J|A|C|S
ARTICLES
Published on Web 05/08/2002

Absolute pK_a Determinations for Substituted Phenols

Matthew D. Liptak,[†] Kevin C. Gross,[‡] Paul G. Seybold,[‡] Steven Feldgus,[†] and George C. Shields*[,†]

Contribution from the Department of Chemistry, Hamilton College, 198 College Hill Road, Clinton, New York 13323, and Department of Chemistry, Wright State University, Dayton, Ohio 45435

Received November 5, 2001. Revised Manuscript Received April 20, 2002

Abstract: The CBS-QB3 method was used to calculate the gas-phase free energy difference between 20 phenols and their respective anions, and the CPCM continuum solvation method was applied to calculate the free energy differences of solvation for the phenols and their anions. The CPCM solvation calculations were performed on both gas-phase and solvent-phase optimized structures. Absolute pK_a calculations with solvated phase optimized structures for the CPCM calculations yielded standard deviations and root-mean-square errors of less than 0.4 pK_a unit. This study is the most accurate absolute determination of the pK_a values of phenols, and is among the most accurate of any such calculations for any group of compounds. The ability to make accurate predictions of pK_a values using a coherent, well-defined approach, without external approximations or fitting to experimental data, is of general importance to the chemical community. The solvated phase optimized structures of the anions are absolutely critical to obtain this level of accuracy, and yield a more realistic charge separation between the negatively charged oxygen and the ring system of the phenoxide anions.

Introduction

The ability to determine the equilibrium constant for the ionization of an acid in water, K_a (or the negative logarithm, pK_a), from first principles would be very useful to the chemical community. Accurate predictions of pK_a values using a coherent, well-defined approach, without external approximations or fitting to experimental data, are of general significance. The acid/base behavior of phenols is of inherent importance in environmental and biochemical contexts. Four groups have published papers where they correlate experimental pK_a values for phenols with quantum mechanical parameters,[1−4] but no one has used gas-phase and liquid-phase quantum chemical calculations to predict absolute pK_a values for phenols. We have recently published calculations of pK_a values accurate to within half a pK_a unit for a set of six carboxylic acids.[5,6] In the present work, we extend our predictions to a set of 20 phenols.

The definition of pK_a is

$$\mathrm{p}K_a = -\log K_a \tag{1}$$

and since

$$\Delta G^\circ = -2.303RT \log K_a \tag{2}$$

$$\mathrm{p}K_a = \Delta G^\circ/2.303RT \tag{3}$$

The calculation of accurate pK_a values is demanding, as a 1.36 kcal/mol error in ΔG° results in an error of 1 pK_a unit. In a previous work we have extensively examined different treatments of H^+ and H_2O in a thermodynamic cycle for pK_a calculations. We concluded that the following cycle is the most effective.[7]

$$\begin{array}{ccccc}
AH_{gas} & \xrightarrow{\Delta G_{gas}} & A^-_{gas} & + & H^+_{gas} \\
\uparrow -\Delta G_s(AH) & & \downarrow \Delta G_s(A^-) & & \downarrow \Delta G_s(H^+) \\
AH_{aq} & \xrightarrow{\Delta G_{aq}} & A^-_{aq} & + & H^+_{aq}
\end{array}$$

In this cycle we calculate ΔG_{gas} with the CBS-QB3 model,[8] which has been shown to be accurate to within 1.2 kcal/mol when compared against gas-phase deprotonation reactions in the National Institute of Technology (NIST) database.[9,10] The ΔG_s values were calculated using CPCM continuum solvation methods.[11] The absolute pK_a values are calculated using the following equation derived from eq 3 where $\Delta G^\circ = \Delta G_{aq}$,

* Corresponding author. E-mail: gshields@hamilton.edu.
[†] Hamilton College.
[‡] Wright State University.

(1) Gruber, C.; Buss, V. *Chemosphere* **1989**, *19*, 1595−1609.
(2) Schuurmann, G. *Quant. Struct.-Act. Relat.* **1996**, *15*, 121−132.
(3) Hanai, T.; Koizumi, K.; Kinoshita, T. *J. Liq. Chromatogr. Relat. Technol.* **2000**, *23*, 363−385.
(4) Gross, K. C.; Seybold, P. G. *Int. J. Quantum Chem.* **2001**, *85*, 569−579.
(5) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114*, 4595−4606.
(6) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314−7319.
(7) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727−741.
(8) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822−2827.
(9) Bartmess, J. E., Ed. *Negative Ion Energetics Data*; National Institute of Standards and Technology: Gaithersburg, MD, 2000; http://webbook.nist.gov.
(10) Pokon, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, *105*, 10483−10487.
(11) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

$$pK_a = \Delta G_{aq}/2.303RT \tag{4}$$

For our thermodynamic cycle,

$$\Delta G_{aq} = \Delta G_{gas} + \Delta\Delta G_{sol} \tag{5}$$

where

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

The value for $G_{gas}(H^+)$, −6.28 kcal/mol, comes from the Sackur−Tetrode equation,[12] and the value for $\Delta G_s(H^+)$, −264.61 kcal/mol, was derived from experiment (see discussion).[6] The calculation of ΔG_{gas} uses a reference state of 1 atm and the calculation of ΔG_s uses a 1 M reference state. Converting the ΔG_{gas} reference state (24.46 L at 298.15 K) from 1 atm to 1 M is accomplished by using

