First Principles Calculations of Aqueous pK_a Values for Organic and Inorganic Acids Using COSMO−RS Reveal an Inconsistency in the Slope of the pK_a Scale

Andreas Klamt,* Frank Eckert, and Michael Diedenhofen

COSMOlogic GmbH&COKG, Burscheider Str. 515, 51381 Leverkusen, Germany

Michael E. Beck

Bayer Cropscience, BCS-R-SCLA, Alfred-Nobel-Str. 50, 40789 Monheim, Germany

Received: March 18, 2003; In Final Form: July 28, 2003

The COSMO−RS method, a combination of the quantum chemical dielectric continuum solvation model COSMO with a statistical thermodynamics treatment for more realistic solvation (RS) simulations, has been used for the direct prediction of pK_a constants of a large variety of 64 organic and inorganic acids. A highly significant correlation of $r^2 = 0.984$ with a standard deviation of only 0.49 between the calculated values of the free energies of dissociation and the experimental pK_a values was found, without any special adjustment of the method. Thus, we have a theoretical a priori prediction method for pK_a, which has the regression constant and the slope as only adjusted parameters. Such a method can be of great value in many areas of physical chemistry, especially in pharmaceutical and agrochemical industry. To our surprise, the slope of pK_a vs ΔG_{diss} is only 58% of the theoretically expected value of 1/RTln(10). A careful analysis with respect to different contributions as well as a comparison with the work of other authors excludes the possibility that the discrepancy is due to weaknesses of the calculation method. Hence, we must conclude that the experimental pK_a scale depends differently on the free energy of dissociation than generally assumed.

Introduction

Proton transfer and the associated dissociation constants of acids and bases in water and in other solvents are of great importance in many areas of organic, inorganic, and biological chemistry.[1] Therefore, the first principle prediction of pK_a values has been the subject of many research projects and scientific papers. In the gas phase, dissociation constants are experimentally difficult to measure and good quantum chemical calculations, i.e., post Hartree−Fock ab initio calculations or state-of-the-art density functional methods, can nowadays be as good or even more accurate than experimental measurements.[2−12] The situation is very different for liquid-phase pK_a, because here a lot of inexpensive standard experimental methods exist and the experimental error for pK_a measurements is usually less than 0.1 pK_a units in the range between $0 \leq \mathrm{p}K_a \leq 14$. Outside this range, empirical correction are required for pK_a measurements, and thus the error may be larger. On the other hand the theoretical calculation of liquid-phase pK_a is much more complicated than that in the gas phase because of the strong solvation effects, which require the application of empirical solvation models on top of the quantum chemical calculations. Considering typical values of −8 kcal/mol for the solvation energies of neutral polar molecules and of even up to −120 kcal/mol for ions, such solvation models have to provide high accuracy for the short range and long-range dipole and monopole electrostatics, and they should also take into account explicit solvation effects such as hydrogen bonding. For a long time, this situation for theoretical calculation of aqueous pK_a appeared to be almost intractable. Therefore, lots of empirical models have been developed, using all kinds of fragment based methods, empirical quantum chemical descriptors, artificial intelligence, and combinations thereof. Because it is impossible to give a complete overview of such models, we only refer to a few examples.[13−18] Such empirical models are well able to reproduce and predict pK_a values of organic compounds if similar acidic sites are well represented in the training set of the method, but they often fail for rare acidic sites with complicated electronic effects, especially if electron delocalization is important in the resulting anions. Such nonlocal effects are hard to describe by empirical models.

Despite of all their theoretical insufficiencies,[19−21] during the past decade dielectric continuum solvation models[22,23] (DCSMs) got more and more often used for more rigorous liquid-phase pK_a calculations based on quantum chemical calculations.[24−77] The reason for this is their high electrostatic accuracy for all kinds of long-range electrostatics in combination with their computational efficiency in the context of quantum chemical programs. However, the applicability of DCSMs for the short-range electrostatics of polar solutes and ions is at least questionable, and hydrogen bonding cannot be described by dielectric theory at all. Nevertheless, moderate success has been reported with this approach by various authors, starting from the simple relation

$$\mathrm{p}K_a = A\frac{\Delta G_{\mathrm{diss}}}{RT\ln(10)} + B \qquad (1)$$

where $A = 1$ is expected from standard electrolyte theory, if ΔG_{diss} would be calculated without any error. Looking in detail into these studies, it appears that either reasonable correlations between the calculated value of ΔG_{diss} and the experimental

---

* Corresponding Author. Tel.: +49−2171−731681. E-mail: klamt@cosmologic.de.

10.1021/jp034688o CCC: \$25.00 © 2003 American Chemical Society
Published on Web 10/14/2003

pK_a is reported for a small class of compounds, or many additional parameters have been specially adjusted to achieve good correlations over a wider range of compounds. The adjustable parameters mainly are atom-type specific cavity radii or special adjustments of other cavity parameters, and often even charge dependent cavity definitions are considered. Almost all studies report a significantly lower slope in the regression of pK_a versus the calculated dissociation free energy ΔG_{diss} than the theoretically expected value of 1/RTln(10), unless large efforts are made to achieve the expected slope by often physically meaningless parameter adjustment. This fact has recently been pointed out by Chipman,[75] who also found a low slope for aqueous pK_a in a study on nine compounds, while he achieved closer agreement with the theoretically expected slope for pK_a in DMSO and acetonitrile.

