QSAR MODELING OF PHENOL DERIVATIVES USING 3D MoRSE DESCRIPTORS AND EIGEN VALUES

Arun K Sikarwar[1] and Sameer Dixit*[2]

1. Department of Chemistry, Govt. Home Science P. G. College Hoshangabad, Madhya Paradesh (INDIA)
2. Department of Chemistry, M. J. P. Govt. Polytechnic College Khandwa, Madhya Paradesh (INDIA)

*Email : arunsikarwar@yahoo.co.in
: dixitsameer1@rediffmail.com

Received November 25, 2011 *Accepted March 5, 2012*

ABSTRACT

The eigen value for distance metrics (Ev) and 3D MoRSE Descriptors have been used for proposing a series of QSAR models, which can be used for estimating physico-chemical properties of phenol derivatives. This paper examines the correlation of molecular modeling technique parameters and the physico-chemical properties of phenol derivatives. The physical properties of phenol derivatives used are Partition Coefficient (LogP), Molar Refractivity (M_R), Polarizability (α). The results obtained that 3D MoRSE Descriptors is better in modeling the properties of phenol derivatives.

Key Words : QSAR, 3D MoRSE Descriptors, Eigen Values, LogP, Molar Refractivity (M_R) Polarizability, Topological indices

INTRODUCTION

QSARs (Quantitative Structure-Property Relationships) deals with relationship between physico-chemical properties & chemical structure. QSARs are based on the assumption that the structure of a molecule (i.e. its geometric, steric and electronic properties) must contain the features responsible for its physical, chemical, and biological properties and on the ability to represent the chemical by one or more numerical descriptor(s). By QSAR models, the biological activity (or property, reactivity, etc.) of a new or untested chemical can be inferred from the molecular structure of similar compounds whose activities (properties, reactivities, etc.) have already been assessed. The QSPR (Quantitative Structure-Property relationship) acronymous is used when a property is modeled. A Physico-chemical property of a compound is directly depending on chemical structure. The quest to understand the relationship between chemical structures and Physico-chemical property has become a subject of enumerable studies.[1] Quantitative structure-activity relationship (QSAR) techniques increase the probability of success and reduce time and cost involvement in drug discovery process[2]. In this paper a theoretical technique has been discussed by which the properties of hypothetical molecules can be measure prior to their synthesis. This technique shall reduce the drug discovery coast, time and efforts.

*Author for correspondence

MATERIAL AND METHODS

Molecular Modelling Methods

Topological indices -Topological descriptors are derived from hydrogen-suppressed molecular graphs, in which the atoms are represented by vertices and the bonds by edges. The connections between the atoms can be described by various types of topological matrices (e.g., distance or adjacency matrices), which can be mathematically

manipulated so as to derive a single number, usually known as graph invariant, graph-theoretical index or topological index. As a result, the topological index can be defined as two-dimensional descriptors that can be easily calculated from the molecular graphs and do not depend on the way the graph is depicted or labeled and no need of energy minimization of the chemical structure.

3D MoRSE Descriptors

3D MoRSE descriptors (3D Molecule Representation of Structures based on Electron diffraction) are derived from Infrared spectra simulation using a generalized scattering function. A typical MoRSE descriptor is denoted by Mor sw where s and w take the values1 and w where, u is unweighted

m is weighted by mass
v is weighted by van der Waals volume
e is weighted by electronegativity
p is weighted by polarizability

The MoRSE descriptor is defined as follows:

$$Mor(s,w) = I(s,w) = \sum_{i-2}^{n}\sum_{j-1}^{i-1} w_i w_j \sin\frac{(sr_{ij})}{(sr_{ij})}, \quad ...(1)$$

where,

r_{ij} is the Euclidean distance between the atoms i and j, and

w_i and w_j are the weights of the atoms i and j respectively.

Absolute eigen value sum from electronegativity weighted distanced matrix (AEige) or (Ev). For a I D-λ I I = 0 is known as characteristic equation and roots of this equation are called the proper value or Eigen value and it denoted by . Principal eigen value is the largest positive eigen value of the distance matrix.

