

## Acidity of Organic Molecules in the Gas Phase and in Aqueous Solvent

Igor A. Topol,\* Gregory J. Tawa, Richard A. Caldwell,† Michael A. Eissenstat,‡ and Stanley K. Burt

Advanced Biomedical Computing Center and Structural Biochemistry Program, SAIC Frederick, NCI Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201, and Department of Chemistry, The University of Texas at Dallas, P.O. Box 830688, Richardson, Texas 75083-0688

Received: May 26, 2000; In Final Form: August 28, 2000

Using a combination of ab initio, DFT and continuum solvation methods, the gas phase and aqueous acidities for a set of weak organic acids with high  $pK_a$  values, which cannot be measured experimentally in aqueous solvent, have been calculated. Comparison of the computed and experimental data for different terms used in the thermodynamic cycle for the calculation of  $pK_a$  values allowed us to estimate that the errors in the  $pK_a$  calculations are of order of 2  $pK_a$  units, i.e., less than 10% of the expected  $pK_a$  values for the studied weak organic acids. It is shown that inclusion of explicit water molecules in the solute cavity of compounds with  $pK_a$  values over 40 could lead to dubious results due to the inappropriate description of the corresponding anion solvation. Acidity trends for compounds in the gas phase and in aqueous solvent were found to be different, due to the effects of aqueous solvation.

## 1. Introduction

Studies of the acidity of organic compounds are of the fundamental importance and play a very important role in the evaluation of reactivity, reaction mechanisms and structure of organic compounds.<sup>1,2</sup> The acidities of organic compounds in the gas phase are rather well-known experimentally<sup>3</sup> and provide an important reference tool for validations of various quantum chemistry methods used for the calculation of thermochemistry parameters of organic compounds. Contemporary high-level post-Hartree–Fock ab initio and density functional theory (DFT) methods can provide gas-phase acidities within 2–3 kcal/mol of the experimental values,<sup>4</sup> and this accuracy is very often within range of the experimental error.<sup>5</sup>

The situation with the acidities of organic compounds in aqueous solvent is quite different. Although many organic chemical processes have been studied in nonaqueous media, acidities are usually discussed in terms of  $pK_a$  values in aqueous solvent. However, such values are problematic, since the acidity of water does not allow direct equilibrium measurements of organic acids with  $pK_a$  values above the  $pK_a$  of water (15.7). The  $pK_a$  values of weak organic acids obtained from the studies in nonaqueous solvents and from the kinetic measurements may be significantly influenced by ion-pairing and aggregation effects.<sup>1,6</sup> Thus, theoretical calculation of  $pK_a$  values can substantially contribute to the tabulation of aqueous acidities of organic compounds and serve as a guideline to the experimentalists.

Unfortunately, there are just a few high level ab initio theoretical investigations on the aqueous  $pK_a$  values of weak organic acids. Noteworthy in this respect are the papers of Jorgensen et al.,<sup>7</sup> where  $pK_a$  values of methanethiol, methanol, acetonitrile, methylamine and ethane in water were calculated using a combination of ab initio methods and Monte Carlo statistical mechanics simulations.

The goal of the present work was to calculate the  $pK_a$  values of weak organic acids in aqueous solvent. We chose a set of compounds with estimated  $pK_a$  values ranging from 17 to over 50. These compounds included acetylene ( $C_2H_2$ ), acetylene hydrate ( $C_2H_2-H_2O$ ), cyclopentadiene ( $C_5H_6$ ), ethane ( $C_2H_6$ ), cyclopropane ( $C_3H_6$ ), toluene ( $C_7H_8$ ), cyclohexane ( $C_6H_{12}$ ) and bicyclo[3.1.0]hexane ( $C_6H_{10}$ ). In the case of toluene we considered two deprotonation sites: one on the methyl group ( $C_M$ ), and the other on a ring carbon ( $C_R$ ). Special attention devoted to the analysis of the results of calculations of different part of the thermodynamic cycle used in the  $pK_a$  calculations. Analysis of uncertainties in calculated gas-phase acidities, C–H bond energies, electron affinities of anion-radicals, and hydration energies of neutral compounds and their deprotonated anions should allow us to evaluate the reliability of empirical estimations of aqueous  $pK_a$  values of weak organic acids.

## 2. Theory

The gas-phase acidity of the protonated molecule AH is described by the enthalpy  $[\Delta H_{acid}(A-H)]$  for the proton abstraction reaction



The gas-phase acidity is related to the A–H bond dissociation energy  $[BDE(A-H)]$  and the electron affinity (EA) of the final radical by the following relationship:

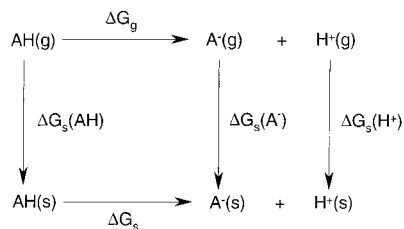
$$\Delta H_{acid}(A-H) = BDE(A-H) + IP(H) - EA(A) \quad (2)$$

where  $IP(H)$  is the well-known hydrogen atom ionization potential (313.58 kcal/mol). Thus, knowing the BDE and EA, the gas-phase acidity can be determined.

Theoretically, the absolute or relative aqueous  $pK_a$  values for the compound AH can be calculated from the generalized thermodynamic cycle

† University of Texas at Dallas.

‡ NCI Frederick Cancer Research and Development Center.



where  $\Delta G_g$  and  $\Delta G_s$  are the free energies for proton abstraction (1) in the gas phase and aqueous solvent respectively, and  $\Delta G_s(\text{AH})$ ,  $\Delta G_s(\text{A}^-)$  and  $\Delta G_s(\text{H}^+)$  are solvation energies of AH,  $\text{A}^-$  and  $\text{H}^+$ , respectively.

