Scientific paper

Relationships Between Aqueous Acidities of Benzene Polycarboxylic Acids and Computed Surface-electrostatic Potentials and Charges

Janez Cerar* and Črtomir Podlipnik

Faculty of Chemistry and Chemical Technology, University of Ljubljana, Aškerčeva 5, 1000 Ljubljana, Slovenia,

** Corresponding author: E-mail: janez.cerar@fkkt.uni-lj.si*

Received: 30-05-2008

Dedicated to the memory of Professor Ljubo Golič

Abstract

A correlation between calculated surface-electrostatic potential of benzene polycarboxylic acids and their conjugate bases on one side and experimental pK_a values on the other side was examined. It was found out that pK_a values for consecutive dissociation steps of polyprotic carboxylic acids are linearly dependent on the minimum of the surface-electrostatic potential of the conjugate bases obtained in the ionization step. Such a relation could be used in evaluation of pK_a values of a polyprotic acid from the titration data. The minimum of the surface-electrostatic potential of the conjugate base and the number of dissociated protons can be used also as two descriptors in the multiple regression model that predicts pK_a values for the whole family of benzene polycarboxylic acids. Replacing the surface-electrostatic potential of the conjugate base with the number of carboxyl groups in the molecule gives a multiple regression model whose results are even better correlated to experimental pK_a values but its use is less universal and, among other things, ignores occurrence of the intramolecular hydrogen bond. Lack of trustworthy experimental data was identified to be one of the bottlenecks in obtaining more reliable pK_a prediction models for polyprotic acids.

Keywords: Polyprotic acids, pK_a, surface-electrostatic potential, multiple regression model, benzene polycarboxylic acids

1. Introduction

Extensive use of computational chemistry in recent years offers opportunity to calculate by computer some of the physico-chemical properties of the compounds that are used in research, routine laboratory work, or in industry. One of such properties is certainly pK_a values for various organic acids. Although in everyday work experimental data are usually preferred over theoretical values because the later ones are less reliable, the experimental data are not always available. Another problem that is often encountered with the use of experimental pK_a values is their reliability and laboratory conditions in which they were obtained. Therefore it is of no surprise that, beside scientific curiosity and increased computer capabilities, more and more efforts are put in theoretical calculation of various physico–chemical properties.

Several programs that are capable of prediction of pK_a values exist.[1] They differ according to approach used in calculation of pK_a, complexity, computer time consumption, accuracy, price, etc. All these programs are learnt on some training sets and if the compound in question is similar to compounds from the training set, the predicted pK_a is usually close to the experimental one. While widely used chemicals and chemicals for which broad interest is shown (e.g. biological and drug-like molecules) are represented relatively well in these training sets, less attention is paid to polyprotic carboxylic acids. Two of limitations, responsible for such situation, are certainly a small number of polyprotic acids that could be used in such training sets, and questionable experimental pK_a values, found in literature. While experimental determination of pK_a values for monoprotic and diprotic acids is usually relatively simple this is less true for polyprotic acids where difficulties with obtaining pK_a rapidly increase with increased number of ionizable protons. In such cases consecutive protonation steps are more and more overlapped and increased number of ex-

perimental points in the titration curve offers only a limited improvement to the reliability of experimental pK_a values. Even more, experimental pK_a values are becoming more disputed with every further ionization step also from the theoretical point of view. While scientific community widely acknowledge that Debye-Hückel theory gives rather good predictions of activity coefficients only for the case of diluted solutions of 1–1 electrolytes, in the lack of more appropriate equations it still uses some form of Debye-Hückel equation for calculation of activity coefficients of ions bearing multiple elemental charge.[2,3]

For the case of polycarboxylic acids, where number of dissociating groups is counted in hundreds or thousands, alternative approach was proposed several decades ago by polyelectrolyte chemists.[4] It is based on the field effect, discovered by Bjerrum, that interprets the decrease in the dissociation constant of a polymeric acid with increasing ionization by the increased difficulty of removing a proton from the field of the negatively ionized carboxyl groups. Following this, K_ν, the ionization constant of (ν–1) times ionized molecule carrying P carboxyl groups, is a function of an intrinsic ionization constant K_0 and of a free energy change ΔG_ν accompanying the removal of one proton from the (ν–1) times ionized molecule[4.]

$$K_\nu = K_0 \frac{P - \nu + 1}{\nu} e^{-\frac{\Delta G_\nu}{RT}} \quad (1)$$

In the upper equation R is the gas constant, T is temperature and term $(P - \nu + 1)/\nu$ represents the statistical factor. Yet further, the free energy change for the removal of a proton (calculated to one mol of protons), carrying the elemental electric charge e_0, is directly related to the electrostatic potential ψ_0 at the site where the proton originates:

$$\Delta G_\nu = -e_0 N_A \psi_0 \quad (2)$$

while N_A stands for Avogadro's constant. In the upper equation ψ_0 is a function of ν and generally increases with the number of removed (dissociated) protons ν. These relations are still the most popular mode for interpreting potentiometric titration curves of polyelectrolytes.

Recently, Ma et al. examined relationships between aqueous acidities and computed surface-electrostatic potentials and local ionization energies of substituted phenols and benzoic acids.[5] More precisely, good correlations were found on one hand between pK_as and the most positive values of the electrostatic potential on the molecular surface, $V_{S,max}$, of the neutral acids and, on the other hand, between pK_as and the most negative values of the electrostatic potential on the molecular surface, $V_{S,min(cb)}$, and the minimum values of local ionization energy computed on the molecular surface, $I_{S,min}$, of the conjugate bases for both sets of molecules. In the paper of Ma et al. only singly charged substituted benzoic acids were investigated.

