Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 97, No. 1, July 1986, pp. 91–95.

QSAR studies on estrogen receptor binding affinity of 2-phenylindoles using first-order valence molecular connectivity

P SINGH

Department of Chemistry, Birla Institute of Technology and Science, Pilani 333031, India

MS received 1 August 1985; revised 10 January 1986

Abstract. The estrogen receptor binding affinity of 2-phenylindoles is found to be significantly correlated with Kier's first-order valence molecular connectivity index. The correlation equations obtained provide a much simple rationale to design more potent compounds and help in predicting the potency of new compounds.

Keywords. Quantitative structure-activity relationship; 2-phenylindoles; estrogen receptor binding affinity; molecular connectivity.

1. Introduction

Recently von Angerer *et al* (1984) synthesised a series of 2-phenylindoles and screened them for their binding affinities for the calf uterine estrogen receptor by a competitive binding assay with 17β-[^{3}H] estradiol. The relative binding affinities (RBA) were given as 100 × the ratio of the molar concentration of 17β-estradiol and indole required to decrease the receptor bound radioactivity by 50%. Their results indicate a very sensitive dependence of binding affinity on structural modifications of these indoles. The maximum in binding affinity was qualitatively rationalized by two contrary effects: the favourable increase in lipophilicity in the centre of the molecule and the steric hindrance by larger groups. Also, the hydroxy groups in two aromatic rings of these compounds, at a certain distance were prerequisites for a strong binding interaction with the receptor site.

These results stimulated me to rationalize the process of structural modifications. In the simplest way this may be achieved by either analysing the relationship of the RBA with well known physicochemical parameters like π, σ, MR, ES, etc. or with molecular parameters like molecular connectivity χ and van der Waals volume V_w. Because of the lack of data on physicochemical parameters at present for these compounds, it is proposed to find the correlation of the RBA with molecular parameters. This shows the applicability of only one molecular parameter which can be well correlated with the RBA. This parameter is the first-order valence molecular connectivity ($^1\chi^v$). Though an attempt was also made to establish the relationship between V_w and RBA, the correlation equations obtained were not statistically sound.

The connectivity index, χ, has several versions (Kier and Hall 1976). The simplest as well as the extended versions in χ ($^n\chi$) are all calculated from a hydrogen-suppressed graph of the molecule. However, with the use of the connectivity index of various orders ($^n\chi$, $n = 0, 1, 2$----), it becomes difficult to interpret the relationship of the activity with the structure. A single term of the lowest possible order, if correlated with the activity, can be expected to describe the structural influence on the activity most vividly,

and of all the χ's the best and most meaningful is the $^1\chi^v$, as it is an index of the lowest order that takes into account the branching in the molecules and is calculated with the consideration of the valence electrons of the atoms as well. In a number of cases (Kier and Hall 1976, 1977; Bindal *et al* 1980; Gupta *et al* 1982, 1983) this parameter was found to be significantly correlated to biological activities. Further this molecular parameter has also been shown (Kier *et al* 1976) to be correlated separately with hydrophobicity, log P and molar refraction, R_m, accounting for steric bulk, for different series of compounds.

2. First-order valence molecular connectivity

The first-order valence molecular connectivity index, $^1\chi^v$, as described by Kier and Hall (1976) encodes information about size, branching, cyclization, unsaturation and heteroatom content. It is defined as

$$^1\chi^v = \sum (\delta_i^v \delta_j^v)^{-1/2} \tag{1}$$

where δ_i^v and δ_j^v are atom connectivity terms indicating the number of non-hydrogenic atoms adjacent or connected to atoms i and j which are formly bonded, and the summation extends to all connections or edges in the hydrogen suppressed graph. δ^v for any atom is defined as

$$\delta^v = Z - N_H \tag{2}$$

where Z is the number of valence electrons of the atom and N_H is the number of hydrogen atoms attached to it. Thus $^1\chi^v$ is calculated for the total structure of a given compound supressing all the hydrogens. The calculated $^1\chi^v$ values, the indicator variables I_1 and I_2 and the log of relative estrogen receptor binding affinities, log RBA, for 2-(4-hydroxyphenyl) indoles and 2-(3-hydroxyphenyl) indoles are given in tables 1 and 2, respectively. I_1 is used to distinguish the position of the hydroxyl function present in the aromatic ring of the indole part. $I_1 = 1$ for X = 6-OH and $I_1 = 0$ for X = 5-OH. Similarly I_2 is used to differentiate the compounds of table 1 from the compounds of table 2.

