

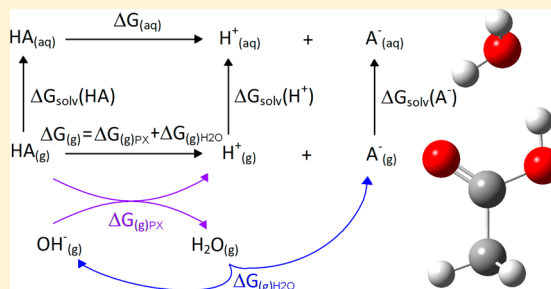
First Principles pK_a Calculations on Carboxylic Acids Using the SMD Solvation Model: Effect of Thermodynamic Cycle, Model Chemistry, and Explicit Solvent Molecules

Catherine C. R. Sutton, George V. Franks, and Gabriel da Silva*

Department of Chemical and Biomolecular Engineering, The University of Melbourne, Victoria 3010, Australia

S Supporting Information

ABSTRACT: Aqueous pK_a values are calculated from first principles for a set of carboxylic acids using the SMD solvation model with various model chemistries, thermodynamic cycles, and treatments of explicit solvation. In all, 108 unique theoretical protocols are examined. The direct (D) and water proton exchange (PX) cycles are trialed along with a new approach, termed the semidirect (SD) cycle. The SD thermodynamic cycle offers some improvements over the D and PX schemes, as it bypasses the gas-phase heterolytic bond dissociation calculation required in the conventional D approach while also avoiding an aqueous OH^- calculation required by the PX method when using water as the reference acid. With all three cycles, the recommended model chemistry employs M05-2X/cc-pVTZ Gibbs energies of solvation with a single discrete water molecule and a high-level composite method for the gas-phase reaction energies. With the SD cycle, these calculations result in a mean unsigned error of less than 1 pK_a units, with respective mean signed error and maximum unsigned error of less than 0.5 and 2 pK_a units. Similar results are obtained with the D and PX cycles, and further improvement is required in both the gas and aqueous phase *ab initio* energy calculations before we can truly discriminate between the thermodynamic cycles investigated here.



1. INTRODUCTION

Electronic structure theory in conjunction with continuum solvation models and thermodynamic cycles has been extensively used to theoretically determine pK_a values. For organic acids, recent works have converged on a mean unsigned error (MUE) of around 2 pK_a units as the state-of-the-art accuracy,^{1–26} although there is no one procedure for these calculations that is agreed upon in the literature. Part of the reason for this lack of consensus is the continual improvement both in theoretical methods and in computational hardware. Beyond this, there are varied approaches taken by different authors without a comprehensive comparison. A range of thermodynamic models have been used in the literature, with the direct cycle^{3,6–10,12,13,15,21,22} and the proton exchange cycle^{3,7,13,17–19} being widely used. Also, while it can be beneficial to include explicit solvent molecules in combination with the continuum solvent model, this too is not implemented in a uniform manner.^{6,7,14,18}

We aim to evaluate the SMD²⁷ solvent model for pK_a calculations on carboxylic acids, testing a wide range of model chemistries, thermodynamic cycles, and explicit solvent representations. The following section describes the different calculation approaches adopted here, followed by a description of the theoretical methods and then the results and discussion.

2. pK_a CALCULATIONS

The dissociation of an acid, HA, can be represented as follows:



$$K_a = \frac{[A^-][H^+]}{[HA]} \quad (2)$$

The acidity constant, K_a , is a measure of the extent of this dissociation. K_a is usually reported in the form pK_a :

$$pK_a = -\log_{10} K_a \quad (3)$$

The pK_a can be calculated from the Gibbs energy in solution of the acid dissociation reaction.³

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

Here $\Delta G_{(aq)}$ is the Gibbs energy of reaction for eq 1, and this is the property that we are interested in obtaining from first principles so as to determine pK_a .

2.1. Thermodynamic Cycles. It is possible to arrive at $\Delta G_{(aq)}$ for the dissociation of an acid through a thermodynamic cycle, so as to avoid performing a calculation on the solvated proton. A number of different thermodynamic cycles have been employed in the literature, all of which determine $\Delta G_{(aq)}$ from a combination of gas-phase Gibbs energies of reaction and solvation Gibbs energies, obtained from theory or experiment. A gas-phase Gibbs energy of reaction can be calculated using

Received: June 15, 2012

Revised: August 23, 2012

Published: August 25, 2012

well-established theoretical protocols, with a correction of $RT \ln(24.46)$ ($1.894 \text{ kcal mol}^{-1}$ at 298 K), placing it in the 1 mol L^{-1} reference state. The Gibbs energy of solvation (ΔG_{solv}) for some solute molecule can be readily determined using continuum solvation models such as SMD that have been parametrized using experimental solvation energies. The solvation Gibbs energy in SMD is calculated as²⁸

$$\Delta G_{\text{solv}} = (E_{\text{soln}} + G_{\text{nes}}) - E_{\text{gas}} \quad (5)$$

where E_{soln} and E_{gas} are the respective electronic energies of the solute in the presence and absence of the solvent field and G_{nes} is the sum of non-electrostatic contributions to the solvation Gibbs energy such as cavitation and dispersion–repulsion interactions. The output of an SMD solvent-phase calculation encompasses $E_{\text{soln}} + G_{\text{nes}}$ such that ΔG_{solv} is defined at the reference state of 1 mol L^{-1} (standard liquid reference state).

2.1.1. Direct (D). The prototypical thermodynamic cycle for $\text{p}K_{\text{a}}$ calculations is the direct (D) cycle, illustrated in Figure 1.³

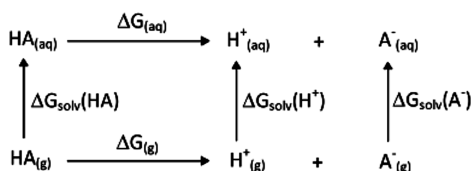


Figure 1. Direct (D) thermodynamic cycle.