During the past 8 years we have developed a unique kind of solvation model, the Conductor-like Screening Model for Real Solvents[19−21,78,79] (COSMO−RS), which combines the electrostatic advantages and the computational efficiency of the DCSM COSMO[80] with a statistical thermodynamics method for local interaction of surfaces, which takes into account local deviations from dielectric behavior as well as hydrogen bonding. In this approach, all information about solutes and solvents is extracted from initial QM/COSMO calculations, and only very few parameters have been adjusted to experimental values of partition coefficients and vapor pressures of a wide range of neutral organic compounds. COSMO−RS is capable of predicting partition coefficients, vapor pressures, solvation free energies of neutral compounds with an error of 0.3 log-units (RMS), and better and a great deal of experience have been gathered during the past years about its surprising ability to predict mixture thermodynamics.[78,79] Although we also gained some experience about its performance for charged species, a quantification of the accuracy of COSMO−RS for ions is missing, because reliable and well defined thermodynamic data for ions is much less available than for neutral compounds. The reported values for ΔG_{hydr} of ions are especially subject to some skepticism because they often are determined indirectly.

Stimulated by the initial COSMO−RS pK_a predictions in the agrochemical industry and motivated by the large number of accurately measured pK_a values for organic compounds commonly available in the literature and databases, we decided to perform a systematic study on the ability of COSMO−RS to predict pK_a values. For that purpose, we calculated ΔG_{hydr} for a broad selection of about 65 organic and inorganic acids, spanning a pK_a range between 0 and 16 and using our standard COSMO−RS method implemented in the COSMOtherm program[82] based on TURBOMOLE DFT/COSMO calculations.[83−85] Without any special adjustments of radii or other parameters, this led to a very good prediction model for pK_a but also to a surprising conclusion regarding the slope of the experimental pK_a versus theoretical ΔG_{hydr}.

Theoretical Calculations

Our theoretical calculations for ΔG_{diss} in water are based on the reaction model

$$\mathrm{AH} + \mathrm{H_2O} \rightarrow \mathrm{A^-} + \mathrm{H_3O^+} \tag{2}$$

Obviously, the exact species of the hydronium is not clearly defined, but approximating it by H_3O^+ should be a reasonable assumption. Because the deviations arising from this approximation should be the same for all kinds of acid dissociation reactions, they should not effect the quality of a regression of pK_a versus ΔG_{hydr}. Hence, we did not put any further effort into this point.

Because we are not interested in the gas phase reaction, we directly calculated the self-consistent state of each species in aqueous solutions. For that, we first applied our standard procedure for COSMO−RS calculations to all 4 species appearing in eq 1, which consists of the following steps:

(1) Full DFT geometry optimization with the Turbomole program package[83,85] using B−P density functional[86,87] with TZVP quality basis set and applying the RI approximation.[88] During these calculations, the COSMO continuum solvation model was applied in the conductor limit ($\epsilon = \infty$). Element-specific default radii from the COSMO−RS parametrizations have been used for the COSMO cavity construction.[78] Such calculations end up with the self-consistent state of the solute in the presence of a virtual conductor, which surrounds the solute outside the cavity. The total energy of the solute in this state is referred to as COSMO energy E_{COSMO}, below. Following the usual COSMO notation, the total energy gain due to the interaction of the solute with the continuum (i.e., half of the interaction energy of the solute with the continuum) is called the dielectric energy E_{diel}.

(2) COSMO−RS calculations have been done using the COSMOtherm program.[82] In these calculations, the deviations of a real solvent, in our case water or DMSO, compared to an ideal conductor are taken into account in a model of pairwise interacting molecular surfaces. For this purpose, electrostatic energy differences and hydrogen bonding energies are quantified as functions of the local COSMO polarization charge densities σ and σ' of the two interacting surface pieces. The chemical potential differences arising from these interactions are evaluated using an exact statistical thermodynamics algorithm for independently pairwise interacting surfaces, which is implemented in COSMOtherm. More detailed descriptions of the COSMO−RS method are given elsewhere.[19−21,78,79]

If more than one conformation was considered to be potentially relevant for the neutral or anionic form of the acid A, several conformations have been calculated in step 1 and a thermodynamic conformation average is consistently calculated by the COSMOtherm program.

For a subset of nine compounds, we also performed single point DFT/COSMO calculations on a larger basis set (aug-cc-pVTZ) using the same BP functional and with a hybrid functional (B3LYP[89]) using the TZVP basis set but without RI approximation. These calculations are used for a sensitivity analysis of the results with respect to the basis set and DFT method. Finally, we also performed single-point HF/COSMO and MP2/COSMO calculations[90] with a TZVPP basis set to prove independence of the results from systematic DFT errors.

For all 65 acids, we calculated the Gibbs free energy of dissociation as the difference of the total free energy of the anion A^- and the protonated acid AH and added the free energy difference of H_3O^+ and H_2O as a constant contribution

$$\Delta G_{diss} = G_{tot}(A^-) - G_{tot}(AH) + [G_{tot}(H_3O^+) - G_{tot}(H_2O)] \tag{3}$$

Zero point vibrational energies are not explicitly taken into account. Here, we make the common assumption that the difference in zero point energy between the neutral acid and the anion is generally small, which has been confirmed by several workers in this area.[48,62,77]

For a subset of seven acids, the equivalent of eq 2 is evaluated for the solvent DMSO instead of water. For that, only the

COSMO−RS calculation needed to be repeated for DMSO instead of water, while the underlying DFT/COSMO calculations are independent of the real solvent.

Data Set. The data set consists of 5 alcohols, 23 carboxylic acids, 8 inorganic acids, 16 phenols, and 12 heterocyclic compounds. Only acids with acidic hydrogens at oxygen or nitrogen are considered in this study. The collection of acids has been built up in several steps, starting from an initial set taken from the work of Schüürmann[39] and later expanded to get a much broader coverage of organic and inorganic compounds and a reasonable coverage of the pK_a range. Some compounds have been added because of reported problems of other methods to see whether our method is able to treat them correctly. No compound has been removed from the data set for any reason during the study. Experimental data are taken from different sources.[14,39,58,62,91,92]

Results

The results for all 64 acids are shown in Table 1 and Figure 1. We clearly see a very good correlation of the experimental aqueous pK_a with the calculated value of ΔG_{diss} with a correlation coefficient of $r^2 = 0.982$ and an rms deviation of 0.5 pK_a-units. Only three acids deviate by slightly more than 1 pK_a-unit from the regression line, all in the same direction. Two of these outliers are succinimide heterocycles, one is 4-nitrophenol. For the latter, we believe we understand the origin of the deviation, because the anion is found to take a chinoidal form, indicated by a strong decrease in the C−O bond length. Therefore, the ionization in this case goes along with much larger changes in the entire molecule than in the other phenols. It is very likely that the level of quantum chemistry, which we are applying here, is insufficient for an accurate quantification of the energy change going along with this reorganization.

