

Research Article



# A QSAR Study on the 4-Substituted Coumarins as Potent Tubulin Polymerization Inhibitors

Leila Dinparast<sup>1</sup> , Siavoush Dastmalchi<sup>2,3,4\*</sup> 

<sup>1</sup>Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

<sup>2</sup>Pharmaceutical Analysis Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.

<sup>3</sup>School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran.

<sup>4</sup>Faculty of Pharmacy, Near East University, POBOX: 99138, Nicosia, North Cyprus, Mersin 10, Turkey.

## Article info

### Article History:

Received: 5 Mar. 2019

Revised: 13 Oct. 2019

Accepted: 9 Nov. 2019

published: 18 Feb. 2020

### Keywords:

- Coumarin
- Cancer
- Antiproliferative
- QSAR
- GA-MLR

## Abstract

**Purpose:** Despite the discovery and synthesis of several anticancer drugs, cancer is still a major life threatening incident for human beings after cardiovascular diseases. Toxicity, severe side effects, and drug resistance are serious problems of available commercial anticancer drugs. Coumarins are synthetic and natural heterocycles that show promising antiproliferative activities against various tumors. The aim of this research is to computationally study the coumarin derivatives in order to develop reliable quantitative structure-activity relationship (QSAR) models for predicting their anticancer activities.

**Methods:** A data set of thirty one coumarin analogs with significant antiproliferative activities toward HepG2 cells were selected from the literature. The molecular descriptors for these compounds were calculated using Dragon, HyperChem, and ACD/Labs programs. Genetic algorithm (GA) accompanied by multiple linear regression (MLR) for simultaneous feature selection and model development was employed for generating the QSAR models.

**Results:** Based on the obtained results, the developed linear QSAR models with three and four descriptors showed good predictive power with  $r^2$  values of 0.670 and 0.692, respectively. Moreover, the calculated validation parameters for the models confirmed the reliability of the QSAR models.

**Conclusion:** The findings of the current study could be useful for the design and synthesis of novel anticancer drugs based on coumarin structure.

## Introduction

Cancer is one of the severe life-threatening human health problems worldwide.<sup>1,2</sup> Despite significant development in cancer chemotherapy in the past 50 years, cancer continues to be the second most frequent cause of death after cardiovascular diseases.<sup>2,3</sup> There are numerous reports in the literature on the discovery of novel anticancer agents, but there is no single drug with 100% efficacy for the cancer treatment.<sup>4</sup> Most of the clinically used drugs have limited effectiveness and selectivity, accompanied with serious toxicity, and unacceptable side effects.<sup>5,6</sup> Moreover, the most common tumors show resistance against the significant number of commercially available anticancer drugs.<sup>6</sup> Therefore, considerable demand for the discovery of efficient new anticancer drugs continues to exist in order to overcome the current chemotherapeutic problems in cancer treatment.<sup>5</sup>

Coumarin and its derivatives are important oxygen containing heterocycles which are found in natural products. Over the past decades, coumarins have attracted great attention because of their interesting biological and

pharmacological activities such as anticoagulant,<sup>7</sup> anti-inflammatory,<sup>8,9</sup> antioxidant,<sup>10</sup> antiviral,<sup>11</sup> antimicrobial,<sup>12,13</sup> antidepressants,<sup>14</sup> and anti-HIV effects.<sup>15-17</sup> Also, they are promising compounds due to their low toxicity, little drug resistance, less side effects, high bioavailability, and ease of chemical synthesis.<sup>18</sup> Several studies have shown that coumarins are potential anticancer agents having growth suppressive effects on many types of cancers such as ovarian,<sup>19</sup> breast,<sup>20,21</sup> skin,<sup>22</sup> prostate,<sup>23</sup> liver,<sup>24,25</sup> and pancreatic.<sup>26,27</sup>

