

An Optimal Characterization of Structure by Means of Several Molecular Connectivity and Complexity Indices

by J.A. Marinich¹, F. Maguna¹, N. Okulik¹ and E.A. Castro^{2*}

¹Facultad de Agroindustrias, Universidad Nacional del Nordeste, R. S. Peña, Chaco, Argentina
e-mail: fmaguna@fai.unn.edu.ar

²CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata,
C.C. 962, La Plata 1900, Argentina
e-mail: castro@daltón.quimica.unlp.edu.ar

(Received April 9th, 2001; revised manuscript November 21st, 2001)

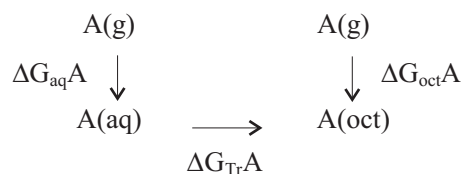
An improved numerical procedure is presented in order to enhance the possibilities of fitting polynomial equations to predict log P data within the realm of the QSAR/QSPR theory. The use of real exponents instead of restricting to integer ones for the variables in the mathematical equations gives better results and a minor number of independent variables are needed to achieve a given accuracy degree. Some possible future extensions of the method are pointed out.

Key words: QSAR/QSPR theory, lipophilicity, molecular connectivity, topological indices, multiple regression analysis

The term lipophilic literally means “oil loving”, and lipophilicity is a measure of the degree to which a given molecule prefers hydrophobic non polar environments to water. The most common experimental measure of lipophilicity is the logarithm of the partition coefficient for a solute distributing itself between water and some organic solvent, such as 1-octanol or chloroform. This quantity is abbreviated as log P and has been measured experimentally for a wide range of organic compounds

$$\log P = \log_{10}([S]_{\text{org}}/[S]_{\text{aq}}) \quad (1)$$

The partition coefficient for octanol-water ($\log P_{\text{ow}}$) has become the preferred measure for lipophilicity in the development of biologically active molecules, in which transport across biological membranes is often critical [1–4] and has become the most popular physicochemical parameter to define the general lipophilicity of compounds in structure-activity studies [2,5–7]. The most theoretically satisfying method for calculating $\log P_{\text{ow}}$ would be directly model the transfer of a solute from water to octanol [8–13]. It also could be calculated directly from the free energies of



$$\log P_{\text{ow}} = (G_{\text{aq}} - \Delta G_{\text{aq}})/2.30RT \quad (2)$$

The solvation free energy, which is the work required to transfer a molecule from a fixed position in the ideal gas phase to a fixed position in solution [14], consists of the free energies of the creation of a cavity in the solution and of the placement of the solute into that cavity. The former contribution is determined by the volume and surface area of the solute, whereas the later is equivalent to establishing the solute electronic configuration in the cavity, usually termed as the charging free energy. The difficulty with this approach is calculating the free energies of solvation [15].

For numerous chemicals $\log P_{\text{ow}}$ values are readily available. However, its experimental determination is time-consuming, expensive, and can yield various technical problems [16,17]. Furthermore, it is practically impossible to measure the $\log P_{\text{ow}}$ for all existing chemicals, and the ability to predict this quantity of yet no synthesized compounds could be very useful. Under these conditions, several theoretical schemes have been proposed and used to calculate $\log P_{\text{ow}}$ [18]. Although these approaches work rather well for simple chemicals, they have limitations for complex molecules since correction factors have to be introduced. This prompted Basak *et al.* [19] to develop models for the prediction of $\log P_{\text{ow}}$ of molecules using parameters which can be calculated directly from structure and resorting to the maximum R^2 criterion for the optimal selection [20]. The authors classify a diverse molecular set ($n = 382$ molecules) into more homogeneous subsets, taking those which lack any hydrogen-bonding potential with respect to the scale of Ou *et al.* [21] ($n = 139$), and they investigated to what extent topological indices could predict $\log P_{\text{ow}}$ values for the homogeneous non-hydrogen bonding group.

