

Prediction of accurate pK_a values of some α -substituted carboxylic acids with low cost of computational methods

M. Abul Kashem Liton*, Sabrina Helen, Mukta Das, Dipa Islam and M. Rabiul Karim

Department of Chemistry, MawlanaBhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh

Received July 2015; Accepted December 2015

ABSTRACT

The acidity constants (pK_a) of thirty four (34) α -substituted carboxylic acids in aqueous solution have been calculated using conductor-like polarizable continuum (C-PCM) solvation model. The gas-phase energies at the Density Functional Theory (DFT-MPW1PW91) and solvation energies at HartreeFock (HF) are combined to estimate the pK_a values which are very close to the experimental values where, and the root-mean-square error ($rmse$) is less than 0.9. In addition, the pK_a values predicted by Multilinear Regression (MLR), Principle Components Regression (PCR) and Partial Least Square Regression (PLSR) methods showed an excellent correlation to the experimental values where the root-mean-square error ($rmse$) decreases from 0.9 to 0.293, 0.326 and 0.319 respectively. Furthermore, the $rmse$ of pK_a values found less than 0.3 when the solvation models of some powerful electron withdrawing substituted acids are employed at the level of MPW1PW91 theory. The computation by different regression methods showed reasonably a better calculation of pK_a values through the lower root-mean-square errors ($rmse$).

Keywords: HF; DFT; C-PCM; pK_a ; Free energy

INTRODUCTION

The activity and the ionizing ability of carboxylic acids are the direct function of the electronic properties and bonding order of the atoms that make up the carboxyl (COOH) moiety. The most significant chemical property of carboxylic acids in medicinal chemistry is their acidic nature. When considering the solution behaviour of a drug in the body, dealing with a dilute solution for which the Brønsted-Lowry acid-base theory is the most appropriate to explain and predict acid-base properties. The most important concept in medicinal chemistry is the acidic properties of drug molecules that influence lipophilicity,

solubility [1], protein binding and permeability, which in turn show the direct effect on absorption, distribution, metabolism and excretion [2-6], and potential incompatibilities between drugs. Procedural formulation for optimizing drug delivery [7] also benefits from the determination of the ionization constant (pK_a). Pharmaceutical products and the formulation of useful dosage forms depend upon an understanding of drug dissociation and the extent of dissociation that will occur in the organism of the body. Mono-, di-, and tri-carboxylic acids are sometimes key intermediates in a variety of metabolic

*Corresponding author: litonchem@hotmail.com

pathways [8].

Numerous reports have been appeared in the literature describing computational methods of pK_a calculations [9-12]. The most commonly used schemes couple quantum chemical calculations of gas-phase deprotonation energy with continuum dielectric models to account for solvent effects. The coupling of Density Functional Theory (DFT) with a continuum solvent model has been shown [12] to provide a computationally viable and accurate pK_a calculation. Several studies have been attempted to accurately calculate pK_a of organic molecules but none of them have achieved chemical accuracy [13-18]. There are also many studies that use a variety of methods to accurately calculate the pK_a values. For an instance, the relative pK_a of carboxylic acids [19] have been calculated at HF and B3LYP methods in aqueous solution using Möller-Plesset Perturbation Theory (MP2) where most of the acids showed a good agreement to the corresponding experimental pK_a values, in some cases, large deviations exist, such as, unsigned errors of pK_a from experimental values for BrCH_2COOH , FCH_2COOH and Cl_2CHCOOH are 0.88, 1.47 and 2.03 pK_a units where unsigned errors of these acids in this study are very small (0.03, 0.04 and 0.06 for BrCH_2COOH , FCH_2COOH and Cl_2CHCOOH). It would thus appear that more accurate methods are necessary for this problematic system. To address this problem, in the present work, the pK_a of carboxylic acids has been successfully calculated using low cost method of MPW1PW91 gas phase energy in conjunction with appropriate models of solvation.

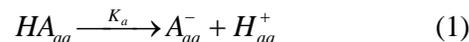
However, in recent years, many improvements in computational chemistry have been observed, discrepancy still exist. The aim of our present study is to establish the theoretical pK_a calculations and to

compare the predicted pK_a s among a variety of methods and basis sets used in C-PCM solvation methods and, also comparison with pK_a s obtained by regression methods.

METHODS AND CALCULATION

Thermodynamic cycle

According to the definition of Brønsted acid in aqueous solution, it dissociates as follows:



The equilibrium constant of the reaction (1) is K_a of HA and pK_a is $-\log K_a$. There is a simple relation between the standard Gibbs free energy changes of reaction (1), ΔG_{aq}^o and K_a .

$$\Delta G_{aq}^o = -2.303RT \log K_a \quad (2)$$

$$\Delta G_{aq}^o = 2.303RT pK_a \quad (3)$$

$$pK_a = \frac{\Delta G_{aq}^o}{2.303RT} \quad (4)$$

Based on thermodynamic [20] cycle (Figure 1), Gibbs-free energy change, ΔG_{aq}^o of reaction (1) in aqueous solution is equal to the sum of gas-phase Gibbs free energy change, ΔG_{gas}^o and change of solvation energies, $\Delta \Delta G_{solv}^o$, i.e.,

$$\Delta G_{aq}^o = \Delta G_{gas}^o + \Delta \Delta G_{solv}^o \quad (5)$$

In this work, ΔG_{gas}^o and $\Delta \Delta G_{solv}^o$ are

$$\Delta G_{gas}^o = G_{(gas,A^-)}^o + G_{(gas,H^+)}^o - G_{(gas,HA)}^o \quad (6)$$

$$\Delta \Delta G_{solv}^o = \Delta G_{(solv,A^-)}^o + \Delta G_{(solv,H^+)}^o - \Delta G_{(solv,HA)}^o \quad (7)$$

where, $G_{(gas,HA)}^o$, $G_{(gas,A^-)}^o$ and $G_{(gas,H^+)}^o$ are the gas-phase Gibbs free energies and $\Delta G_{(solv,HA)}^o$, $\Delta G_{(solv,A^-)}^o$ and $\Delta G_{(solv,H^+)}^o$ are the change of Gibbs free energies of solvation for HA , A^- and H^+ respectively.