Thus,  $\text{p}K_a$  calculations require reliable and highly accurate gas-phase protonation/deprotonation energy calculations as well as solvation energy calculations for both products and reactants. Inaccuracies in the quantum-mechanical calculation of proton affinities even at the post-HF level can lead to errors of several units in relative  $\text{p}K_a$ .<sup>8</sup> Inaccuracies in solvation energy calculations can lead to additional errors. For example, the relative solvation free energies calculated for a set of neutral organic molecules using Monte Carlo simulations were within 1.3 kcal/mol of their experimental values.<sup>7</sup> However, even an error of this size can easily lead to errors in  $\text{p}K_a$  of greater than 50% for relatively strong acids.<sup>9</sup>

Recently, however, the possibilities for accurate  $\text{p}K_a$  calculations have been improved. One area of improvement that is relevant to small and medium size molecules is the usage of DFT<sup>9–11</sup> and correlated G2<sup>4a–c,10–13</sup> theories. These methods consistently yield proton affinities and proton-transfer enthalpies within 1–4 kcal/mol of experiment. Another area of improvement has been in the development of self-consistent reaction field procedures that combine ab initio quantum mechanics with dielectric continuum solvation theory.<sup>14–25</sup> These procedures can give a remarkably accurate representation of the properties of molecules in aqueous environments.

We have recently shown, using a set of substituted imidazoles, that both absolute and relative  $\text{p}K_a$  values for the deprotonation of nitrogen on the imidazole ring can be calculated with an average absolute deviation less than 0.8 units from experiment.<sup>10</sup> This degree of accuracy is possible only if the solutes are treated at the correlated level using either G2 or DFT in the B3LYP approximation. The DFT methods seem to be very effective tools for large molecules that, when combined with a continuum dielectric solvation model, allow the calculation of  $\text{p}K_a$  values for complex systems. Recently, the combination of these two methods was applied to the calculation of  $\text{p}K_a$  values for several small and medium size molecules,<sup>10,11,26</sup> and was shown to be more accurate than most approaches used to calculate absolute  $\text{p}K_a$  values.

Examining reaction 1 and thermodynamic cycle shown above, the  $\text{p}K_a$  for compound AH can be calculated from the following formula:

$$\text{p}K_a = 0.434(RT)^{-1} \{ \Delta G_g + \Delta G_s(\text{A}^-) - \Delta G_s(\text{AH}) + \Delta G_s(\text{H}^+) \} \quad (3)$$

The gas-phase free energy of the proton abstraction reaction can be expressed in terms of the enthalpy and entropy contributions

$$\Delta G_g = \Delta H_g - T\Delta S_g \quad (4)$$

The enthalpy of the proton abstraction reaction in the gas phase (gas-phase acidity),  $\Delta H_g$ , can be calculated by

$$\Delta H_g = E(\text{A}^-) - E(\text{AH}) + E_{\text{vib}}(\text{A}^-) - E_{\text{vib}}(\text{AH}) + \frac{5}{2}RT \quad (5)$$

where  $E$  represents the electronic energy obtained from the standard DFT or post-HF calculations,  $E_{\text{vib}}$  includes zero point energy and temperature corrections to the vibrational enthalpy, and the term  $\frac{5}{2}RT$  includes the translational energy of the proton and the  $\Delta(PV)$  term. The entropy contribution is given by

$$-T\Delta S_g = -T[S(\text{A}^-) + S(\text{H}^+) - S(\text{AH})] \quad (6)$$

For  $T = 298$  K at the standard pressure, the second term  $TS(\text{H}^+) = 7.76$  kcal/mol.<sup>27</sup> Thus,

$$\Delta G_g = \Delta H_g - T[S(\text{A}^-) - S(\text{AH})] - 7.76 \quad (7)$$

Actually, the translational and rotational contributions to the  $\Delta H_g$  should also be included in expression (5); however, for anionic and neutral systems in the ideal gas approximation, these contributions cancel each other.

The solvation free energy,  $\Delta G_s$ , for any of the species present in eq 3 can be written, in general, in terms of the electrostatic free energy and the nonelectrostatic enthalpic and entropic components:

$$\Delta G_s = \Delta G_{\text{el}} + \Delta G_{\text{nel}} \quad (8)$$

where the electrostatic component,  $\Delta G_{\text{el}}$ , is calculated in the dielectric continuum approximation using a self-consistent reaction field (SCRF) cycle.<sup>20,25</sup> The nonelectrostatic term  $\Delta G_{\text{nel}}$  includes a cavity term, solute–solvent dispersion interaction energy, and the entropy change of the solvent due to localized ordering about the solute.

The last term in eq 3, the proton solvation free energy  $\Delta G_s(\text{H}^+)$  is, unfortunately, not known with high precision. The range of proposed values for  $\Delta G_s(\text{H}^+)$  ranges from  $-259^9$  to  $-262.5$  kcal/mol.<sup>28</sup> Our recent calculations on  $\text{p}K_a$  values of substituted imidazoles<sup>10</sup> indicate that the solvation free energy of the proton is closer to the lower end of this range. Moreover, our direct quantum-mechanical calculations using a combination of explicit solvent molecules and dielectric continuum<sup>29</sup> also gave a value of the proton free energy of solvation close to  $-262.5$  kcal/mol. A subsequent related theoretical study by Tissandier et al.,<sup>30</sup> obtained by analysis of certain trends in experimentally determined cluster-ion solvation data, found  $\Delta G_s(\text{H}^+)$  to be  $-263.98 \pm 0.07$  kcal/mol. In the present work we have used the value  $\Delta G_s(\text{H}^+) = -262.5$  kcal/mol. The determination of relative  $\text{p}K_a$  values does not require a knowledge of the proton solvation free energy because of cancellation effects.

