Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 1249–1255
Bioorganic & Medicinal Chemistry Letters

QSAR studies on 1,2-dithiole-3-thiones: modeling of lipophilicity, quinone reductase specific activity, and production of growth hormone

Padmakar Khadikar,* Mona Jaiswal, Madhu Gupta, Dheeraj Mandloi and Raj Singh Sisodia

Research Division, Laxmi Fumigation and Pest Control (P) Ltd, 3 Khatipura, Indore 452 007, India

Received 21 June 2004; accepted 18 November 2004
Available online 24 December 2004

Abstract—Studies on modeling of lipophilicity (log *P*) quinone reductase specific activity (logCDQR) and production of growth hormone (logCDGH) of 1,2-dithiole-3-thiones have been carried out using distance-based topological indices. The regression analysis of the data has shown that the set of compounds exhibit 'familial' relationships in that excellent results are obtained by dividing the data set into two or more classes (families).

1. Introduction

The use of log *P* as a correlating parameter as well as for expressing lipophilicity in biological studies is now well established. It is efficiently used as one of the molecular parameters in structure–activity relationship (SAR) studies related to medicinal chemistry, toxicology, pharmaceutical sciences, biological chemistry, and environmental research. The widespread applications of log *P* to biophysical processes involving xenobiotic explains the urgent need for both valid and quick procedures to quantify molecular lipophilicity.

Burgot et al.[1] reported relationships between detoxification properties of some 1,2-dithiole-3-thiones and their log *P*. This appears to be the first report on such relationships. The compounds used are the potent inducers of enzymes involved in the maintenance of reduced glutathione pools as well as phase-2 enzymes important to electrophile detoxification.[1,2] Furthermore, the compounds exhibit lower mammalian toxicity. They, therefore, have very promising properties.

Burgot et al.[1] classified 19 compounds into four different categories (families): (i) dithiolethiones substituted only in the fourth position (**2**,**3**,**4**,**5**,**6**); (ii) dithiolethiones substituted only in the fifth position (**7**,**8**,**9**,**10**,**11**,**12**,**13**); (iii) 4,5-di substituted dithiolethiones (**14**,**15**,**16**,**17**,**18**,**19**), and (iv) unsubstituted dithiolethiones (**1**). The only statistically significant correlation obtained by Burgot with disubstituted dithiolethiones (**15**,**16**,**17**,**18**) and the unsubstituted one. The disubstituted compounds **14** and **19** gave useless statistics ($r^2 = 0.23$). QSAR models for the compounds **1**,**15**,**16**,**17**,**18** for log 1/CDGH and log 1/CDQR were found by Burgot[1] as below:

$$\log(1/\mathrm{CDGH}) = -0.2911 \log P - 0.087 \quad (r^2 = 0.876) \tag{1}$$

$$\log(1/\mathrm{CDQR}) = -0.2591 \log P + 0.586 \quad (r^2 = 0.776) \tag{2}$$

indicating that the less lipophilic dithiolethiones are, the more active they are.

The report of Burgot[1] is statistically very poor as out of 19 compounds they have used only 5 compounds in arriving at statistically significant models. The proposed models, therefore, accounts for only 26% of compounds. The results of Burgot[1] also led us to conclude that the

Keywords: Lipophilicity; Quinone reluctance specific activity; Productive growth hormones; Dithioles; Thiones; Topological indices.
* Corresponding author. Tel.: +91 731 2531906; fax: +91 731 2701170; e-mail addresses: pvkhadikar@rediffmail.com; tomona@rediffmail.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.11.074

