Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 18 (1) 75-86: Spring 2021
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Quantitative Structure- Property Relationship (QSPR) Study of 2-Phenylindole derivatives as Anticancer Drugs Using Molecular Descriptors

Samira Bahrami, Fatemeh Shafiei*, Azam Marjani, Tahereh Momeni Esfahani

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

Received Februrary2021; Accepted April 2021

ABSTRACT

A QSPR study on a series of 2-Phenylindole derivatives as anticancer agents was performed to explore the important molecular descriptor which is responsible for their thermodynamic properties such as heat capacity (Cv) and entropy(S). Molecular descriptors were calculated using DRAGON software and the Genetic Algorithm (GA) and backward selection procedure were used to reduce and select the suitable descriptors. Multiple Linear Regression (MLR) analysis was carried out to derive QSPR models, which were further evaluated for statistical significance such as squared correlation coefficient (R^2) root mean square error (RMSE), adjusted correlation coefficient (R^2adj) and fisher index of quality (F). The multicollinearity of the descriptors selected in the models were tested by calculating the variance inflation factor (VIF), Pearson correlation coefficient (PCC) and the Durbin–Watson (DW) statistics. The predictive powers of the MLR models were discussed using Leave-One-Out Cross-Validation (LOO_{CV}) and test set validation methods. The best QSPR models for prediction the Cv(J/molK) and S(J/molK), having squared correlation coefficient R^2 =0.907 and 0.901, root mean squared error RMSE=2.019 and RMSE= 2.505, and cross-validated squared correlation coefficient R^2 cv = 0.902 and 0.889, respectively. The statistical outcomes derived from the present study demonstrate good predictability and may be useful in the design of new 2-Phenylindole derivatives.

Keywords: 2-Phenylindole derivatives; structure -property relationship; heat capacity; entropy; genetic algorithm -multiple linear regressions (GA-MLR)

1. INTRODUCTION

structure-activity/ property relationships (QSARs/QSPRs) are mathematical models that relate molecular structure to their biomedicinal, biological activity, toxic potential, physicochemical, and thermodynamic properties. Since experimental properties or activities for the majority of chemicals are not known, QSARs/QSPRs based on calculated descriptors are becoming more popular [1-5].

The molecular descriptors or

*Corresponding author: f-shafiei@iau-arak.ac.ir

fingerprints are the final results of logic and mathematical procedures which transforms chemical / physical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment [6].

Dragon molecular descriptors include different groups: 2D autocorrelations, constitutional descriptors, topological indices, molecular walk counts, GETAWAY descriptors, geometrical descriptors, 3D-MoRSE descriptors, weighted holistic invariant molecular (WHIM) descriptors, empirical descriptors, functional groups, atom-centered fragments, empirical descriptors, and properties. These groups of descriptors have been widely used in the different fields such as pharmaceutical sciences, Modern Computational Drug Design, environmental protection policy, chemoinformatics, health researches [7-10].

2-Phenylindole derivatives have attractive anticancer activities in case of melanoma and lung cancer along with the breast cancer and may be a future hope for better anticancer drugs of low toxicity and high potency [11].

A multiple validated QSAR model has been applied to predict cytotoxic activity of phenylindole derivatives using the molecular fingerprints. This study suggests that the hydrophobic activity of linear alkyl group has a positive effect on cytotoxicity of 102 phenylindole derivatives [12].

Correlation weights and index of ideality of correlation (IIC) have been used for Predicting cytotoxicity of 2-phenylindole derivatives against breast cancer cells [13]

Three-dimensional quantitative structure–activity relationship (3DQSAR) model has been established to study anticancer activity of 2-phenylindole derivatives using the comparative molecular field analysis (CoMFA) method [14].

A QSAR study has been performed to estimate molecular orbital energies (HOMO and LUMO), hydrophobicity (logP), molar volume (V), molecular polarizability (MP), surface area grid (SAG) and molecular mass (MASS) of some 2-phenylindole-3- carbaldehyde derivatives by electrotopological state atom (ETSA) and refracto topological state atom (RTSA) indices[15].

Two general classes of molecular descriptors, namely, atom pairs (APs) and topological indices (TIs) have been used to develop QSARs for anticancer activity of a set of 93 derivatives of 2-phenylindole [16].

Two of the basic thermodynamic properties are heat capacity and entropy. Heat capacity is the measurable physical quantity that characterizes the amount of heat required to increase the temperature by one unit. A knowledge of heat capacity allows computing of other properties such as entropy and enthalpy of a chemical substance [17,18].

Heat capacity can be measured experimentally, but most of methods such as, the ac calorimetry [19] and modulated-bath calorimetry [20] are often expensive and require specialized skill sets and also these are time consuming and costly [21,22].

The properties of chemical species can be estimated theoretically by quantum chemistry calculations and molecular structure descriptors that they are computationally cheaper than experimental methods. Some studies have been conducted on the relationship between the chemical structure and activities of 2 phenylindole derivatives but we did not find a literature that studied the relationship between the chemical structure and physical properties of these derivatives.

