JCTC Journal of Chemical Theory and Computation

pK_a Calculation of Some Biologically Important Carbon Acids - An Assessment of Contemporary Theoretical Procedures

Junming Ho and Michelle L. Coote*

ARC Centre of Excellence for Free-Radical Chemistry and Biotechnology, Research School of Chemistry, Australian National University, Canberra ACT 0200, Australia

Received August 13, 2008

Abstract: In this study, the aqueous pK_a values for 13 neutral, 10 cationic, and 5 anionic carbon acids, including amino acids, peptides, and related species have been calculated using the high level ab initio composite procedure, G3MP2+//BMK, combined with solvation energies that were calculated using the CPCM-(UAKS/UAHF), COSMO-RS, and SM6 continuum models. The pK_as were further calculated using three schemes, namely the direct method and the proton exchange method as well as the inclusion of an explicit solvent water molecule. The results of this study indicate that the direct method is unsuitable for computing the pK_a of carbon acids, whereas the other two schemes perform significantly better with varying degrees of success, depending on the charge of the carbon acid. Specifically, the combination of the proton exchange scheme and CPCM-UAKS model performed particularly well for neutral species, with mean absolute deviations (MADs) of ~1 pK_a unit. The ionic species were more problematic, though the combination of the proton exchange scheme and the SM6 and CPCM-UAKS models performed reasonably well for the cationic and anionic acids, respectively. The inclusion an explicit water molecule generally improved the calculated values for anionic carbon acids.

1. Introduction

Carbon acids are ubiquitous in nature and in the chemistry laboratory. Amino acids and peptides are prominent examples of carbon acids that occur naturally. It is well-known that all living organisms synthesize peptides and proteins that are composed entirely of amino acids in the L-configuration. However, spontaneous racemization can result in the generation of D-residues during the life span of the protein. For example, the accumulation of D-aspartic acid in the brain,[1] tooth enamel,[2,3] bones,[4] and lens proteins[5] has been associated with aging and can contribute to loss of tissue functions. The relative rates of racemization of different amino acid residues at proteins have been reported,[6,7] and this relates to the acidity of the α-C-H protons, which depends on the nature of the amino acid side chain, the amino acid sequence, and the peptide conformation.

D-Amino acids are also widely found in prokaryotic (bacteria) peptides and less commonly in eukaryotic peptides, which are invariably components of venoms.[8−15] Contrary to spontaneous racemization as a result of aging, these D-amino acids are critical for the biological activity of these peptides, and specialized enzymes such as racemases and enolases are used to catalyze the heterolytic cleavage of these very stable α-C-H bonds. The mechanism as to how enzymes activate these bonds for proton transfer, i.e. by increasing the kinetic acidity of these protons, has been a subject of intense research.[16−23]

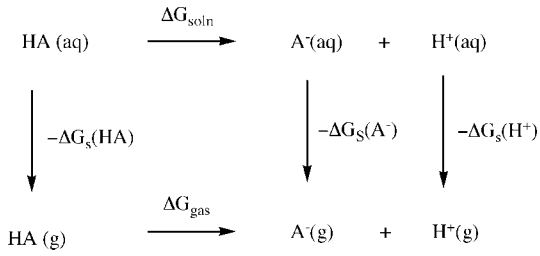
The acidity of α-C-H protons also plays an important role in chemical synthesis. For example, the carbanions generated from deprotonation of the α-carbonyl protons of ketones and aldehydes are routinely used in nucleophilic substitution reactions for forming carbon−carbon bonds. Cyclic dipeptides or diketopiperazines have also been explored as chiral auxiliaries for the stereoselective synthesis of amino acids, where regioselective deprotonation of α-C-H protons is critical.[24−26]

Clearly, an understanding of substituent effects on the acidity of the α-C-H protons has important implications in

* Corresponding author e-mail: mcoote@rsc.anu.edu.au.

10.1021/ct800335v CCC: $40.75
Published on Web 01/09/2009

Scheme 1. pK_a Calculation via the Direct Method



both biological systems as well as in chemical synthesis. The acid dissociation constant, pK_a, is the most common measure of the thermodynamic acidity. However, accurate measurements of pK_a values are further complicated by the extremely weak acidity of these carbon protons, with pK_a values typically in the range of 20 or higher.[27,28] Thus, very sensitive methods are required to determine the dissociation constants of these acids.

In recent years, Richard et al. have pioneered the use of NMR methods for measuring the pK_a values of a wide variety of carbon acids with different functionalities, viz. carboxylic acids, esters, amides, derivatives of amino acids, and heterocycles as well as small peptides in aqueous solution.[17,18,22,29−36] Bordwell et al. have also compiled a large database of pK_a values for various organic acids, including carbon acids, in DMSO.[27,28] This has led to a better understanding of the substituent effects and the mechanisms employed by enzymes to accelerate the deprotonation of the α-protons. For example, pyridoxal 5′-phosphate (PLP) is known to catalyze carbon deprotonation of α-amino acids at enzyme active sites by formation of an imine. Through quantitative measurements of pK_a, Richard et al. have found that a dramatic increase in acidity at the α-carbons, by ∼7 pK_a units, may be achieved through the formation of such an intermediate.[22,37]

In light of the recent reports on acidity constants for carbon acids of amino acid derivatives and peptides and their associated exciting insights into enzymatic mechanisms, the theoretical calculation of pK_a of such carbon acids has become even more appealing. There has been significant effort targeted at making reliable predictions of pK_a values using quantum chemical methods. Liptak and Shields utilized the thermodynamic cycle shown in Scheme 1 in which gas-phase free energies obtained via high level *ab initio* methods (e.g., CBS-QB3 and G-*n* models) are combined with the solvation free energies obtained from continuum solvation models.[38,39] Despite the intrinsically larger errors in continuum solvation calculations, the authors showed that it was possible to predict pK_a values of carboxylic acids to within 0.5 pK_a units, presumably due to systematic error cancelation. More generally, they suggested that accurate pK_a values could be obtained by means of a proton exchange scheme that allowed for further error cancelation.[40] Similar approaches have been used for calculating the pK_a values of carboxylic acid derivatives,[41] alcohols,[42] carbenes,[43] amines,[44] phosphoranes,[45] substituted phenols,[46] and pharmaceutically important compounds.[47] Various assessment and methodology-development studies employing large data sets of various neutral and charged organic and inorganic acids have also been reported.[47−52] In general, these studies have found that accurate pK_a values can be obtained through the combination of high-level ab initio methods with continuum solvation models, particularly when proton exchange reactions are used.

However, because accurate experimental pK_a values for carbon acids are relatively scarce, far fewer studies have been carried out to examine the performance of computational methods on these systems. Brinck et al. have studied the performance of the polarizable continuum model (PCM) for the solvation of small aliphatic carbanions in organic solvents DMF and THF.[53] The deviation with experimental values was found to be quite large (∼20−30 kJ/mol), and this has been attributed to the neglect of short-range solvent effects in continuum models. Fu et al. have made use of continuum solvent model combined with the proton exchange scheme for the pK_a calculation of a large data set of organic molecules, including carbon acids, in DMSO.[51] In their test set, the carbon acids include ketones and substituted aliphatic systems, and the calculated values are generally in good agreement with experiment. Gao et al. have also examined the use of continuum methods and QM/MM-Ewald simulations to calculate the aqueous pK_a of the acetate anion using both direct method and the proton exchange scheme.[54] The authors found that the QM/MM-ewald protocol yields the best result using the latter scheme, with calculated values within 2 pK_a units of experiment. However, further testing of this method is necessary to establish its general applicability to other carbon acids.

In this light, it is thus important to establish whether such procedures are suitable for the calculation of pK_a values for carbon acids, particularly those in biological systems. In this study, we wish to examine the performance of several popular procedures for calculating the pK_a of a range of neutral, cationic, and anionic carbon acids in aqueous solution. It is also worth highlighting that most of these molecules are not included in the training sets for the parametrization of various continuum solvation models and should therefore provide an objective and rigorous test of these methods.