Multiple regression analysis showed that there was an improvement in the prediction of $\log P_{\text{ow}}$ with respect to the original diverse group of 382 chemicals [22]. The topological indices appearing in the best one-, two,, six-variable linear model for the prediction of the dependent variable are those listed in Table 1 (Definitions for these indices are given in [19]).

Table 1. Best topological indices to predict $\log P_{\text{ow}}$.

${}^0\chi^v$	Valence connectivity type path terms of zero order
${}^4\chi^v$	Valence connectivity type path terms of fourth order
${}^4\chi_{\text{PC}}^v$	Valence connectivity type path-cluster terms of fourth order
${}^6\chi_{\text{CH}}$	Valence connectivity type chain or cycle terms of fourth order
${}^5\chi_{\text{C}}^v$	Valence connectivity type cluster terms of fifth order
IC_o	Mean information content or complexity of a graph based on the zero order neighborhood of vertices in a graph

The model had two influential outliers (compounds 14 and 101 in Table 5 of [19] as determined by Cook's D statistic [23] and deletion of these molecules resulted in a highly significant 6-variable linear model. The purpose of this work is to look for a rather simple and direct way to improve Basak *et al's* equations in order to develop at full length the potentialities of the present approach.

METHOD

The present study is the continuation of our ongoing QSAR/QSPR research to improved models for the prediction of molecular properties, making full use of the intrinsic advantages of the regression formulae. In fact, previously we have verified that a satisfactory enough improvement of these relationships can be gotten via the simple resort of employing higher-order equations [24–28] (*i.e.* quadratic, cubic, *etc.* statistical regressions). Non-linear models may be fitted to data sets by the inclusion of functions of physicochemical parameters in a linear regression model or by use of non-linear fitting models. Construction of linear regression models containing non-linear terms is most often prompted when the data is clearly not well fitted by a linear model. A very common example in the field of QSAR involves non-linear relationships with the hydrophobic descriptors, such as log P [29]. Non-linear dependency of biological properties on this parameter became apparent early in the development of QSAR model and a first approach to the solution of these problems involved fitting a parabola in log P [6]. Whatever the cause of such relationships, it is clear that non-linear functions are required in order to model the biological data. An interesting feature of the use of non-linear functions is that it is possible to calculate an optimum value for the physicochemical property involved (usually log P) [30,31].

In order to make a clear and direct comparison with well known linear equations, we have resorted to the Basak *et al's* previous results [19] for prediction of log P_{ow} from topological indices listed in Table 1 *via* a multiple regression analysis. We have made a systematic search of the best one-variable model, the best two-variables model, up to the six-variables model of second and third order fitting equations. We have employed the maximum R² method to identify the prediction model for log P_{ow} [20], as used by Basak *et al.* [19]. The model was developed resorting to the set of 137 compounds which lack any hydrogen-bonding potential with respect to the scale of Ou *et al.* [21] to investigate to what extent higher-order relationships could improve linear fitting equations to predict log P_{ow} values as a function of topological indices for the homogeneous non-hydrogen bonding group.

Calculations were made with the well known MATHEMATICA[®] software, through the employment of the computational program "*Statistic Nonlinear Fit*" contained in the standard statistical analysis packages [32].

RESULTS AND DISCUSSION

In Table 2 we give a summary of multiple regression analysis for prediction of $\log P_{ow}$ from topological indices via higher-order equations for the best one-variable, two-variables, up to four-variables models together with a couple of rather poor results (Eqs. 13 and 14) included here for comparative purposes. Table 3 displays the data for experimental $\log P_{ow}$ and estimated $\log P_{ow}$ via the best fitting high-order equation (Eq. 29 in Table 2) for 137 compounds. We have also determined another complete set of regression equations without the intersection term (*i.e.* $A = 0$), but we deem unnecessary to insert them here because they are slightly less accurate than those with $A \neq 0$. Complete results derived from Eqs. 1-29 (with both $A = 0$ and $A \neq 0$) are available upon request to the corresponding author (EAC).