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

Using these numbers, the pK_a values reported in this work are calculated with eq 8.

$$pK_a = \{G(A^-{}_{gas}) - G(AH_{gas}) + \Delta G_s(A^-) - \Delta G_s(AH) - 269.0\}/1.3644 \tag{8}$$

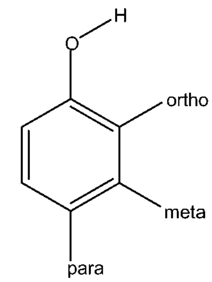
In this paper we report on our absolute calculations for phenol and the meta-, para-, and ortho-substituted aminophenols, chlorophenols, cyanophenols, fluorophenols, hydroxyphenols, methoxyphenols, methylphenols, and nitrophenols. The approach outlined in this work is a general one, using well-defined methods, without fitting or correlating data to obtain "correct" results. We provide a realistic assessment of the errors for each different method used in this study.

Methods

We studied 20 different substituted phenols ranging in size from 50 to 66 electrons (Figure 1). All calculations were performed on SGI Origin 200 servers with Gaussian 98.[13] Initial geometry optimizations were performed with B3LYP/CBSB7, identical to the first step in a CBS-QB3[8] calculation. CBS-QB3 was used to obtain the values for G_{gas} at 298.15 K for a 1 atm standard state. HF/6-31G(d) and HF/6-31+G(d) geometry optimizations were also performed on each species, and the absence of negative frequencies verified that all structures were true minima.

One set of solvated phase geometries was obtained with the CPCM/HF/6-31+G(d) method for each phenol by using the B3LYP/CBSB7 gas-phase geometry as a starting point. In one case, *p*-aminophenoxide, the solvated phase optimization produced a lower energy conformer than the gas-phase optimization. For this species, a CBS-QB3 calculation was performed on the lowest energy solvated phase geometry, and the difference between the lowest energy gas-phase conformation and the lowest energy solvated phase conformation was used to correct ΔG_s.

Phenol Backbone



R Groups

-NH_2	Amino-
-Cl	Chloro-
-CN	Cyano-
-F	Fluoro-
-OH	Hydroxy-
-CH_2OH	Methoxy-
-CH_3	Methyl-
-NO_2	Nitro-

Figure 1. Phenol backbone and the various R groups used in this work.

In addition, convergence criteria on *o*-nitrophenoxide needed to be loosened (RMS force of 0.0017 au; OPT = LOOSE) to achieve convergence in the solvated optimization.

The CBS-QB3 model[8] is one of the Complete Basis Set models developed by Petersson and co-workers.[8,14−18] In these models, a series of calculations are made on a particular geometry, and a complete basis set quadratic configuration interaction model chemistry is defined to include corrections for basis set truncation errors. These methods use fairly 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. This method has been described in more detail in previous papers,[5−7] and has been shown to be accurate to within 1 kcal/mol when compared to the most accurate experimental data for gas-phase deprotonation reactions.[10]

Free energy of solvation values were calculated by using the CPCM[19] implicit solvation model,[11] based on the polarized continuum model (PCM) of Tomasi and co-workers.[20−27] In the Barone and Cossi implementation of this method the cavities are modeled on the molecular shape, using optimized parameters, and both electrostatic and nonelectrostatic contributions to the energies are included. The solute molecules are embedded in cavities using interlocking spheres, and the surface is smoothed. The cavity surface is divided into small domains, known as tesserae, by projecting onto the surface the faces of suitable polyhedra. The total free energy is the sum of the free energies obtained for electrostatic interactions, formation of the cavity in the continuum medium, dispersion interactions, and repulsion interactions. The dispersion and repulsion terms are calculated following the procedure pioneered by Floris and Tomasi[28,29] so that the nonelectrostatic terms in the Barone and Cossi CPCM method are exactly the same as for the PCM method provided that equal sized cavities are used in each case.

(12) McQuarrie, D. M. *Statistical Mechanics*; Harper and Row: New York, 1970.
(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; 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.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
(14) Petersson, G. A.; Allaham, M. A. *J. Chem. Phys.* **1991**, *94*, 6081−6090.
(15) Petersson, G. A.; Tensfeldt, T. G.; Montgomery, J. A. *J. Chem. Phys.* **1991**, *94*, 6091−6101.
(16) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. *J. Chem. Phys.* **1996**, *104*, 2598−2619.
(17) Montgomery, J. A.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1994**, *101*, 5900−5909.
(18) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532−6542.
(19) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(20) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117.
(21) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, *65*, 239.
(22) Cossi, M.; Barone, V.; Cammi, R.; Tomasi, J. *Chem. Phys. Lett.* **1996**, *255*, 327−335.
(23) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032−3041.
(24) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210−3221.
(25) Cossi, M.; Barone, V.; Mennucci, B.; Tomasi, J. *Chem. Phys. Lett.* **1998**, *286*, 253−260.
(26) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404−417.
(27) Mennucci, B.; Cammi, R.; Tomasi, J. *J. Chem. Phys.* **1998**, *109*, 2798−2807.
(28) Floris, F.; Tomasi, J. *J. Comput. Chem.* **1989**, *10*, 616−627.
(29) Floris, F. M.; Tomasi, J.; Ahuir, J. L. P. *J. Comput. Chem.* **1991**, *12*, 784−791.