Generally, three sources of errors in pK_a

calculations [21] are found. The first source of the errors is the model used to calculate the pK_a , which generally involves a thermodynamic cycle shown in Fig. 1 and the second and third major errors found from the calculations of ΔG_{aq}^o and ΔG_{solv}^o energies. In this work, gas-phase Gibbs free energies and the changes in free energy of solvation (aqueous) of neutral molecules (HA), and their anion (A^-) have been calculated with minimal errors. The experimental gas-phase Gibbs free energy (-6.28 kcal/mol[22]) and the change in free energy of solvation (-264.00 kcal/mol[23]) of H^+ in standard state (1 atm and 298.15 K) were employed for the calculation of pK_a s of studied acids.

COMPUTATIONAL DETAILS

All the calculations were carried out on Intel core-i5 Fujitsu laptop computer with 4–8 GB of memory and 60 GB of scratch disk space, using Gaussian03 (v6.0) program [24]. About 12 different combinations of methods (HF [25], B3LYP [26], MPW1PW91 [27], MP2 [28]) and basis sets [6-311g(d,p), 6-311+g(d,p), 6-311g(2d,2p), 6-311+g(2d,2p) and 6-311++g(2d,2p)] were applied in trial calculations of gas-phase Gibbs free energy changes to compare their relative accuracies with the experimental values and next the solvation model [30, 31] (C-PCM) at HF, B3LYP, MP2 and MPW1PW91 methods with small basis sets (6-31g(d) and 6-31+g(d)) were also employed in trial calculations of solvation energies to select the best combination (gas-phase and solvation energies) of methods to determine the pK_a values of the set of α -substituted carboxylic acids. In the solvation model, the radii of the united atom topological model (UA0 and UAHF) and “scfvac” keyword have been chosen as recommended in the Gaussian03 user’s

reference manual, and the rest of the parameters of the models have been kept as default [24] to obtain the free-energy changes of solvation of the studied molecules.

Every structure was optimized in the gas phase at the respective level of theory, and the optimized structures were checked with all real frequencies then the gas phase Gibbs free energy changes, ΔG_{gas}^o have been calculated. Next solvation energies in water were calculated at the level of Hartree-Fock [25] (HF) and Density Functional Theory [26, 27, 29] (DFT-MPW1PW91) using conductor-like polarizable continuum solvation model [29, 30] (C-PCM). Small-sized basis set has been preferred considering the size of studied molecules, computational cost and proof for reliability of these basis sets. Before the calculation of solvation energies, HF/6-311+g(d)-UAHF and B3LYP/6-311+g(d)-UA0 levels of theory in C-PCM aqueous solution method were arbitrary chosen for the optimization of all studied molecules. Next the optimized structures were confirmed through vibrational frequency calculations with all real frequencies. These computational procedures (levels of theory, basis sets, solvation model, etc.) have been designed particularly for the prediction of reliable energies of the studied molecules and have been demonstrated to provide the accuracy within 1-2 kcal/mol when assessed against large test sets of thermo-chemical data [32, 33].

RESULTS AND DISCUSSIONS

Selection and justification of experimental data

Initially, some reported experimental and theoretical values which are used in this work for comparison or as part of the calculations are needed to re-examine. The gas-phase energy and solvation energy of

H^+ are very important in this study for the calculation of pK_a values. The gas-phase free energy, $G_{(gas,H^+)}^o$ of H^+ cannot be determined quantum mechanically due to the unavailability of electrons in proton. The experimental value of -6.28 kcal/mol [22] in the standard state (1 atm 298.15 K) is used to calculate the pK_a s of the studied molecules. Free energy of H^+ was taken from the standard equations of thermodynamics, and based on the Sackur-Tetrode equation [34] where the deduced value was the same as can be found (-6.28 kcal/mol) experimentally from the NIST database. The value of H° (H^+) equal to $5/2(RT)$ or 1.48 kcal/mol was deduced using the translational energy of $1.5RT$ combined with $PV = RT$ and $H = E + PV$. Finally, $G^\circ(H^+) = -6.28$ kcal/mol was obtained by putting the entropy value of TS (H^+) (7.76 kcal/mol at 1 atm pressure 298.15 K) which was already deduced from the Sackur-Tetrode equation and the value of H° (H^+) (1.48 kcal/mol in the equation, $G = H - TS$).

However, to date, several theoretical and experimental values of $\Delta G_{(solv,H^+)}^o$ have been reported: -264.0 [23], -264.6 [21, 22], -265.9 [35, 36] and -266.1 [37] kcal/mol. In addition, Coe and his co-workers [18] had used experimental ion-water clustering data to derive a value for $\Delta G_{(solv,H^+)}^o$ of -264.0 kcal/mol. In the present work, this experimental value of $\Delta G_{(solv,H^+)}^o$ is also used to determine the pK_a values of α -halogenated carboxylic acids in the standard state (1 atm pressure 298.15 K). It was supported that the correct standard state was 1 atm, so that changing to a standard state of 1M, the value of $\Delta G_{(solv,H^+)}^o$ was converted into -265.9 kcal/mol [35, 36]. Further supported by Goddard and his co-workers [38], they included concentration terms to Zhan and

Dixon's high-level ab initio calculations [39] of the hydration free energy of the proton, their value is corrected to -265.63 ± 0.22 kcal/mol.