Some studies have been conducted on the relationship between the chemical structure and activities of 2 phenylindole derivatives but we did not find any article that studied the relationship between the chemical structure and thermodynamic properties of these derivatives. In the present study, we have carried out a QSPR analysis to derive a quantitative relationship between chemical structural of 43 anticancer agents and their thermodynamic properties.

2.MATERIALS AND METHODS

The compounds discussed in this study comprise 43 derivatives of 2-Phenylindole. These data sets were randomly divided into 2 groups: training and test sets consisting of 30 and 13 data point, respectively. The template structure of 2-phenylindole derivatives used in the present study is showed in Fig.1. The Gauss View software was used to draw all the derivatives. Geometry optimizations and vibrational frequencies of 2-Phenylindole compounds were carried out using the Hartree–Fock functional [23] and 6-31G basis sets [24] at the Gaussian 09W[25] and the heat capacity (Cv/ J $mol^{-1}K^{-1}$) and entropy (S/ J mol^{-1} K^{-1}) were calculated. The values of these properties are listed in Table 1.

Different types of numerical descriptors were generated to describe each compound. The molecular descriptors for constructing the best model were calculated using the Dragon5.4- 2006 package [26]. In total, 1826 and 1588 theoretical descriptors were generated for Cv and S respectively.

The Genetic algorithms (GA) are written in MATLAB (version 2010a) and backward method using the Statistical Package for the Social Science (SPSS) software was used to reduce the number of molecular descriptors [27-29].

The multiple linear regression models were obtained by the SPSS [30-32] statistics version 22 to determine the relationship between the thermodynamic properties and molecular descriptors of the 2-Phenylindole derivatives.

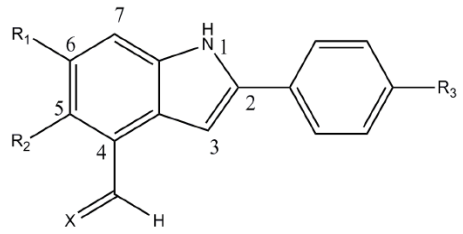

Fig.1. The general structure of 2-phenylindole derivatives used in the present study.

Table 1. The heat capacity and entropy data for substituted 2-phenylindole derivatives discussed in the present study

No.	R_1	R_2	R_3	X	Cv(J/molK)	S(J/molK)
1	H	H	H	$C(CN)_2$	68.2793	141.7525
2	H	H	OCH_3	$C(CN)_2$	87.4661	170.7824
3	H	OCH_3	OCH_3	$C(CN)_2$	88.4451	172.2638
4	OCH_3	H	OCH_3	$C(CN)_2$	90.5298	175.4179
5	H	F	OCH_3	$C(CN)_2$	93.4212	179.7927
6	F	H	OCH_3	$C(CN)_2$	96.2123	184.0157
7	OCH_3	H	CH_3	$C(CN)_2$	86.1098	168.7303
8	H	CH_3	OCH_3	$C(CN)_2$	89.2476	173.4779
9	Cl	CH_3	OCH_3	$C(CN)_2$	102.6057	193.689
10	H	n-Pr	OCH_3	$C(CN)_2$	105.5190	198.0968
11	H	i-Pr	OCH_3	$C(CN)_2$	108.3384	202.3627
12	H	n-Bu	OCH_3	$C(CN)_2$	75.8589	153.2206
13	H	n-Pentyl	OCH_3	$C(CN)_2$	71.9208	147.2622
14	H	n-Hexyl	OCH_3	$C(CN)_2$	71.0050	145.8765

No.	R_1	R_2	R_3	X	Cv(J/molK)	S(J/molK)
15	H	n-Bu	CH_3	$C(CN)_2$	67.7167	140.9013
16	H	n-Bu	CH_2CH_3	$C(CN)_2$	68.4952	142.0791
17	H	n-Bu	CF_3	$C(CN)_2$	69.4873	143.5802
18	H	n-Pentyl	CF_3	$C(CN)_2$	74.0621	150.5020
19	H	n-Hexyl	CF_3	$C(CN)_2$	66.7965	139.5090
20	H	OCH_3	OCH_3	O	73.5061	149.6607
21	OCH_3	H	OCH_3	O	76.0159	153.4582
22	F	H	OCH_3	O	78.0242	156.4968
23	H	F	OCH_3	O	90.2245	174.956
24	Cl	H	OCH_3	O	78.6610	157.4601
25	Cl	CH_3	OCH_3	O	78.7395	157.5789
26	H	CH_3	OCH_3	O	81.3932	161.5941
27	H	Pr	OCH_3	O	81.1185	161.1784
28	H	n-Bu	OCH_3	O	66.6570	139.2978
29	H	sec-Bu	OCH_3	O	62.5313	133.0557
30	H	t-Bu	OCH_3	O	71.6919	146.9158
31	H	n-Pentyl	OCH_3	O	73.2037	148.8596
32	H	n-Hexyl	OCH_3	O	72.8678	148.3345
33	OCH_3	OCH_3	OCH_3	O	85.4458	167.9946
34	OCH_3	H	CH_3	O	82.2398	162.9835
35	H	CH3	CH_3	O	88.0331	172.0387
36	H	n-Bu	CH_3	O	90.5487	175.9708
37	H	n-Bu	CH_2CH_3	O	62.5923	132.2733
38	H	CH_2CH_3	n-Bu	O	55.0414	120.4708
39	H	n-Bu	CF_3	O	80.5129	160.2843
40	H	n-Pentyl	CF_3	O	79.1957	158.2254
41	H	n-Hexyl	CF_3	O	78.0895	156.4963
42	OCH_3	H	H	O	78.4193	157.0118
43	H	H	H	O	85.8289	168.5934