We did not take into account the symmetric multiplicity factors, because we did not feel able to do this consistently for all kinds of acids in the same way. Considering this deficiency, which typically accounts for factors of 2 and resulting pK_a variations of 0.3 pK_a, the rms of 0.5 pK_a-units must be considered as very satisfying. Apparently, the purely predictive calculation of ΔG_{diss} by COSMO−RS reflects the physics of the acid dissociation reasonably. The regression equation for aqueous pK_a reads

$$\mathrm{p}K_a = 0.58(\pm 0.01)\frac{\Delta G_{\mathrm{diss}}}{RT\ln(10)} + 1.66(\pm 0.10) \qquad (4)$$

If we would have omitted the free energy difference of H_3O^+ and H_2O in the definition of ΔG_{diss}, which we calculate as 268.3 kcal/mol in good agreement to experimental and calculated literature values,[75] we would have got a regression constant of −115.2. Hence, the achieved deviation of 3.4 to the theoretical value[94] of $B = -\log[H_2O] = -1.74$ is satisfying, considering the uncertainties in the true state of the positive charge.

Much more important is the fact that we find a considerably low slope compared with the theoretical expectation of 1/RT ln(10), as it was found by many other workers before. To better understand the origin of the low slope we split the calculated ΔG_{diss} into four independent contributions:

- the dielectric energy of the anion A^-, E_{diel}^{A-},

- the dielectric energy of the neutral compound AH, E_{diel}^{AH},

- the residual quantum chemical pseudo gas-phase energy difference,

$$\Delta E_{\mathrm{QC}}^{*} = (E_{\mathrm{COSMO}}^{A-} - E_{\mathrm{diel}}^{A-}) - (E_{\mathrm{COSMO}}^{AH} - E_{\mathrm{diel}}^{AH}) \qquad (4)$$

- and the chemical potential differences arising from the COSMO−RS model

$$\Delta G_{\mathrm{CRS}} = \mu_{\mathrm{H2O}}^{A-} - \mu_{\mathrm{H2O}}^{AH} \qquad (5)$$

We performed a multi-linear regression of pK_a with respect to these four contributions of ΔG_{diss}. Our expectation was that the wrong slope is mainly due to one of the contributions, most likely due to the first, and that this would show up by large deviations in the coefficients in front of the contributions. However, surprisingly, we find all four of the coefficients in front of the energy contributions to be essentially identical

$$\mathrm{p}K_a = (0.01/RT\ln(10))[62(\pm 2)E_{\mathrm{diel}}^{A-} - 62(\pm 2)E_{\mathrm{diel}}^{AH} + 59(\pm 1)\Delta E_{\mathrm{QC}}^{*} + 55(\pm 4)\Delta G_{\mathrm{CRS}}] + 113.5(\pm 2.0) \qquad (6)$$

Despite the larger number of adjustable parameters, the correlation coefficient increases only to 0.986 and the standard deviation decrease to 0.47. The four coefficients vary between 55% and 62% of the expected slope, with uncertainties of 4% or less. Hence, we have to conclude, that the deviation between the physically expected slope and the calculated slope cannot be due to an error of just one of the contributions. It is especially surprising that the slope in front of the dielectric energies is identical, although no special adjustments for ions have been introduced. This appears to invalidate the usual assumption, that DCSM calculations for anions require special radii adjustments.

To further validate the strange finding of the low slopes, we did COSMO calculations with a larger basis set, other calculations using a hybrid density functional (B3LYP), HF calculations, and finally MP2 calculations for just nine acids that span the entire pK_a scale. We find a strong correlation between the results of the modified quantum calculations with the BP/TZVP calculations with $r^2 \geq 0.993$ in all cases. The slope increases by 1.5% of the theoretically expected slope for the larger basis set, but it even decreases by 3−4% for the other methods B3LYP, HF, and MP2. Thus, it is unlikely that the quantum chemical accuracy level may account for the large slope deviation to experiment. All results of these calculations are given in Table 2.

To demonstrate the significance of the COSMO−RS contribution, we did a regression analysis based on the bare COSMO energies instead of the COSMO−RS free energies. Here, the regression coefficient decreased to 0.90, and the rms-deviation increases by 150% compared with the COSMO−RS result. Thus, we find that the COSMO−RS contribution, which takes into account deviations from a purely dielectric solvation model, is highly significant. The slope of this regression is only 50% of the theoretical expectation (i.e., even worse than the slope including COSMO−RS).

To analyze the origin of the slope discrepancy we finally considered the pK_a in DMSO for seven of the acids, for which we could find data in the Bordwell table.[95,96] We can easily do that because the COSMO−RS method allows for the treatment of arbitrary solvents without special parametrization, because cavitation energies are already included in the statistical thermodynamics part. The results are given in Table 3 and displayed in Figure 2. We got a good correlation of $r^2 \geq 0.97$ and rms = 1.76. The regression equation is

$$\mathrm{p}K_a^{\mathrm{DMSO}} = 0.70(\pm 0.05)\frac{\Delta G_{\mathrm{diss}}^{\mathrm{DMSO}}}{RT\ln(10)} + 3.44(\pm 1.59) \qquad (7)$$

Hence, we again find a low slope compared with experiment.