Calculation of the pK_a values

In this study, we have presented a simple approach for the calculation of accurate pK_a values of thirty four α -substituted carboxylic acids (mono-, di- or tri-substituents at α -position), which are shown in Table 1. A series of trial calculations were performed to select the best computational methods (basis sets, level of theory and solvation models) to be used in the pK_a calculations of the set of α -substituted carboxylic acids. The preeminent results of gas-phase Gibb's free energy changes, ΔG_{gas}^o of the acid dissociation calculated at MPW1PW91/6-311+g(2d,2p) level of theory and the experimental [40, 41] gas-phase Gibb's free energy changes at 298.15 K are also reported in Table 1. All the calculated gas-phase Gibbs free energy changes were within the acceptance limit of experimental values and comparison of calculated gas-phase Gibb's free energy changes with experimental gas Gibb's free energy changes show that only small deviations exist with a maximum of 0.59%. The aqueous solvation energy of each molecule and its anion calculated at HF and MPW1PW91 methods with 6-31g(d) and 6-31+g(d) basis sets using C-PCM (conductor-like polarizable continuum model) solvation models are shown in Table 2. It is observed that the solvation energies calculated with 6-31g(d) basis set are comparatively better than that of 6-31+g(d) for both ions and the estimated pK_a s are also shown in Table 2 where the residual errors were incorporated in parenthesis. Among the combinations of C-PCM solvation model at the level of HF and MPW1PW91 theories with 6-31g(d)

and 6-31+g(d) basis sets, the pK_a calculated at HF/6-31g(d)/C-PCM aqueous solvation method showed very good results with lower root-mean-square error (*rmse*) of 0.838.

Examination of calculated pK_a showed that in most cases, the pK_a values are reasonably close to the experimental values. For example, the calculated pK_a of 2-bromopentanoic acid (29) is 2.96 whereas the experimental pK_a is reported to be 2.97. About similar approximation is observed for 2, 2-difluoroethanoic acid (calculated $pK_a = 1.27$ and experimental $pK_a = 1.24$), 2-bromoethanoic acid (calculated $pK_a = 2.89$ and experimental $pK_a = 2.86$) and 2, 2-dichloroethanoic acid (calculated $pK_a = 1.23$ and experimental $pK_a = 1.29$). In some cases, large deviations are observed where some negative values can be seen. These deviations were arisen from a higher number of halogens attached to the carboxylic acid and oxo group (=O) attached directly in the α -position, such as 2, 2, 2-trifluoroethanoic acid (calculated $pK_a = -1.85$ and experimental $pK_a = 0.23$) and 2-oxo-butanoic acid (calculated $pK_a = -0.65$ and experimental $pK_a = 2.50$). The root-mean-square error is 0.838 including those giving high deviations. Since the main source of deviation comes from the solvation energies, and the gas-phase energies should be also considered, so a stepwise multi-linear regression (MLR) was performed [42], and the following relation was found:

Model-1 (MLR):

$$pK_a(\text{MLR}) = 0.3832 \times \Delta G_{\text{gas}}^{\circ} - 0.2933 \times \Delta G_{(\text{solv}, \text{HA})}^{\circ} + 0.3265 \times \Delta G_{(\text{solv}, \text{A}^{-})}^{\circ} - 102.8277 \quad (8)$$

$$n_{\text{train}} = 29; r^2 = 0.982; q^2 = 0.976; p < 0.0001; F = 456.054; sd = 0.168; n_{\text{test}} = 5; r_{\text{pred}}^2 = 0.793$$

with standard deviation (*sd*) of 0.168 and squared correlation coefficient (r^2) of

0.982. The *F-value* of 456.054 indicates the robustness of the model and q^2 of 0.976 and r_{pred}^2 of 0.710 confirm the validity of the regression model internally and externally.

Next additional two regression models PCR (principle components regression where *PC1*, *PC2* and *PC3* are principle components) and PLSR (partial least squares regression) are included to verify the pK_a values obtained by MLR method. All the regression models gave excellent results of pK_a and in most cases; the pK_a values are very close to each other among them.

Model-2 (PCR):

$$pK_a(\text{PCR}) = 0.1282 \times \text{PC1} + 0.0641 \times \text{PC2} - 0.0276 \times \text{PC3} - 40.5388 \quad (9)$$

$$n_{\text{train}} = 29; r^2 = 0.891; F = 68.210; q^2 = 0.860; sd = 0.313; n_{\text{test}} = 5; r_{\text{pred}}^2 = 0.641$$

Model-3 (PLSR):

$$pK_a(\text{PLSR}) = 0.1398 \times \Delta G_{\text{gas}}^{\circ} + 0.0534 \times \Delta G_{(\text{solv}, \text{HA})}^{\circ} - 0.0123 \times \Delta G_{(\text{solv}, \text{A}^{-})}^{\circ} - 43.3755 \quad (10)$$

$$n_{\text{train}} = 29; r^2 = 0.898; F = 73.247; q^2 = 0.868; sd = 0.301; n_{\text{test}} = 5; r_{\text{pred}}^2 = 0.741$$

In the above models, 34 acids are considered, and they are divided into two: a training set of 29 acids and 5 acids that have been randomly chosen as a test set. The gas-phase free energies and solvation energy changes of neutral (HA) and anion (A⁻) species are considered in these analyses. Table 3 represents the predicted values of pK_a obtained by Eqs. (4), (8), (9) and (10) as well as experimental values [43-46]. Significant improvement can be seen in these regressions (MLR, PCR and PLSR) compared with directly measured pK_a values [47] obtained by Eq. (4) and the root-mean-square error (*rmse*) decreased

from 0.838 to 0.293, 0.326 and 0.319 for MLR, PCR and PLSR methods respectively. The pK_a s predicted with MLR, PCR and PLSR showed a high correlation with experimental values through squared correlation coefficients (r^2) of 0.982, 0.891 and 0.898.

