Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 2287–2293
Bioorganic & Medicinal Chemistry

QSAR study on phenolic activity: need of positive hydrophobic term (log *P*) in QSAR

Mamta Thakur,[a] Alok Agarwal,[b] Abhilash Thakur[c] and Padmakar V. Khadikar[d,*]

[a]*Department of Chemistry, Soft Vision College of Biotechnology and Science, 7, Green Park Colony, Indore 452001, India*
[b]*Department of Chemistry, Govt. Girls P.G. College Kilamaidan, Indore 452001, India*
[c]*Department of Chemistry, Bhopal Institute of Technology and Science Bangrasia, Bhopal, India*
[d]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd, 3 Khatipura, Indore 452007, India*

Received 16 November 2003; accepted 6 February 2004

Abstract—The phenolic activity ($\log 1/C$) of a large series of phenols against L1210 leukaemia cells was modelled using physicochemical parameters other than conventional electronic and steric parameters. Attempts have also been made to examine need or otherwise of the hydrophobic parameter, $\log P$, in such studies. The results have shown that contribution of $\log P$ in modelling $\log 1/C$ is favourable.

1. Introduction

Recently Hansch group[1] has published an excellent review on chem-bioinformatics and QSAR in that they critically reviewed QSAR lacking positive hydrophobic terms. Hansch is the first chemist to introduce the concept and technology of QSAR. His main approach is to use the electronic and steric parameters together with hydrophobic parameters based on octanol/water partition coefficient ($\log P$) for developing QSAR models. In the referred review[1] it was stated *''it is timely to examine those instances where hydrophobic terms are not significant''*. Consequently, different cases were critically examined and discussed in this review.[1] The parameters of Hansch interest were $\log P$, π (hydrophobic), Es, MR, MV, η, (electronic and steric); B_1, B_5, L (sterioneal parameters) and σ parameters. However, in the review[1] QSAR study using parameters other than those mentioned here are not discussed.

Consequent to above, we introduce yet another study in that we have used parachor (Pc), surface tension (ST), density (D), polarizability (Pol) parameters. Also, we have included molar refraction (MR), molar volume (MV), index of refraction (η) and hydrophobic parameter ($\log P$) for comparison. The parameter $\log P$ following the view of Hansch,[1] has been particularly used to investigate the need or/otherwise of hydrophobic parameter ($\log P$) in QSAR studies. In the present study, we have used $\log 1/C$ activity against the L1210 leukaemia cells, which are available in the literature.[2]

The present study is carried out under the following headings: (i) regression analysis using all the 69 compounds; (ii) further analysis depending upon the results obtained from the regression analysis and (iii) effect of ortho-substitution.

2. Results and discussion

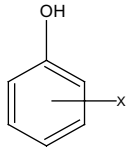
The original set of 69 phenols,[1,2] their activity ($\log 1/C$), $\log P$ and indicator parameters used are given in Table 1. Other physicochemical parameters used are presented in Table 2. The correlation of the parameters and their correlation with $\log 1/C$ is presented in Table 3.

A perusal of Table 1 shows that none of the parameters used singly correlates with the phenolic activity ($\log 1/C$). That is, no simple regression equation is possible for modelling the activity ($\log 1/C$). The results (Table 3) show that all the three physicochemical parameters (MR, MV, Pc) are mutually correlated.

Keywords: QSAR; Leukaemia; Hydrophobicity/lipophilicity; Regression analysis; Phenols; Molecular descriptors.
* Corresponding author. Tel.: +91-731-2531906; fax: +91-731-27636-18; e-mail: pvkhadikar@rediffmail.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.02.010

Table 1. Structural details, observed biological activity ($\log 1/C$), $\log P$ and indicator parameters used in present study



