

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/223129858>

# Experimentation with different thermodynamic cycles used for pKa calculations on carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvat...

ARTICLE in INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY · DECEMBER 2001

Impact Factor: 1.43 · DOI: 10.1002/qua.1703

---

CITATIONS

112

---

READS

29

## 2 AUTHORS:



Matthew D Liptak

University of Vermont

21 PUBLICATIONS 1,306 CITATIONS

SEE PROFILE



George C Shields

Bucknell University

96 PUBLICATIONS 3,935 CITATIONS

SEE PROFILE

---

# 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

---

MATTHEW D. LIPTAK, GEORGE C. SHIELDS

*Contribution from the Department of Chemistry, Hamilton College, 198 College Hill Road, Clinton, New York 13323*

*Received 28 February 2001; revised 21 May 2001; accepted 3 July 2001*

---

**ABSTRACT:** Complete basis set and Gaussian- $n$  methods were combined with Barone and Cossi's implementation of the polarizable conductor model (CPCM) continuum solvation methods to calculate  $pK_a$  values for six carboxylic acids. Four different thermodynamic cycles were considered in this work. An experimental value of  $-264.61$  kcal/mol for the free energy of solvation of  $H^+$ ,  $\Delta G_s(H^+)$ , was combined with a value for  $G_{\text{gas}}(H^+)$  of  $-6.28$  kcal/mol, to calculate  $pK_a$  values with cycle 1. The complete basis set gas-phase methods used to calculate gas-phase free energies are very accurate, with mean unsigned errors of 0.3 kcal/mol and standard deviations of 0.4 kcal/mol. The CPCM solvation calculations used to calculate condensed-phase free energies are slightly less accurate than the gas-phase models, and the best method has a mean unsigned error and standard deviation of 0.4 and 0.5 kcal/mol, respectively. Thermodynamic cycles that include an explicit water in the cycle are not accurate when the free energy of solvation of a water molecule is used, but appear to become accurate when the experimental free energy of vaporization of water is used. This apparent improvement is an artifact of the standard state used in the calculation. Geometry relaxation in solution does not improve the results when using these later cycles. The use of cycle 1 and the complete basis set models combined with the CPCM solvation methods yielded  $pK_a$  values accurate to less than half a  $pK_a$  unit. © 2001 John Wiley & Sons, Inc. *Int J Quantum Chem* 85: 727–741, 2001

*Correspondence to:* G. C. Shields; e-mail: gshields@hamilton.edu.

Contract grant sponsors: Petroleum Research Fund; NIH; Hamilton College; Merck/AAAS Undergraduate Summer Research Program.

**Key words:**  $pK_a$ ; carboxylic acids; CBS-QB3; CBS-APNO; G2; G3; free energy; aqueous solution; thermodynamic cycles

## Introduction

The field of computational chemistry is reaching the point where calculations at the level of chemical accuracy, within 1 kcal/mol, are now possible. Numerous attempts to accurately calculate  $pK_a$  values have been made, but none has achieved chemical accuracy [1–28]. The definition of  $pK_a$  is

$$pK_a = -\log K_a, \quad (1)$$

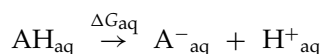
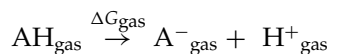
and since

$$\Delta G^0 = -2.303RT \log K_a, \quad (2)$$

$$pK_a = \Delta G^0 / 2.303RT. \quad (3)$$

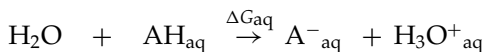
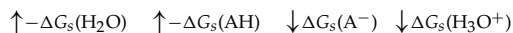
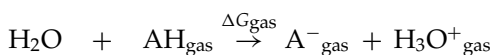
Calculating accurate  $pK_a$  values is demanding, as an error of 1.36 kcal/mol in  $\Delta G^0$  gives an error of 1  $pK_a$  unit. There are at least three sources of error in  $pK_a$  calculations. The first is the model used to calculate  $pK_a$ , which generally involves a thermodynamic cycle such as:

### Cycle 1



or

### Cycle 2



In these cycles  $\Delta G_{\text{gas}}$  is calculated with high-level ab initio or density functional methods, and the  $\Delta G_s$  values are calculated using a solvation method, typically a continuum dielectric approach. The  $pK_a$  values can be determined from Eq. (3), where  $\Delta G^0 = \Delta G_{\text{aq}}$ . The second and third major errors stem from the accuracy of the calculations for  $\Delta G_{\text{gas}}$  and  $\Delta G_s$ . Relative  $pK_a$  calculations allow one to calculate the unknown  $pK_a$  of a molecule BH from the known

$pK_a$  of molecule AH. In these relative calculations only the values for  $G_{AH}$ ,  $G_{A^-}$ ,  $G_{BH}$ ,  $G_{B^-}$ ,  $\Delta G_s(AH)$ ,  $\Delta G_s(A^-)$ ,  $\Delta G_s(BH)$ , and  $\Delta G_s(B^-)$  are required for either thermodynamic cycle, as the values for the species involving  $H^+$  in cycle 1, or  $H_3O^+$  and  $H_2O$  in cycle 2, cancel. For this reason, accuracy in the calculation of relative  $pK_a$  values depends on the second and third main errors. We have recently reported that relative and absolute  $pK_a$  values accurate to within half a  $pK_a$  unit can be calculated for carboxylic acids [29, 30]. The absolute  $pK_a$  values were determined with thermodynamic cycle 1 [30]. In this work we analyze the errors that stem from the use of different thermodynamic cycles in  $pK_a$  calculations. This work is of fundamental importance because different workers have used different treatments of  $H^+/H_3O^+$  in their own calculations.

## Methods

To find the best thermodynamic cycle for calculating absolute  $pK_a$  values, we used six simple carboxylic acids that ranged in size from 24 to 68 electrons: formic acid, acetic acid, cyanoacetic acid, chloroacetic acid, oxalic acid, and pivalic acid. All calculations were performed on Origin 200 SGI servers equipped with 1–2 GB of memory and 36 GB of scratch disk space, using Gaussian 98 [31], Spartan 5.1 [32], and GAMESOL [33] software. The acids and corresponding ions were initially optimized in the gas phase at the Hartree–Fock (HF) level. We then used the CBS-QB3 [34], CBS-APNO [35, 36], G3 [37], and G2 [38] model chemistry methods to calculate  $G_{\text{gas}}$  values for various conformers of all six acids and their corresponding anions. CBS-APNO calculations were not run on chloroacetic acid and the chloroacetate anion as this method is not yet available for chlorine.

The complete basis set methods [34–36] were developed by Petersson and co-workers as a model chemistry that makes use of a complete basis set (CBS) extrapolation of the correlation energy based on the asymptotic convergence of pair natural orbital expansions [39–43]. The philosophy of the method is based on the idea that the major source of error in most ab initio calculations of molecular energies is the truncation of the one-electron

basis set. The CBS-QB3 method uses density functional theory geometries and SPCs are performed at the CCSD(T), MP4SDQ, and MP2 levels with small basis sets. The CBS-APNO method uses HF/6-311G(d,p) geometries for the frequency calculation and SPCs are carried out on QCISD/6-311G(d,p) geometries. The SPCs are expensive and use QCISD(T) and MP2 theory for correlation corrections.

Pople and co-workers [37, 38, 44–47] have developed the Gaussian methods (G1, G2, G3, and MP2 variants), which are extrapolation schemes similar to the CBS model chemistries. The G2, G3, CBS-QB3, and CBS-APNO model chemistries are state-of-the-art models for accurate thermochemistry [34, 37, 48, 49].

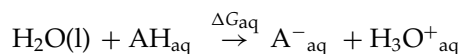
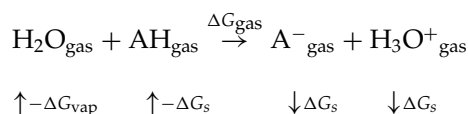
The absence of imaginary frequencies verified that all structures were true minima at their respective levels of calculation. Higher energy conformers were also computed using each method for five of the acids. The absolute value of each energy, in hartrees, have been previously reported [29, 30].