Computer assisted drug design (CADD) has attracted considerable attention in modern drug discovery and development by reducing the time-consuming and expensive synthetic and biological experiments needed to achieve the required results.<sup>28</sup> Quantitative structure-activity relationship (QSAR) studies as part of CADD techniques play a critical role in medicinal chemistry for the design of new therapeutically active compounds.<sup>29-31</sup> QSAR studies are used for the prediction of the biological activity and may also be used for the interpretation of the mode of ligand-receptor interaction. The required time and

\*Corresponding Author: Siavoush Dastmalchi, Tel: +98 41 33364038, Email: dastmalchi.s@tbzmed.ac.ir

cost spent for drug design and discovery are significantly decreased by using various QSAR techniques.<sup>32</sup>

In the current work, a QSAR analysis was conducted on a set of coumarin analogs for which biological activities have been reported in the literature.<sup>33</sup> Using GA-MLR-based two-dimensional QSAR analysis, the cell toxicity of the studied coumarins was correlated to their structural features. Based on the obtained results, the developed linear models showed good predictive power, and can be used in designing new anticancer agents.

## Materials and Methods

### Methodology

#### Data set

The experimental  $IC_{50}$  (nM) values obtained for antiproliferative activities of coumarin derivatives (31 compounds) against HepG2 cell line, reported by Cao et al,<sup>33</sup> were used in the present study. For QSAR analysis, all the biological data were converted into  $pIC_{50}$  (i.e.,  $-\log IC_{50}$ ).

#### Molecular descriptors

The 3D structures of the ligands were built by GaussView 5.0 software.<sup>34</sup> The energy minimization of the structures were conducted initially using the empirical method (i.e., MM+) <sup>35</sup> followed by semi-empirical technique AM1<sup>36</sup> using the Polak-Ribiere algorithm included in HyperChem 7.5 software.<sup>37</sup> The molecular descriptors for the fully optimized molecular structures were calculated using Dragon (version 3.0) program.<sup>38</sup> Log  $P$  and log  $D$  were calculated by ACD/Labs 6.0 program<sup>39</sup> while the molar refractivity, surface area, density, and polarizability were calculated using HyperChem 7.5 software. From the total different molecular descriptors calculated by Dragon software, descriptors with 50% constant values were omitted. Moreover, descriptors were pretreated to remove those with more than 0.95 correlations.<sup>40</sup> These pretreatments on the descriptors were performed using R 3.2.3 software.<sup>41</sup>

#### Methods

Three algorithms were used for dividing the data set into train and test sets. These include Kenard-Stone, Euclidian Distance, and Activity/Property methodologies which are available in a java-based tool.<sup>42,43</sup> For reducing the number of molecular descriptors, as well as selecting the appropriate features, multi linear regression (MLR) method optimized by incorporating the GA algorithm known as GA-MLR was used. This tool is a java-based graphical user interface and proposes an MLR model based on five validation parameters i.e.  $r^2$ ,  $r^2_{Adjusted}$ ,  $q^2$ ,  $\overline{rm}^2$ , and  $\Delta\overline{rm}^2$  with their default values set to  $> 0.6$ ,  $> 0.6$ ,  $> 0.6$ ,  $> 0.5$ , and  $< 0.2$ , respectively.<sup>44</sup> The GA-MLR approach was carried out with its default settings for finding the linear equations with three and four parameters. Although, GA-MLR was only applied on the train set, however for validating the

generated models on the test set compounds, four criteria, i.e.,  $Q^2_{(test)}$ , absolute percentage error (APE), mean absolute percentage error (MAPE), and standard deviation of error of prediction (SDEP) calculated according to equations 1, 2, 3, and 4, were used:

$$Q^2_{(test)} = 1 - \frac{Press}{\sum_{i=1}^N (y_{obs,i} - y_m)^2}$$

$$Press = \sum_{i=1}^N (y_{pred,i} - y_{obs,i})^2 \quad (1)$$

$$APE = \frac{|pIC_{50(pred)} - pIC_{50(obs)}|}{pIC_{50(obs)}} \times 100 \quad (2)$$

$$MAPE = \frac{\sum_{i=1}^N APE}{N} \quad (3)$$

$$SDEP = \sqrt{\frac{\sum_{i=1}^N (y_{obs,i} - y_{pred,i})^2}{n}} \quad (4)$$

Where,  $y_{obs,i}$ ,  $pIC_{50(obs)}$ ,  $y_{pred,i}$  and  $pIC_{50(pred)}$  are the experimental and predicted activities of an individual compound in the test set, respectively.  $N$  is the number of molecules and  $y_m$  is the mean of experimental biological activities of the compounds. PRESS is the predictive residual sum of the squares.