Comp. no.	Substituent (X)	$\log 1/C$	$\log P$	I_1	I_2	I_3	I_X	I_N
1	4-OMe	4.48	1.34	0	0	1	0	0
2	4-OC_2H_5	4.64	1.81	0	0	1	0	0
3	4-OC_3H_7	4.85	2.33	0	0	1	0	0
4	4-OC_4H_9	5.20	2.90	0	0	1	0	0
5	4-OC_6H_{13}	5.50	4.22	0	0	1	0	0
6	H	3.27	1.47	0	0	0	0	0
7	4-NO_2[a]	3.45	1.91	0	0	1	0	0
8	4-Cl[a]	4.29	2.39	0	0	1	1	0
9	4-I	3.86	2.91	0	0	1	1	0
10	4-CHO	3.08	1.35	0	0	1	0	0
11	4-F	3.83	1.77	0	0	1	1	0
12	4-NH_2	5.09	0.04	0	0	1	0	1
13	4-OH	4.59	0.59	0	0	1	0	0
14	4-Me	3.85	1.94	0	0	1	0	0
15	4-C_2H_5	3.86	2.47	0	0	1	0	0
16	4-NHCOMe[a]	3.73	0.51	0	0	1	0	0
17	4-CN[a]	3.44	1.60	0	0	1	0	0
18	4-OC_6H_5	4.97	3.35	0	0	1	0	0
19	BisphenolA	4.07	3.32	0	0	1	0	0
20	4-Br[a]	4.20	2.59	0	0	1	1	0
21	4-CMe_3	4.09	3.31	0	0	1	0	0
22	3-NO_2[a]	3.48	2.00	0	1	0	0	0
23	3-NHCOMe	2.65	0.73	0	1	0	0	0
24	3-Cl[a]	3.87	2.50	0	1	0	1	0
25	3-CMe_3	3.88	3.05	0	1	0	0	0
26	3-Me	3.54	1.96	0	1	0	0	0
27	3-OMe	3.71	1.58	0	1	0	0	0
28	3-NMe_2[a]	4.11	1.56	0	1	0	0	1
29	3-C_2H_5	3.71	2.40	0	1	0	0	0
30	3-Br[a]	3.82	2.63	0	1	0	1	0
31	3-CN	3.11	1.70	0	1	0	0	0
32	3-F	3.46	1.93	0	1	0	1	0
33	3-OH	3.46	0.80	0	1	0	0	0
34	3-NH_2[a]	4.11	0.21	0	1	0	0	1
35	2-Me	3.52	0	1	0	0	0	0
36	2-Cl	3.22	0	1	0	0	1	0
37	2-F	3.20	0	1	0	0	1	0
38	2-OMe	3.78	0	1	0	0	0	0
39	2-C_2H_5	3.75	0	1	0	0	0	0
40	2-OH	4.92	0	1	0	0	0	0
41	2-OH,4-Me	5.03	0	1	0	1	0	0
42	2-NH_2	5.16	0	1	0	0	0	1
43	2-CN[a]	3.30	0	1	0	0	0	0
44	2-NO_2[a]	3.34	0	1	0	0	0	0
45	2-Br[a]	3.44	0	1	0	0	1	0
46	2-CMe_3	4.00	0	1	0	0	0	0
47	4-C_3H_7	4.04	3.00	0	0	1	0	0
48	4-C_4H_9	4.33	3.64	0	0	1	0	0
49	4-C_5H_{11}	4.47	4.06	0	0	1	0	0
50	4-C_8H_{17}	4.62	5.68	0	0	1	0	0
51	4-$CONH_2$[a]	2.48	0.33	0	0	1	0	0
52	4-SO_2NH_2	2.50	0.06	0	0	1	0	0
53	4-C_7H_{15}	4.49	5.15	0	0	1	0	0
54	4-C_9H_{19}	4.75	6.21	0	0	1	0	0
55	Diethyl-stilbesterol	4.68	5.07	0	0	1	0	0
56	â-Estradiol	4.34	4.01	0	1	1	0	0
57	Equilin	4.10	2.90	0	1	1	0	0
58	Estriol	4.01	2.45	0	1	1	0	0
59	Equilenin	4.60	3.12	0	1	1	0	0
60	2-Naphthol	3.82	2.70	0	1	1	1	0
61	2-I[a]	3.95	0	0	0	0	0	0
62	2-SMe	3.70	0	1	0	0	0	0
63	2-$CHMe_2$	3.50	0	1	0	0	0	0
64	2-CH_2CHMe_2	3.90	0	1	0	0	0	0
65	2-CH_2OH	2.70	0	1	0	0	0	0
66	2-C_3H_7	3.49	0	1	0	0	0	0
67	2-CF_3[a]	3.22	0	1	0	0	0	0
68	2-$NHCONH_2$[a]	3.50	0	1	0	0	0	0
69	2-OC_2H_5	3.25	0	1	0	0	0	0

I_1 = substitution at second position, I_2 = substitution at third position, I_3 = substitution at fourth position, I_X = halogen group as substituents, I_N = amide substituents, $\log P$ = octanol/water partition coefficient, $\log 1/C$ = activity of phenols against L1210 leukaemia cells.
[a] Data points not included in calculation in Eqs. 1 and 3.