Table 2. Fitting equations and statistical parameters.

Equation	% R ²	Standard error	Estimated coefficients		
			Coefficient	Standard error	
$A + b IC_0 + c {}^0\chi^v$ (1)	94.00	0.3410	A	-3.3157	0.2366
			b	-2.6152	0.3040
			c	4.8448	4.6179
$A + b ({}^0\chi^v)^2$ (2)	91.17	0.4121	A	-0.1769	0.1286
			b	1.0639	0.0285
$A + b ({}^0\chi^v)^2 + c IC_0$ (3)	94.54	0.3253	A	1.4799	0.2085
			b	1.1999	0.0270
			c	-2.6373	0.2900
$A + b ({}^0\chi^v)^2 + c (IC_0)^2$ (4)	94.43	0.3287	A	0.4309	0.1234
			b	1.2015	0.0275
			c	-1.6303	0.1843
$A + b ({}^0\chi^v)^2 + c ({}^5\chi_c)^2$ (5)	91.19	0.4132	A	-0.2115	0.1448
			b	1.0760	0.0367
			c	-0.3853	0.7347
$A + b ({}^0\chi^v)^2 + c ({}^6\chi_{CH})^2$ (6)	92.22	0.3884	A	-0.4636	0.1387
			b	1.1601	0.0351
			c	-5.0010	1.1790
$A + b ({}^0\chi^v)^2 + c ({}^4\chi_v)^2$ (7)	91.34	0.4098	A	-0.4159	0.1972
			b	1.1525	0.0625
			c	-0.1674	0.1105
$A + b ({}^0\chi^v)^2 + c ({}^4\chi_{PC})^2$ (8)	91.52	0.4054	A	-0.3444	0.1452
			b	1.1296	0.0396
			c	-0.2728	0.1161
$A + b ({}^0\chi^v)^2 + c ({}^6\chi_{CH})^2 + d (IC_0)^3 + e {}^4\chi^v$ (9)		0.3166	A	0.1350	0.1261
			b	1.0756	0.4940
			c	-6.0111	1.0077
			d	-1.2010	0.1368
			e	0.7037	0.1472

Table 2 (continuation)

$A + b({}^0\chi^v)^2 + c(IC_0)^2 + d({}^6\chi_{CH})^2 + e({}^4\chi^v)$ (10)	95.79	0.2879	A	0.4537	0.1362
			b	1.0799	0.0485
			c	-1.5033	0.1630
			d	-6.0439	0.9879
			e	0.6947	0.1445
$A + b({}^0\chi^v)^2 + c IC_0 + d({}^4\chi^v_{PC})$ (11)	94.87	0.3166	A	1.1470	0.2331
			b	1.2750	0.0369
			c	-2.4016	0.2937
			d	-0.3602	0.1240
$A + b({}^0\chi^v)^2 + c(IC_0)^2 + d({}^5\chi_c)^2$ (12)	94.46	0.3289	A	0.3860	0.1335
			b	1.2181	0.0333
			c	-1.6345	0.1845
			d	-0.5179	0.5850
$A + b({}^5\chi_c)^2 + c(IC_0)^2 + d({}^4\chi^v_{PC})^2$ (13)	41.97	1.0647	A	2.7952	0.3809
			b	0.9167	4.2602
			c	1.1584	0.5500
			d	1.7291	0.6466
$A + b({}^5\chi_c)^2 + c(IC_0)^2 + d({}^4\chi_{PC})$ (14)	48.71	1.0010	A	3.1377	0.3666
			b	1.0502	2.5389
			c	-0.0793	0.5947
			d	2.4355	0.4816
$A + b({}^0\chi^v)^2 + c(IC_0)^3$ (15)	94.27	0.3333	A	0.0784	0.1082
			b	1.2012	0.0281
			c	-1.3077	0.1537
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d IC_0$ (16)	94.68	0.3225	A	3.2927	1.0120
			b	1.1882	0.0275
			c	1.8035	0.9855
			d	-6.0834	1.9049
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^5\chi_c)$ (17)	94.34	0.3326	A	-0.0223	0.1343
			b	1.2339	0.0382
			c	-1.2084	0.1549
			d	-0.3807	0.3023
$A + b({}^0\chi)^2 + c(IC_0)^3 + d({}^6\chi_{CH})$ (18)	94.86	0.3167	A	-0.2023	0.1252
			b	1.2796	0.0334
			c	-1.1137	0.1542
			d	-1.5395	0.3921
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^0\chi^v)$ (19)	94.28	0.3343	A	-0.4751	1.1214
			b	1.0641	0.2778
			c	-1.3082	0.1542
			d	0.5573	1.1239
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^4\chi^v_{PC})$ (20)	94.69	0.3220	A	-0.1539	0.1265
			b	1.2873	0.0379
			c	-1.1877	0.1530
			d	-0.4072	0.1248

