

Accurate pK_a Calculations for Carboxylic Acids Using Complete Basis Set and Gaussian-n Models Combined with CPCM Continuum Solvation Methods

Matthew D. Liptak and George C. Shields*

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

Received February 27, 2001

Abstract: Complete Basis Set and Gaussian-n methods were combined with CPCM continuum solvation methods to calculate pK_a values for six carboxylic acids. 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 $G_{gas}(H^+)$ 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. 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.

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

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

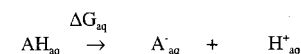
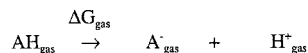
and since

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

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

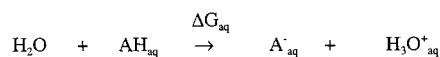
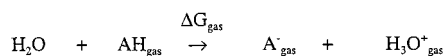
Calculating accurate pK_a values is demanding, as an error of 1.36 kcal/mol in ΔG° gives an error of 1 pK_a unit. There are at least three sources of error in pK_a calculations. The first is the model used to calculate pK_a , which generally involves a thermodynamic cycle such as:

Cycle 1



or

Cycle 2



In these cycles ΔG_{gas} is calculated with high level ab initio or density functional methods, and the ΔG_s values are calculated using a solvation method, typically a continuum dielectric approach. The pK_a values can be determined from eq 3, where $\Delta G^\circ = \Delta G_{aq}$. The second and third major errors stem from the accuracy of the calculations for ΔG_{gas} and ΔG_s . Relative pK_a

(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 Chem. Scand.* **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.

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 pK_a values accurate to within half a pK_a unit can be calculated for carboxylic acids.²⁹ In this paper we report accurate absolute pK_a calculations using thermodynamic cycle 1.

Methods

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.³⁰ 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 GAMESOL³³ 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 G2³⁸ model chemistry

(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.; J. 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.; Carvalho, S.; Paniago, E. B.; Simas, A. M. *J. Pharm. Sci.* **1999**, *88*, 111–120.

(23) Perkayla, M. *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. Acc.* **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) Sargeant E. P.; Dempsey, B. *Ionisation Constants of Organic Acids in Aqueous Solution*; IUPAC no. 23; Pergamon Press: New York, 1979.

(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, Inc.: Pittsburgh, PA, 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.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, S. Koseki, J. H.; 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.

methods to calculate G_{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 model chemistry is defined to include corrections for basis set truncation errors. The accurate calculation of molecular structures and energies requires convergence in the expansion of the basis set and in the degree of correlation, yet increases in the size of the basis set and in the degree of correlation raise the computational cost significantly. However, the main contribution to structure and energy is captured at the HF level, with correlation providing smaller (but critical) corrections to both structure and energy. CBS methods take advantage of the idea that at higher and higher levels of correlation, the contribution of the correction to the total energy can be determined to less accuracy than at lower levels of theory. Thus, these methods use relatively large basis sets for the structure calculation, medium sized basis sets for the second-order correlation correction, and small sized basis sets for higher order correlation corrections. Empirical corrections are also added as necessary.³⁵ 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. In the G2 model an initial geometry is calculated at the HF/6-31G(d) level and then harmonic frequencies are calculated and scaled by 0.8929 to give the zero-point energy. The geometry is then refined at the MP2(full)/6-31G(d) level, and this geometry is used for a series of single-point calculations (SPCs) at the MP2/6-311G(d,p), MP4/6-311G(d,p), QCISD/6-311G(d,p), MP2/6-311+G(d,p), MP4/6-311+G(d,p), and MP2/6-311+G(3df,2p) levels. These values are used, along with empirical corrections, to arrive at a final G2 energy. The G3 model chemistry uses the same geometries, but smaller basis sets for most of the SPCs, except for the last calculation, where the MP2/6-311+G(3df,2p) SPC is replaced by a MP2/G3 large calculation. The G3 large basis set has been modified to change the number of polarization functions used for first and second row atoms.³⁷ 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.