### 3. Computational Methods

In the dielectric continuum approximation the solvated system is pictured as follows: a solute molecule is placed in solution and displaces the solvent, thereby creating a solvent-excluded volume. The boundary of this volume is the solute molecular surface.<sup>31</sup> The region within the solute volume is assigned a dielectric constant  $\epsilon(\mathbf{r}) = 1$ . The rest of space is assigned the measured dielectric constant of the solution,  $\epsilon(\mathbf{r}) = 78.5$ , for aqueous applications. The solute charge density imposes an electric field on the surrounding continuum. This induces a solvent polarization that achieves equilibrium with the solute electric field. The subsequent solute–solvent interaction defines the *electrostatic* component  $\Delta G_{\text{el}}$  of the solvation free energy, or of the solvation enthalpy.<sup>32,33</sup>

**TABLE 1: Gas Phase Acidities and Free Energies of the Proton Abstraction Reaction at 298 K ( $\Delta H$  and  $\Delta G$ , respectively, in kcal/mol) for Studied Organic Molecules**

method		C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> –H <sub>2</sub> O	C <sub>5</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>3</sub> H <sub>6</sub>	C <sub>7</sub> H <sub>8</sub> C <sub>M</sub> <sup>a</sup>	C <sub>7</sub> H <sub>8</sub> C <sub>R</sub> <sup>b</sup>	C <sub>6</sub> H <sub>12</sub>	C <sub>6</sub> H <sub>10</sub>
MP2/6-311++G(d,p)	$\Delta H$	379.3	363.8	348.0	420.9	414.0	384.4	399.7	413.5	409.0
	$\Delta G$	371.4	356.9	340.9	413.2	406.3	377.2	391.5	405.6	401.3
MP2/6-311+G(3df,2p) <sup>c</sup>	$\Delta H$	376.8	361.3	346.5	418.9	411.6	381.3	399.2	413.1	407.4
	$\Delta G$	368.8	355.7	340.4	411.2	404.0	374.0	391.4	405.3	399.7
G2MP2	$\Delta H$	377.4	362.6	354.0	420.5	413.1	382.9	400.5	414.5	409.0
	$\Delta G$	369.4	355.5	347.9	412.9	405.4	374.6	392.7	406.8	401.3
B3LYP/6-311++G(d,p)	$\Delta H$	376.7	360.8	353.7	418.9	413.1	380.9	401.3	411.9	409.2
	$\Delta G$	368.9	354.7	346.1	411.2	405.3	374.3	393.6	404.1	401.3
exp <sup>d</sup>	$\Delta H$	378.0 <sup>e</sup>		353.9	420.1	412.2 <sup>e</sup>	380.8	400.7 <sup>f</sup>	404.0	
	$\Delta G$	369.8		347.7	411.3	408.2	373.7	390.9	398.0	

<sup>a</sup> Proton abstraction from the methyl group of toluene molecule (C<sub>M</sub>). <sup>b</sup> Proton abstraction from the ring carbon (C<sub>R</sub>) located para to the methyl group in the toluene molecule. <sup>c</sup> Calculated at the MP2/6-31G(d) optimized geometries; enthalpic and entropic contributions were taken from the MP2/6-311++G(d,p) calculations. <sup>d</sup> All experimental values from ref 42 unless otherwise noted; experimental accuracy for  $\Delta H$  values is  $\pm 2 \div 3$  kcal/mol. <sup>e</sup> Reference 43. <sup>f</sup> Experimental data for benzene.<sup>42</sup>

The practical implementations of these theoretical descriptions can proceed stepwise. Electronic structure and vibrational frequency calculations are performed using the GAUSSIAN 94 suite of programs<sup>34</sup> on all solutes to obtain the gas-phase acidities and the free energies for the proton abstraction reaction (Table 1). The electronic energies,  $E$  in eq 5, were calculated by second-order Moller–Plesset perturbation theory (MP2), G2MP2<sup>12</sup> and DFT methods. DFT calculations were done using the hybrid exchange–correlation functional of Becke, Lee, Yang, and Parr (B3LYP).<sup>35,36</sup> No scaling factors were used in the MP2 and B3LYP calculations of vibrational frequencies. In all MP2 and B3LYP optimizations and thermodynamic parameter calculations the Gaussian 6-311++G(d,p) basis set was used. We have shown in previous  $pK_a$  calculations<sup>10,11</sup> that this type of basis set, due to fast basis set convergence of the DFT method in general, is a good compromise for accuracy and efficiency. Electronic energies also were calculated by the MP2 method using an extended 6-311+G(3df,2p) basis set at the MP2/6-31G(d) optimized geometries, and corresponding energy corrections were applied to the thermodynamic parameters calculated at the MP2/6-311++G(d,p) level.

The solvation calculations to obtain  $\Delta G_s$  were performed next. The electrostatic solvation free energy,  $\Delta G_{el}$ , was calculated in the B3LYP/6-311+G(d,p) approximation using a self-consistent reaction field cycle.<sup>25</sup> In the SCRF cycle the B3LYP/6-311+G(d,p) calculations are performed on the solutes to obtain an initial electronic energy and charge density. An electrostatic potential fit (ESP) is performed to represent the solute charge density as a set of atom-centered charges. Molecular surfaces (water probe radius = 1.4 Å)<sup>31,37</sup> are then constructed for all solutes. The solvent response to the solute charge distribution (represented by ESP charges) is obtained by solving a Poisson–Boltzmann equation as implemented in the program complex Jaguar.<sup>25,38,39</sup> The electrostatic potential used for fitting of atomic charges was computed on grid formed by merging sets of spherical shells, whose grid points are centered on each nucleus.<sup>23,38</sup> The nonelectrostatic contribution  $\Delta G_{nel}$  to the solvation energy  $\Delta G_s$ , as determined in eq 8, was calculated as a solvation energy of a nonpolar solute of identical size and shape to the actual solute.<sup>23,25</sup> This term was parametrized initially for several solutes on the basis of linear relationship between experimental solvation energies and solute accessible surface area. The reaction field of the solvent is obtained as a set of polarization charges located on the solute molecular surface. The solute Hamiltonian is then augmented with a Coulomb operator representing the interaction of the polarization charges with the electrons and nuclei of the solute. The calculations are repeated until the electronic energy of the solutes become constant.