We computed  $\Delta G_s$  using Barone and Cossi's implementation of the polarizable conductor model (CPCM) [50], which is based on the polarized continuum model (PCM) of Tomasi and co-workers [51–57]. In this method the solute cavities are modeled on the optimized molecular shape and include both electrostatic and nonelectrostatic contributions to the energies. The CPCM calculations were performed as SPCs using the 6-31G(d) and 6-31+G(d) basis sets on the HF/6-31G(d) and HF/6-31+G(d) geometries for each of the six systems. Geometry optimizations were also performed using CPCM/HF/6-31G(d). In all of the CPCM calculations the surface of each sphere was subdivided into 240 triangular tesserae, and the area of the tesserae was set at 0.3 Å<sup>2</sup>. All CPCM calculations were performed using Gaussian 98 [31].

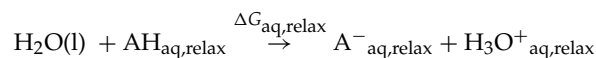
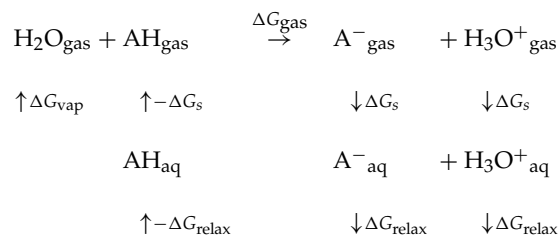
Using different combinations of  $G_{\text{gas}}$  and  $\Delta G_s$  values, and different thermodynamic cycles, we calculated the absolute pK<sub>a</sub> for each acid in its lowest energy gas-phase conformation. In addition, we calculated the conformational average of the aqueous free energies of the acids [59] and computed the absolute pK<sub>a</sub> values for each of the six acids.

We compared four different thermodynamic cycles in this work. Besides the two cycles outlined in the introduction, we used the following cycles, which are based on the work of Nascimento and co-workers [21, 28].

### Cycle 3



### Cycle 4



The difference between cycle 2 and cycle 3 is that the experimental free energy of vaporization of water is used instead of the free energy of solvation. Cycle 4 includes the relaxation of each molecule in solution, from the gas-phase geometry to the solution geometry, obtained by optimizing each molecule in the presence of the dielectric continuum [21, 28].

In Eqs. (2) and (3), our  $\Delta G^0$  is  $\Delta G_{\text{aq}}$  from the thermodynamic cycles. Therefore all of our pK<sub>a</sub> calculations use the formula

$$\text{pK}_a = \Delta G_{\text{aq}} / 2.303 \text{ RT}. \quad (4)$$

In our thermodynamic cycles,

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

and for cycle 1,

$$\Delta \Delta G_{\text{sol}} = \Delta G_s(\text{H}^+) + \Delta G_s(\text{A}^-) - \Delta G_s(\text{AH}). \quad (6)$$

The values for  $G(\text{H}^+_{\text{gas}})$  and  $\Delta G_s(\text{H}^+)$  are derived from experiment. We have used the values  $G(\text{H}^+_{\text{gas}}) = -6.28$  kcal/mol and  $\Delta G_s(\text{H}^+) = -264.61$  kcal/mol. The calculation of  $\Delta G_{\text{gas}}$  uses a reference state of 1 atm, and the calculations of  $\Delta G_s$  use a reference state of 1 M. Converting the  $\Delta G_{\text{gas}}$  reference state (24.46 L at 298.15 K) from 1 atm to 1 M is accomplished using:

$$\Delta G_{\text{gas}}(1 \text{ M}) = \Delta G_{\text{gas}}(1 \text{ atm}) + \text{RT} \ln(24.46). \quad (7)$$

Using these values, the pK<sub>a</sub> values using the first thermodynamic cycle are given by Eq. (8):

$$\begin{aligned} \text{pK}_a = [ & G(\text{A}^-_{\text{gas}}) - G(\text{AH}_{\text{gas}}) + \Delta G_s(\text{A}^-) \\ & - \Delta G_s(\text{AH}) - 269.0 ] / 1.3644. \end{aligned} \quad (8)$$

The  $pK_a$ 's calculated with cycle 2 used Eq. (9):

$$pK_a = [G(\text{H}_3\text{O}^+_{\text{gas}}) - G(\text{A}^-_{\text{gas}}) - G(\text{AH}_{\text{gas}}) - G(\text{H}_2\text{O}_{\text{gas}}) + \Delta G_s(\text{H}_3\text{O}^+) + \Delta G_s(\text{A}^-) - \Delta G_s(\text{AH}) - \Delta G_s(\text{H}_2\text{O})]/1.3644 - 1.74303, \quad (9)$$

where we have subtracted the log of the water concentration at 298 K to convert  $pK$  for cycle 2 to  $pK_a$ .

The  $pK_a$ 's calculated with cycle 3 used Eq. (10):

$$pK_a = [G(\text{H}_3\text{O}^+_{\text{gas}}) - G(\text{A}^-_{\text{gas}}) - G(\text{AH}_{\text{gas}}) - G(\text{H}_2\text{O}_{\text{gas}}) + \Delta G_s(\text{H}_3\text{O}^+) + \Delta G_s(\text{A}^-) - \Delta G_s(\text{AH}) + 2.05]/1.3644 - 1.74303, \quad (10)$$

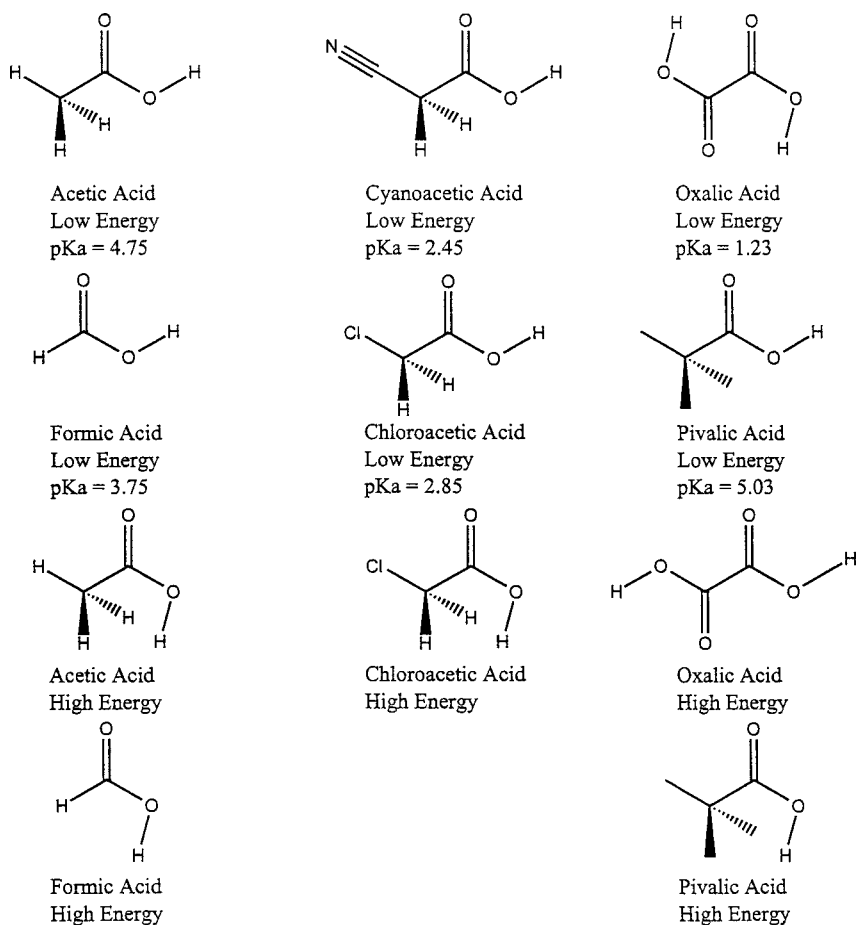
where  $\Delta G_{\text{vap}}$  is 2.05 kcal/mol.