Table 1. CBS-QB3 Energies (in hartrees)

	acid	anion
phenol	−306.968099	−306.412675
m-aminophenol	−362.248941	−361.692804
p-aminophenol	−362.246537	−361.686573
o-chlorophenol	−766.135761	−765.589834
m-chlorophenol	−766.133728	−765.591353
p-chlorophenol	−766.133090	−765.587991
m-cyanophenol	−399.095423	−398.562773
p-cyanophenol	−399.096184	−398.568500
m-fluorophenol	−406.137560	−405.591989
p-fluorophenol	−406.136153	−405.584605
m-hydroxyphenol	−382.119456	−381.568684
p-hydroxyphenol	−382.117048	−381.560697
m-methoxyphenol	−421.332371	−420.779768
p-methoxyphenol	−421.330159	−420.774032
o-methylphenol	−346.202215	−345.648593
m-methylphenol	−346.201903	−345.646543
p-methylphenol	−346.201547	−345.644834
o-nitrophenol	−511.269946	−510.733557
m-nitrophenol	−511.266427	−510.734710
p-nitrophenol	−511.267116	−510.745914

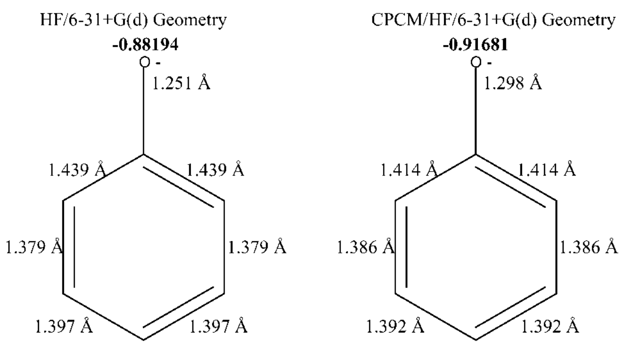


Figure 2. Phenoxide geometries for gas-phase optimized (left) and solvation phase optimized (right) structures. Natural charges on the oxygen are in bold.

The CPCM calculations were performed with the HF/6-31G(d) and HF/6-31+G(d) basis sets on the gas-phase B3LYP/CBSB7, HF/6-31G-(d), and HF/6-31+G(d) geometries and CPCM/HF/6-31+G(d) solvated phase geometry yielding a total of eight different solvation procedures. Additional test calculations with larger basis sets and with MP2 theory were also made. The CPCM calculations were performed with default parameters. All of the final geometries and energies are available as Supporting Information.

Results

The CBS-QB3 gas-phase free energies (in au) for all of the acids and their corresponding anions are given in Table 1 for a standard state of 1 atm. The ΔG_s values for all of the acids and the anions are reported in Table S2 (standard state of 1 M), using eight different solvation procedures, **S1**−**S8**. Procedures **S1**−**S6** use the B3LYP and HF gas-phase structures, while **S7** and **S8** use the CPCM/HF/6-31+G(d) optimized geometries. Geometry optimization in solution has little effect on the phenols, but has a significant effect on the phenoxide anions (Figure 2). Table S2 is available as Supporting Information and Table 2 contains the results for the CPCM optimized geometries along with available experimental values. Table S3 gives the pK_a values calculated with CBS-QB3 gas-phase calculations and the eight different solvation procedures. Table S3 is available as Supporting Information and Table 3 contains the results for the **S7** and **S8** solvation procedures. The reported error is obtained by subtracting the experimental pK_a from the calculated pK_a (Table 4). The mean unsigned error (MUE), standard

Table 2. Solvation Energies of Acids and Their Anions (in kcal/mol)[a]

Acids	exp[34]	S7	S8
phenol	−6.62	−7.21	−7.92
m-aminophenol		−10.83	−11.68
p-aminophenol		−10.85	−11.76
o-chlorophenol		−4.20	−4.61
m-chlorophenol		−7.16	−7.73
p-chlorophenol		−7.50	−8.08
m-cyanophenol		−9.71	−10.48
p-cyanophenol		−10.42	−11.21
m-fluorophenol		−7.63	−8.44
p-fluorophenol		−8.03	−8.92
m-hydroxyphenol		−12.79	−13.81
p-hydroxyphenol		−13.23	−14.32
m-methoxyphenol		−8.37	−9.10
p-methoxyphenol		−9.04	−9.83
o-methylphenol	−5.87	−6.60	−7.18
m-methylphenol	−5.49	−7.01	−7.70
p-methylphenol	−6.14	−7.04	−7.72
o-nitrophenol		−4.44	−5.13
m-nitrophenol		−9.64	−10.54
p-nitrophenol		−10.65	−11.58
Anions	**exp[24,34]**	**S7**	**S8**
phenoxide	−75, −72	−73.26	−73.49
m-aminophenoxide		−77.27	−77.68
p-aminophenoxide		−78.50	−79.47
o-chlorophenoxide		−67.32	−67.58
m-chlorophenoxide		−65.83	−66.34
p-chlorophenoxide		−67.13	−67.80
m-cyanophenoxide		−63.99	−64.55
p-cyanophenoxide		−61.35	−62.10
m-fluorophenoxide		−68.25	−68.52
p-fluorophenoxide		−71.62	−72.06
m-hydroxyphenoxide		−76.64	−77.15
p-hydroxyphenoxide		−80.11	−81.07
m-methoxyphenoxide		−73.46	−73.74
p-methoxyphenoxide		−74.75	−75.46
o-methylphenoxide		−70.65	−70.91
m-methylphenoxide		−72.90	−73.26
p-methylphenoxide		−73.23	−74.06
o-nitrophenoxide		−61.81	−62.39
m-nitrophenoxide		−63.20	−63.89
p-nitrophenoxide		−57.92	−58.59