It is noteworthy that the calculated pK_a s obtained by Eq. (4) with HF/6-31g(d)/C-PCM solvation model for studying acids except tri-halogenated acids (19, 25, 31) and 2-oxo-butanoic acids (34) showed a very good agreement with the experimental values where deviation exists less than 0.61 pK_a units (residual errors of pK_a are shown in Table 2 in a parenthesis). The *rmse* of 0.838 comes mainly from tri-halo substituted acids (19, 25 and 31) and 2-oxo-carboxylic acids (34) where deviations found more than 2.50 pK_a units. On contrary, these powerful electron withdrawing substituted acids (19, 25 and 31) and 2-oxo-carboxylic (34) acids showed an excellent result at MPW1PW91/6-31+g(d)/C-PCM solvation model where root-mean-square error (*rmse*) was 0.231. Next these four acids are separated and the *rmse* for the studied acids (except tri-halo and 2-oxo-substituted carboxylic acids) decreases from 0.838 to 0.299. Finally the *rmse* of thirty four α -substituted carboxylic acids decreases from 0.838 to 0.292 and squared correlation coefficient (r^2) of 0.941 indicates an

excellent calculation of pK_a s which are shown in Table 4. The correlations of pK_a s between experimental and predicted values are shown in Fig. 2.

CONCLUSION

The predicted pK_a s of thirty four α -substituted carboxylic acids by MLR, PCR and PLSR methods using gas-phase energies calculated at MPW1PW91/6-311+g(2d,2p) levels of theory and solvation energies at HF-6-31g(d)/C-PCM model are very close to the experimental values whereas directly measured pK_a s in some cases show large deviations with the root-mean-square error (*rmse*) of 0.838. In the mean while, predicted pK_a by regression methods showed comparatively better results with lower root-mean-square errors (0.293, 0.326 and 0.319). The large deviations are mainly originated from the tri-halogenated acids as well as 2-oxo-carboxylic acids (19, 25, 31 and 34) and the better pK_a s of these four acids are found at MPW1PW91/6-31+g(d)/C-PCM solvation model. It is observed that the values in the calculation of pK_a s are sensitive to the level of theory in comparison with the solvation models at which the solvation energies were calculated. It is also clear that high level of theory is always not necessary for the calculation of reliable pK_a s.

Table 1. Gas-phase Gibbs free energy of each molecule (*HA*) and its anion (*A*⁻) together with Gibbs free energy changes of reaction (1), ΔG_{gas}^o in the gas phase calculated at MPW1PW91/6-311+g(2d,2p) level of theory

Comp ID	Structural Formula	Hartree/particle		kcal/mol		% of errors
		G_{HA}^o	$G_{A^-}^o$	$\Delta G_{gas}^o (cal)$	$\Delta G_{gas}^o (exp)$	
1	H CH ₂ COOH	-229.073255	-228.519965	340.92	341.5±2	0.17
2	CH₃ CH ₂ COOH	-268.363263	-267.810859	340.36	340.4±2	0.01
3	CH ₃ (CH₃)CHCOOH	-307.652783	-307.101652	339.56	339.0±2	-0.17
4	CH ₃ (CH₃) ₂ CCOOH	-346.942264	-346.393126	338.31	338.0±2	-0.09
5	HO CH ₂ COOH	-304.282539	-303.734136	337.85	-	-
6	CH ₃ (OH)CHCOOH	-343.577597	-343.028922	338.02	-	-
7	NH₂ CH ₂ COOH	-284.408176	-283.864205	335.07	334.7±2	-0.11
8	HS CH ₂ COOH	-627.288633	-626.750692	331.29	-	-
9	CH₃S CH ₂ COOH	-666.577284	-666.036382	333.15	-	-
10	C₂H₅S CH ₂ COOH	-705.868082	-705.327110	333.19	-	-
11	C₃H₇S CH ₂ COOH	-745.158378	-744.617138	333.36	-	-
12	CH₃O CH ₂ COOH	-343.558246	-343.011665	336.71	336.0±2	-0.21
13	C₂H₅O CH ₂ COOH	-382.854279	-382.307287	336.97	335.0±2	-0.59
14	C₃H₇O CH ₂ COOH	-422.144214	-421.597353	336.89	-	-
15	NC SCH ₂ COOH	-719.520359	-719.006494	316.18	-	-
16	O₂N CH ₂ COOH	-433.579665	-433.060810	319.31	-	-
17	F CH ₂ COOH	-328.314745	-327.777455	330.88	331.3±2	0.46
18	F₂ CHCOOH	-427.572195	-427.044377	324.94	323.9±2	-0.31
19	F₃ CCOOH	-526.839225	-526.324345	316.82	316.5±2	-0.10
20	Cl CH ₂ COOH	-688.706405	-688.173726	327.99	329.2±2	0.37
21	CH ₃ (Cl)CHCOOH	-728.000753	-727.46597	329.31	330.6±2	0.40
22	CH ₃ CH ₂ (Cl)CHCOOH	-767.290636	-766.756166	329.11	330.9±2	0.53
23	Cl₂ CHCOOH	-1148.341766	-1147.819787	321.27	322.3±2	0.30
24	CH ₃ (Cl₂)CCOOH	-1187.630412	-1187.109674	320.49	-	-
25	Cl₃ CCOOH	-1607.967079	-1607.456525	314.10	-	-
26	Br CH ₂ COOH	-2802.778559	-2802.24762	326.89	327.8±2	0.26
27	CH ₃ (Br)CHCOOH	-2842.072948	-2841.539197	328.66	329.4±2	0.23
28	CH ₃ CH ₂ (Br)CHCOOH	-2881.362732	-2880.829285	328.47	330.1±2	0.51
29	CH ₃ CH ₂ CH ₂ (Br)CHCOOH	-2920.649944	-2920.116980	328.16	-	-
30	Br₂ CHCOOH	-5376.483998	-5375.963879	320.10	321.53±2	0.44
31	Br₃ CCOOH	-7950.177169	-7949.669354	312.38	-	-
32	NC CH ₂ COOH	-321.304950	-320.780980	322.52	323.7±2	0.36
33	CH₃CO CH ₂ COOH	-381.687265	-381.153842	328.45	-	-
34	CH ₃ CH ₂ (O=)CCOOH	-381.676571	-381.147512	325.71	-	-