The  $pK_a$ 's calculated with cycle 4 used Eq. (11):

$$pK_a = [G(\text{H}_3\text{O}^+_{\text{gas}}) - G(\text{A}^-_{\text{gas}}) - G(\text{AH}_{\text{gas}}) - G(\text{H}_2\text{O}_{\text{gas}}) + \Delta G_s(\text{H}_3\text{O}^+) + \Delta G_{\text{relax}}(\text{H}_3\text{O}^+) + \Delta G_s(\text{A}^-) + \Delta G_{\text{relax}}(\text{A}^-) - \Delta G_s(\text{AH}) - \Delta G_{\text{relax}}(\text{AH}) + 2.05]/1.3644 - 1.74303. \quad (11)$$

## Results

The conformers of the carboxylic acids are displayed in Figure 1, along with the experimental  $pK_a$ 's of each acid. The conformationally averaged  $\Delta G_s$  values for the six carboxylic acids and their anions, obtained using the CBS-QB3 gas-phase and CPCM solvation methods, are given in Table I [30]. CPCM/6-31G(d)//HF/6-31G(d), CPCM/6-31+G(d)//HF/6-31G(d), and CPCM/6-31+G(d)//HF/6-31+G(d) are denoted **S1**, **S2**, and **S3**, respectively. Table II contains the previously reported calculated  $pK_a$ 's and mean unsigned errors from thermodynamic cycle 1 and Eq. (8) [30]. Table III contains the calculated  $pK_a$ 's and errors



**FIGURE 1.** Experimental  $pK_a$  values and structures of the conformers of acetic, chloroacetic, cyanoacetic, formic, oxalic, and pivalic acids.

**TABLE I**  
Conformationally averaged  $\Delta G_s$  values for carboxylic acids and anions using CBS-QB3, in kcal/mol.<sup>a,b</sup>

	Acetic	Formic	Cyanoacetic	Chloroacetic	Oxalic	Pivalic	Water	
<b>S1</b>	−6.93	−7.39	−13.24	−9.67	−11.56	−5.97	−6.37	
<b>S2</b>	−7.65	−8.38	−14.32	−10.56	−12.58	−6.70	−7.26	
<b>S3</b>	−7.72	−8.43	−14.52	−10.65	−12.73	−6.78	−7.23	
	Acetate	Formate	Cyanoacetate	Chloroacetate	Oxalate	Pivalate	H <sup>+</sup> <sup>c</sup>	H <sub>3</sub> O <sup>+</sup> <sup>d</sup>
<b>S1</b>	−76.58	−76.15	−68.72	−69.48	−74.92	−71.04	−264.61	−104.00
<b>S2</b>	−77.18	−76.58	−69.51	−70.11	−75.35	−71.97	−264.61	−104.00
<b>S3</b>	−77.58	−77.10	−69.99	−70.57	−75.72	−72.42	−264.61	−104.00

<sup>a</sup> From Ref. [30].<sup>b</sup> **S1** = CPCM/HF/6-31G(d)//HF/6-31G(d); **S2** = CPCM/HF/6-31+G(d)//HF/6-31G(d); **S3** = CPCM/HF/6-31+G(d)//HF/6-31+G(d).<sup>c</sup> From experimental values for acetic acid, see text for details.<sup>d</sup> Experimental value, Ref. [74].

when using thermodynamic cycle 2 and Eq. (9), with all values calculated using one of the five gas-phase methods and one of the three solvation procedures. Table IV contains the calculated pK<sub>a</sub>'s and errors from thermodynamic cycle 3 and Eq. (10). Table V contains the calculated pK<sub>a</sub>'s and errors from thermodynamic cycle 4 and Eq. (11). This cycle differs from cycle 3 only in that the gas-phase geometries are allowed to relax in solution, a procedure first used by Nascimento and co-workers [21, 28]. Table VI contains a pK<sub>a</sub> error analysis of each of five gas-phase computational methods combined with three condensed-phase methods, for each of the four thermodynamic cycles.

## Discussion

### REVIEW OF PREVIOUS WORK USING THERMODYNAMIC CYCLE 1

Table I contains the previously reported conformationally averaged values for  $\Delta G_s$ . The previously reported [30] absolute pK<sub>a</sub> values are presented in Table II for thermodynamic cycle 1 using the values in Table I and the previously reported [29] values for G<sub>gas</sub> for the acids and the anions. The value of −6.28 comes from the Sackur–Tetrode equation [66] to evaluate the entropy combined with inclusion of translational energy at 298 K [67]. The value for  $\Delta G_s(\text{H}^+)$  was obtained using the average of three published experimental values for the gas-phase dissociation of acetic acid in thermodynamic cycle 1 [61–63] the experimental  $\Delta G_s$  values

for acetic acid (−6.69 kcal/mol) and acetate ion (−77 kcal/mol) [68], and the experimental value for  $\Delta G_{\text{aq}}$  ( $2.303RT \text{ p}K_a = 6.48 \text{ kcal/mol}$ ) [69].

The CBS-QB3 gas-phase method and the **S1** solvation procedure predict absolute pK<sub>a</sub> values of 5.34, 3.47, 2.33, 3.44, 0.92, and 5.60 for acetic, formic, cyanoacetic, chloroacetic, oxalic, and pivalic acids, respectively (Table II) [30]. The largest deviation is 0.59 pK<sub>a</sub> units, and the mean unsigned error (MUE) for the six acids is 0.41 pK<sub>a</sub> units using this combination of gas-phase and condensed phase methods. Using the CBS-QB3 gas-phase method, the **S2** solvation procedure gives a MUE for the six acids of 0.37 pK<sub>a</sub> units, while **S3** solvation gives a MUE of 0.24 pK<sub>a</sub> units. Results for the other gas-phase methods are similar. The CBS-APNO method is quite accurate, although it failed for one conformer of pivalic acid [70]. The G2 gas-phase methods and the **S1**, **S2**, and **S3** solvation methods were reliable for all but pivalic acid. The G3 method combined with the **S1**, **S2**, and **S3** solvation methods were least accurate for chloroacetic, oxalic, and pivalic acids.

A key problem for previous workers trying to make absolute pK<sub>a</sub> predictions is the value to use for  $\Delta G_s(\text{H}^+)$  [30, 71, 72]. The difference between the recalculated value of −264 kcal/mol of Tissandier et al. [72] and the commonly used values of −259.5 to −262.5 kcal/mol is in itself enough to make the error in calculated absolute pK<sub>a</sub> calculations as large as one to three pK<sub>a</sub> units. We derived a value of −264.6 kcal/mol from the experimental thermodynamic cycle of acetic acid, where all quantities are known except for  $\Delta G_s(\text{H}^+)$ . As Table II shows, the

**TABLE II**  
**pK<sub>a</sub> values using CBS and Gaussian-*n* models, CPCM solvation calculations, and cycle 1.<sup>a,b</sup>**