3. RESULT AND DISCUSSION

3.1. QSPR models

The genetic approximation-multiple linear regression (GA-MLR) analysis led to the derivation of 3 models for the heat capacity (Cv), with 5-3 descriptors (Table **2**).

The models were evaluated with regression parameters: correlation coefficient (R), squared regression coefficient (R^2), root mean squared error (RMSE), Fisher ratio (F), Durbin- Watson (DW) and Significance (Sig) [33-36].

The statistical coefficients of the three models are almost similar; so, the model 3, which has the lowest number of descriptors, has been selected. QSPR model and statistical parameters for three molecular descriptors is shown as follows:

$$Cv= 8.737+0.026DD1+1.218G1-0.148 Eiglm \quad (1)$$

N=30 R= 0.998 R^2=0.996 R^2_{adj}=0.996 RMSE=0.911 DW= 1.936 $F=3.334\times10^3$ Sig=0.000

As can be seen in Table 3, the three descriptors are useful to predict the heat capacity (Cv), which are: DD1,G1 and Eiglm. These descriptors are classified in 3D matrix-based descriptors (DD1), Geometrical distance matrix(G1), descriptors 2D matrix-based descriptors(Eiglm).

The suitable linear model for the entropy (S) includes three molecular descriptors which are: DD1, G1 and ITH(see Table3). These descriptors are classified in Geometrical distance matrix (G1), 3D matrix-based descriptors (DD1) and GETAWAY descriptors(ITH).

$$S=58.211+1.267G1+0.028DD1+0.275ITH \quad (2)$$

N=30, R= 0.995, R^2=0.991, R^2_{adj}=0.990, RMSE=1.397, DW= 1.998, $F=1.449\times10^3$,Sig=0.000

Table2- Statistical parameters of the models calculated with the SPSS software for the S of 2-Phenylindole derivatives

Model	Descriptor	R	R^2	R^2_{adj}	RMSE	F	Sig
1	DD1,G1, Eiglm, RDF, ITH	0.998	0.997	0.996	0.901	1.405×10^3	0.000
2	DD1,G1, Eiglm, ITH	0.998	0.997	0.996	0.904	1.809×10^3	0.000
3	DD1,G1, Eiglm	0.998	0.996	0.996	0.911	3.334×10^3	0.000

Table3. Statistical parameters of the models calculated with the SPSS software for the S of 2-Phenylindole derivatives

Model	Descriptors	R	R^2	R^2_{adj}	RMSE	F	Sig
1	DD1,G1,ITH, Eiglm RDF0504, RDF040e	0.996	0.991	0.989	1.419	543.754	0000
2	DD1,G1,ITH, RDF0504, Eiglm,	0.996	0.991	0.990	1.405	707.122	0000
3	DD1,G1,ITH, RDF0504,	0.996	0.991	0.990	1.395	971.577	000
4	DD1,G1,ITH	0.995	0.991	0.990	1.397	1.449×10^3	0000

We studied the relationship between the molecular descriptors and the thermodynamic properties of 43 2-Phenylindole derivatives. In present study, to find the best model for predicting the mentioned properties, we will use the following sections.

3.2. Multicollinearity

In statistics, multicollinearity property is a phenomenon in which one predictor variable in a multiple regression model can be linearly predicted from the others with a substantial degree of accuracy. In other words, one predictor variable can be used to predict the other. This property in the model was examined by calculating the Pearson coefficient correlation (PCC), variance inflation factor (VIF) and Durbin–Watson (DW) statistics. The VIF, PCC and DW values were detected by SPSS program. If the VIF value lies between 1 and 10, there is no multicollinearity; if VIF<1 or >10, or PCC> 0.5 there is multicollinearity [37-39]. The VIF is calculated as follows:

$$\mathrm{VIF} = \frac{1}{1 - R^2} \qquad (3)$$

The suitable linear model for QSPR study of the heat capacity (Equation **1**) includes three molecular descriptors (DD1,G1 and Eiglm). The results of the correlation between these descriptors are listed in Table **4**.