I D-λ I I = 0 ……..(2)

Physico-chemical properties

The physico-chemical properties, namely the Partition Coefficient (LogP), Molar Refractivity (M_R), Polarizability (α) for the set of 34 phenol & derivatives, were adopted from the literature and calculated by softwares.

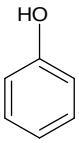


Fig. 1 : Phenol

Regression analysis

The Regression analysis of the data **(Table 1** to **Table 3)** was carried out using Ms-Excel program. Physico-chemical properties (LogP, MR,) of the and its nuclear substituted derivatives used in the present investigation are given in **Table 1**.

RESULT AND DISCUSSION

As started earlier we have used the set of 34 phenol derivatives as **(Table 1)** as they are biological active compounds and also they exhibit interesting physico-chemical properties. The general structure of the compound is shown in **Fig. 1**. The values of the topological indices (3D Morse Descriptors, Ev) and physico-chemical properties (LogP, M_R,) are given in **Table 1**.

It is worthy mentioning that numerous examples illustrating the influence of structure on properties of compounds are available[3-6]. It is apparent that each property bears a relationship to molecular structure, although the nature of relationship is not always known. It is, however, known that the relation between structure and property reveals two general trends: (i) properties, which may as a sum of the corresponding values of the constituents parts called additive. And (ii) properties that depend heavily on details of the arrangement of the constituent atoms called constitutive. So we conclude that topological indices are convenient tools to formulate direct relationship between molecular structure and physico-chemical properties. Recall that we have studies QSPR modeling of three physico-chemical properties of 34 phenol derivatives which are already mention above. The first step in such a study is to investigate co-linearity between topological indices and their correlations with physico-chemical properties used. This can be achieved by obtaining a correlation matrix. Such a correlation matrix obtained in the present case is shown in **Table 3**.

The data presented in **Table 3** indicates that the topological indices Ev and MoRSE D and three properties (LogP, M_R, and α) are mutually highly correlated. This indicates that Ev and MoRSE D have similar correlating properties.

For final conclusion we have to correlate these three indices independently with the properties

Table 1 : Phenol derivatives, their eigen values (Ev) and 3D MoRSE Descriptors physico-chemical properties

S/N	Compound	LogP	$M_R(cm^3)$	α (10-24cm^3)	MoRSE D	Ev
1	4-OC_2H_5	2.049	39.44	15.63	117.868	39.414
2	4-OC_3H_7	2.419	44.07	17.47	251.91	51.067
3	4-OC_4H_9	2.875	48.71	19.31	320.877	64.526
4	4-OC_6H_{13}	3.787	57.97	22.98	483.757	96.437
5	H	1.563	28.13	11.15	77.674	17.559
6	4-NO_2	1.457	34.67	13.74	117.093	34.356
7	4-CHO	1.322	34.88	13.83	109.205	29.881
8	4-F	1.768	28.12	11.15	83.845	22.474
9	4-OH	1.295	30.01	11.89	94.9	22.788
10	4-CH_3	2.049	32.95	13.06	117.868	23.978
11	4-C_2H_5	2.505	37.68	14.93	166.377	32.803
12	4-CN	1.442	32.84	13.02	93.496	29.576
13	4-OC_6H_5	3.123	54.57	21.63	277.548	70.092
14	4-$C(CH_3)_3$	2.963	46.52	18.44	288.342	47.044
15	3-NO_2	1.457	34.67	13.74	117.093	33.636
16	3-$C(CH_3)_3$	2.963	46.52	18.44	288.342	46.171
17	3-CH_3	2.049	32.95	13.06	117.868	23.649
18	3-OCH_3	1.546	34.81	13.8	138.921	29.262
19	3-C_2H_5	2.505	37.68	14.93	166.377	32.244
20	3-CN	1.442	32.84	13.02	93.496	29.03
21	3-F	1.768	28.12	11.15	83.845	22.158
22	3-OH	1.295	30.01	11.89	94.9	22.469
23	2-CH_3	2.049	32.95	13.06	117.868	23.064
24	2-F	1.768	28.12	11.15	83.845	21.603
25	2-OCH_3	1.546	34.81	13.8	138.921	28.289
26	2-C_2H_5	2.505	37.68	14.93	166.377	31.246
27	2-OH	1.295	30.01	11.89	94.9	21.907
28	2-OH,4CH_3	1.781	34.84	13.81	138.921	28.604
29	2-CN	1.442	32.84	13.02	93.496	28.061
30	2-NO_2	1.457	34.67	13.74	117.093	32.362
31	2-$C(CH_3)_3$	2.963	46.52	18.44	288.342	44.613
32	4-C_3H_7	2.961	42.31	16.77	251.91	43.574
33	4-C_4H_9	3.417	46.94	18.61	288.342	56.09
34	4-C_5H_{11}	3.874	51.58	20.44	361.796	70.252