Acetic acid $pK_a = 4.75$				Formic acid $pK_a = 3.75$				Cyanoacetic acid $pK_a = 2.45$			
Solvation	$\Delta\Delta G_{sol}$	$pK_a$	Difference	Solvation	$\Delta\Delta G_{sol}$	$pK_a$	Difference	Solvation	$\Delta\Delta G_{sol}$	$pK_a$	Difference
CBS-QB3											
<b>S1</b>	-334.26	5.34	0.59	<b>S1</b>	-333.37	3.47	-0.28	<b>S1</b>	-320.09	2.33	-0.12
<b>S2</b>	-334.14	5.43	0.68	<b>S2</b>	-332.81	3.88	0.13	<b>S2</b>	-319.80	2.54	0.09
<b>S3</b>	-334.47	5.19	0.44	<b>S3</b>	-333.28	3.53	-0.22	<b>S3</b>	-320.08	2.34	-0.11
CBS-APNO											
<b>S1</b>	-334.26	5.06	0.31	<b>S1</b>	-333.37	3.86	0.11	<b>S1</b>	-320.09	2.31	-0.14
<b>S2</b>	-334.14	5.15	0.40	<b>S2</b>	-332.81	4.27	0.52	<b>S2</b>	-319.80	2.52	0.07
<b>S3</b>	-334.47	4.91	0.16	<b>S3</b>	-333.28	3.93	0.18	<b>S3</b>	-320.08	2.31	-0.14
G2											
<b>S1</b>	-334.26	4.59	-0.16	<b>S1</b>	-333.37	3.29	-0.46	<b>S1</b>	-320.09	1.97	-0.48
<b>S2</b>	-334.14	4.67	-0.08	<b>S2</b>	-332.81	3.70	-0.05	<b>S2</b>	-319.80	2.18	-0.27
<b>S3</b>	-334.47	4.43	-0.32	<b>S3</b>	-333.28	3.35	-0.40	<b>S3</b>	-320.08	1.97	-0.48
G2MP2											
<b>S1</b>	-334.26	4.87	0.12	<b>S1</b>	-333.37	3.63	-0.12	<b>S1</b>	-320.09	2.26	-0.19
<b>S2</b>	-334.14	4.96	0.21	<b>S2</b>	-332.81	4.04	0.29	<b>S2</b>	-319.80	2.48	0.03
<b>S3</b>	-334.47	4.72	-0.03	<b>S3</b>	-333.28	3.70	-0.05	<b>S3</b>	-320.08	2.27	-0.18
G3											
<b>S1</b>	-334.26	5.22	0.47	<b>S1</b>	-333.37	4.16	0.41	<b>S1</b>	-320.09	2.60	0.15
<b>S2</b>	-334.14	5.31	0.56	<b>S2</b>	-332.81	4.57	0.82	<b>S2</b>	-319.80	2.82	0.37
<b>S3</b>	-334.47	5.07	0.32	<b>S3</b>	-333.28	4.23	0.48	<b>S3</b>	-320.08	2.61	0.16
Chloroacetic acid $pK_a = 2.85$				Oxalic acid $pK_a = 1.23$				Pivalic acid $pK_a = 5.03$			
CBS-QB3											
<b>S1</b>	-324.42	3.44	0.59	<b>S1</b>	-327.97	0.92	-0.31	<b>S1</b>	-329.68	5.60	0.57
<b>S2</b>	-324.16	3.63	0.78	<b>S2</b>	-327.38	1.36	0.13	<b>S2</b>	-329.88	5.46	0.43
<b>S3</b>	-324.53	3.36	0.51	<b>S3</b>	-327.60	1.20	-0.03	<b>S3</b>	-330.25	5.19	0.16
CBS-APNO											
Method not available for chlorine				<b>S1</b>	-327.97	1.24	0.01	See Ref. [70]			
				<b>S2</b>	-327.38	1.67	0.44				
				<b>S3</b>	-327.60	1.51	0.28				
G2											
<b>S1</b>	-324.42	3.21	0.36	<b>S1</b>	-327.97	1.15	-0.08	<b>S1</b>	-329.68	6.51	1.48
<b>S2</b>	-324.16	3.40	0.55	<b>S2</b>	-327.38	1.58	0.35	<b>S2</b>	-329.88	6.37	1.34
<b>S3</b>	-324.53	3.13	0.28	<b>S3</b>	-327.60	1.42	0.19	<b>S3</b>	-330.25	6.10	1.07
G2MP2											
<b>S1</b>	-324.42	3.49	0.64	<b>S1</b>	-327.97	1.52	0.29	<b>S1</b>	-329.68	6.76	1.73
<b>S2</b>	-324.16	3.68	0.83	<b>S2</b>	-327.38	1.95	0.72	<b>S2</b>	-329.88	6.62	1.59
<b>S3</b>	-324.53	3.41	0.56	<b>S3</b>	-327.60	1.79	0.56	<b>S3</b>	-330.25	6.34	1.31

(Continued)

**TABLE II**  
(Continued).

Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference
G3											
<b>S1</b>	-324.42	3.69	0.84	<b>S1</b>	-327.97	1.99	0.76	<b>S1</b>	-329.68	7.06	2.03
<b>S2</b>	-324.16	3.88	1.03	<b>S2</b>	-327.38	2.42	1.19	<b>S2</b>	-329.88	6.91	1.88
<b>S3</b>	-324.53	3.61	0.76	<b>S3</b>	-327.60	2.26	1.03	<b>S3</b>	-330.25	6.64	1.61

<sup>a</sup> From Ref. [30].<sup>b</sup> **S1** = CPCM/HF/6-31G(d)//HF/6-31G(d); **S2** = CPCM/HF/6-31+G(d)//HF/S-31G(d); **S3** = CPCM/HF/6-31+G(d)//HF/6-31+G(d).

correct value for  $\Delta G_s(\text{H}^+)$  must be in the range of -264 kcal/mol [30].

The Sackur–Tetrode equation [66] was used to evaluate the entropy,  $G(\text{H}^+)$ , which yields a value of -7.76 kcal/mol for  $\text{TS}(\text{H}^+)$  at 298 K and 1 atm pressure. Including translational energy, the total gas-phase free energy of the proton is -6.28 kcal/mol at 298 K [67].

## THERMODYNAMIC CYCLE 2

Table III contains the results of pK<sub>a</sub> calculations using thermodynamic cycle 2 and Eq. (9). None of the methods are accurate. The MUE in pK<sub>a</sub> ranges from 2.4 to 5.3 pK<sub>a</sub> units. It is no surprise that previous workers who have used cycle 2 have reported that quantitative prediction of absolute pK<sub>a</sub> values are not possible [13, 20, 24]. We can assess the quality of this cycle by using available experimental numbers. Using acetic acid, where  $\Delta G_s(\text{AH})$  is -6.7 kcal/mol [65],  $\Delta G_s(\text{A}^-)$  is -77.0 kcal/mol [65],  $\Delta G_s(\text{H}_2\text{O})$  is -6.3 kcal/mol [73], and  $\Delta G_s(\text{H}_3\text{O}^+)$  is -104 kcal/mol [74],  $\Delta\Delta G_{\text{sol}}$  has a value of -168.1 kcal/mol. Combining this value with the computed value for  $\Delta G_{\text{gas}}$  for the top line in cycle 2 will give us an estimate for the value for  $\Delta G_{\text{aq}}$ , which can be compared to the same value as that determined from the pK<sub>a</sub>. If the data in the thermodynamic cycle were accurate, one would get the same value for  $\Delta G_{\text{aq}}$  as determined from the pK<sub>a</sub>. Acetic acid has a pK<sub>a</sub> of 4.75, so  $\Delta G_{\text{aq}}$  is 6.48 kcal/mol.

Using the CBS-QB3 calculated values for acetic acid (-228.792813 hartrees), water (-76.355104 hartrees), acetate anion (-228.241528 hartrees), and  $\text{H}_3\text{O}^+$  (-76.615647 hartrees), the CBS-QB3  $\Delta G_{\text{gas}}$  value is 182.4 kcal/mol. Combining the CBS-QB3 value for  $\Delta G_{\text{gas}}$  and the experimental value for  $\Delta\Delta G_{\text{sol}}$  yields 13.4 kcal/mol for the value of  $\Delta G_{\text{aq}}$  determined from cycle 2, which is higher than  $\Delta G_{\text{aq}}$  obtained directly from the experimental pK<sub>a</sub> by

about 7 kcal/mol. If we use the  $\Delta G_s$  numbers calculated with the **S3** procedure (-106.27, -77.58, -7.72, and -7.23 for  $\text{H}_3\text{O}^+$ , acetate ion, acetic acid, and water, respectively), then  $\Delta\Delta G_{\text{sol}}$  is -168.9, with no improvement in  $\Delta G_{\text{aq}}$ .