2. pK_a Calculation Methods

Invariably, most pK_a calculations involve the representation of the acid dissociation process as a sum of several intermediate steps such as in the thermodynamic cycle shown below in Scheme 1. By virtue of Hess's law, the free energy of acid dissociation in solution, ΔG_{soln}, is equal to

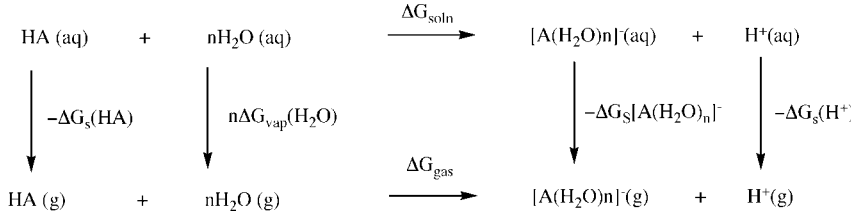
$$\Delta G^*_{soln} = \Delta G^*_{gas} + \Delta\Delta G^*_{solv} \quad (1)$$

where $\Delta\Delta G^*_{solv} = \Sigma\Delta G_{solv,products} - \Sigma\Delta G_{solv,reactants}$. The "*" symbol is used for a standard state of 1 mol/L in any phase. The K_a and pK_a may be obtained through the thermodynamic relationship

$$\Delta G^*_{soln} = -RT \ln K_a \quad (2)$$

Eq 1 allows us to decompose the errors in the acidity constant into a gaseous component, ΔG_{gas}, and a solvation component, $\Delta\Delta G_{solv}$. Gas-phase energies and hence acidities can now be calculated with chemical accuracy provided that electron correlation effects are included through appropriate post-Hartree−Fock or density functional methods.[55−68] Liptak et al. have found that, for gas-phase energies, the CBS-QB3

Scheme 2. pK_a Calculation with Explicit Water Molecules



method gives the most accurate results.[38,39] In this study, we examine the use of various levels of theories, such as CBS-QB3 and G*n* composite methods for calculating ΔG_{gas}.

However, the situation is much less satisfactory in solution, mostly due to the difficulty of treating the solvent−solute interactions rigorously. In particular, the acid dissociation involves the formation of charged species starting from neutral molecules. Short-range intermolecular interactions (e.g., ion-dipole and hydrogen bonding) are considered to be particularly important in solvation of charged species. Since these effects are not explicitly taken into account by continuum solvation models,[50,69−71] the direct calculation of pK_a values via the above thermodynamic cycle is likely to incur significant errors. Two general procedures are used to remedy this deficiency in continuum solvation models.

First, an isodesmic reaction has been shown to yield very accurate pK_a values (±1 pK_a unit) for moderately strong acids.[41,42,72,73] In this study, a proton exchange reaction between the acid and a reference acid (HR) with known pK_a is considered.

$$HA(aq) + R^-(aq) \Leftrightarrow HR(aq) + A^-(aq);\ \Delta G_{EX} \qquad (3)$$

Since the number of charged species is conserved on both sides of the equation, one can expect some cancelation of the errors due to the neglect of short-range solvent effects. Accordingly, the equilibrium constant, K_{EX}, for reaction 3 can be calculated from eqs 1and 2. The acid dissociation constant, K_a(HA), can then be determined by the product of K_a(HR) and $K_{EX.}$ Unfortunately, the success of this approach depends heavily on the choice of reference acid, with best results expected if HR is structurally similar to HA. The accuracy of the calculated value also depends very much on the accuracy of the experimental K_a(HR).

The second approach involves inclusion of explicit solvent molecules. There are several variants of this approach, including the cluster-continuum model[51,74] and the implicit-explicit solvent approach.[48] In the latter approach, the pK_a of an acid is calculated via the thermodynamic cycle in Scheme 2. Based on eq 1, the free energy of acid dissociation in solution is therefore given by eq 4:

$$\Delta G_{soln} = \Delta G_{gas} + \Delta G_s(H^+) + \Delta G_s[(A(H_2O)_n)^-] - \Delta G_s(HA) + n\Delta G_{vap}(H_2O) \qquad (4)$$

Kelly et al. have used this thermodynamic cycle to calculate ΔG_{solv} and pK_a in aqueous solution for a variety of organic acids and found that the agreement with experiment is significantly improved when one explicit water molecule was included in the thermodynamic cycle.[48]

In this study, we examine the three approaches highlighted above, namely the direct method, the proton exchange method, and the implicit-explicit solvent method, for calculating pK_as. The test set of molecules, mainly amino acids and derivatives, has been compiled from a list of studies that were mostly conducted by the research group of Richard's.[20,22,29,30,33−37,75−79] These molecules are further categorized as neutral, cationic, and anionic as shown in Figure 1.

3. Theoretical Procedures

Gaussian 03 software[80] has been used for all gas-phase ab initio molecular orbital theory[81] and density functional theory calculations.[82] The gas-phase acid dissociation free energy of several carbon acids has been calculated at the G3MP2,[83] G3,[62] and CBS-QB3[84] levels of theory. The Gaussian methods (G3, G3MP2, and various other modified versions) approximate QCISD(T) energies with a large triple-ζ basis using cheaper QCISD(T)/6−31G(d) calculations in conjunction with additivity corrections, obtained at the MP2, MP3, and/or MP4 levels of theory.[62,83] The complete basis set methods (CBS) are a model chemistry that makes use of a complete basis set extrapolation of the correlation energy, which is performed at the MP2 level of theory and then corrected to the CCSD(T) level via additivity corrections.[84] These high-level composite procedures have been designed particularly for the prediction of reliable energies of molecules in the gas phase and have been demonstrated to provide an accuracy of 1−2 kcal/mol when assessed against large test sets of thermochemical data.[62,84] In addition to these standard procedures, calculations were also performed using the modified procedures, G3+ and G3MP2+, in which calculations with the 6−31G(d) basis set have been replaced with the 6−31+G(d) basis set, so as to allow for an improved description of anionic species.

In their original forms, the G3MP2 and G3 methods both employ the geometries which are optimized at the MP2(Full)/6−31G(d) level; CBS-QB3 employs B3-LYP/CBSB7 optimized geometries. However, in the present work we sought to identify an optimal procedure for the calculation of geometries that balanced accuracy and computational expense. To this end, proton affinities of a selection of carbon acids were first calculated at a consistent level of theory, MP2(Full)/6−311+G(d, p), using geometries that had been optimized at various levels of theory. On the basis of this initial study, the BMK/6−31+G(d) level of theory was selected for geometry optimizations and frequency calculations for the remainder of the study. Scale factors for the BMK/6−31+G(d,p) vibrational frequencies have been used for the free energy calculations.[85]

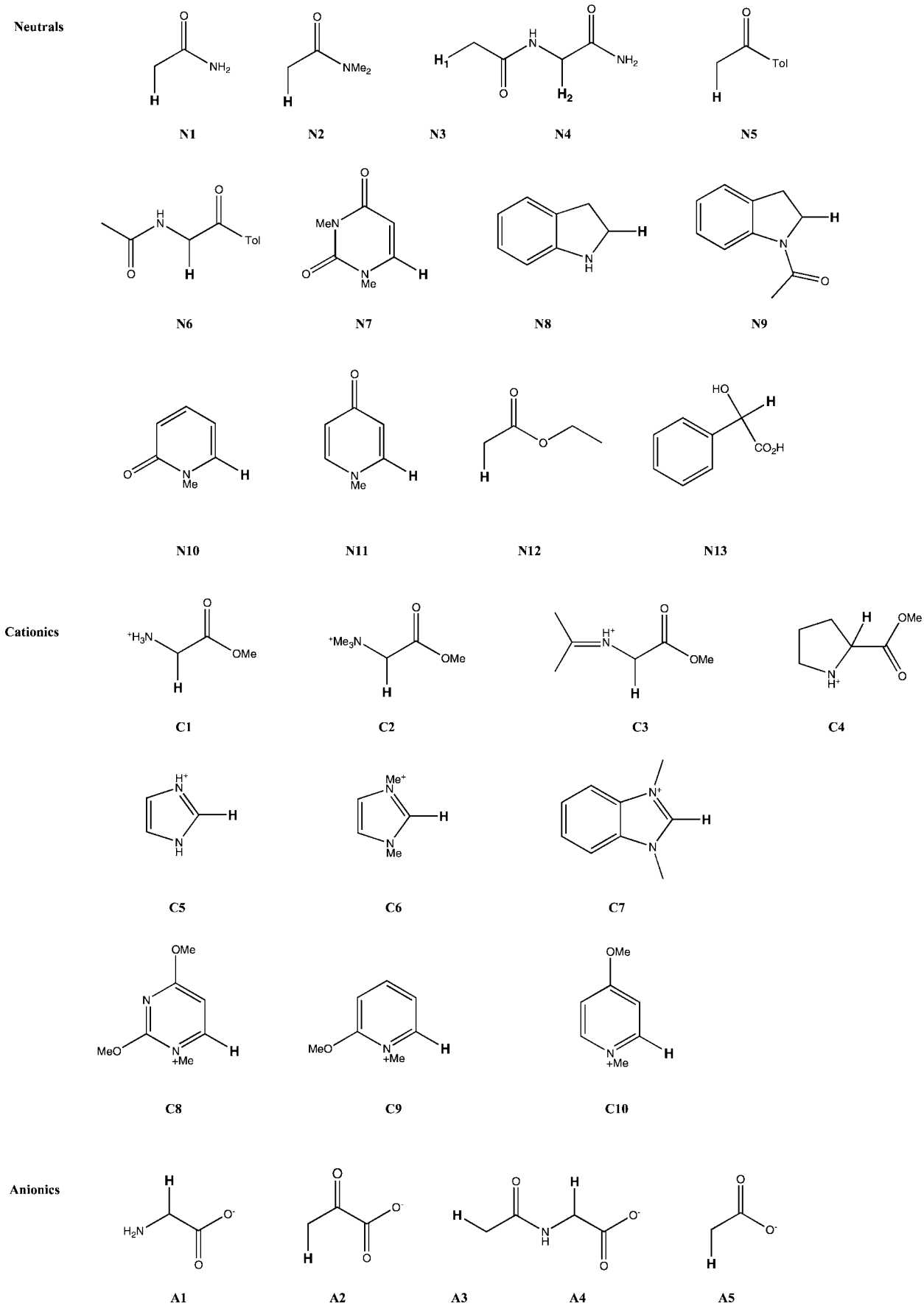
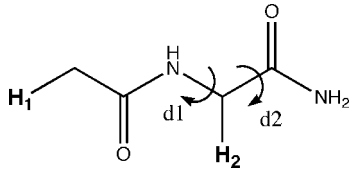


Figure 1. Test set of studied carbon acids.