Our goal in this study is to extend research between pK_as and surface-electrostatic potential to polyprotic acids. As the model compounds the set of benzene polycarboxylic acids with known experimental pK_a values will be used. Semiempirical methods will be applied in the calculation of electrostatic potential. In parallel, a prediction power of MARVIN program as a rather accurate[1] representative of specialized programs for calculations of pK_a values will be examined for the case of polycarboxylic acids.

2. Methods

2. 1. Dataset

Dataset of studied acids consists of 12 acids from the family of carboxylic acids where carboxyl group is directly attached to the benzene ring, and their experimental pK_a values can be found in literature. To minimize other effects than the influence of removed protons and sterical effects compounds with additional functional groups were omitted from this study (Tables 1 and 2).

It has to be said that not all the data in the upper table are of the same quality. While, for example, the data from various literature sources for benzoic acid is quite consistent, the pK_a values for the acids with more than two ionizable protons are much less reliable and can differ considerably from one source to another. In the case of benzenehexacarboxylic acid, where rather extensive and well documented work is available, the recent data of Apelblat[3] were taken as the most reliable. For this, rather special case, a comparison with other data is made in the section Results and discussion. Unfortunately, for all delicate polyprotic acids the data are not documented so well. If several sources were available a consistency of data from these sources was considered as well as perceived reliability of the source although the alternative data might be more closely correlated with the model prediction.

2. 2. Prediction of Aqueous Ionization Constants with MARVIN Program

The MARVIN cheminformatics software, developed by ChemAxon[18], contains the module that predicts aqueous ionization constants of organic molecules on the basis of empirically calculated physico-chemical parameters that are obtained from ionization site-specific regression equations. Since a molecule with N ionizable sites can exists in 2^N possible microspecies, all of them have to be considered. In the first step the micro pK_a value is evaluated for every possible conjugated acid-base pair using empirical increments such as increments of partial charges, polarizability, and structure specific increments. In the case of monoprotic molecules micro pK_a is just equal to macro pK_a which is the final result of calculation. For

Table 1. Dataset of acids considered in this study. The cited pK_a values were taken from the literature. In some cases the temperature for which pK_a values are valid is not written explicitly although it can be from the cited source implicitly understood that it is valid in the range of room temperatures if not even at 25 °C. Where several sources are available for the same compound only the data from the first row were used in statistical analysis.

compound	literature data	temperature [°C]	pK_1	pK_2	pK_3	pK_4	pK_5	pK_6
benzoic acid	Ref.[6, 7]	25	4.204	–	–	–	–	–
1,2-benzenedicarboxylic acid	Ref.[6]	25	2.943	5.432	–	–	–	–
(phthalic acid)	Ref.[8]	25 (?)	2.89	5.51				
	Ref.[9]	25	2.75	5.41				
1,3-benzenedicarboxylic acid	Ref.[6]	25	3.70	4.60	–	–	–	–
(isophthalic acid)	Ref.[7]	25	3.62	4.60				
	Ref.[8]	25 (?)	3.54	4.60				
	Ref.[10]	25	3.50	4.50				
1,4-benzenedicarboxylic acid	Ref.[6]	25	3.54	4.34	–	–	–	–
(terephthalic acid)	Ref.[7]	25	3.54	4.46				
	Ref.[8]	25 (?)	3.51	4.82				
	Ref.[11]	25	3.52	4.46				
1,2,3-benzenetricarboxylic acid	Ref.[12]	25	2.80	4.20	5.87	–	–	–
(hemimellitic acid)	Ref.[7,10,13]	25	2.88	4.75	7.13			
	Ref.[8]	25 (?)	2.98	4.25	5.87			
1,2,4-benzenetricarboxylic acid	Ref.[7,12,13]	25	2.52	3.84	5.20	–	–	–
(trimellitic acid)	Ref.[8]	25 (?)	2.40	3.71	5.01			
	Ref.[10]	25	2.48	4.04	5.54			
1,3,5-benzenetricarboxylic acid	Ref.[14]	25	3.17	3.96	4.70	–	–	–
(trimesitic acid)	Ref.[12]	25	3.12	3.89	4.70			
	Ref.[8]	25 (?)	3.16	3.98	4.85			
	Ref.[10]	25	3.12	4.10	5.18			
	Ref.[7,13(a)]	25	2.12	4.10	5.18			
1,2,3,4–benzenetetracarboxylic acid (mellophanic acid)	Ref.[7,12]	25	2.06	3.25	4.73	6.21	–	–
1,2,3,5–benzenetetracarboxylic acid (prehnitic acid)	Ref.[7,12]	25	2.38	3.51	4.44	5.81		
1,2,4,5–benzenetetracarboxylic acid	Ref.[15]	25 (?)	1.92	2.82	4.49	5.64	–	–
(pyromellitic acid)	Ref.[7]	25	1.92	2.87	4.49	5.63		
	Ref.[8]	25 (?)	2.43	3.13	4.44	5.61		
	Ref.[16]	25 (?)	1.87	2.72	4.30	5.52		
	Ref.[10,13]	25	1.70	3.12	4.92	6.23		
benzenepentacarboxylic acid	Ref.[12]	25	1.80	2.73	3.97	5.25	6.46	–
	Ref.[7]	25	1.80	2.73	3.96	5.25	6.46	
	Ref.[8]	25 (?)	2.34	2.95	3.94	5.07	6.25	
benzenehexacarboxylic acid	Ref.[3(b)]	25	0.68	2.21	3.40	5.09	6.32	7.49
(mellitic acid)	Ref.[7,10(c)]	25	0.68	2.21	3.52	5.09	6.32	7.49
	Ref.[10(d)]	25	1.21	2.19	3.53	5.09	6.31	7.45
	Ref.[8]	25 (?)	2.08	2.46	3.24	4.44	5.50	6.59
	Ref.[12]	25	1.40	2.19	3.31	4.78	5.89	6.96
	Ref.[12(e)]	25	1.31	2.39	3.57	5.08	6.31	7.46
	Ref.[17]	25	0.636	1.66	2.25	3.32	4.07	5.03