According to Figure 1, $\Delta G_{\text{(aq)}}$ is calculated as

$$\Delta G_{\text{(aq)}} = \Delta G_{\text{(g)}} + \Delta G_{\text{(solv)}}(\text{A}^-) + \Delta G_{\text{(solv)}}(\text{H}^+) - \Delta G_{\text{(solv)}}(\text{HA}) \quad (6)$$

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

The value of $\Delta G_{\text{(solv)}}(\text{H}^+)$ was under debate in the literature for some time, but the value now generally accepted is $-265.9 \text{ kcal mol}^{-1}$.^{3,4,7,8,20}

Ho and Coote (2010) conducted an extensive review and benchmarking study of thermodynamic cycles to calculate $\text{p}K_{\text{a}}$.⁷ They reported that, while the direct method is the most commonly implemented, it is less accurate than the proton exchange method for most $\text{p}K_{\text{a}}$ calculations. The exception to this is for the cases of phenols and aliphatic carboxylic acids, where they describe the results as moderately accurate, the felicity of which they attribute to the systematic cancellation of errors. Note that some studies also use a thermodynamic cycle that has the proton represented as the hydronium ion H_3O^+ , balancing this with a water molecule on the other side of the reaction.³ It has been seen in previous studies that this cycle is either equivalent or inferior to the direct thermodynamic cycle and so this procedure is not included in this work.^{7,11,19}

2.1.2. Proton Exchange (PX). Another thermodynamic cycle is the proton exchange method, shown in Figure 2, which avoids having any explicit protons by the inclusion of a reference acid,

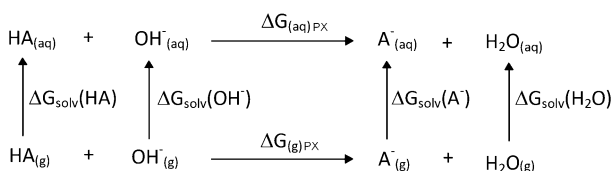


Figure 2. Proton exchange (PX) thermodynamic cycle.

commonly water. The proton exchange cycle is considered more reliable because the number of charged species is conserved on both sides of the equation, which allows for some cancellation of errors.⁷

In this case, the $\text{p}K_{\text{a}}$ is obtained by eq 8:

$$\text{p}K_{\text{a}} = \frac{\Delta G_{\text{(aq)PX}}}{\ln 10RT} + \text{p}K_{\text{a}}(\text{H}_2\text{O}) \quad (8)$$

$$\begin{aligned} \Delta G_{\text{(aq)PX}} &= \Delta G_{\text{(g)}} + \Delta G_{\text{(solv)}}(\text{A}^-) + \Delta G_{\text{(solv)}}(\text{H}_2\text{O}) \\ &\quad - \Delta G_{\text{(solv)}}(\text{HA}) - \Delta G_{\text{(solv)}}(\text{OH}^-) \end{aligned} \quad (9)$$

$$\Delta G_{\text{(g)PX}} = G_{\text{(g)}}(\text{A}^-) + G_{\text{(g)}}(\text{H}_2\text{O}) - G_{\text{(g)}}(\text{HA}) - G_{\text{(g)}}(\text{OH}^-) \quad (10)$$

The reference acid requires an accurate $\text{p}K_{\text{a}}$ of its own and should be structurally similar to the one for which the $\text{p}K_{\text{a}}$ is being calculated, to maximize cancellation of errors.⁷ Many studies have used water as the reference acid and we have also adopted this approach.^{3,5,7} When using water as a reference acid as well as the solvent, a correction must be made from the standard state of liquid water, 55.34 mol L^{-1} .⁷ Others have incorporated this correction by using a 1 mol L^{-1} reference state $\text{p}K_{\text{a}}$ of 15.74 for water. We have used a $\text{p}K_{\text{a}}$ of 14 and made a concentration adjustment of $RT \log[\text{H}_2\text{O}] = 2.38 \text{ kcal mol}^{-1}$ to the solvation energy of the water molecule.²⁹ In this way, reference states are consistent when explicit water molecules are also included in the $\text{p}K_{\text{a}}$ calculations.

While the elimination of the proton in the PX cycle might suggest that the use of a thermodynamic cycle is obsolete, and that the solution phase Gibbs energies are all that are required, this is not borne out in the literature where the PX cycle is indeed used. As the solvent models have been parametrized on Gibbs energies of solvation, not solution phase reaction energies, more accurate results can be obtained when they are used to calculate the energy of solvation, in association with a gas basicity calculation performed at a higher level of theory.⁷

2.1.3. Semi-Direct (SD). In this work, the semi-direct thermodynamic cycle is introduced. Here $\Delta G_{\text{(g)PX}}$ is calculated theoretically and the cycle in Figure 3 is used in combination with

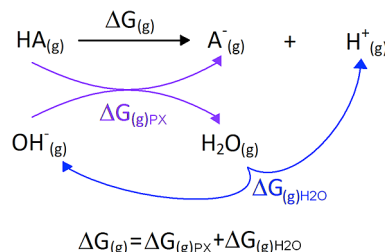


Figure 3. Calculation of $\Delta G_{\text{(g)}}$ in the semi-direct (SD) thermodynamic cycle.

$\Delta G_{\text{H}_2\text{O}}$ to arrive at $\Delta G_{\text{(g)}}$. The experimental Gibbs energy of the dissociation for water ($\Delta G_{\text{(g)H}_2\text{O}}$) is precisely known, the most recent determinations being 383.61^{30} and $383.74 \text{ kcal mol}^{-1}$.^{31,32} Taking the former and accounting for a change of reference state, a value of $385.50 \text{ kcal mol}^{-1}$ for the dissociation of water is arrived at for use in the current calculations. The benefit of the SD cycle is that while still avoiding the calculation of a gas-phase heterolytic bond dissociation energy (as per PX) it also negates

the need to calculate the solvation Gibbs energy of a reference species.

