This is a postprint of the Journal of Computational Chemistry, Volume 30, April 2009, Pages 799-810. The original article can be found under http://onlinelibrary.wiley.com/doi/10.1002/jcc.21103/full

Prediction of Acidity in Acetonitrile Solution with COSMO-RS

Frank Eckert[a,*], *Ivo Leito*[b,*], *Ivari Kaljurand*[b], *Agnes Kütt*[b], *Andreas Klamt*[a,c], *Michael Diedenhofen*[a]

[a]COSMOlogic GmbH & Co KG, Burscheider Str. 515, D-51381 Leverkusen, Germany

[b]University of Tartu, Institute of Chemical Physics, Jakobi 2, Tartu 51014, Estonia

[c]University of Regensburg, Institute of Physical and Theoretical Chemistry, 93040 Regensburg, Germany

[*] To whom correspondence should be addressed. Email: frank.eckert@cosmologic.de. Phone: +492171731680. Fax: +492171731689. Email: ivo.leito@ut.ee. Phone: +3725184176. Fax: +3727375264.

ABSTRACT: The COSMO-RS method, a combination of the quantum chemical dielectric continuum solvation model COSMO with a statistical thermodynamics treatment for realistic solvation simulations, has been used for the prediction of pK_a values in acetonitrile. For a variety of 93 organic acids the directly calculated values of the free energies of dissociation in acetonitrile showed a very good correlation with the pK_a values (r^2 = 0.97) in acetonitrile, corresponding to a standard deviation of 1.38 pK_a units. Thus we have a prediction method for acetonitrile pK_a with the intercept and the slope as the only adjusted parameters. Furthermore, the pK_a values of CH acids yielding large anions with delocalized charge can be predicted with a rmse of 1.12 pK_a units using the theoretical values of slope and intercept resulting in truly *ab initio* pK_a prediction. In contrast to our previous findings on aqueous acidity predictions the slope of the experimental pK_a versus theoretical ΔG_{diss} was found to match the theoretical value 1/RTln(10) very well. The predictivity of the presented method is general and is not restricted to certain compound classes. However, a systematic correction of -7.5 kcal·mol^{-1} is required for compounds that do not allow electron-delocalization in the dissociated anion. The prediction model

was tested on a diverse test set of 129 complex multifunctional compounds from various sources, reaching a root mean square deviation of 2.10 pK_a units.

KEYWORDS: pK_a; acetonitrile; acidity; COSMO; COSMO-RS; density functional theory;

Introduction

Proton transfer is one of the fundamental processes in chemistry and biology. Thus the understanding and the prediction of the thermodynamics of the proton transfer reaction and the dissociation constants of acids and bases in different solvents are of crucial importance in many areas of chemistry and biochemistry. Experimental measurement of aqueous phase pK_a values nowadays has become an inexpensive standard application[1]. The same cannot be said about measurement of pK_a values in nonaqueous solvents. In addition, there are broad classes of chemicals that are not readily amenable to experimental characterization (e.g. reaction intermediates, very strong and very weak acids or bases with a pK_a outside the "natural" pK_a range that can be conveniently measured). Consequently, considerable effort has been devoted to develop first principle prediction methods for pK_a values. Acetonitrile is a useful solvent for ionic reactions, including acid-base reactions. It has a high dielectric constant (ε = 36.0)[2] and thus favors dissociation of ion pairs into ions. At the same time it has low basicity and extremely low acidity resulting in a very low autoprotolysis constant[3] of p$K_{auto} \geq 33$. The low acidity also implies that acetonitrile has very low ability for specific solvation of anions. These properties put together make acetonitrile a very good differentiating solvent, especially for studies of acids. pK_a measurements of acids and bases in acetonitrile date back to the classic works of the groups of Kolthoff[3,4] and Coetzee[2,5] in the 1960s. The pK_a data in acetonitrile published up to 1990 have been gathered in the compilation of Izutsu[6]. During the recent decade spectrophotometric pK_a scales of acids[7] and bases[8] both containing around hundred compounds and spanning for more than 20 orders of

magnitude have been set up in acetonitrile. These are the most consistent datasets of pK_a values currently available in acetonitrile.

The rapid development of efficient quantum chemical (QC) methods in the last years has opened new perspectives for the rigorous prediction of liquid phase pK_a values. Of the different quantum chemical methodologies available for the computation of pK_a values the dielectric continuum solvation methods (DCSMs[9]) have become quite popular in the recent years[10-17] since they are able to describe accurately long range electrostatic interactions of solutes at moderate computational cost in the context of quantum chemical programs. Despite the well known deficiencies of DCSM methods, (i.e. the neglect of hydrogen bonding and the inadequate treatment of the short range electrostatics[10,18-21], which can be much stronger in ions than in neutrals and thus can introduce a large asymmetry to the solvation energy of an acid compared to its conjugate base) it is possible to correlate the quantum chemical dissociation free energy of a solvated molecule ΔG_{diss} with its pK_a via a linear free energy relationship (LFER) [10]:

$$\mathrm{p}K_{\mathrm{a}} = c_1 \frac{\Delta G_{\mathrm{diss}}}{RT\ln(10)} + c_2 \qquad (1)$$

From the basic thermodynamics c_1 is expected to be unity if ΔG_{diss} would be calculated without a systematic error and the LFER axis intercept c_2 is expected to be equal to -log[Solvent][22]. Looking in detail into the DCSM studies,[10-17] in the regression of pK_a values versus the calculated dissociation free energy ΔG_{diss} the studies report slopes that are significantly lower than the theoretically expected value of $1/RT\ln(10)$. Such a behavior has been reported for aqueous[10-12] and non-aqueous acids[10,13-15,23] as well as for bases.[16,17,24] This drawback is common to all simple DCSMs unless considerable effort is taken in the (often physically hardly justifiable) adjustment of numerous additional and often physically doubtful parameters of the DCSM. Atom type or hybridization specific cavity radii and cavity definitions that depend on the charge of the molecule are examples of such parameters[25]. Although such models became quite popular and successful applications for nonaqueous solvents have been reported[26-28], it remains

doubtful if the predictive power of such empirical adjustments persists for more complex chemically multifunctional solutes or for solutes such as free radicals, zwitterions or excited states[23].

Quite some effort has been devoted to the computational prediction of pKa values in Acetonitrile. Most of the works have focused on computation of pK_a values of cationic acids (protonated bases) and to the best of our knowledge all of them use experimental pK_a data to achieve useful predictive power for their approaches. Moreover, the adjustment of these cavity specific parameters (and thus also the quantum chemical DCSM computation of the solute acid and conjugate base) has to be done anew for each new solvent considered, making this approach hardly practical or extensible.

To avoid such problems Chipman[23] proposed a DCSM on isodensity cavities, which claims to describe both cationic and neutral acids by a single correlation line between computational and experimental pK_a values. There are, however, only six data points, which is too few and all the cationic acids included in the correlation have lower pK_a values than any of the neutral acids. Furthermore, in refs. 7 and 29 new, more accurate, pK_a values for acetic acid, benzoic acid and phenol have been published, which are all higher (by up to 2 pK_a units) than the earlier values used by Chipman. Substitution of the new values to the correlation leads to the increase of the rmse of the correlation from 0.3 to 0.6 pK_a units. Thus, as admitted also by Chipman, too far-reaching conclusions should not be made. A related isodensity DCSM approach has been used by the Maksić group in number of computational acetonitrile pK_a studies of bases[30]. Most of their works aim at (and achieve) highly accurate pK_a predictions within groups of closely related compounds and therefore use experimental pK_a values of structurally similar compounds to "calibrate" the computations, thus achieving rmse values down to 0.3 pK_a units.

A promising approach to the pK_a problem, which does not artificially modify the cavity to try and reproduce hydrogen bonding and short-range solute-solvent interaction behavior that is not accounted for by the DCSM, is the addition of explicit solvent molecules to the solute ions[31-34]: a solute anion is represented by a cluster of the anionic solute molecule with one or more surrounding solvent molecules that form a partial or full solvation shell around the ion, accounting for strong solute-solvent interactions

in a physical way. Although this approach has the advantage that the slope of the aqueous pK_a LFER is reported to be significantly closer to the theoretical slope compared to simple DCSMs[31, 32], its practical application leads to some ambiguities and problems, especially in the case of nonaqueous solvents: there is no natural choice of the number of solvent molecules that represent the solvent shell, retaining some level of arbitrariness involved, where a choice has to be taken. However, what in practice might turn out to be the much harder problem, is the optimization of the solute-solvent cluster. For complex, multifunctional solutes, as most chemically or biologically interesting drug-like compounds are, it is very difficult and computationally demanding to find the global minimum of the weakly bonded solute-solvent complex. If the solvent itself is a complex multifunctional compound, or if a mixture of several solvent compounds is used, it easily may become impossible to find the global minimum of the cluster at all. From these practical considerations the computation of the large and complex data sets used below, the explicit solvation approach was outside the scope of this study. In addition, the explicit goal of the study was to provide a methodology that is very simple on the level of the quantum chemistry involved and that the solute compounds computed on the quantum chemistry level are "transferable", meaning that they can be used for pK_a, predictions in other solvents or even solvent mixtures as well, without the need of recomputing them (as the modified cavity and the explicit solvation models would demand). Thus we chose an approach different from the ones already mentioned: the Conductor-like Screening Model for Real Solvents (COSMO-RS).