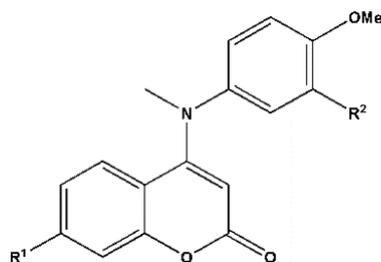
The applied fitness function (i.e.,  $F$ ) in this approach is as follow (for more details readers may be referred to the manual of the GA-MLR):

$$F = \sum_{i=0}^{i=k} \frac{P_i - \hat{P}_i}{P_{i,Max} - \hat{P}_{i,Min}} \quad (5)$$

## Results and Discussion

The structures of coumarin analogs used in the current study are shown in Table 1. The size, lipophilicity, and electronic features of the substituents are different. For extracting chemical information from the data set compounds, computing a wide range of structural descriptors is essential for any successful QSAR analysis. In the various fields of chemometrics, it is clear that utilizing an effective variable selection method which results in reducing the complexity of the model, can improve the interpretability and the predictive ability of the developed model.<sup>45-47</sup> For developing a QSAR model that explains the antiproliferative activities of the compounds shown in Table 1 on HepG2 cells, large number of structural parameters belonging to different classes of descriptors such as those listed in Table 2, are used.

The total set of compounds was randomly divided into train (21 compounds, 70% of the whole data set) and test sets (10 compounds, 30% of whole data set) for the generation of QSAR models and validating the developed models, respectively. For this purpose, the hybrid methodology developed in Roy's Lab known as GA-MLR was used on

**Table 1.** Chemical structures, experimental and predicted  $pIC_{50}$  values of coumarin analogs used in this study which data collected from Cao et al for model construction and absolute percentage errors for the developed models