remaining compounds are outliers. Furthermore, Burgot[1] has not given any convincing reason for the omission of other compounds from the regression procedure. However, we believe that log *P* used by Burgot[1] is not a good parameter for modeling log(1/CDGH) as well as log(1/CDQR). Our earlier studies[3–7] show that perhaps the use of topological indices might help to solve the problem. This was, therefore, our primary aim of present investigation. The distance-based topological indices are usually found useful in such cases. Also, that the Wiener indexes (W)[8] is the oldest but by far not the best. The Randic as well as Kier & Hall connectivity[9–12] indices are the best and widely used. The Balaban index (J)[13], commonly refereed to as extended connectivity index, is the by far the most discriminating index. However, comparatively Balaban index is less used in QSAR studies. In many cases this index is not found that useful. In the present study also we observed that both Kier and Hall as well as Balaban indices are not useful and that the Winer and Randic indices alone yielded statistically significant models. It is worthy to mention that all these distance-based topological indices are highly correlated. Furthermore, they either belong to the first-generation indices or to the second-generation indices. According to Balaban,[14,15] such indices in spite of their observed degeneracy can be successfully used in QSAR studies. These indices are, therefore, fit for the present QSAR study in that pairs of compounds (**5** and **6** and **11** and **12**) have the same values for all these indices used in the present study. The occurrence of highly linearly correlated indices in the proposed models likewise will be dealt with in accordance with the recommendations made by Randic.

2. Results and discussion

The 1,2-dithiole- 3- thiones used in the present study along with their adopted[1] log *P*, logCDQR, logCDGH values are presented in Table 1. The calculated values of the topological indices (W, $^0\chi$, $^1\chi$, $^2\chi$) are given in Table 2. Using distance-based topological indices, we have applied Burgot[1] methodology to all the compounds, which they have classified under different categories (families), but we could not obtain any statistically significant model. This may be due to the fact that either the classifications made by Burgot[1] are not justified and/or log *P* is not an appropriate parameter to be used to model the lipophilic behavior of the compounds used. Furthermore, the statement made by Burgot[1] that 'the biological results obtained with functionalized di-thiolethiones must be read with caution because of possible metabolism may not be ignored', need also be taken care of.

Classification of compounds as made by Burgot[1] is not that uncommon. Randic[16] also made such classifications and called it as 'family' and that such 'familial' behavior are frequently used in developing quantitative structure–property relationships (QSPR/QSAR).[17–23] Following Burgot[1] and by using W, $^0\chi$, $^1\chi$, and $^2\chi$ indices we also observed that the data set can be split into two or three categories (Table 3). However, the number of categories and the compounds included there in and in the present study are different for different topological indices used for modeling the activities (Table 3). This might be due to different types of information content in the topological indices used. Another reason might be that different types of information content in the molecules control biological activity as well as their mechanisms of action. However, the results obtained in the present study are in favor of Burgot findings that better results are only obtained by splitting the data set in to two or more categories. It is interesting to record that the results obtained in the present study indicates that lipophilicity (log *P*) can be modeled using distance-based topological indices. Earlier,[24,25,23] we have also shown

Table 1. Structural details of 1,2-dithiole-3-thiones their log *P*, logCDQR, logCDGH values, and estimated topological indices

Compd No.	R_1	R_2	log *P*	logCDQR	logCDGH	W	$^0\chi$	$^1\chi$	$^2\chi$
1	H	H	1.58	0.176	0.415	26	4.4058	2.8938	2.3896
2	H	CH_3	2.18	1.079	1.114	40	5.276	3.3045	2.8855
3	H	C_2H_5	2.67	0.903	1.491	61	5.9831	3.8425	3.0765
4	H	C_6H_5	3.20	0.301	1.255	191	8.3889	5.877	4.9701
5	H	CO_2H	1.33	1.380	1.477	84	6.8534	4.2152	3.8161
6	H	$CONH_2$	0.30	1.662	1.579	84	6.8534	4.2152	3.8161
7	CH_3	H	1.85	1.230	1.114	41	5.276	3.2877	3.0234
8	C_2H_5	H	2.31	0.903	1.415	63	5.9831	3.8257	3.1925
9	C_6H_5	H	3.67	1.301	1.322	197	8.3889	5.8602	5.0764
10	$C(CH_3)_3$	H	3.25	0.699	1.477	113	7.776	4.499	4.9812
11	CO_2H	H	1.05	1.903	1.903	87	6.8534	4.1984	3.9223
12	$CONH_2$	H	0.01	1.301	0.653	87	6.8534	4.1984	3.9223
13	C_6H_4(P)OCH_3	H	3.82	1.255	1.000	318	9.9663	6.792	5.8674
14	CH_3	CH_3	2.45	1.903	1.875	58	6.1463	3.7152	3.391
15		CH_3	2.79	1.342	0.903	238	9.2591	6.2877	5.4757
16	$(CH_2)_4$	—	3.10	0.698	0.929	106	6.9747	4.877	4.263
17	C_6H_5(P)OCH_3	CH_3	4.10	1.580	1.301	374	10.8365	7.2195	6.2667
18		—	1.80	0.342	0.778	104	7.1378	4.7877	4.437
19	$(CH_2)_8$	—	2.53	−0.602	−1.000	77	6.2676	4.377	3.9095