(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.

The geometries of all the stationary points and absolute energies in hartrees of each stationary point at each level of theory are available as Supporting Information.

We computed ΔG_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 (see refs 29 and 50 for a more complete discussion of the method). 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. In all of the CPCM calculations the area of the tesserae was set at 0.3 Å². All CPCM calculations were performed using Gaussian 98.³¹

Using different combinations of G_{gas} and ΔG_s values, and thermodynamic cycle 1, we calculated the absolute pK_a 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,⁵⁸ and computed the absolute pK_a values for each of the six acids.

In eqs 2 and 3, our ΔG° is ΔG_{aq} from the thermodynamic cycles. Therefore all of our pK_a calculations use the formula

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

For thermodynamic cycle 1,

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

and

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

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 ΔG_{gas} uses a reference state of 1 atm and the calculations of ΔG_s use a reference state of 1 M. Converting the ΔG_{gas} reference state (24.46 L at 298.15 K) from 1 atm to 1 M is accomplished using:

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

Using these values, the pK_a values using the first thermodynamic cycle are given by eq 8.

$$pK_a = [G(\text{A}^-_{\text{gas}}) - G(\text{AH}_{\text{gas}}) + \Delta G_s(\text{A}^-) - \Delta G_s(\text{AH}) - 269.0]/1.3644 \quad (8)$$

Results

The conformers of the carboxylic acids are displayed in Figure 1, along with the experimental pK_a values of each acid.³⁰ The conformationally averaged ΔG_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 1. 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 2 contains the calculated pK_a values, and

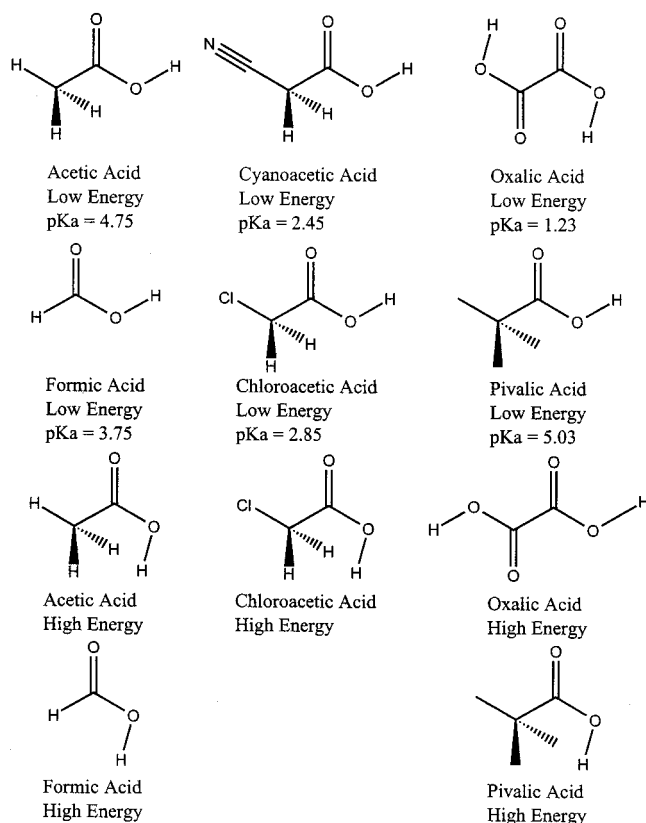


Figure 1. Experimental pK_a values and structures of the conformers of acetic, chloroacetic, cyanoacetic, formic, oxalic, and pivalic acids.