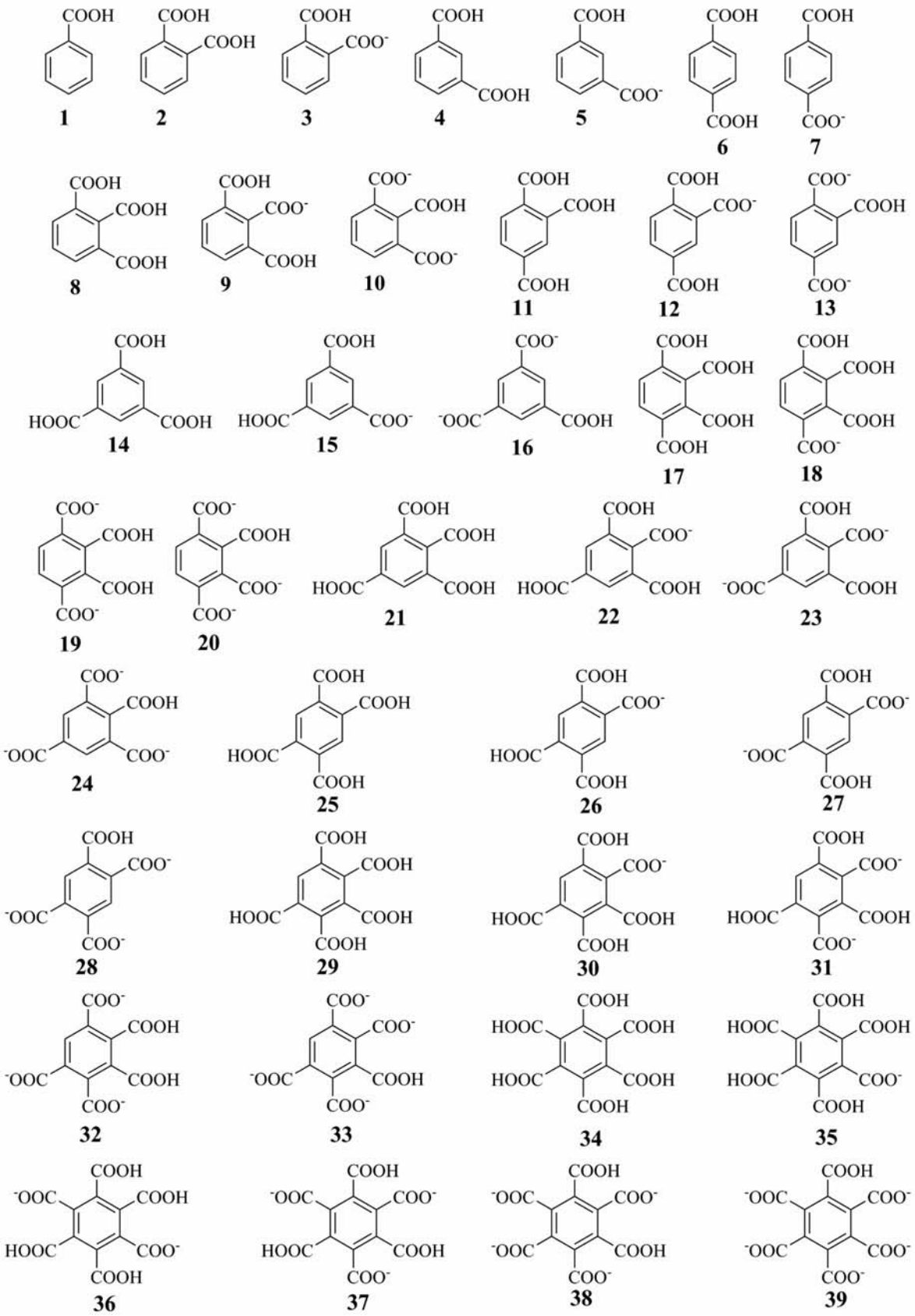
[(a)] the value of pK_1 in Ref.[7] was probably mistyped from Ref.[10] and consequently taken from Ref.[7] in Ref.[13]
[(b)] constants are obtained from the conductivity data using constants from Ref.[10] as a point of departure
[(c)] obtained using a Harned cell arrangement
[(d)] obtained using a glass electrode-calomel electrode assembly
[(e)] experimental data from Ref.[12] as recalculated in Ref.[10]

molecule that has more than one ionizable proton (polyprotic molecule), the program calculates macro pK_a according to the global mass and charge conservation law. Calculated values for the compounds considered in this paper are presented in Table 3.

2. 3. Semiempirical Calculations

For the generation of different ionization states of benzene polycarboxylic acids in pH range between 0 and 14 as well as for subsequent initial minimization of obtained structures with OPLSAA force field[19] Schrödinger

Table 2. Set of benzene carboxylic acids used in this study. For every acid only the most stable prototropic tautomer (obtained on the basis of energy calculations) is shown.



LigPrep utility[20] was used. These structures were written into input files for all semiempirical calculations that were performed with Spartan'04 program[21] at the PM3 level[22] with SM5.4 model[23] to account for solvation effects. Descriptors that we have tried to use in building pK_a predictivity model for benzene polycarboxylic acids were the maximum and the minimum values of electrostatic potential on isoelectronic density surface of 0.002 e bohr^{-3} of the acid and of its conjugate base, respectively.

In cases when multiple structures of the same compound with the same net charge were possible, Boltzmann population analysis was used to calculate the average value of electrostatic potential descriptors taking into account all possible proton tautomer forms.