Table 2. Regression models for modeling log *P*, logCDQR, and logCDGH along with their regression parameters and quality of correlation

Modeling of log P	
Wiener index (*W*)	
I.	$\log P = 1.3507 + 0.0204W$
II.	$\log P = -0.4119 + 0.0202W$
III.	$\log P = -0.9862 + 0.0145W$
Zero-order connectivity index ($^0\chi$)	
I.	$\log P = -0.1820 + 0.4547\,^0\chi$
II.	$\log P = -0.0990 + 0.3922\,^0\chi$
III.	$\log P = -2.9630 + 0.6252\,^0\chi$
First-order connectivity index ($^1\chi$)	
I.	$\log P = -0.0340 + 0.9940\,^1\chi$
II.	$\log P = 0.3123 + 0.5583\,^1\chi$
III.	$\log P = -0.1984 + 0.7649\,^1\chi$
Second-order connectivity index ($^2\chi$)	
I.	$\log P = 0.2829 + 0.6971\,^2\chi$
II.	$\log P = -0.1919 + 0.6858\,^2\chi$
III.	$\log P = -2.6027 + 0.9847\,^2\chi$
Modeling of log CDQR	
Wiener index (*W*)	
I.	$\log \mathrm{CDQR} = -0.5900 + 0.0430W$
II.	$\log \mathrm{CDQR} = -0.0324 + 0.0061W$
III.	$\log \mathrm{CDQR} = -0.7083 + 0.0060W$
Zero-order connectivity index ($^0\chi$)	
I.	$\log \mathrm{CDQR} = -0.9461 + 0.3981\,^0\chi$
II.	$\log \mathrm{CDQR} = -1.9318 + 0.4762\,^0\chi$
III.	$\log \mathrm{CDQR} = -2.3298 + 0.4055\,^0\chi$
IV.	$\log \mathrm{CDQR} = -3.7118 + 0.4902\,^0\chi$
First-order connectivity index ($^1\chi$)	
I.	$\log \mathrm{CDQR} = -1.5706 + 0.8270\,^1\chi$
II.	$\log \mathrm{CDQR} = -2.7213 + 0.9768\,^1\chi$
III.	$\log \mathrm{CDQR} = -2.4766 + 0.6230\,^1\chi$
IV.	$\log \mathrm{CDQR} = -4.0863 + 0.7780\,^1\chi$
Second-order connectivity index ($^2\chi$)	
I.	$\log \mathrm{CDQR} = -0.9249 + 0.7015\,^2\chi$
II.	$\log \mathrm{CDQR} = -0.8902 + 0.5745\,^2\chi$
III.	$\log \mathrm{CDQR} = -0.8163 + 0.3942\,^2\chi$
IV.	$\log \mathrm{CDQR} = -3.0105 + 0.7250\,^2\chi$
Modeling of log CDGH	
Wiener index (*W*)	
I.	$\log \mathrm{CDGH} = 0.9431 + 0.0059W$
II.	$\log \mathrm{CDGH} = 0.2648 + 0.0053W$
III.	$\log \mathrm{CDGH} = -1.3984 + 0.0077W$
Zero-order connectivity index ($^0\chi$)	
I.	$\log \mathrm{CDGH} = -1.5198 + 0.4983\,^0\chi$
II.	$\log \mathrm{CDGH} = -1.0714 + 0.3567\,^0\chi$
III.	$\log \mathrm{CDGH} = -1.7208 + 0.3587\,^0\chi$
IV.	$\log \mathrm{CDGH} = -4.2099 + 0.5146\,^0\chi$
First-order connectivity index ($^1\chi$)	
I.	$\log \mathrm{CDGH} = 0.1612 + 0.3307\,^1\chi$
II.	$\log \mathrm{CDGH} = 0.0317 + 0.3303\,^1\chi$
III.	$\log \mathrm{CDGH} = 0.2516 + 0.1519\,^1\chi$
IV.	$\log \mathrm{CDGH} = 0.0392 + 0.1423\,^1\chi$
Second-order connectivity index ($^2\chi$)	
I.	$\log \mathrm{CDGH} = -3.1112 + 1.4769\,^2\chi$
II.	$\log \mathrm{CDGH} = -3.1307 + 1.4240\,^2\chi$
III.	$\log \mathrm{CDGH} = -2.3856 + 0.7759\,^2\chi$
IV.	$\log \mathrm{CDGH} = -3.0575 + 0.8650\,^2\chi$
V.	$\log \mathrm{CDGH} = -4.8337 + 1.0009\,^2\chi$