mean unsigned errors are from thermodynamic cycle 1 and eq 8. To simplify the discussion of results we have focused on the **S3** method. The results for all methods are available in Table 2a (Tables 2a and 2b are available as Supporting Information). Table 2b contains a pK_a error analysis of each of five gas-phase computational methods combined with three condensed-phase methods. To better understand the errors in pK_a we analyzed the errors in the gas-phase and solvation calculations separately. Table 3 shows the errors for the gas-phase calculation of ΔG_{gas} for cycle 1 using the Complete Basis Set and Gaussian-n methods for the five carboxylic acids where experimental data are available.^{59–64} Table 4 contains the errors for the calculation of $\Delta \Delta G_{\text{sol}}$ for cycle 1 using the CPCM/HF/6-31G(d)//HF/6-31G(d), CPCM/HF/6-31+G(d)//HF/6-31G(d), and CPCM/HF/6-31+G(d)//HF/6-31+G(d) solvation procedures. The experimental values for $\Delta \Delta G_{\text{sol}}$ were obtained from the experimental values³⁰ for ΔG_{aq} ($\Delta G_{\text{aq}} = 2.303RTpK_a$) and ΔG_{gas} ,^{59–64} as $\Delta \Delta G_{\text{sol}} = \Delta G_{\text{aq}} - \Delta G_{\text{gas}}$.

Discussion

Review of Errors in Gas Phase and Condensed-Phase Calculations from Relative pK_a Work.

In our previous work

(50) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
 (51) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.

(52) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239–245.

(53) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327–335.

(54) Cancès, M. T.; Mennucci, V.; 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. *Chem. Phys. Lett.* **1998**, *286*, 253–260.

(57) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404–417.

(58) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147–7157.

(59) Bartmess, J. E. Negative Ion Energetics Data. In *NIST Chemistry WebBook*; NIST Standard Reference Database No. 69; Mallard, W. G., Linstrom, P. J., Ed.; National Institute of Standards and Technology: Gaithersburg, MD, February 2000 (<http://webbook.nist.gov>).

(60) Taft, R. W.; Topsom, R. D. *Prog. Phys. Org. Chem.* **1987**, *16*, 1–83.

(61) Cumming, J. B.; Kebarle, P. *Can. J. Chem.* **1978**, *56*, 1–9.

(62) Fujio, M.; McIver, R. T., Jr.; Taft, R. W. *J. Am. Chem. Soc.* **1981**, *103*, 4017–29.

(63) Caldwell, G.; Renneboog, R.; Kebarle, P. *Can. J. Chem.* **1989**, *67*, 611–18.

(64) Cai, J.; Topsom, R. D.; Headley, A. D.; Koppel, I.; Mishima, M.; Taft, R. W.; Veji, S. *J. Mol. Struct. (THEOCHEM)* **1988**, *168*, 141–6.

Table 1. Conformationally Averaged DGs Values for Six Carboxylic Acids and Their Anions Using CBS-QB3 (kcal/mol)

	acetic	formic	cyanoacetic	chloroacetic	oxalic	pivalic	water
S1^b	-6.93	-7.39	-13.24	-9.67	-11.56	-5.97	-6.37
S2^c	-7.65	-8.38	-14.32	-10.56	-12.58	-6.70	-7.26
S3^d	-7.72	-8.43	-14.52	-10.65	-12.73	-6.78	-7.23
	acetate	formate	cyanoacetate	chloroacetate	oxalate	pivalate	H ⁺ ^a
S1^b	-76.58	-76.15	-68.72	-69.48	-74.92	-71.04	-264.61
S2^c	-77.18	-76.58	-69.51	-70.11	-75.35	-71.97	-264.61
S3^d	-77.58	-77.10	-69.99	-70.57	-75.72	-72.42	-264.61

^a From experimental values for acetic acid, see text for details. ^b **S1**: CPCM/HF/6-31G(d)//HF/6-31G(d). ^c **S2**: CPCM/HF/6-31+G(d)//HF/6-31G(d). ^d **S3**: CPCM/HF/6-31+G(d)//HF/6-31+G(d).