3. Results and Discussion

One of the problems that are met when one is seeking for pK_a values of organic acids which are not very common is reliability of experimental pK_a data. This difficulty is both evidently and exemplarily represented by the case of benzenehexacarboxylic acid, a highly symmetric six protic carboxylic acid for which pK_a data from various authors are available.

In Figure 1 experimental pK_a values from several sources and the values calculated by the MARVIN program are plotted as a function of the dissociation step of benzenehexacarboxylic acid.

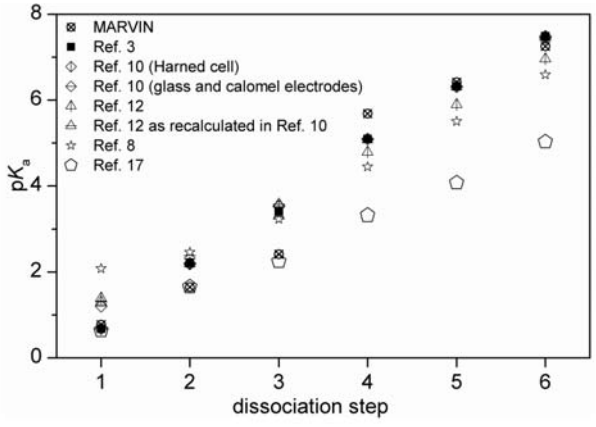


Figure 1. Comparison of published experimental pK_a values for six protic benzenehexacarboxylic acid. Values, calculated by the one of commercial programs (MARVIN), are added for comparison.

As one can see, the span of experimental values of pK_a values for the given dissociation step is between 0.8 for the removal of the second and 2.46 for the last proton, respectively. The differences among pK_a values from the different sources can be, as already noted by Beutler and Stebler[17] and later restated by Apelblat et al.,[3] attributed to the different ionic strength of solutions during titration experiments.

Regarding pK_a values, obtained by MARVIN, it is evident that theoretical numbers for the case of benzenehexacarboxylic acid lie more or less within limits of the experimental values with the exception for the pK_a value of the fourth dissociation step. Because the program takes into account number of microspecies present and calculates macro ionization constants, these values are directly comparable to experimental literature values. The best agreement of MARVIN's pK_a values for benzenehexacarboxylic acid is found with the recent data of Apelblat.[3] Considering care and plentifulness of experimental work done in the study of Apelblat et al. as well as the low concentrations of acid used in this research, this data are probably the most relevant for judging theoretical values and will be therefore used in these comparisons.

The next feature that can be noticed in Figure 1 is relatively large difference between pK_3 and pK_4 values, both calculated by MARVIN. This step apparently divides six pK_a values into two subsets of three points with approximately the same linear dependency of pK_a values on the number of removed protons ν. Because the program does not offer opportunity to get deeper insight into calculation of given pK_a value the reason for such a shift can not be identified without further information. Contrary to this, in all sets of experimental pK_a values rather constant increase of pK_a values can be found.

Since MARVIN program (as well as similar programs) is parameterized on the basis of a certain group of compounds that are well characterized and possibly widely used, it is expected that it would predict rather accurately pK_a values of benzoic and all three benzenedicarboxylic (= phtalic) acids. In Figure 2 a comparison between experimental pK_as and MARVIN's predictions is made for the

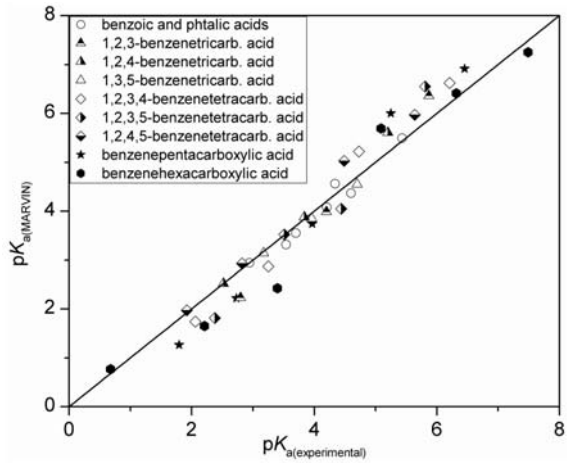


Figure 2. Comparison between calculated pK_a values, calculated by the MARVIN program, and experimental data. Data points that belong to different compounds are denoted distinctively. Benzoic and three phtalic acids are represented by the same symbol. The plotted line indicates the ideal match between calculated and experimental pK_as.

whole set of benzene (poly)carboxylic acids by distinct designation of data points belonging to different compounds.

As can be seen, predicted pK_as of mono- and dicarboxylic acids are all very close to the literature values. In the case of acids with more than two carboxyl groups only predicted pK_as for 1,3,5-benzenetricarboxylic acid are practically identical to experimental data. It is interesting that practically all the calculated pK_as with the value below 4.4 are lower than the corresponding experimental values and, similarly, practically all the calculated pK_as with the value higher than 4.4 are overestimated. Nevertheless, a great advantage of this program is that the results are quickly obtained and that they are good enough to be used when semiquantitative estimations are needed.