TABLE 1: 64 Acids with Exp. and Calculated PK_a Values (Eq 4) and Calculated Free Energy Contributions (in kcal/mol)[a]

acid	pK_a^{exp}	ref.	pK_a^{CRS}	ΔpK_a	G_{H2O}^{AH}	G_{H2O}^{A-}	ΔG_{diss}	ΔE_{QC}^{*}	E_{diel}^{AH}	E_{diel}^{A-}	ΔG_{CRS}
tert-butyl alcohol	18.00	93	17.55	−0.45	−146698.2	−146392.6	37.31	388.44	−6.43	−77.73	−11.58
H2O	15.74	91	16.26	0.52	−47993.3	−47690.8	34.26	408.28	−9.49	−100.6	−14.66
methanol	15.54	91	16.46	0.92	−72657.2	−72354.2	34.74	394.26	−6.92	−85.83	−12.37
ethanol	16.00	91	16.31	0.31	−97338.2	−97035.5	34.38	390.65	−6.77	−82.62	−12.18
2,2,2-trichloroethanol	12.02	91	11.62	−0.40	−962650.8	−962359.2	23.34	357.66	−9.25	−66.19	−9.13
formicacid	3.77	91	3.57	−0.20	−119145.3	−118872.6	4.41	351.01	−9.16	−75.57	−11.95
aceticacid	4.76	91	4.88	0.12	−143830.4	−143554.6	7.50	357.30	−9.30	−77.99	−12.86
chloroaceticacid	2.86	91	3.11	0.25	−432266.9	−431995.3	3.34	341.01	−12.17	−71.16	−10.44
dichloroaceticacid[b]	1.29	91	2.08	0.79	−720701.0	−720431.8	0.91	333.13	−10.91	−66.48	−8.40
trichloroaceticacid	0.65	91	1.14	0.49	−1009131.0	−1008864.1	−1.31	326.86	−8.68	−61.86	−6.73
n-pentanoicacid	4.84	92	5.16	0.32	−217862.7	−217586.3	8.14	357.35	−9.06	−77.31	−12.71
2,2-dimethylpropanoicacid	5.03	92	5.43	0.40	−217861.6	−217584.5	8.79	355.20	−8.37	−74.46	−12.07
benzoicacid	4.20	91	4.68	0.48	−264189.4	−263914.1	7.03	351.96	−9.95	−75.69	−10.95
oxalicacid[b]	1.23	91	1.69	0.46	−237535.3	−237267.1	−0.02	323.55	−16.37	−65.87	−5.82
maleicacid[b]	1.93	91	1.34	−0.59	−286127.4	−285860.0	−0.83	319.74	−18.39	−62.68	−8.03
fumaricacid	3.02	91	3.72	0.70	−286129.9	−285856.9	4.77	344.47	−15.11	−75.83	−10.73
acrylicacid[b]	4.25	91	4.64	0.39	−167733.8	−167458.6	6.92	354.29	−9.22	−76.48	−11.85
phthalicacid[b]	2.98	91	2.42	−0.56	−382578.6	−382308.7	1.71	344.73	−17.77	−78.46	−14.08
bromoaceticacid[b]	2.86	91	3.06	0.20	−1758938.8	−1758667.4	3.22	340.29	−11.06	−70.02	−9.86
chloroaceticacid[b]	2.86	91	3.06	0.20	−432267.2	−431995.7	3.23	341.76	−11.08	−71.17	−10.19
cyanoaceticacid[b]	2.43	91	2.82	0.39	−201734.4	−201463.5	2.66	334.77	−17.56	−71.37	−10.05
fluoroaceticacid[b]	2.66	91	3.14	0.48	−206125.6	−205854.0	3.40	345.01	−11.29	−73.90	−10.74
iodoaceticacid[b]	3.12	91	2.95	−0.17	−150640.0	−150368.8	2.96	339.13	−10.94	−68.69	−10.18
2-hydroxypropanoicacid[b]	3.86	91	3.14	−0.72	−215733.4	−215461.8	3.41	339.85	−12.11	−69.82	−10.48
2-chloropropanoicacid[b]	2.83	92	3.53	0.70	−456946.8	−456674.2	4.31	344.76	−10.78	−72.27	−10.71
3-chloropropanoicacid[b]	3.98	92	4.28	0.30	−456949.2	−456674.9	6.08	344.96	−10.62	−69.91	−11.34
nitroaceticacid[b]	1.68	91	1.85	0.17	−272222.9	−271954.3	0.37	330.44	−15.26	−68.12	−8.95
2-acetylbutanedioicacid[b]	2.86	14	2.44	−0.42	−382726.9	−382456.9	1.77	350.60	−19.15	−83.29	−16.45
carbonicacid[b]	3.58	91	3.47	−0.11	−166391.1	−166118.7	4.18	346.30	−12.66	−77.00	−9.54
hypochlorousacid	7.40	92	8.22	0.82	−336389.3	−336105.7	15.34	367.41	−6.39	−79.42	−10.79
hypobromousacid	8.55	92	9.09	0.54	−1663069.5	−1662783.8	17.40	367.35	−6.50	−77.91	−10.30