Table 2. Structural parameters for phenol derivatives used in present study

Comp. no.	MR	MV	Pc	η	ST	D	Pol
1	34.81	111.8	278.9	1.534	38.6	1.109	13.80
2	39.44	128.3	318.7	1.526	38.0	1.076	15.63
3	44.07	144.8	358.5	1.520	37.5	1.050	17.47
4	48.71	161.3	398.3	1.515	37.1	1.029	19.31
5	57.97	194.3	477.8	1.508	36.5	0.999	22.98
6	28.13	87.80	222.2	1.553	40.9	1.071	11.15
7	34.67	99.70	277.7	1.612	60.2	1.395	13.74
8	33.02	99.80	258.1	1.575	44.7	1.287	13.09
9	41.04	109.9	297.8	1.669	53.9	2.001	16.27
10	34.88	99.50	267.3	1.618	52.0	1.226	13.83
11	28.12	92.00	229.4	1.523	38.5	1.217	11.15
12	32.37	90.10	248.1	1.637	57.4	1.210	12.83
13	30.01	86.20	237.3	1.612	57.1	1.275	11.89
14	32.95	104.1	259.9	1.545	38.8	1.038	13.06
15	37.68	120.6	298.8	1.536	37.6	1.012	14.93
16	42.40	120.9	326.0	1.618	52.8	1.249	16.81
17	32.84	97.20	268.2	1.590	57.8	1.220	13.02
18	54.57	158.4	413.4	1.605	46.3	1.175	21.63
19	68.16	199.5	519.7	1.598	46.0	1.143	27.02
20	35.82	104.0	272.7	1.604	47.2	1.662	14.20
21	46.52	154.5	370.3	1.513	32.9	0.971	18.44
22	34.67	99.70	277.7	1.612	60.2	1.395	13.74
23	42.40	120.9	326.0	1.618	52.8	1.249	16.81
24	33.02	99.80	258.1	1.575	44.7	1.287	13.09
25	46.52	154.5	370.3	1.513	32.9	0.971	18.44
26	32.95	104.1	259.9	1.545	38.8	1.038	13.06
27	34.81	111.8	278.9	1.534	38.6	1.109	13.80
28	42.44	125.8	324.3	1.589	44.0	1.089	16.82
29	37.68	120.6	298.8	1.536	37.6	1.012	14.93
30	35.82	104.0	272.7	1.604	47.2	1.662	14.20
31	32.84	97.20	268.2	1.590	57.8	1.220	13.02
32	28.12	92.00	229.4	1.523	38.5	1.217	11.15
33	30.01	86.20	237.3	1.612	57.1	1.275	11.89
34	32.37	90.10	248.1	1.637	57.4	1.210	12.83
35	32.95	104.1	259.9	1.545	38.8	1.038	13.06
36	33.02	99.80	258.1	1.575	44.7	1.287	13.09
37	28.12	92.00	229.4	1.523	38.5	1.217	11.15
38	34.81	111.8	278.9	1.534	38.6	1.109	13.80
39	37.68	120.6	298.8	1.536	37.6	1.012	14.93

Table 2 (*continued*)

Comp. no	MR	MV	Pc	η	ST	D	Pol
40	30.01	86.20	237.3	1.612	57.1	1.275	11.89
41	34.84	102.5	274.9	1.594	51.6	1.210	13.81
42	32.37	90.10	248.1	1.637	57.4	1.210	12.83
43	32.84	97.20	268.2	1.590	57.8	1.220	13.02
44	34.67	99.70	277.7	1.612	60.2	1.395	13.74
45	35.82	104.0	272.7	1.604	47.2	1.662	14.20
46	46.52	154.5	370.3	1.513	32.9	0.971	18.44
47	42.31	137.1	338.6	1.529	37.1	0.992	16.77
48	46.94	153.6	378.4	1.522	36.7	0.977	18.61
49	51.58	170.2	418.2	1.518	36.4	0.964	20.44
50	65.48	219.7	537.5	1.507	35.8	0.939	25.95
51	37.06	106.5	296.7	1.612	60.1	1.286	14.69
52	40.71	116.8	328.2	1.614	62.3	1.482	16.14
53	60.84	203.2	497.8	1.510	36.0	0.946	24.12
54	70.11	236.2	577.3	1.505	35.6	0.932	27.79
55	83.22	242.2	631.8	1.603	46.2	1.107	32.99
56	79.50	232.6	615.4	1.599	48.9	1.170	31.51
57	77.47	218.7	590.6	1.626	53.1	1.220	30.71
58	81.09	229.6	622.1	1.624	53.8	1.255	32.14
59	79.39	214.2	582.9	1.662	54.7	1.242	31.47
60	45.97	121.9	326.1	1.677	51.0	1.181	18.22
61	41.04	109.9	297.8	1.669	53.9	2.001	16.27
62	40.92	117.4	311.2	1.613	49.2	1.190	16.22
63	42.31	137.9	337.3	1.525	35.7	0.987	16.77
64	46.90	154.0	375.8	1.520	35.4	0.975	18.59
65	34.58	101.6	275.8	1.595	54.1	1.220	13.71
66	42.31	137.9	337.3	1.525	35.7	0.987	16.77
67	33.11	121.3	279.4	1.457	28.1	1.335	13.12
68	41.30	108.1	315.2	1.689	72.3	1.407	16.37
69	39.44	128.3	318.7	1.526	38.0	1.076	15.63

MR = molar refractivity, MV = molar volume, Pc = parachor, η = refractive index, ST = surface tension, D = density, Pol = polarizability.

Thus, if any two of them or all the three are present in the regression expression then the model may suffer from the defect due to colinearly. However, their occurrence will be dealt with according to the recommendations made by Randic.[3]

No mutual correlation exists between molecular descriptors (MR, MV, Pc) and the indicator parameters used. It means that use of indicator parameters along with MR, MV and Pc will not result into models having co-linearity defect. However, model containing any two of the parameters MR, MV and Pc will suffer from the defect due to colinearity. Such cases will be dealt with following the recommendation made by Randic.[3]

The above discussion recommended to undertake multiple regression analysis for yielding statistically significant models for modelling the phenolic activity ($\log 1/C$). Selassie et al.[2] considered only 52 phenols from the data set for modelling the activity. Our results support such a selection of 52 compounds.