[a] **S7** = CPCM/HF/6-31G(d)//CPCM/HF/6-31+G(d). **S8** = CPCM/HF/6-31+G(d)//CPCM/HF/6-31+G(d).

deviation (STDEV), and root-mean-square error (RMS) are listed for each solvation procedure. Table 5 presents the errors in the gas-phase calculations as compared to experiment,[30−33] for a standard state of 1 atm. The experimental error bars are reported as ±2 kcal/mol. Table 6 analyzes the errors in the solvation calculations. The experimental values for $\Delta\Delta G_s$ are derived from experimental values for ΔG_{gas} and pK_a by using our thermodynamic cycle and eqs 4−6. All values in Table 6 are for a 1 M standard state.

The accuracy of the raw gas-phase data reported in Table 1 is evaluated by comparison to experiment in Table 5. [30−33] All of the ΔG_{gas} values reported are for the reaction HA → H^+ + A^-, using a standard state of 1 atm, and the experimental values are taken from the NIST website.[9] The CBS-QB3 value for ΔG_{gas} was compared to an average of all experimental values

(30) Fujio, M.; McIver, R. T., Jr.; Taft, R. W. *J. Am. Chem. Soc.* **1981**, *103*, 4017−4029.
(31) Cumming, J. B.; Kebarle, P. *Can. J. Chem.* **1978**, *56*, 1−9.
(32) Kebarle, P.; McMahon, T. B. *J. Am. Chem. Soc.* **1977**, *99*, 2222-.
(33) Bartmess, J. E.; Scott, J. A.; McIver, R. T., Jr. *J. Am. Chem. Soc.* **1979**, *101*, 6046−6056.

Table 3. Absolute pK_a Values and Signed Error from Experiment

	pK_a	experiment	error		pK_a	experiment	error
phenol				*m*-cyanophenol			
S7	9.88	9.98	−0.10	**S7**	8.03	8.61	−0.58
S8	10.23	9.98	0.25	**S8**	8.19	8.61	−0.42
m-aminophenol				*p*-cyanophenol			
S7	9.92	9.87	0.05	**S7**	8.21	7.95	0.26
S8	10.25	9.87	0.38	**S8**	8.24	7.95	0.29
p-aminophenol				*m*-fluorophenol			
S7	10.80	10.30	0.50	**S7**	9.33	9.28	0.05
S8	10.75	10.30	0.45	**S8**	9.73	9.28	0.45
o-chlorophenol				*p*-fluorophenol			
S7	7.66	8.56	−0.90	**S7**	9.90	9.95	−0.05
S8	7.77	8.56	−0.79	**S8**	10.23	9.95	0.28
m-chlorophenol				*m*-hydroxyphenol			
S7	9.29	9.02	0.27	**S7**	9.36	9.44	−0.08
S8	9.33	9.02	0.31	**S8**	9.73	9.44	0.29
p-chlorophenol				*p*-hydroxyphenol			
S7	9.84	9.38	0.46	**S7**	9.70	9.96	−0.26
S8	9.77	9.38	0.39	**S8**	9.80	9.96	−0.16
m-methoxyphenol				*p*-methylphenol			
S7	9.29	9.65	−0.36	**S7**	10.37	10.14	0.23
S8	9.62	9.65	−0.03	**S8**	10.26	10.14	0.12
p-methoxyphenol				*o*-nitrophenol			
S7	10.45	10.21	0.24	**S7**	7.49	7.23	0.26
S8	10.51	10.21	0.30	**S8**	7.57	7.23	0.34
o-methylphenol				*m*-nitrophenol			
S7	10.52	10.29	0.23	**S7**	8.13	8.40	−0.27
S8	10.75	10.29	0.46	**S8**	8.29	8.40	−0.11
m-methylphenol				*p*-nitrophenol			
S7	9.97	10.08	−0.11	**S7**	7.91	7.15	0.76
S8	10.21	10.08	0.13	**S8**	8.10	7.15	0.95

Table 4. Summary of Errors in Absolute pK_a Calculation[a]

	MUE	STDEV	RMS error
S1	3.34	3.76	3.67
S2	3.56	3.96	3.86
S3	2.92	3.22	3.13
S4	3.13	3.39	3.30
S5	2.62	2.92	2.85
S6	2.81	3.08	3.00
S7	0.30	0.39	0.38
S8	0.35	0.42	0.41

