Russian Journal of General Chemistry, Vol. 73, No. 11, 2003, pp. 1792–1798. From Zhurnal Obshchei Khimii, Vol. 73, No. 11, 2003, pp. 1893–1899.

Improved QSAR Analysis of the Toxicity of Aliphatic Carboxylic Acids[1]

F. P. Maguna, M. B. Núñez, N. B. Okulik, and E. A. Castro

Facultad de Agroindustrias, Universidad Nacional del Nordeste, Pcia, Roque Sáenz Peña, Chaco, Argentina
CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Argentina

Received June 4, 2002

Abstract — The results of a QSPR study of the toxicity of carboxylic acids in aqueous solution are reported. The molecular set comprises 35 carboxylic acids with the corresponding pK_a values in water. The set of molecular and topological parameters includes electrotopological state of the carboxy and methyl groups, molar refractivity, refractive index, *n*-octanol–water partition coefficient $\log K_{o/w}$, surface tension, and polarizability. Quite reasonable estimates are obtained, which improve the results of previous theoretical calculations.

Since scientists began to measure physical and biological properties of the natural world, they also sought for a pattern of relationships between the measurements they made. However, it was not until 1930s that knowledge of the extent and rates of chemical processes, together with properties of the reacting molecules (shape, size, and electronic properties) allowed correlations to be made between the nature of molecules and their reactivity. Surprisingly, even though similar types of measurements were possible in the biological world, particularly related to drug potency and toxicity, very few attempts had been made to connect biological activity and physical properties.

Later, in 1950s, Hansch developed a hydrophobic parameter and used regression analysis to correlate biological activity with molecular properties. Concurrently since then, scientists have used more sophisticated statistical methods and developed other forms of pattern recognition, such as cluster analysis, principal components analysis, and factor analysis, in the search for pattern between biological and physical data [1].

The aim of Quantitative Structure–Activity Relationships (QSAR) is to develop correlations between biological activity and substituent properties, i.e.,

$$\text{Biological activity} = f(\text{properties}). \quad (1)$$

According to the Brønsted definition, any compound having a hydrogen atom is an acid, since it may be lost as an acidic proton. Depending on the molecule, this process requires more or less energy, and in some cases, the process may be spontaneous. As proton transfer reactions are crucial in chemistry, it is important to quantify the tendency of a molecule to lose its hydrogen atom as an acidic proton. This is the role played by the quantity defined as pK_a. The dissociation equilibrium of a Brønsted acid depends on the interaction of the acid and its conjugate base with the solvent molecules. Therefore, the pK_a value depends on the medium were it was measured, and any reference to pK_a of a certain compound is meaningful only if the solvent is specified. The most experimentally studied medium is water, which justifies the choice of the medium used in this paper. Although water itself is a Brønsted acid, the processes under study should not be affected by self-ionization of the solvent since even the least acidic compound considered is still approximately 10^9 times more acidic than pure water. The experimental pK_a values of several compounds, mainly organic acids in water [2, 3] are determined through very well established methods [4], such as spectroscopy, potentiometry, conductometry, competing reactions, etc. A detailed discussion of the importance of pK_a in chemistry, as well as of the role of proton in organic chemistry can be found in two seminal papers [5, 6].

Carboxylic acids contain a terminal carboxy group (COOH) and an alkyl or aryl group. The uniqueness of substituent lies in the combination of a carbonyl group (C=O) and a hydroxy group (OH). Carboxylic acids are polar molecules capable of forming hydrogen bonds with each other and other molecules. However, increasing hydrophobicity serves to reduce the reactivity of the molecule. In water, carboxylic

[1] The original article was submitted by the authors in English.

1070-3632/03/7311-1792 $25.00 © 2003 MAIK "Nauka/Interperiodica"

acids readily undergo ionization to give carboxylate ion. The latter is unique in that the carbon atom therein is joined to three other atoms by σ bonds. This allows orbitals to overlap in such a way that hybrid carbon–oxygen bonds are formed. Electrons are thus bound to three rather than two nuclei (one carbon and two oxygen atoms). Therefore, electrons are held more tightly, the bonds are stronger, and the anion is more stable than the neutral species. Since the resulting electrons are involved in more than one bond, the electron cloud is delocalized. The acidity of a carboxylic acid originates from the resonance stabilization of its anion. The resulting acidity and stabilization are possible because of the presence of the carbonyl group. Carboxylic acids are more acidic than other organic acids (e.g., alcohols or acetylene) or water. Therefore, aqueous hydroxides readily convert carboxylic acids into their respective salts. In turn, aqueous mineral acids convert these salts back into carboxylic acids. Acid–base equilibrium studies in water can furnish very useful information related to many chemical processes, and so far the main source of this information comes from experimental work.