Where is the error? Our previous work showed that relative pK<sub>a</sub> calculations for these same carboxylic acids were quite accurate, and in these calculations the subtraction of one thermodynamic cycle from another meant that the results from using either thermodynamic cycle 1 or thermodynamic cycle 2 would be equivalent [29]. In relative pK<sub>a</sub> calculations the only unique numbers are the gas-phase values  $G_{\text{gas}}$  for each acid and its corresponding anion, and the  $\Delta G_s$  values for each acid and each anion. The values for  $G(\text{H}^+_{\text{gas}})$  and  $\Delta G_s(\text{H}^+)$  in thermodynamic cycle 1, and  $G(\text{H}_3\text{O}^+_{\text{gas}})$ ,  $\Delta G_s(\text{H}_3\text{O}^+)$ , and  $\Delta G_s(\text{H}_2\text{O})$  in thermodynamic cycle 2, all cancel in relative pK<sub>a</sub> calculations. The error in using thermodynamic cycle 2 for absolute pK<sub>a</sub> calculations must therefore be in one or more of the following terms:  $G(\text{H}_2\text{O}_{\text{gas}})$ ,  $G(\text{H}_3\text{O}^+_{\text{gas}})$ ,  $\Delta G_s(\text{H}_3\text{O}^+)$ , or  $\Delta G_s(\text{H}_2\text{O})$ . Nascimento and co-workers have suggested that the error is in  $\Delta G_s(\text{H}_2\text{O})$  [21, 28]. They reason that the thermodynamic definition of solvent energy [73, 78] is inappropriate when the solute and the solvent are the same species, as a solvated water molecule is indistinguishable from any other water molecule [28]. They proposed that it would be more appropriate to identify the transformation of water in the thermodynamic cycle with the vaporization process, and that using  $\Delta G_s(\text{H}_2\text{O})$  instead of  $\Delta G_{\text{vap}}(\text{H}_2\text{O})$  in the cycle represents a source of error that leads to an overestimation of pK<sub>a</sub> values [21]. They have used thermodynamic cycle 4 for their work on carboxylic acids [21] and aliphatic acids, thiols, and halogenated carboxylic acids [28]. We will describe how their cycle, and a modified cycle, works with our calculations in the next section.



**TABLE III**  
**pK<sub>a</sub> values using CBS and Gaussian-*n* models, CPCM solvation calculations, and cycle 2.<sup>a</sup>**

Acetic acid				Formic acid				Cyanoacetic acid			
Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference
CBS-QB3											
S1	-167.28	9.37	4.62	S1	-166.39	7.50	3.75	S1	-153.11	6.36	3.91
S2	-166.27	10.11	5.36	S2	-164.94	8.56	4.81	S2	-151.93	7.22	4.77
S3	-166.63	9.85	5.10	S3	-165.44	8.19	4.44	S3	-152.24	7.00	4.55
CBS-APNO											
S1	-167.28	8.21	3.46	S1	-166.39	7.01	3.26	S1	-153.11	5.45	3.00
S2	-166.27	8.95	4.20	S2	-164.94	8.07	4.32	S2	-151.93	6.31	3.86
S3	-166.63	8.68	3.93	S3	-165.44	7.70	3.95	S3	-152.24	6.09	3.64
G2											
S1	-167.28	8.22	3.47	S1	-166.39	6.92	3.17	S1	-153.11	5.60	3.15
S2	-166.27	8.96	4.21	S2	-164.94	7.99	4.24	S2	-151.93	6.47	4.02
S3	-166.63	8.70	3.95	S3	-165.44	7.62	3.87	S3	-152.24	6.24	3.79
G2MP2											
S1	-167.28	8.51	3.76	S1	-166.39	7.27	3.52	S1	-153.11	5.90	3.45
S2	-166.27	9.25	4.50	S2	-164.94	8.33	4.58	S2	-151.93	6.76	4.31
S3	-166.63	8.98	4.23	S3	-165.44	7.96	4.21	S3	-152.24	6.53	4.08
G3											
S1	-167.28	8.61	3.86	S1	-166.39	7.55	3.80	S1	-153.11	5.99	3.54
S2	-166.27	9.35	4.60	S2	-164.94	8.62	4.87	S2	-151.93	6.86	4.41
S3	-166.63	9.09	4.34	S3	-165.44	8.25	4.50	S3	-152.24	6.63	4.18
Chloroacetic acid				Oxalic acid				Pivalic acid			
CBS-QB3											
S1	-157.44	7.47	4.62	S1	-160.99	4.95	3.72	S1	-162.70	9.63	4.60
S2	-156.29	8.31	5.46	S2	-159.51	6.04	4.81	S2	-162.01	10.14	5.11
S3	-156.69	8.02	5.17	S3	-159.76	5.86	4.63	S3	-162.41	9.84	4.81
CBS-APNO											
				S1	-160.99	4.38	3.15				
				S2	-159.51	5.47	4.24				
				S3	-159.76	5.29	4.06				
G2											
S1	-157.44	6.85	4.00	S1	-160.99	4.79	3.56	S1	-162.70	10.15	5.12
S2	-156.29	7.69	4.84	S2	-159.51	5.87	4.64	S2	-162.01	10.66	5.63
S3	-156.69	7.40	4.55	S3	-159.76	5.69	4.46	S3	-162.41	10.36	5.33
G2MP2											
S1	-157.44	7.13	4.28	S1	-160.99	5.15	3.92	S1	-162.70	10.40	5.37
S2	-156.29	7.97	5.12	S2	-159.51	6.23	5.00	S2	-162.01	10.90	5.87
S3	-156.69	7.68	4.83	S3	-159.76	6.05	4.82	S3	-162.41	10.61	5.58

(Continued)

**TABLE III**  
(Continued).

Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference
G3											
<b>S1</b>	-157.44	7.08	4.23	<b>S1</b>	-160.99	5.38	4.15	<b>S1</b>	-162.70	10.45	5.42
<b>S2</b>	-156.29	7.93	5.08	<b>S2</b>	-159.51	6.46	5.23	<b>S2</b>	-162.01	10.96	5.93
<b>S3</b>	-156.69	7.63	4.78	<b>S3</b>	-159.76	6.28	5.05	<b>S3</b>	-162.41	10.66	5.63

<sup>a</sup> **S1** = CPCM/HF/6-31G(d)//HF/6-31G(d); **S2** = CPCM/HF/6-31+G(d)//HF/6-31G(d); **S3** = CPCM/HF/6-31+G(d)//HF/6-31+G(d).

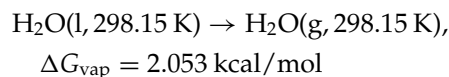
### THERMODYNAMIC CYCLES 3 AND 4

Thermodynamic cycle 3 is the same as cycle 2, except with  $-\Delta G_s(\text{H}_2\text{O})$  replaced by  $\Delta G_{\text{vap}}(\text{H}_2\text{O}) = 2.053 \text{ kcal/mol}$  [77]. Table IV contains the results of using cycle 3 and Eq. (10). The improvement over cycle 2 is dramatic, and we see that simply replacing the free energy of solvation of water with the free energy of vaporization gives reasonable predictions of the pK<sub>a</sub>'s of most of the molecules with most of the methods. The Gaussian-*n* models do not work well for pivalic acid, where the MUEs exceed 1 pK<sub>a</sub> unit. Overall the results are not as good as those for cycle 1 (Table II) but are clearly superior to those for cycle 2 (Table III). Clearly, replacing  $\Delta G_s(\text{H}_2\text{O})$  of  $-6.3 \text{ kcal/mol}$  with  $\Delta G_{\text{vap}}(\text{H}_2\text{O})$  of  $2.053 \text{ kcal/mol}$  leads to a dramatic improvement in calculating pK<sub>a</sub>'s when using a cycle that includes an explicit water.

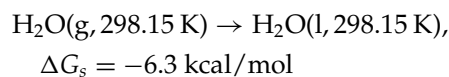
Table V contains the results of using cycle 4 and Eq. (11). In this cycle we have included the relaxation energy obtained by a CPCM/HF/6-31G(d) **S1** geometry optimization of each molecule in aqueous solution. Nascimento and co-workers have argued that inclusion of this term in their work was responsible for the accuracy of their pK<sub>a</sub> calculations [21, 28]. However, we find that inclusion of geometry relaxation for these carboxylic acids has little effect but tends to make the calculated pK<sub>a</sub>'s more negative. Geometry optimization results in lower pK<sub>a</sub> values, so that whenever cycle 3 gives results that are too positive (such as CBS-QB3 and **S1** for acetic acid) then the pK<sub>a</sub> values are improved by geometry optimization in aqueous solution. The exception to the trend is for formic acid, where the pK<sub>a</sub> values increase for all methods. This is a consequence of the small size of formic acid and formate anion, so that the relaxation energy is very small for these two species and the overall  $\Delta\Delta G_{\text{sol}}$  is dominated by the  $\Delta G_s(\text{H}_3\text{O}^+)$  term.