* LogP- Partition Coefficient , M_R- Molar Refractivity; α - Polarizability

Table 2 : Estimated and observed LogP, M_R, α values of Phenol derivatives (Values are estimated from 3D MoRSE Descriptors and Ev*)

		Estimated from 3D Morse Descriptor			Estimated from Eigen Values		
S/N	**Compound**	**LogP(cMD)**	**M_R(cMD)**	**α(cMD)**	**LogP(cEv)**	**M_R(cEv)**	**α(cEv)**
1	4-OC_2H_5	1.754876	33.53684	13.29017	2.235511	38.84686	15.39974
2	4-OC_3H_7	2.69317	43.85807	17.37846	2.660846	43.9602	17.42736
3	4-OC_4H_9	3.175939	49.16853	19.48195	3.152099	49.86601	19.76922
4	4-OC_6H_{13}	4.316099	61.71029	24.44979	4.316851	63.86856	25.32174
5	H	1.473518	30.4419	12.06426	1.437804	29.25689	11.59697
6	4-NO_2	1.749451	33.47716	13.26654	2.050894	36.62741	14.51964
7	4-CHO	1.694235	32.86979	13.02595	1.887557	34.66378	13.74099
8	4-F	1.516715	30.91707	12.25247	1.617201	31.41359	12.45218
9	4-OH	1.5941	31.7683	12.58965	1.628662	31.55137	12.50681
10	4-CH_3	1.754876	33.53684	13.29017	1.672097	32.07355	12.71387
11	4-C_2H_5	2.094439	37.27203	14.7697	1.99421	35.94596	14.24942
12	4-CN	1.584272	31.66019	12.54683	1.876424	34.52995	13.68792
13	4-OC_6H_5	2.872636	45.8322	18.16041	3.355258	52.30837	20.73771
14	4-$C(CH_3)_3$	2.948194	46.66333	18.48963	2.514006	42.19491	16.72736
15	3-NO_2	1.749451	33.47716	13.26654	2.024614	36.31148	14.39436
16	3-$C(CH_3)_3$	2.948194	46.66333	18.48963	2.482142	41.81183	16.57545
17	3-CH_3	1.754876	33.53684	13.29017	1.660089	31.92918	12.65663
18	3-OCH_3	1.902247	35.15792	13.93229	1.864963	34.39217	13.63329
19	3-C_2H_5	2.094439	37.27203	14.7697	1.973806	35.70067	14.15216
20	3-CN	1.584272	31.66019	12.54683	1.856495	34.29036	13.59292
21	3-F	1.516715	30.91707	12.25247	1.605667	31.27493	12.39719
22	3-OH	1.5941	31.7683	12.58965	1.617019	31.4114	12.45131
23	2-CH_3	1.754876	33.53684	13.29017	1.638736	31.67248	12.55484
24	2-F	1.516715	30.91707	12.25247	1.58541	31.0314	12.30062
25	2-OCH_3	1.902247	35.15792	13.93229	1.829449	33.96521	13.46399
26	2-C_2H_5	2.094439	37.27203	14.7697	1.937379	35.26274	13.9785
27	2-OH	1.5941	31.7683	12.58965	1.596506	31.16479	12.35352
28	2-OH,4CH_3	1.902247	35.15792	13.93229	1.840946	34.10344	13.5188
29	2-CN	1.584272	31.66019	12.54683	1.821127	33.86517	13.42431
30	2-NO_2	1.749451	33.47716	13.26654	1.978113	35.75245	14.17269
31	2-$C(CH_3)_3$	2.948194	46.66333	18.48963	2.425275	41.12818	16.30436
32	4-C_3H_7	2.69317	43.85807	17.37846	2.387351	40.67227	16.12358
33	4-C_4H_9	2.948194	46.66333	18.48963	2.844185	46.16429	18.30136
34	4-C_5H_{11}	3.462372	52.31929	20.72998	3.361098	52.37858	20.76555

