CROATICA CHEMICA ACTA
CCACAA, ISSN 0011-1643, e-ISSN 1334-417X
Croat. Chem. Acta **87** (4) (2014) 385–395.
http://dx.doi.org/10.5562/cca2472

Original Scientific Article

Basicities of Strong Bases in Water: A Computational Study†

Karl Kaupmees,[a] Aleksander Trummal,[b] and Ivo Leito[a,*]

[a]*University of Tartu, Institute of Chemistry, 14a Ravila str, 50411 Tartu, Estonia*
[b]*National Institute of Chemical Physics and Biophysics, 10 Rävala Blvd., Tallinn 10143, Estonia*

RECEIVED APRIL 29, 2014; REVISED AUGUST 13, 2014; ACCEPTED AUGUST 18, 2014

Abstract. Aqueous pK_a values of strong organic bases – DBU, TBD, MTBD, different phosphazene bases, *etc* – were computed with CPCM, SMD and COSMO-RS approaches. Explicit solvent molecules were not used. Direct computations and computations with reference pK_a values were used. The latter were of two types: (1) reliable experimental aqueous pK_a value of a reference base with structure similar to the investigated base or (2) reliable experimental pK_a value in acetonitrile of the investigated base itself.

The correlations of experimental and computational values demonstrate that direct computations do not yield pK_a predictions with useful accuracy: mean unsigned errors (MUE) of several pK_a units were observed. Computations with reference bases lead to MUE below 1 pK_a unit and are useful for predictions. Recommended aqueous pK_a values are proposed for all investigated bases taking into account all available information: experimental pK_a values in acetonitrile and water (if available), computational pK_a values, common chemical knowledge.

Keywords: superbases, quantum chemistry in solution, aqueous basicity, DBU

INTRODUCTION

Strong non-ionic bases are indispensable reagents in many chemical processes, most importantly organic synthesis.[1–9]

A core characteristic of a base B is its basicity, referring to the following equation

$$BH^+ + S \overset{K_a}{\rightleftharpoons} B + HS^+ \qquad (1)$$

and expressed as the pK_a value of its conjugate acid HB^+:[10]

$$pK_a = -\log \frac{a(HS^+) \cdot a(B)}{a(BH^+)} \qquad (2)$$

The pK_a values are different in different solvents. Out of all possible solvents used water is by far the most important and basicity data in water are important for several reasons. Firstly, many of these bases are also used in water. Secondly, water is the "champion-solvent" by the availability of pK_a data of medium strength bases. An as diverse as possible range of bases with available pK_a values in any one solvent is very useful for development of different prediction and computation methods, such as e.g. QSAR. The pK_a data of strong bases are currently scarce in water, so, additional data would be very welcome. Thirdly, for any base, especially the well-known ones, it is beneficial to know its basicity in the most important solvents and water certainly is one of those.

There are significant gaps in our knowledge concerning the basicity of strong and superstrong bases in water. Strong bases like phosphazenes, amidines, *etc* are difficult to study in aqueous solution due to low solubility of the nonpolar compounds and their very high basicity. Relatively high acidity (proton donicity) of water results in levelling effect. The bases with pK_a in water higher than ca 13, are all almost fully protonated in water, even if their basicities actually differ by orders of magnitude.

A correlation equation was proposed by Kaljurand *et al.*[11] for relating pK_a values measured in acetonitrile (MeCN) and water:

$$pK_a(AN) = 5.20 + 1.313\ pK_a(H_2O) \quad (n = 40,\ R^2 = 0.957,\ S = 0.98) \qquad (3)$$

that can be transferred to

$$pK_a(H_2O) = 0.762\ pK_a(AN) - 3.96 \qquad (4)$$

† Dedicated to Dr. Mirjana Eckert-Maksić on the occasion of her 70th birthday.
* Author to whom correspondence should be addressed. (E-mail: ivo.leito@ut.ee)

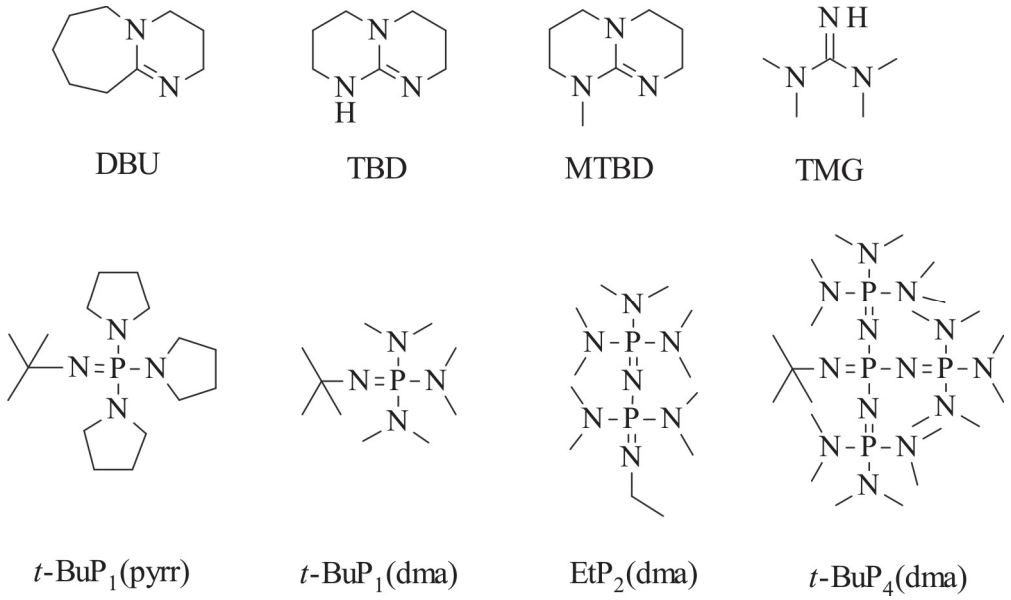


Figure 1. Structures of some of the investigated bases.

Equation 4 gives some insight into the basicities of strong bases in water. For example in the same paper[11] pK_a values of DBU and *t*-BuP$_1$(pyrr) in MeCN are measured as 24.34 and 28.42 respectively. Using equation 4 gives the corresponding pK_a values in water as 14.6 and 17.7. These estimated aqueous pK_a values can only be considered very approximate because of the quite high scatter of points around the regression line in Eq. 3 and, especially, because the highest aqueous pK_a value used in regression analysis was that of phenyltetramethylgua-nidine, leading to a strong extrapolation. Nevertheless, these estimates show that the superbasic region lies mainly around and above the basicity of the hydroxide ion (pK_a of H_2O in water can be calculated to be 15.74 using equations 1 and 2 as well as the autoprotolysis constant of water[12] $K_w = 10^{-14}$). Experimental measurements in that region are very difficult and need several approximations like measuring in solution with high concentrations of alkali hydroxide or in mixed solvents.[13] Another issue is the low solubility of many strong bases in water. Surfactants have been used for overcoming this problem,[14] altering somewhat the properties of the solvent. Given these difficulties it is unlikely that accurate aqueous pK_a values can be measured for strong bases unless a breakthrough in pK_a measurement methods is made.