Table 2 (continuation)

$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^6\chi_{CH})^2$ (21)	94.87	0.3165	A	-0.1582	0.1189
			b	1.2663	0.0314
			c	-1.2237	0.1475
			d	-3.8414	0.9708
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^6\chi_{CH})^3$ (22)	94.73	0.3208	A	0.0887	0.1151
			b	1.2468	0.0302
			c	-1.2652	0.1485
			d	-8.5263	2.4959
$A + b({}^0\chi^v)^3 + c(IC_0)^3 + d({}^6\chi_{CH})^3$ (23)	94.22	0.3359	A	1.5432	0.0976
			b	0.4042	0.0103
			c	-1.2153	0.1550
			d	-11.4924	2.6499
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^6\chi_{CH})^2 + e({}^4\chi^v)^{0.5}$ (24)	95.81	0.2871	A	-0.1097	0.1082
			b	1.1244	0.0386
			c	-1.1799	0.1340
			d	-4.5725	0.8908
			e	0.6668	0.1225
$A + b({}^0\chi^v)^2 + c(IC_0)^3 + d({}^6\chi_{CH})^2 + e({}^4\chi^v)^{0.6}$ (25)	95.80	0.2875	A	-0.0547	0.1097
			b	1.1098	0.0407
			c	-1.1849	0.1342
			d	-4.8590	0.9020
			e	0.6951	0.1288
$A + b({}^0\chi^v)^2 + c(IC_0)^{2.8} + d({}^6\chi_{CH})^2 + e({}^4\chi^v)^{0.6}$ (26)	95.84	0.2862	A	0.0092	0.1107
			b	1.1104	0.0405
			c	-1.2281	0.1374
			d	-4.8685	0.8975
			e	0.6949	0.1282
$A + b({}^0\chi^v)^2 + c(IC_0)^{2.4} + d({}^6\chi_{CH})^2 + e({}^4\chi^v)^{0.6}$ (27)	95.91	0.2837	A	0.1053	0.1143
			b	1.1114	0.0401
			c	-1.3365	0.1461
			d	-4.8899	0.8889
			e	0.6950	0.1271
$A + b({}^0\chi^v)^2 + c(IC_0)^{2.4} + d({}^6\chi_{CH}) + e({}^4\chi^v)^{0.6}$ (28)	96.40	0.2662	A	-0.0432	0.1129
			b	1.1029	0.0371
			c	-1.0834	0.1450
			d	-2.6105	0.3612
			e	0.9299	0.1280
$A + b({}^0\chi^v)^{2.89089} + c(IC_0)^{-2.932} + d({}^6\chi_{CH}) + e({}^4\chi^v)^{0.345562}$ (29)	97.11	0.2384	A	-0.0055	0.0952
			b	0.4109	0.0103
			c	0.2232	0.0252
			d	-2.7957	0.3058
			e	1.0079	0.0936

Results show that, in general, present equations give very good predictions of the physicochemical property $\log P_{ow}$ for the set of 137 chemicals, where R^2 is higher than 0.94 in all cases. When comparing results arising from linear fitting equations

with those derived from higher order ones, we see there is a significant improvement (for example, compare statistical data for Eqs. 1, 3, 4, 15, 16 and 19). In some cases, results are very inaccurate, such as those associated to Eqs. 13 and 14.