1 Hartree/particle = 627.5095 kcal/mol
α-substituents are indicated with bold

Table 2. pK_a values were calculated using gas-phase Gibb's free energies (ΔG_{gas}^o) at MPW1PW91/6-311+g(2d,2p) level of theory in conjunction with C-PCM aqueous solvation energies (ΔG_{solv}^o) at HF and MPW1PW91 methods using 6-31g(d) and 6-31+g(d) basis sets

Comp. ID	ΔG_{gas}^o	hf/6-31g(d)			mpw1pw91/6-31+g(d)			pKa (exp.)
		ΔG_{solv}^o		pKa (cal.)	ΔG_{solv}^o		pKa (cal)	
		HA	A ⁻		HA	A ⁻		
1	340.92	-7.49	-77.55	5.03 (-0.27)	-7.37	-73.63	7.81 (-3.05)	4.76
2	340.36	-7.18	-76.11	5.45 (-0.58)	-7.47	-72.48	8.32 (-3.45)	4.87
3	339.56	-6.94	-75.04	5.47 (-0.59)	-6.87	-71.29	8.17 (-3.29)	4.88
4	338.31	-6.39	-72.98	5.66 (-0.61)	-6.30	-68.53	8.86 (-3.81)	5.05
5	337.85	-16.08	-84.41	4.05 (-0.22)	-16.28	-80.70	6.91 (-3.08)	3.83
6	338.02	-14.23	-82.56	4.17 (-0.30)	-14.41	-79.01	6.91 (-3.04)	3.87
7	335.07	-14.24	-81.35	2.90 (-0.55)	-14.72	-77.76	5.89 (-3.54)	2.35
8	331.29	-10.95	-73.04	3.81 (-0.14)	-10.87	-69.62	6.26 (-2.59)	3.67
9	333.15	-10.79	-74.54	3.96 (-0.24)	-10.58	-70.95	6.43 (-2.71)	3.72
10	333.19	-11.21	-75.08	3.90 (-0.16)	-11.05	-71.49	6.41 (-2.67)	3.74
11	333.36	-10.43	-74.24	4.07 (-0.30)	-10.27	-70.59	6.62 (-2.85)	3.77
12	336.71	-12.09	-80.45	3.19 (0.35)	-12.04	-76.33	6.17 (-2.63)	3.54
13	336.97	-12.01	-80.35	3.39 (0.21)	-12.07	-76.44	6.30 (-2.70)	3.6
14	336.89	-11.28	-79.55	3.38 (0.27)	-11.34	-75.62	6.31 (-2.66)	3.65
15	321.55	-12.46	-64.36	4.14 (-0.56)	-12.11	-60.42	6.77 (-3.19)	3.58
16	319.31	-17.84	-70.4	2.02 (-0.34)	-16.34	-66.50	3.78 (-2.10)	1.68
17	330.88	-11.01	-74.15	2.74 (-0.08)	-10.69	-70.10	5.47 (-2.81)	2.66
18	324.94	-10.39	-69.60	1.26 (0.02)	-10.06	-65.76	3.84 (-2.60)	1.24
19	316.82	-7.95	-63.29	-1.85 (2.08)	-7.62	-59.86	0.42 (-0.19)	0.23
20	327.99	-10.07	-70.26	2.78 (0.08)	-9.64	-66.42	5.28 (-2.42)	2.86
21	329.31	-9.32	-70.99	2.67 (0.16)	-8.98	-66.34	5.83 (-3.00)	2.83
22	329.11	-7.63	-68.63	3.01 (-0.13)	-7.30	-64.92	5.49 (-2.61)	2.88
23	321.27	-9.62	-65.22	1.23 (0.06)	-9.19	-61.73	3.47 (-2.18)	1.29
24	320.49	-7.79	-62.05	1.64 (0.20)	-8.27	-58.50	4.59 (-2.75)	1.84
25	314.10	-5.79	-57.42	-1.12 (1.77)	-5.44	-54.46	0.79 (-0.14)	0.65
26	326.89	-10.52	-69.48	2.88 (-0.02)	-10.05	-65.34	5.57 (-2.71)	2.86
27	328.66	-9.14	-69.17	3.39 (-0.39)	-8.72	-65.43	5.83 (-2.83)	3.00
28	328.47	-7.44	-67.68	3.10 (-0.16)	-7.31	-64.12	5.61 (-2.67)	2.94
29	328.16	-7.86	-67.98	2.96 (0.01)	-7.45	-64.23	5.41 (-2.44)	2.97
30	320.10	-10.32	-64.62	1.32 (0.16)	-9.88	-60.65	3.91 (-2.43)	1.48
31	312.38	-6.16	-56.28	-1.27 (1.93)	-5.84	-53.47	0.55 (0.11)	0.66
32	322.52	-14.75	-70.04	2.37 (0.07)	-14.10	-66.15	4.74 (-2.30)	2.44
33	328.45	-12.58	-71.90	3.76 (-0.18)	-11.49	-67.91	5.89 (-2.31)	3.58
34	325.71	-8.40	-71.00	-0.65 (3.15)	-7.55	-66.37	2.12 (0.38)	2.50
<i>rmse</i>	-	-	-	0.838	-	-	2.639	-