Then, $\Delta G_{\text{(aq)}}$ is calculated as

$$\Delta G_{\text{(aq)}} = \Delta G_{\text{(g)PX}} + \Delta G_{\text{(g)H}_2\text{O}} + \Delta G_{\text{(solv)}}(\text{A}^-) + \Delta G_{\text{(solv)}}(\text{H}^+) - \Delta G_{\text{(solv)}}(\text{HA}) \quad (11)$$

2.2. Solvent Representation. All calculations in the gas phase are for an ideal gas. When the species of interest resides in a solvent, consideration must be given to how the intermolecular interactions are represented. One approach to handle this is to explicitly include a great many individual solvent atoms in the calculation, such that they interact with the solute, both affecting it and being affected themselves. This approach can be implemented in molecular mechanics calculations, but these are limited to Newtonian interactions and do not sufficiently account for electronic effects. The scale of such an approach with an *ab initio* model has deterred its pursuit due to the computational intensity. Instead, implicit solvent models have been developed and used to great success, where the solvent is represented as a polarizable dielectric continuum. Within the continuum, a cavity is defined, which the solute inhabits, the shape of which depends on how it is defined by the particular solvent model.^{27,33} Solvent models are continually being improved upon, and new versions and models frequently appear. In this work, Marenich, Cramer, and Truhlar's solvent model SMD is used.²⁷ It has been widely used and is particularly suited to calculations where nonelectrostatic interactions are important.^{21,22,34}

As continuum methods are not able to explicitly account for hydrogen-bonding, discrete solvent molecules are sometimes included to improve the modeling of solute–solvent interactions.⁷ The discrete solvent molecules and the solute form a complex which is then encased in the cavity in the solvent field. The inclusion of explicit solvent molecules can improve the accuracy of calculations but requires more computational time.³⁵ Furthermore, there is no clear way to determine the best number of solvent molecules to include.^{5–7} The effect of the number of solvent molecules is beyond the scope of this paper; this work considers the inclusion of one explicit water molecule. In the literature, there are differing views as to how explicit solvent molecules should be included in calculations. In this work, the three different explicit solvent treatments described below are considered and compared.

2.2.1. Explicit H₂O Treatment. The first situation considered (NO) represents the solvent by the implicit SMD model but does not actually use any explicit water molecules. The calculations are as above in the discussion on the thermodynamic cycles. This is calculated as a benchmark.

The second scenario (ALL) is where an explicit water molecule is included as a complex with all neutral and ionic species, both in the gas phase and aqueous phase. ALL is illustrated for the direct thermodynamic cycle in Figure 4.

Third, the explicit solvent molecule is included as a complex with the aqueous acid and anion but not the gas phase species (AQ), as shown in Figure 5. This method reduces the computational cost vs ALL but still incorporates some treatment of discrete solvent molecule effects.

With this solvent description, the solvation Gibbs energy terms are calculated as

$$\Delta G_{\text{solv}} = (E_{\text{soln}} + G_{\text{nes}})(\text{X} \cdot \text{H}_2\text{O}) - E_{\text{gas}}(\text{X}) - E_{\text{gas}}(\text{H}_2\text{O}) \quad (12)$$

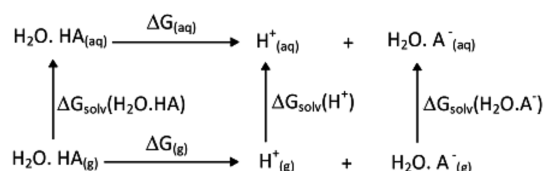


Figure 4. Explicit solvent molecules on all acid and anion species, in the direct thermodynamic cycle.

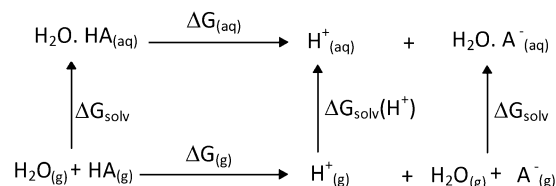


Figure 5. Explicit solvent molecules on aqueous acid and anion species, in the direct thermodynamic cycle.

where X is the solute. Note that this term now differs from the solvation Gibbs energy of X, as it also incorporates a complexation energy.

The fourth option (AN) is to only have water complexed with the anion (in solution and in gas) but not the acid.^{5,7} This is illustrated as in Figure 6.

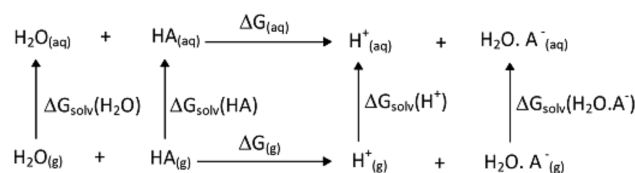


Figure 6. Explicit solvent molecules on the anion, in the direct thermodynamic cycle.