As can be seen in Table 4, the VIF for two descriptors (Eiglm, G1) are bigger than 10, and the PCC between them is close to unity therefore there is linearity between these descriptors. After removing Eiglm, and the next step DD1 from this model, we corrected Equation (**1**) as follows:

Cv=4.559+2.181G1 (4)

N=30, R=0.951, R^2 =0.907, R^2_{adj}=0.901, RMSE=2.019, F=268.563, DW=1.872

For the entropy (Equation **2**), the Pearson correlation coefficient between ITH and G1 is close to the unit and VIF for these descriptors are bigger than 10(see Table **5**). After removing ITH, and the next step DD1 from this model, we corrected Equation (**2**) as follows:

S=52.938+2.997G1 (5)

N=30, R=0.949, R^2 =0.901, R^2_{adj}=0.897, RMSE=2.505, F=54.934, DW=1.887

In our all models, the value of Durbin-Watson statistic at the 0.05 level of significance is smaller than 2 (see Table 6) and bench the error are uncorrelated [38].

Table 5. Correlation between the molecular descriptors in Equation (2) for the entropy

Descriptor	ITH	DD1	G1	Tolerance	VIF_1	VIF_2	VIF_3
ITH	1	0.333	0.912	0.083	13.473	-	-
DD1		1	0.29	0.445	2.272	3.200	-
G1			1	0.074	11.987	2.020	1

3.3 Validations

In statistics and chemometrics several validation techniques have been proposed to estimate the model prediction capability. The model prediction capability is something different from the model fitness capability, i.e. the ability of the model to estimate the response for objects that do not participate to the calculated model.

The leave-one-out cross-validation technique (LOO_{CV}) is the simplest way to obtain the predictive power of the model [40-42]. In LOOcv method, one of the dataset is randomly deleted. This procedure is repeated until 20% of the data set is removed, then Q^2 is calculated for each deletion. Finally, Q^2 is obtained for the data set by the average of Q^2s.

The Q^2_{LOO} values of the heat capacity and entropy models (Equations (4, 5)) calculated 0.902, and 0.889 respectively.

In order to external validation test, the data set of 43 compounds and randomly separated into a training set of 30 compounds (70%), that was applied to test the made model a prediction set of 13 compounds (30%).

Statistical factors such as correlation coefficient (R), squared multiple correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), Fisher ratio (F) and root mean square error (RMSE), of final models(Equations (4, 5)) for training and external validation sets which that was used to create and test model are listed in Table 6.

These factors show potential of the GA-MLR model for prediction of the C_V and S values of 2-Phenylindole derivatives as anticancer agents and this method could simulate the complicated linear relationship between the studied properties and molecular descriptors.

The predicted and observed values of the Cv and S of 2-Phenylindole derivatives and comparison between them are listed in Table **7**. Figures **2, 3** present the linear correlation between the observed and predicted values of the above mentioned properties that were obtained using Equations 4, 5.

3.4 Residuals

Residuals, in the context of regression models, are the difference between the observed (experimental) value of the target variable (y) and the predicted (calculated) value (ŷ), i.e. the error of the prediction. The residuals plot shows the difference between residuals on the vertical axis and the dependent variable on the horizontal axis [43]. The residual plots can be used to assess the quality of the regression and the underlying statistical assumptions about residuals such as constant variance, independence of variables and normality of the distribution. For these assumptions to hold true for a particular regression model, the residuals would have to be randomly distributed around zero.

As can be seen in figures 4 and 5, the points in residual plots are randomly scattered around the horizontal axis therefore a linear model provides a decent fit to the data.

The residual value of the Cv and S expressed by Equations **4** and **5** are shown in Table **7**.

Table 6. Statistical factors obtained by the GA- MLR model for the heat capacity and entropy (Equations.**4,5**)

Dataset	Property	N	R	R^2	R^2_{adj}	RMSE	DW	F	Sig
Training	Cv	30	0.951	0.907	0.901	2.019	1.872	268.563	0.000
Test	Cv	13	0.924	0.853	0.840	2.087	1.839	63.969	0.000
Training	S	30	0.949	0.901	0.897	2.505	1.887	254.934	0.000
Test	S	13	0.934	0.877	0.858	2.517	1.219	73.804	0.000

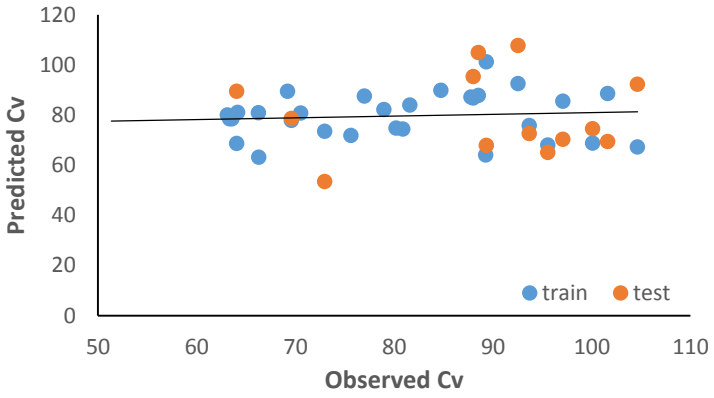

Fig. 2. Comparison between the predicted and observed heat capacity by the GA-MLR method for the training and test sets.