Table 3. Experimental and theoretical log P.

Sequence #	Molecule	log P (exp)	log P [19]	log P (Eq.29)
1	1,1,1-trichloroethane	2.481	2.797	2.420
2	1,1,2,2-tetrachloroethane	2.644	2.872	2.915
3	1,2,3,4-tetrachlorobenzene	4.994	4.875	4.901
4	1,2,3,4-tetramethylbenzene	4.738	4.707	4.592
5	1,2,3,5-tetrachlorobenzene	4.994	4.912	4.935
6	1,2,3,5-tetramethylbenzene	4.738	4.712	4.623
7	1,2,3-trichlorobenzene	4.281	4.165	4.182
8	1,2,3-trimethylbenzene	4.089	4.067	3.969
9	1,2,4,5-tetrachlorobenzene	4.994	4.896	4.905
10	1,2,4,5-tetramethylbenzene	4.738	4.711	4.595
11	1,2,4-trichlorobenzene	4.281	4.113	4.181
12	1,2,4-trimethylbenzene	4.089	3.991	3.967
13	1,2-dibromobenzene	3.588	4.063	4.518
14	1,2-dichlorobenzene	3.568	3.414	3.452
15	1,2-dichloroethane	1.458	1.862	1.725
16	1,2-diphenylethane	4.888	4.931	4.958
17	1,3,5-trichlorobenzene	4.281	4.195	4.255
18	1,3,5-trimethylbenzene	4.089	4.005	4.034
19	1,3-dichlorobenzene	3.568	3.417	3.503
20	1,3-dimethylnaphthalene	4.614	4.657	4.569
21	1,4,5-trimethylnaphthalene	5.263	5.259	5.168
22	1,4-dibromobenzene	3.868	4.172	4.506
23	1,4-dichlorobenzene	3.568	3.269	3.443
24	1,4-dimethylnaphthalene	4.614	4.698	4.557
25	1,5-dimethylnaphthalene	4.614	4.716	4.566
26	1-butene	2.266	1.627	1.831
27	1-chlorobutane	2.523	2.917	2.911
28	1-chloroheptane	4.110	4.719	4.409
29	1-chlorohexane	3.581	4.175	3.920
30	1-chloronaphthalene	4.029	4.166	4.030
31	1-chloropentane	3.052	3.568	3.414
32	1-chloropropane	1.994	1.920	1.778
33	1-ethylnaphthalene	4.494	4.513	4.398
34	1-hexene	3.324	3.220	3.311
35	1-isopropyl-4-methylbenzene	4.368	4.323	4.327
36	1-methylbenz(a)anthracene	6.313	6.178	6.332

Table 3 (continuation)