With the exception of some work on a small number of benzoic acids, carboxylic acids have been conspicuously absent from the QSAR literature [7–9]. Furthermore, toxicity data in QSAR area on aliphatic carboxylic acids are scattered and fragmentary. At present, just minor efforts have been made to derive QSAR to predict toxic potency of aliphatic carboxylic acids. In one of the papers devoted to this issue, Seward and Schultz [9] investigated the aquatic toxicity of aliphatic mono- and dicarboxylic acids and sodium salts of selected acids, with the specific aims of (a) assessing the relative hazard of these chemicals in the *Tetrahymena* inhibitory growth assay [10], and (b) developing quantitative structure–activity relationships for the toxicity of acids. The purpose of this study is to improve those results resorting to a wider set of molecular descriptors in order to obtain better correlation equations.

The paper is organized as follows: the next section deals with the presentation of the molecular descriptors. Then, we give the molecular data set and a brief sketch of the method. After that, we display results and compare the estimations obtained from the present approach with previous data and discuss the relative merits of each procedure. Finally, we state the main conclusions derived from the present study, pointing out some possible further extensions.

Molecular descriptors. The topological indices (or topological descriptors) are numerical quantities derived from molecular graphs representing molecules. The algorithms transforming mathematical representations of molecular graphs into topological indices can be divided into three groups: simple, combinatorial, and complex. The first group includes algorithms performing simple functions on matrix elements or polynomial coefficients such as counting, multiplying, squaring, etc. The second group includes algorithms additionally performing a combinatorial analysis over elements of graph representations. Algorithms of the third group are based on complicated transformations (diagonalization) of the graph matrix representation [11]. Topological indices have been used so far in the correlation and prediction of a host of molecular properties, such as physicochemical, thermodynamic, biophysical, and physiological properties. Up to now, more than 1000 various topological descriptors have been put forward in the chemical literature, though only a handful of them have been widely employed for correlation or/and predictive studies [12].

Table 1. Molecular descriptors

Symbol	Molecular descriptors
$\log K_{o/w}$	Partition coefficient
pK_a	Dissociation constant
E_{LUMO}	Energy of the lowest unoccupied molecular orbital
E_{carb}	Electrotopological state of the carboxy group
E_{meth}	Electrotopological state of the methyl group
MR	Molar refractivity
n_D	Refractive index
T_s	Surface tension
P	Polarizability

When resorting to application of molecular descriptors, it is convenient to distinguish those dealing with physical chemistry properties, electronic indices, and topological features. Table 1 lists the molecular descriptors employed in this work, which are related to the above categories.

Electrotopological indices are particularly important. In fact, electrotopological states, or *E*-states, are fictitious atomic states, i.e., nonquantum states characterizing properties of an atom due to its local and global molecular environments [13]. *E*-State methods are applicable to modeling of many biological properties, such as toxicity and bioaccumulation, as well as of physicochemical properties, including aqueous solubility and chromatographic retention indices. An enormous benefit of this broadly based

and widely applicable paradigm is the opportunity to investigate a range of property modeling problems while operating in the same structure representation genre. A common process in QSAR is blending of different categories of structure descriptors to fully express structure variations, as well as to achieve a quality model. Among these multivariate analyses, there are a number of studies using *E*-state indices as a single parameter. These are usually commingled with molecular connectivity topological indices, physical chemistry descriptors, and electronic parameters. This idea is consistent with the common use of a pharmacophor and a molecular hydrophobic character to describe essential features of a molecule invoking some biological activity, e.g., toxicity.

A natural extension of the *E*-state method is the possibility for calculating an approximate value for a collection of atoms in a group and then for defining a single group parameter in order to estimate the influence of this group compared to that of other common groups. The concept of a group intrinsic state value is useful when comparisons among groups are necessary. This situation is frequently encountered with decision trees or substituent effect considerations.

Electrotopological *E*-state indices were calculated by E-cal software [14]; molar refractivity, polarizability, refractive indices, and surface tension were computed with ACD/ILAB software [15]; and water–octanol partition coefficients were the same as those calculated by Seward and Schultz [9].