The preliminary regression analysis made by us indicates that the data set of 69 compounds needs to be divided (split) into two or more sets. The preliminary regression analysis indicated division of 69 compounds in two sets (categories). One of these categories contains 52 compounds and other one 17 compounds. We are not in a position to provide any structural evidence for such splitting. However, out of all probabilities the splitting may be because of their different mechanism of action. Such type of division into different sets is reported earlier by Rouvray and El-Basil,[4] Dais[5] and also by us.[6–12] We have, therefore, carried out regression on these divided sets. Stepwise regression analysis gave a penta-parametric model containing 51 compounds with significant statistics (one compound **12** has to be deleted being outlier):

Table 3. Correlation table- of various parameters and biological activity ($\log 1/C$)[a]

	$\log 1/C$	MR	MV	Pc	η	ST	D	Pol
$\log 1/C$	1.0000							
MR	0.3851	1.0000						
MV	0.4193	0.9694	1.0000					
Pc	0.3926	0.9935	0.9871	1.0000				
η	−0.1009	0.0431	−0.1924	−0.0465	1.0000			
ST	−0.2114	−0.0949	−0.2973	−0.1488	0.8936	1.0000		
D	−0.2416	−0.2181	−0.3669	−0.2790	0.6549	0.5815	1.0000	
Pol	0.3850	1.0000	0.9694	0.9935	0.0431	−0.0949	−0.2180	1.0000
I_1	−0.2124	−0.2587	−0.2594	−0.2687	0.0269	0.0144	0.0887	−0.2587
I_2	−0.1374	0.1434	0.0686	0.1107	0.2216	0.1431	0.0375	0.1433
I_3	0.4517	0.4432	0.4434	0.4567	−0.0172	−0.0113	−0.0863	0.4432
I_X	−0.1253	−0.2584	−0.2876	−0.2899	0.1068	−0.0394	0.4563	−0.2582
I_N	0.2619	−0.1376	−0.1797	−0.1579	0.2464	0.2088	−0.0202	−0.1377
$\log P$	0.4687	0.6523	0.7067	0.6755	−0.2185	−0.3096	−0.2641	0.6523

	I_1	I_2	I_3	I_X	I_N	$\log P$
I_1	1.0000					
I_2	−0.3356	1.0000				
I_3	−0.6121	−0.3214	1.0000			
I_X	0.0418	0.1019	−0.1124	1.0000		
I_N	−0.0366	0.1350	−0.1204	−0.1080	1.0000	
$\log P$	−0.6585	0.1655	0.5644	0.0245	−0.1880	1.0000

[a] The neutral correlation of the parameters is in term of r value.

$$\log 1/C = -9.53\ (\pm 1.49) + 0.07\ (\pm 0.01)\ \mathrm{MV} - 0.02\ (\pm 0.0043)\ \mathrm{Pc} + 8.03\ (\pm 1.58)\eta + 0.39\ (\pm 0.07) I_3 + 0.44\ (\pm 0.14)\ I_N$$
$$n = 51, \quad \mathrm{Se} = 0.1893, \quad r = 0.9084, \quad F = 42.481, \quad Q = 4.7987 \tag{1}$$

This model still contained two outliers (**8**,**24**), the removal of which gave improved statistics:

$$\log 1/C = -0.90\ (\pm 1.12) + 0.06\ (\pm 0.01)\ \mathrm{MV} - 0.02\ (\pm 0.0040)\ \mathrm{Pc} + 7.65\ (\pm 1.46)\eta + 0.36\ (\pm 0.06) I_3 + 0.46\ (\pm 0.06) I_N$$
$$n = 49, \quad \mathrm{Se} = 0.1742, \quad r = 0.9246, \quad F = 50.691, \quad Q = 5.3077 \tag{2}$$

No other higher parametric models, other than the above model (Eq. 2) gave improved statistics. Before discussing the model expressed by Eqs. 1 and 2 it is necessary to provide explanation why MV and Pc, which exhibits mutual correlation, are used in deriving the equations. Such situations are well discussed by Randic.[13] He has shown that selection of descriptors to be used in QSAR, QSPR and QSTR studies should not be delegated solely to the computers, although the statistical criteria will continue to be useful for preliminary screening of descriptors taken from large pool. Often in an automated selection of descriptors, descriptors will be discarded because it is highly correlated with another descriptor already selected, but what is important is not two descriptors parallel to one another, that is duplicate much of the same structural information, but whether they in those parts that are important for QSAR, QSPR and QSTR correlations. If they differ in the domains, which are important for the activity, property or toxicity considered, both descriptors should be retained. If they differ in parts that are not relevant for the correlation of considered property, activity or toxicity then one of them can be discarded. Furthermore Randic[13] has stated that one should preliminary be aware of a common fit all in regression analysis in describing descriptors that are highly inter-correlated. He stated that by discarding one of the descriptor, which commonly duplicates another we may be discarding a descriptor that nevertheless may carries useful structural information in the parts in which it does not parallel with the another descriptors. If the regression equation (model) containing highly mutually correlated descriptors have respective coefficients higher than their standard deviation then they are considered statistically significant. In view of these recommendations the presence of MV and Pc, which have high mutual correlation, can be used in deriving Eqs. 1 and 2 as above.