### STANDARD STATES AND FREE ENERGIES OF VAPORIZATION AND SOLVATION

We have seen that using a value for  $\Delta G_{\text{vap}}$  of  $2.053 \text{ kcal/mol}$  works in cycles 3 and 4 while using a value for  $\Delta G_s$  of  $-6.3 \text{ kcal/mol}$  does not work for cycle 2. The discrepancy between



and



must be investigated. The standard state for a pure liquid or solid is the substance in the condensed phase under a pressure of one atmosphere [77]. The values for  $\Delta G_f^0$  listed in the tables in the *CRC Handbook of Chemistry and Physics* [77] represent the change in the appropriate thermodynamic quantity when one gram-formula weight of the substance in its standard state is formed isothermally at the indicated temperature from the elements, each in its appropriate reference state. This means that  $\Delta G_f^0[\text{H}_2\text{O}(\text{l}, 298.15 \text{ K})]$  in the *CRC Handbook* has a standard state of 55.53 M and  $\Delta G_f^0[\text{H}_2\text{O}(\text{g}, 298.15 \text{ K})]$  has a standard state of 1 atm. Using the free energies of formation of liquid and gaseous water yields the value of 2.053, when the reference state is 55.53 M, 298.15 K for liquid water and 1 atm, 298.15 K for water vapor. Winget et al. have pointed out that the relationship between intermolecular interactions and free energies of transfer between phases is most direct if one uses the same standard concentration for both phases [78]. This is most often a standard state of one mole per liter for both the gas- and liquid-phase components, which is the standard state that we have tried to use in our thermodynamic cycles. Converting  $\Delta G_M$  from a standard state of  $\text{H}_2\text{O}(\text{l}, 298.15 \text{ K}, 55.53 \text{ M})$  and

**TABLE IV**  
**pK<sub>a</sub> values using CBS and Gaussian-*n* models, CPCM solvation calculations, and cycle 3.<sup>a</sup>**

Acetic acid				Formic acid				Cyanoacetic acid			
Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference
CBS-QB3											
S1	-171.60	6.20	1.45	S1	-170.71	4.33	0.58	S1	-157.43	3.19	0.74
S2	-171.48	6.29	1.54	S2	-170.15	4.74	0.99	S2	-157.14	3.40	0.95
S3	-171.81	6.05	1.30	S3	-170.62	4.40	0.65	S3	-157.42	3.20	0.75
CBS-APNO											
S1	-171.60	5.04	0.29	S1	-170.71	3.84	0.09	S1	-157.43	2.28	-0.17
S2	-171.48	5.13	0.38	S2	-170.15	4.25	0.50	S2	-157.14	2.50	0.05
S3	-171.81	4.89	0.14	S3	-170.62	3.91	0.16	S3	-157.42	2.29	-0.16
G2											
S1	-171.60	5.06	0.31	S1	-170.71	3.76	0.01	S1	-157.43	2.44	-0.01
S2	-171.48	5.14	0.39	S2	-170.15	4.17	0.42	S2	-157.14	2.65	0.20
S3	-171.81	4.90	0.15	S3	-170.62	3.82	0.07	S3	-157.42	2.44	-0.01
G2MP2											
S1	-171.60	5.34	0.59	S1	-170.71	4.10	0.35	S1	-157.43	2.73	0.28
S2	-171.48	5.43	0.68	S2	-170.15	4.51	0.76	S2	-157.14	2.94	0.49
S3	-171.81	5.19	0.44	S3	-170.62	4.17	0.42	S3	-157.42	2.74	0.29
G3											
S1	-171.60	5.45	0.70	S1	-170.71	4.39	0.64	S1	-157.43	2.83	0.38
S2	-171.48	5.54	0.79	S2	-170.15	4.80	1.05	S2	-157.14	3.04	0.59
S3	-171.81	5.29	0.54	S3	-170.62	4.45	0.70	S3	-157.42	2.84	0.39
Chloroacetic acid				Oxalic acid				Pivalic acid			
CBS-QB3											
S1	-161.76	4.30	1.45	S1	-165.31	1.79	0.56	S1	-167.02	6.47	1.44
S2	-161.50	4.49	1.64	S2	-164.72	2.22	0.99	S2	-167.22	6.32	1.29
S3	-161.87	4.22	1.37	S3	-164.94	2.06	0.83	S3	-167.59	6.05	1.02
CBS-APNO											
				S1	-165.31	1.22	-0.01				
				S2	-164.72	1.65	0.42				
				S3	-164.94	1.49	0.26				
G2											
S1	-161.76	3.68	0.83	S1	-165.31	1.62	0.39	S1	-167.02	6.98	1.95
S2	-161.50	3.87	1.02	S2	-164.72	2.05	0.82	S2	-167.22	6.84	1.81
S3	-161.87	3.60	0.75	S3	-164.94	1.89	0.66	S3	-167.59	6.57	1.54
G2MP2											
S1	-161.76	3.96	1.11	S1	-165.31	1.98	0.75	S1	-167.02	7.23	2.20
S2	-161.50	4.15	1.30	S2	-164.72	2.42	1.19	S2	-167.22	7.08	2.05
S3	-161.87	3.88	1.03	S3	-164.94	2.26	1.03	S3	-167.59	6.81	1.78

(Continued)

**TABLE IV**  
(Continued).

Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	pK <sub>a</sub>	Difference
G3											
<b>S1</b>	-161.76	3.92	1.07	<b>S1</b>	-165.31	2.21	0.98	<b>S1</b>	-167.02	7.29	2.26
<b>S2</b>	-161.50	4.11	1.26	<b>S2</b>	-164.72	2.64	1.41	<b>S2</b>	-167.22	7.14	2.11
<b>S3</b>	-161.87	3.84	0.99	<b>S3</b>	-164.94	2.48	1.25	<b>S3</b>	-167.59	6.87	1.84

<sup>a</sup> **S1** = CPCM/HF/6-31G(d)//HF/6-31G(d); **S2** = CPCM/HF/6-31+G(d)//HF/6-31G(d); **S3** = CPCM/HF/6-31+G(d)//HF/6-31+G(d).

**TABLE V**  
pK<sub>a</sub> values using CBS and Gaussian-*n* models, CPCM solvation calculations, and cycle 4.<sup>a</sup>

Acetic acid				Formic acid				Cyanoacetic acid			
Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference	Solvation	$\Delta\Delta G_{\text{sol}}$	$\text{p}K_{\text{a}}$	Difference
CBS-QB3											
S1	−172.44	5.59	0.84	S1	−170.45	4.52	0.77	S1	−158.25	2.59	0.14
CBS-APNO											
S1	−172.48	4.40	−0.35	S1	−170.26	4.17	0.42	S1	−158.25	1.68	−0.77
G2											
S1	−172.48	4.41	−0.34	S1	−170.23	4.11	0.36	S1	−158.25	1.84	−0.61
G2MP2											
S1	−172.48	4.69	−0.06	S1	−170.24	4.44	0.69	S1	−158.25	2.13	−0.32
G3											
S1	−172.47	4.81	0.06	S1	−170.29	4.70	0.95	S1	−158.25	2.23	−0.22
Chloroacetic acid				Oxalic acid				Pivalic acid			
CBS-QB3											
S1	−162.51	3.75	0.90	S1	−166.45	0.95	−0.28	S1	−167.96	5.78	0.75
CBS-APNO											
				S1	−166.09	0.65	−0.58				
G2											
S1	−162.65	3.03	0.18	S1	−166.45	0.79	−0.44	S1	−168.02	6.25	1.22
G2MP2											
S1	−162.64	3.32	0.47	S1	−166.44	1.16	−0.07	S1	−168.03	6.49	1.46
G3											
S1	−162.69	3.24	0.39	S1	−166.42	1.40	0.17	S1	−168.01	6.56	1.53

<sup>a</sup> **S1** = CPCM/HF/6-31G(d)//CPCM/HF/6-31G(d).