3. COMPUTATIONAL METHODS

All calculations were performed in Gaussian 09.³⁴ The energies reported are always for the optimized lowest-energy conformer, and frequency calculations were performed on all structures to confirm the absence of imaginary vibrational modes. Note that structures were optimized independently in both gas and aqueous phases, and although no reordering of conformers was observed, some species did demonstrate significant structural changes for their water complexes (for example, see the acetic acid and acetate structures provided in Figure 7). The final

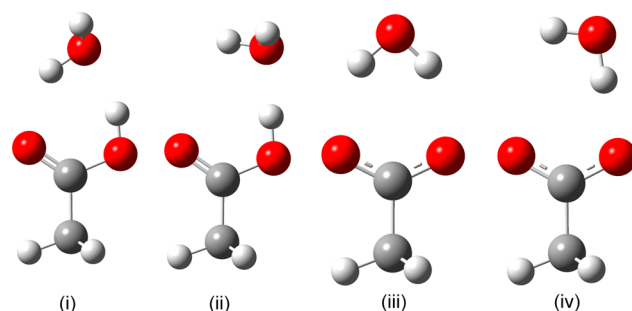


Figure 7. Optimized geometries at the M05-2X/cc-pVTZ level of theory: (i) acetic acid with one water in vacuum; (ii) acetic acid with one water in SMD; (iii) acetate with one water in vacuum; (iv) acetate with one water in SMD.

model chemistry used to determine pK_a values arises from a combination of *ab initio* gas-phase and solvation Gibbs energies, as described below. We refer to the complete model chemistry using the notation “gas/solvent”.

Gas-phase reaction Gibbs energy calculations make use of the composite G3SX method as well as the reduced-order G3SX(MP3) variant. These methods are based on B3LYP/6-31G(2df,p) optimized structures and scaled zero point energies, along with a series of higher-level wave function theory single-point energy calculations from HF through QCISD(T) theory that are combined with scaling corrections to arrive at the final electronic energy. The entropy component of the Gibbs energy is determined by applying the rigid-rotor harmonic-oscillator approximation to the internal and external degrees of freedom described by the B3LYP vibrational frequencies and moments of inertia. The G3SX and G3SX(MP3) calculations are expected to provide energies accurate to below 1 kcal mol⁻¹,^{7,36} on average.

Solvation Gibbs energies (ΔG_{solv}) are obtained using SMD, which is the default model for solvation Gibbs energies implemented in Gaussian 09. Calculations were performed using the M05-2X density functional with the 6-31G(d) (i.e., 6-31G*), 6-31+G(d,p) (6-31+G**), and cc-pVTZ basis sets; these represent the highest-level method/basis set combinations used to parametrize SMD. Two additional methods were also investigated: (i) B3LYP/6-31G(2df,p) (for which gas-phase energies are already available from the G3SX calculations) and (ii) M05-2X/aug-cc-pVTZ (to further explore the effect of diffuse functions). With SMD, the 298 K solvation Gibbs energy is defined as the difference between the solvent and gas *electronic* energies (0 K, no ZPE), necessitating a gas-phase calculation at each level of theory. A recent contribution has shown that using a gas-phase treatment of entropy for solution-phase species is an accurate approximation.³⁷ Using SMD, solvation Gibbs energies should be accurate to around 1 kcal mol⁻¹ for neutral solutes and 4 kcal mol⁻¹ for ionic species, at an appropriate level of theory.⁷

In order to enable the reproduction of our calculations, a spreadsheet is provided as Supporting Information which illustrates how the pK_a of acetic acid is calculated with the various thermodynamic cycles and solvent descriptions using the G3SXISMD/M05-2X/cc-pVTZ model chemistry.

4. RESULTS AND DISCUSSION

We have performed pK_a calculations using three thermodynamic cycles, four different treatments of explicit solvation (or lack thereof) and nine model chemistries. This results in 108 unique computational protocols. Because our aim is to explore a large number of different protocols, we have restricted ourselves to a test set of eight representative carboxylic acids for which accurate experimental pK_a values are available. These acids, along with their pK_a values, are listed in Table 1. This set of acids encompasses the six acids used by Liptak and Shields (2001)³ with the addition of propanoic and hexanoic acids.³⁸ All 864 calculated pK_a values are listed as Supporting Information. The error in pK_a , defined as the difference between the experimental value and the calculated value (i.e., negative when the calculated value is smaller than experiment), has then been obtained for each data point. With each pK_a protocol, the mean signed error (MSE), mean unsigned error (MUE), maximum unsigned error (MaxUE), and standard deviation in signed error (SDSE) across the test set have been obtained, with the values reported in Tables 2–5 for G3SX theory (the corresponding

Table 1. Experimental pK_a Values of Acids in the Test Set^{3,38}

acid	pK_a
acetic	4.75
chloroacetic	2.85
cyanoacetic	2.45
formic	3.75
hexanoic	4.88
oxalic	1.23
pivalic	5.03
propanoic	4.87

Table 2. Mean Signed Error in Computed pK_a Values^a

		mean signed error				
G_{gas}		G3SX				
ΔG_{solv}		B3LYP/6-31G(2df,p)	M05-2X/6-31G(d)	M05-2X/6-31+G(d,p)	M05-2X/cc-pVTZ	M05-2X/aug-cc-pVTZ
NO	D	1.81	0.08	1.63	0.75	1.83
	PX	1.76	−0.12	−5.88	−0.16	−6.32
	SD	1.65	−0.08	1.47	0.60	1.67
ALL	D	0.36	0.10	1.52	0.27	1.07
	PX	0.31	−0.10	−5.99	−0.65	−7.08
	SD	0.20	−0.06	1.36	0.11	0.91
AQ	D	2.06	0.11	1.97	0.76	1.89
	PX	2.01	−0.08	−5.54	−0.16	−6.26
	SD	1.90	−0.04	1.81	0.60	1.73
AN	D	2.02	2.62	4.06	2.34	3.03
	PX	1.97	2.43	−3.44	1.43	−5.11
	SD	1.86	2.46	3.91	2.18	2.87

^aNO = no explicit solvent; ALL = an explicit water on all species; AQ = an explicit water on aqueous species; AN = an explicit water on anions; D = direct thermodynamic cycle; PX = proton exchange thermodynamic cycle; SD = semi-direct thermodynamic cycle. The bold-italic values are those in the family of procedures that this work recommends for the determination of pK_a .