In this paper we use computational methods as well as available experimental data for obtaining estimates of reasonable reliability for the aqueous pK_a values of a series of strong and very strong neutral organic bases. Among others the following are included: DBU, TBD, MTBD, TMG, *t*-BuP$_1$(pyrr), *t*-BuP$_1$(dma), EtP$_2$(dma), *t*-BuP$_4$(dma) (see Figure 1 for base structures).

Computational methods are free of the abovementioned problems: compounds with low solubility and high basicity can be studied. During the last decade continuum solvation models (CSM)[15] became an important tool for addressing the solvation phenomena, enabling researchers to establish Gibbs free energies of solvation and calculate pK_a values with reasonable accuracy. Just few examples: substituted phenols and carboxylic acids in water[16] using CPCM,[17] substituted phenols in dimethyl sulfoxide and acetonitrile[18] using IEF-PCM,[19] and various CH and NH superacids in 1,2-dichloroethane[23] using SMD,[24] as well as guanidine-based superbases in acetonitrile using IPCM[20] and amines in aqueous solution using SVPE,[21] PCM,[21] IEF-PCM,[21] CPCM,[22] SMD[22] and SM8[22] methods. The results indicate that the accuracy of CSM-based pK_a predictions is often in the range of 0.4−0.7 pK_a units, although sometimes worse accuracy is observed.[21] The results of Liptak and Shields[16] are especially encouraging. They demonstrated that even pure continuum approach could be a method of choice when modeling solvation in water, the difficult solvent that typically implies the so-called cluster – continuum representation[25] due to strong specific solvation and short range solute – solvent interactions.

Eckert *et al.*[26] applied the COSMO-RS procedure[27] combining polarized continuum theory with a statistical thermodynamics treatment to calculate pK_a values for the different classes of organic acids in acetonitrile. The method predicts pK_a values of substituted phenols in MeCN with the MUE of 0.8 pK_a units. Similar MUE was later found by Heldebrant *et al.* for carboxylic acids.[28] Klamt, *et al.*[29] used COSMO-RS to predict pK_a values of organic and inorganic acids in

water. The error of 0.5 pK_a units is reported to measure RMS deviation between pK_a estimates from linear regression and corresponding experimental values. For bases of low and medium strength in aqueous solution a RMS accuracy of 0.66 pK_a units was reported.[30]

For computing aqueous pK_a values a special care should be taken in choosing the appropriate thermodynamic cycle. In the recent report[31] Ho and Coote explored different pK_a calculation strategies and arrived at the conclusion that direct thermodynamic cycle involving deprotonation equilibrium is generally unsuitable for pK_a calculations in water. In contrast, the proton exchange scheme using an acid with established pK_a value as a reference yielded reasonably accurate results and, therefore, should be considered as a more viable alternative.

The purpose of the present study is a prediction of aqueous pK_a values for a number of strong guanidine and phosphazene bases using popular COSMO-RS, CPCM and SMD protocols and different thermodynamic cycles.

EXPERIMENTAL

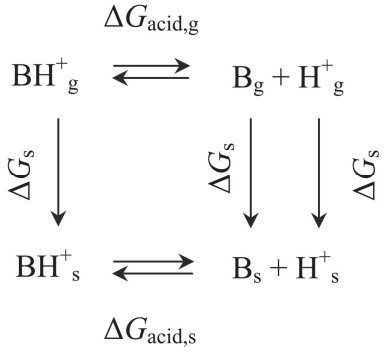
Computational Methods

pK_a computations with the COSMO-RS approach[32] were done similarly as in Ref. 33 using Turbomole[34] version 6.5 and COSMOthermX[35] version C30 with parametrization 1401. The two-step COSMO-RS computation protocol[32] was used. COSMO BP/TZVP geometry optimizations within RI approximation were carried out first in the conductor limit with Turbomole software package[34] for the studied base and corresponding conjugated acid. As the second step, for the resulting solvated structures COSMO-RS calculations were performed taking water as a real solvent and computing the deviations from ideal conductor by evaluating the differences in electrostatic and H-bonding energies according to the default procedure implemented in the COSMO*therm* software.[35] All stable conformers were taken into account and statistically weighted as is customary in the COSMO-RS procedure.

From the first step of the COSMO-RS protocol a σ-surface is obtained, that can be used to quantitatively describe the charge delocalization in ions.[36] In the case of cations the *Weighted Average Negative Sigma* (*WANS*)[36] parameter is used:

$$WANS = \frac{\int_{\sigma=-\infty}^{0} \sigma \cdot p(\sigma) d\sigma}{A \int_{\sigma=-\infty}^{0} p(\sigma) d\sigma} \quad (5)$$

where σ is the polarization charge density; $p(\sigma)$ the probability function of σ and A the surface area of the cation. The more extensive is charge delocalization in a cation the lower is its *WANS* value.



Scheme 1. Direct Thermodynamic Cycle for the CPCM/SMD calculations.

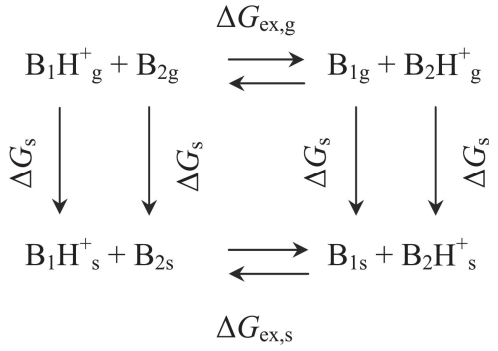
The CPCM[17] and SMD[24] calculations of pK_a values of bases B were based on the thermodynamic cycles presented in Scheme 1 and Scheme 2 involving the gas-phase acidities ($\Delta G_{acid,g}$) of BH^+, equal to the gas-phase basicities (GB) of B. The gas-phase basicity GB of the base B is defined as Gibbs free energy of deprotonation equilibrium of the conjugated acid BH^+.