This formulation places the entire solute charge density in the form of ESP charges inside the solute cavity. Although some penetration of charges outside the cavity, especially in the cases of large neutral molecules and negatively charged ions, is obvious, this approximation works remarkably well.<sup>20,25,33,40</sup> We discussed the problems of charge penetration outside the cavity and its effect on solvation energy calculations in our previous paper,<sup>10</sup> and we refer readers to this publication. Very recently, we have shown<sup>41</sup> that the use of explicit water molecules inside the cavity in dielectric continuum can substantially reduce the errors in the calculated solvation energies of positively and negatively charged ions by minimizing the charge penetration effects. We have applied this methodology in the current work to investigate changes in the solvation energies and  $pK_a$  values of some weak organic acids due to inclusion of explicit water molecules into the solute cavity of the dielectric continuum model. In the case of clusters created by explicit water molecules and the corresponding organic molecule, the lowest energy structures for a given number of water molecules was used in the solvation energy and  $pK_a$  calculation.

#### 4. Results and Discussion

As mentioned above, the accuracy of calculated  $pK_a$  values could be checked by analyzing consequently the accuracy achieved for all components of the thermodynamic cycle used for calculations (eq 3).

The calculated gas phase acidities and free energies for the proton abstraction reaction in the gas phase are shown in Table 1. As can be seen, there is generally good agreement between theoretical and experimental data. The difference between results for gas phase acidities and free energies obtained using different quantum chemistry methods is generally within the experimental error (2–3 kcal/mol). For most of studied compounds the B3LYP results are close to the G2MP2 data. There is just one example when our calculation data deviated more than by 2–3 kcal/mol from the empirical estimations—the case of cyclohexane. The G2MP2 gas-phase acidity of 414.5 kcal/mol and the corresponding B3LYP value are larger than experimental estimation<sup>42</sup> by 10.5 and 7.9 kcal/mol, respectively. On the other hand, we have very good agreement between experimental and theoretical acidities for ethane and cyclopropane. Besides, the empirical value of the gas-phase acidity of cyclopentane (416.1 kcal/mol),<sup>42</sup> which is expected to be close to the acidity of cyclohexane, also is much closer to our theoretical value than to the empirical gas-phase acidity of cyclohexane. Moreover, as we will show below, the C–H bond dissociation energy (BDE) calculated for cyclohexane agrees well with the experi-



**TABLE 2: Electron Affinities (in kcal/mol) of Radicals, Calculated in Different MP2 and B3LYP Approximations Using the 6-311++G(d,p) Basis Set<sup>a</sup>**

method	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> -H <sub>2</sub> O	C <sub>5</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>3</sub> H <sub>6</sub>	C <sub>7</sub> H <sub>8</sub> C <sub>M</sub> <sup>b</sup>	C <sub>7</sub> H <sub>8</sub> C <sub>R</sub> <sup>c</sup>	C <sub>6</sub> H <sub>12</sub>	C <sub>6</sub> H <sub>10</sub>
UMP2 <sup>d</sup>	40.47	91.43	47.22	-11.35	3.80	37.26	43.53	-6.17	8.29
ROMP2 <sup>e</sup>	65.64	83.24	40.26	-11.60	3.64	17.39	24.46	-6.56	8.17
B3LYP	71.58	89.98	40.80	-6.73	6.16	19.23	23.20	-3.60	9.16
exp	68.5 <sup>f</sup>		41.2 <sup>g</sup>			21.0 <sup>f</sup>	25.3 <sup>h</sup>		

<sup>a</sup> Deprotonated neutral radicals R correspond to the molecules RH that are shown in the first row of the table. <sup>b</sup> Proton abstraction from the methyl group of the toluene molecule (C<sub>M</sub>). <sup>c</sup> Proton abstraction from the ring carbon (C<sub>R</sub>) located para to the methyl group of the toluene molecule.

<sup>d</sup> Unrestricted MP2 approximation. <sup>e</sup> Restricted open shell MP2 calculation at the UMP2 optimized geometries. <sup>f</sup> Reference 43. <sup>g</sup> Reference 44.

<sup>h</sup> Value for benzene.<sup>43</sup>

**TABLE 3: C-H Bond Dissociation Energies (in kcal/mol) Calculated in Different MP2 and B3LYP Approximations Using the 6-311++G(d,p) Basis Set<sup>a</sup>**

method	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>2</sub> -H <sub>2</sub> O	C <sub>5</sub> H <sub>6</sub>	C <sub>2</sub> H <sub>6</sub>	C <sub>3</sub> H <sub>6</sub>	C <sub>7</sub> H <sub>8</sub> C <sub>M</sub> <sup>b</sup>	C <sub>7</sub> H <sub>8</sub> C <sub>R</sub> <sup>c</sup>	C <sub>6</sub> H <sub>12</sub>	C <sub>6</sub> H <sub>10</sub>
UMP2 <sup>d</sup>	106.2	141.7	81.7	96.0	104.2	108.0	129.6	93.8	103.7
ROMP2 <sup>e</sup>	131.4	133.5	74.7	95.7	104.0	88.2	110.6	93.4	103.6
B3LYP	134.7	137.2	80.9	98.6	105.6	86.5	110.9	94.7	104.8
exp <sup>f</sup>	133.1		81.2 <sup>g</sup>	101.1	106.3	88.5	111.2 <sup>h</sup>	95.6	