Table 2. pK_a Values Using CBS and Gaussian-n Models, CPCM Solvation Calculations, and Cycle 1

acetic acid $pK_a = 4.75$				formic acid $pK_a = 3.75$				cyanoacetic acid $pK_a = 2.45$			
solvation ^a	$\Delta\Delta G_{sol}$	pK_a	difference	solvation ^a	$\Delta\Delta G_{sol}$	pK_a	difference	solvation ^a	$\Delta\Delta G_{sol}$	pK_a	difference
	CBS-QB3				CBS-QB3				CBS-QB3		
S3	-334.47	5.19	0.44	S3	-333.28	3.53	-0.22	S3	-320.08	2.34	-0.11
	CBS-APNO				CBS-APNO				CBS-APNO		
S3	-334.47	4.91	0.16	S3	-333.28	3.93	0.18	S3	-320.08	2.31	-0.14
	G2				G2				G2		
S3	-334.47	4.43	-0.32	S3	-333.28	3.35	-0.40	S3	-320.08	1.97	-0.48
	G2MP2				G2MP2				G2MP2		
S3	-334.47	4.72	-0.03	S3	-333.28	3.70	-0.05	S3	-320.08	2.27	-0.18
	G3				G3				G3		
S3	-334.47	5.07	0.32	S3	-333.28	4.23	0.48	S3	-320.08	2.61	0.16
	chloroacetic acid $pK_a = 2.85$				oxalic acid $pK_a = 1.23$				pivalic acid $pK_a = 5.03$		
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				CBS-QB3				CBS-QB3		
S3	-324.53	3.36	0.51	S3	-327.60	1.20	-0.03	S3	-330.25	5.19	0.16
	CBS-APNO				CBS-APNO				CBS-APNO		
	method not available for chlorine			S3	-327.60	1.51	0.28		see ref 68		
	G2				G2				G2		
S3	-324.53	3.13	0.28	S3	-327.60	1.42	0.19	S3	-330.25	6.10	1.07
	G2MP2				G2MP2				G2MP2		
S3	-324.53	3.41	0.56	S3	-327.60	1.79	0.56	S3	-330.25	6.34	1.31
	G3				G3				G3		
S3	-324.53	3.61	0.76	S3	-327.60	2.26	1.03	S3	-330.25	6.64	1.61

^a **S3**: CPCM/HF/6-31+G(d)//HF/6-31+G(d).

on relative pK_a calculations of the same six carboxylic acids²⁹ we analyzed the errors in the gas-phase and condensed-phase calculations. We compared the relative gas-phase values (Table 9 of ref 29, where we have canceled the role of H⁺ by comparing two acids, HA and HB, pairwise) against the NIST database,^{59–64} and we found that the mean unsigned errors and standard deviations (MUE,STDEV) for the CBS-QB3, CBS-APNO, G2, G2MP2, and G3 methods were (0.4,0.5), (0.5,0.7), (0.8,1.1), (0.8,1.0), and (0.8,1.0), respectively. The relative $\Delta\Delta G_{sol}$ experimental and calculated values (Tables 10 and 11 of ref 29, again comparing two acids AH and BH pairwise, so that the values for $\Delta G_s(H^+)$ canceled) were compared for CPCM/6-31G(d)//HF/6-31G(d) (**S1**), CPCM/6-31+G(d)//HF/6-31G(d) (**S2**), CPCM/6-31+G(d)//HF/6-31+G(d) (**S3**), CPCM/6-31G(d)//QCISD/6-311G(d,p), and CPCM/6-31+G(d)//QCISD/6-311G(d,p). The MUEs and STDEVs (MUE,STDEV) were (0.8,0.9), (0.5,0.6), (0.5,0.7), (0.5,0.7), and (0.4,0.6). These results lead naturally to the prediction that the CBS-QB3 model used for the calculation of ΔG_{gas} , combined with solvation methods that include diffuse functions, should give the most accurate relative pK_a values. In this study we have used the previously reported²⁹ gas-phase and condensed-phase values for the six carboxylic acids and their anions, along with experimental numbers for $G_{gas}(H^+)$ and $\Delta G_s(H^+)$, and eq 8 for absolute pK_a calculations. An absolute error analysis of the present work is reported in Tables 2b–4 and discussed later in this paper.