Both straightforward relationships of the surface-electrostatic potentials with pK_as are poor and of approximately same quality, the one showing $V_{S,min(cb)}$ as a function of pK_a hoving slightly more regular structure. Because it turned out that the linear relationship between $V_{S,max}$ and pK_a is a little bit worse also when this correlation was sought for every single compound included in this research, our attention will be focused to relationship between $V_{S,min(cb)}$ and pK_a. To start examining this correlation, a graph of the most negative value of the surface-electro-

Table 3. Overview of MARVIN program predictions (p$K_{a(MARVIN)}$), and the most positive values of the surface-electrostatic potential of the dissociating acidic species ($V_{S,max}$) as well as the most negative values of the surface-electrostatic potential of the conjugate bases ($V_{S,min(cb)}$) obtained in the dissociating process (the later two were calculated by SPARTAN program). In the last column pK_a values (p$K_{a(model)}$), obtained on the basis of the multiple regression model (Equation 3), are presented.

compound	**dissociation step (= number of) removed protons)**	**p$K_{a(MARVIN)}$**	**$V_{S,max}$ (kcal/mol)**	**$V_{S,min(cb)}$ (kcal/mol)**	**p$K_{a(model)}$**
benzoic acid	1	4.08	28.46	–197.53	3.817
1,2-benzenedicarboxylic acid	1	2.94	31.94	–184.59	2.867
(o-phtalic acid)	2	5.49	–36.99	–266.57	4.979
1,3-benzenedicarboxylic acid	1	3.55	31.43	–195.818	3.691
(m-phtalic acid)	2	4.37	–20.64	–255.71	4.182
1,4-benzenedicarboxylic acid	1	3.32	31.69	–192.74	3.466
(p-phtalic acid)	2	4.56	–16.93	–253.22	4.000
1,2,3-benzenetricarboxylic acid	1	2.23	34.25	–175.23	2.181
	2	3.99	–24.10	–254.63	4.103
	3	6.36	–93.16	–335.55	6.136
1,2,4-benzenetricarboxylic acid	1	2.52	34.60	–182.30	2.700
	2	3.88	–13.83	–249.12	3.698
	3	5.61	–94.61	–320.76	5.051
1,3,5-benzenetricarboxylic acid	1	3.14	35.22	–189.44	3.224
	2	3.85	–15.82	–250.66	3.812
	3	4.55	–67.57	–310.84	4.324
1,2,3,4-benzenetetracarboxylic acid	1	1.74	36.48	–184.40	2.854
	2	2.87	–12.79	–241.35	3.129
	3	5.22	–80.04	–317.54	4.815
	4	6.62	–142.81	–394.54	6.561
1,2,3,5-benzenetetracarboxylic acid	1	1.81	36.42	–183.69	2.802
	2	3.52	–11.13	–242.06	3.181
	3	4.05	–78.68	–302.18	3.688
	4	6.55	–137.09	–388.87	6.145
1,2,4,5-benzenetetracarboxylic acid	1	1.97	36.35	–174.06	2.095
	2	2.93	–10.02	–237.29	2.831
	3	5.03	–98.38	–312.25	4.427
	4	5.97	–182.52	–375.53	5.167
benzenepentacarboxylic acid	1	1.26	40.37	–168.54	1.690
	2	2.22	–9.49	–233.41	2.546
	3	3.74	–71.78	–313.32	4.505
	4	6.00	–132.08	–383.22	5.731
	5	6.91	–214.94	–451.52	6.838
benzenehexacarboxylic acid	1	0.77	39.82	–165.38	1.458
(mellitic acid)	2	1.65	–7.73	–232.14	2.453
	3	2.42	–63.41	–304.56	3.863
	4	5.69	–118.63	–378.25	5.366
	5	6.41	–198.24	–443.28	6.234
	6	7.25	–296.09	–507.74	7.060

static potential of the conjugate bases against pK_a is plotted and a distinct designation is used for the conjugate bases with a different number of removed protons (Figure 3).

From such a picture it is easily to recognize that the relationship between calculated surface-electrostatic potentials and pK_a values still exists but it has to be judged separately for every dissociation step. The slopes of the apparent lines through data of the groups with the same number of removed protons (due to a limited number of polyprotic acids with the number of ionizable protons higher than three this can be recognized only for the first three dissociation steps) seems to be very similar. Even more, these groups are rather regularly separated one from

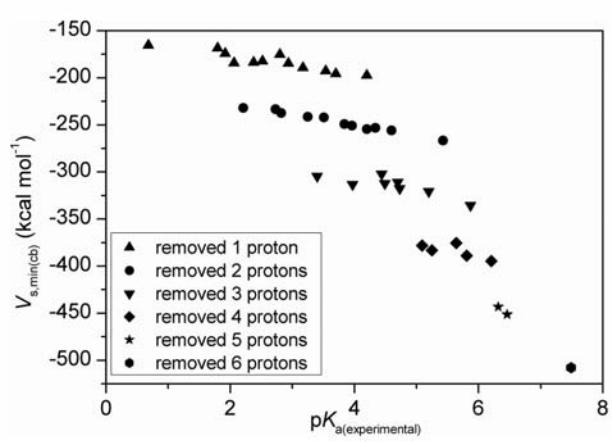


Figure 3. Plot of calculated minimum values of the surface-electrostatic potential of the conjugated bases as a function of experimental pK_a values. Points in the graph are grouped according to the charge of conjugate bases (= number of removed protons).

another by the same value. Therefore, to improve the correlation between modeled pK_a values and the experimental data we referred to the multiple linear regression, using the minimum value of the surface-electrostatic potential of the conjugate bases and the number of removed protons as two descriptors.

The corresponding correlation of the minimum values of the surface-electrostatic potential of the conjugate bases $V_{S,min(cb)}$ and the number of removed protons ν with the pK_a can be represented by equation

$$pK_a = (-6.77 \pm 0.76) - (3.90 \pm 0.43) \cdot \nu - \\ - (0.0734 \pm 0.0065) \frac{mol}{kcal} \cdot (V_{S,min(cb)}) \qquad (3)$$

$n = 39$, $R = 0.969$, $s = 0.38$, $F = 278$

and it is graphically shown in Figure 4.