COSMO-RS,[18-21] goes beyond the DCSM concept in that it combines the electrostatic advantages and the computational efficiency of the DCSM COSMO[35] 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 QC-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, and solvation free energies of neutral compounds with a root mean square error (rmse) of 0.3 log-units and better and a lot of experience has

been gathered during the past years about its surprising ability to predict mixture thermodynamics[18-20]. Stimulated by the successful COSMO-RS predictions of aqueous acidity[10] and basicity[24] as well as some preliminary studies in nonaqueous solvents,[10] we decided to perform a systematic study on the ability of COSMO-RS to predict pK_a values of acids in acetonitrile. For that purpose we calculated ΔG_{diss} for a broad selection of 93 organic acids in acetonitrile, spanning a pK_a range between 3 and 27, and using the standard COSMO-RS method implemented in the COSMO*therm* program[36] based on Turbomole DFT/COSMO calculations[37-39].

Theoretical calculations

Our theoretical calculations of ΔG_{diss} of acids in acetonitrile are based on the reaction model

$$AH + CH_3CN \rightarrow A^- + CH_3CNH^+ \qquad (2)$$

Since we are not interested in the gas phase reaction, we directly calculated the free energy of each species in acetonitrile solutions. For that we first applied our standard procedure for COSMO-RS calculations to all four species appearing in eq. 2, which consists of two steps:

1) Full DFT geometry optimization with the Turbomole program package[39] using B-P density functional[40,41] with TZVP quality basis set using the RI approximation.[42] During these calculations the COSMO continuum solvation model was applied in the conductor limit ($\varepsilon = \infty$). Element-specific default radii from the COSMO-RS parameterizations have been used for the COSMO cavity construction.[19,20] Such calculations end up with the self-consistent state of the solute in the presence of a virtual conductor, that surrounds the solute outside the cavity.

2) COSMO-RS calculations have been done using the COSMO*therm* program[36]. In these calculations the deviations of the real solvent, in this case acetonitrile, compared to an ideal conductor are taken into account in a model of pair-wise 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 pair-wise interacting surfaces, which is implemented in COSMO*therm*. More detailed descriptions of the COSMO-RS method are given elsewhere[18-21].

If more than one conformation or different deprotonation sites were considered to be potentially relevant for the neutral or anionic form of the acid AH, several conformations were calculated in step 1 and a thermodynamic Boltzmann average over the total Gibbs free energies of the conformers was consistently calculated by the COSMO*therm* program in step 2.

For all acids AH, the Gibbs free energy of dissociation (ΔG_{diss}) has been calculated as the difference of the total free energy of the anion A^- and the neutral acid AH. To this free energy difference the free energy difference of CH_3CNH^+ and CH_3CN has been added as a constant contribution:

$$\Delta G_{diss} = G_{tot}(A^-) - G_{tot}(AH) + \left[G_{tot}(CH_3CNH^+) - G_{tot}(CH_3CN)\right] \quad (3)$$

From the calculation procedure described above, we get $G_{tot}(CH_3CNH^+)$ - $G_{tot}(CH_3CN)$ = 253.48 kcal·mol^{-1}. This value is in good agreement with literature estimates[23, 43]. Zero point vibrational energies are not taken into account. Consequently, the geometries optimized in step 1 were not analyzed for the nature of the stationary point of the optimized geometry. We make the common assumption that the difference in zero point energy between the neutral and the deprotonated acid is generally small[10]. Moreover, we did not take into account the symmetric multiplicity factors of the compounds conformations, because we did not feel able to do this consistently for all kinds of acids in the same way.

Fit Data Set

For the purpose of finding the LFER coefficients of eq. 1, a data set of 93 acids in acetonitrile was used. The data were taken from ref. 7. The pK_a values in the lower end of the scale (below pK_a = 9, i.e. starting from TosOH) of ref. 7 have been corrected downwards by 0.1 to 0.15 pK_a units because we

discovered an error in the data of ref. 7 in the region of pK_a values 7 to 9. The reason for this is twofold: (a) in the region of pK_a values from 7 to 9 there are only five compounds in the scale (resulting in a smaller number of overlapping ΔpK_a measurements than in other parts of the scale) and even more importantly (b) three out of these five compounds (TosOH, 4-Cl-$C_6H_4SO_3H$ and $C_6H_5CHTf_2$) are inconvenient for measurements as they have not very suitable spectral properties and in addition TosOH and 4-Cl-$C_6H_4SO_3H$ undergo homoconjugation in MeCN, which, although taken into account, complicates measurements and reduces their accuracy. Because the scale is anchored to the pK_a value of picric acid (pK_a = 11.0), the error in the region of pK_a values 7 to 9 influenced the pK_a values of all the acids that are stronger. The error was discovered by additional careful ΔpK_a measurements. Although unfortunate, this shift in pK_a values is quite small and has no influence in most applications. The pK_a values range between 3 and 27. The dataset consists of (a) 23 OH acids, namely 5 sulfonic acids, 14 aromatic alcohols, 1 aliphatic alcohol and 2 carboxylic acids; (b) 32 NH acids, namely 3 aromatic secondary amines, 1 aniline, 21 sulfonimides and 7 carbonylsulfonimides and (c) 38 CH acids, namely 31 trisubstituted methanes, 6 fluorenes and 1 cyclopentadiene. The results for all 93 acids in the fit data set are shown in Table 1. The regression of the calculated Gibbs free energy of dissociation (ΔG_{diss}) vs. experimental pK_a in acetonitrile is depicted in Figure 1.

TABLE 1: Fit data set for COSMO-RS acid pK_a calculations in acetonitrile.[a]