For some of the larger carbon acids such as **N3/N4**, several conformations are possible. Because the acids examined in this study are relatively small, we have performed a grid search on these molecules. This involves optimization of all

possible conformers generated from combinations of rotations about certain chemical bonds. The rotations were examined at 120° and 180° resolution for sp^2 and sp^3 hybridized centers, respectively. For instance, the rotations examined for the acid **N3/N4** are illustrated where d1 and d2 were examined at 120° and 180° resolution, respectively. The amide bond was fixed in its thermodynamically preferred trans-configuration.



In an earlier study by Liptak and Shields, the authors showed that calculations using continuum models on the lowest energy gas-phase conformer and the conformationally averaged structure gave comparable results.[40] Accordingly, in this study we only consider the solvation free energies on the lowest energy gas-phase conformer. The lowest energy gas-phase optimized geometries and their associated conformers are provided in the Supporting Information.

The free energies of solvation were evaluated using several popular procedures as recommended by several earlier studies of other types of acid.[38,39,43,46] The conductor-polarizable continuum model (CPCM)[86,87] was used to compute solvation free energies at the HF and B3LYP levels of theory in conjunction with various basis sets. These calculations were carried out using GAUSSIAN 03 software[80] using the radii of the united atom topological model, optimized for the Hartree−Fock and DFT methods (UAHF and UAKS), and default values for the other parameters. All geometries of the studied species have been optimized fully in the presence of solvent using different basis sets at the level of HF and B3LYP.

In addition, the free energies of solvation were also computed using the COSMO-RS[88−90] and SM6[91] models. The COSMO-RS model is a variant of the CPCM model (conductor-like screening model for real solvents) that describes the interactions in a fluid as local interaction of molecular surfaces, the interaction energies being quantified by the values of the two screening charge densities that form a molecule contact.[88,89] The resulting energies are presumably more accurate than a typical PCM calculation because the real character of the solvent is taken into account and not a simple homogeneous continuum. The ADF package[92] was used to compute the COSMO-RS solvation free energies on the gas-phase geometries at the BP/TZP level of theory, with the rest of the parameters (e.g., atomic cavity radii, radius of the probing sphere, and cavity construction) kept as default values.[92] We have computed the ADF COSMO-RS solvation free energies for a selection of molecules and compared them with values obtained from the original paper,[90] where the COSMO-RS model was parametrized slightly differently. The agreement is generally very good (within 0.5 kcal/mol), and the data are tabulated in Table S3 in the Supporting Information.

The SM6 model is based on a generalized born approach which uses a dielectric continuum to treat bulk electrostatics effects combined with atomic surface tensions to account for first shell solvent effects, and it has been shown to give aqueous solvation free energies accurate to within ~2 kJ/mol for neutral species.[91] The SM6 free energy of solvation is calculated on the gas-phase geometries at the B3LYP level using the GAMESSPLUS program.[93]

Table 1. Absolute Deviations of Proton Affinities[a] (kJ/mol) Calculated on Geometry Optimized at Various Levels of Theory Relative to MP2(full)/6-311+G(d,p)

geometry	N1	N3	N4	C4	A4
RHF/6−31+G(d)	0.58	0.31	1.78	2.03	5.04
RHF/6−31+G(d,p)	0.26	0.00	1.75	1.78	4.88
RHF/6−311+G(d,p)	0.86	0.70	0.72	1.40	4.74
B3LYP/6−31+G(d)	2.31	0.55	0.31	0.27	1.16
B3LYP/6−31+G(d,p)	0.70	0.49	0.13	0.20	1.35
B3LYP/6−311+G(d,p)	0.75	0.60	0.08	0.11	1.20
BMK/6−31+G(d)	1.04	0.85	0.32	0.26	0.46
BMK/6−31+G(d,p)	0.84	0.72	0.08	0.02	0.51
BMK/6−311+G(d,p)	1.05	0.84	0.02	0.07	0.28
MP2(full)/6−31+G(d)	0.13	0.15	0.05	0.03	0.24
MP2(full)/6−31+G(d,p)	0.23	0.17	0.42	0.04	0.21
MP2(full)/6−311+G(d,p)	0.00	0.00	0.00	0.00	0.00

[a] Proton affinities are calculated as E_e(conjugate base) − E_e(conjugate acid).

As noted above, the pK_a values for the carbon acids in this assessment study have been calculated using the direct method and the proton exchange method. In the direct method, we have used the most recent experimental-theoretical values of −26.3 kJ/mol[38] and −1112.5 kJ/mol[94] for the gas-phase Gibbs free energy of H^+, $G^o(g, H^+)$, and solvation energy of H^+ in water, $\Delta G_s^o(H^+)$. Calculation of the gas-phase energies are for a standard state of 1 atm, but solvation energies use a standard state of 1 mol/L; therefore, the value of 7.9 kJ/mol which corresponds to $RT\ln(24.46)$ has been added to gas-phase energies in Scheme 1.

4. Results

Gas-Phase Proton Affinities. Although the largest source of error in pK_a calculations is likely to be the treatment of solvation effects (and how this error is mitigated through the pK_a calculation method), it is nonetheless necessary to ensure that the chosen electronic-structure procedures used are capable of delivering accurate gas-phase proton affinities in a cost-effective manner. To this end, we first investigated the effect of the level of theory used in the geometry optimization on the proton affinities for a small selection of neutral, cationic, and anionic carbon acids. Table 1 shows the proton affinities as calculated at a consistent level of theory, using geometries obtained using HF, B3LYP, BMK, and MP2(full) in conjunction with a variety of basis sets. As shown in Table 1, the proton affinities calculated on these geometries appear to be relatively insensitive to the level of theory since significant error cancelation is expected in the reaction energies. Even at the lowest level of theory studied, RHF/6−31+G(d) geometries are generally within 2 kJ/mol of the corresponding MP2(full)/6−311+G(d,p) level, though there are some exceptions to this. The lowest cost procedure that consistently delivered proton affinities within 1 kJ/mol of the corresponding benchmark level was BMK/6−31+G(d), hence this procedure has been adopted for the remainder of the study.

Table 2. Experimental and Calculated[a] Gas-Phase Free Energies (kJ/mol) of the Deprotonation Reaction at 298 K

level of theory	acetamide (**N1**)	N-methylpyrazole	N-methylimidazole	1,3-dimethyluracil[b]
G3MP2//BMK	1551.8 (24.6)	1574.8 (2.1)	1583.5 (7.2)	1540.3 (3.6)
G3MP2+//BMK	1553.7 (26.5)	1576.5 (0.4)	1585.0 (5.7)	1540.3 (3.6)
G3//BMK	1554.1 (26.9)	1577.2 (0.3)	1586.6 (4.1)	1541.7 (2.2)
G3+//BMK	1555.3 (28.1)	1577.6 (0.7)	1586.3 (4.4)	1542.1 (1.8)
CBS-QBMK	1554.9 (27.7)	1577.9 (1.0)	1585.6 (5.1)	1539.6 (4.3)
W1/BMK	1554.0 (26.8)	–	–	–
experimental	1527.2 ± 12.6[95]	1576.9 ± 2.9[100]	1590.7 ± 4.2[100]	1543.9 ± 8.4[101]

[a] Based on BMK/6−31+G(d) optimized geometries and BMK/6−31+G(d,p) scaled frequencies. Absolute deviation from experiment in parentheses. [b] ΔH_{acid}.