37	1-methylfluorene	4.874	4.946	4.722
38	1-methylnaphthalene	3.965	4.136	3.961
39	1-pentene	2.795	2.510	2.794
40	12-methylbenz(a)anthracene	6.313	6.237	6.329
41	2,2',4,5-tetrachlorobiphenyl	6.882	6.462	6.725
42	2,2',4-trichlorobiphenyl	6.169	5.964	6.107
43	2,2,4-trimethylpentane	4.536	4.701	5.047
44	2,3,4,5-tetrachlorobiphenyl	6.882	6.451	6.707
45	2,3-dimethylnaphthalene	4.614	4.632	4.538
46	2,4'-dichlorobiphenyl	5.456	5.385	5.452
47	2,4,5-trichlorobiphenyl	6.169	5.957	6.078
48	2,4,6-trichlorobiphenyl	6.169	6.024	6.121
49	2,5-dichlorobiphenyl	5.456	5.404	5.448
50	2,6-dichlorobiphenyl	5.456	5.546	5.479
51	2,6-dimethylnaphthalene	4.614	4.476	4.546
52	2-chlorobiphenyl	4.743	5.016	4.847
53	2-chloronaphthalene	4.029	4.049	4.016
54	2-chlorophenanthrene	5.203	5.221	5.240
55	2-chlorotoluene	3.504	3.283	3.326
56	2-methylantracene	5.139	5.140	5.167
57	2-methylbutane	3.209	2.496	2.661
58	2-methylhexane	4.267	4.080	4.256
59	2-methylnaphthalene	3.965	4.015	3.948
60	2-methylpentane	3.738	3.592	3.831
61	2-methylphenanthrene	5.139	5.192	5.175
62	3-chlorotoluene	3.504	3.254	3.372
63	4-chlorotoluene	3.504	3.118	3.316
64	5,6-dimethylchrysene	6.962	6.657	6.857
65	5-methylchrysene	6.313	6.231	6.329
66	6-methylbenzo(e)pyrene	6.773	6.687	6.838
67	6-methylchrysene	6.313	6.244	6.325
68	7-ethylbenz(a)anthracene	6.842	6.466	6.731
69	7-methylbenz(a)anthracene	6.313	6.249	6.325
70	9,10-dimethylantracene	5.788	5.832	5.736
71	9-methylantracene	5.139	5.316	5.174
72	acenaphthene	4.070	4.026	3.832
73	adamantane	3.982	4.376	4.366
74	anthracene	4.490	4.815	4.627
75	benz(a)anthracene	5.664	5.826	5.802
76	benz(a)anthracene	5.664	5.783	5.797
77	benzene	2.142	2.031	2.252
78	benzo(a)fluorene	5.399	5.298	5.218

Table 3 (continuation)

79	benzo(a)pyrene	6.124	6.299	6.341
80	benzo(b)fluoranthrene	6.124	5.984	5.962
81	benzo(b)fluorene	5.399	5.232	5.192
82	benzo(e)pyrene	6.124	6.346	6.335
83	benzo(ghi)perylene	6.584	6.732	6.850
84	benzo(j)fluoranthrene	6.124	5.986	5.991
85	benzo(k)fluoranthrene	6.124	5.942	5.969
86	biphenyl	4.030	4.529	4.354
87	bromobenzene	3.005	3.230	3.335
88	carbon tetrachloride	2.875	3.544	3.572
89	chloroanthrene	6.418	6.065	6.200
90	chlorobenzene	2.855	2.663	2.787
91	chrysene	5.664	5.877	5.808
92	cycloheptane	3.913	4.160	3.902
93	cyclohexane	3.354	3.198	3.120
94	cyclohexene	2.810	2.786	2.830
95	cyclooctane	4.472	4.713	4.356
96	cyclopentane	2.795	2.836	2.998
97	cyclopentene	2.251	2.371	2.690
98	dibenz(ah)anthracene	6.838	6.659	6.925
99	dibenz(aj)anthracene	6.838	6.659	6.926
100	dimethyl sulfide	0.842	1.622	1.607
101	ethyl chloride	1.465	1.167	1.347
102	ethylbenzene	3.320	3.129	3.163
103	fluoranthrene	4.950	5.058	4.811
104	fluorene	4.225	4.461	4.167
105	fluorobenzene	2.285	2.069	2.278
106	fluorotrichloromethane	2.435	2.455	2.129
107	hexachlorobenzene	6.420	6.534	6.628
108	hexamethylbenzene	6.036	5.966	5.824
109	iodobenzene	3.265	3.579	3.707
110	isopropylbenzene	3.719	3.802	3.706
111	naphthalene	3.316	3.582	3.401
112	pentachlorobenzene	5.707	5.667	5.672
113	pentachloroethane	3.627	3.844	3.636
114	pentamethylbenzene	5.387	5.365	5.221
115	perylene	6.124	6.458	6.404
116	phenanthrene	4.490	4.866	4.635
117	pyrene	4.950	5.379	5.202
118	tetrachloroethylene	3.020	3.394	3.317
119	toluene	2.791	2.601	2.727
120	trichloroethylene	2.267	2.193	2.207