that PI (Padmakar–Ivan) index can be used in place of log *P* for modeling lipophilicity. In view of this, and using 'familial' relationships we have modeled lipophilicity(log *P*). In addition, we extended our methodology for modeling quinone reductance specific activity (logCDQR) as well as production of growth hormones (logCDGH) for the set of 1,2-dithiole-3-thiones used (Table 1, Fig. 1):

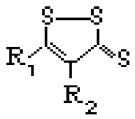

Figure 1. General structure of 1,2-dithiole-3-thiones (ref. Table 1).

In the parameters logCDQR and logCDGH, CD stands for the concentration of dithiolethiones, QR is specific activity of quinone reductance and GH is the production of growth hormone in peculiar cells.

2.1. Modeling of log *P*

The adopted values of log *P*, logCDQR, and logCDGH are presented in Table 1. The information regarding topological indices used (Table 2), regression models obtained (Table 3) and splitting of the data set into different categories and regression parameters for modeling log *P* of the 1,2-dithiole-3-thiones are presented in Table 4. This Table 4 shows that in modeling log *P* and by using the referred set of topological indices the data set splits into three categories: I, II, and III. In all the three categories, log *P* is linearly related to the topological index used. That is, lipophilicity of the compound increases with the increase in the value of the topological index under consideration. The data presented in Table 4 also show that using W index for modeling log *P* excellent results are obtained in each of the splitted categories. It is found to be the case when Randic connectivity indices ($^0\chi$, $^1\chi$, $^2\chi$) are used for modeling log *P*. However, since different compounds are included in the splitted categories depending upon the use of topological indices we cannot compare the results. To resolve this difficulty we have estimated log *P* values and obtained their correlation with the observed values of log *P*. This resulted into predictive correlation coefficients (R_{pred}) the use of which is then made to establish relative correlation potential of *W*, $^0\chi$, $^1\chi$, and $^2\chi$ for modeling log *P*.

Using the respective regression expressions and the topological indices under splitted condition we have estimated values of log *P*. The following correlations are obtained between observed (experimental) and calculated (estimated) log *P*:

Using wiener index (*W*)

$$\log P_{(\mathrm{obs})} = 0.0025 + 1.0016(\pm 0.0380)\log P_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1794,\ r = 0.9880,\ Q = 5.5077 \qquad (3)$$