hypoiodousacid	10.50	92	9.64	−0.86	−54779.3	−54492.3	18.70	368.48	−7.07	−77.13	−11.47
nitrousacid	3.29	91	2.42	−0.87	−129156.8	−128886.9	1.70	347.36	−6.14	−72.88	−10.67
sulfurousacid	1.90	91	1.23	−0.67	−392331.0	−392063.9	−1.10	333.77	−12.80	−70.48	−8.93
phosphoricacid[b]	2.16	92	2.04	−0.12	−404358.8	−404089.7	0.82	336.72	−21.95	−81.89	−7.71
boric acid	9.23	91	9.51	0.28	−158528.5	−158241.8	18.39	361.68	−16.93	−80.81	−11.17
pentachlorophenol	4.90	39	5.10	0.20	−1635212.8	−1634936.5	8.01	329.18	−5.59	−52.28	−6.23
2,3,4,6-tetrachlorophenol	5.62	39	5.71	0.09	−1346778.5	−1346500.8	9.44	333.29	−5.85	−54.93	−6.53
2,3,4-trichlorophenol[b]	7.1	39	7.32	0.22	−1058341.5	−1058060.0	13.25	340.59	−5.91	−57.90	−7.11
2,3-dichlorophenol[b]	7.76	39	7.75	−0.01	−769903.9	−769621.4	14.25	345.83	−5.87	−61.67	−7.52
2,4,5-trichlorophenol[b]	7.07	39	7.08	0.01	−1058343.8	−1058062.9	12.67	338.53	−5.86	−56.63	−6.84
2,4,6-trichlorophenol	6.21	39	6.38	0.17	−1058343.2	−1058064.0	11.02	337.79	−6.11	−57.77	−6.86
2,4-dichlorophenol[b]	8.09	39	7.95	−0.14	−769905.7	−769622.7	14.71	345.42	−5.90	−60.78	−7.58
2,5-dichlorophenol[b]	7.51	39	7.53	0.02	−769906.2	−769624.2	13.74	343.69	−5.92	−60.39	−7.24
2,6-dichlorophenol	6.79	39	6.99	0.20	−769903.8	−769623.1	12.45	344.21	−6.59	−62.84	−7.26
2-chlorophenol[b]	8.29	39	8.39	0.10	−481465.1	−481181.1	15.75	352.57	−5.81	−65.95	−8.42
3,4-dichlorophenol	8.68	39	8.51	−0.17	−769904.8	−769620.5	16.03	342.20	−9.39	−59.89	−7.43
3,5-dichlorophenol	8.27	39	8.03	−0.24	−769907.1	−769624.0	14.90	339.78	−8.53	−58.23	−6.94
3-chlorophenol	8.78	39	8.84	0.06	−481466.3	−481181.2	16.81	347.90	−8.94	−63.74	−8.04
4-chlorophenol	9.14	39	9.28	0.14	−481465.9	−481179.8	17.84	349.67	−9.03	−64.22	−8.40
phenol	9.82	39	9.47	−0.35	−193024.2	−192737.6	18.29	357.60	−8.42	−69.69	−9.79
4-nitrophenol	7.14	91	5.99	−1.15	−321426.2	−321147.8	10.11	327.17	−15.46	−57.58	−6.70
5-nitrouracil	5.30	58	5.65	0.35	−388832.9	−388555.4	9.30	339.22	−28.65	−77.98	−12.34
thymine	9.75	58	9.48	−0.27	−285120.9	−284834.4	18.31	342.43	−19.96	−67.50	−8.34
trans-5-formyluracil	6.84	58	6.83	−0.01	−331586.9	−331306.6	12.07	344.87	−23.17	−75.74	−11.99
uracil	9.42	58	9.08	−0.34	−260440.6	−260155.0	17.38	339.85	−20.94	−66.97	−8.20
fluorouracil[b]	8	62	8.13	0.13	−322734.7	−322451.3	15.15	349.43	−21.67	−69.85	−17.86
methylthiouracil[b]	8.2	62	7.72	−0.48	−487789.7	−487507.2	14.17	354.74	−19.99	−69.09	−23.22
phenytoin[b]	8.3	62	8.28	−0.02	−526588.2	−526304.4	15.48	350.00	−19.03	−76.66	−8.64
3,3-methylphenylglutarimide[b]	9.2	62	8.97	−0.23	−420825.9	−420540.5	17.12	358.89	−16.31	−77.57	−12.26
3,3-dimethylsuccinimide[b]	9.5	62	8.35	−1.15	−275792.9	−275509.0	15.66	354.31	−14.14	−72.95	−11.58
dimethadione[b]	6.1	62	6.11	0.01	−298345.7	−298067.1	10.38	344.28	−14.98	−70.30	−10.33
phthalimide[b]	8.3	62	8.34	0.04	−322116.1	−321832.2	15.63	352.68	−13.26	−71.80	−10.26
succinimide[b]	9.6	62	8.34	−1.26	−226437.4	−226153.5	15.64	353.75	−15.49	−73.67	−11.68