<sup>a</sup> For the calculation of the C-H bond dissociation energies the  $\Delta H$  values from Table 1 and electron affinities from Table 2 were used. <sup>b</sup> C-H bond at the methyl group of the toluene molecule (C<sub>M</sub>). <sup>c</sup> C-H bond for the ring carbon (C<sub>R</sub>) located opposite to the methyl group in the toluene molecule. <sup>d</sup> Unrestricted MP2 approximation. <sup>e</sup> Restricted open shell MP2 calculation at the UMP2 optimized geometries. <sup>f</sup> Experimental values from ref 43. <sup>g</sup> Reference 44. <sup>h</sup> Value for benzene.<sup>43</sup>

mental estimation. These results suggest that the empirical gas-phase acidity and free energy for the proton abstraction reaction for cyclohexane according to eq 2 are substantially underestimated.

The experimental acidity for the proton abstraction reaction from the ring carbon in toluene is not known. However, it is reasonable to assume that acidity of toluene for the ring carbon located para to the methyl group (C<sub>R</sub>) should be close to the acidity of benzene. This value for benzene is close to the calculated C<sub>R</sub> acidity of toluene. For the proton abstraction energetics in the gas phase, the key factor is the distribution of electron density at the deprotonation site (i.e., C<sub>R</sub> toluene is similar to benzene and the cyclopropane, cyclohexane and bicyclo[3.1.0]hexane acidities should be similar too). As can be seen from Table 1, the differences between gas-phase acidities and between free energies for the proton abstraction reaction for these compounds were within 4 kcal/mol (~1%).

Another way to study the accuracy of the gas-phase portion of the acidity calculations is to calculate the electron affinity of neutral radicals (EA(A) in eq 2) and to use these values together with calculated gas-phase acidities to determine C-H bond dissociation energies (BDE) for ionizable groups according to eq 2. To verify the reliability of these calculations, we used different theoretical methods: spin-unrestricted MP2 approximation (UMP2), restricted open-shell MP2 approximation (ROMP2) and spin-unrestricted B3LYP approach. The results of these calculations are shown in Tables 2 and 3.

As can be seen from Table 2, the calculated EA's are very sensitive to the calculation method. As the UMP2 results demonstrate, there can be very large discrepancies between EA's obtained by UMP2 and ROMP2 methods for a given geometry of the molecules. These discrepancies are caused mostly by the spin contamination of the open shell wave function of the radical. This effect is weak in the DFT representation of the electron density and calculated DFT EA's are close to the experimental data. For the compounds with known experimental EA's the ROMP2 data agree well with the B3LYP results. There are negative EA's for ethane and cyclohexane and a small value of positive EA for cyclopropane. The accuracy of these negative EA's is very important for the determination of BDE according to eq 2.

Table 3 shows BDE of the ionizable C-H bonds in the studied compounds. The BDE's calculated by the B3LYP approach agree very well with the experimental values, much better than the corresponding UMP2 and ROMP2 results. The reason for these discrepancies is determined mostly by uncertainties in the EA values of radicals (see Table 2). The good agreement between experimental and B3LYP data in Tables 1-3 demonstrate the good predictive power of the DFT approach for calculations of gas-phase acidities, EA's of radicals and BDE's of organic compounds.

Given the results shown in Tables 1-3, we can expect that the gas-phase data involved in the calculation of the aqueous pK<sub>a</sub> values of weak organic acids according to eq 3 could be theoretically determined within 2-3 kcal/mol. Such uncertainty should lead to deviations in the pK<sub>a</sub> values of, at most, ~2 pK<sub>a</sub> units. Since the pK<sub>a</sub> values of the studied compounds exceed 15, we expect that any uncertainties in the gas-phase part of calculations should lead to uncertainties in the calculated pK<sub>a</sub> values of less than 2 pK<sub>a</sub> units.

Let us further analyze the results of solvation energy calculations ( $\Delta G_s(A^-)$  and  $\Delta G_s(AH)$  values in eq 3). These results were obtained using the B3LYP approximation and are shown in Table 4. The calculated solvation energies for ethane, cyclopropane, cyclohexane and bicyclo[3.1.0]hexane are positive and agree with empirical results within ~1 kcal/mol. The calculated small negative solvation energy of toluene also is close to the experimental value. The experimental solvation energies for the anions of acetylene and toluene are within 10% of our calculation data. Since the uncertainty of the experimental data is on the same order we expect that the error in the pK<sub>a</sub> values due to the solvation calculations will not exceed 1-2 pK<sub>a</sub> units.

Calculated and empirical pK<sub>a</sub> values for these hydrocarbons are shown in Table 5. The data obtained in the B3LYP and G2MP2 approximations are very close to each other. Thus, we further discuss our results based mostly on the B3LYP calculations. While the empirical data for acetylene and cyclopentadiene are more or less consistent from several sources, the pK<sub>a</sub> values for other compounds are quite elusive. For example, the suggested pK<sub>a</sub> value of ethane (50)<sup>1</sup> represents empirical estimations for this value that range from 40 to 60.<sup>1,47,48</sup> The