G3SX(MP3) results are in the Supporting Information). Note that the SDSE is independent of the thermodynamic cycle used within a given model chemistry. There are few major differences in the MSEs and SDSEs, typically coming in at around 1 pK_a unit or better, testifying to the precision of the pK_a protocols (excluding the AN solvent description and the PX cycle with diffuse functions on the ΔG_{solv} basis set, as addressed below). As such, the following discussion focuses on the MUE and MaxUE. The best choices of model chemistry, solvent description, and thermodynamic cycle when performing pK_a calculations on carboxylic acids using SMD are highlighted below, and our recommendations are noted in bold-italic typeface in Tables 2–5.

First, we consider the effect that the gas-phase reaction Gibbs energy has on the pK_a calculations. Note that all three thermodynamic cycles employ a gas-phase reaction Gibbs energy; for the D cycle, this is a heterolytic O–H bond dissociation forming an anion (A^-) and a cation (H^+), but for the PX and SD cycles, it is a proton exchange reaction that balances charge. Because of the absence of ion solvation *in vacuo*, the heterolytic bond dissociation energies are large (ca. 1400 kJ mol⁻¹), whereas the isodesmic proton exchange reactions proceed with much smaller reaction energies (ca. 200 kJ mol⁻¹), which merely reflect the *difference* in the heterolytic O–H bond dissociation energies between the target (HA) and reference (H_2O) acids. As a consequence, the PX and SD cycles offer significant cancellation of error in the gas-phase component,

Table 3. Mean Unsigned Error in Computed pK_a Values^a

		mean unsigned error					
G_{gas}		G3SX					
ΔG_{solv}		B3LYP/6-31G(2df,p)	M05-2X/6-31G(d)	M05-2X/6-31+G(d,p)	M05-2X/cc-pVTZ	M05-2X/aug-cc-pVTZ	average
NO	D	2.58	1.34	1.88	1.34	2.00	1.79
	PX	2.55	1.29	5.88	1.01	6.32	2.68
	SD	2.47	1.30	1.76	1.23	1.89	1.69
ALL	D	1.00	0.93	1.52	0.68	1.10	1.03
	PX	0.98	0.93	5.99	0.70	7.08	2.15
	SD	0.96	0.93	1.36	0.64	0.98	0.97
AQ	D	2.40	1.02	1.97	0.97	1.89	1.59
	PX	2.37	0.97	5.54	0.60	6.26	2.37
	SD	2.29	0.98	1.81	0.87	1.73	1.49
AN	D	2.35	2.83	4.06	2.39	3.03	2.91
	PX	2.31	2.69	3.44	1.71	5.11	2.54
	SD	2.23	2.72	3.91	2.27	2.87	2.78
average		2.04	1.50	3.26	1.20	3.35	

^aNO = no explicit solvent; ALL = an explicit water on all species; AQ = an explicit water on aqueous species; AN = an explicit water on anions; D = direct thermodynamic cycle; PX = proton exchange thermodynamic cycle; SD = semi-direct thermodynamic cycle. The bold-italic values are those in the family of procedures that this work recommends for the determination of pK_a .

relative to the direct cycle. This is borne out in the results of additional pK_a calculations reported as Supporting Information, in which the gas-phase G3SX/G3SX(MP3) Gibbs energy of reaction is replaced by a value determined at the same level of theory as that of the SMD solvation Gibbs energy. At the M05-2X/6-31+G(d,p) level of theory, the SD cycle returns a MUE of 0.77 pK_a units averaged across the four different solvent treatments, whereas the D cycle yields a MUE of 1.13 pK_a units (the SD cycle also performs better than D for all four individual cases).³⁹ Using the larger aug-cc-pVTZ basis set, this effect is less pronounced, as would be expected, with respective MUEs of 0.87 and 0.94 pK_a units. It is necessary, however, to include diffuse functions on the basis set used for the gas-phase calculations with the SD cycle in order to adequately describe the OH^- anion (which does not appear in the D cycle).

Interestingly, when comparing between the more accurate G3SX and G3SX(MP3) methods, the D cycle is relatively insensitive to the choice of method but the PX and SD pK_a values change by around 0.5–1.0 pK_a units. However, this difference in pK_a typically corresponds to an improvement when moving from reduced-order G3SX(MP3) theory to the higher-level G3SX method. For example, the G3SX pK_a values have a MUE about 0.8 pK_a units below those obtained with G3SX(MP3) theory when using the accurate M05-2X/cc-pVTZ solvation Gibbs energies. Any number of composite protocols (Gn, CBS), or wave function theory methods such as CCSD(T) with at least a good triple- ζ quality basis set, should be reliable, although testing them is beyond the scope of this work. We do however identify that reduced perturbation order variants of the Gn methods may be appropriate when dealing with large molecules, as they can reduce the computational time required in the expensive gas-phase calculations by almost an order of magnitude. It is also noted that quite reasonable results are obtained with the SD cycle (but not PX) when a lower level of theory is used for the gas-phase energy calculations as well as the solvation Gibbs energy, provided diffuse functions are present, such as is the case for M05-2X/6-31+G(d,p).