[a] **S1** = CPCM/HF/6-31G(d)//B3LYP/CBSB7, **S2** = CPCM/HF/6-31+G(d)//B3LYP/CBSB7, **S3** = CPCM/HF/6-31G(d)//HF/6-31G(d), **S4** = CPCM/HF/6-31+G(d)//HF/6-31G(d), **S5** = CPCM/HF/6-31G(d)//HF/6-31+G(d), **S6** = CPCM/HF/6-31+G(d)//HF/6-31+G(d), **S7** = CPCM/HF/6-31G(d)//CPCM/HF/6-31+G(d), **S8** = CPCM/HF/6-31+G(d)//CPCM/HF/6-31+G(d).

for a particular reaction. Since the reported experimental uncertainties for all of these experiments is ±2 kcal/mol, all of the calculated CBS-QB3 values fall within experimental error. Of the 20 phenols, 18 have calculated values that are more negative than the average experiment. This raises the possibility of a systematic error in either the experimental data, the CBS-QB3 method, or the value of $G_{gas}(H^+)$.

Table 2 reports the ΔG_s values for all of the acids and anions (in kcal/mol). The use of a diffuse basis set in the wave function of the CPCM calculation consistently lowers the ΔG_s value. Switching from **S1** to **S2** lowers the values by an average of 0.84 kcal/mol for the acids and 0.54 kcal/mol for the anions (Table S2, Supporting Information). Changing from **S3** to **S4** lowers the values by an average of 0.75 kcal/mol for the acids and 0.47 kcal/mol for the anions. Switching from **S5** to **S6** also lowers the values by an average of 0.75 kcal/mol for the acids while lowering the value of the anions by 0.49 kcal/mol. Finally, changing from **S7** to **S8** lowers the values by an average of 0.77 kcal/mol for the acids and the anions by 0.55 kcal/mol. All of these examples illustrate the larger effect diffuse functions have on the acids than on the anions. We have also seen this effect in our work on carboxylic acids and their anions.[6] Barone and Cossi have acknowledged this limitation, which they attribute to larger basis sets allowing more solute charge to escape the cavity.[11,25] Since Table 6 shows that the error in $\Delta\Delta G_s$ is almost always positive, it follows that the larger decrease in the acid than the anion should increase the error in $\Delta\Delta G_s$ (eq 6). This is supported by the error analysis in Table 6, as the errors get worse going from **S1** to **S2**, **S3** to **S4**, **S5** to **S6**, and **S7** to **S8**.

Changing the level of theory used for the gas-phase geometry optimization affects the CPCM solvation SPCs (**S1**−**S6**). The addition of diffuse functions is unimportant for the acid geometries; changing from HF/6-31G(d) to HF/6-31+G(d) only lowers the ΔG_s value of the acids by an average of 0.02 kcal/mol. As expected, reoptimizing with diffuse functions has a much greater effect on the anions, lowering their ΔG_s value by an average of 0.86 kcal/mol. Changing from B3LYP/CBSB7 to HF/6-31G(d) geometries raises the ΔG_s value by 2.37 kcal/mol for the acids and just 1.23 kcal/mol for the anions. Changing from B3LYP/CBSB7 to HF/6-31+G(d) geometries raises the ΔG_s by 2.35 kcal/mol for the acids and only 0.36 kcal/mol for the anions. The addition of diffuse functions has a negligible effect on the acid geometry, but a very significant effect upon the anion geometry. There is also a very large difference between the B3LYP and the HF geometries, especially for the acids. As mentioned previously, since the error in $\Delta\Delta G_s$ is almost always positive (Table 6), the solvation method with the least negative ΔG_s for the acids and most negative ΔG_s for the anions will give the most accurate results. It follows that the HF/6-31+G(d) geometry should produce the best results if a gas-phase geometry is used for the solvation calculation. By

Table 5. Analysis of Errors in Gas-Phase Calculations (in kcal/mol)[a]

	Exp. 1	Exp. 2	Exp. 3	Exp. 4	av	CBS-QB3		error
phenol	342.3	343.4			342.9	342.3		−0.6
m-aminophenol	343.7		344.3		344.0	342.7		−1.3
p-aminophenol	345.6		347.6		346.6	345.1		−1.5
o-chlorophenol			337.1		337.1	336.3		−0.8
m-chlorophenol	335.0		335.5		335.3	334.1		−1.2
p-chlorophenol	336.2		336.8		336.5	335.8		−0.7
m-cyanophenol	328.9		329.0		329.0	328.0		−1.0
p-cyanophenol	325.3		325.7		325.5	324.8		−0.7
m-fluorophenol	336.8		337.6		337.2	336.1		−1.1
p-fluorophenol	339.9		340.8		340.4	339.8		−0.5
m-hydroxyphenol	339.8		338.3		339.1	339.3		0.3
p-hydroxyphenol	343.1				343.1	342.8		−0.3
m-methoxyphenol	341.1		341.9		341.5	340.5		−1.0
p-methoxyphenol	343.5		344.2		343.9	342.7		−1.2
o-methylphenol	342.0		342.7		342.4	341.1		−1.2
m-methylphenol	342.7		343.8		343.3	342.2		−1.0
p-methylphenol	343.4		344.7	343.4	343.8	343.1		−0.8
o-nitrophenol			329.5		329.5	330.3		0.8
m-nitrophenol	327.5		327.7		327.6	327.4		−0.2
p-nitrophenol	320.9				320.9	320.8		−0.1
							MUE	0.8
							STDEV	0.9
							RMS	0.9

[a] Exp. 1, ref 30. Exp. 2, ref 31. Exp. 3, ref 32. Exp. 4, ref 33.