Many correlations do not need to be linear. In general, one should also test multivariate regression analysis for larger than linear polynomial order and, if warranted, for other functional dependence [16]. We have computed several fitting polynomial orders and have found that it is not necessary to go beyond second order to improve significantly the final results. The calculations were performed resorting to the standard software STATGRAPHICS® Plus [17], and the fitting procedure in the multivariate regression analysis was made for first- and second-order polynomials.

The chemicals tested here were the same aliphatic mono- and dicarboxylic acids and sodium salts as chosen by Seward and Schultz [9]. Table 2 lists the complete set of acids together with molecular descriptors and aquatic toxicity data.

Seward and Schultz [9] have shown that distinct class-based relationships exist for saturated mono acids, saturated diacids and sodium salts of monoacids. However, for some particular choice of molecular descriptors, these authors could build a response surface that encompassed all the tested acids. The main results reported previously [9] were as follows:

Saturated aliphatic monocarboxylic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.31\log K_{\mathrm{o/w}} - 0.74; \quad (2)$$
$$n\ 17,\ r^2\ 0.927,\ s\ 0.10,\ F\ 190.$$

Saturated aliphatic monocarboxylic acids except for undecanoic acid:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.27\log K_{\mathrm{o/w}} - 0.68; \quad (3)$$
$$n\ 16,\ r^2\ 0.943,\ s\ 0.07,\ F\ 233.$$

Saturated aliphatic dicarboxylic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.19\log K_{\mathrm{o/w}} - 0.66; \quad (4)$$
$$n\ 9,\ r^2\ 0.951,\ s\ 0.08,\ F\ 135.$$

Unsaturated aliphatic monocarboxylic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.23\log K_{\mathrm{o/w}} - 0.38; \quad (5)$$
$$n\ 12,\ r^2\ 0.450,\ s\ 0.321,\ F\ 8.18.$$

All aliphatic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.26\log K_{\mathrm{o/w}} - 0.21E_{\mathrm{LUMO}} - 0.46; \quad (6)$$
$$n\ 38,\ r^2\ 0.727,\ s\ 0.219,\ F\ 46.6.$$

All aliphatic acids, except for 4-pentenoic, acrylic, and 2-nonynoic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.27\log K_{\mathrm{o/w}} - 0.12E_{\mathrm{LUMO}} - 0.57; \quad (7)$$
$$n\ 35,\ r^2\ 0.848,\ s\ 0.157,\ F\ 89.3.$$

These results are rather modest, so that we have tried other possibilities in order to look for better descriptions. First, we have tried for the molecular set employed in Eq. (8) molecular descriptors given in Table 1 to obtain the following linear relationships:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.26\log K_{\mathrm{o/w}} - 0.64; \quad (8)$$
$$n\ 35,\ r^2\ 0.832,\ s\ 0.162,\ F\ 163.20;$$
$$\log(\mathrm{IGC}_{50}^{-1}) = 0.16\mathrm{p}K_{\mathrm{a}} - 0.89; \quad (9)$$
$$n\ 35,\ r^2\ 0.076,\ s\ 0.380,\ F\ 2.70;$$
$$\log(\mathrm{IGC}_{50}^{-1}) = 0.1E_{\mathrm{LUMO}} - 0.20; \quad (10)$$
$$n\ 35,\ r^2\ 0.001,\ s\ 0.396,\ F\ 0.00;$$
$$\log(\mathrm{IGC}_{50}^{-1}) = 0.34E_{\mathrm{carb}} - 6.11; \quad (11)$$
$$n\ 35,\ r^2\ 0.339,\ s\ 0.322,\ F\ 16.90;$$

Table 2. Molecular descriptors and aquatic toxicity data for aliphatic mono- and dicarboxylic acids