Table 3. Regression parameters and quality of correlation's

S. No.	Property/activity	T.I.	Splitting	S_x	S_y	R
1.	log P	W	I. 2,3,14, 1,7,8,10,16,19	11.3578	0.2454	0.9453
			II. 4,5,9,11,18	56.6005	1.1619	0.9854
			III. 6,12,13,15,17	132.1257	1.9358	0.9889
		$^0\chi$	I. 2,3,9,10,16,19	1.1643	0.5410	0.9786
			II. 1,4,7,8,13,14,17	2.4609	0.9688	0.9961
			III. 5,11,15,18	1.1626	0.7638	0.9516
		$^1\chi$	I. 2,3,10,14	0.4935	0.4549	0.9940
			II. 1,4,7,8,9,13,16,17,19	1.5329	1.1086	0.7719
			III. 5,11,15,18	0.9826	0.7638	0.9840
		$^2\chi$	I. 1,2,3,8,9,14,16	0.9089	0.7367	0.8600
			II. 4,7,10,13,17,19	1.2078	0.8288	0.9995
			III. 5,11,15,18	0.7587	0.7638	0.9781
2.	logCDQR	W	I. 2,7,14, 3,5,6,8,11,12	10.1160	0.4387	0.9924
			II. 1,9,10,15,16,18	15.5081	0.4798	0.9545
			III. 4,13,17,19	162.7933	0.9864	0.9953
		$^0\chi$	I. 2,6,7,11,14	0.9107	0.3807	0.9524
			II. 1,3,5,8,12	0.9997	0.4770	0.9979
			III. 9,10,15,16,18	0.9406	0.4318	0.8831
			IV. 4,13,17,19	2.0061	0.9864	0.9969
		$^1\chi$	I. 2,7,11,14	0.5210	0.4387	0.9822
			II. 1,3,5,6,8,12	0.5102	0.5203	0.9579
			III. 9,10,15,16,18	0.7387	0.4853	0.9482
			IV. 4,13,17,19	1.2578	0.9864	0.9921
		$^2\chi$	I. 2,6,7,11,14	0.5329	0.3807	0.9820
			II. 3,5,8,12	0.4290	0.2546	0.9680
			III. 1,9,15,16	0.6180	0.3606	0.9632
			IV. 4,10,13,17,18,19	0.7436	0.5653	0.9537
3.	logCDGH	W	I. 2,5,6,7,8,10	28.3790	0.1997	0.8418
			II. 1,4,9,12,16,18	63.3200	0.3509	0.9852
			III. 13,15,17,19	129.1778	1.0478	0.9548
		$^0\chi$	I. 2,3,7,8,11,14	0.6515	0.3258	0.9964
			II. 1,5,6,10	1.4448	0.5501	0.9367
			III. 4,9,12,16,18	0.7736	0.2930	0.9473
			IV. 13,15,17,19	2.4260	1.2507	0.9982
		$^1\chi$	I. 3,6,8,11,14	0.2202	0.0821	0.9035
			II. 2,5,7,10	0.6235	0.2096	0.9825
			III. 4,9,16,17	1.1754	0.2028	0.8804
			IV. 1,12,13,15,18	1.5806	0.2282	0.9858
		$^2\chi$	I. 2,3,14	0.2553	0.3805	0.9909
			II. 1,5,6,7,8,11	0.6025	1.1888	0.7217
			III. 10,12,16	0.5405	0.4194	0.9999
			IV. 4,9,18	0.3426	0.2966	0.9991
			V. 13,15,17,19	1.0320	1.0478	0.9859

Using zero-order connectivity index ($^0\chi$)

$$\log P_{(\mathrm{obs})} = 5.8310 \times 10^{-4} + 0.9998(\pm 0.0384) \log P_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1342,\ r = 0.9855,\ Q = 7.3699 \quad (4)$$

Using first-order connectivity index ($^1\chi$)

$$\log P_{(\mathrm{obs})} = 0.0976 + 0.7348(\pm 0.2232) \log P_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.6952,\ r = 0.6476,\ Q = 0.9315 \quad (5)$$

Using second-order connectivity index ($^2\chi$)

$$\log P_{(\mathrm{obs})} = -0.0510 + 0.9984(\pm 0.0444) \log P_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1549,\ r = 0.9855,\ Q = 6.3608 \quad (6)$$

The above-mentioned regression expressions show that all the topological indices have similar power for modeling log P. The results also show that the splitting of the data set into different categories (families) is well justified.