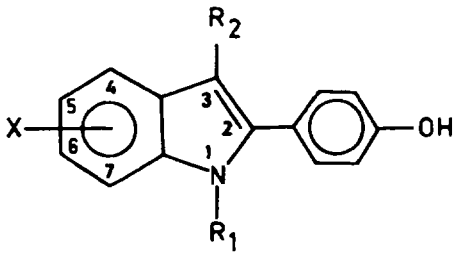
3. Results and discussions

With the use of data as given in table 1, the most significant correlation equation, that is surfaced by regression analysis, is (3). This equation correlates log RBA with $^1\chi^v$ in its first and second powers and I_1.

$$\log \text{RBA} = 12{\cdot}585\ ^1\chi^v + 0{\cdot}880(^1\chi^v)^2 + 0{\cdot}127\ I_1 - 43{\cdot}825,$$
$$n = 26,\ r = 0{\cdot}893,\ s = 0{\cdot}481,\ F(3,22) = 28{\cdot}853, \tag{3}$$

where n is the number of compounds considered for analysis, r is the correlation coefficient, s is the standard deviation and F is the F ratio between the variances of calculated and observed activities. For the present study however, compound 3 (table 1) is excluded because it does not fit into the behaviour of the remaining compounds. The reason for its low binding affinity can be attributed to the fact that the ethyl group at

Table 1. $^1\chi^v$ and relative estrogen binding affinity data for 2-(4-hydroxyphenyl) indoles.

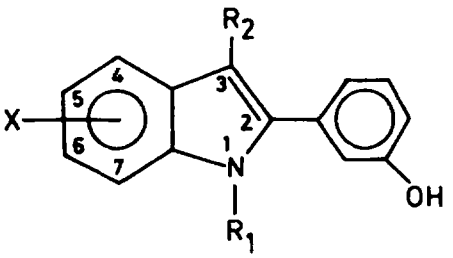


Compound number	R_1	R_2	$X=OH$ position	$^1\chi^v$	I_1	I_2	log RBA Obsd.[a]	Calc. (4)	Calc. (7)
1	H	H	6	5·334	1	1	−2·00	−1·68	−1·67
2	H	CH_3	6	5·756	1	1	−1·22	0·47	0·47
3	H	C_2H_5	6	6·317	1	1	−0·89[b]	—	—
4	CH_3	H	6	5·728	1	1	0·58	−0·54	−0·54
5	C_2H_5	H	6	6·304	1	1	1·20	0·62	0·62
6	C_3H_7	H	6	6·804	1	1	0·93	1·15	1·15
7	C_4H_9	H	6	7·304	1	1	0·63	1·23	1·22
8	CH_3	CH_3	6	6·150	1	1	1·00	0·37	0·37
9	C_2H_5	CH_3	6	6·726	1	1	1·52	1·10	1·10
10	C_3H_7	CH_3	6	7·226	1	1	1·11	1·24	1·24
11	i-C_3H_7	CH_3	6	7·116	1	1	1·11	1·25	1·25
12	CH_3	C_2H_5	6	6·711	1	1	0·77	1·08	1·08
13	C_2H_5	C_2H_5	6	7·287	1	1	1·32	1·23	1·23
14	C_3H_7	C_2H_5	6	7·797	1	1	1·28	0·87	0·87
15	H	H	5	5·334	0	1	−2·00	−1·68	−1·67
16	H	CH_3	5	5·756	0	1	−1·22	−0·47	−0·47
17	CH_3	H	5	5·728	0	1	−0·10	−0·54	−0·54
18	C_2H_5	H	5	6·304	0	1	0·76	0·62	0·62
19	C_3H_7	H	5	6·804	0	1	1·26	1·15	1·15
20	CH_3	CH_3	5	6·150	0	1	0·66	0·37	0·37
21	C_2H_5	CH_3	5	6·726	0	1	0·98	1·10	1·10
22	C_3H_7	CH_3	5	7·226	0	1	1·20	1·24	1·24
23	i-C_3H_7	CH_3	5	7·116	0	1	0·54	1·25	1·25
24	C_4H_9	CH_3	5	7·726	0	1	0·66	0·94	0·94
25	C_5H_{11}	CH_3	5	8·226	0	1	0·36	0·19	0·19
26	C_2H_5	C_2H_5	5	7·287	0	1	1·36	1·23	1·23
27	C_3H_7	C_3H_7	5	8·287	0	1	0·23	0·06	0·06