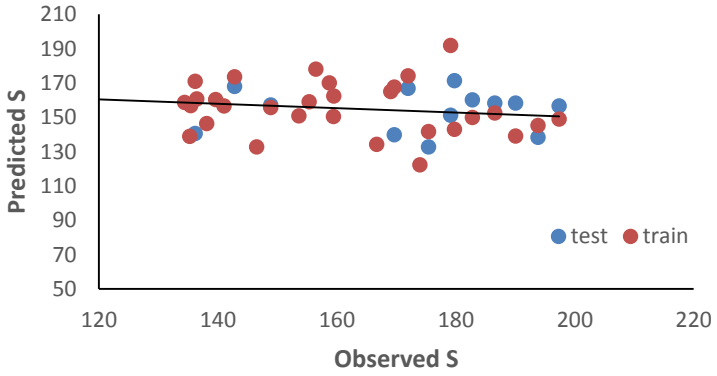

Fig. 3. Comparison between the predicted and observed entropy by the GA-MLR method for the training and test sets.

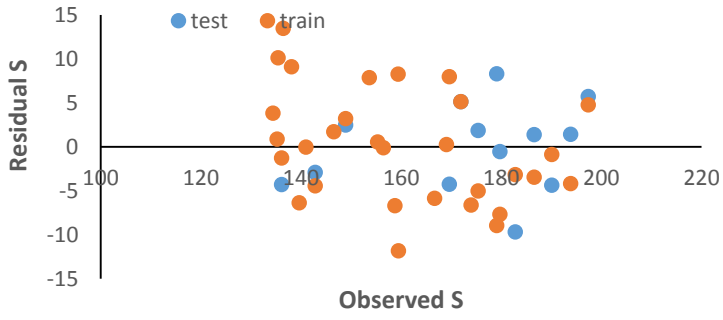

Fig. 4.Residuals plotted against the observed entropy for training and test sets of 2- Phenylindole derivatives.

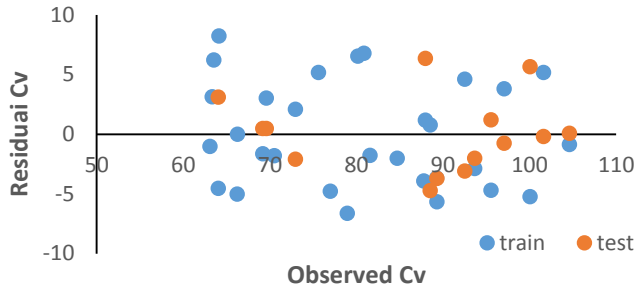

Fig .5.Residuals plotted against the observed heat capacity for training and test sets of 2- Phenylindole derivatives.

Table 7. The observed, predicted and residual value for training and test sets of 2- Phenylindole derivatives

Observed S(J/molK)	Predicted S	Residual S	Observed Cv(J/molK)	Predicted C_V	Residual C_V
136.183	141.752	-5.569	64.039	68.279	-4.240
171.990	170.782	1.208	88.001	87.466	0.535
169.659	172.263	-2.604	88.532	88.445	0.087
179.142	175.418	3.724	92.546	90.530	2.016
186.584	179.793	6.791	97.079	93.421	3.658
193.873	184.016	9.857	101.62	96.212	5.408
175.449	168.730	6.7187	89.332	86.110	3.222
179.778	173.478	6.300	93.681	89.248	4.433
182.820	193.689	-10.869	95.558	102.606	-7.048
190.110	198.097	-7.987	100.097	105.519	-5.422
197.396	202.363	-4.967	104.638	108.338	-3.700
148.924	153.221	-4.297	72.974	75.859	-2.885
142.852	147.262	-4.410	69.593	71.921	-2.328
141.026	145.876	-4.850	69.213	71.005	-1.792
135.459	140.901	-5.442	63.352	67.717	-4.365
136.482	142.079	-5.597	63.549	68.495	-4.946
135.299	143.580	-8.281	64.139	69.487	-5.348
146.566	150.502	-3.936	70.529	74.062	-3.533
138.161	139.509	-1.348	66.291	66.796	-0.505
153.661	149.661	4.000	75.625	73.506	2.119
159.440	153.458	5.982	80.195	76.016	4.179
158.765	156.497	2.268	80.870	78.024	2.846
169.110	174.956	-5.846	87.792	90.224	-2.432
156.484	157.460	-0.976	81.586	78.661	2.925
166.746	157.579	9.167	84.738	78.740	5.998
174.017	161.594	12.423	89.279	81.393	7.886
159.526	161.178	-1.652	78.978	81.118	-2.140
139.676	139.298	0.378	66.242	66.657	-0.415