The predictive power of the model can be established on the basis of the magnitude of quality factor, Q. This is

Table 4. Calculated values of VIF, PE, SDR, R^2_{pred}

Model (eq)	PE	R^2_{pred}
1 (3)	0.0037	0.9761
2 (4)	0.0044	0.9712
3 (5)	0.0156	0.8980
4 (6)	0.0044	0.9712
5 (8)	0.0086	0.9438
6 (9)	0.0054	0.9651
7 (10)	0.0058	0.9624
8 (11)	0.0073	0.9526
9 (13)	0.0098	0.9526
10 (14)	0.0037	0.9761
11 (15)	0.0048	0.9688
16 (16)	0.0175	0.9688

found to be the highest for the model based on $^0\chi$. In fact, based on Q-values following order of predictive power is obtained:

$$^0\chi > {}^2\chi > W > {}^1\chi \tag{7}$$

The results, therefore, show that the number and type of atoms are mainly responsible for the exhibition of lipophilicity (log P).

2.2. Modeling of logCDQR

The situation for modeling quinone reductance specific activity (logCDQR) is like modeling of log P. Here also, the data set splits into three or four categories (Table 4). Table 4 shows that $^1\chi$ and $^2\chi$ appears to be good parameters for modeling logCDQR. In other two cases comparatively lower quality statistics is obtained. For W this category is the I containing: **1**, **2**, **3**, **5**, **6**, **7**, **8**, **11**, **12**, **14** compounds. In case of $^0\chi$ such category is the category III containing compounds: **9**, **10**, **15**, **16**, and **18**. The effect of such a lower statistics on the overall effect of modeling could not be judged from these results. What we need further is to calculate logCDQR activity for the entire set compounds using the models under the splitted categories. This will provide predictive correlation coefficient ($R_{pred}{}^2$), which will help us to decide which topological index is favorable for modeling logCDQR.

Using W, $^0\chi$, $^1\chi$, and $^2\chi$ indices, the following correlation is observed between observed and calculated logCDQR:

Using wiener index (W)

$$\log \mathrm{CDQR}_{(\mathrm{obs})} = -0.1275 + 1.1023(\pm 0.0653)\log \mathrm{CDQR}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.15582,\ r = 0.9715,\ Q = 6.2344 \tag{8}$$

Using zero-order connectivity index ($^0\chi$)

$$\log \mathrm{CDQR}_{(\mathrm{obs})} = -2.2406 \times 10^{-4} + 1.0002(\pm 0.0475)\log \mathrm{CDQR}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1190,\ r = 0.9824,\ Q = 8.2536 \tag{9}$$

Using first-order connectivity index ($^1\chi$)

$$\log \mathrm{CDQR}_{(\mathrm{obs})} = 7.7593 \times 10^{-5} + 1.0000(\pm 0.0511)\log \mathrm{CDQR}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.0.1271,\ r = 0.9810,\ Q = 7.7163 \tag{10}$$

Using second-order connectivity index ($^2\chi$)

$$\log \mathrm{CDQR}_{(\mathrm{obs})} = 7.9552 \times 10^{-4} + 0.9999(\pm 0.0576)\log \mathrm{CDQR}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1109,\ r = 0.9760,\ Q = 8.8051 \tag{11}$$

The above regression expressions once again show that $^0\chi$ is the most appropriate topological index for modeling logCDQR. Also, that the splitting of the compounds in different categories (families) is once again well justified. The value of Q favors this finding. Based on Q value the following order of predictive power is obtained:

$$^0\chi > {}^2\chi > {}^1\chi > W \tag{12}$$

This order is little bit different from that obtained for log P. Unlike log P, here $^1\chi$ is found better than W index. That is, number of atoms and first-order branching play their role in the exhibition of logCDQR.

2.3. Modeling of logCDGH

Like the earlier two cases the modeling of logCDGH have to be considered under two or more categories (Table 4). In order to find out which topological index is better for modeling logCDGH, again we have estimated the activities of all the 19 compounds under splitted condition and compared them with the experimental values.