The choice of theoretical method in the ΔG_{solv} calculations also significantly impacts the accuracy of the pK_a calculations. Using G3SX theory for the gas-phase reaction Gibbs energies,

the best methods in terms of MUE, when averaged across the 12 combinations of thermodynamic cycles and solvent descriptions, are M05-2X/cc-pVTZ (1.20) and M05-2X/6-31G(d) (1.50), as per columns 2 and 4 of Table 3. The use of diffuse functions with either basis set significantly degrades the average MUE (to 3.26 and 3.35, in columns 3 and 5 of Table 3), which can be attributed predominantly to results obtained with the PX thermodynamic cycle. This problem has been traced to the Gibbs energy of solvation of OH^- , which is not required by the D and SD cycles.⁴⁰ When the PX results are excluded, neither the aug-cc-pVTZ nor 6-31+G(d) basis set offers any improvement over their nonaugmented counterpart (average MUEs of 1.93 and 2.28), providing similar results to B3LYP/6-31G(2df,p) (MUE of 2.04, column 1 in Table 3). From our results, we recommend using M05-2X/cc-pVTZ theory for SMD solvation Gibbs energies, although M05-2X/6-31G(d) is a good cost-effective alternative. Given the need for high-level gas-phase reaction Gibbs energies, though, the computational time required for the M05-2X/cc-pVTZ solvation Gibbs energies is unlikely to be prohibitive.

We now turn our attention to the choice of explicit solvent representation, i.e., no explicit water molecules (NO) or one explicit water on all species (ALL), on the aqueous species (AQ), or on the anions (AN). Figure 8 illustrates the effect of the different solvent representations on the pK_a of each acid, for the case where the direct thermodynamic cycles is used with G3SXISMD/M05-2X/cc-pVTZ. Inclusion of a discrete water molecule does appear to offer a small improvement (vs NO) in pK_a , consistent with previous findings. This improvement is clear when one explicit solvent molecule is introduced either on all species (ALL) or only on the aqueous species (AQ) but not as much so when the water is only on the anion (AN). Using the above recommendation of a G3SXISMD/M05-2X/cc-pVTZ model chemistry, the MUE averaged across the three thermodynamic cycles is 1.2 for NO, around 0.4 pK_a units worse than the ALL and AQ approaches (see column 5 of Table 3). There are similar differences in the MaxUE (Table 4). Furthermore, ALL and AQ have lower standard deviation in signed error than NO and AN (rows 2 and 3 compared with 1 and 4 of Table 5). It is difficult to discriminate between the

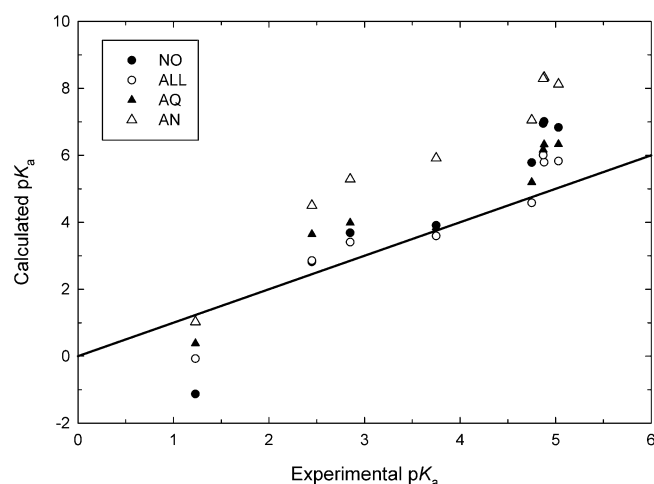


Figure 8. pK_a calculated with G3SX for G_{gas} , M05-2X/cc-pVTZ for ΔG_{solv} and direct thermodynamic cycle, in pK_a units. The solid line represents $y = x$.

ALL and AQ approaches, with the inclusion of one explicit water molecule on all species vs only those in solution offering a negligible improvement in pK_a of around 0.1. The AQ method does, however, offer the practical benefit of removing an explicit water molecule from the computationally intensive gas-phase calculations.

Finally, we need to consider the selection of thermodynamic cycle, from the direct (D), water proton exchange (PX), and semidirect (SD) cycles implemented here. With our recommended G3SX/SMO5-2X/cc-pVTZ model chemistry and inclusion of an explicit water (either as ALL or AQ), there is little difference between the three options, with MUEs of around 0.6–1.0 pK_a units and MaxUEs of better than 2.3 (see the bold-italic entries in column 5 of Tables 3 and 4). The best cycle is either SD (ALL) or PX (AQ) depending on the choice of explicit solvent treatment. To further explore differences within these six theoretical protocols, we have plotted calculated pK_a values for the eight individual acids against

their known experimental values in Figure 9. We note that the pK_a of the most acidic acid (oxalic acid, $pK_a = 1.23$) is almost always under-predicted and in most cases is responsible for the MaxUE. Oxalic acid is the only diprotic acid considered here, and we suspect that it may require an additional explicit water molecule to effectively describe hydrogen bonding at the protonated carboxyl group. It is known that diprotic acids present challenges for pK_a calculations.²⁹ The present results make it difficult to discriminate between the three thermodynamic cycles, although PX and SD may offer some slight improvement over D. Generalizing, the PX and SD cycles are expected to offer better cancellation of errors than D in the gas-phase reaction Gibbs energy. The SD cycle introduced here is useful in that it avoids a sometimes problematic solvation Gibbs energy calculation on a second anion (as seen here for OH^-), although the corollary of this is that the PX cycle could benefit from cancellation of error in the ion solvation Gibbs energies.

Table 5. Standard Deviation in Signed Error (SDSE) in Computed pK_a Values Using Various Model Chemistries and Explicit Solvent Descriptions^a

standard deviation in signed error						
G _{gas}	G3SX					
ΔG _{solv}	B3LYP/6-31G(2df,p)	M05-2X/6-31G(d)	M05-2X/6-31+G(d,p)	M05-2X/cc-pVTZ	M05-2X/AUG-cc-pVTZ	average
NO	2.60	1.92	1.19	1.46	1.15	1.79
ALL	1.26	1.29	0.66	0.79	0.57	1.00
AQ	1.95	1.30	0.50	0.80	0.48	1.14
AN	1.61	1.59	1.06	1.17	0.95	1.36
average	1.86	1.52	0.85	1.06	0.79	

^aNO = no explicit solvent; ALL = an explicit water on all species; AQ = an explicit water on aqueous species; AN = an explicit water on anions; D = direct thermodynamic cycle; PX = proton exchange thermodynamic cycle; SD = semi-direct thermodynamic cycle. The bold-italic values are those in the family of procedures that this work recommends for the determination of pK_a .

