2005, 1, 70–7.
8. 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(7), 2493–9.
9. Bryantsev, V.S., Diallo, M.S., Goddard, W.A. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *J. Phys. Chem. B* 2008, 112(32), 9709–19.

10. Tissandier, M.D., Cowen, K.A., Feng, W.Y., Gundlach, E., Cohen, M.H., Earhart, A.D., Coe, J.V., Tuttle, T.R. The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. *J. Phys. Chem. A* 1998, 102(40), 7787–94.
11. 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(47), 10795–7.
12. 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(32), 16066–81.
13. Pliego, J.R., Riveros, J.M. Gibbs energy of solvation of organic ions in aqueous and dimethyl sulfoxide solutions. *Phys. Chem. Chem. Phys.* 2002, 4(9), 1622–7.
14. Pliego, J.R., Riveros, J.M. Theoretical calculation of $pK(a)$ using the cluster-continuum model. *J. Phys. Chem. A* 2002, 106(32), 7434–9.
15. McQuarrie, D.M. *Statistical Mechanics*, Harper and Row, New York, 1970, p. 86.
16. Jorgensen, W.L., Briggs, J.M., Gao, J. A priori calculations of pK_a s for organic compounds in water—the pK_a of ethane. *J. Am. Chem. Soc.* 1987, 109(22), 6857–8.
17. Jorgensen, W.L., Briggs, J.M. A priori pK_a calculations and the hydrations of organic-anions. *J. Am. Chem. Soc.* 1989, 111(12), 4190–7.
18. Lim, C., Bashford, D., Karplus, M. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* 1991, 95(14), 5610–20.
19. Schuurmann, G., Cossi, M., Barone, V., Tomasi, J. Prediction of the $pK(a)$ of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF. *J. Phys. Chem. A* 1998, 102(33), 6706–12.
20. 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(10), 4595–606.
21. Tawa, G.J., Topol, I.A., Burt, S.K., Caldwell, R.A., Rashin, A.A. Calculation of the aqueous free energy of the proton. *J. Chem. Phys.* 1998, 109(12), 4852–63.
22. Zhan, C.G., Dixon, D.A. Absolute hydration free energy of the proton from first-principles electronic structure calculations. *J. Phys. Chem. A* 2001, 105(51), 11534–40.
23. Cramer, C.J. *Essentials of Computational Chemistry: Theories and Models*, 2nd edn., John Wiley & Sons Ltd, Chichester, England, 2004, p. 579.
24. Palascak, M.W., Shields, G.C. Accurate experimental values for the free energies of hydration of H^+ , OH^- , and H_3O^+ . *J. Phys. Chem. A* 2004, 108(16), 3692–4.
25. Kelly, C.P., Cramer, C.J., Truhlar, D.G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *J. Chem. Theory. Comput.* 2005, 1(6), 1133–52.
26. Marenich, A.V., Cramer, C.J., Truhlar, D.G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* 2009, 113(18), 6378–96.
27. Fu, Y., Liu, L., Li, R.C., Liu, R., Guo, Q.X. First-principle predictions of absolute $pK(a)$'s of organic acids in dimethyl sulfoxide solution. *J. Am. Chem. Soc.* 2004, 126(3), 814–22.
28. De Abreu, H.A., De Almeida, W.B., Duarte, H.A. $pK(a)$ calculation of poliprotic acid: Histamine. *Chem. Phys. Lett.* 2004, 383(1–2), 47–52.
29. Gao, D.Q., Svoronos, P., Wong, P.K., Maddalena, D., Hwang, J., Walker, H. $pK(a)$ of acetate in water: A computational study. *J. Phys. Chem. A* 2005, 109(47), 10776–85.
30. Caballero, N.A., Melendez, F.J., Munoz-Cara, C., Nino, A. Theoretical prediction of relative and absolute $pK(a)$ values of aminopyridines. *Biophys. Chem.* 2006, 124(2), 155–60.
31. Namazian, M., Halvani, S. Calculations of $pK(a)$ values of carboxylic acids in aqueous solution using density functional theory. *J. Chem. Thermodyn.* 2006, 38(12), 1495–502.
32. Ho, J.M., Coote, M.L. $pK(a)$ calculation of some biologically important carbon acids—an assessment of contemporary theoretical procedures. *J. Chem. Theory. Comput.* 2009, 5(2), 295–306.
33. Bickmore, B.R., Tadanier, C.J., Rosso, K.M., Monn, W.D., Eggett, D.L. Bond-valence methods for $pK(a)$ prediction: Critical reanalysis and a new approach. *Geochim. Cosmochim. Acta* 2004, 68(9), 2025–42.