Based on W, $^0\chi$, $^1\chi$, and $^2\chi$ the following correlations are obtained between observed and estimated logCDGH:

Using wiener index (W)

$$\log \mathrm{CDGH}_{(\mathrm{obs})} = 0.0028 + 1.0003(\pm 0.0699)\log \mathrm{CDGH}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1623,\ r = 0.9675,\ Q = 5.9620 \tag{13}$$

Using zero-order connectivity index ($^0\chi$)

$$\log \mathrm{CDGH}_{(\mathrm{obs})} = 2.301 \times 10^{-4} + 0.9999(\pm 0.0404)\log \mathrm{CDGH}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.1034,\ r = 0.9880,\ Q = 9.5515 \tag{14}$$

Using first-order connectivity index ($^1\chi$)

$$\log \mathrm{CDGH}_{(\mathrm{obs})} = 0.0078 + 0.9800(\pm 0.0488)\log \mathrm{CDGH}_{(\mathrm{cal})}$$
$$n = 19,\ Se = 0.0633,\ r = 0.9843,\ Q = 15.5395 \tag{15}$$

Using second-order connectivity index ($^2\chi$)

$$\log \mathrm{CDGH}_{(\mathrm{obs})} = 0.1706 + 0.7715(\pm 0.0673) \log \mathrm{CDGH}_{(\mathrm{cal})} \quad (16)$$
$$n = 19,\ Se = 0.2230,\ r = 0.9410,\ Q = 4.2197$$

The above regression expressions show that almost 1:1 correlation is estimated between observed and calculated logCDGH. This, once again favors the splitting of the data set. Unlike, earlier two cases, here the first-order connectivity index ($^1\chi$) is found most appropriate for modeling logCDGH. That is, here also the number and kind of atoms together with first-order branching favors the exhibition of logCDGH.

Based on the values of quality factor (Q) the following order of predictive power is obtained.

$$^1\chi > {}^0\chi > W > {}^2\chi \quad (17)$$

It is worthy to mention that excellent correlations between observed and estimated parameters (log P, logCDQR, logCDGH) indicate that the splitting of the data set into two or more categories is well justified. In spite of such success one may argue that the methodology used by us may not be good for predicting log P, logCDQR, and logCDGH. However, such an approach is well known[19,20] and favors different types of mechanism of action at the receptor site.

Further support in favor of the methodology used by us is obtained through the calculations of variance inflation factor (VIF), probable error of the coefficient of correlation (PE), deviation of and predictive R^2 (R^2_{pred}). The calculated values of these parameters are presented in Table 4.

The physico-chemical significance attached to the proposed mathematical models:

The results discussed above indicate that the mathematical modeling is well done and it is, therefore, needed that they should be correlated with the physico-chemical significance of the topological indices involved in these mathematical models. The Winer index (W) accounts for the size, shape and branching. Randic indices ($^0\chi$, $^1\chi$, $^2\chi$) on the other hand takes care of the number of atoms in the molecule and the order of branching.

The $^0\chi$ index is responsible for the former and the other two indices, that is, $^1\chi$ and $^2\chi$ convey more information about the number of atoms in the molecule and incode more information about branching. The results summarized in Eq. 7 indicates that $^0\chi$ is the best index for modeling log P meaning there by that the number of atoms and consequently the size of the compound is favorable for the exhibition of log P. The same is found to be the case for modeling logCDQR (Eq. 12). However, the physico-chemical significance attached in modeling logCDGH is bit different. Here, Eq. 17 shows that the most suitable topological index in this case is the $^1\chi$ index. That is, increased number of atoms and first-order branching are responsible for the exhibition of logCDGH.

3. Conclusions

From the results and discussion made above we conclude that methodology used by us is far superior to the Burgot[1] methodology and agrees with his argument of splitting of the data set into different categories. However, Burgot[1] approach was incorrect. Firstly, because out of 19 data set (compounds) they have used only five compounds in proposing their QSAR model. Secondly, they have not given any convincing reason for the omission of other compounds from the regression procedure used for them. Our methodology on the other hand seems to be more useful. In our case all the compounds are included in obtaining QSAR model and the number of splitted categories is quite a small. It is worth to mention that such splitting of the data set into different categories is not uncommon. This was observed earlier by Randic[15], Dias,[21] Rouvary,[22] and recently by us also.[5,23–25] The splitting of the data set further suggests different mechanism of action of the compounds belonging to that particular category.