The above results shows that the first category gave pentaparametric model having MV, Pc, η, I_3 and I_N as the correlating parameters and ultimately containing 49 compounds. Eqs. 1 and 2 show that the phenolic activity ($\log 1/C$) is mainly governed by the various modes of substitution and the parameters MV and Pc.

The coefficients of indicator parameters (I_3 and I_N) in Eq. 2 are positive. This indicates that substitution at 2,3 and 4 positions are favourable for the exhibition of the phenolic activity.

With regards to other parameters involved in Eq. 2, it is interesting to record that both of them (MV and Pc) are linearly correlated and that MV is contained in the definition of Pc.[13] In the present case once again we have to argue for the highly mutual correlation between MV and Pc. The problems caused by multi-colinearity, and how to deal with them, continue to be of prime concern to theoretical satisfaction. From a discussion maker's view point, one should be aware of that multi-colinearity can exist and recognize the basic problem it can cause. Some of the most obvious problems and indication of highly mutual correlation are as follows:

(i) incorrect signs on the coefficients
(ii) change in the values of the previous coefficient when a new variable is added in the model and
(iii) an increase in the standard error of the estimation when a variable is added to the model.

The negative coefficient of Pc in Eq. 2 is therefore, because of the high mutual correlation between Pc and MV.

Basically, both these parameters (MV and Pc) account for the polarizability effect. Their positive coefficients in models (Eqs. 1 and 2) indicate that polarizability is responsible for the exhibition of phenolic activity ($\log 1/C$).

In order to investigate the role of hydrophobic effect ($\log P$) in the exhibition of $\log 1/C$ we have used $\log P$ as an addition correlating parameters. Now, the parameters used being: MV, Pc, η, I_3, I_N and $\log P$. Successive regression analysis using these six parameters ultimately resulted into the following statistically significant model:

$$\log 1/C = -6.6\ (\pm 1.46) + 0.05\ (\pm 0.01)\ \mathrm{MV} - 0.02\ (\pm 0.0043)\mathrm{Pc} + 0.048\ (\pm 0.03) \log P + 6.15\ (\pm 1.49)\eta + 0.30\ (\pm 0.08) I_3 + 0.75\ (\pm 0.13) I_N$$
$$n = 52, \quad \mathrm{Se} = 0.2005, \quad r = 0.9107, \quad F = 36.401, \quad Q = 4.5421 \tag{3}$$

This model (Eq. 3) contains three outliers (**9**,**12**,**28**), the deletion of which resulted into following regression model:

$$\log 1/C = -8.63 + 0.06\ (\pm 0.01)\ \mathrm{MV} - 0.02\ (\pm 0.0038)\mathrm{Pc} + 0.08(\pm 0.02) \log P + 7.47\ (\pm 1.38)\eta + 0.22\ (\pm 0.07) I_3 + 0.53(\pm 0.17) I_N$$
$$n = 49, \quad \mathrm{Se} = 0.1646, \quad r = 0.9340, \quad F = 47.819, \quad Q = 5.67 \tag{4}$$

This shows that statistics (Eq. 2) is improved by the addition of hydrophobic parameter($\log P$). This led us to conclude that in the case considered, that is, for the first category containing 52 compounds, the introduction of hydrophobic term is beneficial for QSAR modelling. Hence, the positive hydrophobic term ($\log P$) is a favourable requirement for the exhibition of $\log 1/C$. In other words, in addition to polarizability, hydrophobicity is favourable for exhibiting phenolic activity ($\log 1/C$).

We now discuss the results obtained for the category second containing 17 compounds. All these compounds taken together result into a model according to the following equation:

$$\log 1/C = 4.31\ (\pm 1.25) + 0.14\ (\pm 0.03)\ \mathrm{MV} - 0.06\ (\pm 0.02)\ \mathrm{Pc} + 0.99\ (\pm 0.51) I_1$$
$$n = 17, \quad \mathrm{Se} = 0.7904, \quad r = 0.7440, \quad F = 5.371, \quad Q = 0.9413 \tag{5}$$

No other regression resulted into a better regression expression than the above (Eq. 5).

Once again, for investigating the need of hydrophobic term in QSAR we have used $\log P$ as additional correlating parameter for obtaining model for this set of 17 compounds.

Addition of $\log P$ to above model (Eq. 5) gave significant improvement in the statistics yielding the following regression expression:

$$\log 1/C = 4.72\ (\pm 1.83) + 0.10\ (\pm 0.05)\ \mathrm{MV} - 0.04\ (\pm 0.02)\mathrm{Pc} + 0.48\ (\pm 0.31) \log P + 1.38\ (\pm 0.55) I_1$$
$$n = 17, \quad \mathrm{Se} = 0.7516, \quad r = 0.7910, \quad F = 5.050, \quad Q = 1.0542 \tag{6}$$

For a data set of 17 compounds, this model (Eq. 6) is a marginal case of rule of thumb. Furthermore, on passing from (Eq. 5) to (Eq. 6), the adjusted R^2 is increased from 0.4504 to 0.5031. This shows that use of $\log P$ term in Eq. 6 is well justified statistically. The physicochemical significant attached to models expressed by Eqs. 5 and 6 is similar to those discussed above.