All the values of ΔG_{gas}^o and ΔG_{solv}^o are in kcal/mol

The values in parenthesis are residual errors of pK_a between experimental and calculated

Table 3. Calculated pK_a for α -substituted carboxylic acid on the best combination of gas-phase energies at MPW1PW91 /6-311+g(2d,2p) method and solvation energies at HF/6-31g(d) as well as at different regression methods

Comp ID	pK_a					^a Errors	^b Errors	^c Errors	^d Errors
	Cal	MLR	PCR	PLSR	Exp				
Training set									
1	5.03	4.69	4.82	4.84	4.76	-0.27	0.07	-0.06	-0.08
2	5.45	4.86	4.73	4.76	4.87	-0.58	0.01	0.14	0.11
3	5.47	4.83	4.62	4.65	4.88	-0.59	0.05	0.26	0.23
4	5.66	4.86	4.43	4.47	5.05	-0.61	0.19	0.62	0.58
5	4.05	3.79	4.07	4.03	3.83	-0.22	0.04	-0.24	-0.20
6	4.17	3.92	4.16	4.13	3.87	-0.30	-0.05	-0.29	-0.26
8	3.81	3.49	3.24	3.25	3.67	-0.14	0.18	0.43	0.42
9	3.96	3.66	3.53	3.54	3.72	-0.24	0.06	0.19	0.18
10	3.90	3.62	3.53	3.53	3.74	-0.16	0.12	0.21	0.21
11	4.07	3.73	3.57	3.58	3.77	-0.30	0.04	0.20	0.19
12	3.19	3.48	4.07	4.04	3.54	0.35	0.06	-0.53	-0.50
13	3.39	3.59	4.10	4.08	3.6	0.21	0.01	-0.50	-0.48
14	3.38	3.60	4.12	4.10	3.65	0.27	0.05	-0.47	-0.45
16	2.02	1.78	1.19	1.18	1.68	-0.34	-0.10	0.49	0.50
17	2.74	2.98	3.22	3.20	2.66	-0.08	-0.32	-0.56	-0.54
19	-1.85	0.24	1.31	1.27	0.23	2.08	-0.01	-1.08	-1.04
20	2.78	2.87	2.80	2.80	2.86	0.08	-0.01	0.06	0.06
21	2.67	2.92	3.04	3.04	2.83	0.16	-0.09	-0.21	-0.21
22	3.01	3.12	3.05	3.07	2.88	-0.13	-0.24	-0.17	-0.19
24	1.64	2.01	1.76	1.77	1.84	0.20	-0.17	0.08	0.07
25	-1.12	0.49	0.94	0.93	0.65	1.77	0.16	-0.29	-0.28
26	2.88	2.84	2.61	2.61	2.86	-0.02	0.02	0.25	0.25
27	3.39	3.21	2.92	2.93	3.00	-0.39	-0.21	0.08	0.07
28	3.10	3.13	2.96	2.98	2.94	-0.16	-0.19	-0.02	-0.04
29	2.96	3.03	2.90	2.92	2.97	0.01	-0.06	0.07	0.05
30	1.32	1.76	1.62	1.62	1.48	0.16	-0.28	-0.14	-0.14
31	-1.27	0.31	0.66	0.66	0.66	1.93	0.35	0.00	0.00
32	2.37	2.22	1.79	1.79	2.44	0.07	0.22	0.65	0.65
33	3.76	3.25	2.74	2.75	3.58	-0.18	0.33	0.84	0.83
Test set									
7	2.90	3.19	2.85	2.71	2.35	-0.55	-0.84	-0.50	-0.36
15	4.14	3.03	2.66	2.70	3.58	-0.56	0.55	0.92	0.88
18	1.26	2.01	2.37	2.10	1.24	-0.02	-0.77	-1.13	-0.86
23	1.23	1.81	1.83	1.83	1.29	0.06	-0.52	-0.54	-0.54
34	-0.65	1.98	2.63	2.58	2.50	3.15	0.52	-0.13	-0.08
<hr/>									
<i>rmse</i>	-	-	-	-	-	0.838	0.293	0.326	0.319
r^2	0.982	0.891	0.898	-	-	-	-	-	-

pK_a (Cal) values were calculated directly using Eq. 4 through thermodynamic cycle shown in Fig. 1

pK_a (MLR) values were calculated using Eq. 8

pK_a (PCR) values were calculated using Eq. 9

pK_a (PLS) values were calculated using Eq. 10

^aErrors, ^bErrors, ^cErrors and ^dErrors are residual errors for Cal, MLR, PCR and PLSR respectively