4. Experimental

4.1. Physiological activity

Lipophilicity (log P), quinone reductase specific activity (log CDQR) and production of growth hormone (log CDGH) were adopted as reported by Burgot et al.[1]

4.2. Topological indices

Topological indices (W, $^0\chi$, $^1\chi$, $^2\chi$) used in the present study were computed from the hydrogen-suppressed graphs in that all the carbon–hydrogen as well as heteroatom–hydrogen bonds are deleted. The details for the calculation of these indices are available in the literature;[26–28] however, below we give the expressions used for their calculations.

Wiener index (W)—The Wiener index (W) of a molecular graph can be estimated from the following expression:

$$W = W(G) = 1/2 \sum (Dij) \quad (18)$$

where the summation ($\sum$) of (Dij) (the ijth element of distance matrix (D)) denotes the shortest graph theoretic distance between j and I vertices (atoms) of a molecular graph, G.

Randic connectivity indices ($^0\chi$, $^1\chi$, $^2\chi$): The general expression used for the calculations of $^0\chi$, $^1\chi$, and $^2\chi$ is as below:

$$^m\chi^R = \sum_{(i,j,k)} (\delta_i, \delta_j, \delta_k)^{-0.5} \quad (19)$$

Regression analysis: The regression analysis is performed using maximum R^2 method in a stepwise fashion.[29]

Software's: The topological indices were calculated using the software prepared by one of us (RSS), while the regression analysis were performed using Regress-1 program of Lukovits.

Acknowledgements

Authors thanks are due to Prof. Lukovits for providing software. One of the authors (P.V.K.) is thankful to Prof. Ivan Gutman for introducing him (P.V.K.) to the fascinating field: Chemical Topology and Graph Theory.

References and notes

1. Burgot, G.; Bona, N.; Christen, M. V.; Burgot, J. L. *Int. J. Pharm.* **1996**, *129*, 295.
2. Dimitrov, N. V.; Benett, J. L.; Millan, J.; Perlott, M.; Leece, C. M.; Malove, W. *Invest. New Drugs* **1992**, *10*, 289.
3. Khadikar, P. V.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1125.
4. Khadikar, P. V.; Agrawal, V. K.; Karmarkar, S. *Bioorg. Med. Chem.* **2002**, *10*, 3499.
5. Agrawal, V. K.; Singh, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3981.
6. Agrawal, V. K.; Bano, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *11*, 4039.
7. Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmarkar, S.; Khadikar, P. V. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57.
8. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
9. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6604.
10. Randic, M. *J. Mol. Graphics Modeling* **2001**, *10*, 19.
11. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies: Litchworth, UK, 1986.
12. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Research Studies: Litchworth, UK, 1976.
13. Balaban, A. T. *J. Chem. Phys. Lett.* **1982**, *189*, 399.
14. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
15. Balaban, A. T. *Math. Chem.* **1986**, *21*, 115.
16. Randic, M. *Theor. Chem. Acta* **1983**, *62*, 485.
17. Jaiswal, M.; Khadikar, P. V. *Bioorg. Med. Chem.* **2004**, *12*, 1793.
18. Thakur, M.; Thakur, A.; Khadikar, P. V. *Bioorg. Med. Chem.* **2004**, *12*, 82.
19. Dias, J. R. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 124.
20. Rouvary, D. H.; EL-Basil, S. *J. Mol. Str. (Theochem)* **1988**, *165*, 9.
21. Mathur, K. C.; Gupta, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *13*, 1915.
22. Khadikar, P. V.; Bajaj, A. V.; Mandloi, D. *Indian. J. Chem.* **2002**, *41A*, 2065.
23. Singh, S.; Joshi, S.; Shrivastava, A.; Khadikar, P. V. *J. Sci. Ind. Res.* **2002**, *61*, 961.
24. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
25. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 143.
26. Trinajstic, N. In *Chemical Graph Theory*; CRC: Boca Raton, FL, 1983; Vols. I and II.
27. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
28. Diudea, M. V.; Khadikar, P. V. *Molecular Topology and Its Applications*; New Age Int: New Delhi, in press.
29. Chatterji, S.; Hadi, S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.