To calculate absolute aqueous pK_a values from the direct thermodynamic cycle the following equation was applied together with the corresponding expansion for $\Delta G_{acid,s}$:

$$pK_a = \frac{\Delta G_{acid,s}}{RT\ln(10)} \quad (6)$$

$$\Delta G_{acid,s} = GB(B) + \Delta Gs(B) + \Delta Gs(H^+) - \Delta Gs(BH^+) + RT\ln(24.46) \quad (7)$$

The relative pK_a calculations are based on the proton exchange cycle presented in Scheme 2 with the follow-



Scheme 2. Proton Exchange Thermodynamic Cycle for the CPCM/SMD calculations.

ing expressions for pK_a and $\Delta G_{ex,s}$:

$$\mathrm{p}K_a(B_1) = \frac{\Delta G_{ex,s}}{RT\ln(10)} + \mathrm{p}K_a(B_2) \quad (8)$$

$$\begin{aligned}\Delta G_{ex,s} = {} & \mathrm{GB}(\mathrm{B_1}) + \Delta Gs(\mathrm{B_1}) - \Delta Gs(\mathrm{B_1H^+}) \\ & - \mathrm{GB}(\mathrm{B_2}) - \Delta Gs(\mathrm{B_2}) + \Delta Gs(\mathrm{B_2H^+})\end{aligned} \quad (9)$$

The absolute aqueous free energy of solvation of the proton with appropriate standard state correction $\Delta G_S(H^+)$ is based on the results of Tissandier *et al*.[37] and equals –265.9 kcal mol^{-1}. *RT* ln(24.46) reflects the change in the standard conditions from 1 atm to 1 mol L^{-1} and provides the necessary correction to GB values. The geometries were optimized both in solution (CPCM/HF/6-31G* with default cavities based on UFF radii and SMD/M05-2X/6-31G* with default cavities based on intrinsic atomic Coulomb radii) and in the gas phase using the same functional and basis set combination. The $\Delta G_S(B)$ and $\Delta G_S(BH^+)$ values are defined as the differences in SCF energy of the structure in solution and in the gas phase.[38] For the SMD calculations both electrostatic and non-electrostatic SCF energy terms were taken into account. The latter term represents cavity formation, dispersion interactions and the changes in solvent structure, and is usually denoted as CDS energy. When available, experimental GB values were used in this study for pK_a calculations. For the bases with unknown experimental gas-phase basicity, GB values were calculated at B3LYP/6-311G** level of theory. All geometry optimizations, both in the gas phase and in solution, were followed by frequency calculations to confirm the optimized structures to be the true minima on the potential energy surface. All thermal corrections were calculated for the standard state of 1 atm at 298.15 K.

All CPCM, SMD and GB calculations were carried out with the Gaussian09 software package.[39]

To supplement the computational methods, an alternate scheme was used to predict pK_a values in water based on reliable experimental pK_a values in MeCN, and Gibbs free energies of solvation of all the species both in water and MeCN.

$$\mathrm{p}K_a(\mathrm{H_2O}) = \mathrm{p}K_a(\mathrm{MeCN}) - Eff_{solv} + X_{corr} \quad (10)$$

where Eff_{solv} is the solvation effect between MeCN and H_2O in pK_a units:

$$Eff_{solv} = \frac{\Delta\Delta G_{solv}}{RT\ln(10)} = \frac{\Delta G_{solv}^{MeCN} - \Delta G_{solv}^{H_2O}}{RT\ln(10)} \quad (11)$$

The ΔG_{solv} in a given solvent is defined in the case of bases as

$$\Delta G_{solv} = G(\mathrm{H^+}) + G(\mathrm{neutral}) - G(\mathrm{cation}) \quad (12)$$

The solvation Gibbs free energies of neutrals and cations were calculated (see Supporting Information) using the SMD/M05-2X/6-31G* method. The *G* values of the proton were taken from experiments.[40,41] The X_{corr} in equation 10 is a correction term derived from the same calculation for PhTMG, for which reliable pK_a values are known both in MeCN and in water.

RESULTS AND DISCUSSION

Altogether 27 strong neutral bases were investigated, with base strength varying by 16 pK_a units. Table 1 presents the aqueous pK_a values computed with COSMO-RS, CPCM and SMD methods along with experimental aqueous and MeCN pK_a data as well as gas-phase basicities from literature where available. The pK_a values of some bases with reliable pK_a values available in MeCN were computed according to Eq. 10. Because of the difficulties with measurements mentioned above reliable experimental data in water can be found only for the less basic region of the investigated bases.

Correlation Analysis and Errors of Computational Methods

In Table 2 the data of linear regressions between all used computational methods and the experimental values are presented alongside with error analysis of the computational methods. When interpreting the correlation analysis data it is important to keep in mind that in the higher basicity region, e.g. if the pK_a of a base is higher than ca 12, the experimental values can also contain significant errors.

From the first section it is evident that if the dataset is not divided into compound groups none of the methods seems to reproduce the experimental values satisfactorily. COSMO-RS is by a narrow margin the best with R^2 = 0.74 and *Mean Unsigned Error* (MUE) of 1.04 pK_a units. It is evident from Figure 2, that the dataset seems to contain in broad terms two compound groups – phosphazenes and amidines/guanidines – which is also chemically and structurally reasoned. This reasoning is supported by the *WANS* values, which for phosphazenes are below 2.4 and for amidines/guanidines above 2.4. To put the *WANS* values into perspective, the *WANS* values for some common small cations are as follows: H_3O^+ 50.11; NH_4^+ 38.22; Li^+ 104.23; Na^+ 60.32; K^+ 25.93; guanidinium 12.94; trimethylammonium 7.85; triethylammonium 4.18; tetramethylammonium 6.13; tetraethylammonium 3.19.