Compound[b]	CAS-RN	Type	Class	Delocalized[c]	ΔG_{diss}	pK_a^{Exp}	pK_a^{Calc}	pK_a^{Calc} (corr)
$(C_6F_5)CH(CN)COOEt$	2340-87-6	CH	methane	yes	27.93	17.75	16.15	19.20
$(4\text{-}CF_3\text{-}C_6F_4)CH(CN)COOEt$	32251-53-9	CH	methane	yes	24.47	16.08	13.46	16.73
$4\text{-}Me\text{-}C_6H_4CH(CN)_2$	33534-88-2	CH	methane	yes	23.89	17.59	13.01	16.31
$(C_6H_5)(C_6F_5)CHCN$	42238-33-5	CH	methane	yes	37.30	26.14	23.44	25.90
$(C_6F_5)_2CHCN$	42238-34-6	CH	methane	yes	31.22	21.10	18.72	21.56
$(4\text{-}CF_3\text{-}C_6F_4)(C_6F_5)CHCN$	42238-35-7	CH	methane	yes	26.33	18.14	14.91	18.06
$(4\text{-}Cl\text{-}C_6F_4)(C_6F_5)CHCN$	42238-36-8	CH	methane	yes	30.00	20.36	17.76	20.68
$(4\text{-}H\text{-}C_6F_4)(C_6F_5)CHCN$	42254-09-1	CH	methane	yes	31.34	21.11	18.80	21.64
$(4\text{-}Me\text{-}C_6F_4)(C_6F_5)CHCN$	52345-34-3	CH	methane	yes	32.60	21.94	19.78	22.54
$(4\text{-}Cl\text{-}C_6F_4)CH(CN)COOEt$	55810-56-5	CH	methane	yes	27.16	17.39	15.55	18.65
$(4\text{-}NC_5F_4)CH(CN)COOEt$	55810-61-2	CH	methane	yes	23.29	14.90	12.54	15.88
$4\text{-}H\text{-}C_6F_4CH(CN)_2$	55810-63-4	CH	methane	yes	19.79	12.98	9.82	13.38
$(4\text{-}H\text{-}C_6F_4)CH(CN)COOEt$	55852-22-7	CH	methane	yes	28.48	18.08	16.58	19.60
$4\text{-}CF_3\text{-}C_6F_4CH(CN)_2$	55852-24-9	CH	methane	yes	14.75	10.19	5.90	9.77
$(4\text{-}Me\text{-}C_6F_4)_2CHCN$	58432-44-3	CH	methane	yes	33.96	22.80	20.85	23.52
$(4\text{-}CF_3\text{-}C_6F_4)_2CHCN$	58432-55-6	CH	methane	yes	22.27	16.13	11.75	15.15
$(4\text{-}Me\text{-}C_6F_4)(C_6H_5)CHCN$	58432-62-5	CH	methane	yes	38.99	26.96	24.76	27.11
$(2\text{-}C_{10}F_7)CH(CN)COOEt$	62325-34-2	CH	methane	yes	27.14	17.50	15.54	18.64
$2\text{-}C_{10}F_7CH(CN)_2$	62325-35-3	CH	methane	yes	18.55	12.23	8.85	12.49
$(4\text{-}NC_5F_4)(2\text{-}C_{10}F_7)CHCN$	62325-37-5	CH	methane	yes	23.98	16.02	13.08	16.38
$(2\text{-}C_{10}F_7)_2CHCN$	62325-38-6	CH	methane	yes	28.08	19.32	16.27	19.31
$(4\text{-}NC_5F_4)(C_6F_5)CHCN$	62325-51-3	CH	methane	yes	23.44	16.40	12.66	15.99
$(2\text{-}C_{10}F_7)(C_6F_5)CHCN$	64934-68-5	CH	methane	yes	29.47	20.08	17.35	20.30
$(2,4,6\text{-}Cl_3\text{-}C_6F_2)(C_6F_5)CHCN$	64934-69-6	CH	methane	yes	29.76	20.13	17.58	20.51
$4\text{-}Me\text{-}C_6F_4CH(CN)_2$	64934-71-0	CH	methane	yes	21.14	13.87	10.87	14.34
$(4\text{-}NC_5F_4)_2CHCN$	64934-72-1	CH	methane	yes	19.79	13.46	9.82	13.38
$C_6F_5CH(CN)_2$	719-38-0	CH	methane	yes	19.88	13.01	9.89	13.44
$3\text{-}CF_3\text{-}C_6H_4CH(CN)_2$	99726-60-0	CH	methane	yes	20.10	14.72	10.06	13.60
$4\text{-}NO_2\text{-}C_6H_4CH(CN)_2$	7077-65-8	CH	methane	yes	16.09	11.61	6.94	10.73
$C_6H_5CHTf_2$	40906-82-9	CH	methane	yes	13.78	7.85	5.14	9.08
$(C_6F_5)CH(COOEt)_2$	1582-05-4	CH/OH[d]	methane	yes	35.12	22.85	21.75	24.34
2,3,5-tricyanocyclopentadiene	215395-09-8	CH	cyclopentadiene	yes	8.88	4.16	1.33	5.57
$9\text{-}C_6F_5\text{-}Fluorene$	73482-93-6	CH	fluorene	yes	42.17	28.11	27.24	29.39
Fluoradene	205-94-7	CH	fluorene	yes	36.15	23.90	22.55	25.08
9-COOMe-Fluorene	3002-30-0	CH	fluorene	yes	35.04	23.53	21.68	24.28
9-CN-Fluorene	1529-40-4	CH	fluorene	yes	30.80	21.36	18.39	21.25
$9\text{-}C_6F_5\text{-}Octafluorofluorene$	63264-80-2	CH	fluorene	yes	31.41	18.88	18.86	21.69
Octafluorofluorene	27053-34-5	CH	fluorene	yes	40.09	24.49	25.62	27.90
$2,4,6\text{-}Br_3\text{-}Phenol$	118-79-6	OH	phenol	no	33.18	20.35	20.24	17.59
$4\text{-}NC_5F_4\text{-}OH$	2693-66-5	OH	phenol	no	27.42	15.40	15.75	13.47
$4\text{-}CF_3\text{-}2,3,5,6\text{-}F_4\text{-}Phenol$	2787-79-3	OH	phenol	no	28.39	16.62	16.51	14.16
$4\text{-}C_6F_5\text{-}2,3,5,6\text{-}F_4\text{-}Phenol$	2894-87-3	OH	phenol	no	31.29	18.11	18.76	16.24
$1\text{-}C_{10}F_7OH$	5386-30-1	OH	phenol	no	33.08	19.72	20.16	17.52
$2,3,4,5,6\text{-}Br_5\text{-}Phenol$	608-71-9	OH	phenol	no	28.95	17.83	16.95	14.56
$2\text{-}C_{10}F_7OH$	727-49-1	OH	phenol	no	31.68	18.50	19.07	16.52
$2,3,5,6\text{-}F_4\text{-}Phenol$	769-39-1	OH	phenol	no	34.44	20.12	21.22	18.49
$2,3,4,5,6\text{-}F_5\text{-}Phenol$	771-61-9	OH	phenol	no	33.84	20.11	20.75	18.06
$2,3,4,5,6\text{-}Cl_5\text{-}Phenol$	87-86-5	OH	phenol	no	29.71	18.02	17.53	15.11
$2,4,6\text{-}(SO_2F)_3\text{-}Phenol$	882492-01-5	OH	phenol	no	11.33	5.53	3.23	1.96
$2\text{-}NO_2\text{-}Phenol$	88-75-5	OH	phenol	no	39.41	22.85	25.09	22.05
$2,4\text{-}(NO_2)_2\text{-}Phenol$	51-28-5	OH	phenol	no	29.54	16.66	17.41	14.99
Picric acid	88-89-1	OH	phenol	no	20.30	11.00	10.21	8.37
$2,4,6\text{-}Tf_3\text{-}Phenol$	71571-37-4	OH	phenol	no	13.50	4.80	4.92	3.51
$(CF_3)_3COH$	2378-02-01	OH	alcohol	no	32.71	20.55	19.88	17.26
Acetic acid	64-19-7	OH	carboxylic acid	no	41.89	23.51	27.01	23.82
Benzoic acid	65-85-0	OH	carboxylic acid	no	38.28	21.51	24.20	21.24
TosOH	104-15-4	OH	sulfonic acid	no	21.71	8.45	11.31	9.38
$4\text{-}NO_2\text{-}C_6H_4SO_3H$	138-42-1	OH	sulfonic acid	no	17.26	6.60	7.85	6.20

1-$C_{10}H_7SO_3H$	85-47-2	OH	sulfonic acid	no	20.23	7.89	10.16	8.33
3-NO_2-$C_6H_4SO_3H$	98-47-5	OH	sulfonic acid	no	17.77	6.65	8.25	6.57
4-Cl-$C_6H_4SO_3H$	98-66-8	OH	sulfonic acid	no	19.60	7.16	9.67	7.88
4-Me-C_6H_4C(=O)NHTf	343337-70-2	NH	carbonylsulfonamide	no	23.78	11.46	12.92	10.87
C_6H_5C(=O)NHTf	39062-91-4	NH	carbonylsulfonamide	no	23.62	11.06	12.80	10.75
4-NO_2-C_6H_4C(=O)NHTf	39062-98-1	NH	carbonylsulfonamide	no	20.90	9.49	10.68	8.80
4-Cl-C_6H_4C(=O)NHTf	39062-99-2	NH	carbonylsulfonamide	no	22.57	10.36	11.98	10.00
4-F-C_6H_4C(=O)NHTf	39063-00-8	NH	carbonylsulfonamide	no	23.57	10.65	12.76	10.71
4-MeO-C_6H_4C(=O)NHTf	39063-05-3	NH	carbonylsulfonamide	no	24.57	11.60	13.54	11.43
Saccharin	81-07-2	NH	carbonylsulfonamide	no	29.83	14.57	17.63	15.19
4-NO_2-$C_6H_4SO_2$NHTos	100724-78-5	NH	sulfonimide	no	24.20	10.04	13.25	11.16
$C_6H_5SO_2$NHTf	174788-87-5	NH	sulfonimide	no	18.24	5.89	8.61	6.90
4-Cl-$C_6H_4SO_2$NHTf	174788-89-7	NH	sulfonimide	no	17.93	5.34	8.37	6.68
4-NO_2-$C_6H_4SO_2$NHTf	174788-91-1	NH	sulfonimide	no	16.49	4.39	7.25	5.65
4-Cl-C_6H_4SO(=NTf)NHTos	174788-93-3	NH	sulfonimide	no	17.43	5.14	7.98	6.32
4-Cl-C_6H_4SO(=NTf)NH$SO_2C_6H_4$-4-Cl	174788-95-5	NH	sulfonimide	no	16.69	4.34	7.40	5.79
4-Cl-C_6H_4SO(=NTf)NH$SO_2C_6H_4$-4-NO_2	174788-97-7	NH	sulfonimide	no	13.51	3.62	4.93	3.52
4-Cl-3-NO_2-$C_6H_3SO_2$NHTos	215395-06-5	NH	sulfonimide	no	24.02	9.71	13.11	11.04
TosNHTf	215395-07-6	NH	sulfonimide	no	18.94	6.17	9.16	7.40
$(C_6H_5SO_2)_2$NH	2618-96-4	NH	sulfonimide	no	26.40	11.34	14.96	12.74
$(4\text{-Cl-}C_6H_4SO_2)_2$NH	2725-55-5	NH	sulfonimide	no	24.80	10.20	13.72	11.60
Tos_2NH	3695-00-9	NH	sulfonimide	no	27.22	11.97	15.60	13.32
$(4\text{-}NO_2\text{-}C_6H_4SO_2)_2$NH	4009-06-7	NH	sulfonimide	no	21.46	8.19	11.11	9.20
4-MeO-C_6H_4C(=NTf)NHTf	500721-87-9	NH	sulfonimide	no	17.75	6.41	8.23	6.55
4-Me-C_6H_4C(=NTf)NHTf	500721-89-1	NH	sulfonimide	no	18.79	6.19	9.04	7.30
C_6H_5C(=NTf)NHTf	500721-91-5	NH	sulfonimide	no	18.25	6.04	8.62	6.91
4-F-C_6H_4C(=NTf)NHTf	500721-93-7	NH	sulfonimide	no	18.14	5.66	8.54	6.83
4-Cl-C_6H_4C(=NTf)NHTf	500721-95-9	NH	sulfonimide	no	17.24	5.56	7.84	6.19
4-NO_2-C_6H_4C(=NTf)NHTf	500721-97-1	NH	sulfonimide	no	16.88	5.13	7.55	5.93
4-Cl-$C_6H_4SO_2$NHTos	69173-28-0	NH	sulfonimide	no	25.77	11.10	14.47	12.29
4-NO_2-$C_6H_4SO_2$NH$SO_2C_6H_4$-4-Cl	95468-16-9	NH	sulfonimide	no	23.02	9.17	12.33	10.32
$(4\text{-}NC_5F_4)(C_6H_5)$NH	39077-43-5	NH	amine(sec)	no	42.88	26.34	27.79	24.53
$(4\text{-}Me_2N\text{-}C_6F_4)(C_6F_5)$NH	80588-34-7	NH	amine(sec)	no	41.52	25.12	26.73	23.56
$(4\text{-Me-}C_6F_4)(C_6F_5)$NH	80588-36-9	NH	amine(sec)	no	40.77	24.94	26.15	23.02
2,4,6-$(SO_2F)_3$-Aniline	133213-11-3	NH	aniline	no	34.85	19.66	21.54	18.79

[a] ΔG_{diss}: Gibbs free energies of dissociation calculated from eq. 3 in kcal·mol^{-1}; pK_a^{Exp}: experimental pK_a value in acetonitrile, taken from Ref. 7; pK_a^{Calc}: pK_a value calculated by Eq. 4; p$K_a^{Calc}{}_{(corr)}$: pK_a value calculated by Eq. 7

[b] Tf denotes CF_3-SO_2- Tos denotes 4-Me-C_6H_4-SO_2-

[c] Formal notation, see text.