The results obtained indicate that the compounds belonging to second category has poor statistics; this may be attributed to different mechanism of action of the compounds belonging to this category.

We now discuss the effect of ortho-substitution on the exhibition of $\log 1/C$. In the present study, there are 21 ortho-substituted phenols. The regression study on these 21 ortho-substituted compounds after deleting compounds 40 and 65 gave the following statistically significant model for modelling $\log 1/C$:

$$\log 1/C = 2.19\ (\pm 0.73) + 0.04\ (\pm 0.01)\ \mathrm{MR} + 1.60\ (\pm 0.21) I_3 + 1.82\ (\pm 0.21) I_N$$
$$n = 19, \quad \mathrm{Se} = 0.1987, \quad r = 0.9449, \quad F = 41.076, \quad Q = 4.755 \tag{7}$$

Introduction of $\log P$ term in the above model (Eq. 7) did not show any appreciable improvement in the statistics. The results (Eq. 7) show that when all the 21 ortho-substituted compounds are considered as a separate class, then there is no need of hydrophobic term ($\log P$) for the exhibition of phenolic activity ($\log 1/C$), that is, QSAR modelling related to phenolic activity. This may be attributed to be due to ortho-effect. This is in favour of earlier results,[1,2] which also favoured that an ortho-substitution did not require $\log P$ for modelling the activity.

The predictive power of the proposed models is discussed by calculating quality factor Q.[14,15] The Q value that is, R/Se indicates that the predictive power goes on increasing with increase in the R and decrease in Se. The same is followed in the models discussed above and shows that predictive powers follows the statistical qualities.

Table 4. Estimated biological activity ($\log 1/C$) from various models discussed in the text

Comp. no.	$\log 1/C$	$\log 1/C^a$	$\log 1/C^b$	$\log 1/C^c$	$\log 1/C^d$
1	4.48	3.83[e]	3.74[e]	4.51	5.03[e]
2	4.68	3.90[e]	3.84[e]	4.66	5.19[e]
3	4.85	3.40[e]	3.95[e]	4.84	5.36[e]
4	5.20	4.10[e]	4.07[e]	5.05	5.52[e]
5	5.50	4.33[e]	4.36[e]	5.54	5.85[e]
6	3.27	3.38	3.48	4.62[e]	3.19[e]
7	3.45	3.67	3.68	3.65[e]	5.03[e]
8	4.29	3.85[e]	3.87[e]	4.71	4.97[e]
9	3.86	4.29[e]	4.35[e]	4.27[e]	5.25[e]
10	3.08	3.94[e]	3.89[e]	3.80	5.03[e]
11	3.83	3.61	3.59	4.87[e]	4.79[e]
12	5.09	4.39[e]	4.32[e]	3.07[e]	6.77[e]
13	4.59	3.73[e]	3.65[e]	3.40	4.86[e]
14	3.85	3.85	3.83	4.84[e]	4.96[e]
15	3.86	3.95	3.94	5.06[e]	5.13[e]
16	3.73	3.96	3.83	3.01[e]	5.30[e]
17	3.44	3.56	3.54	3.66[e]	4.96[e]
18	4.97	4.25[e]	4.28[e]	4.34	5.73[e]
19	4.07	4.37	4.36	3.85[e]	6.22[e]
20	4.20	4.00	4.03	4.60[e]	5.07[e]
21	4.09	4.30	4.29	5.76[e]	5.45[e]
22	3.48	3.30	3.46	3.69[e]	3.42[e]
23	2.65	3.59[e]	3.62[e]	3.12	3.70[e]
24	3.87	3.48[e]	3.65	4.76[e]	3.36[e]
25	3.88	3.93	4.05	5.63[e]	3.84[e]
26	3.54	3.49	3.61	4.85[e]	3.36[e]
27	3.71	3.46	3.54	4.62[e]	3.43[e]
28	4.11	4.19	4.33[e]	4.06[e]	5.52[e]

(*continued on next page*)

Table 4 (*continued*)