Table 1. Computed and experimental pK_a values of strong neutral bases

Base	*WANS*	GB[d]	Experimental pK_a		pK_a (H_2O) (COSMO-RS)	pK_a (CPCM)				pK_a (SMD)				pK_a *	pK_a (H_2O) (recommended)[c]
			H_2O[a]	(MeCN)[a]		Direct	PhTMG As ref.	PhP1(pyrr) As ref	PhP1(dma) As ref	Direct	PhTMG As ref.	PhP1(pyrr) As ref	PhP1(dma) As ref		
t-BuP_4(dma)	0.63	287.5[h]		42.7[49]	28.3	31.7	33.1	30.8	30.6	31.9	30.4	28.4	28.1		30.0 ± 4.0
EtP_2(dma)	1.08	264.6[42]		32.94[49]	21.5	20.8	22.1	19.8	19.6	21.6[i]	20.2[i]	18.2[i]	17.8[i]		20.0 ± 2.0
PhP_3(dma)	0.83	273.6[h]		31.48	20.9	21.6	23.0	20.7	20.5	22.6	21.1	19.1	18.8		20.0 ± 2.0
t-BuP_1(pyrr)	1.25	258.7[42]		28.42	17.0	19.3	20.6	18.3	18.1	21.1	19.7	17.7	17.3		17.5 ± 1.0
t-BuP_1(dma)	1.71	252.9[42]		26.98	16.8	18.1	19.4	17.1	16.9	21.1	19.7	17.7	17.3		17.0 ± 1.0
PhP_2(dma)	1.08	261.7[42]		26.46	16.1	17.8	19.1	16.8	16.6	19.3	17.9	15.9	15.5		16.2 ± 1.0
PMG	3.00	242.6[43]	15.6[2]	25.00[50]	15.9	14.3	15.6	13.3	13.1	15.9	14.4	12.4	12.0		15.5 ± 1.5
2-Cl-$C_6H_4P_2$(dma)	1.06	260.5[42]			15.4	16.3	17.6	15.3	15.1	18.4	17.0	15.0	14.6		15.0 ± 1.0
MTBD	2.88	246.2[42]	13.0[44]	25.49	15.3	14.9	16.3	14.0	13.8	16.4	15.0	13.0	12.6	15.0[f]	15.0 ± 1.0
TBD	3.49	244.3[43]	14.5[45]	26.03	15.0	14.4	15.7	13.4	13.2	16.0	14.5	12.5	12.2	15.2[f]	15.2 ± 1.0
HP_1(pyrr)	1.68	255.2[42]	13.93[b]	27.01	14.65	16.4	17.7	15.4	15.2	19.4	17.9	15.9	15.6	15.8[g]	15.8 ± 1.0
HP_1(dma)	2.35	249.7[42]	13.32[b]	25.85	13.9	15.7	17.0	14.7	14.5	19.4	17.9	15.9	15.5	15.1[g]	15.1 ± 1.0
DBU	3.02	242.7[43]	11.5[46] 11.9[44]	24.34	14.2	13.6	14.9	12.6	12.4	15.4	14.0	12.0	11.6	14.2[f]	13.5 ± 1.5
TMG	3.72	238.4[43]	13.6[47]	23.37	13.8	11.7	13.0	10.7	10.5	14.2	12.7	10.7	10.3		13.0 ± 1.0
4-NMe_2-$C_6H_4P_1$(pyrr)	1.09	257.5[42]	12.00	23.88	13.4	14.6	15.9	13.6	13.4	16.9[i]	15.4[i]	13.4[i]	13.0[i]		12.0 ± 0.2
4-OMe-$C_6H_4P_1$(pyrr)	1.20	254.9[42]	11.94	23.12	12.7	13.6	14.9	12.6	12.4	16.7	15.3	13.3	12.9		11.9 ± 0.2
PhP_1(pyrr)	1.30	251.7[42]	11.52	22.34	12.4	12.5	13.8	11.5	11.3	15.0	13.5	11.5	11.2		11.5 ± 0.2
PhTMG	2.47	240.4[43]	11.77 12.18[48]	20.84	11.6	10.5	11.8	9.5	9.3	13.2	11.8	9.8	9.4		11.8 ± 0.2
4-Br-$C_6H_4P_1$(pyrr)	1.27	249.3[42]	11.23	21.19	11.4	12.6	13.9	11.6	11.4	15.4	14.0	12.0	11.6		11.2 ± 0.2
PhP_1(dma)	1.74	246.2[42]	10.64	21.25	11.4	11.8	13.1	10.8	10.6	14.5	13.0	11.0	10.6		10.6 ± 0.2
2-Cl-$C_6H_4P_1$(pyrr)	1.23	250.8[42]	9.98[e]	20.17	11.1	10.8	12.1	9.8	9.6	13.6[i]	12.1[i]	10.1[i]	9.8[i]		10.0 ± 0.2
2,6-Cl_2-$C_6H_3P_1$(pyrr)	1.25	245.3[42]	9.00[e]	18.56	10.9	8.3	9.6	7.3	7.1	12.0	10.5	8.5	8.2		9.0 ± 0.2
4-CF_3-$C_6H_4P_1$(pyrr)	1.23	245.4[42]	10.65	20.16	10.7	10.6	11.9	9.7	9.5	13.4	11.9	9.9	9.6		10.7 ± 0.2
Ph guanidine	5.24	236.7[h]	12.18[2]		10.6	11.2	12.5	10.2	10.0	14.9	13.4	11.4	11.1		12.2 ± 0.2
2-Cl-$C_6H_4P_1$(dma)	1.68	243.8[42]		19.07	9.9	8.8	10.1	7.8	7.6	11.6	10.2	8.2	7.8		9.1 ± 0.2
2,5-Cl_2-$C_6H_3P_1$(pyrr)	1.24	248.3[42]	9.21[e]	18.52	8.1	10.2	11.5	9.3	9.1	12.9	11.4	9.4	9.0		9.2 ± 0.2
4-NO_2-$C_6H_4P_1$(pyrr)	1.30	241.2[h]	9.22	18.51	7.0	9.0	10.3	8.0	7.8	11.1[i]	9.6[i]	7.6[i]	7.3[i]		9.2 ± 0.2

* pK_a according to Eq. 10. [a] The aqueous pK_a values are from Ref. 14 if not specified otherwise, the MeCN values are from Ref. 11, if not specified otherwise. [b] Estimated from the pK_a values determined in MeCN and THF, from Ref. 14. [c] The estimated uncertainties correspond to *ca* 90 % probability (Ref. 46). [d] Experimental values if not indicated otherwise. [e] From Ref. 14, using 0.1% aqueous TWEEN solution. [f] Phenyl tetramethylguanidine was used as reference base. [g] PhP_1(pyrr) was used as reference base. [h] Computational values from this work, c.f. Computational methods. [i] Estimated value, based on lowest energy states during geometry optimization. The effects of extrapolation are estimated to be within 0.1 pK_a unit margins for the corresponding pK_a values.

Table 2. Statistical data of regression analysis between experimental and computational pK_a data for both compound groups

Without grouping (N = 17)[a]	CPCM direct	CPCM *vs* PhTMG	CPCM *vs* PhP1(pyrr)	CPCM *vs* PhP1(dma)	SMD direct	SMD *vs* PhTMG	SMD *vs* PhP1(pyrr)	SMD *vs* PhP1(dma)	COSMO -RS
Slope	0.70	0.70	0.70	0.70	0.77	0.77	0.77	0.77	0.66
s(Slope)	0.15	0.15	0.15	0.15	0.20	0.20	0.20	0.20	0.10
R^2	0.59	0.59	0.59	0.59	0.49	0.49	0.49	0.49	0.74
Mean Error	−0.38	−1.70	0.59	0.79	−2.91	−1.44	0.57	0.91	−0.44
MUE	1.16	1.77	1.12	1.18	2.91	1.69	1.05	1.19	1.04
RMSE	1.33	2.13	1.41	1.50	3.20	1.95	1.44	1.60	1.28
Phosphazenes (N = 10)[a]									
Slope	0.55	0.55	0.55	0.55	0.56	0.56	0.56	0.57	0.47
s(Slope)	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07	0.12
R^2	0.90	0.90	0.90	0.90	0.89	0.89	0.89	0.89	0.67
Mean Error	−0.84	−2.17	0.12	0.32	−3.60	−2.13	−0.12	0.22	−0.37
MUE	1.05	2.17	0.70	0.73	3.60	2.13	0.70	0.69	1.03
RMSE	1.25	2.36	0.93	0.97	3.71	2.30	0.88	0.89	1.22
Amidines/guanidines (N =7)[a]									
Slope	0.46	0.46	0.46	0.46	0.63	0.63	0.63	0.63	0.53
s(Slope)	0.30	0.30	0.30	0.30	0.50	0.50	0.49	0.50	0.22
R^2	0.32	0.32	0.32	0.32	0.24	0.24	0.24	0.24	0.53
Mean Error	0.29	−1.04	1.25	1.45	−1.92	−0.45	1.55	1.90	−0.54
MUE	1.32	1.20	1.74	1.82	1.92	1.06	1.57	1.90	1.04
RMSE	1.45	1.76	1.90	2.03	2.28	1.30	1.97	2.26	1.37

[a] experimental values from Table 1, excluding the ones that are obtained from correlation analysis from other solvent (HP_1(dma) and HP_1(pyrr)). For DBU the experimental value of 11.9 is used.