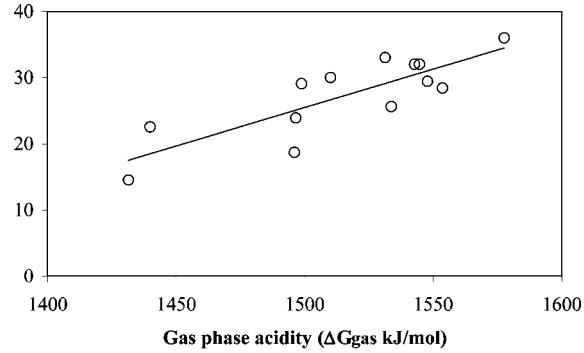


Figure 2. The correlation between gas-phase (G3MP2+//BMK/6−31+G*) and aqueous acidities of neutral carbon acids. Least-squares correlation $pK_a = 0.12\Delta G_{gas} - 148.55$ ($r = 0.68$).

Using the BMK/6−31+G(d) geometries, we then investigated the accuracy of the various ab initio procedures for the calculation of the gas-phase acidities. The calculated and experimental gas-phase proton affinities (as free energies) are provided in Table 2. The free energies computed using the various composite methods are all in good agreement with one another. Furthermore, with the exception of **N1**, the calculated values are generally within the error margins of the experimental values. It has been acknowledged by Hare et al.[95] that the true value for **N1** is likely to be on the high side of their experimental value, which would be in accordance with our calculated values. To verify this, a more accurate W1[96] calculation has been performed on **N1**, and there is good agreement with the value calculated from the other composite methods. Thus, on the basis of accuracy and computational cost, we have chosen the G3MP2+//BMK method for calculating the gas-phase free energies of the carbon acids in this study and estimate that the gas-phase errors associated with this level of theory are of the order of 5 kJ mol^{-1} (i.e., chemical accuracy).

We were also interested in whether there is any correlation between the gas-phase and solution acidities. As shown in Figure 2, there is a relatively strong linear correlation between gas-phase and aqueous acidities for the neutral carbon acids ($r^2 = 0.68$ and $pK_a = 0.12 \cdot \Delta G_{gas} - 148.6$). Of greater interest is that substituent effects on aqueous acidities are also manifested in the gas-phase data. Specifically, the activating effect of the N-acetyl group is clearly seen in **N9** (*cf.* **N8**) and **N6** (*cf.* **N5**), where the electron-withdrawing group is expected to stabilize the adjacent carbanionic charge via inductive effects.[36] The experimental pK_a of **N9**, **N8**, **N6**, and **N5** are 33, 36, 14.5, and 18.7, and their corresponding gas-phase acidities (ΔG_{gas}) are 1531, 1577, 1431, and 1496 kJ/mol, respectively. The weaker aqueous acidity of an α-carbonyl proton in an amide compared to an ester is also mirrored in the gas-phase data. A similarly strong correlation was observed in the cationic systems ($r^2 = 0.71$), although a somewhat weaker correlation was observed in the anionic systems ($r^2 = 0.33$). Presumably, this is due to the divalent anionic conjugate bases of these acids where the solvation energies are expected to dominate the trends in aqueous acidities.

Solvation Energies and pK_a Values. Having identified suitable electronic structure methods for calculating the gas-phase proton affinities, these procedures were then used in conjunction with a variety of solvation models to calculate the corresponding pK_a values. In calculating the solvation energies and pK_a values, there are a number of additional variables to consider, including the choice of solvation model, the level of theory at which it is applied, whether or not explicit solvent molecules are included in the calculation, and whether the pK_a value is calculated via a direct or proton exchange approach. To simplify the experimental design, in this section we consider only the CPCM model and study the effect of level of theory on the (directly calculated) pK_a values. Then, having selected an appropriate level of theory for the solvation energy calculations, in subsequent sections we explore the effect of solvation model, pK_a calculation method on the accuracy of the results.

Table 3 shows the pK_a values for a selection of neutral, anionic, and cationic carbon acids, in which the gas-phase energies were calculated at a consistent high level of theory (G3MP2+//BMK), and solvation energies were calculated using CPCM at various levels of theory. The geometry was fully optimized in the presence of solvent at each of the studied levels of theory. As shown in Table 3, increasing the basis set has a minimal effect on the accuracy of the pK_a for the neutral, cationic, and anionic carbon acids, and the smaller basis set 6−31+G(d) is sufficient for the solvation free energy calculations. However, there are significant differences between the CPCM results at the HF and B3LYP levels of theory in a number of acids (such as **N6** and **A4**), and thus both methods are retained for the remainder of the study.

The COSMO-RS model in ADF has been parametrized to some extent, and the BP/TZP level of theory was used as recommended.[92] The SM6 is a density functional theory continuum solvation model and can be used in conjunction with any good density functional, including the mPW0, B3LYP, and B3PW91 functionals.[91] As such, the SM6 solvation free energies have been computed at the B3LYP/

Table 3. Aqueous pK_a Values for Selected Neutral, Cationic, and Anionic Carbon Acids Calculated Using CPCM Solvent Models with Various Basis Sets at 298 K[a]

method	level of theory	N1	N2	N6	C1	C2	A1	A3	A4
CPCM	B3LYP/6−31+G(d)	35.6	37.0	21.4	24.6	24.0	43.6	38.9	42.0
CPCM	B3LYP/6−31+G(d,p)	35.7	37.1	21.5	24.8	24.1	43.7	39.1	42.2
CPCM	B3LYP/6−311+G(d,p)	35.4	36.9	21.4	23.5	24.0	42.8	38.5	41.8
CPCM	HF/6−31+G(d)	35.3	36.3	16.9	25.0	24.4	42.5	40.5	39.8
CPCM	HF/6−31+G(d,p)	35.3	36.4	17.1	25.2	24.6	42.0	40.5	39.9
CPCM	HF/6−311+G(d,p)	35.2	36.4	17.2	25.2	24.4	41.7	40.3	39.7

[a] All associated gas-phase calculations performed using the G3MP2+//BMK/6−31+G(d) level of theory. pK_a values calculated via the direct method. B3LYP calculations used UAKS radii; HF calculations used UAHF radii.

Table 4. Calculated[a] and Experimental Aqueous Acid Dissociation Constants at 298 K for Neutral Carbon Acids[b]

carbon acid	direct method				proton exchange method				
	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	expt
N1	35.6 (7.2)	35.3 (6.9)	28.6 (0.2)	28.1 (−0.3)	reference	reference	reference	reference	28.4 ± 0.5[33]
N2	37.0 (7.6)	36.3 (6.9)	31.7 (2.3)	28.5 (−0.9)	29.8 (0.4)	29.4 (0.0)	31.5 (2.1)	28.8 (−0.6)	29.4 ± 0.5[33]
N3	35.9 (6.8)	39.3 (10.2)	31.2 (2.1)	29.5 (0.4)	28.6 (−0.5)	32.3 (3.2)	31.0 (1.9)	29.8 (0.7)	29.1[36]
N4	32.1 (8.2)	35.9 (12.0)	24.7 (0.8)	25.8 (1.9)	24.9 (1.0)	29.0 (5.1)	24.5 (0.6)	26.1 (2.2)	23.9[36]
N5[e]	25.6 (6.9)	21.8 (3.1)	22.8 (4.1)	20.5 (1.8)	18.4 (−0.3)	14.9 (−3.8)	22.7 (4.0)	20.8 (2.1)	18.7[36]
N6	21.4 (6.9)	16.9 (2.4)	16.7 (2.2)	17.7 (3.2)	14.2 (−0.3)	10.0 (−4.5)	16.5 (2.0)	18.0 (3.5)	14.5[36]
N7	38.4 (8.4)	38.2 (8.2)	35.6 (5.6)	32.5 (2.5)	31.1 (1.1)	31.2 (1.2)	35.5 (5.5)	32.8 (2.8)	∼30[77]
N8	44.3 (8.3)	42.8 (6.8)	48.0 (12.0)	39.9 (3.9)	37.1 (1.1)	35.9 (−0.1)	47.9 (11.9)	40.2 (4.2)	∼36[78]
N9	41.3 (8.3)	40.6 (7.6)	41.3 (8.3)	36.1 (3.1)	34.1 (1.1)	33.7 (0.7)	41.2 (8.2)	36.4 (3.4)	∼33[78]
N10[e]	41.7 (9.7)	41.2 (9.2)	40.2 (8.2)	36.2 (4.2)	34.4 (2.4)	34.3 (2.3)	40.1 (8.1)	36.5 (4.5)	32 ± 2[77]
N11[e]	38.1 (6.1)	35.7 (3.7)	39.1 (7.1)	36.0 (4.0)	30.8 (−1.2)	28.8 (−3.2)	38.9 (6.9)	36.3 (4.3)	32 ± 2[77]
N12	33.9 (8.3)	33.4 (7.8)	28.3 (2.7)	24.7 (−0.9)	26.6 (1.0)	26.4 (0.8)	28.2 (2.6)	25.0 (−0.6)	25.6 ± 0.5[30]
N13	29.7 (7.1)	27.9 (5.3)	23.4 (0.8)	18.8 (−3.8)	22.4 (−0.2)	21.0 (−1.6)	23.2 (0.6)	19.1 (−3.5)	22.6[79]
AD_{max}	9.7	12.0	12.0	4.2	2.4	5.1	11.9	4.5	–
MAD	7.7	6.9	4.3	2.4	0.9	2.2	4.5	2.7	–

[a] Signed errors are shown in brackets. [b] All associated gas-phase calculations performed using the G3MP2+//BMK/6−31+G(d) level of theory. All solvation energy calculations performed at the B3LYP/6−31+G(d) level of theory unless noted otherwise. [c] Solvation energy calculations performed at the HF/6−31+G(d) level of theory. [d] Solvation energy calculations performed at the BP/TZP level of theory. [e] MP2/6−31+G(d) geometries, frequencies, and corresponding scale factors were used due to convergence problems during geometry optimization.