Table 4. Calculated and experimental pK_a with residual errors

Comp ID	$pK_a(cal)$	$pK_a(exp)$	Residual Errors
Class-A			
1	5.03	4.76	-0.27
2	5.45	4.87	-0.58
3	5.47	4.88	-0.59
4	5.72	5.05	-0.67
5	4.05	3.83	-0.22
6	4.17	3.87	-0.30
7	2.90	2.35	-0.55
8	3.81	3.67	-0.14
9	3.96	3.72	-0.24
10	3.90	3.74	-0.16
11	4.07	3.77	-0.30
12	3.19	3.54	0.35
13	3.39	3.6	0.21
14	3.38	3.65	0.27
15	4.14	3.58	-0.56
16	2.02	1.68	-0.34
17	2.74	2.66	-0.08
18	1.26	1.24	-0.02
20	2.78	2.86	0.08
21	2.67	2.83	0.16
22	3.01	2.88	-0.13
23	1.23	1.29	0.06
24	1.64	1.84	0.20
26	2.88	2.86	-0.02
27	3.39	3.00	-0.39
28	3.10	2.94	-0.16
29	2.96	2.97	0.01
30	1.32	1.48	0.16
32	2.37	2.44	0.07
33	3.76	3.58	-0.18
Class-B			
19	0.42	0.23	-0.19
25	0.79	0.65	-0.14
31	0.55	0.66	0.11
34	2.12	2.50	0.38
rmse	-	-	0.292
r^2			0.941

Class-A: All mono and di-substituted acid (at α -position)Class-B: Tri-halo and 2-oxo substituted acid (at α -position)

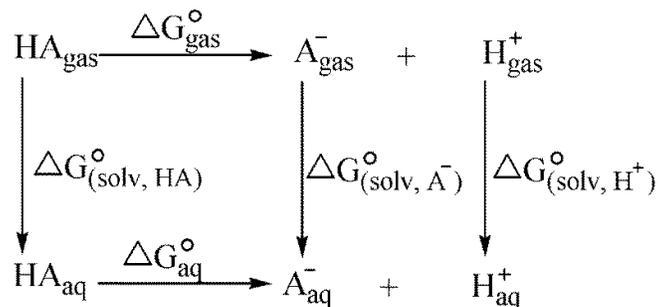


Fig. 1. Thermodynamic cycle used to calculate $\Delta G_{\text{aq}}^{\circ}$ of reaction (1) from its components.

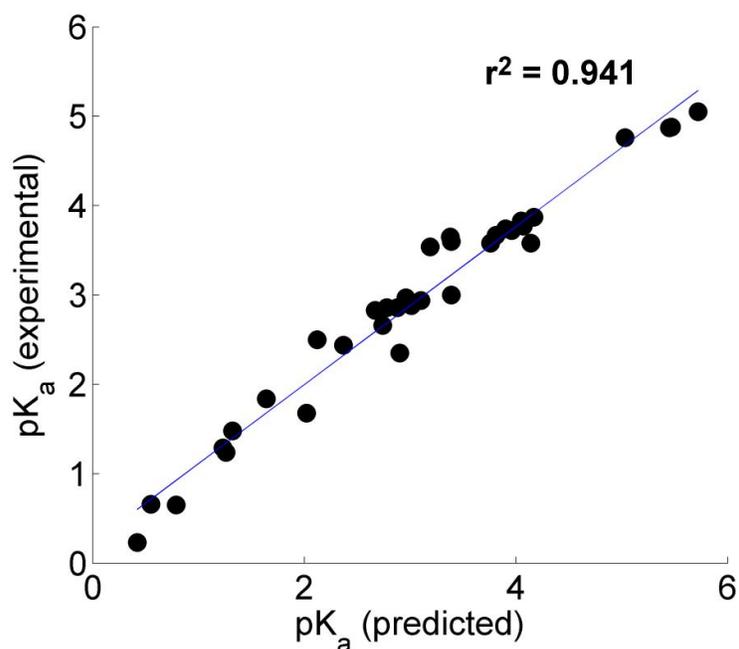


Fig. 2. Correlation between experimental and predicted pK_a s.

REFERENCES

- [1]. W. Q. Tong and G. Whitesell, *Pharm. Dev. Tech.* 3 (1998) 215.
- [2]. E. H. Kerns and L. Di, *Drug Discov. Today: Technol.* 1 (2004) 343.
- [3]. A. Avdeef, *Curr. Top. Med. Chem.* 1 (2001) 277.
- [4]. X. Xie, S. H. Steiner and M. H. Bickel, *Drug Metab. Dispos.* 19 (1991) 15.
- [5]. T. Jones and G. Taylor, *Proc. Eur. Congr. Biopharm. Pharmacokinet.* 3rd. 2 (1987) 181.
- [6]. G. M. Mitani, I. Steinberg, E. J. Lien, E. C. Harrison and U. Elkayam, *Clin. Pharmacokinet* 12 (1987) 253.
- [7]. D. T. Manallack, *Perspect. Med. Chem.* 1 (2007) 25.
- [8]. C. Cunchillos and G. Lecointre, *Biochimie* 89 (2007) 555.
- [9]. C. Lim, D. Bashford and M. Karplus, *J. Phys. Chem.* 65 (1991) 5610.
- [10]. C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.A* 110 (2006) 2493.