Acid	$\log(IGC_{50}^{-1})$[a]	$\log K_{o/w}$[a]	E_{LUMO}[a]	pK_a[a]	E_{carb}	E_{meth}	MR	n_D	T_s	P
Monocarboxylic acids										
2-Ethylbutyric	–0.15	1.68	1.0000	4.81	17.8635	1.4838	31.36	1.425	31.3	12.43
Isobutyric	–0.33	0.94	0.9800	4.68	16.9491	0	22.10	1.408	30.4	8.76
Isovaleric	–0.34	1.16	0.9600	4.86	17.1806	0.2778	26.73	1.418	30.9	10.59
2-Propyl-pentanoic	0.03	2.75	1.0000	4.83	18.4794	3.5836	40.63	1.435	31.8	16.10
Butyric	–0.57	0.79	0.9700	4.80	16.8022	1.0232	22.14	1.411	32.5	8.77
2-Ethyl-hexanoic	0.08	2.64	1.0100	4.90	18.3937	3.7145	40.63	1.435	31.8	16.10
Valeric	–0.27	1.39	0.9600	4.82	17.1101	2.0815	26.77	1.420	32.7	10.61
Trimethyl-acetic	–0.25	1.47	1.0200	4.62	17.5139	0	26.74	1.419	30.2	10.60
Propionic	–0.51	0.33	0.9500	4.74	16.3449	0.2222	17.51	1.397	32.3	6.94
Heptanoic	–0.11	2.42	0.9400	4.88	17.7382	4.5544	36.04	1.432	32.9	14.28
Nonanoic	0.35	3.47	0.9700	4.97	17.7382	7.2507	45.30	1.440	33.1	17.96
Decanoic	0.51	4.09	0.9600	4.83	17.8253	8.6425	49.94	1.443	33.1	19.79
Undecanoic	0.90	4.42	0.9600	4.86	17.8983	10.0532	54.57	1.445	33.2	21.63
Lauric	[b]	4.60	0.9600	4.86	–	–	–	–	–	–
3-Methylvaleric	–0.23	1.75	0.9800	4.86	17.4886	1.2528	31.36	1.425	31.3	12.43
4-Methylvaleric	–0.27	1.75	0.9700	4.86	17.3705	1.0902	31.36	1.425	31.3	12.43
Octanoic	0.08	3.05	0.9400	4.90	17.6321	5.8845	40.67	1.437	33.0	16.12
Hexanoic	–0.21	1.92	0.9700	4.87	17.3322	3.2770	31.41	1.427	32.8	12.45
Dicarboxylic acids										
Glutaric	–0.64	–0.29	0.8100	4.35	16.8734	0.0866	28.34	1.476	56.1	1.476
Adipic	–0.61	0.08	0.8100	4.35	17.1579	1.0178	32.97	1.476	52.4	1.476
Succinic	–0.94	–0.59	0.6000	3.76	16.4629	–0.5926	23.70	1.477	61.6	1.477
1,12-Dodecane-dicarboxylic	0.08	4.13	0.9069	4.44	18.0218	11.2898	70.03	1.474	41.7	1.474
1,10-Decandi-carboxylic	–0.09	3.07	0.8757	4.44	17.9095	8.5144	60.77	1.474	43.1	1.474
Malonic	–0.71	–0.58	0.1296	2.77	15.8195	–0.8056	19.07	1.478	70.5	1.478
Pimelic	–0.58	0.43	0.7474	4.40	17.3666	2.1002	37.60	1.476	49.8	1.476
Sebacic	–0.27	2.01	0.8548	4.44	17.7543	5.8246	51.50	1.475	44.9	1.475
Suberic	–0.51	0.95	0.7951	4.42	17.5262	3.281	42.24	1.475	47.8	1.475
Unsaturated carboxylic acids										
trans-2-Pentenoic	–0.28	1.41	–0.1600	4.72	16.7559	0.7840	26.83	1.452	34.1	1.452
4-Pentenoic	0.12	0.87	0.9200	4.74	16.9781	1.8572	26.50	1.438	33.1	1.438
trans-2-Hexenoic	–0.13	1.94	–0.1900	4.72	16.9432	4.1320	31.46	1.455	34.0	1.455
trans-3-Hexenoic	–0.22	1.40	0.6500	4.54	17.1430	1.0567	31.46	1.455	34.0	1.455
Crotonic	–0.54	0.72	–0.1900	4.62	16.4480	0	22.20	1.448	34.2	1.448
Acrylic	0.16	0.35	–0.2000	1.86	15.8727	2.9606	17.23	1.422	32.8	1.422
2-Octenoic	0.21	3.00	–0.1700	4.72	17.2779	4.3525	40.73	1.459	34.0	1.459
2-Octynoic	–0.06	2.12	0.1200	2.67	16.9238	4	38.68	1.470	40.9	1.470
2-Propynoic	–0.29	–0.52	0.0600	1.86	15.4004	0	15.33	1.455	52.9	1.455
2-Nonynoic	0.71	2.65	0.1200	2.39	17.0299	5.3118	43.31	1.47	40.1	1.470
2-Nonenoic	0.60	3.53	–0.1700	4.39	17.384	5.6883	45.36	1.46	33.9	1.460
3-Butenoic	–0.64	0.64	0.8000	4.39	16.5416	3.2756	21.87	1.432	33.0	1.432

[a] Data of [9]. [b] Not toxic in a saturated solution.