[d] Tautomeric equilibrium, see text

Correlation of the complete fit data set results in a correlation coefficient of $r^2 = 0.857$. The regression equation for acids pK_a in solvent acetonitrile reads

$$\mathrm{p}K_a = 1.06(\pm 0.01)\frac{\Delta G_{diss}}{RT\ln(10)} - 5.6(\pm 0.1) \qquad (4)$$

The calculated axis intercept of –5.6 is in reasonable concordance with the theoretical value of $c_{2,ideal}$ = -log[CH_3CN] = -1.28. If we would have omitted the free energy difference of CH_3CNH^+ and CH_3CN,

which we calculate as -253.48 kcal·mol^{-1}, in the definition of ΔG_{diss} we would have received a regression constant of $\hat{c}_2$ = 191.6. In contrast to previous findings on aqueous acidity[10] and basicity[24] and dimethylsulfoxide acidity[10], we found that the slope of the regression is close to the theoretical value of 1/RTln(10). Application of eq. 4 to predict the pK_a values of the fit set yields a rmse of 2.53 pK_a units.

A closer look at the regression Table 1 and Figure 1 reveals that there are systematic deviations: the regression splits into two distinct groups with slightly different slopes (which both are close to the theoretical slope) and significantly different axis intercepts. This is an interesting behavior, which is not observed in the COSMO-RS-ΔG_{diss} vs. pK_a correlations for solvent water (neither for acids[10] nor for bases[24]) and acids in nonaqueous solvent dimethylsulfoxide[10].

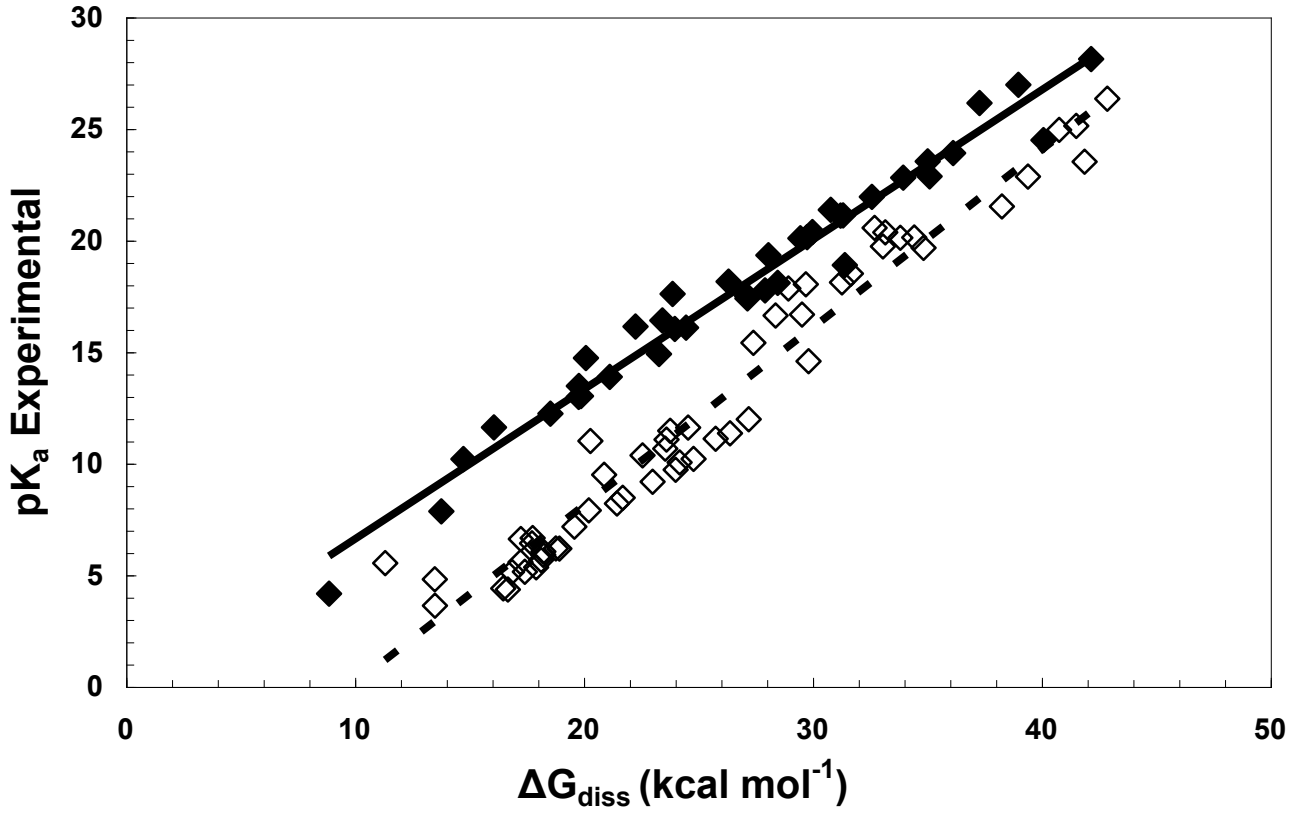

FIGURE 1: Fit data set. Calculated Gibbs free energy of dissociation vs. experimental acids pK_a in solvent acetonitrile. Filled rhombus: acids yielding charge-delocalized anions on dissociation. Solid line: regression line (eq. 5) for delocalized anion acids (r^2 = 0.971, c_1=0.91, c_2 = -0.1). Open rhombus: acids where the charge of the anion remains localized on the deprotonated atom or group. Dotted line: regression line (eq. 6) for delocalized anion acids (r^2 = 0.958, c_1=1.08, c_2 = -7.8).

Analysis of the electronic structure of the molecules involved suggests the presence of two groups of acids. The classification of the compounds to these groups is correlated with the level of charge delocalization in the anions. The anions with localized charges have strong interactions with solvent molecules, which results in strong influence of solvation on the pK_a values. This influence is not fully taken into account by the calculations. At the same time the acids (especially CH acids) that yield anions with delocalized charges are less affected by solvation and their acidities are better predicted.

If one compares the *ab initio* pK_a values of the CH acids (calculated directly from eq. 1 using theoretical values of the *c* coefficients $c_{1,ideal} = 1$ and $c_{2,ideal} = -\log[CH_3CN] = -1.28$) to the experimental p$K_a$ values then it can be seen that the agreement is very good. Only the two acids that give the most charge-localized anions (Octafluorofluorene and $(C_6F_5)CH(COOEt)_2$) deviate by more than 2 pK_a units. The rmse is 1.12 pK_a units. If we exclude these two acids then we arrive at rmse = 0.86 pK_a units, which is excellent, keeping in mind that the pK_a value are not adjusted in any way!

All the acids dissociating from a carbon atom (CH acids) included in the dataset derive their acidity from an extensive charge delocalization that stabilizes the anion. The anionic centre is conjugated to one or more aromatic systems and those are substituted by electronegative (in most cases heavily: perfluorinated) or resonance acceptor groups. All CH acids with the exception of octafluorofluorene can be regarded as trisubstituted methanes. Octafluorofluorene can be regarded as a disubstituted methane and it is the most deviating point of the CH acid cloud. It is important to note that $(C_6F_5)CH(COOEt)_2$, although formally a CH acid, is able to form a tautomeric structure, which has a planar central carbon atom and is protonated on one of the carbonyl oxygen atoms of the ester groups, thus being an OH acid. In addition, the proton is strongly chelated by the oxygen atom of the second carbonyl group, resulting in a stable 6-membered cycle. If this tautomeric equilibrium is taken into account in the computation of ΔG_{diss} by means of pseudo conformer equilibrium of the tautomers in COSMO-RS, the regression of this compound neatly falls into the CH acids group. Due to the highly delocalized charge in the anions and

thus low sensitivity to moisture and other ions in the solution (and also very suitable spectral properties) we rate the pK_a values of CH acids as the most reliable of the three acid groups in the fit data set.

The acids dissociating from an oxygen atom (OH acids) have to be considered at greater detail. Most of them are phenols that are heavily substituted by electronegative and electron acceptor substituents (the least substituted one is 2,4,6-tribromophenol). Conjugation of the OH center with the aromatic system provides possibility for delocalization of the charge, although not nearly as efficient as in the CH acids group, due to the higher electronegativity of oxygen compared to carbon and due to the fact that just one substituent is attached to the oxygen compared to three substituents attached to carbon atom in the CH acids. For some of the phenols the possibility for delocalization of the charge in the anion is even further diminished by the steric hindrance of bulky electron-acceptor groups such as nitro (NO_2) or (to a lesser extent) trifluoromethanesulfonyl (Tf), which try to avoid contact and are bent out of the ring plane and thus fail to conjugate efficiently with the $–O^–$ (deprotonated OH) center. Therefore phenols form a distinct second cloud on the figure, lower than the CH acids cloud. There are 8 OH acids in the set that form anions with a localized charge. These are all 5 sulfonic acids, two carboxylic acids (benzoic acid and acetic acid) and perfluoro-*tert*-butyl alcohol. All sulfonic acids included here and benzoic acid do have an aromatic system. But these aromatic systems are not conjugated with the OH acidity centre, but are separated by an SO_2 or a CO fragment. Acetic acid and perfluoro-*tert*-butyl alcohol do not have an aromatic system. Consequently, all of these 8 acids are distinctly separated from the "delocalized" CH acids in the ΔG_{diss} vs. pK_a plot and are also slightly lower than the cloud of substituted phenols Figure 1.