Thermodynamic Cycle 1. Table 1 contains the conformationally averaged values for ΔG_s . The absolute pK_a values are presented in Table 2 for thermodynamic cycle 1 using the values in Table 1 and the previously reported²⁹ values for G_{gas} for the acids and the anions. The value of -6.28 comes from the Sackur–Tetrode equation⁶⁵ to evaluate the entropy combined with inclusion of translational energy at 298 K.⁶⁶ The value for $\Delta G_s(H^+)$ was obtained using the average of three published experimental values for the gas-phase dissociation of acetic acid in thermodynamic cycle 1,^{60–62} the experimental ΔG_s values for acetic acid (-6.69 kcal/mol) and acetate ion (-77 kcal/mol),⁶⁷ and the experimental value for ΔG_{aq} ($2.303RTpK_a = 6.48$ kcal/mol).³⁰

The CBS-QB3 gas-phase method and the **S3** solvation procedure predict absolute pK_a values of 5.19, 3.53, 2.34, 3.36, 1.20, and 5.19 for acetic, formic, cyanoacetic, chloroacetic, oxalic, and pivalic acids, respectively (Table 2). The largest deviation is 0.51 pK_a units, and the MUE for the six acids is 0.24 pK_a units using this combination of gas-phase and condensed-phase methods. Results for the other gas-phase

(65) McQuarrie, D. M. *Statistical Mechanics*; Harper and Row: New York, 1970; p 86.

(66) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Chem. Phys.* **1999**, *111*, 10998–11014.

(67) Christen, H. R.; Vogtle, F. *Organische Chemie-Von den Grundlagen zur Forschung*, Organic Chemistry-From Fundamentals to Research; Otto Salle Verlag: Frankfurt, Germany, 1988; Vol. 1, p 419.

Table 3. Analysis of Errors in Gas-Phase Calculations Using Thermodynamic Cycle 1

Gas-Phase Calculations, in kcal/mol: AH → A ⁻ + H ⁺						
	CBS-QB3	CBS-APNO	G2	G2MP2	G3	
acetic acid	341.6	341.2	340.5	340.9	341.4	
formic acid	338.1	338.6	337.9	338.3	339.1	
cyanoacetic acid	323.3	323.2	322.8	323.2	323.6	
chloroacetic acid	329.1		328.8	329.2	329.5	
pivalic acid	337.3		338.6	338.9	339.3	
Experimental Values, in kcal/mol: ^a AH → A ⁻ + H ⁺						
	exp. 1 ^b	exp. 2 ^c	exp. 3 ^d	exp. 4 ^e	exp. 5 ^f	av. exp.
acetic acid	341.1	341.5	341.7			341.4
formic acid		338.2	338.4	338.3		338.4
cyanoacetic acid	323.7					323.7
chloroacetic acid		328.8	328.8	329.0		328.9
pivalic acid				337.6	338.0	337.8
Error Evaluation, in kcal/mol: AH → A ⁻ + H ⁺						
	CBS-QB3	CBS-APNO	G2	G2MP2	G3	
acetic acid	0.1	-0.3	-0.9	-0.5	-0.0	
formic acid	-0.2	0.3	-0.5	-0.0	0.7	
cyanoacetic acid	-0.4	-0.5	-0.9	-0.5	-0.1	
chloroacetic acid	0.2		-0.1	0.3	0.6	
pivalic acid	-0.5		0.8	1.1	1.5	
MSE	-0.2	-0.1	-0.3	0.1	0.5	
MUE	0.3	0.3	0.6	0.5	0.6	
STDEV	0.4	0.4	0.8	0.7	0.9	

^a Reference 59, all values corrected to $T = 298$ K. ^b Reference 60. ^c Reference 61. ^d Reference 62. ^e Reference 63. ^f Reference 64.