[a] The total free energies in water are from COSMO−RS calculations. The meaning of the partial contributions explained in the text. [b] More than one conformation was taken into account in the COSMO−RS calculations.

It appears that the slope for the DMSO pK_a-scale is slightly higher than the one for the aqueous pK_a-scale, but due to the small number of acids considered for DMSO, we cannot clearly rule out the possibility that both slopes may be the same.

Discussion and Conclusions

Dissociation free energies calculated by the quantum chemically based solvation model COSMO−RS, which combines the advantages of dielectric continuum solvation models with a statistical thermodynamics treatment of the deviations from dielectric solvation due to specific local interactions, yield a very good description of experimental pK_a values of a broad range of acids, including standard organic acids, heterocyclic compounds, and even inorganic acids. This result was achieved without any adjustment of the COSMO−RS method, which was originally developed and parametrized for solvation properties

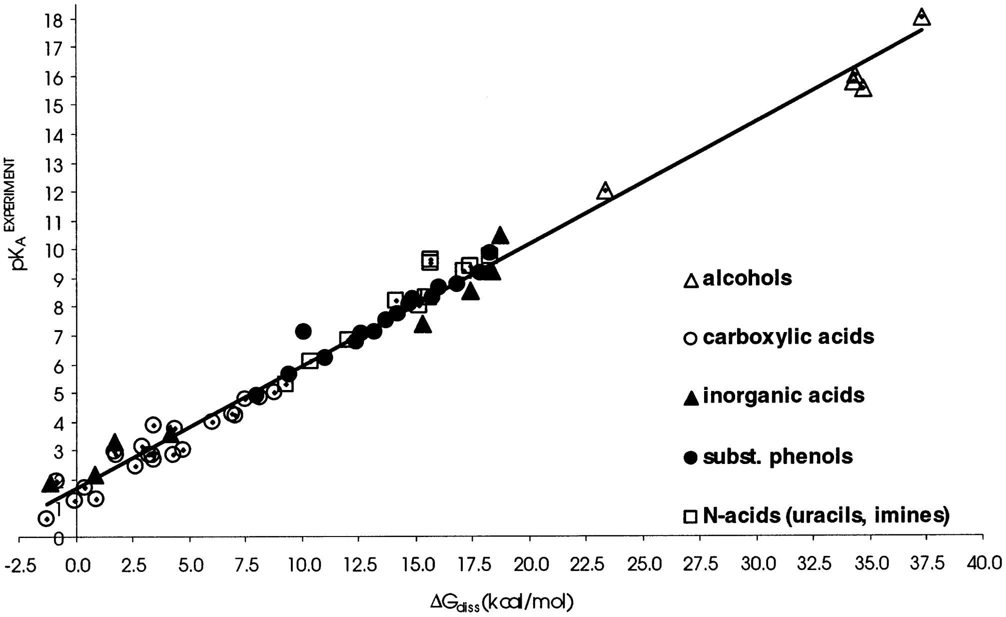

Figure 1. Experimental pK_a vs calculated free energy of dissociation.

TABLE 2: COSMO Energy Differences Anion and Neutral Acid with Different Quantum Chemical Levels and Basis Sets

method basis set	BP TZVP	BP aug-cc-pVTZ	B3LYP TZVP	HF TZVPP	MP2 TZVPP
2,2,2-trichloroethanol	302.65	302.87	307.20	314.82	307.68
chloroaceticacid	282.03	283.16	284.67	290.94	286.69
ethanol	314.81	314.28	318.99	325.35	320.53
formicacid	284.61	285.17	286.81	292.44	287.97
phenol	296.33	297.77	300.33	306.64	301.98
sulfurousacid	276.09	275.71	278.45	284.60	278.04
thymine	294.90	296.23	298.04	304.61	298.22
trichloroaceticacid	273.67	275.33	276.24	283.20	279.06
hypoiodousacid	298.42		301.49		
regress vs BP-TZVP					
slope	1.00	0.98	1.05	1.07	1.05
stat error of slope	0.00	0.02	0.01	0.03	0.03
r^2	1.00	0.997	0.999	0.996	0.993

TABLE 3: PK_a in DMSO for Eight Acids with Experimental[94,95] and Calculated PK_a and Calculated free energy contributions (in kcal/mol)

acid	pK_a^{exp}	ref	pK_a^{CRS}	ΔpK_a	G^{AH}_{H2O}	G^{A-}_{H2O}	ΔG_{diss}
H2O	31.4	94,95	33.60	2.20	−47994.2	−47656.6	58.83
ethanol	29.8	94,95	27.53	−2.27	−97340.1	−97014.3	46.98
methanol	29	94,95	28.48	−0.52	−72658.5	−72330.8	48.86
phenol	18	94,95	16.22	−1.78	−193030.5	−192726.7	24.93
aceticacid	12.6	94,95	13.60	1.00	−143833.7	−143535.0	19.83
benzoicacid	11.1	94,95	11.76	0.66	−264195.9	−263900.8	16.24
succinimide[a]	14.7	94,95	15.41	0.71	−226439.3	−226137.1	23.35

[a] More than one conformation was taken into account in the COSMO−RS calculations.

of neutral compounds. This proves that the COSMO−RS model combined with DFT quantum chemistry is well able to correctly describe the physics of acid dissociation in water and in other solvents, although obviously, room is left for further improvements. Since DFT/COSMO−RS calculations can be applied to almost any kind of organic compounds with moderate computational costs, it provides a tool for the reliable prediction of acidities of complex organic compounds as they often appear in the design and development of drugs and agrochemical agents. Beyond the total pK_a, partial acidities of different acidic sites of a compound can also be easily evaluated. This information is often needed for a better understanding of molecular interactions, especially in drug design.

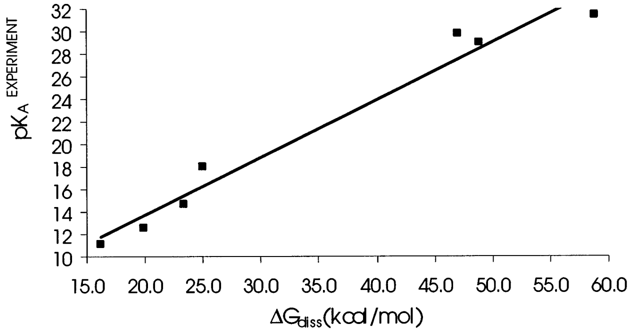

Figure 2. Experimental pK_a vs calculated free energy of dissociation in DMSO.

Despite the good correlation, we clearly find a slope of the experimental pK_a-scale versus the calculated free energy of dissociation, which is only about 60% of the usually expected slope of 1/RTln(10). Such low slope has been found in several other ab initio calculations of pK_a before, but so far the authors either assumed deficiencies of the calculations to be responsible for this results, or they distorted the parameters of the solvation models, until the final results gave reasonable agreement with the expected slope. Indeed, one of the authors also found such low slope 10 years ago in an unpublished study based on semiempirical AM1/COSMO calculations, but at that time, the level of quantum chemical accuracy and the trust in the applied solvation model were by far not sufficient to draw any conclusion from this deviation. In the present study, the statistical uncertainty as well as the error bars in the theoretical calculation of ΔG_{diss} are far too small for achieving agreement with the standard expectation for the slope of the pK_a-scale.