Table 4. Max Unsigned Error in Computed pK_a Values^a

G_{gas}	max unsigned error					
	G3SX					average
	ΔG_{solv}	B3LYP/6-31G(2df,p)	M05-2X/6-31G(d)	M05-2X/6-31+G(d,p)	M05-2X/cc-pVTZ	
NO	D	5.06	4.19	2.79	2.36	3.60
	PX	5.01	4.38	8.50	3.28	5.29
	SD	4.90	4.34	2.63	2.52	3.60
ALL	D	2.14	2.64	2.34	1.31	2.10
	PX	2.19	2.83	7.37	2.22	3.65
	SD	2.30	2.79	2.18	1.46	2.18
AQ	D	4.73	2.58	2.42	1.44	2.79
	PX	4.68	2.78	6.66	1.76	3.97
	SD	4.57	2.74	2.26	1.29	2.71
AN	D	3.81	4.06	5.07	3.45	4.10
	PX	3.77	3.87	5.79	2.54	3.99
	SD	3.65	3.90	4.91	3.29	3.94
average		3.90	3.43	4.41	2.24	4.45

^aNO = no explicit solvent; ALL = an explicit water on all species; AQ = an explicit water on aqueous species; AN = an explicit water on anions; D = direct thermodynamic cycle; PX = proton exchange thermodynamic cycle; SD = semi-direct thermodynamic cycle. The bold-italic values are those in the family of procedures that this work recommends for the determination of pK_a .

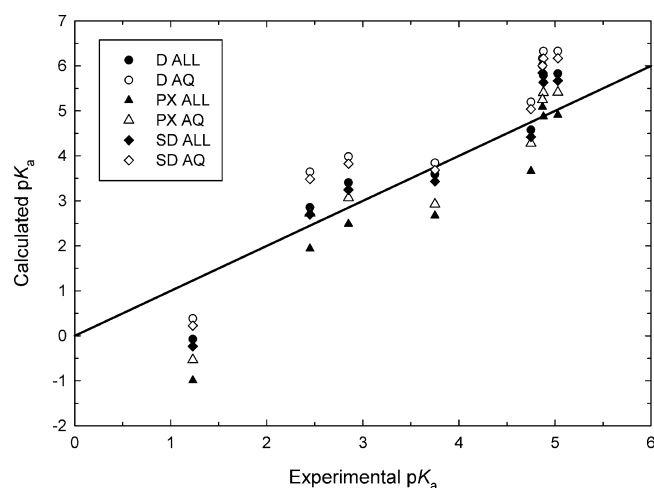


Figure 9. Comparison of calculated and experimental pK_a values for the carboxylic acid test set. Calculated values use the G3SXIM05-2X/cc-pVTZ model chemistry and different thermodynamic cycles (D, PX, SD) and explicit solvent descriptions (ALL, AQ). The solid line represents $y = x$.

The PX cycle also offers improved flexibility when used with reference acids other than H_2O . Ultimately, significant improvements are required in both the gas-phase reaction Gibbs energy and solvation Gibbs energy components of the pK_a calculations before the choice of thermodynamic cycle is the limiting factor for accuracy, and pragmatic criteria should guide their selection.

The procedure that we have now recommended is of comparable accuracy to other works that use a range of solvent models, including CPCM, CPCM-UAQS, CPCM-UAHF, SM6, IPCM, and COSMO-RS.^{3,7,22,23} Most importantly, different options for performing the calculations have been explored, and a number of theoretical protocols that do not result in accurate pK_a predictions have been identified. The SMD solvent model, which is the recommended choice for computing ΔG of solvation in Gaussian 09, has however been shown to perform well for organic acid pK_a values when using well-chosen theoretical methods.

5. CONCLUSIONS

The effect of thermodynamic cycle, model chemistry, and inclusion of an explicit water molecule on the calculation of organic acid pK_a values has been conducted. The most accurate calculations occur when a high-level composite method (G3SX) is used for gas phase calculations and M05-2X/cc-pVTZ is used for solvation calculations. Inclusion of an explicit water molecule on the aqueous acid and anion leads to more accurate pK_a values than the continuum solvent model alone. Inclusion of an explicit water molecule on all acid and anion species also gives this improvement, but no more so than when only on aqueous species. While all three thermodynamic cycles generally give reasonable results, the semidirect method avoids problems that can manifest in the direct and proton exchange thermodynamic cycles.

Therefore, the recommended procedure for calculating the pK_a of organic acids with the SMD solvent model is to employ the semidirect thermodynamic cycle, to include a single explicit water molecule on aqueous species, to apply G3SX for calculating gas basicity and M05-2X/cc-pVTZ for calculating the solvation energy. In this work, the above procedure has resulted in a mean signed error of less than 0.5 pK_a units, a mean unsigned

error of less than 1 pK_a unit, a maximum unsigned error of 2 pK_a units, and a standard deviation of error of 0.8 pK_a units.

■ ASSOCIATED CONTENT

Supporting Information

The mean signed error, mean unsigned error, max unsigned error, and standard deviation in signed error for all model chemistries (as well as the case that the gas basicity is calculated using the same model chemistry as the solvation energies), thermodynamic cycles, and representations of the solvent. The pK_a values for each acid for all model chemistries, thermodynamic cycles, and representations of the solvent. Also provided is a sample calculation of pK_a for acetic acid with the G3SXIM05-2X/cc-pVTZ model chemistry. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +61 3 8344 6627. Fax: +61 3 8433 4153. E-mail: gdsilva@unimelb.edu.au.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is supported by the Australian Research Council through Discovery Project DP098570 (GVF). C.C.R.S. is the recipient of an APA scholarship.