Table 4. Analysis of Errors in Solvation Calculations for Thermodynamic Cycle 1

$\Delta\Delta G_{\text{sol}}$ Calculations Using CBS-QB3 Conformational Averaging, in kcal/mol				
	S1 ^a	S2 ^b	S3 ^c	exptl
acetic	-334.26	-334.14	-334.47	-334.92
formic	-333.37	-332.81	-333.28	-333.28
cyanoacetic	-320.09	-319.8	-320.08	-320.36
chloroacetic	-324.42	-324.16	-324.53	-325.01
pivalic	-329.68	-329.88	-330.25	-330.94
Error Evaluation, in kcal/mol				
	S1 ^a	S2 ^b	S3 ^c	
acetic	0.7	0.8	0.4	
formic	-0.1	0.5	0.0	
cyanoacetic	0.3	0.6	0.3	
chloroacetic	0.6	0.9	0.5	
pivalic	1.3	1.1	0.7	
MSE	0.5	0.7	0.4	
MUE	0.6	0.7	0.4	
STDEV	0.8	0.9	0.5	

^a S1: CPCM/HF/6-31G(d)//HF/6-31G(d). ^b S2: CPCM/HF/6-31+G(d)//HF/6-31G(d). ^c S3: CPCM/HF/6-31+G(d)//HF/6-31+G(d).

methods are similar. The CBS-APNO method is quite accurate, although it failed for one conformer of pivalic acid.⁶⁸ 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. The pK_a values

(68) The high-energy conformer (Figure 1) ran successfully but the low-energy conformer crashed repeatedly, with the following 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 15GB to 24GB in increments of 0.5 GB, but could not get the low-energy conformer to run with CBS-APNO.

calculated using the S1 and S2 solvation methods can be found in Table 2a in the Supporting Information.

Comparison with Previous Work and Comments on $\Delta G_s(\text{H}^+)$ and $G(\text{H}^+)$. The accuracy of cycle 1 and eq 8 for calculations of pK_a values by these methods allows for a comparison of previous work that used cycle 1.^{3,9,14,15,19,25} A key problem for previous workers trying to make absolute pK_a predictions is the value to use for $\Delta G_s(\text{H}^+)$. Karplus and co-workers outlined the problem in their work on models for ionizable side chains of amino acids, pointing out that the range of experimental $\Delta G_s(\text{H}^+)$ values based on the absolute potential correction for the standard hydrogen standard electrode is from -254 to -261 kcal/mol, and they used a value of -262.2 kcal/mol in their own work.³ The average value for five separate measurements of the standard hydrogen electrode is -259.5 kcal/mol, which is the value used by two other research groups to determine the pK_a values of various aliphatic, alicyclic, and aromatic amines,¹⁵ and for hydroxybenzoic acids.¹⁹ Noodleman and co-workers use a value of -260.5 kcal/mol for their pK_a studies of hydrated transition metal cations⁹ and organic acids.¹⁴ Topol et al. use a value of -262.5 kcal/mol for $\Delta G_s(\text{H}^+)$ in their calculation of pK_a values for pyrone and dihydropyrones,²⁵ based on their calculated aqueous solvation energy of the proton.⁶⁹ Their calculation of $\Delta G_s(\text{H}^+)$ was determined by adding explicit waters of hydration coupled to a continuum dielectric, with the hydration free energy converging for four to six waters of hydration at -262.23 kcal/mol.⁶⁹ This led them to conclude that the proton hydration free energy is at the lower end of the range of values proposed in the literature.⁶⁹ A recent examination of the experimental data by Tissandier et al. leads to an even lower value for $\Delta G_s(\text{H}^+)$.⁷⁰ They used a cluster-pair-based approximation to determine $\Delta G_s(\text{H}^+)$ to be -263.98 ± 0.07 kcal/mol. The uncertainty in this value of ±0.07 is smaller than expected because the cluster data of 20 different pairings of oppositely charged ions are folded into the determination.⁷⁰ The difference between the recalculated value of -264 kcal/mol of Tissandier et al. 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_a calculations as large as one to three pK_a units! It is understandable that previous work on pK_a calculations using thermodynamic cycle 1 is not highly accurate,^{3,9,14,15,19} with the best results coming from calculations by Topol et al., who state that their relative pK_a calculations are robust but their absolute pK_a calculations may have fortuitous agreement with experiment.²⁵ A recent report used B3LYP in combination with a Poisson-Boltzmann continuum solvation method to calculate pK_a values for 5-substituted uracil derivatives.⁷¹ Excellent correlation was obtained by varying the dielectric constant of the solute region (0.92) which led to a value of $\Delta G_s(\text{H}^+)$ of -258.32 kcal/mol. The value of -258.32 minimizes the root-mean-square deviation between the calculated and experimental pK_a values for the substituted 5-uracils.⁷¹ Because the method was scaled to obtain the best correlation, other values of the dielectric constant and ΔG_s yield similar results (Appendix of ref 71). We derive a value of -264.61 kcal/mol from the experimental thermodynamic cycle of acetic acid, where all quantities are known except for $\Delta G_s(\text{H}^+)$. As Table 2 shows, the correct value for $\Delta G_s(\text{H}^+)$ must be in the range of -264 kcal/mol.