**TABLE 4: Solvation Energies at  $\epsilon = 78.4$  for Studied Organic Molecules and Their Deprotonated Anions (all Values in kcal/mol)**

molecule	$\Delta G_s^a$	exp <sup>b</sup>
C <sub>2</sub> H <sub>2</sub>	-2.55	
[C <sub>2</sub> H] <sup>-</sup>	-75.25	-73 <sup>c</sup>
C <sub>2</sub> H <sub>2</sub> -H <sub>2</sub> O	-9.24	
[C <sub>2</sub> H-H <sub>2</sub> O] <sup>-</sup>	-69.48	
C <sub>5</sub> H <sub>6</sub>	-1.03	
[C <sub>5</sub> H <sub>5</sub> ] <sup>-</sup>	-60.42	
C <sub>2</sub> H <sub>6</sub>	+1.95	+1.83
[C <sub>2</sub> H <sub>5</sub> ] <sup>-</sup>	-75.39	
C <sub>3</sub> H <sub>6</sub>	+0.34	+0.75
[C <sub>3</sub> H <sub>5</sub> ] <sup>-</sup>	-71.34	
C <sub>7</sub> H <sub>8</sub>	-1.38	-0.89
[C <sub>7</sub> H <sub>7</sub> ] <sup>-</sup> (C <sub>M</sub> ) <sup>d</sup>	-55.74	-59 <sup>c</sup>
[C <sub>7</sub> H <sub>7</sub> ] <sup>-</sup> (C <sub>R</sub> ) <sup>e</sup>	-66.44	
C <sub>6</sub> H <sub>12</sub>	+2.32	+1.23
[C <sub>6</sub> H <sub>11</sub> ] <sup>-</sup>	-60.50	
C <sub>6</sub> H <sub>10</sub>	+1.43	
[C <sub>6</sub> H <sub>9</sub> ] <sup>-</sup>	-66.91	

<sup>a</sup> Free energy of aqueous solvation in eq 8. <sup>b</sup> Experimental values for the free energy of aqueous solvation from ref 45 unless otherwise noted. <sup>c</sup> Data taken from ref 46. <sup>d</sup> Proton abstraction from the methyl group of the toluene molecule (C<sub>M</sub>). <sup>e</sup> Proton abstraction from the ring carbon (C<sub>R</sub>) located para to the methyl group of the toluene molecule.

**TABLE 5: Difference of the Free Energies of Solvation (kcal/mol) of Studied Neutral Molecules and Their Anions ( $\Delta\Delta G_{\text{solv}}$  in Water, and  $pK_a$  Values**

molecule	$\Delta\Delta G_{\text{solv}}^a$	$pK_a(\text{B3LYP})^b$	$pK_a(\text{G2MP2})^c$	$pK_a(\text{exp})^d$
C <sub>2</sub> H <sub>2</sub>	72.70	24.7	25.1	25
C <sub>2</sub> H <sub>2</sub> -H <sub>2</sub> O	60.24	23.5	24.0	
C <sub>5</sub> H <sub>6</sub>	59.38	17.8	19.1	14–16
C <sub>2</sub> H <sub>6</sub>	75.39	53.8	55.0	48–50
C <sub>3</sub> H <sub>6</sub>	71.68	52.2	52.3	38–46
C <sub>7</sub> H <sub>8</sub> (C <sub>M</sub> )	54.4	42.1	42.4	35–41
C <sub>7</sub> H <sub>8</sub> (C <sub>R</sub> )	65.07	48.4	47.8	37–43 <sup>e</sup>
C <sub>6</sub> H <sub>12</sub>	62.82	57.8	59.8	45
C <sub>6</sub> H <sub>10</sub>	68.3	51.7	51.7	

<sup>a</sup>  $\Delta\Delta G_{\text{solv}} = \Delta G_{\text{solv}}(\text{AH}) - \Delta G_{\text{solv}}(\text{A}^-)$ . <sup>b</sup> Free energies in the gas phase and solvation energies were calculated at the B3LYP/6-311++G(d,p) level. <sup>c</sup> Free energies in the gas phase were calculated by the G2MP2 method,  $\Delta\Delta G_{\text{solv}}$  values were taken from the column 2. <sup>d</sup> Range of the estimated empirical  $pK_a$  values from refs 1, 2, and 49. <sup>e</sup> Value for benzene.

$pK_a$  value of 53.8 for ethane obtained in the present calculations is very close to the  $pK_a$  value of  $54.4 \pm 0.2$  calculated by Jorgensen et al.<sup>7</sup> from relative  $pK_a$  estimations. The calculated  $pK_a$  values for cyclopropane and toluene also are higher than the empirical values. Explicit hydration of acetylene lowers the calculated  $pK_a$  value from 24.7 to 23.5. The  $pK_a$  value of bicyclo[3.1.0]hexane 51.7 is close to the corresponding value for cyclopropane (52.2) showing that the additional ring has little effect on the cyclopropane  $pK_a$ . The different acidity trends in the gas phase (cyclohexane > cyclopropane > ethane) compared to the aqueous phase (cyclopropane > ethane > cyclohexane) are noteworthy. This difference is determined by the different solvation energies of the corresponding anions (Table 4).

As mentioned in the previous section, the problem of “charge leakage” outside the can be partly addressed in the solvation calculation by the inclusion of explicit water molecules in the solute cavity. This allows explicitly take into account hydrogen bond interactions between the solute and solvent for the ionizable group in the neutral molecule as well as its anion. We performed such calculations by adding 1–4 water molecules to each solute. Results of solvation energy and  $pK_a$  calculations