Among the amidines/guanidines group there seem to be two outliers, DBU and MTBD that are by their published experimental pK_a values seemingly better grouped with phosphazenes. However, the *WANS* values of their cations do not support their exclusion from the amidines group.

Taking into account the two groups additional group-wise correlations were made. For phosphazenes the correlation improved drastically for all CPCM and SMD computations, the R^2 being around 0.9 and the best MUE being around 0.7 pK_a units, if computational schemes relative to phosphazenes are used. Both the direct SMD models and schemes relative to PhTMG give considerably worse MUEs, up to 3.6 pK_a units. The direct CPCM models gives still good results with MUE = 1.05. COSMO-RS differs from CPCM and SMD models by worse R^2 value 0.67, but seems still to have acceptable errors (MUE = 1.03). The poor correlation is mostly due to the least basic phosphazenes (4-NO_2 and 2,5-Cl_2 substituted PhP1(pyrr)), which deviate strongly but no concrete reason could be found.

For amidines/guanidines the correlation and error characteristics remain poor (R^2 = 0.2 .. 0.5) because of deviation of DBU and MTBD.

Inspection of the results of correlation between calculated and experimental aqueous pK_a values presented in Table 2 reveals that the regression line slopes are rather low for all computational methods. Similar

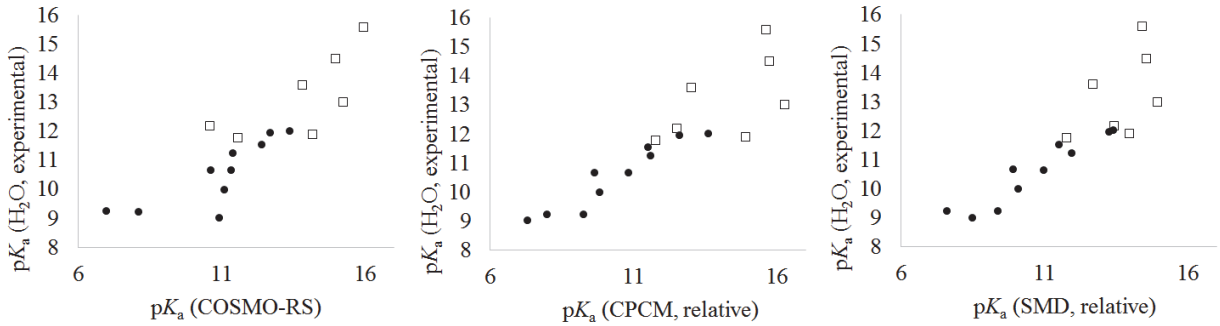


Figure 2. Correlations between computed and experimental pK_a values. Full dots represent phosphazenes and emty squeres amidines/guanidines; Relative CPCM and SMD methods use PhP1(pyrr) and PhTMG as references for phosphazenes and amidines/guanidines, respectively.

observations regarding low slope values for aqueous pK_a calculations using implicit solvation approach have been reported on several occasions.[51] It was shown by Adam[51] that adding one explicit water molecule to the anions of phenols increased the value of the slope of pK_a regression line from 0.50 to 0.88 while for aliphatic carboxylic acids adding two water molecules to the anions changed the value of slope from 0.50 to 1.01. In contrast, the pK_a regression for unhydrated anilinium ions was characterized by the slope value of 0.70 which is higher than the slope value for unhydrated phenols and carboxylic acids and seems to be insensitive to hydration. Kelly *et al.*[51] studied the effects of adding explicit water molecules to the anions of monoprotic acids and also arrived at a conclusion that in terms of slope of the pK_a regression equation the performance of pure polarized continuum model is improved after implementing the cluster–continuum approach. The same authors argued that adding explicit water molecules is usually justified in case of small size anions and anions with significant charge localization. They also noted that the addition of water molecules does not always lead to improved calculation accuracy and that for several acids reliable pK_a values were obtained using pure continuum treatment of aqueous medium.[51] The results of Chipman[51] obtained for neutral OH and cationic NH acids are consistent with those reported by Adam.[51] It is evident that implicit solvation treatment of the latter group of acids yields aqueous pK_a values that are in reasonable agreement with the experiment while for the former acid group characterized by small to medium size anions with high degree of charge localization pure polarized continuum approach fails.[51]

In this respect it is important to note that the present study is all about alkylated guanidine and phosphazene bases. The ionic species – protonated bases – involved are bulky and the charge is extensively delocalized in the cations. This is evidenced by their *WANS* values being in general below 4 (only phenyl guanidine above 5). *WANS* values of charge-localized cations are significantly higher as evidenced by the *WANS* examples given above. Under these circumstances the cluster–continuum protocol has not been considered a mandatory choice in this study. However, the slope value for the group of phosphazene bases is still low and this finding deserves further attention. In particular, it would be important to discriminate between deficiencies of implicit solvation approach and other possible reasons, most importantly the uncertainties of experimental pK_a and GB values.

Assigning Recommended pK_a Values to the Bases

The following criteria were taken into account when assigning the recommended pK_a values:

1. The experimental data of moderately basic compounds (pK_a around 11 or below) are much more reliable than computational values. At the same time experimental values of bases with high basicity are not very reliable and due to the specifics of the pK_a measurement methods tend to be underestimated, rather than overestimated.
2. Computations using reference bases are generally more reliable than direct computations, because the errors in solvation energy partially cancel, and the reliability increases with increasing similarity of the reference base and investigated base.
3. Computations via MeCN pK_a values according to Eq. 10 are more reliable than computations via gas-phase basicities (Eq. 9), because (a) the same base is used, (b) pK_a values in MeCN are more reliably known than GB values and (c) MeCN as a medium is more similar to water than the gas phase. The only counterargument is that solvation energies in two different solvents are used, instead of just one solvent as in Eq. 9.
4. The basicity order in water can differ significantly from the gas phase but not too much from acetonitrile.
5. Several sources of experimental pK_a data have very limited or completely missing experimental parts. This precludes judging their reliability and decreases their trustworthiness.
6. Correlations between the computational and available experimental data range from poor to fair. In addition they cover the low to medium basicity range only. These two factors together make these correlations of little use for correcting/adjusting the predicted pK_a values of strong bases and consequently were not used.