**TABLE VI**  
**Summary of errors by method.**

$G_{\text{gas}}$	$\Delta\Delta G_{\text{sol}}$	MSE	MUE	STDEV	$G_{\text{gas}}$	$\Delta\Delta G_{\text{sol}}$	MSE	MUE	STDEV
Thermodynamic cycle 1					Thermodynamic cycle 2				
CBS-QB3	<b>S1</b>	0.17	0.41	0.49	CBS-QB3	<b>S1</b>	4.20	4.20	4.63
CBS-QB3	<b>S2</b>	0.37	0.37	0.51	CBS-QB3	<b>S2</b>	5.05	5.05	5.54
CBS-QB3	<b>S3</b>	0.12	0.24	0.33	CBS-QB3	<b>S3</b>	4.78	4.78	5.25
CBS-APNO	<b>S1</b>	0.07	0.15	0.21	CBS-APNO	<b>S1</b>	3.22	3.22	3.72
CBS-APNO	<b>S2</b>	0.36	0.36	0.46	CBS-APNO	<b>S2</b>	4.15	4.15	4.80
CBS-APNO	<b>S3</b>	0.12	0.19	0.23	CBS-APNO	<b>S3</b>	3.89	3.89	4.50
G2	<b>S1</b>	0.11	0.51	0.75	G2	<b>S1</b>	3.75	3.75	4.17
G2	<b>S2</b>	0.31	0.44	0.68	G2	<b>S2</b>	4.60	4.60	5.07
G2	<b>S3</b>	0.06	0.46	0.59	G2	<b>S3</b>	4.33	4.33	4.77
G2MP2	<b>S1</b>	0.41	0.51	0.84	G2MP2	<b>S1</b>	4.05	4.05	4.49
G2MP2	<b>S2</b>	0.61	0.61	0.88	G2MP2	<b>S2</b>	4.90	4.90	5.39
G2MP2	<b>S3</b>	0.36	0.45	0.69	G2MP2	<b>S3</b>	4.63	4.63	5.10
G3	<b>S1</b>	0.78	0.78	1.08	G3	<b>S1</b>	4.17	4.17	4.61
G3	<b>S2</b>	0.98	0.98	1.20	G3	<b>S2</b>	5.02	5.02	5.52
G3	<b>S3</b>	0.73	0.73	0.96	G3	<b>S3</b>	4.75	4.75	5.23
Thermodynamic cycle 3					Thermodynamic cycle 4				
CBS-QB3	<b>S1</b>	1.04	1.04	1.22	CBS-QB3	<b>S1</b>	0.52	0.61	0.74
CBS-QB3	<b>S2</b>	1.23	1.23	1.39					
CBS-QB3	<b>S3</b>	0.99	0.99	1.12					
CBS-APNO	<b>S1</b>	0.05	0.14	0.20	CBS-APNO	<b>S1</b>	-0.32	0.53	0.64
CBS-APNO	<b>S2</b>	0.34	0.34	0.44					
CBS-APNO	<b>S3</b>	0.10	0.18	0.21					
G2	<b>S1</b>	0.58	0.58	0.98	G2	<b>S1</b>	0.06	0.53	0.68
G2	<b>S2</b>	0.78	0.78	1.04					
G2	<b>S3</b>	0.53	0.53	0.82					
G2MP2	<b>S1</b>	0.88	0.88	1.20	G2MP2	<b>S1</b>	0.36	0.51	0.77
G2MP2	<b>S2</b>	1.08	1.08	1.31					
G2MP2	<b>S3</b>	0.83	0.83	1.07					
G3	<b>S1</b>	1.00	1.00	1.28	G3	<b>S1</b>	0.48	0.55	0.83
G3	<b>S2</b>	1.20	1.20	1.42					
G3	<b>S3</b>	0.95	0.95	1.17					

$\text{H}_2\text{O}(\text{g}, 298.15 \text{ K}, 1 \text{ atm})$ , to a standard state of  $\text{H}_2\text{O}(\text{l}, 298.15 \text{ K}, 1 \text{ M})$  and  $\text{H}_2\text{O}(\text{g}, 298.15 \text{ K}, 1 \text{ M})$  can be achieved using Eq. (7), adding  $RT \ln(24.46)$  to convert the water vapor reference state, and then by adding  $RT \ln(55.5)$  to convert the liquid water reference state. Adding  $RT \ln(24.46)$  and  $RT \ln(55.5)$  to 2.053 yields 6.3 kcal/mol for the vaporization of water. Thus  $\Delta G_{\text{vap}}$  and  $\Delta G_{\text{s}}$  are the same numbers, but of opposite sign, as one would expect. Thus we are left with the conclusion that cycles 3 and 4 are get-

ting the right result for the wrong reason and that cycle 2 does not work. Further investigation into these cycles is warranted.

### ERROR ANALYSIS

Table VI contains a summary of the errors obtained by each of the four thermodynamic cycles combined with the CBS and Gaussian- $n$  gas phase and **S1**, **S2**, and **S3** solvation procedures, for the

five acids where  $\Delta G_{\text{gas}}$  is known experimentally. This analysis reveals that the most accurate results for the absolute calculation of pK<sub>a</sub> values for formic, acetic, cyanoacetic, chloroacetic, oxalic, and pivalic acids are obtained using CBS methods and cycle 1. Combining CBS-QB3 and CBS-APNO with the CPCM solvation methods **S1**, **S2**, and **S3** yields MUEs less than 0.42 pK<sub>a</sub> units and STDEVs less than 0.52 pK<sub>a</sub> units. Good results are also obtained with cycle 3 and all of the gas-phase and solvation methods, with MUEs ranging from 0.30 to 0.60 pK<sub>a</sub> units and STDEVs ranging from 0.40 to 0.82 pK<sub>a</sub> units. Because of standard state problems, we believe that this agreement is fortuitous. Use of cycle 4, which includes geometry relaxation in solution using the **S1** solvation procedure does not make much of a difference in the results. This finding makes sense since the CPCM method was parameterized based on gas phase geometries [50–58].

Overall the error analysis gives confidence that these methods can predict accurate pK<sub>a</sub> values, with the CBS-QB3 and CBS-APNO methods accurate to less than half a pK<sub>a</sub> unit when using thermodynamic cycle 1. Cycle 1 is more accurate than the other cycles as there are fewer terms (each term with its own uncertainty) in this cycle compared to cycles 2–4.

## Conclusion

The CBS-QB3 [34] and CBS-APNO [35, 36] methods can be combined with CPCM [50] continuum solvation methods to calculate pK<sub>a</sub> values for carboxylic acids accurate to within half a pK<sub>a</sub> unit [30]. Thermodynamic cycles that include an explicit water in the cycle are not accurate when the free energy of solvation of a water molecule is used, but appear to become accurate when the experimental free energy of vaporization of water is used [21, 28]. However, this apparent agreement seems to be an artifact of the standard states used in considering  $\Delta G_{\text{vap}}$  and  $\Delta G_s$ . Geometry relaxation in solution does not improve the results. The reason why cycle 2 is not accurate is not readily apparent, but it is clear that cycle 1 has fewer terms, and each term adds to the error in the overall calculation of a pK<sub>a</sub> value.

## ACKNOWLEDGMENTS

We thank Steve Feldgus and Ed Sherer for helpful discussions. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the ACS, and to the NIH, and to Hamilton

College for support of this work. M.D.L. acknowledges support from the Merck/AAAS Undergraduate Summer Research Program.