Comp. no.	$\log 1/C$	$\log 1/C^{a}$	$\log 1/C^{b}$	$\log 1/C^{c}$	$\log 1/C^{d}$
29	3.71	3.58	3.71	5.03[e]	3.53[e]
30	3.82	3.64	3.81	4.62[e]	3.46[e]
31	3.11	3.19	3.33	3.70[e]	3.36[e]
32	3.46	3.25	3.38	4.94[e]	3.19[e]
33	3.46	3.37	3.44	3.50[e]	3.25[e]
34	4.11	4.02	4.11	3.15[e]	5.16[e]
35	3.52	3.49	3.46	5.30[e]	3.36
36	3.22	3.48	3.46	4.96[e]	3.36
37	3.20	3.25	3.23	5.41[e]	3.19
38	3.78	3.46	3.42	5.25[e]	3.43
39	3.75	3.58	3.53	5.27[e]	3.53
40	4.92	3.37[e]	3.38[e]	4.51	3.25[e]
41	5.03	3.77[e]	3.62[e]	4.51	5.03
42	5.16	4.02[e]	4.10[e]	4.43	5.16
43	3.30	3.19	3.20	4.28[e]	3.36
44	3.34	3.30	3.31	4.12[e]	3.42
45	3.44	3.64	3.61	4.75[e]	3.46
46	4.00	3.93	3.82	5.57[e]	3.84
47	4.04	4.03	4.04	5.25[e]	5.30[e]
48	4.33	4.12	4.16	5.48[e]	5.46[e]
49	4.47	4.23	4.28	5.63[e]	5.63[e]
50	4.62	4.57	4.68	6.20[e]	6.12[e]
51	2.48	3.66[e]	3.55[e]	2.76	5.11[e]
52	2.50	3.61[e]	3.47[e]	2.31	5.24[e]
53	4.49	4.45	4.54	6.01[e]	5.96[e]
54	4.75	4.69	4.82	6.38[e]	6.29[e]
55	4.68	4.56	4.64	4.13[e]	6.75[e]
56	4.34	4.29	4.32	3.38[e]	6.62[e]
57	4.10	4.17	4.15	2.54[e]	6.55[e]
58	4.01	4.13	4.06	2.06[e]	6.68[e]
59	4.60	4.34	4.33	2.53[e]	6.62[e]
60	3.82	4.47[e]	4.49[e]	4.15	5.43[e]
61	3.95	3.93	3.91	2.89[e]	3.65
62	3.70	3.68	3.64	4.44[e]	3.64
63	3.50	3.72	3.64	5.34[e]	3.69
64	3.90	3.83	3.72	5.29[e]	3.86
65	2.70	3.34[e]	3.34[e]	4.39	3.42[e]
66	3.49	3.72	3.64	5.34[e]	3.69
67	3.22	3.48	3.40	6.16[e]	3.36
68	3.50	3.56	3.57	3.36[e]	3.66
69	3.25	3.54	3.48	5.18[e]	3.59

[a] From Eq. 2.
[b] From Eq. 4.
[c] From Eq. 6.
[d] From Eq. 7.
[e] Data points not included in the corresponding models.

In order to confirm our findings, we have calculated predictive correlation coefficient (R^2_{pre}) by correlating estimated phenolic activity with the experimental activity. The data needed for such calculations are presented in Table 4 and used in Figures 1–4. The obtained predictive correlation coefficients R^2: 0.8549, 0.8724 and 0.8928 respectively for correlations demonstrated by Figures 1–4, and the closeness of estimated activity with the observed one confirms our findings discussed above.

Finally it is worthy to compare our results with those reported earlier.[1,2] The exact comparison is not possible as different parameters are used by us and by the earlier workers. For the ortho-substituted compounds our model is better than the earlier model, particularly because the number of compounds involved in our model is 19 as compared to 14 compounds used earlier. Furthermore, the earlier report concern with 52 compounds only and that they have not attempted QSAR for the remaining set of 17 compounds. That is, earlier, out of 69 compounds only 52 compounds were used for

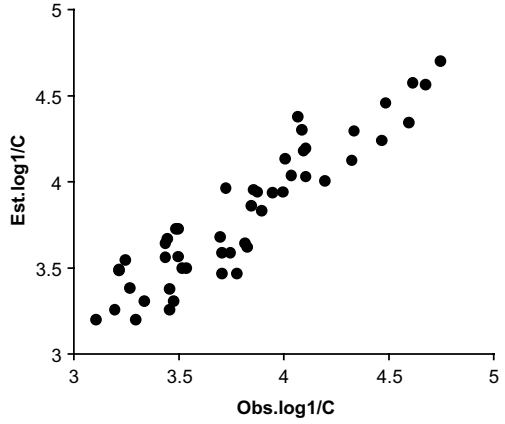


Figure 1. Correlation between observed and estimated activities ($\log 1/C$) using Eq. 2.

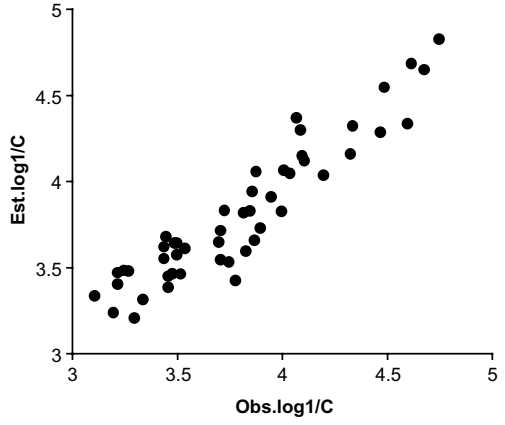


Figure 2. Correlation between observed and estimated activities ($\log 1/C$) using Eq. 4.