The assigned recommended values are presented in the last column of Table 1. Comments on some of the more important bases follow.

The recommended pK_a value 13.5 for DBU has been assigned taking into account all computations but ignoring the experimental values of 11.5 and 11.9. The experimental value of PhTMG 11.77 and comparison of these two bases in MeCN implies that both experimental aqueous pK_a values of DBU are most probably underestimated. The sources of the experimental values do not contain any descriptions of experimental pK_a determination.

The recommended pK_a values of MTBD and TBD are primarily based on Eq. 10. Their basicity order does not match that of most computations and also not the basicity order in the gas phase, but matches the experimental basicity order in MeCN. More efficient solvation of $TBDH^+$ is the reason for TBD being more basic in MeCN than MTBD. The same is expected in water. They fall nicely in the correct area: for MTBD COS-

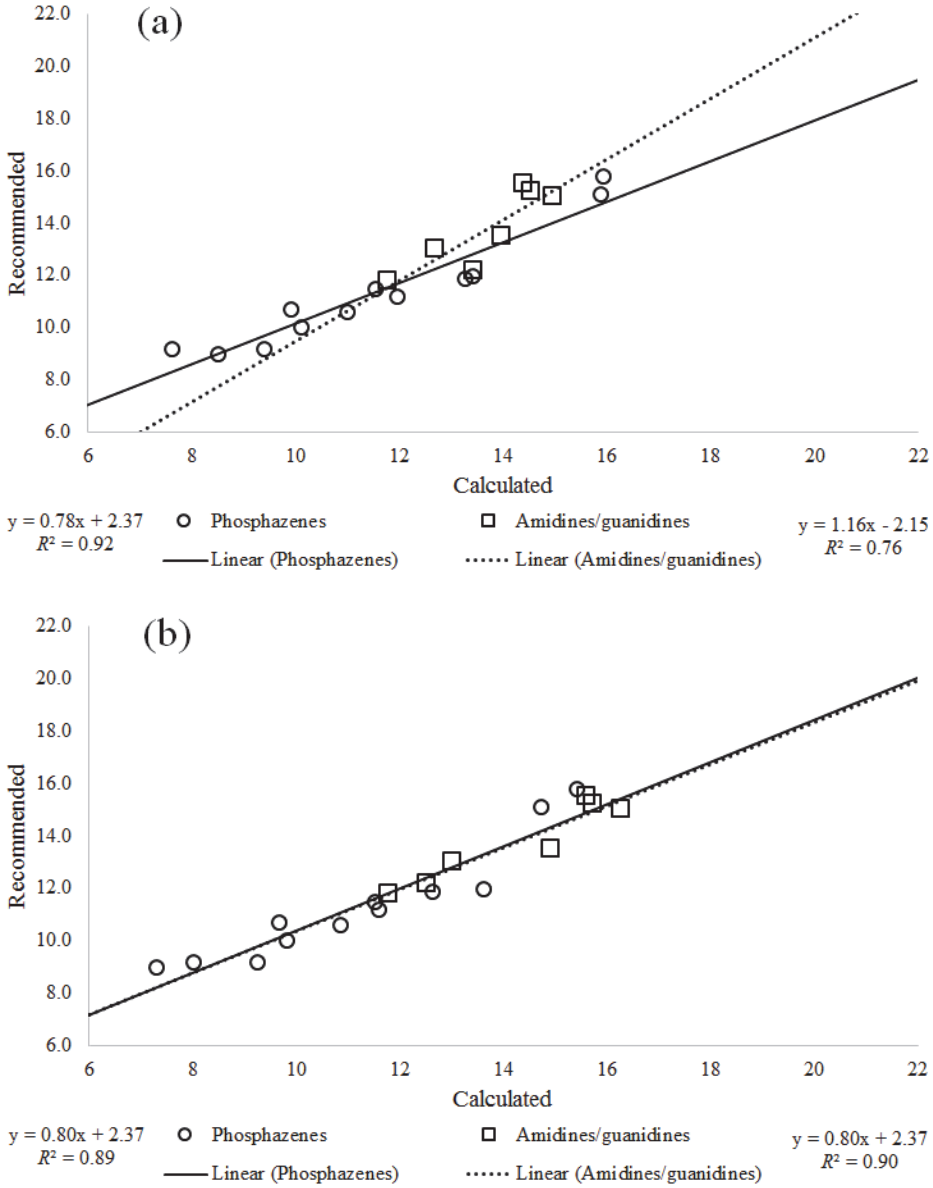


Figure 3. Correlation between the pK_a values obtained with the SMD (a) and CPCM (b) relative schemes and recommended pK_a values. Relative methods use PhP1(pyrr) and PhTMG as references for phosphazenes and amidines/guanidines, respectively.

MO-RS and SMD (relative to PhTMG) values are 15.3 and 15.0, respectively, for TBD 15.0 and 14.5.

Assigning the new recommended values to DBU and MTBD removes the above described problem that these bases seriously fall off from the correlations between experimental and computational values.

The published aqueous pK_a values of HP_1(dma) and HP_1(pyrr) (13.32 and 13.93, respectively) were obtained from correlation analysis from values in MeCN and THF.[14] The results of the present calculations do not support those values. Without exception, all computed values are higher. This can be due to the much less hindered basicity center in the cations and thus more efficient solvation stability of the protonated forms of the bases than in the case of the phosphazene bases that were used for correlation analysis in Ref. 14. The recommended values are based first of all on Eq. 10, but are also well supported by CPCM and SMD calculations if phosphazenes are used as reference bases. With the new recommended values these two phosphazene bases drift away from the phosphazene series of the correlations described in the previous section. The reason is that the basicity centers of these two bases are less sterically screened than those of any other phosphazenes in this study.