*LogP(cMD)- LogP value calculated from the 3D MoRSE Descriptors ; M_R(cMD)- M_R value calculated from the 3D MoRSE Descriptors; (cMD)- value calculated from the 3D MoRSE Descriptors; LogP(cEv)- LogP value calculated from the Absolute eigen value; M_R(cEv)-Ev value calculated from the Absolute eigen value ; α (cMD)-α value calculated from the Absolute eigen value

Table 3 : Correlation matrix for the inter correlation of Morse D and Ev as well as for their correlation with physico-chemical properties of phenol derivatives (LogP, MR, and α)

	MoRSE D	Ev	LogP	M_R (cm_3)	α
MoRSE D	1				
Ev	0.941652	1			
LogP	0.918824	0.835556	1		
M_R	0.956704	0.956186	0.891954	1	
α	0.956736	0.95625	0.891913	0.999999	1

used. The data are then subjected to the regression analysis using the method of least squares. The 3D MoRSE Descriptors (MoRSE D or MD) & eigen values (Ev) correlate with Partition Coefficient (Log P) as follows:

$$LogP = 0.007\ MD + 0.9298 \quad \text{-----(3)}$$

n =34, sd= 0.306726, r^2 = 0.8442, r = 0.9188

$$LogP = 0.0365\ Ev + 0.7969 \quad \text{-----(4)}$$

n =34, sd= 0.426985, r^2 = 0.6982, r = 0.8355

By above results indicate that the potential of the 3D MoRSE Descriptors in modeling LogP is better than Ev.

In the case of modeling the Molar Refractivity (M_R) the following statistically significant correlation are obtained

$$M_R = 0.077\ MD + 24.461 \quad \text{-----(5)}$$

n =34, sd= 2.378109, r^2 = 0.9153, r = 0.956704

$$PR = 0.4388\ Ev + 21.552 \quad \text{-----(6)}$$

n =34, sd= 2.391985, r^2 = 0.9143, r=0.956186

The result indicate that 3D MoRSE Descriptors and eigen values have similar potential in modeling Molar Refractivity (M_R) but Morse gives slightly better result than Ev..

Similar results were also obtained for modeling Polarizability (α). The corresponding regression expressions are found as follows:

$$= 0.0305\ MD + 9.6952 \quad \text{----- (7)}$$

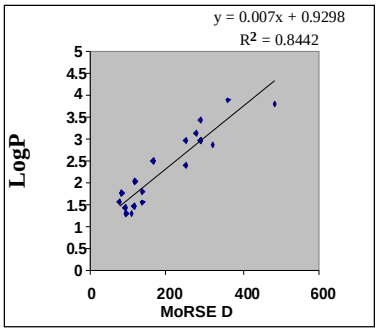
n =34, sd= 0.942312, r^2 = 0.9153, r = 0.956736

$$= 0.174\ Ev + 8.5417 \quad \text{----- (8)}$$

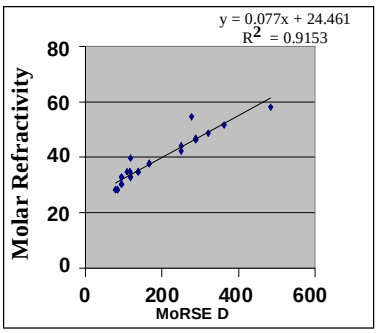
n =34, sd=0.947469, r^2 = 0.9144, r=0.95625

From the above regression expressions we conclude that using 3D MoRSE and Ev for three physico-chemical properties can be estimated with a considerable precision. Such calculated (estimated) values are presented in **Table 2**. The comparison of the results shows that the estimated properties are much closed to the observed values. Further, the difference between the observed and estimated properties is consistent with the **QSPR models** proposed above.