Table 5. Calculated[a] and Experimental Aqueous Acid Dissociation Constants at 298 K for Cationic Carbon Acids[b]

carbon acid	direct method				proton exchange method				
	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	expt
C1	24.6 (3.6)	25.0 (4.0)	27.2 (6.2)	26.2 (5.2)	reference	reference	25.3 (4.3)	reference	21 ± 1[34]
C2	24.0 (6.0)	24.4 (6.4)	19.9 (1.9)	24.5 (6.5)	20.4 (2.4)	20.5 (2.5)	reference	19.3 (1.3)	18 ± 1[34]
C3	18.9 (4.9)	17.4 (3.4)	13.5 (−0.5)	17.2 (3.2)	15.3 (1.3)	13.5 (−0.5)	11.6 (−2.4)	12.0 (−2.0)	14 ± 1[22]
C4	23.7 (2.7)	24.5 (3.5)	25.5 (4.5)	28.4 (7.4)	20.1 (−0.9)	20.6 (−0.4)	23.6 (2.6)	23.3 (2.3)	21 ± 1[75]
C5	24.4 (0.6)	22.0 (−1.8)	26.6 (2.8)	30.9 (7.1)	20.8 (−3.0)	18.0 (−5.8)	24.7 (0.9)	25.7 (1.9)	23.8 ± 0.5[29]
C6	24.1 (1.1)	24.4 (1.4)	24.1 (1.1)	29.7 (6.7)	20.5 (−2.5)	20.5 (−2.5)	22.2 (−0.8)	24.6 (1.6)	23.0 ± 0.5[29]
C7	21.6 (0.0)	22.3 (0.7)	21.4 (−0.2)	24.9 (3.3)	18.0 (−3.6)	18.4 (−3.2)	19.5 (−2.1)	19.8 (−1.8)	21.6 ± 0.5[29]
C8	28.8 (−4.2)	28.6 (−4.4)	29.0 (−4.0)	32.3 (−0.7)	25.2 (−7.8)	24.6 (−8.4)	27.1 (−5.9)	27.2 (−5.8)	33 ± 2[76]
C9	30.8 (−3.2)	30.8 (−3.2)	32.1 (−1.9)	36.9 (2.9)	27.2 (−6.8)	26.9 (−7.1)	30.2 (−3.8)	31.7 (−2.3)	34 ± 2[76]
C10	29.8 (−3.2)	29.6 (−3.4)	31.3 (−1.7)	36.7 (3.7)	26.2 (−6.8)	25.6 (−7.4)	29.4 (−3.6)	31.5 (−1.5)	33 ± 2[76]
AD_{max}	6.0	6.4	6.2	7.4	7.8	8.4	5.9	5.8	–
MAD	2.9	3.2	2.5	4.7	3.9	4.2	2.9	2.3	–

[a] Signed errors are shown in brackets. [b] All associated gas-phase calculations performed using the G3MP2+//BMK/6−31+G(d) level of theory. All solvation energy calculations performed at the B3LYP/6−31+G(d) level of theory unless noted otherwise. [c] Solvation energy calculations performed at the HF/6−31+G(d) level of theory. [d] Solvation energy calculations performed at the BP/TZP level of theory.

6−31+G(d) level of theory. Having selected appropriate levels of theory for the gas- and solution-phase calculations, we now examine the effects of solvation model and pK_a calculation method on the accuracy of the results for the neutral, cationic, and anionic acids. These results, including mean absolute deviation (MAD), maximum absolute deviations (AD_{max}), and signed errors (in parentheses), are presented in Tables 4, 5 and 7, respectively. It should be noted that the MAD for the proton-exchange pK_a values also provides a measure of the accuracy of the *relative* values of pK_a, as obtained by the direct method.

5. Discussion

Neutral Carbon Acids. Shown in Table 4 are the calculated acid dissociation constants for the neutral carbon acids using the direct and proton exchange methods. In the latter approach, acetamide was chosen as the reference acid.

Table 6. Multipole Derived Atomic Charges Computed at the BP/TZP Level of Theory

carbon acid	atomic charge[a]	formal charge	carbon acid	atomic charge[b]	formal charge
N1	+0.33	−1	C1	+0.6	+1
N2	+0.28	−1	C2	−0.53	+1
N3	+0.32	−1	C3	−0.15	+1
N4	+0.33	−1	C4	+0.25	+1
N5	+0.30	−1	C5	−0.10	+1
N6	+0.27	−1	C6	−0.42	+1
N7	−0.27	−1	C7	−0.46	+1
N8	−0.36	−1	C8	−0.45	+1
N9	−0.32	−1	C9	−0.42	+1
N10	−0.29	−1	C10	−0.42	+1
N11	−0.34	−1	–	–	–
N12	+0.36	−1	–	–	–
N13	+0.27	−1	–	–	–

[a] Atomic charge on the α carbon of the conjugate base of the neutral carbon acid. [b] Atomic charge on the nitrogen adjacent to the α carbon in the cationic acid.

Using the direct method, the mean absolute deviation (MAD) from experiment for CPCM-UAKS and CPCM-UAHF are in excess of 7 pK_a units, which is unacceptably large. On the other hand, the COSMO-RS and, in particular, the SM6 model perform significantly better with MADs of 4.3 and 2.4, respectively.

As mentioned earlier, gas-phase free energies calculated using high-level composite methods have an intrinsic error of about 5 kJ/mol. On the other hand, a recent assessment study conducted by Houk et al. also showed that the MADs in the CPCM free energy of solvation are ∼5 kJ/mol and ∼16 kJ/mol for neutral and anionic species, respectively.[52] Assuming that these errors are additive, a crude estimate of the error in ΔG_{soln} for the overall reaction, as calculated via the direct method, would be about 25 kJ/mol energy or approximately 5 pK_a units. On this basis, we note that the MADs in Table 4 are in the right range.

It is interesting to note that COSMO-RS performed quite poorly for carbon acids **N7** to **N11** (AD > 6) but otherwise performs rather well, with calculated values within ∼2 units of experiment. The conjugate bases of these acids are somewhat different due to the absence of an adjacent carbonyl group resulting in a highly localized anionic charge at the α carbon. Very recently, Eckert and co-workers have attempted to use the COSMO-RS model for calculating the pK_a of carbon acids in acetonitrile.[97] The authors noted that the COSMO-RS model performed better on acids that produced anions with delocalized charges as they are less affected by solvation. Conversely, the short-range interactions in the solvation of anions with localized charges are not fully accounted for in the COSMO-RS model.[97] In this light, we examined the atomic charge on the anionic carbon, which has a formal charge −1, using multipole derived charge analysis[98] (as implemented in ADF) to assess the degree of charge delocalization in these species. These charges are tabulated in Table 6. As shown, the atomic charges on the anionic carbon in the conjugate bases of **N7** to **N11** are negative, whereas the corresponding charges in the other acids are ∼+0.3, indicating a high degree of charge delocalization in the latter and further supports the argument by Eckert and co-workers.

In the proton exchange method, marked improvement in the MADs was observed for all of the solvation models, except COSMO-RS and SM6, where the MAD is similar in both approaches. The performance of the CPCM-UAKS model is most noteworthy, with an MAD and AD_{max} of about 1 and 2.5 pK_a units, respectively. Moreover, the neutral carbon acids studied here included both cyclic and acyclic systems with a range of functionalities (amides, amines, and ketones), and the performance of the CPCM-UAKS model was relatively insensitive to these structural variations. Thus, when used in conjunction with a proton exchange scheme, this model should provide a useful strategy for accurate pK_a calculation of a wide range of neutral carbon acids in biological systems as well as in chemical synthesis.