Observed S(J/molK)	Predicted S	Residual S	Observed Cv(J/molK)	Predicted C_V	Residual C_V
134.430	133.056	1.374	63.096	62.531	0.565
155.357	146.916	8.441	76.990	71.692	5.298
158.697	148.860	9.837	81.317	73.204	8.113
158.569	148.334	10.235	81.209	72.868	8.341
163.321	167.995	-4.674	83.235	85.446	-2.211
165.262	162.983	2.278	82.142	82.240	-0.098
170.618	172.039	-1.421	87.778	88.033	-0.255
177.915	175.971	1.944	92.320	90.549	1.771
128.304	132.273	-3.969	60.354	62.592	-2.238
115.685	120.471	-4.786	51.364	55.041	-3.677
150.441	160.284	-9.843	75.546	80.513	-4.967
153.837	158.225	-4.388	75.817	79.196	-3.379
162.315	156.496	5.819	79.120	78.089	1.031
159.547	157.012	2.535	79.030	78.419	0.611
165.026	168.593	-3.567	82.787	85.829	-3.042

3.5.Interpretation of the best descriptors

As it is shown in the results and discussion section, only one descriptors (G1) can be used successfully for modeling and predicting the heat capacity (Cv) and entropy(S) of 2- Phenylindole derivatives. G1 is molecular descriptor derived from a geometrical representation that are called geometrical or 3D-descriptors [44-46]. Because a geometrical representation involves knowledge of the relative positions of the atoms in 3D space, i.e. the (x, y, z) atomic coordinates of the molecule atoms, geometrical descriptors usually provide more information and discrimination power than topological descriptors for similar molecular structures and molecule conformations.

The different sets of geometrical descriptors are the Weighted Holistic Invariant Molecular (WHIM), the Geometry, Topology, and Atom-Weights Assembly (GETAWAY), the 3-Dimensional- Molecule Representation of Structures based on Electron diffraction (3D-MoRSE), the Radial Distribution Function (RDF), the alignment-independent (EVA) and the electronic eigenvalue (EEVA) descriptors.

These descriptors have been applied to study relationships between the knowledge of the 3D structure of the molecule and complex properties/activities, by exploiting their large information content [47-59].

The heat capacity at constant volume (Cv) and entropy(S) are strongly depend on (x,y,z)coordinates of the molecule atoms and other quantities derived from the coordinates, e.g. interatomic distances or distances from a specified origin. Many of these are derived from the molecular geometry matrix G (or geometric distance matrix) defined by all the geometrical distances r_{ij} between atom pairs. The geometry matrix G contains information about molecular configurations and conformations, and it can be calculated as the following:

$$G \equiv \begin{vmatrix} 0 & r_{12} & \cdots & r_{1A} \\ r_{21} & 0 & \cdots & r_{2A} \\ \cdots & \cdots & \cdots & \cdots \\ r_{A1} & r_{A2} & \cdots & 0 \end{vmatrix} \qquad (6)$$

where A is the number of atoms in the molecule. Diagonal entries are always zero.

The different sets of geometrical descriptors are the Weighted Holistic Invariant Molecular (WHIM) , the Geometry, Topology, and Atom-Weights Assembly (GETAWAY), the 3-Dimensional- Molecule Representation of Structures based on Electron diffraction (3D-MoRSE), the Radial Distribution Function (RDF), the alignment-independent (EVA) and the electronic eigenvalue (EEVA) descriptors.

4.CONCLUSION

In present study, QSPR models have been used to predict the heat capacity ($logP_{ow}$) and entropy of 2-Phenylindole derivatives as anticancer agents by the genetic algorithm -multiple linear regressions (GA-MLR).

DRAGON software has been used to calculate a variety of molecular descriptor derived from different molecular representations. The backward stepwise multiple linear regression and genetic algorithm techniques were applied to select the best molecular descriptors.

Leave-One-Out Cross-Validation (LOOCV), and external validation were proposed to verify the predictive efficiency of constructed QSPR models. The results of QSPR study show that 3D or geometrical descriptor contains important structural information sufficient to develop useful predictive models for thermodynamic properties.

The results of validations and high statistical quality of GA-MLR models indicate that the generated models can be applied to predict the Cv and S of the studied compounds. Also these developed QSPR models can be used to predict thermodynamic properties of new 2-Phenylindole derivatives as anticancer agents.

REFERENCES

[1] E. N. Muratov, J.ajorath, R. P. Sheridan, I. V. Tetko, D. Filimonov, V. Poroikov, T.I. Oprea, I. I. Baskin, A. Varnek, Chem. Soc. Rev., 49 (2020) 3716.

[2] N.Ahmadinejad, F. Shafiei, T. Momeni Isfahani, Comb. Chem. High Throughput Screen. 21 (2018) 533.

[3] E. Pourbasheer, R. Aalizadeh, M.R. Ganjali, P. Norouzi, Med. Chem. Res. 23 (2014) 57.

[4] D. E. Arthur, A. Uzairu, P. Mamza, E. Abechi, G. Shallangwa, Albanian . J. Pharm .Sci .3(2016)4.

[5] D. Joudaki, F. Shafiei, Curr. Comput-Aided. Drug Des. 16 (2020) 571.