[a] Relative binding affinity for the calf uterine estrogen receptor, taken from von Angerer *et al* (1984);
[b] compound is excluded in the derivation of (4) and (7).

position 3 of the indole offers steric bulk and the hydrogen at position 1 of the indole probably decreases the hydrophobicity, as suggested by Fritsche (1964), as it forms hydrogen bridges with water molecules and prevents the hydrophobic interaction with the receptor site. The increase in steric hindrance and decrease in hydrophobicity, therefore, tends to decrease the binding action of this compound. Though (3) is quite

Table 2. $^1\chi^v$ and relative binding affinity data for 2-(3-hydroxyphenyl) indoles.



Compound number	R_1	R_2	X=OH position	$^1\chi^v$	I_1	I_2	log RBA Obsd.[a]	log RBA Calc. (6)	log RBA Calc. (7)
1	C_2H_5	H	6	6·304	1	0	0·23	0·05	0·06
2	CH_3	CH_3	6	6·150	1	0	−0·26	0·07	−0·18
3	C_2H_5	CH_3	6	6·726	1	0	0·48	0·37	0·54
4	C_3H_7	CH_3	6	7·226	1	0	0·54	0·75	0·68
5	C_2H_5	H	5	6·304	0	0	0·23	0·05	0·07
6	CH_3	CH_3	5	6·150	0	0	−0·22	−0·07	−0·18
7	C_2H_5	CH_3	5	6·726	0	0	0·34	0·37	0·54
8	C_3H_7	CH_3	5	7·226	0	0	0·87	0·75	0·68

[a] See footnotes under table 1.

sound statistically, the coefficient associated with I_1 indicates that its contribution to the activity is very small. This implies that the binding action of the molecule will not be much effected by the presence of hydroxyl group either at position 5 or position 6 of indole. Therefore, deleting the variable I_1 an equally statistically significant parabolic equation (4) is obtained.

$$\log \text{RBA} = 12{\cdot}835\ {}^1\chi^v + 0{\cdot}899({}^1\chi^v)^2 - 44{\cdot}563,$$

$$n = 26,\ r = 0{\cdot}891,\ s = 0{\cdot}475,\ F(2, 23) = 44{\cdot}168. \quad (4)$$

The F value of the above equation is significant at 99% level [$F^2_{23}(0{\cdot}01) = 5{\cdot}66$] and accounts for 79% of variance ($r^2 = 0{\cdot}794$). The calculated values of log RBA listed in table 1 are in close agreement with observed values.