The success of the proton exchange scheme relies on several factors. First, the accuracy of the experimental pK_a of the reference acid is critical since any errors in the experimental data would propagate into the calculations. Acetamide was chosen as the reference acid for the present study as it is the smallest neutral carbon acid and has a relatively small experimental uncertainty of ±0.5 pK_a units. Second, the errors in the solvation model must be systematic, i.e. it needs to consistently over- or underestimate the experimental values so as to allow for optimal error cancellation. In practice, this would involve choosing a reference acid that is structurally similar to the carbon acid of interest. For example, one would not choose a cationic acid as a reference for a neutral carbon acid since the magnitude and sign of the errors incurred by the continuum solvation model is likely to be very different for the two species. As shown in Figure 3, there is a very strong correlation ($r^2 = 0.98$) between the experimental pK_a and those calculated using the direct approach (CPCM-UAKS). The unsigned errors in Table 4 also indicate that this approach systematically overestimates the pK_as of the neutral carbon acids. As such, a marked improvement is obtained using the proton exchange scheme. On the other hand, using the COSMO-RS and SM6 models, the errors in the direct method are less systematic, and, depending on the choice of reference, the use of a proton exchange reaction reduces the errors in only some of the species of the test set.

Cationic Carbon Acids. Shown in Table 5 are the calculated acid dissociation constants for the cationic acids using the direct and proton exchange methods, respectively. In contrast to the neutral acids, inspection of the MAD values reveal that the direct method performs better for cationic systems, with MADs generally under 3 pK_a units across the various solvation models. Unfortunately, the method still fails in several cases, such as in carbon acid **C2**, which accounts for the AD_{max} of >5 pK_a units in the CPCM models. Closer inspection of the data reveals that the smaller MAD observed in these systems is partly due to the good agreement between the calculated and experimental values for carbon acids **C5**, **C6**, and **C7**, where the agreement is generally within 1 pK_a unit. It is also interesting to note that the good performance of the SM6 model on neutral systems is not reflected in the cationic carbon acids.

The COSMO-RS model, which performed best for these species, has an MAD and AD_{max} of 2.5 and 6.2, respectively.

Table 7. Calculated[a] and Experimental Aqueous Acid Dissociation Constants at 298 K for Anionic Carbon Acids[b]

carbon acid	direct method				proton exchange method				
	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	CPCM/UAKS	CPCM/UAHF[c]	COSMO-RS[d]	SM6	expt
A1	43.6 (9.6)	42.5 (8.5)	35.4 (1.4)	26.1 (−7.9)	35.0 (1.0)	32.3 (−1.7)	38.8 (4.8)	31.7 (−2.3)	∼34[20]
A2	26.4 (9.4)	23.5 (6.5)	11.8 (−5.2)	12.9 (−4.1)	17.8 (0.8)	13.3–(−3.7)	15.1 (−1.9)	18.6–(1.6)	17[102]
A3	38.9 (8.6)	40.5 (10.2)	26.9 (−3.4)	24.6 (−5.7)	reference	reference	reference	reference	30.3[36]
A4	42.0 (11.2)	39.8 (9.0)	30.7 (−0.1)	30.5 (−0.3)	33.4 (2.6)	29.6–(−1.2)	34.1 (3.3)	36.1 (5.3)	30.8[36]
A5	44.9 (11.4)	42.3 (8.8)	36.7 (3.2)	20.6 (−12.9)	36.3 (2.8)	32.1 (−1.4)	40.1–(6.6)	26.3 (−7.2)	33.5[33]
AD_{max}	11.4	10.2	5.2	12.9	2.8	3.7	6.6	7.2	–
MAD	10.1	8.6	2.7	6.2	1.8	2.0	4.2	4.1	–

[a] Signed errors are shown in brackets. [b] All associated gas-phase calculations performed using the G3MP2+//BMK/6-31+G(d) level of theory. All solvation energy calculations performed at the B3LYP/6-31+G(d) level of theory unless noted otherwise. [c] Solvation energy calculations performed at the HF/6−31+G(d) level of theory. [d] Solvation energy calculations performed at the BP/TZP level of theory.

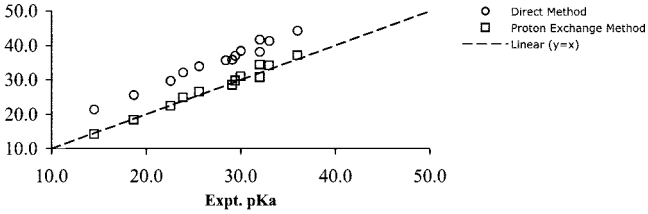


Figure 3. The correlation between experimental and calculated (direct and proton exchange methods using CPCM/UAKS model) aqueous acidities of neutral carbon acids at 298 K. Least-squares correlation for (a) direct method: $pK_a(Calc)=1.06pK_a(Expt) + 6.03$; $r^2= 0.98$ and (b) proton exchange method: $pK_a(Calc)=1.06pK_a(Expt) - 1.19$; $r^2= 0.98$.

Furthermore, we observed that the method performed rather well for most of the carbon acids with the exception of **C1** and **C4**, which might be related to charge distribution in these systems. The atomic charges for the nitrogen with a formal +1 charge are tabulated in Table 6 where it is seen that the cationic charge is strongly localized in **C1** and **C4**; all the other systems exhibit negative charges on the nitrogen adjacent to the α carbon. The poorer performance of the COSMO-RS in these two carbon acids is thus consistent with our earlier observations in the neutral systems. Accordingly, it appears that neutral and cationic carbon acids are amenable to moderately accurate pK_a calculations (∼2 pK_a unit of experiment) using the COSMO-RS solvation model provided the charges on these systems are 'sufficiently delocalized'. Admittedly, a consistent and direct measurement of charge delocalization is required for the appropriate application of the COSMO-RS model. In this work, we have used the *sign* of the atomic charge for this purpose; however, more extensive studies need to be carried out to establish the general applicability of this approach.

It should also be brought to the readers' attention that the discrimination between localized and delocalized systems is not observed in the other solvation models. This may be attributed to the different approaches adopted by the various models to account for explicit intermolecular interactions. For example, the CPCM-UAHF model indirectly accounts for these effects by utilizing cavity radii optimized at the HF/6−31G(d) level of theory,[99] whereas the SM6 model uses different parameters such as atomic surface tensions and a different set of atomic radii that are fitted against a much larger data set of experimental solvation free energies.[91] In this light, we have chosen glycine methyl ester (**C1**) as the reference acid for the CPCM and SM6 models and betaine methyl ester (**C2**) for the COSMO-RS model in the proton exchange approach.

As mentioned earlier, the success of the proton exchange scheme depends on the nature of the errors incurred in the solvation models. Inspection of the signed errors in direct approach suggest that these errors are specific to the functionality of the carbon acid; the CPCM model overestimates the acidity constants for **C1** to **C4** (amino acids) and underestimates the values for **C8** to **C10** (pyriminidiums). This observation is in contrast to the neutral systems. Accordingly, the proton exchange method has resulted in an increase in the MAD of the CPCM solvation models. Specifically, the choice of glycine methyl ester, **C1**, as the reference acid resulted in reasonably accurate results for carbon acids **C2**, **C3**, and **C4** but led to deviations that are much larger than those encountered using the direct method for the remaining carbon acids **C5** to **C10**.

On the other hand, the SM6 model appears to perform better in the proton exchange scheme. In the direct method, the SM6 method consistently overestimates the experimental pK_as, and the proton exchange reaction ameliorates some of this error, reducing the MAD from 4.7 to ∼2 pK_a unit. The performance of the COSMO-RS model in the exchange scheme is less satisfactory because of its sensitivity to the charge distribution in these acids. On the basis of these results, the combination of the SM6 model and a proton exchange scheme is most likely to give the best results.

Anionic Carbon Acids. The data for the anionic acids are summarized in Table 7. The performance of the direct method is clearly unsatisfactory for these systems. In particular, the CPCM model consistently overestimates the pK_a by ∼10, suggesting that the solvation free energies of divalent anions are significantly underestimated. Interestingly, despite the high charge density of the divalent anions, the COSMO-RS model faired reasonably well, with MAD of ∼3, although its performance was less consistent, as was the SM6 model.

In the proton exchange scheme, **A3** was used as the reference acid. Not surprisingly, there is an enormous improvement in the CPCM values, where the MADs were reduced to 2 or less. This result is encouraging in view of the much larger errors incurred by continuum models for multi-valent ions. For reasons explained earlier, the proton exchange scheme was less effective when used in combination with the COSMO-RS and SM6 models.