(69) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Caldwell, R. A.; Rashin, A. A. *J. Chem. Phys.* **1998**, *109*, 4852-4863.

(70) 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.

(71) Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105*, 274-280.

There is less uncertainty in the value of $G(H^+)$. The Sackur–Tetrode equation⁶⁵ was used to evaluate the entropy, which yields a value of -7.76 kcal/mol for $TS(H^+)$ at 298 K and 1 atm of pressure. Including translational energy, the total gas-phase free energy of the proton is -6.28 kcal/mol at 298 K.⁶⁶ To check this number we also evaluated $G_{\text{gas}}(H^+)$ from data available on the NIST web site combined with the high-level CBS-QB3 calculations (which have a MUE and STDEV of 0.3 and 0.4 kcal/mol, respectively, Table 3) of the acids and their anions. The gas-phase values for ΔG for the reaction $AH \rightarrow H^+ + A^-$ for five of the six carboxylic acids we have studied are known,^{59–64} and the CBS-QB3 values for $G_{\text{gas}}(AH)$ and $G_{\text{gas}}(A^-)$ for the same five carboxylic acids have been tabulated.²⁹ Five separate determinations of $G_{\text{gas}}(H^+)$ yielded the final value of -6.09 ± 0.5 kcal/mol. Using either value of $G_{\text{gas}}(H^+)$ works well for pK_a calculations on these carboxylic acids.

Error Analysis. Table 2b (Supporting Information) contains a summary of the errors obtained using thermodynamic cycle 1 combined with the CBS and Gaussian-n gas phase and **S1**, **S2**, and **S3** solvation procedures, for the five acids where ΔG_{gas} is known experimentally. This analysis reveals that the most accurate results for the absolute calculation of pK_a values for formic, acetic, cyanoacetic, chloroacetic, oxalic, and pivalic acids are obtained using CBS methods. Combining CBS-QB3 and CBS-APNO with the CPCM solvation methods **S1**, **S2**, and **S3** yields MUEs less than 0.42 pK_a units and STDEVs less than 0.52 pK_a units.