■ REFERENCES

- Piiego, J. R. *Chem. Phys. Lett.* **2003**, 367, 145–149.
- da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *Chem. Phys. Lett.* **2003**, 381, 244–245.
- Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, 123, 7314–7319.
- Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, 373, 411–415.
- Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2006**, 110, 2493–2499.
- Wang, X. X.; Fu, H.; Du, D. M.; Zhou, Z. Y.; Zhang, A. G.; Su, C. F.; Ma, K. S. *Chem. Phys. Lett.* **2008**, 460, 339–342.
- Ho, J.; Coote, M. L. *Theor. Chem. Acc.* **2010**, 125, 3–21.
- Borstnar, R.; Choudhury, A. R.; Stare, J.; Novic, M.; Mavri, J. *THEOCHEM* **2010**, 947, 76–82.
- da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, 103, 11194–11199.
- Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, 104, 2402–2409.
- Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, 124, 6421–6427.
- Piiego, J. R.; Riveros, J. M. *J. Phys. Chem. A* **2002**, 106, 7434–7439.
- De Abreu, H. A.; De Almeida, W. B.; Duarte, H. A. *Chem. Phys. Lett.* **2004**, 383, 47–52.
- Fu, Y.; Liu, L.; Li, R. C.; Liu, R.; Guo, Q. X. *J. Am. Chem. Soc.* **2004**, 126, 814–822.
- Gao, D. Q.; Svoronos, P.; Wong, P. K.; Maddalena, D.; Hwang, J.; Walker, H. *J. Phys. Chem. A* **2005**, 109, 10776–10785.
- Namazian, M.; Kalantary-Fotooh, F.; Noorbala, M. R.; Searles, D. J.; Coote, M. L. *THEOCHEM* **2006**, 758, 275–278.
- Namazian, M.; Halvani, S. *J. Chem. Thermodyn.* **2006**, 38, 1495–1502.
- Pratungdejkul, J.; Nosoongnoen, W.; Guerin, G. A.; Loric, S.; Conti, M.; Launay, J. M.; Manivet, P. *Chem. Phys. Lett.* **2006**, 420, 538–544.
- Sadlej-Sosnowska, N. *Theor. Chem. Acc.* **2007**, 118, 281–293.

- (20) Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L. *Chem. Phys. Lett.* **2008**, *451*, 163–168.
- (21) Rayne, S.; Forest, K. *THEOCHEM* **2010**, *949*, 60–69.
- (22) Lee, T. B.; McKee, M. L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10258–10269.
- (23) Zevatskii, Y. E.; Samoilov, D. V. *Russ. J. Org. Chem.* **2011**, *47*, 1445–1467.
- (24) Magill, A. M.; Yates, B. F. *Aust. J. Chem.* **2004**, *57*, 1205–1210.
- (25) Ali-Torres, J.; Rodriguez-Santiago, L.; Sodupe, M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 7852–7861.
- (26) da Silva, G.; Kennedy, E. M.; Dlugogorski, B. Z. *J. Phys. Chem. A* **2006**, *110*, 11371–11376.
- (27) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.
- (28) Ho, J.; Klamt, A.; Coote, M. L. *J. Phys. Chem. A* **2010**, *114*, 13442–13444.
- (29) Marenich, A. V.; Ding, W. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. Lett.* **2012**, *3*, 1437–1442.
- (30) Ruscic, B.; Wagner, A. F.; Harding, L. B.; Asher, R. L.; Feller, D.; Dixon, D. A.; Peterson, K. A.; Song, Y.; Qian, X. M.; Ng, C. Y.; et al. *J. Phys. Chem. A* **2002**, *106*, 2727–2747.
- (31) Smith, J. R.; Kim, J. B.; Lineberger, W. C. *Phys. Rev. A* **1997**, *55*, 2036–2043.
- (32) Bartmess, J. E. Negative Ion Energetics Data. In *NIST Chemistry WebBook*; Linstrom, P. J., Mallard, W. G., Eds; NIST Standard Reference Database Number 69; National Institute of Standards and Technology: Gaithersburg, MD; <http://webbook.nist.gov> (retrieved March 14, 2012).
- (33) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027–2094.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian, Inc.: Wallingford, CT, 2009.
- (35) Konstantinov, I. A.; Broadbelt, L. J. *Mol. Simul.* **2010**, *36*, 1197–1207.
- (36) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, No. 084108.
- (37) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2011**, *115*, 14556–14562.
- (38) Vanysek, P. *CRC Handbook of Chemistry and Physics*, 67th ed.; Weast, R. C., Astle, M. J., Beyer, W. H., Eds.; CRC Press, Inc: Boca Raton, FL, 1986; p D-162.
- (39) Reliable results are never obtained using the PX cycle with the same model chemistry for ΔG_{gas} and ΔG_{solv} because of the need to use diffuse functions in the gas phase which causes problems with the solvation Gibbs energy of OH^- (*vide infra*).
- (40) Using either M05-2X/6-31+G(d,p) or M05-2X/aug-cc-pVTZ for ΔG_{solv} gives particularly large errors with proton exchange. This is due to the $\Delta G_{\text{solv}}(\text{OH}^-)$ calculation; whereas the values in the G-31G(d) and cc-pVTZ cases are -106.0 and -104.9 kcal/mol, respectively, the 6-31+G(d,p) and aug-cc-pVZ cases return inaccurate results of -97.3 and -94.8 kcal/mol. Across all model chemistry options, B3LYP/6-31G(2df,p) has poor performance. M05-2X/cc-pVTZ has superior performance to all others.