34. da Silva, R.R., Ramalho, T.C., Santos, J.M., Figueroa-Villar, J.D. On the limits of highest-occupied molecular orbital driven reactions: The frontier effective-for-reaction molecular orbital concept. *J. Phys. Chem. A* 2006, 110(3), 1031–40.
35. Habibi-Yangjeh, A., Danandeh-Jenagharad, M., Nooshyar, M. Application of artificial neural networks for predicting the aqueous acidity of various phenols using QSAR. *J. Mol. Model.* 2006, 12(3), 338–47.
36. Seybold, P.G., May, M., Bagal, U.A. Molecular-structure property relationships. *J. Chem. Educ.* 1987, 64(7), 575–81.
37. Needham, D.E., Wei, I.C., Seybold, P.G. Molecular modeling of the physical properties of the alkanes. *J. Am. Chem. Soc.* 1988, 110(13), 4186–94.
38. Seybold, P.G. Explorations of molecular structure-property relationships. *SAR QSAR Environ. Res.* 1999, 10(2–3), 101–15.
39. Gross, K.C., Seybold, P.G. Substituent effects on the physical properties and $pK(a)$ of aniline. *Int. J. Quantum Chem.* 2000, 80(4–5), 1107–15.
40. Gross, K.C., Seybold, P.G. Substituent effects on the physical properties and $pK(a)$ of phenol. *Int. J. Quantum Chem.* 2001, 85(4–5), 569–79.
41. Gross, K.C., Seybold, P.G., Peralta-Inga, Z., Murray, J.S., Politzer, P. Comparison of quantum chemical parameters and Hammett constants in correlating $pK(a)$ values of substituted anilines. *J. Org. Chem.* 2001, 66(21), 6919–25.
42. Gross, K.C., Seybold, P.G., Hadad, C.M. Comparison of different atomic charge schemes for predicting $pK(a)$ variations in substituted anilines and phenols. *Int. J. Quantum Chem.* 2002, 90(1), 445–58.
43. Hollingsworth, C.A., Seybold, P.G., Hadad, C.M. Substituent effects on the electronic structure and $pK(a)$ of benzoic acid. *Int. J. Quantum Chem.* 2002, 90(4–5), 1396–403.
44. Ma, Y.G., Gross, K.C., Hollingsworth, C.A., Seybold, P.G., Murray, J.S. Relationships between aqueous acidities and computed surface-electrostatic potentials and local ionization energies of substituted phenols and benzoic acids. *J. Mol. Model.* 2004, 10(4), 235–9.
45. Peterangelo, S.C., Seybold, P.G. Synergistic interactions among QSAR descriptors. *Int. J. Quantum Chem.* 2004, 96(1), 1–9.
46. Seybold, P.G. Analysis of the $pK(a)$ s of aliphatic amines using quantum chemical descriptors. *Int. J. Quantum Chem.* 2008, 108(15), 2849–55.
47. Kreye, W.C., Seybold, P.G. Correlations between quantum chemical indices and the $pK(a)$ s of a diverse set of organic phenols. *Int. J. Quantum Chem.* 2009, 109(15), 3679–84.
48. Montgomery, J.A., Frisch, M.J., Ochterski, J.W., Petersson, G.A. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J. Chem. Phys.* 1999, 110(6), 2822–7.
49. Ochterski, J.W., Petersson, G.A., Montgomery, J.A. A complete basis set model chemistry. 5. Extensions to six or more heavy atoms. *J. Chem. Phys.* 1996, 104(7), 2598–619.
50. Montgomery, J.A., Ochterski, J.W., Petersson, G.A. A complete basis-set model chemistry. 4. An improved atomic pair natural orbital method. *J. Chem. Phys.* 1994, 101(7), 5900–09.
51. Bartmess, J.E. Negative Ion Energetics Data. <http://webbok.nist.gov> (accessed January 20, 2006).
52. Pokon, E.K., Liptak, M.D., Feldgus, S., Shields, G.C. Comparison of CBS-QB3, CBS-APNO, and G3 predictions of gas phase deprotonation data. *J. Phys. Chem. A* 2001, 105, 10483–7.
53. Pickard, F.C., Griffith, D.R., Ferrara, S.J., Liptak, M.D., Kirschner, K.N., Shields, G.C. CCSD(T), W1, and other model chemistry predictions for gas-phase deprotonation reactions. *Int. J. Quantum Chem.* 2006, 106(15), 3122–8.
54. Liptak, M.D., Shields, G.C. Comparison of density functional theory predictions of gas-phase deprotonation data. *Int. J. Quantum Chem.* 2005, 105(6), 580–7.
55. Adam, K.R. New density functional and atoms in molecules method of computing relative $pK(a)$ values in solution. *J. Phys. Chem. A* 2002, 106(49), 11963–72.
56. Kaminski, G.A. Accurate prediction of absolute acidity constants in water with a polarizable force field: Substituted phenols, methanol, and imidazole. *J. Phys. Chem. B* 2005, 109(12), 5884–90.