To better understand the pK_a results using thermodynamic cycle 1, we analyzed the gas-phase and solvation calculations separately. In Table 3 we have the calculated gas-phase values for each acid dissociating into its anion and H^+ , where we have used the value for $G(H^+)$ of -6.28 kcal/mol, as discussed previously. The calculated values are compared with the experimental values available from the NIST database.^{59–64} The MUEs are 0.3 kcal/mol for the CBS-QB3 and CBS-APNO (only 3 data points) methods, 0.6 kcal/mol for the G2 and G3 methods, and 0.5 kcal/mol for the G2MP2 method. The STDEVs are 0.4, 0.4, 0.8, 0.7, and 0.9 kcal/mol for the CBS-QB3, CBS-APNO, G2, G2MP2, and G3 model chemistries, respectively.

The solvation results are displayed in Table 4. Here we have tabulated the experimental values for $\Delta\Delta G_{\text{sol}}$ along with the calculated values, where we have used the CBS-QB3 method in the conformational analysis procedure.^{29,58} These results are the same when the other gas-phase methods are used in the conformational analysis procedure. Here we find that the **S3** procedure (CPCM/HF/6-31+G(d)//HF/6-31+G(d)) is slightly better than the **S1** procedure (CPCM/HF/6-31G(d)//HF/6-31G(d)), which is slightly better than the **S2** procedure (CPCM/HF/6-31+G(d)//HF/6-31G(d)). The **S3** procedure has a MUE of 0.4 kcal/mol and a STDEV of 0.5 kcal/mol. We would expect, based on simple propagation of errors, that the error in ΔG_{aq} for combining the CBS-QB3 method with the **S3** method would be 0.5 kcal/mol, which would produce an error in pK_a of

approximately 0.4 pK_a units. For the G2 method the same analysis predicts an error in ΔG_{aq} of 0.73 kcal/mol, which would produce an error of approximately half a pK_a unit. Table 3 shows that the G3 method predicts the most positive gas-phase energies (MSE = 0.5 kcal/mol), which contributes to making the calculated pK_a too large. Overall the error analysis gives confidence that these methods can predict accurate pK_a values, with the CBS-QB3 and CBS-APNO methods accurate to less than half a pK_a unit when using thermodynamic cycle 1. We have also used cycle 2 to calculate pK_a values, and our experiments with other thermodynamic cycles show that the classic cycle is the best for pK_a calculations for carboxylic acids.⁷²

Conclusion

The CBS-QB3³⁴ and CBS-APNO^{35,36} methods can be combined with CPCM⁵⁰ continuum solvation methods to calculate pK_a values for carboxylic acids accurate to within half a pK_a unit. These calculations use a classic thermodynamic cycle for proton dissociation, with values of -6.28 kcal/mol for $G_{\text{gas}}(H^+)$ and -264.61 kcal/mol for $\Delta G_s(H^+)$. The value for $G_{\text{gas}}(H^+)$ from the Sackur–Tetrode equation^{65,66} is within 0.19 kcal/mol of the value obtained from the gas-phase deprotonation data available from NIST.^{59–64} The value of -264.61 kcal/mol for $\Delta G_s(H^+)$ is an experimental number that comes from the acetic acid system.^{30,60–62,67} This value is more negative than previous values used for pK_a calculations, but is only 0.63 kcal/mol lower than a recently redetermined value from a cluster-pair-based approximation method.⁷⁰ The accuracy of our calculations suggests that the actual value for $\Delta G_s(H^+)$ is between -264 and -265 kcal/mol. The CBS gas-phase methods^{34–36} are very accurate, with mean unsigned errors of 0.3 kcal/mol and standard deviations of 0.4 kcal/mol. The CPCM solvation calculations are slightly less accurate, with the largest basis sets leading to a mean unsigned error and standard deviation of 0.4 and 0.5 kcal/mol, respectively.

Acknowledgment. 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, 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.

Supporting Information Available: Optimized geometries and energies for all stationary points discussed in the text and in Table 1 of ref 29, and Tables 2a and 2b, which contain pK_a values for the **S1** and **S2** solvation methods and an error analysis comparison of all three solvation methods (PDF). This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA010534F

(72) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.*, in press.