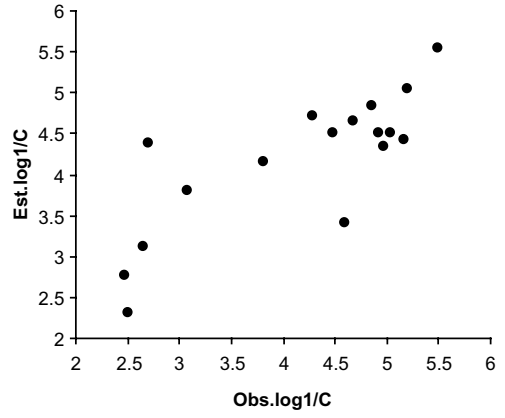


Figure 3. Correlation between observed and estimated activities ($\log 1/C$) using Eq. 6.

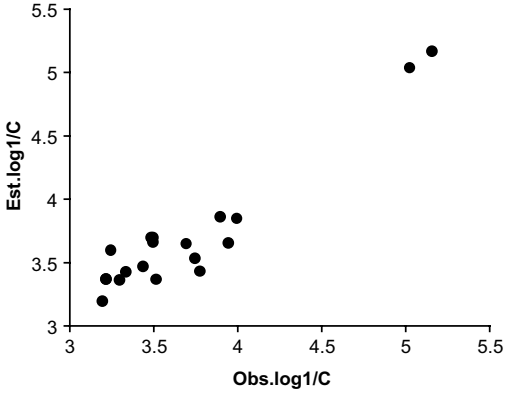


Figure 4. Correlation between observed and estimated activities ($\log 1/C$) using Eq. 7.

QSAR analysis. Our report on the other hand considered all the 69 compounds under split condition. However, for the set of 52 compounds, our statistics ($r = 0.934$) is slightly worse than the earlier statistics ($r = 0.959$).

3. Conclusion

From results and discussion made above we conclude that:

1. It is not possible to model the $\log 1/C$ activity of all the 69 phenols taken together; for obtaining statistically significant models the data set has to be split (divided) into two different categories (sets).
2. The compounds belonging to the separated data set requires a positive hydrophobic term ($\log P$) for exhibition of their phenolic activity ($\log 1/C$).
3. When ortho-substituted compounds are grouped as a separate class, then there is no need of any hydrophobic term ($\log P$) for the exhibition of the activity ($\log 1/C$).

4. Experimental

Activity—The phenolic activity ($\log 1/C$) of the phenols against L1210 leukaemia cells as well as hydrophobicity ($\log P$) were adopted from the literature.[1,2]

Molar volume (MV) and parachor (Pc)—The molar volume (MV) and parachor (Pc) for the set of phenols were calculated from ACD Lab software.[16]

Indicator parameters—These are the dummy parameters sometimes used for accounting those structural feature not covered in any molecular descriptor used. They assumed only two values 1 or 0. If the assumed structural feature is present; then the indicator parameters is 1 otherwise it is 0. The details of such parameters, used in the present study are already given in the Result and Discussion section (ref. Table 1).

Statistical analysis—Maximum R^2 method together with stepwise regression[17] was carried for arriving at statistically significant models.

Acknowledgements

Thanks are due to Prof. Hansch for providing information to carry out the present study. Special thanks are due to Dr. Rajni Garg for help extended by her. One of the authors (PVK) is thankful to Prof. Ivan Gutman for introducing him (PVK) to the fascinating field: chemical topology and graph theory and consequently QSAR.

References and notes

1. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. *Chem. Rev.* **2001**, *101*, 619.
2. Selassie, C. D.; Shusterman, A. J.; Kapur, S.; Verma, R. P.; Zharg, L.; Hansch, C. *J. Chem. Soc. Perkin Fren.* **1999**, 2729.
3. Randic, M. *Theor. Chem. Acta* **1983**, *62*, 485.
4. Rouvray, D. H.; El-Basil, S. *J. Mol. Str. (Theochem)* **1988**, *165*, 9.
5. Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 124.
6. Mathur, K. C.; Gupta, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *11*, 1915.
7. Khadikar, P. V.; Shrivastava, A.; Agrawal, V. K.; Shrivastava, S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3009.
8. Agrawal, V. K.; Bano, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *11*, 4039.
9. Khadikar, P. V.; Mandloi, D.; Shrivastava, A.; Phadnis, A. *Oxid. Commun.* **2003**, *26*, 161.
10. Khadikar, P. V.; Bajaj, A. V.; Mandloi, D. *Indian J. Chem.* **2002**, *10*, 2065.
11. Khadikar, P. V.; Pathre, S.; Shrivastava, A. *Bioorg. Med. Chem. Lett.* **2002**, *52*, 2673.
12. Thakur, A.; Thakur, M.; Khadikar, P. V. *Bioorg. Med. Chem.*, in press.
13. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
14. Pogliani, L. *Amino Acid* **1994**, *6*, 141.
15. Pogliani, L. *J. Phar. Chem.* **1996**, *100*, 18065.
16. ACD-Lab software for calculating the referred physicochemical parameters; Chemsketh S.O. www.Acdlabs.com.
17. Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; John Wiley & Sons: New York, 2000.