The acids dissociating from a nitrogen atom (NH acids) all are sulfonimides or carbonylsulfonamides, except four of them being aromatic amines. All the amides and imides are quite similar to sulfonic acids in that the charges in the anion are rather localized (although somewhat more delocalized than in sulfonic acids). Due to this it is not surprising that these acids form a joint group with sulfonic and carboxylic acids. The four aromatic amines have one or two substituted aromatic rings connected to the NH acidity center. These aromatic amines are a borderline case between the CH acids and OH acids with localized-charge anions: the charge delocalization is similar to that of phenols. Thus it is not

surprising that in Figure 1 they do not fit visually into the group of “delocalized” CH acids, and just like the phenols they do not fully fall into the group of the strictly “localized” acids like carboxylic or sulfonic acids.

Based on the above considerations and in order to avoid too extensive splitting of the data set and considering that phenols and aromatic amines do not deviate strongly from the rest of the OH and NH acids we split it in two: CH acids giving anions with highly delocalized charges and all other acids that have less extensive delocalization of charge in their anions. The assignment of the compounds to these groups, formally called as "delocalized" and "localized" is given in the fifth column of Table 1. The regression of the experimental acetonitrile acid pK_a values with the calculated values of ΔG_{diss} was repeated independently for the two compound families.

There are 38 compounds in the fit data set that allow for delocalization of the charge over the molecule structure in their anionic form, all of which are CH acids (with the exception of compound $(C_6F_5)CH(COOEt)_2$, as explained above). The pK_a vs. ΔG_{diss} regression of this compound family results in a correlation coefficient of $r^2 = 0.971$. The regression equation for the pK_a of acids forming charge-delocalized anions in solvent acetonitrile reads

$$pK_a^{\text{delocalized}} = 0.91(\pm 0.01)\frac{\Delta G_{diss}}{RT\ln(10)} - 0.1(\pm 0.1) \qquad (5)$$

The calculated axis intercept of –0.1 is in reasonable concordance with the theoretical value of $c_{2,ideal}$ = -log[CH_3CN] = -1.28. If we would have omitted the free energy difference of CH_3CNH^+ and CH_3CN, which we calculate as -253.48 kcal·mol^{-1}, in the definition of ΔG_{diss} we would have received a regression constant of $\hat{c}_2$ = 169.9. Application of eq. 5 to predict the pK_a of the family of “delocalized anion” compounds in the fit set yields a rmse of 0.91 pK_a units. Only two acids octafluorofluorene and 9-C_6F_5-octafluorofluorene deviate by more than 2 pK_a units and these are the acids that happen to have to strongest charge-localization in their anions. If these two acids are excluded from the regression we arrive at rmse = 0.74 pK_a units, which is excellent, keeping in mind that the pK_a value are not adjusted in

any way. Excluding these two compounds, the regression results in coefficients c_1 = 0.94, and c_2 = -0.53 with r^2 = 0.981. Both the c coefficients are now in better agreement with their theoretical values.

In the remaining 55 compounds of the fit data set the anionic charge can not be delocalized over the anion's structure. The pK_a vs. ΔG_{diss} regression of this compound family results in a correlation coefficient of r^2 = 0.958. The regression equation for the pK_a of acids forming charge-localized anions in solvent acetonitrile reads

$$\mathrm{p}K_a^{\mathrm{localized}} = 1.08(\pm 0.01)\frac{\Delta G_{diss}}{RT\ln(10)} - 7.8(\pm 0.1) \qquad (6)$$

Considering the typical accuracy of the underlying DFT method, the calculated axis intercept of –7.8 is still in reasonable concordance with the theoretical value. If we would have omitted the free energy difference of CH_3CNH^+ and CH_3CN, which we calculate as -253.48 kcal·mol^{-1}, in the definition of ΔG_{diss} we would have received a regression constant of $\hat{c}_2$ = 194.2. Application of eq. 6 to predict the pK_a values of the family of "localized anion" compounds in the fit set yields a rmse of 1.38 pK_a units. Six compounds deviate by more than 2 pK_a-units from the regression line. Four of them are phenols with a large number of strongly electronegative substituents. As discussed above, the anionic charge of these compounds must be considered as partly delocalized and thus they show a systematic deviation from the regression line. The remaining outliers are acetic acid and $(CF_3)_3COH$. Excluding these six compounds, the regression results in coefficients c_1 = 1.10, and c_2 = -8.91 with r^2 = 0.976 and we arrive at rmse = 0.99 pK_a units.

These results are at least in part the reason why in the earlier work[3] on acid pK_a values in water and DMSO no splitting of the regression was observed: the dataset of ref. 10 did not include CH acids. The second reason may be that water is capable of solvating anions with very high efficiency, so that some effects that are visible in acetonitrile can be masked in water. This effect is also present, although less

pronounced, in dimethylsulfoxide, which also has considerably stronger anion-solvating abilities than acetonitrile.

A major goal of this report is to provide a simple and practical prediction methodology for pK_a values of acids in acetonitrile. The systematic deviations observed above lead us to the conclusion that a simple heuristic correction to ΔG_{diss} that accounts for the different behaviour of acids giving anions of different level of charge delocalization, should lead to an improved correlation as well as to a simple and practical LFER method in eq. 1. From the separate regressions of "delocalized" and "localized" anion acids in eq. 5 and 6, it can be concluded that the significant difference is the axis intercept of the regression, not the slope. Thus the addition of a simple shift value will be sufficient. If ΔG_{diss} for compounds with localized anions is corrected by a value of -7.5 kcal·mol^{-1}, while the "delocalized" anion compounds remain untouched, the linear regression for the experimental acetonitrile pK_a with the corrected calculated values of ΔG_{diss} results in a correlation coefficient of r^2 = 0.957. The regression equation with the thus corrected ΔG_{diss} reads:

$$\mathrm{p}K_a = 0.92(\pm 0.01)\frac{\Delta G_{diss}}{RT\ln(10)} - 0.1(\pm 0.1) \qquad (7)$$

Application of eq. 7 to predict the pK_a of the complete fit data set yields a rmse of 1.38 pK_a units. The calculated results are listed in the ninth column of Table 1. The strong outliers of the prediction with eq. 7 are the same as for the separate fits above. If they are removed from the fit set the regression results in coefficients c_1 = 0.95, and c_2 = -0.53 with r^2 = 0.970 and we arrive at rmse = 1.13 pK_a units. It is interesting to note that many of the "borderline" compounds (with respect to charge delocalization in anions) give a better fit with eq 4 than with eq 7.

Test Data Set

To be able to get an independent test of the pK_a prediction methods deployed, literature data for 129 compound acidities in acetonitrile was collected. The bulk of the test set data was taken from the review book of Izutsu[6]. From the 102 acid solutes in Izutsu's collection 100 were used in the test data set. Two compounds from Izutsu's set (acetic acid and benzoic acid) were used in the fit data set already and thus excluded from the test set. The remaining 29 test data pK_a values were taken from publications 44-49. The pK_a values of the test set range between 1 and 29. Of the 129 compounds 33 are considered to have charge-delocalization in their anionic state, while the remaining 96 compounds have charge-localized anions.

Two aspects are important in characterizing the test set:

1. The majority of the data in the test set are for carboxylic acids while there are only two compounds of this type in the training set. Moreover, there are almost no pK_a values of CH acids (which we considered as most reliable in the fit date set) in the literature. Thus the test set is broader and quite different, qualitatively, from the fit set.
2. The quality and especially the consistency of the data in the test set is not as high as in the fit set. In particular, it was demonstrated in ref. 7 that starting from pK_a value of ca 16 there is a marked contraction of the virtual scale formed by the literature values compared to the values of ref. 7. The reasons for that were analyzed. As one consequence of this the experimental pK_a values of acetic and benzoic acids in Izutsu's collection are by 1.2 and 0.8 pK_a units lower than those reported in ref. 7, which was used in the fit data set. As noted in ref. 7 there are good reasons to prefer the values given there over Izutsu's data.

It thus should be a challenging trial for the predictive qualities of the methodology developed in the previous section. Formal assignment of the acids as "delocalized" and "localized" was done on the basis of the considerations outlined above. A special case is formed by dicarboxylic acids forming stable intramolecular hydrogen bonds in their monoanions. In these anions there is efficient delocalization of charge across the formed cyclic structure and these were assigned into the group of acids with charge-

delocalized anions. In addition, thiophenols, perchloric and fluorosulfuric acid were also assigned to the same group based on the analysis of the anions COSMO surfaces. The prediction results for all 129 acids in the test data set are listed in Table 2 and depicted in Figure 2.