Again, similar to the case of MARVIN program, predicted pK_as of mono- and dicarboxylic acids are quite close to the literature values although agreement is a bit

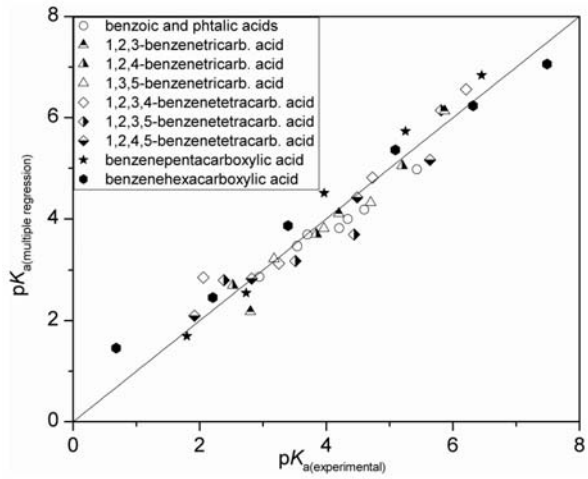


Figure 4. Comparison of pK_as obtained from the multiple regression analysis and experimental pK_as. Two parameters, used in regression analysis, are the minimum values of the surface-electrostatic potential of the conjugate bases and the number of removed protons. The plotted line represents ideal correlation.

worse as in the former case. For the majority of the acids with three or more ionizable protons the concordance was improved. This is probably not a coincidence and can be partly attributed to the fact that MARVIN program is most probably not parametrized with the dataset of benzene polycarboxylic acids with more than three ionizable protons.

When physical interpretation of Equation 3 is considered it is obvious that the negative sign of term containing number of removed protons ν is in contradiction with Figure 3. It is also in disagreement with conclusion made from Equations 1 and 2: increased number of removed protons strengthens the electrostatic field around polyion thus increases the free energy required for removing further protons away from the polyion. Evidently, the correlation between $V_{S,min(cb)}$ and ν (visible in Figure 3) is too high to give physically meaningful results for the family of benzene polycarboxylic acids.

A close inspection of the correlation between $V_{S,min(cb)}$ and pK_a for the compounds with the same number of removed protons and use of a distinct designation for compounds with a different number of carboxyl groups (in Figure 5 such a correlation is shown for the case of singly charged acids) reveals that $V_{S,min(cb)}$ generally increases and pK_a generally decreases with the number of present carboxyl groups. These observations are hardly a surprise if a recent finding of Ma et al. that pK_a values increase as $V_{S,min(cb)}$ of benzoic acids decrease[5] and remark of Maxwell and Partington that one carboxyl can promote the ionization of another by several mechanisms[12] are taken into account. In addition, lower pK_a values for the same ionization step are expected for the acids bearing higher number of carboxyl groups also due to the statistical factor (Equation 1).

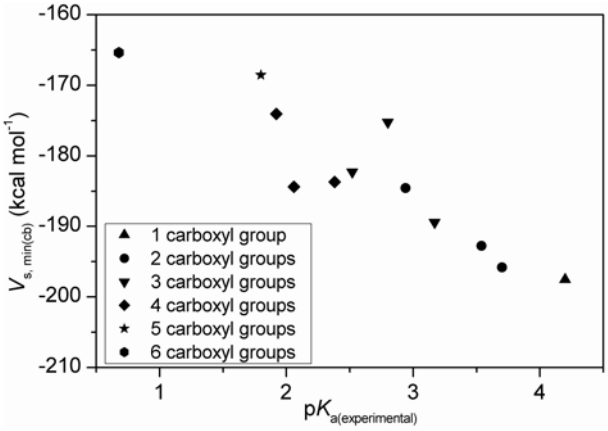


Figure 5. Correlation between pK_a and the computed $V_{S,min(cb)}$ of the conjugated bases of the singly charged benzene polycarboxylic acids. The data, belonging to the acids with the different number of carboxyl groups *P* are designated distinctly.

From Figure 5 can be seen that decreasing of pK_a values is rather well correlated with the increased number of carboxyl groups in the molecule. The only exception is 1,2-benzenedicarboxylic acid where favored formation of intramolecular hydrogen bond between two carboxyl groups in *ortho* position additionally promotes dissociation of the first proton. Contrary to the number of carboxyl groups, calculated $V_{S,min(cb)}$ correctly predicts lowering of pK_a of singly charged 1,2-benzenedicarboxylic acid. On the other hand $V_{S,min(cb)}$ is less successful in predicting first pK_a value of 1,2,3-benzenetricarboxylic acid and 1,2,3,4-benzenetetracarboxylic acid.

As far as the correlation between pK_a and $V_{S,min(cb)}$ for the data presented in Figure 5 is considered, it can be described by the equation

$$pK_a = (-12.9 \pm 2.2) - (0.085 \pm 0.012)\frac{mol}{kcal} \cdot \cdot (V_{S,min(cb)}) \quad (4)$$

$n = 12$, $R = -0.910$, $s = 0.42$, $F = 48$

which agrees with the one given by Ma et al. for the substituted benzoic acids[5] within error limits.

The correlation between pK_a and $V_{S,min(cb)}$ is the best for the doubly charged benzene polycarboxylic acids (see Figure 3). In this case the corresponding relationship can be given as

$$pK_a = (-17.7 \pm 1.1) - (0.0868 \pm 0.0045)\frac{mol}{kcal} \cdot (V_{S,min(cb)}) \quad (5)$$

$n = 11$, $R = -0.988$, $s = 0.15$, $F = 379$

and the factor representing dependence of pK_a on $V_{S,min(cb)}$ is again quite concordant with the one obtained by Ma et al. (–0.09092 ± 0.00836).[5] Obtaining reliable relationships between p*K*a and $V_{S,min(cb)}$ for further dissociation steps is quite difficult since the number of representative data is rapidly diminishing with every proton removed, and, additionally to this, the quality of available data is becoming more and more questionable.

Having all this in mind it is tempting to test correlation of pK_a with descriptors ν (the number of removed protons) and *P* (the number of carboxyl groups attached to the molecule). The result is shown in Figure 6.