Table 6. Analysis of Errors in $\Delta\Delta G_s$ (in kcal/mol)[a]

		S1	S2	S3	S4	S5	S6	S7	S8	exp.
phenol		5.1	5.5	4.2	4.6	3.9	4.2	0.5	0.9	−331.1
m-aminophenol		7.0	7.5	5.6	5.9	5.1	5.5	1.2	1.7	−332.3
p-aminophenol		11.9	12.1	10.2	10.3	9.8	9.9	2.6	2.6	−334.9
o-chlorophenol		−0.5	−0.2	0.6	0.9	0.3	0.6	−2.4	−2.2	−325.4
m-chlorophenol		3.0	3.2	3.5	3.8	3.2	3.4	0.2	0.3	−323.5
p-chlorophenol		5.9	5.9	4.8	4.7	4.3	4.2	0.5	0.4	−324.8
m-cyanophenol		1.6	2.1	1.4	1.8	0.4	0.8	−1.7	−1.5	−317.2
p-cyanophenol		0.9	1.1	0.9	1.2	0.6	0.9	−1.8	−1.7	−313.8
m-fluorophenol		3.5	4.2	3.2	3.9	3.0	3.7	0.2	0.8	−325.5
p-fluorophenol		6.3	6.8	4.7	5.0	4.3	4.7	0.4	0.9	−328.6
m-hydroxyphenol		4.9	5.4	3.7	4.1	3.2	3.7	−1.1	−0.6	−327.3
p-hydroxyphenol		5.2	5.3	6.6	6.8	6.2	6.3	−0.1	0.0	−331.4
m-methoxyphenol		5.5	5.9	3.7	4.1	3.3	3.7	0.1	0.5	−329.8
p-methoxyphenol		8.1	8.2	7.1	7.1	6.6	6.5	1.8	1.9	−332.1
o-methylphenol		6.9	7.2	6.1	6.4	5.1	5.3	2.0	2.3	−330.6
m-methylphenol		6.2	6.5	4.9	5.2	4.6	4.9	1.0	1.4	−331.5
p-methylphenol		7.6	7.4	6.1	6.0	5.7	5.5	1.3	1.2	−332.1
o-nitrophenol		−2.7	−2.6	−2.4	−2.2	−2.5	−2.3	−4.2	−4.1	−317.8
m-nitrophenol		0.4	0.8	0.7	1.1	0.3	0.7	−2.3	−2.1	−315.9
p-nitrophenol		−0.8	−0.4	−0.8	−0.4	−1.0	−0.6	−2.7	−2.4	−309.2
	MUE	4.7	4.9	4.1	4.3	3.7	3.9	1.4	1.5	
	STDEV	5.7	5.9	4.9	5.0	4.5	4.6	1.8	1.8	
	RMS	5.5	5.8	4.7	4.9	4.4	4.5	1.8	1.8	

[a] **S1** = CPCM/HF/6-31G(d)//B3LYP/CBSB7, **S2** = CPCM/HF/6-31+G(d)//B3LYP/CBSB7, **S3** = CPCM/HF/6-31G(d)//HF/6-31G(d), **S4** = CPCM/HF/6-31+G(d)//HF/6-31G(d), **S5** = CPCM/HF/6-31G(d)//HF/6-31+G(d), **S6** = CPCM/HF/6-31+G(d)//HF/6-31+G(d), **S7** = CPCM/HF/6-31G(d)//CPCM/HF/6-31+G(d), **S8** = CPCM/HF/6-31+G(d)//CPCM/HF/6-31+G(d).

this reasoning, the CPCM calculation using the HF/6-31G(d) basis set on the HF/6-31+G(d) geometry should give the best results for methods not using a solvated phase geometry optimization. This is true as Table 6 shows that **S5** has the smallest error. The analysis presented in Table 6 also shows that the geometry has a much more important effect on the CPCM calculations than the basis set of the CPCM/HF wave function. **S5** and **S6** give the best results followed by **S3** and **S4**. The B3LYP geometries, **S1** and **S2**, give the worst results.

Discussion

Since the HF/6-31+G(d) geometry gave the most accurate results for the solvation calculations among the gas-phase geometries, we tried the equivalent solvated phase geometry optimization, CPCM/HF/6-31+G(d). Changing from the gas-phase HF/6-31+G(d) geometry to the solvated phase CPCM/HF/6-31+G(d) geometry lowers the ΔG_s by 0.37 kcal/mol for the acids and 3.93 kcal/mol for the anions. This leads to a change in $\Delta\Delta G_s$ of 3.56 kcal/mol. Using CPCM/HF/6-31+G(d) optimized structures yields accurate pK_a calculations. Since the experimental values for $\Delta\Delta G_s$ are derived from the experimental gas-phase deprotonation data, the errors in the $\Delta\Delta G_s$ data are also 2 kcal/mol. This means that the vast majority of the **S7** and **S8** values reported in Table 2 are within experimental error. Also, the STDEV and RMS errors are within experimental error for the two data sets. For the same reason that **S5** is slightly better than **S6**, **S7** is slightly better than **S8**.