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.09E_{\mathrm{meth}} - 0.49; \quad (12)$$

$$n\ 35,\ r^2\ 0.577,\ s\ 0.257,\ F\ 44.95;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.02MR - 0.91; \quad (13)$$

$$n\ 35,\ r^2\ 0.428,\ s\ 0.299,\ F\ 25.69;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -1.39n_{\mathrm{D}} + 1.81; \quad (14)$$

$$n\ 35,\ r^2\ 0.007,\ s\ -0.394,\ F\ 0.23;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.02\ T_{\mathrm{s}} + 0.54; \quad (15)$$

$$n\ 35,\ r^2\ 0.246,\ s\ 0.344,\ F\ 10.75;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.05P - 0.91; \quad (16)$$

$$n\ 35,\ r^2\ 0.428,\ s\ 0.299,\ F\ 24.69.$$

Analysis of the preceding equations clearly shows that only $\log K_{\mathrm{o/w}}$ is a fairly suitable variable for description of aquatic toxicity of carboxylic acids through one-variable relationships. These data suggest that multi-variable regression equations must be employed with the purpose to ameliorate previous results. One of such equations was given previously (Eq. 8), although improvement is nor very significant [cf. statistical parameters given for Eqs. (8) and (9)]. Thus we have tried other possibilities, and some results are as follows:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.368\log K_{\mathrm{o/w}} - 0.028P - 0.131\mathrm{p}K_{\mathrm{a}} + 0.011E_{\mathrm{carb}} - 0.044; \quad (17)$$

$$n\ 35,\ r^2\ 0.893,\ s\ 0.135,\ F\ 277.5;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.37\log K_{\mathrm{o/w}} - 0.01P - 0.14E_{\mathrm{carb}} - 0.02E_{\mathrm{meth}} + 1.82; \quad (18)$$

$$n\ 35,\ r^2\ 0.872,\ s\ 0.148,\ F\ 225.6;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.38\log K_{\mathrm{o/w}} - 0.13\mathrm{p}K_{\mathrm{a}} + 4.95P - 0.01E_{\mathrm{meth}} + 0.001E_{\mathrm{carb}} - 1.97MR + 0.05; \quad (19)$$

$$n\ 35,\ r^2\ 0.895,\ s\ 0.139,\ F\ 281.6;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.30\log K_{\mathrm{o/w}} - 0.12\mathrm{p}K_{\mathrm{a}} - 0.01E_{\mathrm{carb}} - 7.7110^{-8}P^5 + 0.0005(E_{\mathrm{meth}})^3 + 0.05; \quad (20)$$

$$n\ 35,\ r^2\ 0.930,\ s\ 0.112,\ F\ 432.1.$$

These results make clear that a suitable improvement can be achieved through the use of linear relationships with several variables. Of particular interest are the two last equations, which are the best ones among those where *E*-state indices appear as independent variables. Regarding previous results given by Seward and Schultz [9] for the same molecular set, we show that better estimations can be obtained through application of complementary molecular descriptors, as defined in Table 1.

According to Seward and Schulz [9], distinct class-based relationships exist for saturated monoacids, saturated diacids, and monoacid sodium salts; therefore, we believe it reasonable to report results for each separate molecular set. The most meaningful fitting equations are as follows:

Monocarboxylic acids, except for undecanoic and lauric acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.677 + 0.273\log K_{\mathrm{o/w}}; \quad (21)$$

$$n\ 16,\ r^2\ 0.943,\ s\ 0.07,\ F\ 232.95;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -8.098 + 1.649\mathrm{p}K_{\mathrm{a}}; \quad (22)$$

$$n\ 16,\ r^2\ 0.245,\ s\ 0.259,\ F\ 4.54;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.348 + 0.216E_{\mathrm{LUMO}}; \quad (23)$$

$$n\ 16,\ r^2\ 0.03,\ s\ 0.298,\ F\ 0.00;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.270 + 0.288\log K_{\mathrm{o/w}} - 0.284\mathrm{p}K_{\mathrm{a}} + 0.406E_{\mathrm{LUMO}}; \quad (24)$$

$$n\ 16,\ r^2\ 0.951,\ s\ 0.071,\ F\ 77.93;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.938 + 0.29\log K_{\mathrm{o/w}} - 0.341\mathrm{p}K_{\mathrm{a}}; \quad (25)$$

$$n\ 16,\ r^2\ 0.950,\ s\ 0.069,\ F\ 124.20;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.625 + 0.114(K_{\mathrm{o/w}})^{1.560}; + (E_{\mathrm{LUMO}})^{-2.780}; \quad (26)$$

$$n\ 16,\ r^2\ 0.946,\ s\ 0.069,\ F\ 243.27;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.345 + 0.0025(\log K_{\mathrm{o/w}})^{4.224} + (\mathrm{p}K_{\mathrm{a}})^{2.069}; \quad (27)$$

$$n\ 16,\ r^2\ 0.7775,\ s\ 0.141,\ F\ 48.93;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.2529 + 0.00059(\log K_{\mathrm{o/w}})^{5.224}; \quad (28)$$

$$n\ 16,\ r^2\ 0.701,\ s\ 0.163,\ F\ 32.93.$$

Dicarboxylic acids, except for succinic acid:

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.624 + 0.170\log K_{\mathrm{o/w}}; \quad (29)$$

$$n\ 8,\ r^2\ 0.992,\ s\ 0.028,\ F\ 743.76;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.045 + 0.159\log K_{\mathrm{o/w}} - 0.269\mathrm{p}K_{\mathrm{a}} + 0.640E_{\mathrm{LUMO}}; \quad (30)$$

$$n\ 8,\ r^2\ 0.997,\ s\ 0.020,\ F\ 494.20;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.8 + 0.043(\log K_{\mathrm{o/w}})2 + \mathrm{p}K_{\mathrm{a}} + E_{\mathrm{LUMO}}; \quad (31)$$

$$n\ 8,\ r^2\ 0.935,\ s\ 0.079,\ F\ 86.56;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.792 + 0.043(\log K_{\mathrm{o/w}})^2 + \mathrm{p}K_{\mathrm{a}} + (E_{\mathrm{LUMO}})^2; \quad (32)$$

$$n\ 8,\ r^2\ 0.936,\ s\ 0.08,\ F\ 86.50.$$

Unsaturated carboxylic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.2022 + 0.305\log K_{o/w} - 0.186\mathrm{p}K_a + 0.121E_{LUMO}; \quad (33)$$

$$n\ 12,\ r^2\ 0.697,\ s\ 0.266,\ F\ 6.13;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.207 + 0.095(\log K_{o/w})^2 - 0.171\mathrm{p}K_a + 0.419(E_{LUMO})^2; \quad (34)$$

$$n\ 12,\ r^2\ 0.733,\ s\ 0.25,\ F\ 7.32.$$

Unsaturated carboxylic acids, except for 4-pentenoic acid:

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.431 + 0.247\log K_{o/w}; \quad (35)$$

$$n\ 11,\ r^2\ 0.497,\ s\ 0.321,\ F\ 8.89;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = 0.268 - 0.215\mathrm{p}K_a - 0.156E_{LUMO} + 0.317\log K_{o/w}; \quad (36)$$

$$n\ 11,\ r^2\ 0.863,\ s\ 0.190,\ F\ 14.68;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.014 - 0.023\ (\mathrm{p}K_a)^2 - 0.160(E_{LUMO})^2 + 0.088(\log K_{o/w})^2; \quad (37)$$

$$n\ 11,\ r^2\ 0.835,\ s\ 0.208,\ F\ 11.84.$$

Unsaturated, monocarboxylic, and dicarboxylic acids, except for 4-pentenoic, acrylic, 2-nonynoic, 1,12-dodecanedicarboxylic, 1,10-decanedicarboxylic, undecanoic, 2-nonenoic, lauric, and propynoic acids:

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.66 + 0.273\log K_{o/w} - 0.027E_{LUMO}; \quad (38)$$

$$n\ 30,\ r^2\ 0.944,\ s\ 0.079,\ F\ 229.33;$$

$$\log(\mathrm{IGC}_{50}^{-1}) = -0.790 + 0.069(\log K_{o/w})^2 + 0.014(\mathrm{p}K_a)^2 - 0.052\ (E_{LUMO})^2; \quad (39)$$

$$n\ 30,\ r^2\ 0.858,\ s\ 0.129,\ F\ 52.54.$$

Analysis of these results demonstrates that fairly satisfactory estimations of the toxicity of aliphatic carboxylic acids and salts can be obtained with a relatively high correlation coefficient (r^2 value) on the basis of the chosen molecular descriptors. The predictions for unsaturated acids are not very good [cf. Eqs. (33), (34), (36), and (37)], but they are clearly superior to those presented previously [Eq. (5)]. Particularly noticeable are the excellent fitting equations derived for dicarboxylic acids [Eqs. (29), (30)] where correlation coefficients higher than 0.990 were obtained.

Our study have shown that highly significant QSAR can be obtained on the basis of a rather modest set of molecular descriptors for each subclass of carboxylic acids and salts to estimate their aquatic toxicity [$\log(\mathrm{IGC}_{50}^{-1})$] tested in the *Tetrahymena* population growth assay reported before [9]. Some outliers were detected in the computation of the fitting equations; their exclusion leads to significantly improved estimations. The use of several-variable equations and looking for optimal polynomials of higher order give rather satisfactory predictive relationships which can be useful for practical purposes. Particularly important is the employment of *E*-state indices to arrive at good results. The above equations make up a remarkable improvement with respect to previous results for the same molecular set. Since our search for optimal molecular descriptors has not been exhaustive, there is a room for further improvement efforts. Work along this line is being performed at our laboratories and results will be presented elsewhere in the forthcoming future.

REFERENCES

1. Hansch, C. and Leo, A., *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology. ACS Professional Reference Book,* Washington: Am. Chem. Soc., 1995.
2. Christen, H.R. and Vogtle, F., *Organische Chemie—Von den Grundlagen zur Forschung,* Frankfurt: Otto Salle, 1988, vol. 1, p. 419.
3. March, J., *Advanced Organic Chemistry. Reactions, Mechanisms, and Structure,* New York: Wiley, 1992, ch. 8.
4. Cookson, R.F., *Chem. Rev.,* 1974, vol. 74, p. 5.
5. Schüürmann, G., *Quant. Struct.–Act. Relat.,* 1996, vol. 15, p. 121.
6. Stewart, R., *The Proton: Applications to Organic Chemistry,* Wasserman, H.H., Ed., New York: Academic, 1985.
7. Russom, C.L., Bradbury, S.P., Brtoderius, S.J., Hammermeister, D.E., and Drummond, R.A., *Environ. Toxicol. Chem.,* 1997, vol. 16, p. 948.
8. Konemann, H. and Musch, A., *Toxicology,* 1981, vol. 19, p. 223.
9. Seward, J.R. and Schultz, T.W., *SAR–QSAR Environ. Res.,* 1999, vol. 10, p. 557.
10. Schultz, T.W., *Toxicol. Meth.,* 1997, vol. 7, p. 289.
11. Mekenyan, O. and Basak, S.S., *Graph Theoretical Approaches to Chemical Reactivity,* Bonchev, D. and Mekenyan, O., Eds., Dordrecht: Kluwer Academic, 1994, p. 224.

12. Devillers, J. and Balaban, A.T., *Topological Indices and Related Descriptors in QSAR and QSPR,* Amsterdam: Gordon and Breach, 1999.

13. Kier, L.B. and Hall, L.H., *Molecular Structure Description. The Electrotopological State,* San Diego: Academic, 1999.

14. *E-Cal Version 1.1,* Sci Vision and Lowell H. Hall. Sci Vision, Inc., 200 Wheeler Road, Burlington, MA 01803.

15. ACD/ILAB Software. http://www.acdlabs.com. Advanced Chemistry Development, 1996–2001.

16. Randic, M. and Basak, S.C., *Some Aspects of Mathematical Chemistry,* Sinha, D.K., Basak, S.C., Mohany, R.K., and Busa Mallic, J.N., Eds.: Santiniketan (India): Visva-Bharati Univ., 2000, p. 24.

17. *STATGRAPHICS Plus. Ver. 4.0. Demo version,* Manugistic, Inc. 2115 East Jefferson Street Rockeville, Md 20852. www.statgraphics.com. 1994–1999.