Figure 3 displays the correlations between pK_a values given as recommended in Table 1 and computed

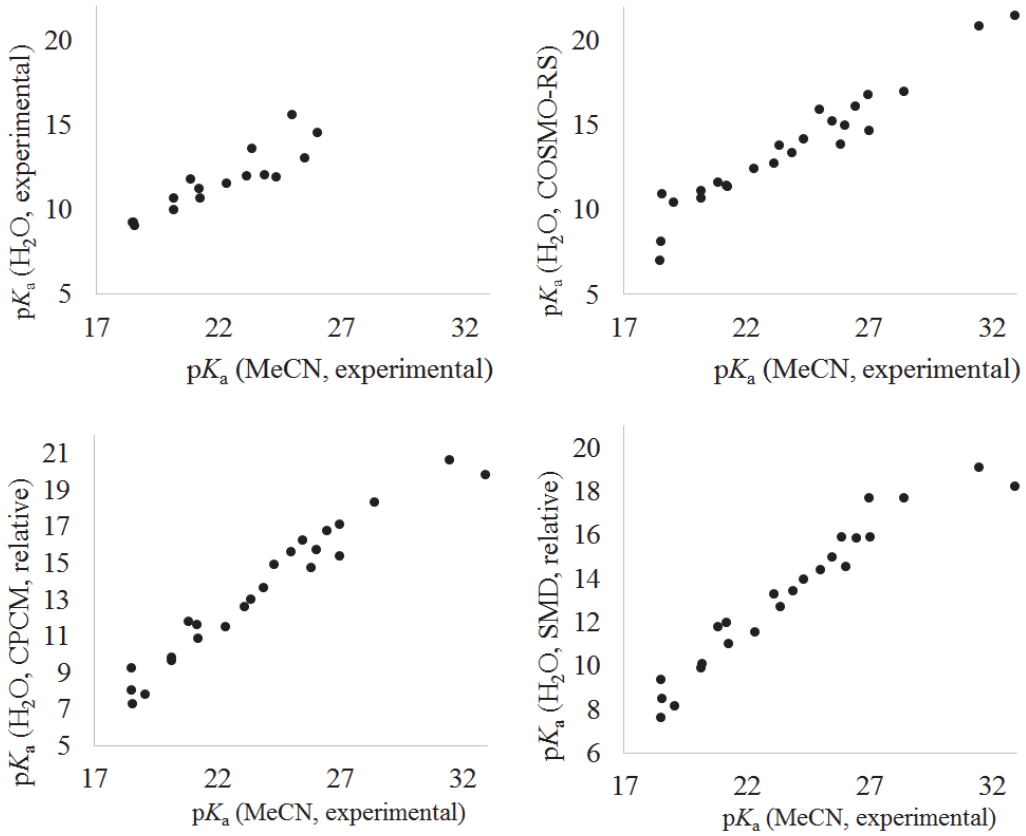


Figure 4. Correlation between pK_a values in water and experimental values in MeCN. Relative CPCM and SMD methods use PhP1(pyrr) and PhTMG as references for phosphazenes and amidines/guanidines, respectively. *t*-BuP$_4$(dma) is left out for clarity reasons.

by the relative SMD and CPCM methods (Eq. 9, using PhTMG and PhP1(pyrr) as reference bases). Only those compounds (12 phosphazenes and 7 amidines/guanidines) that had either an experimental value prior to this study or their recommended values have been obtained using Eq. 10 have been used in the correlation. The SMD method demonstrates slope values between 0.78 and 1.16. The R^2 of phosphazenes is good 0.92, that for amidines/guanidines is worse with 0.76. For CPCM the slope values for both groups are identical with 0.80. The R^2 values are very similar as well with 0.89 and 0.90.

Correlations of Aqueous pK_a Data with MeCN pK_a Data

In order to gain further insight into the quality of the computed aqueous pK_a values they were correlated with the experimental pK_a values in MeCN. MeCN was chosen as a reference solvent because (1) reliable experimental pK_a values of nearly all investigated bases are available from the literature, (2) there is a fairly good correlation between pK_a values in water and acetonitrile as the equation 3 suggests and (3) the true ionic basicities can be measured in acetonitrile, unlike THF, where the actually measured values refer to ion-pair basicities.[52]

Figure 4 shows that the basicity region investigated in this work is almost fully covered by measured pK_a values in MeCN (except EtP$_2$(dma) and *t*-BuP$_4$(dma), which are too basic to be directly measured in MeCN and the MeCN pK_a values have been estimated from THF data[49]). The computed aqueous pK_a values correlate quite well with the experimental data in MeCN, thereby indirectly confirming that the computed aqueous pK_a values do not contain major errors. The correlation between experimental values between water and MeCN shows that dispersion of points around regression line increases along with the increase of basicity.

CONCLUSIONS

In the pK_a range above 12 neither experiments nor computations by any single approach are sufficiently reliable for assigning reliable pK_a values for bases. The best estimates of pK_a values are obtained by combining

knowledge from experiments in water, in other solvents and from the gas phase with different computational methods and chemical reasoning taking into account the expected reliability of experiments and computations, as well as the chemical properties of the involved species.

Acknowledgements. This work was in part supported by institutional research funding IUT14-20 (TLOKT14014I) and IUT23-7 from the Estonian Ministry of Education and Research.