The absolute pK_a calculations are displayed in Tables 3 and S3 (Supporting Information). All of the calculated pK_a values for **S1**−**S6** are higher than experiment (Table S3). As stated earlier, the error in ΔG_{gas} is usually a small negative error and the error in $\Delta\Delta G_s$ is a large positive error. Since the error in $\Delta\Delta G_s$ is so much larger than ΔG_{gas}, this error dominates the pK_a values and the error in pK_a is always positive.

The calculated pK_a values for **S7** and **S8** are quite good (Table 3). The errors for both ΔG_{gas} and $\Delta\Delta G_s$ are generally within experimental error. Also, the large error in *p*-aminophenol is fixed by this treatment. The species *p*-aminophenoxide was found to have significantly different lowest energy conformations between the gas phase and the solvated phase. This difference accounted for the abnormally large error in *p*-aminophenol with **S1**−**S6**. There would be no way to fix this problem using the gas-phase geometry solvation calculations.

Table 4 is an analysis of the total errors in the absolute pK_a calculations. All of the calculations used CBS-QB3 for the gas-phase calculation, so the errors are broken down by solvation method. As expected, **S7** gives the best values followed by **S8**, **S5**, **S6**, **S3**, **S4**, **S1**, and **S2**. This is the exact same order as the accuracy in the solvation calculations.

There are a few experimental solvation values available in the literature for the phenols.[34] These values are listed in Table 2 for phenol, *o*-methylphenol, *m*-methylphenol, and *p*-methylphenol. **S1** and **S2** are off by 1.5 to 3 kcal/mol for these values. The methods using the gas-phase HF geometries are better. For **S3** and **S5** the errors are 0.3 to 1.2 kcal/mol. This is significantly less than the error in $\Delta\Delta G_s$ reported in Table 6. Therefore, the majority of the error must be in the solvation calculations for the anions. The only experimental value for ΔG_s of an anion is −75 kcal/mol for the phenoxide anion. The most negative value we can calculate using CPCM with a gas-phase geometry is −69.92 kcal/mol with **S2**. SM5.42R is another continuum solvation method that uses gas-phase geometries.[34,35] The most negative value reported for the SM5.42R test set is −67.2 kcal/mol for the HF/cc-pVDZ parametrization while the most negative value using a DFT parametrization is −63.3 kcal/mol for B3LYP/MIDI!.[34] SM5.42R may have the same difficulties in this situation as CPCM.

The use of solvated phase geometries yields a significant improvement. For the neutrals the errors are 0.6 to 1.5 kcal/mol for **S7**. This is on a similar scale as **S5** (Supporting Information, Table S2). However, the error in the phenoxide anion is significantly less. For **S7** the ΔG_s value for phenoxide is −73.26 kcal/mol. This is much more negative than any of the ΔG_s values found using gas-phase geometries. Our value of ΔG_s for the phenoxide anion using a solvated phase geometry agrees with the two reported experimental values of −72[24] and −75[34] kcal/mol.

We believe that the entire issue of standard states is worth additional discussion. The value we used for $\Delta G_s(H^+)$, −264.61 kcal/mol, was derived from an experimental thermodynamic cycle for acetic acid.[6] Referring to the thermodynamic cycle in the Introduction, ΔG_{aq} is 6.48 kcal/mol for a pK_a of 4.75, ΔG_s for acetic acid and acetate anion are known,[24,34] and the value of ΔG_{gas} of 341.4 kcal/mol was obtained using the average of three experimental determinations.[30,31,36] Solving for $\Delta G_s(H^+)$ yields a value of −264.61 kcal/mol. This number can be criticized because of the intrinsic errors in the experimental values used to derive this number. However, we can compare the value of −264.6 with the most recent values available. Our value is the standard state free energy of solvation, defined as the free energy of transfer of 1 M solute from the gas phase to 1 M aqueous solution, at 298.15 K. This is the convention used for implicit solvation programs that calculate the standard state free energy of solution.[34,37] We have previously discussed how the value for ΔG_{vap} for water of 2.053 kcal/mol is equivalent to the value for ΔG_s for water of −6.3 kcal/mol when the standard state for ΔG_{vap} [H_2O(l, 298.15 K, 55.53 M) ⇒ H_2O(g, 298.15 K, 1 atm)] is converted to [H_2O(l, 298.15 K, 1 M) ⇒ H_2O(g, 298.15 K, 1 M)].[7] In a similar manner, converting our value for $\Delta G_s(H^+)$ from a standard state of 1 M in the gas phase to 1 atm in the gas phase changes the value of $\Delta G_s(H^+)$ to −262.72 kcal/mol. This number is in good agreement with the most recent experimental and theoretical values.[6,38−40] Two groups have used high-level ab initio calculations in a combined supermolecule-continuum approach to calculate the free energy of hydration of the proton.[38,40] They calculate values of −262.5 and −262.4 for $\Delta G_s(H^+)$ using a reference state of 1atm in the gas phase. Converting their values to the 1 M gas-phase reference state yields values of −264.4 and −264.3 kcal/mol.[40] These calculated results are in excellent agreement with the best experimental estimates of the free energy of hydration of the proton[40] and strengthen support of our use of the value of −264.6 kcal/mol, which we derived from available experimental data on deprotonation of acetic acid.[6]