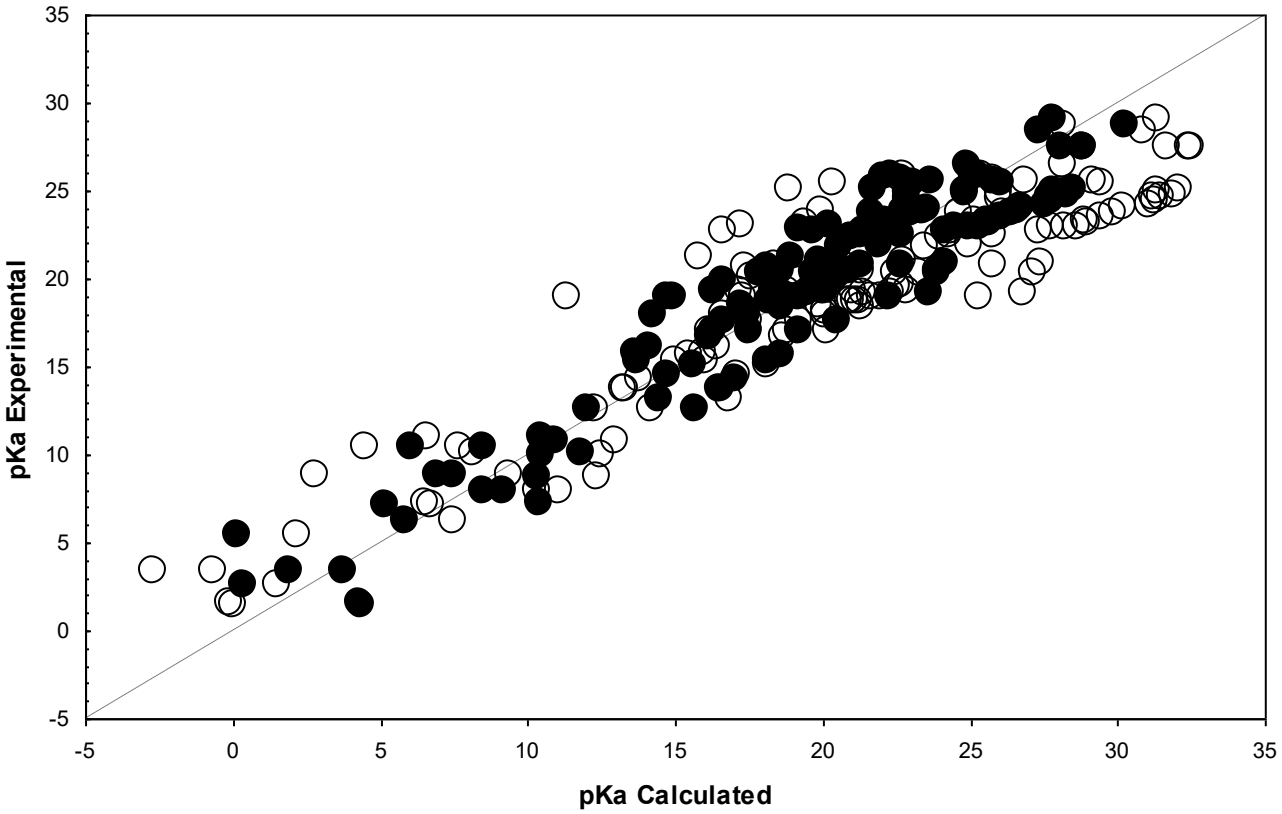

FIGURE 2: Test data set. Calculated vs. experimental acids pK_a in solvent acetonitrile. Open circle: pK_a calculated by Eq. 4. Filled circle: pK_a calculated by Eq. 7.

TABLE 2: Test data set for COSMO-RS acid pK_a calculations in acetonitrile.[a]

Compound	Type	Class	Delocalized	ΔG_{diss}	pK_a^{Exp}	Ref.	pK_a^{Calc}	p$K_a^{Calc}{}_{(corr)}$
trinitromethane	CH	aliphatic	yes	15.59	7.3	7	6.55	10.37
1-(4-nitrophenyl)-1-nitropropane	CH	aliphatic	yes	32.85	23.9	44	19.98	22.72
1-(4-nitrophenyl)-2-methyl-1-nitropropane	CH	aliphatic	yes	36.34	25.9	44	22.70	25.22
1,2-dicyanocyclopentadiene	CH	cyclopentadiene	yes	17.63	10.17	45	8.14	11.83
1,2,3-tricyanocyclopentadiene	CH	cyclopentadiene	yes	7.22	1.44	45	0.04	4.38
1,2,4-tricyano-3-methylcyclopentadiene	CH	cyclopentadiene	yes	6.30	3.4	45	-0.68	3.73
9-cyanofluorene	CH	fluorene	yes	30.84	20.8	44	18.42	21.28
pentakis(trifluoromethyl)-phenylmalonitrile	CH	methane	yes	10.79	8.86	29	2.82	6.94
2,3,4,6-tetrakis(trifluoromethyl)-phenylmalonitrile	CH	methane	yes	12.98	10.45	29	4.52	8.50
pentakis(trifluoromethyl)-toluene	CH	methane	yes	43.37	28.7	29	28.17	30.25
pentacyanotoluene	CH	methane	yes	29.74	20.14	46	17.56	20.50
2,4,6-trinitrotoluene	CH	methane	yes	32.15	23.2	44	19.44	22.22
4-nitrophenylacetonitrile	CH	methane	yes	33.35	25.4	44	20.37	23.07
4-nitrophenylphenylacetonitrile	CH	methane	yes	28.57	22.7	44	16.65	19.66
4-bromophenyl-4-nitrophenylacetonitrile	CH	methane	yes	27.52	21.3	44	15.83	18.91
4-nitrophenyl-4-methoxyphenylacetonitrile	CH	methane	yes	29.35	23.1	44	17.26	20.22
bis(4-nitrophenyl)acetonitrile	CH	methane	yes	21.70	19	44	11.31	14.74
bis(4-nitrophenyl)ethylacetate	CH	methane	yes	31.40	25.1	44	18.85	21.68
4-nitrophenylnitromethane	CH	methane	yes	29.51	20.7	44	17.38	20.33
phthalicacid	OH	carboxylic acid	yes	24.86	14.3	7	13.77	17.01
2,2-diphenicacid	OH	carboxylic acid	yes	27.08	15.7	7	15.49	18.59
3-methylphthalicacid	OH	carboxylic acid	yes	27.93	17	7	16.15	19.20
propanedioicacid	OH	carboxylic acid	yes	26.42	15.3	7	14.98	18.12
succinicacid	OH	carboxylic acid	yes	29.70	17.6	7	17.53	20.47
tetrahydroxysuccinicacid	OH	carboxylic acid	yes	24.26	13.7	7	13.30	16.58
tartronicacid	OH	carboxylic acid	yes	24.19	13.8	7	13.24	16.52
2,6-dihydroxybenzoicacid	OH	carboxylic acid	yes	23.00	12.6	7	12.32	15.67
perchloricacid	OH	perchloric acid	yes	7.05	1.57	7	-0.10	4.26
fluorosulfuricacid	OH	sulfuric acid	yes	3.73	3.38	7	-2.68	1.89
2,4,6-trinitrothiophenol	SH	thiophenol	yes	15.67	11	7	6.61	10.43
o-mercaptophenol	SH	thiophenol	yes	29.20	19.34	7	17.14	20.11
1,1,3,3-tetranitrobutane	CH	aliphatic	no	20.42	8	7	10.31	8.46
pentakis(trifluoromethyl)-aniline	NH	aniline	no	40.54	24.59	29	25.96	22.85
1,3,7,9-tetranitrophenoxazine	NH	phenoxazine	no	34.15	18.8	7	20.99	18.29
1,3,7-trinitrophenoxazine	NH	phenoxazine	no	33.52	20.3	7	20.50	17.84
1,3,9-trinitrophenoxazine	NH	phenoxazine	no	39.23	21.9	7	24.95	21.92
1,3-dinitrophenoxazine	NH	phenoxazine	no	37.92	22.4	7	23.93	20.98
1,7-dimethyl-3-nitrophenoxazine	NH	phenoxazine	no	39.79	25.9	7	25.38	22.32
1-methyl-3-nitrophenoxazine	NH	phenoxazine	no	39.45	25.8	7	25.12	22.08
1-nitrophenoxazine	NH	phenoxazine	no	46.83	28.4	7	30.86	27.36
3,7-dinitrophenoxazine	NH	phenoxazine	no	35.46	22.8	7	22.01	19.22
3-nitrophenoxazine	NH	phenoxazine	no	40.29	25.7	7	25.77	22.68
3,4-dichlorobenzenesulfonamide	NH	sulfonamide	no	44.35	23.29	7	28.93	25.58
4-methylbenzenesulfonamide	NH	sulfonamide	no	48.12	24.82	7	31.86	28.28
benzenesulfonamide	NH	sulfonamide	no	47.29	24.61	7	31.21	27.68
m-chlorobenzenesulfonamide	NH	sulfonamide	no	45.49	23.8	7	29.82	26.40
m-cyanobenzenesulfonamide	NH	sulfonamide	no	44.42	23.23	7	28.98	25.63
m-methoxybenzenesulfonamide	NH	sulfonamide	no	47.35	24.48	7	31.26	27.73
m-nitrobenzenesulfonamide	NH	sulfonamide	no	43.97	22.95	7	28.63	25.31
m-toluenesulfonamide	NH	sulfonamide	no	47.59	24.67	7	31.45	27.90
m-trifluoromethylbenzenesulfonamide	NH	sulfonamide	no	44.97	23.53	7	29.41	26.02
o-xylene-4-sulfonamide	NH	sulfonamide	no	47.46	25.01	7	31.35	27.81
p-brombenzenesulfonamide	NH	sulfonamide	no	45.99	24.04	7	30.21	26.76
p-fluorobenzenesulfonamide	NH	sulfonamide	no	47.11	24.19	7	31.08	27.56
p-methoxybenzenesulfonamide	NH	sulfonamide	no	48.42	25.09	7	32.10	28.49
p-nitrobenzenesulfonamide	NH	sulfonamide	no	43.47	22.91	7	28.25	24.95
4-bromobenzoicacid	OH	carboxylic acid	no	37.07	20.3	7	23.27	20.37
4-hydroxybenzoicacid	OH	carboxylic acid	no	40.29	20.8	7	25.77	22.68