Our calculations suggest that the slope A may be in the range of 0.6 (i.e., 60% of the theoretical expectation). Chipman[75] reported a slope of 0.47. The average slope found by Klicic' et al.[76] for DFT calculations in combination with a Langevin dipole solvation method is 0.60 for all acids and bases, and 0.48 if only acids are included, but they find a large scatter of A, because they are partly fitting on very small data sets for compound classes covering only a narrow pK_a range.

There are two recent studies that appear to achieve agreement with the theoretical slope by inclusion of a few (1−3) explicit water molecules in the DCSM calculations for the free energy of solvation of the ions, while they stay with standard DCSM calculations for the neutral species. The study of Pliego et al. reports an almost one to one correlation with an rms deviation of 2.1 of calculated pK_a versus experimental pK_a for 15 acids and bases covering about 38 pK_a units, without any special fitting apart from the addition of explicit water molecules. However, a regression analysis shows that their exact slope corresponds to $A = 0.90$. This slope is strongly influenced by the single value of the acid NH_3 with experimental pK_a of 33, which is definitely not measured in pure water. If we exclude all values from that study that are outside the natural aqueous pK_a scale of 0−16, the slope reduces to $A = 0.70$ ($\pm$0.09), which is quite compatible with our findings in this pK_a range. In a study by Adam,[97] the number of explicit water molecules added to the anions is increases individually for different classes of acids, until the right slope is achieved. This appears to be quite biased. It does not appear that the results are converged with respect to explicit solvent molecules. In addition, we doubt that the balance between explicitly treated water molecules and implicit solvation is sufficiently good to justify such treatment. Finally, there is considerable bias arising from the exact placement of the explicit water molecules around the ions.

Considering the very good correlation, the consistency of the slope of our regression with respect to the different free energy contributions and the good agreement of our finding with almost all other first principle pK_a calculation studies, we thus have to draw the conclusion that the experimental pK_a-scale does not correspond to the free energy of dissociation in infinite dilution of an acid in pure water and probably even so in other solvents. Being sure about the computational significance of the low slope, we do not know whether to search for the origin of the discrepancy on the theoretical side or on the experimental side, and indeed we consider both as equally unlikely. Therefore, we hereby pass the problem forward to the scientific community for further experimental and theoretical research.

References and Notes

(1) Bell, R. P. *The Proton in Chemistry*; Chapman and Hall: London, 1973.

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

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

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

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

(6) Bagno, A.; Scorrano, G. *J. Phys. Chem.* **1996**, *100*, 1536.

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

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

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

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

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

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

(13) Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. *Quantum Struct. Act. Rel* **1995**, *14*, 348.

(14) Perrin, D. D.; Dempsey, B.; Sergeant, E. P. *pK_a Prediction for Organic Acids and Bases*; Chapman and Hall: New York, 1981.

(15) ACD/pK_a DB, Advanced Chemistry Development: Toronto ON, Canada, 2003, www.acdlabs.com.

(16) PKalc 4.0 User's Guide CompuDrug Ltd. Budapest: Princetown, 1998. Also available: www.compudrug.com.

(17) Klopman, G.; Fercu, D. *J. Comput. Chem.* **1994**, *15*, 1041.

(18) Xing, L.; Glen, R. C. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 796−805.

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

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

(21) Klamt, A: COSMO and COSMO−RS. In *Encyclopedia of Computational Chemistry*, P. v. R. Schleyer, Ed.; Wiley: New York, 1998.

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

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

(24) Rullmann, J. A. C.; van Duijnen, P. Th. *Mol. Phys.* **1988**, *63*, 451.

(25) Guissani, Y.; Guillot, B.; Bratos, S. *J. Chem. Phys.* **1988**, *88*, 5850.

(26) Tanaka, Y.; Shiratori, Y.; Nakagawa, S. *Chem. Phys. Lett.* **1990**, *169*, 513.

(27) Gao, J.; Li, N.; Freinforf, M. *J. Am. Chem. Soc.* **1996**, *118*, 4912.

(28) Figueirido, F.; Del Buono, G. S.; Levy, R. M. *J. Phys. Chem.* **1996**, *100*, 6389.

(29) Byun, K.; Mo, Y.; Gao, J. *J. Am. Chem. Soc.* **2001**, *123*, 3974.

(30) Rashin, A. A.; Rabinowitz, J. R.; Banfelder, J. R. *J. Am. Chem. Soc.* **1990**, *112*, 4133.

(31) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8552.

(32) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610.

(33) Chen, J. L.; Noodleman, L.; Case, D. A.; Bashford, D. *J. Phys. Chem.* **1994**, *98*, 11059.

(34) Tawa, G. J.; Pratt, L. R. *J. Am. Chem. Soc.* **1995**, *117*, 1625.

(35) Andzelm, J.; Kölmel, C.; Klamt, A. *J. Chem. Phys.* **1995**, *103*, 9312.

(36) Peräkylä, M. *J. Org. Chem.* **1996**, *61*, 7420.

(37) Li, J.; Fisher, C. L.; Chen, J. L.; Bashford, D.; Noodleman, L. *Inorg. Chem.* **1996**, *35*, 4694.

(38) Richardson, W. H.; Peng, C.; Bashford, D.; Noodleman, L.; Case, D. A. *Int. J. Quantum Chem.* **1997**, *61*, 207.

(39) Schüürmann, G. *Quant. Struct.-Act. Relat.* **1996**, *15*, 121−132.

(40) Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. *J. Phys. Chem. A* **1997**, *101*, 10075.

(41) Shapley, W. A.; Bacskay, G. B.; Warr, G. G. *J. Phys. Chem. B* **1998**, *102*, 1938.

(42) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.