[6] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors; Wiley-VCH: Weinheim, 2000.

[7] V. Kamath, A. Pai, J. Pharm. and Tech. 10 (2017) 3237.

[8] K. Vorčáková, M. Májeková, E. Horáková, P. Drabina, M. Sedlák, Š. Štěpánková, Bioorg. Chem.78(2018) 280.

[9] J. Verma, V. M. Khedkar, E. C. Coutinho, Curr. Top. Med. Chem. 10 (2010) 95.

[10] H. Kubinyi, G. Folkers, Y. C. Martin, 3D-QSAR in Drug Design: Volume 3: Recent Advances, Kluwer Academic Publishers. New York, 2002.

[11] S. S. El-Nakkady, M. M. Hanna, H. M. Roaiah, I. A. Ghannam, Eur. J. Med. Chem. 47 (2012) 387.

[12] R. Gaikwad, S.A. Amin, N. Adhikari, S. Ghorai, T. Jha, S. Gayen, Struct. Chem.29(2018) 1095.

[13] A.A.Toropovand, A.P. Toropova, Anticancer. Res. 38 (2018) 6189.

[14] S. Y. Liao, L. Qian, T.F. Miao, H. L. Lu, K. C. Zheng, Eur. J. Med. Chem. 44 (2009) 2822.

[15] A. K. Halder, N. Adhikari, T. Jha, Bioorganic Med. Chem. Lett. 19 (2009) 1737.
[16] S. C. Basak, Q. Zhu, D. Mills, Curr. Comput. Aided. Drug Des. 7 (2011) 98.
[17] M. N. Aldosari, K. K. Yalamanchi, X. Gao, S. Mani Sarathy, Energy and AI. 4 (2021) 100054.
[18] D. Jaramillo, G. Plascencia, Basic Thermochemistry in Materials Processing. Springer International Publishing, 1st, 2017.
[19] J. P. Lowe, K. A. Peterson, Quantum chemistry third edition. 3rd ed. Elsevier, 2006.
[20] P. Sullivan, G. Seidel, Phys. Rev. 173(1968) 679.
[21] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh. Nature. 1 (2018) 547.
[22] K. K. Yalamanchi, V. C. O. Van Oudenhoven, F. Tutino, M. Monge-Palacios, A. Al- shehri, X. Gao, S.M. Sarathy, J. Phys. Chem. A.,123(2019) 8305.
[23] C. C. J. Roothaan, Rev. Mod. Phys. 23 (1951) 69.
[24] J. S. Binkley, J. A. Pople, W. J. Hehre, J. Am. Chem. Soc. 102 (1980) 939.
[25] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. A. Pople, Gaussian, Inc., Wallingford CT, 2009.
[26] Talete srl, Dragon (ver. 5.4), Milano, Italy. Web site: www.talete.mi.it/products/software.htm
[27] I. Dohoo, C. Ducrot, C. Fourichon, A. Donald, D. Hurnik, Prev. Vet. Med. 29 (1997) 221.
[28] S. J. Cho, M. A. Hermsmeier, J. Chem. Inf. Comput. Sci. 42 (2002) 927.
[29] K. H. Baumann, H. Albert M. V. Korff, J. Chemometr. 16 (2002) 339.
[30] P. Gramatica, P. Pilutti, E. Papa, SAR. QSAR. Environ. Res. 13 (2002) 743.
[31] M. V. Diudea, QSPR/QSAR studies for molecular descriptors, Ed Nova Science Hunting don, New York. 2000.
[32] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269.
[33] N. R. Hateka, Tests for Detecting Autocorrelation. Principles of Econometrics: An Introduction (Using R). SAGE Publications, 2010.
[34] R. Benigni , C. Bossa, J. Chem. Inf. Model.48(2008) 971.
[35] T. A. Craney, J. G. Surles, Qual. Eng. 14 (2002) 391.
[36] D. G. Kleinbaum, Applied regression analysis and other multivariable methods; Australia, Belmont, CA: Brooks/Cole, 2008.
[37] A. Fisher, Statistical Methods for Research Workers; Oliver and Boyd: Edinburgh, UK, 1925.
[38] B. Reisfeld, A. N. Mayeno, Computational Toxicology: Volume 21, On the Development and Validation of QSAR Models, Springer: Science+Business Media, LLC, 2013.
[39] S. Chatterjee, J. Simonoff, Handbook of Regression Analysis. John Wiley & Sons: New York, 2013.
[40] M. Zhao, D. Wei, Exploring the ligand-protein networks in traditional Chinese medicine: current databases, methods and applications. In Advance in Structural Bioinformatics, Springer, Dordrecht, 2015.
[41] G. C. Siontis, I. Tzoulaki, P. J. Castaldi, J. P. Ioannidis, J. P. J. Clin. Epidemiol. 68 (2015) 25.
[42] F. K. Martens, J. G. Kers, A. C. Janssens, J. Clin. Epidemiol. 68 (2015) 25.
[43] Y. Dodge, The Concise Encyclopedia of Statistics. Springer, 2008.