**TABLE 6: Free Energy (in kcal/mol) of the Proton Abstraction Reaction in the Gas Phase ( $\Delta G_g$ ), Difference between the Free Energy of Solvation and Their Deprotonated Anions  $\Delta\Delta G_s$  (in kcal/mol) for Several Hydrocarbon–Water (W) Clusters, and Corresponding  $pK_a$  Values<sup>a</sup>**

cluster	$\Delta G_g$	$\Delta\Delta G_s$	$pK_a$
C <sub>2</sub> H <sub>2</sub> _0W	368.9	72.7	24.7
C <sub>2</sub> H <sub>2</sub> _1W	354.7	60.2	23.5
C <sub>2</sub> H <sub>2</sub> _2W	343.7	49.6	23.2
C <sub>5</sub> H <sub>6</sub> _0W	346.1	59.4	17.8
C <sub>5</sub> H <sub>6</sub> _1W	337.0	49.1	18.6
C <sub>5</sub> H <sub>6</sub> _2W	328.6	41.3	18.1
C <sub>7</sub> H <sub>8</sub> _0W (C <sub>M</sub> )	374.3	54.4	42.1
C <sub>7</sub> H <sub>8</sub> _1W (C <sub>M</sub> )	365.6	45.6	42.2
C <sub>7</sub> H <sub>8</sub> _2W (C <sub>M</sub> )	357.9	40.8	40.1
C <sub>7</sub> H <sub>8</sub> _0W (C <sub>R</sub> )	393.6	65.1	48.4
C <sub>7</sub> H <sub>8</sub> _1W (C <sub>R</sub> )	376.4	52.1	45.3

<sup>a</sup>  $\Delta G_g$  and  $\Delta\Delta G_s$  values were calculated at the B3LYP/6-311++G(d,p) level.

for some of these cluster models are shown in Table 6. The inclusion of explicit water molecules does not significantly change the  $pK_a$  values for acetylene, cyclopentadiene or toluene. Those changes are within 1–3  $pK_a$  units (less than 10%). The inclusion of explicit water molecules for ethane, cyclopropane and cyclohexane lead to an interesting result. Optimization of the anionic cluster with a single water molecule for each of these molecules lead to the detachment of a proton from water molecule and creation of a cluster containing neutral molecule and OH<sup>-</sup>. The corresponding  $pK_a$  values were decreased approximately by 50%. Further addition of N water molecules always led to clusters containing neutral solute, ion OH<sup>-</sup> and (N - 1) water molecule. The  $pK_a$  of these clusters with 3–4 water molecules was always close to 13–15, i.e., the  $pK_a$  value of water in the presence of neutral solute. Thus, this refined model with explicit water molecules in the solute cavity does not necessary lead to better description of anion solvation for compounds that have  $pK_a$  values over 40. The transfer of a proton from water molecule to the anion of weak organic acids just means that probability of these compounds to be deprotonated in liquid water is very low what leads to high  $pK_a$  values. For such systems the traditional continuum model gives better and more reliable results.

The last point in our analysis concerns the solvation energy of a proton ( $\Delta G_s(\text{H}^+)$  term in the eq 3). We discussed the uncertainties of this value in the section 2 and in previous publications.<sup>10,29</sup> We used the theoretically derived  $\Delta G_s(\text{H}^+)$  value (-262.5 kcal/mol),<sup>29</sup> which is at the far end of the proposed range of  $\Delta G_s(\text{H}^+)$  values. If this solvation energy were larger, then the calculated  $pK_a$  values, shown in Table 5, would be larger. On the other hand if we assume that the  $\Delta G_s(\text{H}^+)$  value is lower (Tissandier et al.<sup>30</sup> obtained a theoretical estimation which is 1.4 kcal/mol lower than our value) than the suggested  $pK_a$  values will be lower by no more than 1  $pK_a$  units. We have shown that all other components of the thermodynamic cycle in the case of acetylene, as an example, are consistent and agree with the corresponding experimental values. Thus, the  $\Delta G_s(\text{H}^+)$  value of the proton used in this work should not lead to errors in the calculated  $pK_a$  values of more than by 2  $pK_a$  units in the direction of an increase in the proposed  $pK_a$ .

## 5. Conclusions

Using a combination of ab initio, DFT and continuum solvation methods, we have calculated gas phase and aqueous acidities for a set of hydrocarbons with high  $pK_a$  values, that

cannot be measured experimentally in aqueous solvent. The analysis of different terms directly used in the thermodynamic cycle for calculating  $pK_a$  values, such as gas-phase acidity, free energy of the proton abstraction reaction, and solvation energy, allows us to estimate that the errors in the  $pK_a$  value calculations do not exceed 1–2  $pK_a$  units, i.e., less than 10% of value of the expected  $pK_a$ 's. Inclusion of explicit water molecules in the solute cavity of compounds with  $pK_a$  values over 40 leads to inappropriate description of the corresponding anion solvation. In these cases the traditional dielectric continuum model leads to more accurate and reliable results. We find that acidity trends for several compounds in the gas phase and aqueous solvent are different due to the effects of aqueous solvation. We hope that these calculations could be helpful for experimentalists and might be used for the improving of the acidity scale for such weak organic acids.

**Acknowledgment.** We thank the staff and administration of the Advanced Biomedical Computing Center and the National Cancer Institute for their support of this project. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract No. N01-CO-56000. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government.