Similarly using the data of table 2 for 2-(3-hydroxyphenyl) indoles the correlation equation (5) is obtained.

$$\log \text{RBA} = 9{\cdot}439\ {}^1\chi^v + 0{\cdot}648({}^1\chi^v)^2 - 0{\cdot}057\ I_1 - 33{\cdot}638,$$

$$n = 8,\ r = 0{\cdot}926,\ s = 0{\cdot}189,\ F(3,4) = 8{\cdot}08. \quad (5)$$

In this equation the coefficients of the $({}^1\chi^v)^2$ and I_1 variables show that activity is not going to be much effected if both of these are deleted. Thus a linear equation (6) is obtained.

$$\log \text{RBA} = 0{\cdot}761\ {}^1\chi^v - 4{\cdot}749,$$

$$n = 8, r = 0{\cdot}897,\ s = 0{\cdot}181,\ F(1, 6) = 24{\cdot}642. \quad (6)$$

The F value of (6) is significant at 99% level [$F_6^1(0{\cdot}01) = 13{\cdot}74$] and the calculated values of log RBA are tabulated in table 2. From the above discussion and from (4) and (6), it is quite evident that I_1 makes no significant contribution to the activity. Therefore, in 2-(4-hydroxyphenyl) or 2-(3-hydroxyphenyl) indoles the position of the hydroxyl group, either at 6 or 7 of the indole moiety will hardly matter as far as the binding action is concerned. Further (4) is parabolic and (6) is linear, therefore, it is considered appropriate to consider 34 compounds all together by taking another indicator variable I_2, which will differentiate the two sets of congeners. The second set of 8 compounds having linear variation of binding affinity with $^1\chi^v$ will fall on the ascending part of the parabolic variation of the first set of 24 compounds, as the sign of the coefficient of $^1\chi^v$ in both equations is the same. I_2 is assigned a value 1 for each of the compound belonging to a 2-(4-hydroxyphenyl) indoles and a value 0 for each of the remaining 2-(3-hydroxyphenyl) indoles. Therefore, I_2 will differentiate the positions of the hydroxyl groups either 4 or 3, in phenyl ring. Considering $^1\chi^v$, $(^1\chi^v)^2$ and I_2 as independent variables and activity as a dependent variable for all 34 compounds from table 1 and table 2 the derived correlation is (7).

$$\log \text{RBA} = 12{\cdot}801\,{}^1\chi^v + 0{\cdot}897\,(^1\chi^v)^2 + 0{\cdot}556\ I_2 - 44{\cdot}996,$$
$$n = 34,\ r = 0{\cdot}893,\ s = 0{\cdot}422,\ F(3, 30) = 39{\cdot}343, \qquad (7)$$

where the F value is again significant at the 99% level [$F_{30}^3(0{\cdot}01) = 4{\cdot}51$] and accounts for 79% of variance. This highly significant equation is therefore used to calculate the activity values for all the compounds of tables 1 and 2. The calculated values of activity, listed in these tables, are in close agreement with observed values. The positive coefficient of I_2 shows that there will be a significant increase in binding affinity of a compound, if it has a hydroxyl group in the para position of the phenyl ring. Similarly, as evident from the coefficients of $^1\chi^v$ and $(^1\chi^v)^2$, the binding action of a compound increases as its molecular connectivity increases. But this does not mean that affinity can take any value as $^1\chi^v$ goes on increasing. After a certain optimum value of $^1\chi^v$, the affinity again starts decreasing, since there exists a parabolic variation between connectivity and binding affinity.

Thus $^1\chi^v$ plays an important role in the binding action of 2-phenylindole with estrogen receptor proteins. The correlation equations mentioned above describe the direct structural influence on binding affinity and are very useful in substituent selection while designing more potent compounds.

References

Bindal M C, Singh P and Gupta S P 1980 *Arzneim.-Forsch.* **30** 234
Fritsche H 1964 *Ber. Bunsenges. Phys. Chem.* **68** 459
Gupta S P, Bindal M C and Singh P 1982 *Arzneim.-Forsch.* **32** 1223
Gupta S P, Prabhakar Y S and Handa A 1983 *Res. Commun. Chem. Pathol. Pharmacol.* **42** 455
Kier L B and Hall L H 1976 *Molecular connectivity in chemistry and drug research*, (New York: Academic Press)
Kier L B and Hall L H 1977 *J. Med. Chem.* **20** 1631
Von Angerer E, Prekajac J and Strohmeier J 1984 *J. Med. Chem.* **27** 1439