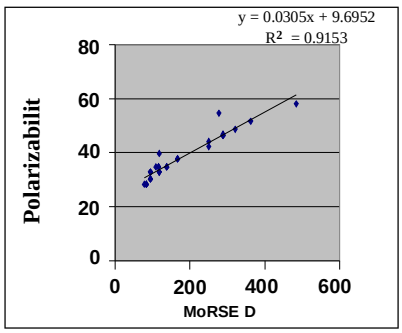
In order to obtain further evidence in the favour of our results we have calculated quantity factor Q for each of the QSPR models obtained as above. Note that the quality factor[7] Q is define as the ratio of correlation coefficient (r) to the standard error of estimation (sd), are the higher will be the Q values, and the better will be the proposed QSPR models.



LogP Vs MoRSE D



MoRSE D Vs Molar Refractivity



MoRSE D Vs Polarizability

QSPR Models

The Q values calculated for the QSPR models represented by Equation (3) to (8) are found as: 2.9955, 1.9567, 0.402296, 0.39987, 1.01528 and 1.00934 respectively. These Q values indicate that the correlation coefficient (r) is not solely responsible in determining the quantity of correlation and the standard error of estimation is also a important factor in determining the quantity of QSPR models.

It is worthy mentioning that the studies compounds are phenols derivatives and they are differing in the number and the position of various substitutions in the aromatic skeleton of the phenol. Therefore, change in the property as well as changes in the topological indices (3D Morse and Ev) may be attributed to the change in the substitution pattern.

It is also worth recording that in all the 34 phenol-derivatives only one cycle is present, i.e. they all are monocyclic graphs of ring size carrying zero, one, two, …… etc. simple tree like attachment of the possible situation pattern. The composition of the graph sample is thus strange. The graphs (phenol-derivatives) chosen here, however, all consist of one and the same cycle. They differ in acyclic part (substituents) only. Therefore any variance of 3D Morse and Ev is caused by the acyclic parts, for which the correlation between 3D Morse and Ev is trivial.

CONCLUSION

From the above result and discussion we conclude that the structure-property models based on the 3D MoRSE Descriptors are better in modeling the properties (Log P, M_R, α) of phenol derivatives used. In modeling of LogP 3D Morse Descriptors is much better than Ev. While in modeling of Molar Refractivity and Polarizability (MR, α) of phenol derivatives 3D Morse Descriptors is slightly better than Eigen Value (Ev) of Phenols derivatives.

ACKNOWLEDGEMENT

Authors are very thankful to Mr A. P. Sakalle Principal M. J. P. Govt. Polytechnic College Khandwa Madhya Pradesh, India .for providing facilities and motivation in the work.

REFERENCES

1. Trinajstic N., Chemical Graph Theory. **2** (1), 2-10, **(1992)**.
2. F. Choplin, Computers and the medicinal chemist, Comp. Medi. Chemi., **4** (1), 33-58, **(1990)**
3. Wiener H., Structural determination of paraffin boiling points, J. Am. Chem. SOC., **69** (1), 17-20. **(1947)**
4. Kier. L. B.and Hall L. H., Molecular Connectivity in Chemistry and Drug Research. Academic Press, New Yark, **1** (1), 46-61, **(1976)**.
5. SabljiC A. and Trinajsti C.N., Quantitative Structure-Activity Relationships: The Role of Topological Indices. Acta Pharm. Jugosl. **31** (1), 189-214 **(1981)**
6. Lukovits I., Wiener Indices and Partition Coefficients of Unsaturated Hydrocarbons Quant. Stuct. Act. Relat **9** (3), 227-231, **(1990)**
7 POGLIANI L.: Structure Property Relationships of Amino Acids and Some Dipeptides, Amino Acids, **6** (1), 141-153, **(1994)**