## References

1. Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J Am Chem Soc* 1987, 109, 6857–6858.
2. Jorgensen, W. L.; Briggs, J. M. *J Am Chem Soc* 1989, 111, 4190–4197.
3. Lim, C.; Bashford, D.; Karplus, M. *J Phys Chem* 1991, 95, 5610–5620.
4. Potter, M. J.; Gilson, M. K.; McCammon, J. A. *J Am Chem Soc* 1994, 116, 10298–10299.
5. Rajasekaran, E.; Jayaram, B.; Honig, B. *J Am Chem Soc* 1994, 116, 8238–8240.
6. Daasbjerg, K. *Acta Chemica Scandinavica* 1995, 49, 878–887.
7. Perakyla, M. *J Org Chem* 1996, 61, 7420–7425.
8. Figueirido, F.; Del Buono, G. S.; Levy, R. M. *J Phys Chem* 1996, 100, 6389–6392.
9. Li, J.; Fisher, C. L.; Chen, J. L.; Bashford, D.; Noodleman, L. *Inorg Chem* 1996, 35, 4694–4702.
10. Amyes, T. L.; Richard, J. P. *J Am Chem Soc* 1996, 118, 3129–3141.
11. Thibaudeau, C.; Plavec, J.; Chattopadhyaya, J. *J Org Chem* 1996, 61, 266–286.
12. Gao, J.; Li, N.; Freindorf, M. *J Am Chem Soc* 1996, 118, 4912–4913.
13. Schuurmann, G. *Quant Struct Act Relat* 1996, 15, 121–132.
14. Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, L.; Case, D. A. *Int J Quantum Chem* 1997, 61, 207–217.
15. Kallies, B.; Mitzner, R. *J Phys Chem B* 1997, 101, 2959–2967.
16. Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J Phys Chem A* 1997, 101, 10075–10081.
17. Perakyla, M. *J Am Chem Soc* 1998, 120, 12895–12902.
18. Chen, J.; McAllister, M. A.; Lee, J. K.; Houk, K. N. *J Org Chem* 1998, 63, 4611–4619.
19. Shapley, W. A.; Bacskey, G. B.; Warr, G. G. *J Phys Chem B* 1998, 102, 1938–1944.
20. Schuurmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J Phys Chem A* 1998, 102, 6706–6712.
21. da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J Phys Chem A* 1999, 103, 11194–11199.
22. Duarte, H. A.; Charvalho, S.; Paniago, E. B.; Simas, A. M. *J Pharm Sci* 1999, 88, 111–120.
23. Perakyla, M. *J Phys Chem Chem Phys* 1999, 1, 5643–5647.
24. Schuurmann, G. *Chem Phys Lett* 1999, 302, 471–479.
25. Topol, I. A.; Burt, S. K.; Rashin, A. A.; Erickson, J. W. *J Phys Chem A* 2000, 104, 866–872.
26. Billeter, S. R.; van Gunsteren, W. F. *J Phys Chem A* 2000, 104, 3276–3286.
27. Chen, I. J.; MacKerell, A. D. *Theor Chem Acct* 2000, 103, 483–494.
28. Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J Phys Chem A* 2000, 104, 2402–2409.

29. Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J Chem Phys* 2001, 114, 4595–4606.
30. Liptak, M. D.; Shields, G. C. *J Am Chem Soc* 2001, 123, 7314–7319.
31. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98 (Revision A.7)*; Gaussian: Pittsburgh, 1998.
32. Spartan version 5.1.3, Wavefunction, Inc., 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612, 1998.
33. Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Liotard, D. A.; Rinaldi, D.; Cramer, C. J.; Truhlar, D. J. *GAMESOL*, version 2.2.4, University of Minnesota, Minneapolis, 1999, based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J Comput Chem* 1993, 14, 1347.
34. Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J Chem Phys* 1999, 110, 2822–2827.
35. Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J Chem Phys* 1996, 104, 2598–2619.
36. Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J. *J Chem Phys* 1998, 109, 10570–10579.
37. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J Chem Phys* 1998, 109, 7764–7776.
38. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J Chem Phys* 1991, 94, 7221–7230.
39. Nyden, M. R.; Petersson, G. A. *J Chem Phys* 1981, 75, 1843–1862.
40. Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. *J Chem Phys* 1988, 89, 2193–2218.
41. Petersson, G. A.; Al-Laham, M. A. *J Chem Phys* 1991, 94, 6081–6090.
42. Petersson, G. A.; Tenfeldt, T. G.; Montgomery, J. A. *J Chem Phys* 1991, 94, 6091–6101.
43. Montgomery, J. A.; Ochterski, J. W.; Petersson, G. A. *J Chem Phys* 1994, 101, 5900–5909.
44. Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. *J Chem Phys* 1989, 90, 5622–5629.
45. Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. *J Chem Phys* 1990, 93, 2537–2545.
46. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J Chem Phys* 1999, 110, 4703–4709.
47. Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J Chem Phys* 1999, 110, 7650–7657.
48. Ochterski, J. W.; Petersson, G. A.; Wiberg, K. B. *J Am Chem Soc* 1995, 117, 11299–11308.
49. Petersson, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, J. A. *J Chem Phys* 1999, 109, 10570–10579.
50. Barone, V.; Cossi, M. *J Phys Chem A* 1998, 102, 1995–2001.
51. Miertus, S.; Scrocco, E.; Tomasi, J. *J Chem Phys* 1981, 55, 117–129.
52. Miertus, S.; Tomasi, J. *J Chem Phys* 1982, 65, 239–245.
53. Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *J Chem Phys Lett* 1996, 255, 327–335.
54. Cancès, M. T.; Mennucci, B.; Tomasi, J. *J Chem Phys* 1997, 107, 3032–3041.
55. Barone, V.; Cossi, M.; Mennucci, B.; Tomasi, J. *J Chem Phys* 1997, 107, 3210–3221.
56. Cossi, M. T.; Barone, V.; Mennucci, B.; Tomasi, J. *J Chem Phys Lett* 1998, 286, 253–260.
57. Barone, V.; Mennucci, B.; Tomasi, J. *J Chem Phys Lett* 1998, 286, 253–260.
58. Barone, V.; Cossi, M.; Tomasi, J. *J Comput Chem* 1998, 19, 404–417.
59. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem B* 1997, 101, 7147–7157.
60. Bartmess, J. E. In *NIST Chemistry WebBook*; NIST Standard References Database Number 69; Mallard, W. G.; Linstrom, P. J., Eds.; National Institute of Standards and Technology: Gaithersburg, MD, February 2000; <http://webbook.nist.gov>.
61. Taft, R. W.; Topsom, R. D. *Prog Phys Org Chem* 1987, 16, 1–83.
62. Cumming, J. B.; Kebarle, P. *Can J Chem* 1978, 56, 1–9.
63. Fujio, M.; McIver, Jr., R. T.; Taft, R. W. *J Am Chem Soc* 1981, 103, 4017–4029.
64. Caldwell, G.; Renneboog, R.; Kebarle, P. *Can J Chem* 1989, 67, 611–618.
65. Cai, J.; Topsom, R. D.; Headley, A. D.; Koppel, I.; Mishima, M.; Taft, R. W.; Veji, S. *J Mol Struct (Theochem)* 1988, 168, 141–146.
66. McQuarrie, D. M. *Statistical Mechanics*; Harper and Row: New York, 1970; p. 86.
67. Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J Chem Phys* 1999, 111, 10998–11014.
68. Christen, H. R.; Vogtle, F. *Organische Chemie—Von den Grundlagen zur Forschung (Organic Chemistry—From Fundamentals to Research)*, Vol. 1; Otto Salle: Frankfurt, Germany, 1988; p. 419.
69. Sargeant, E. P.; Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution*; IUPAC No. 23; Pergamon Press: New York, 1979.
70. The high-energy conformer (Fig. 1) ran successfully, but the low-energy conformer crashed repeatedly, with the error message: *PickT4: no shell combinations can fit!* The problem according to J. Ochterski at Gaussian, Inc., is that the estimates of how much memory is needed are off. We tried using many different values of MAXDISK, from 15 GB to 24 GB in increments of 0.5 GB but could not get the low-energy conformer to run with CBS-APNO.

71. Tawa, G. J.; Topol, I. A.; Burt, S. K.; Caldwell, R. A.; Rashin, A. A. *J Chem Phys* 1998, 109, 4852–4863.
72. Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J Phys Chem A* 1998, 102, 7787–7794.
73. Value found in Ref. [20].
74. Pearson, R. G. *J Am Chem Soc* 1986, 108, 6109–6114.
75. Ben-Naim, A.; Marcus, Y. *J Chem Phys* 1984, 81, 2016–2027.
76. Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987; Chapter 1.
77. Lide, D. R., Editor in chief, *CRC Handbook of Chemistry and Physics*, 72nd ed.; CRC Press: Boston, 1991.
78. Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J Phys Chem B* 2000, 104, 4726–4734.