Compound	R <sup>1</sup>	R <sup>2</sup>	$pIC_{50}$ (Exp.)	$pIC_{50}$ (Pred.)	APE (Eq. 6)	APE (Eq. 7)
1	OCH <sub>3</sub>	OH	6.406	6.340	1.021	0.431
2*	H	O-C(O)CH <sub>3</sub>	7.833	8.054	4.818	2.529
3	H	O-C(O)CHCH <sub>2</sub>	7.504	7.866	1.657	0.724
4	H	O-C(O)C(CH <sub>3</sub> ) <sub>3</sub>	8.167	8.032	3.726	4.489
5	H	O-C(O)CH <sub>2</sub> Cl	8.119	7.817	3.254	3.073
6	H	O-C(O)CH <sub>2</sub> Br	7.963	7.703	0.051	0.556
7*	H	O-C(O)CH <sub>2</sub> CH <sub>3</sub>	7.889	7.615	9.305	9.618
8	H	O-C(O)(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	7.456	7.452	0.056	0.875
9	H	O-C(O)C(CH <sub>2</sub> ) <sub>2</sub> CH <sub>3</sub>	7.221	7.893	0.057	0.344
10	H	O-C(O)CHC(CH <sub>3</sub> ) <sub>2</sub>	7.676	7.680	3.746	5.386
11	H	O-C(O)CHCHCH <sub>3</sub>	7.845	7.849	3.092	0.027
12	H	O-C(O)CCH <sub>2</sub> CH <sub>2</sub> C	7.407	7.684	0.775	1.126
13*	H	O-C(O)CH <sub>2</sub> CH <sub>2</sub> C(O)(O)CH <sub>2</sub> CH <sub>3</sub>	7.168	7.690	0.114	2.433
14	H	O-C(O)(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>	7.087	7.306	0.198	2.809
15	H	O-C(O)(CH <sub>2</sub> ) <sub>8</sub> CH <sub>3</sub>	6.612	6.560	5.313	4.060
16	H	O-C(O)(CH <sub>2</sub> ) <sub>3</sub> C(O)(O)CH <sub>3</sub>	7.565	7.557	5.367	3.592
17	H	O-C(O)(CH <sub>2</sub> ) <sub>3</sub> CH <sub>2</sub> Cl	7.442	7.428	0.113	2.529
18*	H	O-C(O)CH <sub>2</sub> C(CH <sub>3</sub> ) <sub>3</sub>	7.140	7.209	1.088	1.399
19*	H	O-C(O)(CH <sub>2</sub> ) <sub>2</sub> CHCH <sub>2</sub>	7.629	7.509	9.202	7.212
20	H	O-C(O)C(CH <sub>3</sub> )CHCH <sub>2</sub> CH <sub>3</sub>	8.056	7.628	0.486	5.691
21	H	O-C(O)CHCH(CH <sub>2</sub> ) <sub>3</sub>	7.991	7.562	12.410	8.225
22	H	NH <sub>2</sub>	8.481	8.491	2.825	2.086
23	H	HN-C(O)CH <sub>2</sub> CH <sub>2</sub> C(O)(O)CH <sub>2</sub> CH <sub>3</sub>	7.529	7.611	3.482	6.674
24	H	HN-C(O)(CH <sub>2</sub> ) <sub>3</sub> C(O)(O)CH <sub>3</sub>	8.018	7.280	7.279	2.640
25*	H	HN-C(O)(CH <sub>2</sub> ) <sub>4</sub> CH <sub>3</sub>	7.318	7.196	0.958	0.004
26	H	HN-C(O)(CH <sub>2</sub> ) <sub>8</sub> CH <sub>3</sub>	6.543	6.574	1.569	5.705
27*	H	HN-C(O)(CH <sub>2</sub> ) <sub>3</sub> CH <sub>2</sub> Cl	7.564	7.303	1.659	1.358
28*	H	HN-C(O)CHCHCH <sub>3</sub>	7.676	7.596	3.452	0.591
29*	H	HN-C(O)(CH <sub>2</sub> ) <sub>2</sub> CHCH <sub>2</sub>	7.582	7.541	1.033	3.408
30	H	HN-C(O)C(CH <sub>3</sub> )CHCH <sub>2</sub> CH <sub>3</sub>	6.233	7.006	0.543	3.393
31*	H	OH	8.523	8.443	0.940	3.897

\* Test set.

**Table 2.** Details of four most important descriptors were used in model construction

Symbol	Descriptor Block	Description
RDF030u	Radial Distribution Function descriptors	Radial Distribution Function - 030 / unweighted
LP1	Topological (2D matrix-based descriptors)	Lovasz-Pelikan index (leading eigenvalue)
EEig02x	Topological (Edge adjacency indices)	Eigenvalue 02 from edge adj. matrix weighted by edge degrees
Mor04p	3D-MoRSE	Signal 04 / weighted by polarizability

**Table 3.** The statistical parameters of developed models

Models	N	r <sup>2</sup>	r <sup>2</sup> <sub>Adjusted</sub>	q <sup>2</sup> <sub>(LOO)</sub>	r <sup>2</sup> -q <sup>2</sup> <sub>(LOO)</sub>	Q <sup>2</sup> <sub>(test)</sub>	SEE	SDEP	$\overline{r}m^2$	$\Delta\overline{r}m^2$	P value	MAPE (Train set)	MAPE (Test set)
Eq. 6	31	0.689	0.634	0.483	0.206	0.670	0.376	0.437	0.378	0.094	8.08×10 <sup>-9</sup>	0.149	0.237
Eq. 7	31	0.749	0.686	0.530	0.210	0.692	0.349	0.416	0.417	0.138	1.89×10 <sup>-9</sup>	0.152	0.298