Table 3 (continuation)

121	triphenylene	5.664	5.924	5.795
122	m-xylene	3.440	3.300	3.367
123	n-decane	5.984	5.743	5.547
124	n-nonane	5.455	5.297	5.099
125	n-octane	4.926	4.813	4.645
126	n-undecane	6.513	6.153	5.989
127	n-heptane	4.397	4.277	4.179
128	n-butane	2.810	2.098	2.181
129	n-butylbenzene	4.378	4.191	4.072
130	n-hexane	3.868	3.681	3.705
131	n-pentane	3.339	3.035	3.231
132	n-propylbenzene	3.849	3.729	3.631
133	o-xylene	3.440	3.359	3.325
134	p-xylene	3.440	3.176	3.316
135	tert-amylbenzene	4.647	4.609	4.731
136	tert-butylbenzene	4.118	4.506	4.277
137	trans-1,2-dichloroethylene	1.514	1.567	1.548
	<i>Average absolute deviation</i>	–	0.20	0.16

Perhaps a close comparison between our best results and Basak *et al*'s best ones is in order. In fact, those authors presented their predictions for the following highly significant 6-variable model [19]:

$$\log P = -3.127 - 1.644(IC_0) + 2.120(\chi_C^5) - 2.914(\chi_{CH}^6) + 4.208(\chi^0) + 1.060(\chi^4) - 1.020(\chi_{PC}^4) \quad (n = 137, R^2 = 0.97, se = 0.26) \quad (3)$$

which can be compared with our present optimal 4-variables model:

$$\log P = -0.005 + 0.411(\chi^0)^{2.89089} + 0.223(IC_0)^{-2.932} - 2.796(\chi_{CH}^6) + 1.008(\chi^4)^{0.345562} \quad (n = 137, R^2 = 0.97, se = 0.24) \quad (4)$$

We have not accomplished a complete search of optimal non-integer exponents, but we deem that present results seem to highlight the convenience of looking for such real exponents in order to get suitable fitting equations. Evidently, a more thorough numerical search for optimal parameters most probably would yield even better predictive equations. At present, work along this line is being carried out in our laboratory and results will be presented elsewhere in the forthcoming future.

CONCLUSIONS

The primary purpose of this paper was to develop an improved calculation scheme to compute log P data on the basis of previous Basak *et al.*'s results using algorithmically defined structural variables. The success of our efforts is evident from results presented in Tables 2 and 3, which show that resorting to the employment of higher-order and non-integer exponents for the variables can efficiently predict log P values of a relatively homogeneous and large group of weakly hydrogen bonding chemicals. When going beyond linear relationships, a minor set of variables is needed in order to achieve a desirable accuracy (Compare statistical results corresponding to Eq. (9) in [19] with those of Eq. (29) in Table 2 of this paper). We have used graph invariants defined on linear graphs and multigraphs. While the list of molecular descriptors, chosen for this study, is the same as that employed by Basak *et al.* [19] and it is not exhaustive, the set of indices used here appear to optimally characterize the most important aspects of molecular structure related to the prediction of log P(octanol-water).

Further studies with other properties and other predictors are necessary in order to determine the significance of this procedure in the prediction of physicochemical properties and the characterization of molecular structure. At present we are involved in this sort of analysis and results will be presented elsewhere in the near future.

Acknowledgments

Authors thank very much several valuable referee's suggestions to improve the final version of this paper.