## References and Notes

- (1) March, J. *Advanced Organic Chemistry: Reactions, Mechanisms, and Structure*, 2nd ed.; McGraw-Hill Book Co.: New York, 1977.
- (2) Cram, D. J. *Fundamentals of Carbanion Chemistry*; Academic Press: New York, 1965.
- (3) Lias, S. G.; Bartmess, J. E.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. *J. Phys. Chem. Ref. Data, Suppl 1* **1988**, 17.
- (4) (a) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, 94, 7221. (b) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1998**, 109, 42. (c) Petersen, G. A.; Malick, D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frish, M. J. *J. Chem. Phys.* **1998**, 109, 10570. (d) Nicolaides, A.; Rauk, A.; Glukhotsev, M. N.; Radom, L. *J. Phys. Chem.* **1996**, 100, 17460. (e) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, 106, 1063. (f) Bauschlicher, C. W.; Partridge, H., Jr. *J. Chem. Phys.* **1995**, 103, 1788. (g) Wiberg, K. B.; Castejon, H. *J. Org. Chem.* **1995**, 60, 6327. (h) Merrill, G. N.; Kass, S. R. *J. Phys. Chem.* **1996**, 100, 17465. (i) Saunders, W. H., Jr. *J. Phys. Org. Chem.* **1994**, 7, 268.
- (5) Berkowitz, J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, 98, 2744.
- (6) Lowry, T. H.; Richardson, K. S. *Mechanism and theory in Organic Chemistry*, 3rd ed.; Harper & Row: New York, 1987.
- (7) (a) Jorgensen, W. L.; Briggs, J. M.; Gao, J. *J. Am. Chem. Soc.* **1987**, 109, 6857. (b) Jorgensen, W. L.; Briggs, J. M. *J. Am. Chem. Soc.* **1989**, 111, 4190.
- (8) Rashin, A. A.; Rabinowitz, J. R.; Banfelder, J. R. *J. Am. Chem. Soc.* **1990**, 112, 4133.
- (9) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, 95, 5610.
- (10) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Phys. Chem. A* **1997**, 101, 10075.
- (11) Topol, I. A.; Burt, S. K.; Rashin, A. A.; Erickson, J. W. *J. Phys. Chem. A* **2000**, 104, 866.
- (12) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys.* **1993**, 98, 1293.
- (13) Smith, B. J.; Radom, L. *J. Am. Chem. Soc.* **1993**, 115, 4885.
- (14) Miertus, S.; Tomasi, J. *Chem. Phys.* **1982**, 65, 239.
- (15) Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. *J. Mol. Struct.* **1988**, 180, 267.
- (16) Mikkelsen, K. V.; Agren, H.; Jensen, H. J. A.; Helgaker, T. *J. Chem. Phys.* **1988**, 89, 3086.
- (17) Rashin, A. A.; Bukatin, M. A.; Andzelm, J.; Hagler, A. T. *Biophys. Chem.* **1994**, 51, 375.
- (18) Chudinov, G. E.; Napolov, D. V.; Basilevsky, M. V. *Chem. Phys.* **1992**, 160, 41.
- (19) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, 94, 2027.
- (20) Rashin, A. A.; Young, L.; Topol, I. A. *Biophys. Chem.* **1994**, 51, 359.
- (21) Stefanovich, E. V.; Truong, T. N. *Chem. Phys. Lett.* **1995**, 244, 65.
- (22) Bachs, M.; Luque, F. J.; Orozco, M. *J. Comput. Chem.* **1994**, 15, 446.
- (23) Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Ringnalda, M.; Goddard, W. A., III; Honig, B. *J. Am. Chem. Soc.* **1994**, 116, 11875.
- (24) Tawa, G. J.; Martin, R. L.; Pratt, L. R.; Russo, T. V. *J. Phys. Chem.* **1996**, 100, 1515.
- (25) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, 100, 11775.
- (26) Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. *J. Phys. Chem.* **1994**, 98, 11059.
- (27) Levin, I. N. *Physical Chemistry*, 3rd ed.; McGraw-Hill: New York, 1988; Chapter 22.
- (28) Gilson, M. K.; Honig, B. *Proteins; Struct. Funct. Genet.* **1988**, 4, 7.
- (29) Tawa, G. J.; Topol, I. A.; Burt, S. K.; Caldwell, R. A.; Rashin, A. A. *J. Chem. Phys.* **1998**, 109, 4852.
- (30) Tissandier, M. D.; Cowen, K. A.; Feng, W. Y.; Gundlach, E.; Cohen, M. H.; Earhart, A. D.; Coe, J. V. *J. Phys. Chem. A* **1998**, 102, 7787.
- (31) Connolly, M. L. *J. Appl. Crystallogr.* **1985**, 18, 499.
- (32) Rashin, A. A. *J. Phys. Chem.* **1990**, 94, 1725.
- (33) Rashin, A. A.; Namboodiri, K. *J. Phys. Chem.* **1987**, 91, 6003.
- (34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, N.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, Revision D.4; Gaussian, Inc.: Pittsburgh, PA, 1995.
- (35) Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648.
- (36) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.
- (37) Murphy, R.; Pollard, W. T.; Friesner, R. A. *J. Chem. Phys.* **1997**, 106, 5073.
- (38) Jaguar 3.5; Schroedinger, Inc.: Portland, OR, 1998.
- (39) Murphy, R.; Beachy, M.; Friesner, R. A.; Ringnalda, M. *J. Chem. Phys.* **1995**, 103, 1481.
- (40) Corcelli, S. A.; Kress, J. D.; Pratt, L. R.; Tawa, G. J. *Pacific Symposium on Biocomputing '96*; World Scientific: NJ, 1995; p 143.
- (41) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Chem. Phys.* **1999**, 111, 10998.
- (42) Bartmess, J. E. *NIST Standard Reference Database*, 19B. NIST negative ion energetics database, Version 3.00. Standard Reference Data, National Institutes of Standards and Technology, Gaithersburg, MD 20899.
- (43) Berkowitz, J.; Ellison, G. B.; Gutman, D. *J. Phys. Chem.* **1994**, 98, 2744.
- (44) Engelking, P. C.; Lineberger, W. C. *J. Chem. Phys.* **1977**, 67, 1412–1417.
- (45) (a) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, 81, 2016. (b) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, 40, 287. (c) Cabani, S.; Gaianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, 10, 563.
- (46) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, 102, 3257.
- (47) Stewart, R. *The Proton: Applications to Organic Chemistry*; Academic Press: New York, 1985.
- (48) Juan, B.; Schartz, J.; Breslow, R. *J. Am. Chem. Soc.* **1980**, 102, 5741.
- (49) Kosower, E. M. *An Introduction to Physical Organic Chemistry*; John Wiley and Sons: New York, 1968.