Table 8. Effect of Adding an Explicit Water Molecule[a] on Accuracy of Calculated Acid Dissociation Constants of Anionic Carbon Acids[b]

carbon acid	$n = 0$	$n = 1$	expt
A1	26.1 (7.9)	33.3 (0.7)	34[20]
A2	12.9 (4.1)	16.5 (0.5)	17[102]
A3	24.6 (5.7)	28.6 (1.7)	30.3[36]
A4	30.5 (0.3)	33.9 (3.1)	30.8[36]
A5	20.6 (12.9)	28.6 (4.9)	33.5[33,79]
AD_{max}	12.9	4.9	–
MAD	6.2	2.2	–

[a] n = number of water molecules. [b] All associated gas-phase calculations performed using the G3MP2+//BMK/6−31+G(d) level of theory. pK_a values calculated via the direct method. All solvation energy calculations performed using SM6 at the B3LYP/6−31+G(d) level of theory.

Given these problems, we examined whether the results could be improved through the inclusion of an explicit water molecule as shown in Scheme 2. Previously, Kelly and coworkers have reported some success using this approach for the divalent carbonate anion.[48] The configuration of the aqua-complexes has been chosen such that the water molecule is hydrogen bonded to the carbonyl oxygen and/or the α carbon where the anionic charge is likely to reside based on resonance structures. Where several configurations are possible, the lowest energy gas-phase conformer was used.

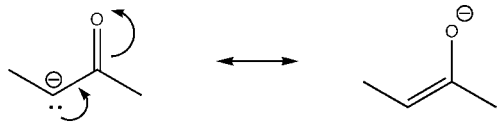


Table 8 shows the results obtained for the anionic carbon acids using the combined SM6 implicit-explicit solvent approach (Scheme 2). As shown, adding a single water molecule to the anion significantly improves the calculated values, where the MAD is reduced to 2.2 as compared to 6.2 for the direct method. Despite the promise of this approach, there are other issues relating to the number of water molecules to add and the conformational sampling problems associated with larger ion-clusters. In particular, we note that this approach led to a larger deviation for carbon acids **A4** and **A5**, and the agreement is likely to worsen with the addition of more water molecules.

6. Conclusions

There are a wide variety of carbon acids in biological systems, including amino acids, peptides, esters, and ketones, which exist in neutral, charged, and zwitterionic forms. As such, the computation of accurate pK_a values of these acids poses a serious challenge for continuum models since their solvation patterns are very different. We find that gas-phase acidities can be accurately obtained using G3MP2+//BMK/6−31+G(d); however, solvation energies are subject to much larger errors. In particular, the direct approach (Scheme 1) yields unacceptably large errors for all three categories of carbon acids in this study.

Nevertheless, the pK_a values of *neutral* carbon acids can be accurately obtained to within ∼1 unit of experiment via the combination of a proton exchange scheme with the CPCM-UAKS solvation model. Furthermore, the accuracy of this approach is not sensitive to the structure of the (neutral) reference acid and should therefore be useful for the pK_a calculation of a wide range of neutral carbon acids. Alternatively, moderately accurate results may be obtained through the direct approach using the SM6 and COSMO-RS solvent models.

Ionic carbon acids are more problematic, where the success of the proton exchange scheme is highly sensitive to the choice of reference acid. This limits the applicability of this approach for studying charged carbon acids. For cationic systems, the SM6 model combined with a proton exchange scheme delivered the best results. For anionic acids, the combination of the proton exchange scheme with the CPCM-UAKS model also gave reasonably good results, and the addition of an explicit water molecule using the SM6 model significantly improved the computed pK_as for anionic acids (∼3 fold reduction in MAD) compared to using the direct method.

Admittedly, the pK_a calculation strategies that have emerged from this work are somewhat *ad hoc*, as they do not directly address the problems associated with the solvation of ionic species. Nevertheless, it is also intended that this work helps to identify limitations of present continuum solvation models and to spur further research aimed at improving the presented results. In this regard we note that the COSMO-RS model provided the best overall performance in the direct method pK_a calculations in all three classes of carbon acids, although there are problems associated with ionic species having highly localized charges. As noted before, the COSMO-RS model is a more sophisticated variant of the CPCM model and takes the real character of the solvent (rather than a simple continuum) into account. Thus, its success could possibly indicate the importance of explicit consideration of real character of the solvent in the future development of solvation models beyond the continuum approximation.

Acknowledgment. We gratefully acknowledge support from the Australian Research Council under their Centres of Excellence program and generous allocations of computing time on the National Facility of the Australian Partnership for Advanced Computing.

Supporting Information Available: Complete BMK/6−31+G(d) optimized gas-phase geometries of all species and their associated conformers (Table S1), corresponding electronic energies and solvation energies (Table S2), and COSMO-RS solvation energies computed using ADF and corresponding values from ref 90 is also available (Table S3). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Man, E. H.; Sandhouse, M. E.; Burg, J.; Fisher, G. H. *Science* **1983**, *220*, 1407.

(2) Helfman, P. M.; Bada, J. L. *Nature* **1976**, *262*, 279.

(3) Helfman, P. M.; Bada, J. L. *Proc. Natl. Acad. Sci., U.S.A.* **1975**, *72*, 297.

(4) Bada, J. L.; Kvenvolden, K. A.; Peterson, E. *Nature* **1973**, *245*, 308.

(5) Masters, P. M.; Bada, J. L.; Zigler, J. S. *Nature* **1977**, *262*, 71.

(6) Liardon, R.; Friedman, M. *J. Agric. Food Chem.* **1987**, *35*, 661.

(7) Liardon, R.; Ledermann, S. *J. Agric. Food Chem.* **1986**, *34*, 557.

(8) Auvynet, C.; Seddiki, N.; Dunia, I.; Nicolas, P.; Amiche, M.; Lacombe, C. *Eur. J. Cell Biol.* **2006**, *85*, 25.

(9) Jilek, A.; Mollay, C.; Tippelt, C.; Grassi, J.; Mignogna, G.; Muellegger, J.; Sander, V.; Fehrer, C.; Barra, D.; Kreil, G. *Proc. Natl. Acad. Sci., U.S.A.* **2005**, *102*, 4235.

(10) Kreil, G. *Annu. Rev. Biochem.* **1997**, *66*, 337.

(11) Kreil, G. *Science* **1994**, *266*, 996.

(12) Montecucchi, P. C.; de Castiglione, R.; Piani, S.; Gozzini, L.; Erspamer, V. *Int. J. Pept. Protein Res.* **1981**, *17*, 275.

(13) Torres, A. M.; Menz, I.; Alewood, P. F.; Bansal, P.; Lahnstein, J.; Gallagher, C. H.; Kuchel, P. W. *FEBS Lett.* **2002**, *524*, 172.

(14) Torres, A. M.; Tsampazi, C.; Geraghty, D. P.; Bansal, P. S.; Alewood, P. F.; Kuchel, P. W. *Biochem. J.* **2005**, *391*, 215.

(15) Torres, A. M.; Tsampazi, M.; Tsampazi, C.; Kennett, E. C.; Belov, K.; Geraghty, D. P.; Bansal, P. S.; Alewood, P. F.; Kuchel, P. W. *FEBS Lett.* **2006**, *580*, 1587.

(16) Tanner, M. E. *Acc. Chem. Res.* **2002**, *35*, 237.

(17) Richard, J. P. *J. Am. Chem. Soc.* **1984**, *106*, 4926.

(18) Richard, J. P. *Biochemistry* **1998**, *37*, 4305.

(19) Richard, J. P.; Amyes, T. L. *Bioorg. Chem.* **2004**, *32*, 354.

(20) Richard, J. P.; Amyes, T. L. *Curr. Opin. Chem. Biol.* **2001**, *5*, 626.

(21) Richard, J. P.; Amyes, T. L.; Toteva, M. M. *Acc. Chem. Res.* **2001**, *34*, 981.

(22) Rios, A.; Crugeiras, J.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2001**, *123*, 7949.

(23) Sievers, A.; Wolfenden, R. *J. Am. Chem. Soc.* **2002**, *124*, 13986.

(24) Bull, S. D.; Davies, S. G.; Garner, A. C.; Parkes, A. L.; Roberts, P. M.; Sellers, T. G. R.; Smith, A. D.; Tamayo, J. A.; Thomson, J. E.; Vickers, R. J. *New J. Chem.* **2007**, *31*, 486.

(25) Davies, S. G.; Garner, C. A.; Ouzman, J. V. A.; Roberts, P. M.; Smith, A. D.; Snow, E. J.; Thomson, J. E.; Tamayo, J. A.; Vickers, R. J. *Org. Biomol. Chem.* **2007**, *5*, 2138.

(26) Davies, S. G.; Rodriguez-Solla, H.; Tamayo, J. A.; Garner, A. C.; Smith, A. D. *Chem. Commun.* **2004**, 2502.