Compound	Type	Class	Delocalized	ΔG_{diss}	pK_a^{Exp}	Ref.	pK_a^{Calc}	$pK_a^{Calc\ (corr)}$
chloroaceticacid	OH	carboxylic acid	no	34.01	18.8	7	20.88	18.18
cyanoaceticacid	OH	carboxylic acid	no	32.99	18	7	20.09	17.45
dichloroaceticacid	OH	carboxylic acid	no	28.78	13.2	7	16.82	14.44
fumaricacid	OH	carboxylic acid	no	35.85	19.2	7	22.31	19.50
oxalicacid	OH	carboxylic acid	no	29.18	14.5	7	17.12	14.73
salicylicacid	OH	carboxylic acid	no	31.13	16.7	7	18.64	16.13
trichloroaceticacid	OH	carboxylic acid	no	23.86	10.75	7	12.99	10.93
trifluoroaceticacid	OH	carboxylic acid	no	25.37	12.65	7	14.16	12.00
1,3-benzenedicarboxylicacid	OH	carboxylic acid	no	36.59	19.3	7	22.89	20.03
1,4-benzenedicarboxylicacid	OH	carboxylic acid	no	36.29	19.7	7	22.66	19.81
1,8-naphthalicacid	OH	carboxylic acid	no	37.36	21.8	7	23.49	20.58
2,3-dibromopropionicacid	OH	carboxylic acid	no	33.09	17.1	7	20.17	17.53
2,4,6-trimethylbenzoicacid	OH	carboxylic acid	no	37.85	20.5	7	23.87	20.93
2,4-dichlorobenzoicacid	OH	carboxylic acid	no	34.54	18.4	7	21.30	18.56
2,4-dinitrobenzoicacid	OH	carboxylic acid	no	28.30	16.1	7	16.44	14.10
2,5-dichlorobenzenesulfonicacid	OH	carboxylic acid	no	16.80	6.2	7	7.49	5.87
2,6-dichlorobenzoicacid	OH	carboxylic acid	no	31.88	17.6	7	19.23	16.66
2,6-dinitrobenzoicacid	OH	carboxylic acid	no	27.69	15.8	7	15.96	13.66
2-chloro-benzoicacid	OH	carboxylic acid	no	34.97	19	7	21.63	18.87
3,4-dichlorobenzoicacid	OH	carboxylic acid	no	35.42	19	7	21.98	19.19
3,4-dimethylbenzoicacid	OH	carboxylic acid	no	39.69	19	7	25.30	22.25
3,5-dichlorobenzoicacid	OH	carboxylic acid	no	34.24	18.7	7	21.07	18.35
3,5-dinitrobenzoicacid	OH	carboxylic acid	no	31.33	17	7	18.79	16.26
3-bromobenzoicacid	OH	carboxylic acid	no	36.14	19.5	7	22.54	19.71
3-nitrobenzoicacid	OH	carboxylic acid	no	34.62	19.2	7	21.36	18.62
4-chloro-3-nitrobenzoicacid	OH	carboxylic acid	no	32.68	18.5	7	19.85	17.24
4-dimethylaminobenzoicacid	OH	carboxylic acid	no	42.81	23	7	27.73	24.48
4-nitrobenzoicacid	OH	carboxylic acid	no	34.49	18.7	7	21.26	18.53
butyricacid	OH	carboxylic acid	no	42.30	22.7	7	27.34	24.12
hexanedioicacid	OH	carboxylic acid	no	42.01	20.3	7	27.11	23.91
hydracrylicacid	OH	carboxylic acid	no	36.41	21	7	22.75	19.90
hydroxy-aceticacid	OH	carboxylic acid	no	34.24	19.3	7	21.06	18.35
nonanedioicacid	OH	carboxylic acid	no	42.38	20.9	7	27.40	24.17
o-nitrobenzoicacid	OH	carboxylic acid	no	32.97	18.2	7	20.08	17.44
pentanedioicacid	OH	carboxylic acid	no	41.64	19.2	7	26.82	23.64
tartaricacid	OH	carboxylic acid	no	30.44	15.1	7	18.11	15.63
nitricacid	OH	nitric acid	no	23.02	8.8	7	12.33	10.32
2-bromophenol	OH	phenol	no	41.33	23.92	7	26.58	23.42
3,4,5-trichlorophenol	OH	phenol	no	38.37	22.5	7	24.28	21.31
3,4-dichlorophenol	OH	phenol	no	41.46	24	7	26.68	23.51
3,5-dichlorophenol	OH	phenol	no	39.53	23.3	7	25.18	22.13
3-chlorophenol	OH	phenol	no	43.27	25	7	28.09	24.81
4-bromophenol	OH	phenol	no	44.63	25.53	7	29.15	25.79
4-chlorophenol	OH	phenol	no	45.04	25.44	7	29.47	26.07
4-nitrophenol	OH	phenol	no	33.95	20.7	7	20.83	18.14
p-cresole	OH	phenol	no	48.85	27.45	7	32.43	28.80
phenol	OH	phenol	no	47.49	29.14	29	31.37	27.83
2-methylphenol	OH	phenol	no	47.88	27.5	7	31.67	28.11
3,4-dinitrophenol	OH	phenol	no	28.51	17.9	7	16.60	14.25
3-chloro-4-nitrophenol	OH	phenol	no	31.87	19.9	7	19.22	16.65
3-nitrophenol	OH	phenol	no	40.75	23.8	7	26.13	23.01
3-trifluoromethyl-4-nitrophenol	OH	phenol	no	31.31	19.3	7	18.78	16.26
4-chloro-2,6-dinitrophenol	OH	phenol	no	27.75	15.3	7	16.01	13.71
4-cyanophenol	OH	phenol	no	38.50	22.7	7	24.38	21.39
m-trifluoromethylphenol	OH	phenol	no	43.31	24.9	7	28.12	24.84
4-(1,1-dimethylethyl)-phenol	OH	phenol	no	48.90	27.48	7	32.47	28.84
3,5-dinitrophenol	OH	phenol	no	34.55	20.5	47	21.31	18.57
2,3,5,6-tetrafluoro-4-methylphenol	OH	phenol	no	36.08	20.3	48	22.49	19.66
2,4,6-trichlorophenol	OH	phenol	no	40.29	22.5	47	25.77	22.68
2-trifluoromethylphenol	OH	phenol	no	40.55	24.88	29	25.97	22.86
3-trifluoromethylphenol	OH	phenol	no	43.37	26.5	29	28.17	24.88

Compound	Type	Class	Delocalized	ΔG_{diss}	pK_a^{Exp}	Ref.	pK_a^{Calc}	$pK_a^{Calc}{}_{(corr)}$
4-trifluoromethylphenol	OH	phenol	no	41.69	25.54	29	26.86	23.68
3,5-bis(trifluoromethyl)-phenol	OH	phenol	no	38.83	23.78	29	24.63	21.63
2,6-bis(1,1-dimethylethyl)-4-nitrophenol	OH	phenol	no	29.44	19	47	17.33	14.92
pentakis(trifluoromethyl)-phenol	OH	phenol	no	17.00	10.46	29	7.65	6.02
4-methylbenzenesulfonic acid	OH	sulfonic acid	no	21.40	8.01	7	11.07	9.16
methanesulfonic acid	OH	sulfonic acid	no	23.19	9.97	7	12.47	10.45
trifluoromethanesulfonic acid	OH	sulfonic acid	no	9.13	2.6	7	1.52	0.38
H_2SO_4	OH	sulfuric acid	no	15.81	7.2	7	6.72	5.16
HBr	BrH	atom	no	10.01	5.5	7	2.20	0.14
HCl	ClH	atom	no	19.24	8.9	7	9.39	7.50

[a] ΔG_{diss}: Gibbs free energies of dissociation calculated from eq. 3 in kcal·mol^{-1}; pK_a^{Exp}: experimental pK_a value in acetonitrile; *Ref.*: experimental pK_a value data source reference; pK_a^{Calc}: pK_a value calculated by Eq. 4; $pK_a^{Calc}{}_{(corr)}$: pK_a value calculated by Eq. 7.

pK_a predictions using the LFER parameters of the uncorrected (raw) fit of equation 4 are given in column 8 of Table 2. The test data set is predicted with a mean signed error of -1.32 pK_a units, and a rmse of 3.63 pK_a units. If ΔG_{diss} values of acids giving charge-localized anions is corrected by a value of -7.5 kcal·mol^{-1} and the according LFER parameters of the "corrected" fit of equation 7 is used to predict the pK_a of the test data set, the mean signed error of this data set reduces to –0.04 pK_a units, and the rmse reduces to 2.10 pK_a units. Taking into account the uncertainties of the experimental data and the diversity of the data sources, this prediction quality is satisfactory.