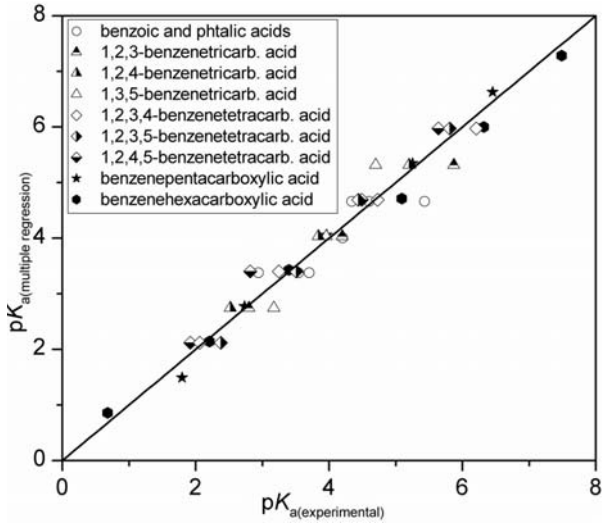


Figure 6. Comparison of pK_as obtained from the multiple regression analysis and experimental pK_as. Two parameters, used in regression analysis, are the number of removed protons and the total number of carboxyl groups in the molecule. The plotted line represents ideal correlation.

The equation, characterizing this correlation, is expressed as

$$pK_a = (3.35 \pm 0.15) + (1.285 \pm 0.043) \cdot \nu - (0.630 \pm 0.042) \cdot P \quad (6)$$

$n = 39$, $R = 0.980$, $s = 0.30$, $F = 448$

and has a better prediction power than all preceding ones. Physical interpretation of the meaning of ν in the upper equation is that it appraises the strength of the "global" electrostatic field around molecules, formed due to the excess of negative charge in the molecules as a consequence of removed protons while *P* is a measure for both of the statistical factor and of the fine electrostatic structure. The predicted pK_a that deviates the most from the experimental data (4.66 against 5.49) is pK_a value of the second dissociation step of 1,2-benzenedicarboxylic acid. The reason is, as already discussed, formation of intramolecular hydrogen bond that impedes dissociation of the second proton, trapped between two vicinal carboxyl groups. Si-

milarly, the calculated pK_a value for the first dissociation step for the same compound (3.38) is also one of the values that depart the most from experimental data.

Since values of v and P can be only discrete and have a limited range, the predicted pK_a values for different compound often coincide. Although a replacement of P as a descriptor with a logarithm of statistical factor from Equation 1 (= $log((P - v + 1)/v)$) offers more possible values and thus more fine description of the relationship, the correlations of pK_a on v and $log((P - v + 1)/v)$ is considerably worse than that in Equation 6. This implies that the hypothesis used in evaluation of the statistical factor (physico-chemical equivalence of carboxyl groups) is of quite limited significance for use in description of pK_a in benzene polycarboxylic acids and emphasizes a need to use in evaluation of pK_a at least one descriptor that is related to the strength of the local electrostatic field.

As it was shown above, a simple linear relation between $V_{S,min(cb)}$ and pK_a values exists also in the case of benzene polycarboxylic acids, however, this is true only for the acids bearing the same electric charge. On the other hand it is interesting that a linear relationship between

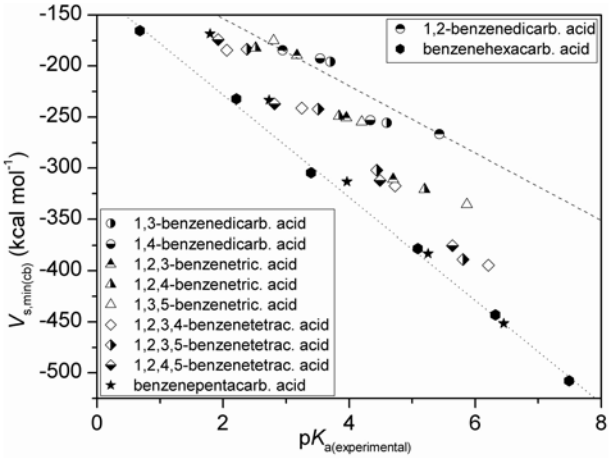


Figure 7. Plot of calculated minimum values of the surface-electrostatic potential of the conjugated bases as a function of experimental pK_a values. Points, belonging to the different compounds, are designated distinctly. In the graph two lines are added: the dotted one represents linear relation between $V_{S,min(cb)}$ and pK_a for the case of benzenehexacarboxylic acid while the dashed one characterizes the same relation for the case of 1,2-benzenedicarboxylic acid. See text.

$V_{S,min(cb)}$ and consecutive pK_a values of the same acid exists for all benzene polycarboxylic acids. The slope $\Delta V_{S,min(cb)}/\Delta pK_a$ is rather similar for all benzene polycarboxylic acids. Lines, representing this correlation (Figure 7), are plotted only for benzenehexacarboxylic acid as a representative with the highest number of dissociation steps, and for 1,2-benzenedicarboxylic acid for which the slope characterizing this relationship deviates the most from the "average" slope for the benzene polycarboxylic acids. This is probably not a coincidence since the formation of the intramolecular hydrogen bond[24,25] has a significant influence on dissociation constants.