57. Purvis, G.D., Bartlett, R.J. A full coupled-cluster singles and doubles model—the inclusion of disconnected triples. *J. Chem. Phys.* 1982, 76(4), 1910–8.

58. Watts, J.D., Gauss, J., Bartlett, R.J. Coupled-cluster methods with noniterative triple excitations for restricted open-shell Hartree-Fock and other general single determinant reference functions—energies and analytical gradients. *J. Chem. Phys.* 1993, 98(11), 8718–33.
59. Lee, Y.S., Kucharski, S.A., Bartlett, R.J. A coupled cluster approach with triple excitations. *J. Chem. Phys.* 1984, 81(12), 5906–12.
60. Watts, J.D., Bartlett, R.J. The inclusion of connected triple excitations in the equation-of-motion coupled-cluster method. *J. Chem. Phys.* 1994, 101(4), 3073–78.
61. Dunning, T.H. Gaussian-basis sets for use in correlated molecular calculations. 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* 1989, 90(2), 1007–23.
62. Krishnan, R., Pople, J.A. Approximate 4th-order perturbation theory of electron correlation energy. *Int. J. Quantum Chem.* 1978, 14(1), 91–100.
63. Helgaker, T., Klopper, W., Koch, H., Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* 1997, 106(23), 9639–46.
64. Halkier, A., Helgaker, T., Jorgensen, P., Klopper, W., Koch, H., Olsen, J., Wilson, A.K. Basis-set convergence in correlated calculations on Ne, N-2, and H₂O. *Chem. Phys. Lett.* 1998, 286(3–4), 243–52.
65. Bak, K.L., Jorgensen, P., Olsen, J., Helgaker, T., Klopper, W. Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations. *J. Chem. Phys.* 2000, 112(21), 9229–42.
66. Curtiss, L.A., Raghavachari, K., Redfern, P.C., Rassolov, V., Pople, J.A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* 1998, 109(18), 7764–76.
67. Martin, J.M.L., de Oliveira, G. Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory. *J. Chem. Phys.* 1999, 111(5), 1843–56.
68. Pickard, F.C., Pokon, E.K., Liptak, M.D., Shields, G.C. Comparison of CBS-QB3, CBS-APNO, G2, and G3 thermochemical predictions with experiment for formation of ionic clusters of hydronium and hydroxide ions complexed with water. *J. Chem. Phys.* 2005, 122(2), 7.
69. Pickard, F.C., Dunn, M.E., Shields, G.C. Comparison of model chemistry and density functional theory thermochemical predictions with experiment for formation of ionic clusters of the ammonium cation complexed with water and ammonia; atmospheric implications. *J. Phys. Chem. A* 2005, 109(22), 4905–10.
70. Cunningham, A.J., Payzant, J.D., Kebarle, P. A kinetic study of the proton hydrate $H+(H_2O)_n$ equilibria in the gas phase. *J. Am. Chem. Soc.* 1972, 94(22), 7627–32.
71. Meot-Ner, M.(Mautner), Sieck, L.W. Relative acidities of water and methanol and the stabilities of the dimer anions. *J. Phys. Chem.* 1986, 90, 6687–90.
72. Kebarle, P. Gas phase ion thermochemistry based on ion-equilibria. From the ionosphere to the reactive centers of enzymes. *Int. J. Mass Spectrom.* 2000, 200, 313–30.
73. da Silva, G., Kennedy, E.M., Dlugogorski, B.Z. Ab initio procedure for aqueous-phase pK_a calculation: The acidity of nitrous acid. *J. Phys. Chem. A* 2006, 110(39), 11371–6.
74. Ervin, K.M., Ho, J., Lineberger, W.C. Ultraviolet photoelectron spectrum of NO₂[−]. *J. Phys. Chem.* 1988, 92(19), 5405–12.
75. Range, K., Riccardi, D., Cui, Q., Elstner, M., York, D.M. Benchmark calculations of proton affinities and gas-phase basicities of molecules important in the study of biological phosphoryl transfer. *Phys. Chem. Chem. Phys.* 2005, 7(16), 3070–9.
76. Bryantsev, V.S., Diallo, M.S., Goddard, W.A. pK(a) calculations of aliphatic amines, diamines, and aminoamides via density functional theory with a Poisson-Boltzmann continuum solvent model. *J. Phys. Chem. A* 2007, 111(20), 4422–30.
77. Klamt, A., Schuurmann, G. COSMO—A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans.* 1993, 2(5), 799–805.
78. Klamt, A., Jonas, V., Burger, T., Lohrenz, J.C.W. Refinement and parameterization of COSMO-RS. *J. Phys. Chem. A* 1998, 102(26), 5074–85.