REFERENCES

1. T. Ishikawa and T. Kumamoto *Synthesis* **5** (2006) 737−752.
2. T. Ishikawa, *Superbases for Organic Synthesis*, Ed. John Wiley and Sons Ltd, West Sussex, UK. 2009.
3. Z. B. Maksić, B. Kovačević, and R. Vianello, *Chem. Rev.* **112** (2012) 5240–5270.
4. K. Vazdar, R. Kunetskiy, J. Saame, K. Kaupmees, I. Leito, and U. Jahn, *Angew. Chem. Int. Edit.* **53** (2014) 1435–1438.
5. J. F. Kögel, B. Oelkers, B. Kovačević, and J. Sundermeyer, *J. Am. Chem. Soc.* **135** (2013) 17768−17774.
6. M. Eckert-Maksic, Z. Glasovac, P. Troselj, A. Kutt, T. Rodima, I. Koppel, and I. A. Koppel, *Eur. J. Org. Chem.* **30** (2008) 5176−5184.
7. M. P. Coles, *Chem. Comm.* (2009) 3659−3676.
8. S. Boileau and N. Illy, *Prog. Polym. Sci.* **36** (2011) 1132−1151.
9. T. Ishikawa and L. M. Harwood, *Synlett* **24** (2013) 2507−2509.
10. As is customary among organic chemists, the term "pK_a of base X" is used in this paper with the actual meaning of pK_a of the conjugate acid of base X.
11. I. Kaljurand, A. Kütt, L. Sooväli, T. Rodima, V. Mäemets, I. Leito, and I. A. Koppel, *J. Org. Chem.* **70** (2005) 1019–1028.
12. R. G. Bates, *Determination of pH: Theory and Practice*. Wiley, New York, 1964.
13. R. A. Cox and R. Stewart, *J. Am. Chem. Soc.* **98** (1976) 488−494.
14. L. Sooväli, T. Rodima, I. Kaljurand, A. Kütt, I. A. Koppel, and I. Leito, *Org. Biomol. Chem.* **4** (2006) 2100–2105.
15. J. Tomasi, B. Mennucci, and R. Cammi, *Chem. Rev.* **105** (2005) 2999−3093.
16. (a) M. D. Liptak and G. C. Shields, *J. Am. Chem. Soc.* **123** (2001) 7314−7319; (b) M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus, and G. C. Shields, *J. Am. Chem. Soc.* **124** (2002) 6421−6427.
17. (a) V. Barone and M. Cossi, *J. Phys. Chem. A* **102** (1998) 1995−2001; (b) M. Cossi, N. Rega, G. Scalmani and V. Barone *J. Comp. Chem.* **24** (2003) 669−681.
18. A. Trummal, A. Rummel, E. Lippmaa, P. Burk, and I. A. Koppel *J. Phys. Chem. A* **113** (2009) 6206−6212.
19. (a)E. Cancès, B. Mennucci, and J. Tomasi, *J. Chem. Phys.* **107** (1997) 3032-3041; (b) B. Mennucci and J. Tomasi, *J. Chem. Phys.* **106** (1997) 5151−5158; (c) B. Mennucci, E. Cancès and J. Tomasi, *J. Phys. Chem. B* **101** (1997) 10506-10517; (d) J. Tomasi, B. Mennucci and E. Cancès, *J. Mol. Struct. (Theochem)* **464** (1999) 211−226.
20. B. Kovačević and Z. B. Maksić, *Org. Lett.* **3** (2001) 1523−1526.
21. H. Lu, X, Chen and C.-G. Zhan, *J. Phys. Chem. B* **111** (2007) 10599−10605.
22. S. Gangarapu, A. T. M. Marcelis, and H. Zuilhof, *ChemPhysChem* **14** (2013), 990–995.
23. A. Trummal, A. Rummel, E. Lippmaa, I. Koppel, and I. A. Koppel, *J. Phys. Chem. A* **115** (2011) 6641−6645.
24. A. V. Marenich, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. B* **113** (2009) 6378−6396.
25. (a) J. R. Pliego Jr and J. M. Riveros, *J. Phys. Chem. A* **106** (2002) 7434−7439; (b) A. V. Marenich, W. Ding, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. Lett.* **3** (2012) 1437−1442.
26. F. Eckert, I. Leito, I. Kaljurand, A. Kütt, A. Klamt, and M. Diedenhofen, *J. Comput. Chem.* **30** (2009) 799−810.
27. (a) F. Eckert and A. Klamt, *AlChE Journal* **48** (2002) 369−385; (b) A. Klamt and F. Eckert *Fluid Phase Equilibr.* **172** (2000) 43−72.
28. D. J. Heldebrant, P. K. Koech, J. E. Rainbolt, F. Zheng, T. Smurthwaite, C. J. Freeman, M. Oss, and I. Leito, *Chem. Eng. J.* **171** (2011) 794−800.
29. A. Klamt, F. Eckert, M. Diedenhofen, and M. E. Beck, *J. Phys. Chem. A* **107** (2003) 9380−9386.
30. F. Eckert and A. Klamt, *J. Comput. Chem.* **27** (2006) 11−19.
31. J. Ho and M. L. Coote, *Theor. Chem. Acc.* **125** (2010) 3−21.
32. A. Klamt, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*; Elsevier Science Ltd: Amsterdam, The Netherlands, 2005.
33. E. Raamat, K. Kaupmees, G. Ovsjannikov, A. Trummal, A. Kütt, J. Saame, I. Koppel, I. Kaljurand, L. Lipping, T. Rodima, V. Pihl, I. A. Koppel, I. Leito, *J. Phys. Org. Chem.* **26** (2013) 162−170.
34. TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989−2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com
35. F. Eckert and A. Klamt COSMOtherm, Version C3.0, Release 14.01; COSMOlogic GmbH&CoKG, Leverkusen, Germany, 2013; Available at http://www.cosmologic.de/
36. a) K. Kaupmees, I. Kaljurand, and I. Leito, *J. Phys. Chem. A* **114** (2010) 11788–11793, (b) K. Kaupmees, I. Kaljurand, and I. Leito, *J. Sol. Chem.* **43** (2014) 1270−1281.
37. M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, and T. R. Tuttle Jr., *J. Phys. Chem. A* **102** (1998) 7787−7794.
38. J. Ho, A. Klamt, and M. L. Coote, *J. Phys. Chem. A* **114** (2010) 13442−13444.
39. Gaussian 09, Revision B.01, M. J. Frisch, G.W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.
40. D. Himmel, S. K. Goll, I. Leito, and I. Krossing, *Angew. Chem. Int. Ed.* **49** (2010) 6885–6888.
41. Y. Marcus, M. J. Kamlet, and R. W. Taft, *J. Phys. Chem.* **92** (1988) 3613−3622.
42. I. Kaljurand, I. A. Koppel, A. Kütt, E.-I. Rõõm, T. Rodima, I. Koppel, M. Mishima, and I. Leito, *J. Phys. Chem. A* **111** (2007) 1245−1250.
43. E. P. L. Hunter and S. G. Lias, *J. Phys. Chem. Ref. Data* **27** (1998) 413−656.
44. R. Srivastava, *J. Mol. Catal. A: Chem.* **264** (2007) 146–152.
45. K. M. K. Yu, I. Curcic, J. Gabriel, H. Morganstewart, and S. C. Tsang, *J. Phys. Chem. A* **114** (2010) 3863–3872.

46. F. Ravalico, S. L. James, and J. S. Vyle, *Green Chem.* **13** (2011) 1778−1783.
47. D. D. Perrin, *Dissociation Constants of Organic Bases in Aqueous Solution*: Published as a Supplement to Pure and Applied Chemistry, Butterworth, London, 1965.
48. P. Przuszynski, *Can. J. Chem.* **65** (1987) 626−629.
49. R. Schwesinger, H. Schlemper, C. Hasenfratz, J. Willaredt, T. Dambacher, T. Breuer, C. Ottaway, M. Fletschinger, J. Boele, H. Fritz, D. Putzas, H. W. Rotter, F. G. Bordwell, A. V. Satish, G.-Z. Ji, E.-M. Peters, K. Peters, H. G. von Schnering, and L. Walz, *Liebigs Ann*. (1996) 1055−1081.
50. R. Schwesinger, J. Willaredt, H. Schlemper, M. Keller, G. Schmittand, and H. Fritz, *Chem. Ber.* **127** (1994) 2435−2454.
51. (a) K. Adam, *J. Phys. Chem. A* **106** (2002) 11963−11972; (b) J. J. Klicić, R. A. Friesner, S.-Y. Liu, and W. C. Guida, *J. Phys. Chem. A* **106** (2002) 1327−1335; (c) D. M. Chipman, *J. Phys. Chem. A* **106** (2002) 7413−7422; (d) C. P. Kelly, C. J. Cramer, and D. G. Truhlar, *J. Phys. Chem. A* **110** (2006) 2493−2499.
52. T. Rodima, I. Kaljurand, A. Pihl, V. Mäemets, I. Leito, and I. A. Koppel, *J. Org. Chem.* **67** (2002) 1873–1881.