At the suggestion of a referee we performed a series of test calculations on the CPCM/HF/6-31+G(d) phenol and phenoxide geometries. Use of 6-311G(2df,p) and 6-311+G(2df,p) basis sets yielded values for phenol of −7.23 and −7.58 kcal/mol, results very similar to those obtained with the **S7** and **S8** procedures (Table 2). Using these same two basis sets for phenoxide gave values of −72.43 and −72.55 kcal/mol, about 1 kcal/mol more positive than the **S7** and **S8** procedures (Table 2). These results are very similar to the results reported previously by Barrone and Cossi, who stated that more solute charge escapes from the cavity for ions than for neutrals, because of the exponential decay of electronic tails.[11] We also performed CPCM/MP2 SPCs using the 6-31G(d), 6-31+G(d), 6-311G-(2df,p), and 6-311+G(2df,p) basis sets on the same two geometric structures. Results were within a few hundredths of a kcal/mol to the HF calculations, resulting in no significant differences. We note that the size of the molecular cavities in the CPCM method were optimized at the Hartree−Fock level by using small basis sets, so the use of procedures **S7** and **S8** makes sense within the context of this particular method.

In our previous work on pK_a calculations of carboxylic acids, geometry optimization in aqueous solution was not necessary to obtain accurate pK_a values.[5−7] Yet for phenols, geometry

(34) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9−63.
(35) Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293−298.
(36) Taft, R. W.; Topsom, R. D. *Prog. Phys. Org. Chem.* **1987**, *16*, 1−83.
(37) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys Chem. B* **2000**, *104*, 4726−4734.
(38) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Caldwell, R. A.; Rashin, A. A. *J. Chem. Phys.* **1998**, *109*, 4852−4863.
(39) Tissiander, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlack, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V.; Tuttle, T. R. *J. Phys Chem. A* **1998**, *102*, 7787−7794.
(40) Zhan, C.-G.; Dixon, D. A. *J. Phys Chem. A* **2001**, *105*, 11534−11540.

optimization is absolutely essential, particularly for the anions, to obtain accurate pK_a results. We can explain this using simple delocalization arguments. Figure 2 displays the geometry of the phenoxide anion, with the bond distances for geometry optimization at the HF/6-31+G* and CPCM/HF/6-31+G* levels of theory. The natural charges from natural population analysis[41] are also presented. The figure shows that geometry optimization in the gas phase results in a short carbon−oxygen bond, and a ring structure with alternating carbon−carbon bond lengths. One of the lone pairs on oxygen conjugates with the π system of the phenyl ring, donating electron density into a molecular orbital that shortens the C−O bond and strengthens two C−C bonds at the expense of the other four. This removes electron density from oxygen, decreasing its attractive interaction with solvent molecules. In contrast, geometry optimization in water favors the accumulation of electron density on the oxygen, so delocalization into the π system is disfavored, resulting in a strikingly longer C−O bond and a more uniform ring structure. In an earlier study the natural charge on the phenoxide oxygen has been shown to be a good parameter to relate to pK_a.[4] The natural charges in Figure 2 confirm that there is more charge localization on the oxygen atom for the CPCM reoptimized geometry. For the carboxylic acids that we have studied previously, formic, acetic, cyanoacetic, chloroacetic, oxalic, and pivalic acids, there are no ring systems available to diffuse the charge. These carboxylate anions spread the charge evenly over both oxygens and furthermore lack the freedom to redistribute the charge that an aromatic ring provides in the present series. Hence they do not display the dramatic differences between their gas-phase and solution structures that one finds in the phenols. Geometry optimization of the formate anion in solution lengthens each C−O bond length by 0.0056 Å, and the natural charges on the oxygens increase by 0.0037. The present work suggests that for any molecules that allow significant charge delocalization, geometry optimization in solution will be important.

Conclusion

In this work we have presented a coherent method that uses readily available computational techniques to accurately estimate pK_a values of substituted phenols. Absolute pK_a calculations for phenols are not accurate when using the CPCM continuum solvation method on gas-phase optimized geometries. For the best gas-phase geometry solvation method, **S5**, the MUE is 2.62, the STDEV is 2.92, and the RMS error is 2.85 pK_a units. However, excellent predictions result from using solvated phase geometries in the solvation calculations. For **S7**, the MUE is 0.30, the STDEV is 0.39, and the RMS error is 0.38 pK_a unit. Geometry optimization of the anions in aqueous solution using the CPCM method is essential for accurate pK_a predictions of phenols. The dramatic improvement in the accuracy of the pK_a calculations results from the charge redistribution that occurs in going from the gas phase to solution.

Acknowledgment. The authors thank the Petroleum Research Fund, administered by the ACS, the Camille and Henry Dreyfus Foundation, the NIH, and Hamilton College for support of this work. M.D.L. acknowledges support from the Merck/AAAS USRP and the Bristol-Myers Squibb Fund. S.F. and G.C.S acknowledge support from the Camille and Henry Dreyfus Foundation Scholar/Fellow program. We thank Chris Cramer and Ed Sherer for helpful discussions.

Supporting Information Available: Optimized geometries and energies for all stationary points discussed in the text, and Tables S2 and S3, which contain results for eight different solvation procedures (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA012474J

(41) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735−746.