Conclusions

A computational method for the computational quantum chemical prediction of the acidity of organic and inorganic acids in solvent acetonitrile has been deployed. Acetonitrile pK_a values of acids were predicted via a thermodynamic cycle, utilizing Gibbs free energies of dissociation in acetonitrile solution as computed by the COSMO-RS theory on the basis of quantum chemical DFT/COSMO calculations. Without any special adjustments of radii or other parameters this led to a prediction model for acid pK_a values in acetonitrile. In contrast to our findings on aqueous acidity predictions[10] the slope of the experimental pK_a versus theoretical ΔG_{diss} was found to match the theoretical value $1/RT\ln(10)$ well. No unique linear free energy relationship between the calculated Gibbs free energy and the experimental acids pK_a values was found. Instead, the linear free energy relationship splits into two major acid groups. The affiliation of acids to these families is based on the degree of localization of charge in the anion produced on acid dissociation. For acids with strongly delocalized charges in the anions both slope and axis intercept of the linear free energy relationship are very close to their theoretical value thus allowing for direct *ab initio* prediction without intermediate LFER correlation. The rmse of the acids with strongly delocalized charges in the anion predicted by the theoretical values for both slope and axis intercept of the linear free energy relationship is 1.12 pK_a units compared to 0.91 pK_a units that are achieved by fitting the LFER parameters. For acids with weakly delocalized or localized charges in the anion the slope of the linear free energy relationship also is very close to its theoretical value, but the axis intercept differs by about –7.5 kcal·mol^{-1}. For these compounds a LFER correlation based prediction is possible with good quality. From the given considerations it is possible to unify the prediction for both families of compounds into one practical prediction methodology, which applies a correction term for the free energy of dissociation.

The prediction of pK_a of acids and bases in solvents water[10,24] and dimethylsulfoxide[10] differs from the current findings in solvent acetonitrile in two aspects: first, no partitioning into groups was observed, and second, the slope of the pK_a vs. ΔG_{diss} regression was significantly lower than the theoretical value.

The first difference is at least in part caused by the absence of CH acids in the dataset of ref. 10. An additional reason is that water is capable of solvating anions with very high efficiency, so that some effects that are visible in acetonitrile can be masked in water. The same, although in a less pronounced way, holds for solvent dimethylsulfoxide, which also has considerably stronger anion-solvating abilities than acetonitrile. This suggests that both findings are, at least in part, related to the capability of the solvent to solvate and thus stabilize the anions, which is not captured sufficiently by the quantum mechanical method used. This is further corroborated by recent reports claiming that the addition of explicit solvent molecules to the continuum solvation model calculations of aqueous pK_a results in a slope of the pK_a vs. ΔG_{diss} regression, which is very close to the expected theoretical slope[31].

The results of this work also demonstrate that urgent and at present yet not satisfied need exists for reliable experimental data of physicochemical parameters in order to develop and validate computational approaches for their prediction.

Acknowledgments

This work was supported by the grants 7374 and 6699 from the Estonian Science Foundation.

References

1. Albert, A.; Serjeant, E. P. The Determination of Ionization Constants; Chapman and Hall: New York, 1984.

2. Coetzee, J. F. Prog. Phys. Org. Chem. 1967, 4, 45-92.

3. Kolthoff, I. M.; Chantooni, M. K. Jr. J. Phys. Chem. 1968, 72, 2270-2272.

4. (a) Kolthoff, I. M.; Bruckenstein, S.; Chantooni, M. K., Jr. J. Am. Chem. Soc. 1961, 83, 3927-3935. (b) Kolthoff, I. M.; Chantooni, M. K. Jr. J. Am. Chem. Soc. 1965, 87, 4428-4436. (c) Kolthoff, I. M.; Chantooni, M. K. Jr.; Bhowmik, S. J. Am. Chem. Soc. 1966, 88, 5430-5439.

5. Coetzee, J. F.; Padmanabhan, G. R. J. Phys. Chem. 1965, 69, 9, 3193-3196.

6. Izutsu, K. Acid-Base Dissociation Constants in Dipolar Aprotic Solvents; IUPAC Chemical Data Series; Blackwell Science Inc.: Oxford, UK, 1990.

7. Kütt, A.; Leito, I.; Kaljurand, I.; Sooväli, L.; Vlasov, V. M.; Yagupolskii, L. M.; Koppel, I. A. J. Org. Chem. 2006, 71, 2829.

8. Kaljurand, I; Kütt, A.; Sooväli, L.; Rodima, T.; Mäemets, V.; Leito, I.; Koppel, I. A. J. Org. Chem. 2005, 70, 1019-1028.

9. Cramer, C. J.; Truhlar, D. G. Chem. Rev. 1999, 99, 2161.

10. Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. J. Phys. Chem. A 2003, 107, 9380.

11. da Silva, G.; Kennedy, E. M.; Dlugogorski, B.Z.; J. Phys. Chem. A 2006, 110, 11371.

12. Gossens, C.; Dorcier, A.; Dyson, P. J.; Rothlisberger, U. Organometallics 2007, 26, 3969.

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

14. Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. J. Phys. Chem. A 2004, 108, 166.

15. Yu, A.; Liu, Y.; Li, Z.; Cheng, J.-P. J. Phys. Chem. A 2007, 111, 9978.

16. Mujika, J. I.; Mercero, J. M.; Lopez, X. J. Phys. Chem. A 2003, 107, 6099.

17. Lu, H.; Chen, X.; Zhan, C.-G., J. Phys. Chem. B 2007, 111, 10599.

18. Eckert, F.; Klamt, A. AIChE J. 2002, 48, 369.

19. A. Klamt, A.; Eckert, F. Fluid Phase Equilibria 2000, 172, 43.

20. Klamt, A: COSMO and COSMO-RS; P. v. R. Schleyer Ed.; Encyclopedia of Computational Chemistry; Wiley: New York, 1998.

21. (a) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, J. C. W. J. Phys. Chem. 1998, 102, 5074. (b) Klamt, A. J. Phys. Chem. 1995, 99, 2224.

22. Pliego, J. R., Jr. Chem. Phys. Letters 2003, 367, 145.

23. Chipman, D. M. J. Phys. Chem. A. 2002, 106, 7413.

24. Eckert, F.; A. Klamt, A. J. Comput. Chem. 2006, 27, 11.

25. Barone, V.; Cossi, M.; Tomasi, J. J. Chem. Phys. 1997, 107, 3210.

26. Westphal, E.; Pliego, J. R., Jr. J. Phys. Chem. A 2007, 111, 10068.

27. Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. J. Phys. Chem. A 2004, 108, 166.

28. Li, J.-N.; Fu, Y.; Liu, L.; Guo, Q.-X. Tetrahedron 2006, 62, 11801.

29. Kütt, A.; Movchun, V.; Rodima,T.; Dansauer, T.; Rusanov, E. B.; Leito, I.; Kaljurand, I.; Koppel, J.; Pihl, V.; Koppel, I.; Ovsjannikov, G.; Toom, L.; Mishima, M.; Medebielle, M.; Lork, E.; Röschenthaler,G.-V.; Koppel, I. A.; Kolomeitsev, A. A. J. Org. Chem. 2008, 73, 2607.

30. (a) Kovacevic, B.; Maksic, Z. B. Org. Lett. 2001, 3, 1523. (b) Kovacevic, B.; Maksic, Z. B. Chem. Eur. J. 2002, 8, 1694. (c) Kovacevic, B.; Glasovac, Z.; Maksic, Z. B. J. Phys. Org. Chem. 2002, 15, 765. (d) Kovacevic, B.; Maksic, Z. B. Tetrahedron Lett. 2006, 47, 2553. (e) Kovacevic, B.; Maksic, Z. B. Chem. Comm. 2006, 1524. (f) Despotovic, I.; Kovacevic, B.; Maksic, Z. B. New J. Chem. 2007, 31, 447.

31. Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. J. Phys. Chem. A 2006, 110, 2493.

32. Pliego, J. R., Jr.; Riveros, J. M. J. Phys. Chem. A, 2002, 106, 7434.

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

34. Westphal, E.; Pliego, J. R., Jr. J. Chem. Phys. 2005, 123, 74508.

35. Klamt, A.; Schüürmann, G. J. Chem. Soc. Perkins Trans.2 1993, 799 .

36. Eckert, F.; Klamt, A. COSMOtherm, Version C2.1, Revision 01.06; COSMOlogic GmbH&CoKG, Leverkusen, Germany, 2006; see also URL: http://www.cosmologic.de.

37. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Chem. Phys. Letters 1989, 162, 165.

38. Schäfer, A.; Klamt, A.; Sattel, D.; Lohrenz, J. C. W.; Eckert, F.; Phys. Chem. Chem. Phys. 2000, 2, 2187.

39. 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.; Hattig, 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 5.8, 2005.

40. Becke, A. D. Phys. Rev. A 1988, 38, 3098.

41. Perdew, J. P. Phys. Rev. B 1986, 33, 8822.

42. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Chem. Phys. Letters 1995, 242, 652.

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

44. Galezowski, W.; Stanczyk, M.; Jarczewski, A. Can. J. Chem. 1997, 75, 285.

45. Webster, O. W.; J. Am. Chem. Soc. 1966, 88, 3046.

46. Kütt, A. This work.

47. Chantooni, M. K.; Kolthoff, I. M. J. Phys. Chem. 1976, 80, 1306.

48. Vlasov, V. M.; Sheremet, O. P. Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim. 1982, 5, 114.