(27) Matthews, W. S.; Bares, J. E.; Bartmess, J. E.; Bordwell, F. G.; Cornforth, F. J.; Drucker, G. E.; Margolin, Z.; McCallum, R. J.; McCollum, G. J.; Vanier, N. R. *J. Am. Chem. Soc.* **1975**, *97*, 7006.

(28) Taft, R. W.; Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 463.

(29) Amyes, T. L.; Diver, S. T.; Richard, J. P.; Rivas, F. M.; Toth, K. *J. Am. Chem. Soc.* **2004**, *126*, 4366.

(30) Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **1996**, *118*, 3129.

(31) Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **1992**, *114*, 10297.

(32) Richard, J. P.; Williams, G.; Gao, J. *J. Am. Chem. Soc.* **1999**, *121*, 715.

(33) Richard, J. P.; Williams, G.; O'Donoghue, A. C.; Amyes, T. L. *J. Am. Chem. Soc.* **2002**, *124*, 2957.

(34) Rios, A.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2000**, *122*, 9373.

(35) Rios, A.; Richard, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 8375.

(36) Rios, A.; Richard, J. P.; Amyes, T. L. *J. Am. Chem. Soc.* **2002**, *124*, 8251.

(37) Crugeiras, J.; Rios, A.; Riveiros, E.; Amyes, T. L.; Richard, J. P. *J. Am. Chem. Soc.* **2008**, *130*, 2041.

(38) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.

(39) Liptak, M. D.; Shields, G. C. *Int. J. Quantum Chem.* **2001**, *85*, 727.

(40) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114*, 4595.

(41) Namazian, M.; Kalantary-Fotooh, F.; Noorbala, M. R.; Searles, D. J.; Coote, M. L. *THEOCHEM* **2006**, *758*, 275.

(42) Namazian, M.; Heidary, H. *THEOCHEM* **2003**, *620*, 257.

(43) Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717.

(44) Kallies, B.; Mitzner, R. *J. Phys. Chem. B* **1997**, *101*, 2959.

(45) Lopez, X.; Schaefer, M.; Dejaegere, A.; Karplus, M. *J. Am. Chem. Soc.* **2002**, *124*, 5010.

(46) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.

(47) Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **2002**, *106*, 1327.

(48) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 2493.

(49) Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. *J. Phys. Chem. A* **2004**, *108*, 166.

(50) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.

(51) Fu, Y.; Liu, L.; Li, R.-Q.; Liu, R.; Guo, Q.-X. *J. Am. Chem. Soc.* **2004**, *126*, 814.

(52) Takano, Y.; Houk, K. N. *J. Chem. Theory Comput.* **2005**, *1*, 70.

(53) Brinck, T.; Larsen, A. G.; Madesen, K. M.; Daasbjerg, K. *J. Phys. Chem. B* **2000**, *104*, 9887.

(54) Gao, D.; Wong, P. K.; Maddalena, D.; Hwang, J.; Walker, H. *J. Phys. Chem. A* **2005**, *109*, 10776.

(55) Smith, B. J.; Radom, L. *J. Phys. Chem.* **1991**, *95*, 10549.

(56) Smith, B. J.; Radom, L. *J. Am. Chem. Soc.* **1993**, *115*, 4885.

(57) Smith, B. J.; Radom, L. *Chem. Phys. Lett.* **1994**, *231*, 345.

(58) Martin, J. M.; Lee, T. J. *Chem. Phys. Lett.* **1996**, *258*, 136.

(59) Ochterski, J. W.; Petersson, G. A.; Wiberg, K. B. *J. Am. Chem. Soc.* **1995**, *117*, 11299.

(60) Wiberg, K. B. *J. Org. Chem.* **2002**, *67*, 4787.

(61) Peterson, K. A.; Xantheas, S. S.; Dixon, D. A.; Dunning, T. H. J. *J. Phys. Chem. A* **1998**, *102*, 2449.

(62) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764.

(63) Remko, M. *J. Phys. Chem. A* **2002**, *106*, 5005.

(64) Pokin, E. K.; Liptak, M. D.; Feldgus, S.; Shields, G. C. *J. Phys. Chem. A* **2001**, *105*, 10483.

(65) Seo, Y.; Kim, Y.; Kim, Y. *Chem. Phys. Lett.* **2001**, *340*, 186.

(66) Burk, P.; Koppel, I. A.; Koppel, I.; Leito, I.; Travnikova, O. *Chem. Phys. Lett.* **2000**, *323*, 482.

(67) Ervin, K. M.; DeTuri, V. F. *J. Phys. Chem. A* **2002**, *106*, 9947.

(68) Hammerum, S. *Chem. Phys. Lett.* **1999**, *300*, 529.

(69) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161.

(70) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.

(71) Chipman, D. M. *J. Chem. Phys.* **2003**, *118*, 9937.

(72) Namazian, M.; Halvani, S. *J. Chem. Thermodyn.* **2006**, *38*, 1495.

(73) Namazian, M.; Halvani, S.; Noorbala, M. R. *THEOCHEM* **2004**, *711*, 13.

(74) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.

(75) Williams, G.; Maziarz, E. P.; Amyes, T. L.; Wood, T. D.; Richard, J. P. *Biochemistry* **2003**, *42*, 8354.

(76) Wong, F. M.; Capule, C.; Chen, D. X.; Gronert, S.; Wu, W. *Org. Lett.* **2008**, *2008*, 2757.

(77) Wong, F. M.; Capule, C.; Wu, W. *Org. Lett.* **2006**, *8*, 6019.

(78) Reutov, O. A.; Beletskaya, I. P.; Butin, K. P. A guide to all existing problems of CH-acidity with new experimental methods and data, including indirect electrochemical, kinetic and thermodynamic studies; CH acids: A guide; Crompton, T. R., Ed.; Pergamon Press Ltd.: New York, 1978.

(79) Chiang, Y.; Kresge, A. J.; Popik, V. V.; Schepp, N. P. *J. Am. Chem. Soc.* **1997**, *119*, 10203.

(80) Frisch, M. J. T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.

(81) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.

(82) Parr, R. G.; Yang, W. *Density Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989.

(83) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703.

(84) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.

(85) Merrick, J. P.; Moran, D.; Radom, L. *J. Phys. Chem. A* **2007**, *111*, 11683.

(86) Klamt, A.; Schueuermann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(87) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24*, 669.

(88) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224.

(89) Klamt, A. *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier Science Ltd.: Amsterdam, The Netherlands, 2005.

(90) Klamt, A.; Jonas, V.; Burger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074.

(91) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 1133.

(92) (a) te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Baerends, E. J.; Snijders, J. G.; Zigler, T. Chemistry with ADF. *J. Comput. Chem.* **2001**, *22*, 931. (b) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391. (c) Pye, C.; Louwen, J. N.; van Lenthe, E. Manuscript in preparation. (d) ADF 2008.01, COSMO-RS, SCM. Available via the Internet at www.scm.com, accessed October 2008.

(93) Higtashi, M.; Marenich, A. V.; Olson, R. M.; Chamberlin, A. C.; Pu, J.; Kelly, C. P.; Thompson, J. D.; Xidos, J. D.; Li, J.; Zhu, T.; Hawkins, G. D.; Chuang, Y.-Y.; Fast, P. L.; Lynch, B. J.; Liotard, D. A.; Rinaldi, D.; Gao, J.; Cramer, C. J.; Truhlar, D. G. GAMESSPLUS - version 2008-2; University of Minnesota, MN, 2008, based on the General Atomic and Molecular Electronic Structure System (GAMESS) as described in Schmidt, M. W.; Baidridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(94) Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2006**, *110*, 16066.

(95) Hare, M. C.; Marimanikkuppam, S. S.; Kass, S. R. *Int. J. Mass Spectrom.* **2001**, *210/211*, 153.

(96) Martin, J. M.; Parthiban, S. In *Quantum Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; Kluwer-Academic: Dordrecht, The Netherlands, 2001; p 31.

(97) Eckert, F.; Leito, I.; Kaljurand, I.; Kütt, A.; Klamt, A.; Diedenhofen, M. *J. Comput. Chem.* **2008**, . in press.

(98) Swart, M.; Duijnen, P. T. v.; Snijders, J. G. *J. Comput. Chem.* **2001**, *22*, 79.

(99) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.

(100) Villano, S. M.; Gianola, A. J.; Eyet, N.; Ichino, T.; kato, S.; Bierbaum, V. M.; Lineberger, W. C. *J. Phys. Chem. A* **2007**, *111*, 8579.

(101) Kurinovich, M. A.; Lee, J. K. *J. Am. Soc. Mass. Spectrom* **2002**, *13*, 985.

(102) Chiang, Y.; Kresge, A. J.; Pruszynski, P. *J. Am. Chem. Soc.* **1992**, *114*, 3103.

CT800335V