REFERENCES

1. Carrupt P., Testa B. and Gaillard P., in *Reviews in Computational Chemistry*, Lipkowitz K.B. and Boyd D.B. (Eds), Wiley-VCH, NY, 1997, **11**, 241–345.
2. Hansch C. and Leo A.J., *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, NY, 1979.
3. Hansch C. and Leo A., *Exploring QSAR Fundamentals and Applications in Chemistry and Biology*, American Chemical Society, Washington, DC, 1995.
4. Martin Y.C., *Quantitative Drug Design: A Critical Introduction*, Marcel Dekker, NY, 1978.
5. Hansch C., Maloney P.P., Fujita T. and Mui R., *Nature*, **194**, 178 (1963).
6. Hansch C., *Acc. Chem. Res.*, **2**, 232 (1969).
7. Leo A., Hansch C. and Elkins D., *Chem. Rev.*, **71**, 525 (1971).
8. Essex J.W., Reynolds C.A. and Richards W.G., *J. Chem. Soc., Chem. Commun.*, 1152 (1989).
9. Jorgensen W.L., Briggs J.M. and Contreras M.L., *J. Phys. Chem.*, **94**, 1683 (1990).
10. Eksterowicz J.E., Miller J.L. and Kollman P.A., *J. Phys. Chem. B*, **101**, 10971 (1997).
11. Nagy P. I., Dunn III, W.J. and Nicholas J.B., *J. Chem. Phys.*, **91**, 3707 (1989).
12. Reynolds C.H., *J. Chem. Inf. Comput. Sci.*, **35**, 738 (1995).
13. Dunn III, W.J. and Nagy P.I., *J. Comput. Chem.*, **13**, 468 (1992).
14. Ben-Naim A., *Solvation Thermodynamics*, Plenum Press, NY, 1987.
15. Reynolds C.H. and Best S.A., *Chemtec*, November 1998, 28.
16. Dearden J.C. and Bresnen G.M., *Quant. Struct.-Act. Relat.*, **7**, 133 (1988).
17. Gago F., Álvarez-Bonilla J. and Elguero J., *J. Liq. Chromat.*, **10**, 1031 (1987).
18. Domine D., Devillers J. and Karcher W., Chapter 2, in *Neural Networks in QSAR and Drug Design*, Devillers J., Ed., Academic Press, London, 1996, pp. 47–63 and references therein.

19. Basak S.C., Niemi G.J. and Veith G.D., *J. Math. Chem.*, **4**, 185 (1990).
20. SAS/STAT User Guide, release 6.03 edition, SAS Institute Inc., Cary, NC (1988), p. 1028.
21. Ou X.-C., Ouyang Y. and Lien E.J., *J. Mol. Sci. (China)*, **4**, 89 (1986).
22. Basak S.C., Niemi G.J. and Veith G.D., in *Computational Chemical Graph Theory*, Rouvray D.H., Ed., NY, 1989.
23. Cook R.D., *Technometrics*, **19**, 15 (1977).
24. Duchowicz P. and Castro E.A., *J. Korean Chem. Soc.*, **44**(6), 501 (2000).
25. Romanelli G.P., Cafferata L.F.R. and Castro E.A., *Khim. Fizika*, **19**(9), 105 (2000).
26. Duchowicz P. and Castro E.A., *Acta Chem. Slov.*, **47**, 281 (2000).
27. Firpo M., Gavernet L. and Castro E.A., *Rom. J. Phys.*, **44**(1,2), 181 (1999).
28. Castro E.A., *Comp. Chem.*, **21**(5), 305 (1997).
29. Livingstone D., *Data Analysis for Chemists*, Oxford Science Publications, Oxford University Press, 1995, Oxford, pp. 127–129.
30. Hansch C., Steward A.R., Anderson S.M. and Bentley J., *J. Med. Chem.*, **11**, 1 (1968).
31. Van de Waterbeemd H. and Kansy M., *Chimia*, **46**, 299 (1992).
32. Wolfram S., *The MATHEMATICA® Book*, Fourth Edition, Mathematica Version 4, Wolfram Media, Cambridge University Press, NY, 1999, p.106.