- [11]. J. Klicic, R. A. Friesner, S. Y. Liu and W. C. Guida, *J. Phys. Chem. A* 106 (2002) 1327.
- [12]. (a) Y. H. Jang, L. C. Sowers, T. Cagin and W. A. III Goddard, *J. Phys. Chem. A* 105 (2001) 274. (b) K. N. Rogstad, Y. H. Jang, L. C. Sowers and W. A. III Goddard, *Chem. Res. Toxicol.* 16 (2003) 1455. (c) S. Hwang, Y. H. Jang and D. S. Chung, *Bull. Korean Chem. Soc.* 26 (2005) 585.
- [13]. W. L. Jorgensen, J. M. Briggs and J. Gao, *J. Am., Chem. Soc.* 109 (1987) 6857.
- [14]. W. L. Jorgensen and J. M. Briggs, *J. Am. Chem. Soc.* 111 (1989) 4190.
- [15]. C. Lim, D. Bashford and M. Karplus, *J. Phys. Chem.* 95 (1991) 5610.
- [16]. M. J. Potter, M. K. Gilson and J. A. McCammon, *J. Am. Chem. Soc.* 116 (1994) 10298.
- [17]. E. Rajasekaran, B. Jayaram and B. Honig, *J. Am. Chem. Soc.* 116 (1994) 8238.
- [18]. K. Daasbjerg, *Acta Chem. Scand.* 49 (1995) 878.
- [19]. M. Namazian and S. Halvani, *J. Iran. Chem. Soc.* 2(1) (2005) 65.
- [20]. C. Lim, D. Bashford and M. Karplus, *J. Phys. Chem.* 85(14) (1991) 5610.
- [21]. M. D. Liptak and G. C. Shields, *J. Am. Chem. Soc.* 123 (2001) 7314.
- [22]. M. D. Liptak, K. C. Gross, P. G. Seybold, S. Feldgus and G. C. Shields, *J. Am. Chem. Soc.* 124(22) (2002) 6421.
- [23]. a) M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe and T. R. Tuttle, *J. Phys. Chem. A* 102(40) (1998) 7787. b) M. W. Palascak and G. C. Shields, *J. Phys. Chem. A* 108 (2004) 3692.
- [24]. M. J. Frisch *et al.*, Gaussian 03, Revision B.01, Gaussian Inc., Pittsburgh PA, 2003
- [25]. a) C. C. J. Roothan, *Rev. Mod. Phys.* 23 (1951) 69. b) J. A. Pople and R. K. Nesbet, *J. Chem. Phys.* 22 (1954) 571. c) R. McWeeny and G. Dierksen, *J. Chem. Phys.* 49 (1968) 4852.
- [26]. A. D. Becke, *J. Chem. Phys.* 98 (1993) 5648.
- [27]. a) K. Burke, J. P. Perdew and Y. Wang, in *Electronic Density Functional Theory: Recent Progress and New Directions*, Ed. J. F. Dobson, G. Vignale and M. P. Das, Plenum, 1998 b) J. P. Perdew, in *Electronic Structure of Solids '91*, Ed. P. Ziesche and H. Eschrig, Akademie Verlag, Berlin, 1991, p. 11. c) J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, *Phys. Rev. B* 46 (1992) 6671. d) J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, *Phys. Rev. B* 48 (1993) 4978. e) J. P. Perdew, K. Burke and Y. Wang, *Phys. Rev. B* 54 (1996) 16533. f) C. Adamo and V. Barone, *J. Chem. Phys.* 108 (1998) 664.
- [28]. a) M. Head-Gordon, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.* 153 (1988) 503. b) M. J. Frisch, M. Head-Gordon and J. A. Pople, *Chem. Phys. Lett.* 166 (1990) 275. c) M. J. Frisch, M. Head-Gordon and J. A. Pople, *Chem. Phys. Lett.* 166 (1990) 281. d) M. Head-Gordon and T. Head-Gordon, *Chem. Phys. Lett.* 220 (1994) 122. e) S. Saebo and J. Almlöf, *Chem. Phys. Lett.* 154 (1989) 83.
- [29]. P. Hohenberg and W. Kohn, *Phys. Rev.* 136 (1964) B864.
- [30]. S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.* 55 (1981) 117.

- [31]. a) V. Barone and M. Cossi, *J. Phys. Chem. A* 102 (1998) 1995. b) M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comp. Chem.* 24 (2003) 669.
- [32]. L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov and J. A. Pople, *J. Chem. Phys.* 109 (1998) 7764.
- [33]. J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.* 110 (1999) 2822.
- [34]. D. M. McQuarrie, *Statistical Mechanics*, Harper and Row: New York, 1970, p 86.
- [35]. D. M. Camaioni and C. A. Schwerdtfeger, *J. Phys. Chem. A* 109(47) (2005) 10795.
- [36]. C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B* 110(32) (2006) 16066.
- [37]. C. P. Kelly, C. J. Cramer and D.G. Truhlar, *J. Phys. Chem. B* 111 (2007) 408.
- [38]. V. S. Bryantsev, M.S. Diallo and W. A. Goddard, *J. Phys. Chem. B* 112(32) (2008) 9709.
- [39]. C. G. Zhan and D. A. Dixon, *J. Phys. Chem. A* 105(51) (2001) 11534.
- [40]. J. B. Cumming and P. Kebarle, *Can. J. Chem.* 56 (1978) 1.
- [41]. G. Caldwell, R. Renneboog and P. Kebarle, *Can. J. Chem.* 67 (1989) 661.
- [42]. M. Namazian and H. Heidary, *J. Mol. Struct.: Theochem* 620 (2003) 257.
- [43]. J. F. J. Dippy, S. R. C. Hughes and A. Rozanski, *J. Chem. Soc.* (1959) 2492.
- [44]. G. Schrüümann, M. Cossi, V. Barone and J. Tomasi, *J. Phys. Chem. A* 102 (1998) 6706.
- [45]. M. B. Smith, *Organic Chemistry, Acid-Base Approach*, Taylor and Francis group, CRS press, USA, 2011.
- [46]. a) N. H. Nelson and S. D. Faust, *Environ. Sci. Technol.* 3(11) (1969) 1186. b) R. Haque, P. C. Kearny and V. H. Freed, in *Pesticides in Aquatic Environments*, M. A. Q. Khan, Ed., Plenum Press, New York, 1976, p. 39. c) A. G. Hornsby, R. D. Wauchope and A. E. Herner, *Pesticide properties in the environment*, Springer-Verlag, Berlin, Germany, 1996
- [47]. a) M. A. K. Liton, M. I. Ali and M. T. Hossain, *Comput. Theor. Chem.* 999 (2012) 1. b) M. A. K. Liton, U. Salma and M. B. Hossain, *Am. J. Chem.* 3(3) (2013) 37.