(43) Peräkylä, M. *J. Am. Chem. Soc.* **1998**, *120*, 12895.

(44) Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.

(45) Schüürmann, G. *J. Chem. Phys.* **1998**, *109*, 9523.

(46) Kubicki, J. D. *J. Phys. Chem. A* **1999**, *103*, 903.

(47) Peräkylä, M. *Phys. Chem. Chem. Phys.* **1999**, *1*, 5643.

(48) da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.

(49) Lee, I.; Kim, C. K.; Han, I. S.; Lee, H. W.; Kim, W. K.; Kim, Y. B. *J. Phys. Chem. B* **1999**, *103*, 7302.

(50) Lee, I.; Kim, C. K.; Lee, I. Y.; Kim, C. K. *J. Phys. Chem. A* **2000**, *104*, 6332.

(51) Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. *J. Phys. Chem. A* **2000**, *104*, 2402.

(52) Chen, I.-J.; MacKerell, Jr., A. D. *Theor. Chem. Acc.* **2000**, *103*, 483.

(53) Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. *J. Phys. Chem. A* **2000**, *104*, 9619.

(54) Zhan, C.-G.; Zheng, F. *J. Am. Chem. Soc.* **2001**, *123*, 2835.

(55) Bartberger, M. D.; Fukuto, J. M.; Houk, K. N. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 2194.

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

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

(58) Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A., III *J. Phys. Chem. A* **2001**, *105*, 271.

(59) Kubicki, J. D. *J. Phys. Chem. A* **2001**, *105*, 8756.

(60) Topol, I. A.; Nemukhin, A. V.; Dobrogorskaya, Y. A.; Burt, S. K. *J. Phys. Chem. B* **2001**, *105*, 11341.

(61) Li, H.; Hains, A. W.; Everts, J. E.; Robertson, A. D.; Jensen, J. H. *J. Phys. Chem. B* **2002**, *106*, 3486.

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

(63) Florian, J.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 5583.

(64) In refs 62 and 63, a Langewin dipole solvation model was used, which is in practice quite similar to dielectric continuum solvation.

(65) Kawata, M.; Ten-no, S.; Kato, S.; Hirata, F. *Chem. Phys.* **1996**, *203*, 53.

(66) Kawata, M.; Ten-no, S.; Kato, S.; Hirata, F. *J. Phys. Chem.* **1996**, *100*, 1111.

(67) Sato, H.; Hirata, F. *J. Phys. Chem. A* **1998**, *102*, 2603.

(68) Sato, H.; Hirata, F. *J. Phys. Chem. B* **1999**, *103*, 6596.

(69) Sakurai, M.; Furuki, T.; Inoue, Y. *J. Phys. Chem.* **1995**, *99*, 17789.

(70) Wiberg, K. B.; Castejon, H.; Keith, T. A. *J. Comput. Chem.* **1996**, *17*, 185.

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

(72) Kovacevic', B.; Maksic', Z. B. *Org. Lett.* **200**1, *3*, 1523.

(73) Maksic', Z. B.; Vianello, R. *J. Phys. Chem. A* **2002**, *106*, 419.

(74) Batoulis, J.; Klamt, A. Bayer research report 1993, unpublished.

(75) Chipman, D. M. *J. Phys. Chem. A, Vol. 106, No. 32, 2002.*

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

(77) Himo, F.; Noodleman, L.; Bloomberg, M. R. A.; Siegbahn, P. E. M. *J. Phys. Chem. A* **2002**, *106*, 8757−8761.

(78) A. Klamt, A.; Eckert, F. *Fluid Phase Equilibria* **2000**, *172*, 43.

(79) Eckert, F.; Klamt, A. *Ind. Eng. Chem. Res.* **2001**, *40*, 2371−2378.

(80) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkins Trans.2*, 1993, 799.

(81) Beck, M. E. *Prediction of Acidity for Agrochemicals*, oral presentation on the EuroQSAR 2002 Conference, 2002, Bournemouth, UK.

(82) Klamt, A.; Eckert, F. COSMOtherm, Version C1.2-Revision 07.02; COSMO*logic* GmbH&CoKG, Leverkusen, Germany 2001. See also URL: http://www.cosmologic.de.

(83) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.

(84) Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F.; *Phys. Chem. Chem. Phys.* **2000**, *2*, 2187.

(85) Ahlrichs, R.; Bär, M.; Baron, H.-P.; Bauernschmitt, R.; Böcker, S.; Ehrig, M.; Eichkorn, K.; Elliott, S.; Furche, F.; Haase, F.; Häser, M.; Horn, H.; H.attig, C.; Huber, C.; Huniar, U.; Kattannek, M.; Köhn, M.; Kölmel, C.; Kollwitz, M.; May, K.; Ochsenfeld, C.; Öhm, H.; Schäfer, A.; Schneider, U.; Treutler, O.; von Arnim, M.; Weigend, F.; Weis, P.; Weiss, H.: Turbomole Version 5.6 2002.

(86) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(87) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822.

(88) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *242*, 652.

(89) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(90) Diedenhofen, M.; Klamt, A. To be published.

(91) R. Williams, taken from: http://icg.harvard.edu/~chem206/Fall_2001/lectures/18_Acid_Based_Properties_of_Organic_Molecules/pK_a_compilation.pdf.

(92) *CRC Handbook of Chemistry and Physics*, Lide, D. R., Ed.; CRC Press: New York, CD-ROM version, 2000.

(93) This is an averaged value because pK_a values ranging from 17−19.2 are reported in different sources.

(94) Pliego, J. R., Jr. *Chem. Phys. Lett.* **2003**, *36*, 145−149.

(95) Bordwell, F. G. *Acc. Chem. Res.* **1988**, *21*, 456−463.

(96) Reich, H. J.: http://www.chem.wisc.edu/areas/reich/pkatable.

(97) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963−11972.