79. Miertus, S., Scrocco, E., Tomasi, J. Electrostatic interaction of a solute with a continuum—A direct utilization of ab initio molecular potentials for the prevision of solvent effects. *Chem. Phys.* 1981, 55(1), 117–29.

80. Barone, V., Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* 1998, 102(11), 1995–2001.
81. Barone, V., Cossi, M., Tomasi, J. Geometry optimization of molecular structures in solution by the polarizable continuum model. *J. Comput. Chem.* 1998, 19(4), 404–17.
82. Cossi, M., Scalmani, G., Rega, N., Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* 2002, 117(1), 43–54.
83. Cramer, C.J., Truhlar, D.G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* 1999, 99, 2161–200.
84. Marenich, A.V., Olson, R.M., Kelly, C.P., Cramer, C.J., Truhlar, D.G. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. *J. Chem. Theory. Comput.* 2007, 3(6), 2011–33.
85. Marenich, A.V., Cramer, C.J., Truhlar, D.G. Universal solvation model based on the generalized born approximation with asymmetric descreening. *J. Chem. Theory. Comput.* 2009, 5(9), 2447–64.
86. Marenich, A.V., Cramer, C.J., Truhlar, D.G. Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies. *J. Phys. Chem. B* 2009, 113(14), 4538–43.
87. Zhan, C.G., Bentley, J., Chipman, D.M. Volume polarization in reaction field theory. *J. Chem. Phys.* 1998, 108(1), 177–92.
88. Zhan, C.G., Chipman, D.M. Cavity size in reaction field theory. *J. Chem. Phys.* 1998, 109(24), 10543–58.
89. Zhan, C.G., Chipman, D.M. Reaction field effects on nitrogen shielding. *J. Chem. Phys.* 1999, 110(3), 1611–22.
90. Liu, J., Kelley, C.P., Goren, A.C., Marenich, A.V., Cramer, C.J., Truhlar, D.G., Zhan, C.G. Free energies of solvation with surface, volume, and local electrostatic effects and atomic surface tensions to represent the first solvation shell. *J. Chem. Theory. Comput.* 2010, 6(4), 1109–1117.
91. Klamt, A., Eckert, F., Diedenhofen, M., Beck, M.E. First principles calculations of aqueous $pK(a)$ values for organic and inorganic acids using COSMO-RS reveal an inconsistency in the slope of the $pK(a)$ scale. *J. Phys. Chem. A* 2003, 107(44), 9380–6.
92. Eckert, F., Klamt, A. Accurate prediction of basicity in aqueous solution with COSMO-RS. *J. Comput. Chem.* 2006, 27(1), 11–9.
93. Jia, Z.K., Du, D.M., Zhou, Z.Y., Zhang, A.G., Hou, R.Y. Accurate $pK(a)$ determinations for some organic acids using an extended cluster method. *Chem. Phys. Lett.* 2007, 439(4–6), 374–80.
94. Shields, G.C., Kirschner, K.N. The limitations of certain density functionals in modeling neutral water clusters. *Synthesis Reactivity Inorg. Metal-Organic Nano-Metal Chem.* 2008, 38(1), 32–6.
95. Sadlej-Sosnowska, N. Calculation of acidic dissociation constants in water: Solvation free energy terms. Their accuracy and impact. *Theor. Chem. Acc.* 2007, 118(2), 281–93.
96. Foresman, J.B., Keith, T.A., Wiberg, K.B., Snoonian, J., Frisch, M.J. Solvent effects. 5. Influence of cavity shape, truncation of electrostatics, and electron correlation ab initio reaction field calculations. *J. Phys. Chem.* 1996, 100(40), 16098–104.
97. 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(30), 7241–7.
98. Andzelm, J., Kolmel, C., Klamt, A. Incorporation of solvent effects into density-functional calculations of molecular energies and geometries. *J. Chem. Phys.* 1995, 103(21), 9312–20.
99. Cossi, M., Rega, N., Scalmani, G., Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* 2003, 24(6), 669–81.
100. Yu, A., Liu, Y.H., Wang, Y.J. Ab initio calculations on $pK(a)$ values of benzo-quinuclidine series in aqueous solvent. *Chem. Phys. Lett.* 2007, 436(1–3), 276–9.
101. Chambers, C.C., Hawkins, G.D., Cramer, C.J., Truhlar, D.G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* 1996, 100(40), 16385–98.
102. Pople, J.A., Head-Gordon, M., Fox, D.J., Raghavachari, K., Curtiss, L.A. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.* 1989, 90(10), 5622–29.
103. Curtiss, L.A., Raghavachari, K., Trucks, G.W., Pople, J.A. Gaussian-2 theory for molecular-energies of 1st-row and 2nd-row compounds. *J. Chem. Phys.* 1991, 94(11), 7221–30.