5. Conclusions

Use of public (commercial or non-commercial) computer programs offers an elegant way to obtain estimates of pK_a values for compounds for which experimental data are not available. These predictions are usually good when compound considered has a structure similar to the set of chemical substances whose experimental data were used in the evaluation of the program. When compound is not so common and differs considerably from the molecules in the training set agreement is generally less satisfactory. In such cases more accurate estimates of pK_as may be found by focusing on the set of similar compounds and examining their molecular features. For the prediction of pK_a values of benzene polycarboxylic acids two multiple regression models that both apply two descriptors were established. The first model uses as descriptors the number of dissociated protons and the calculated surface-electrostatic potential minimum of the conjugate base of the dissociating acid while the second one predicts pK_a values on the bases of the number of dissociated protons and the number of carboxyl groups attached to the benzene ring. Although the obtained multiple regression models might be of limited use, found linear relationship between $V_{S,min(cb)}$ and consecutive pK_a values of the same polycarboxylic acid may be very useful for determination of pK_a values of polyprotic acids e.g. from titration curves. Such an evaluation of twelve pK_a values of dodecaprotic fullerenehexamalonic acid[24] is currently under way.

6. Acknowledgment

This work was supported by the Slovenian Research Agency (ARRS) through Physical Chemistry Research Program P1-0201, and Research Project J1-6653.

7. References

1. M. Meloun, S. Bordovska, *Anal. Bioanal. Chem.* **2007,** *389,* 1267–1281.
2. D. Litchinsky, N. Purdie, M. B. Tomson, W. D. White, *Anal. Chem.* **1969**, *41*, 1726–1730.
3. A. Apelblat, M. Bešter-Rogač, J. Barthel, R. Neueder, *J. Phys. Chem. B* **2006**, *110*, 8893–8906.
4. A. Katchalsky, J. Gillis, *Rec. Trav. Chim. Pays-Bas* **1949**, *68*, 879–897.
5. Y. G. Ma, K. C. Gross, C. A. Hollingsworth, P. G. Seybold, J. S. Murray, *J. Mol. Model.* **2004**, *10*, 235–239.
6. D. R. Lide (Ed.): CRC Handbook of Chemistry and Physics, CRC Press, Inc., Boca Raton, **2007–2008**, p. 2640.

7. J. G. Speight, Lange's Handbook of Chemistry*;* McGraw-Hill*,* New York, **2005**, pp 2.627–2.628.
8. J. Rudat, Thesis "*Enzymatische Decarboxylierung von Benzenpolycarbonsäuren"*, Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn, 2006.
9. United Nations Environment Programme Chemical Branch, OECD Initial Assessment Reports For High Production Volume Chemicals including Screening Information Datasets (SIDS), http://www.chem.unep.ch/irptc/sids/OECDSIDS/85449.pdf, (accessed: 31. March 2008).
10. N. Purdie, M. B. Tomson, N. Riemann, *J. Solut. Chem.* **1972**, *1*, 465–476.
11. United Nations Environment Programme Chemical Branch, OECD Initial Assessment Reports For High Production Volume Chemicals including Screening Information Datasets (SIDS), http://www.chem.unep.ch/irptc/sids/OECDSIDS/100-21-0.pdf, (accessed: 31. March 2008).
12. W. R. Maxwell, J. R. Partington, *Trans. Faraday Soc.* **1937**, *33*, 670–678.
13. S. Vairam, S. Govindarajan, *Thermochim. Acta* **2004**, *414*, 263–270.
14. A. Apelblat, *J. Mol. Liq.* **2002**, *95*, 99–145.
15. K. Ohta, K. Tanaka, *J. Chromatogr. A* **1998**, *804*, 87–93.
16. E. G. Zhizhina, V. F. Odyakov, K. I. Matveev, *React. Kinet. Catal. Lett.* **2007**, *91*, 325–332.
17. A. Beutler, A. Stebler, *Chimia* **1977**, *31*, 179–180.
18. J. Szegezdi, F. Csizmadia, Prediction of dissociation constant using microconstants, http://www.chemaxon.com/conf/Prediction_of_dissociation_constant_using_microconstants.pdf, (accessed: 31 March 2008).
19. W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
20. *LigPrep* (software), Schrödinger, LLC, New York, NY, 2005.
21. "*Spartan ´04"* (software), Wavefunction Inc., Irvine, CA, 2003.
22. J. J. P. Stewart, *J. Comput. Chem.* **1989**, *10*, 209–220.
23. C. C. Chambers, G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem.* **1996**, *100*, 16385–16398.
24. J. Cerar, J. Cerkovnik, J. Škerjanc, *J. Phys. Chem. B* **1998**, *102*, 7377–7381.
25. J. Cerar, J. Škerjanc, *Acta Chim. Slov.* **1998**, *45*, 103–110.

Povzetek

Preiskovali smo zvezo med elektrostatskim potencialom na površini molekul benzenovih (poli)karboksilnih kislin (oziroma njihovih konjugiranih baz) in v literaturi navajanimi eksperimentalnimi pK_a vrednostmi teh spojin. Ugotovili smo, da obstaja linearna zveza med pK_a vrednostmi zaporednih disociacijskih stopenj benzenovih polikarboksilnih kislin in minimumi elektrostatskega potenciala na površini v disociacijskih procesih nastalih konjugiranih baz. Tovrstna korelacija lahko olajša določevanje pK_a vrednosti poliprotičnih kislin iz titracijskih krivulj. Nadalje, minimum elektrostatskega potenciala na površini kislini konjugirane baze in število disociiranih protonov se lahko uporabita kot dva deskriptorja v multivariantnem regresijskem modelu, ki napoveduje pK_a vrednosti za celotno skupino benzenovih (poli)karboksilnih kislin. Če v tej zvezi namesto minimuma elektrostatskega potenciala uporabimo kot deskriptor kar število karboksilnih skupin v molekuli, je korelacija za navedeni primer še boljša, a manj splošna, saj ne upošteva tvorbe eventualnih intramolekularnih vodikovih vezi. Pomanjkanje zanesljivih eksperimentalnih podatkov o pK_a vrednostih poliprotičnih kislin je eden od vzrokov, ki omejujejo izdelavo bolj zanesljivega modela za napoved pK_a vrednosti poliprotičnih kislin.