[44] D. G. Kleinbaum, Applied regression analysis and other multivariable methods, Australia; Belmont, CA: Brooks/Cole, 2008.
[45] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, J. Chem. Inform. Comput. Sci. 42 (2002) 693.
[46] S. Sahoo, C. Adhikari, M. Kuanar, B.K. Mishra, Curr. Comput- Aided. Drug Des. 12 (2016) 181.
[47] A. T. Balaban, From Chemical Topology to Three-Dimensional Geometry, A. T. Balaban (Ed.), Plenum Press, New York (NY), 1997.
[48] B. Hu, Z. Kun Kuang , S. Y. Feng, D. Wang, S. B. He, D. Xin Kong, Molecules. 21 (2016) 1554.
[49] J. Verma, V. M. Khedkar, E. C. Coutinho, Curr. Topics Med. Chem. 10 (2010) 95.
[50]E. Estrada, I. Perdomo-López, J. J. Torres-Labandeira. J. Chem. Inform. Comput. Sci. 41 (2001) 1561.
[51] A. Rybi ´nska-Fryca, A. Sosnowska, T. Puzyn, Materials. 13(2020) 2500.
[52] N. Ahmadinejad, Shafiei, Comb. Chem. High. Throughput. Screen. 22 (2019) 387.
[53] F. Ghaemdoost, F. Shafiei, Curr. Comput- Aided. Drug. Des. 6 (2020) 25.
[54] H. Kušić, B. Rasulev, D. Leszczynska, J. Leszczynski, N. Koprivanac, Chemosphere. 75 (2009) 1128.
[55] J. Schuur, J. Gasteiger, Anal. Chem. 69 (1997) 2398.
[56] L. Saíz-Urra, Y. Pérez-Castillo, M. Pérez González, R. Molina Ruiz, M. Cordeiro,J.E.Rodríguez-Borges, X. García-Mera, QSAR & Comb. Sci. 28 (2009) 98.
[57] L. Saíz-Urra, M. P. González, M. Teijeira, Biorg. Med. Chem. 14 (2006) 7347.
[58] P. R. Duchowicz, M. G. Vitale, E. A. Castro, M. Fernández, J. Caballero, Biorg. Med. Chem. 15 (2007) 2680.
[59] Z. Cheng, Y. Zhang, W. Fu, Eur. J. Med. Chem. 45 (2010) 3970.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۸، شماره ۱، بهار ۱۴۰۰
ISSN ۱۷۳۵-۲۱۲۶

مطالعه ارتباط کمی ساختار–خاصیت مشتقات ۲–فنیل ایندول به عنوان داروهای ضد سرطان با استفاده از توصیف کننده های مولکولی

سمیرا بهرامی ، فاطمه شفیعی*، اعظم مرجانی، طاهره مومنی اصفهانی

گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

چکیده

مطالعه ارتباط کمی ساختار–خاصیت (QSPR)بر روی مجموعه ای از مشتقات ۲–فنیل ایندول به عنوان عوامل ضد سرطان انجام شده است تا توصیف کننده مولکولی مهمی که مرتبط با خواص ترمودینامیکی آنها مانند ظرفیت گرمایی (Cv) و آنتروپی (S) می باشند، تعیین شوند. توصیف کننده های مولکولی بدست آمده از نرم افزار دراگون (DRAGON) با استفاده از روش های الگوریتم ژنتیک (GA) و برگشتی کاهش داده شده و توصیف کننده های مناسب انتخاب شدند. تجزیه و تحلیل رگرسیون چند متغیره خطی(MLR) برای به دست آوردن مدل‌های QSPR مورد استفاده قرار گرفته است. این مدل‌ها با استفاده از ضرایب آماری مانند، مربع ضریب همبستگی (R^2)، ریشه میانگین مربع خطا (RMSE) ، ضریب همبستگی تعدیل شده(R^2adj) و شاخص کیفیت فیشر (F) مورد ارزیابی قرار گرفتند. قدرت پیش بینی مدلهای MLR با استفاده از روش‌های اعتبارسنجی متقابل (LOOCV) و مجموعه آزمون مورد بحث قرار گرفت. بهترین مدل های QSPR برای پیش‌بینی Cv (J/molK) و S (J/molK) بترتیب دارای R^2 = 0.907 و 0.901، RMSE = 2.019 و2.505، و مربع ضریب همبستگی اعتبارسنجی متقابل R^2cv = 0.902 و 0.889 می باشند. نتایج آماری حاصل از مطالعه حاضر، قابلیت پیش‌بینی خوبی را نشان می‌دهد و ممکن است در طراحی مشتقات جدید ۲–فنیل ایندول مفید باشد.

کلید واژه‌ها: مشتقات ۲–فنیل ایندول، ارتباط کمی ساختار– خاصیت، ظرفیت گرمایی، آنتروپی، الگوریتم ژنتیک –رگرسیون چند متغیره خطی

* مسئول مکاتبات: f-shafiei@iau-arak.ac.ir