three train sets differently selected based on the three data division algorithms mentioned in Methods section. The model building processes for each set were run for thirty times for generating three-parameter models. This was led to total of 90 different models with three parameters. For all of these models the r<sup>2</sup> values were compared and the best data division method was identified as being Euclidean Distance method. Then, for achieving better results on this set, twenty further runs using the same settings were performed. Moreover, for generating four-parameter equations, the best three-parameter equation was used such that the fourth parameter was added one at a time from the whole pool of descriptors (i.e., in an all-walk manner) to identify the best four-parameter model. Equations 6 and 7 are the best three- and four-parameter models, respectively.

$$pIC_{50} = 425.48937 - 0.08785 (RDF030u) - 181.78082 (LP1) + 6.80549 (EEig02x) \quad (6)$$

$$pIC_{50} = 511.93488 - 0.12088 (RDF030u) - 218.19055 (LP1) + 7.61787 (EEig02x) + 0.49323 (Mor04p) \quad (7)$$

Where N, r<sup>2</sup>, Q<sup>2</sup><sub>(test)</sub> and MAPE are the number of compounds, the squared correlation coefficient of train set, the squared correlation coefficient of test set, and the mean absolute percentage error, respectively. Table 2 shows the statistical parameters of two developed models with more details. The numerical values and detailed information about the selected descriptors are listed in Tables 2 and 3. Correlation matrix of selected descriptors is represented in Table 4. The three-parameter model (Eq. 6) predicts the antiproliferative activities of the studied coumarins using RDF030u, LP1, and EEig02x descriptors. RDF030u (radial distribution function 3.0/unweighted) belongs to the group of Radial Distribution Function descriptors that are obtained by radial basis functions centered on different interatomic distances ranging from 0.5 to 15.5 Å.<sup>48</sup> The Radial Distribution Function in a system of particles (atoms, molecules, colloids, etc), describes how density varies as a function of distance from a reference

particle. When studying the chemical properties of a compound, the probability distribution of atoms scattered in a spherical volume with radius of 3.0 Å is regarded as an important factor.<sup>49,50</sup> The LP1 feature, which belongs to the topological descriptors, is one of the 2D matrix-based descriptors, and is calculated by eigenvalues of a square (usually symmetric) matrix representing a molecular graph.<sup>51</sup> Du and colleagues have reported that small and large values of LP1 are indicative of compounds with less and more branches, respectively.<sup>52</sup> On the other hand, LP1 is a molecular branching index. The negative coefficient of LP1 in both of the developed equations indicates that the pIC<sub>50</sub> is inversely related to this descriptor, which suggests that 4-substituted coumarins with lesser branches in the overall structure may show the higher antiproliferative activity. The next feature (i.e., EEig02x), also belonging to the topological descriptors has been derived from the edge adjacency matrix weighted by edge degrees.<sup>53</sup> This descriptor is associated with molecular polarity and describes the electronic effects as well as the hydrophobic properties of molecule.<sup>54</sup>

The second QSAR model (Eq. 7) describes the activities of coumarin analogs using one extra parameter added to the three previously explained features. The new parameter, i.e., Mor04p belongs to 3D-MorSE group of descriptors, and is calculated by incorporating the polarizability-based weighting of the scattering features of the molecules.<sup>55</sup> The presence of Mor04p descriptor in the developed model can be regarded as an evidence for the importance of the 3D arrangement influence of the molecules extracted from electron diffraction studies<sup>56</sup> on the antiproliferative activities of the studied compounds. The increase of pIC<sub>50</sub> directly correlates to the shape and size of the studied 4-substituted coumarin derivatives.

The predictive power of the developed models was evaluated using internal and external validation measurements. For this purpose, the squared correlation coefficient (r<sup>2</sup>), leave one out cross-validated correlation coefficient (q<sup>2</sup><sub>(LOO)</sub>), the squared adjusted correlation coefficient (r<sup>2</sup><sub>Adjusted</sub>), the standard error of estimate (SEE), the SDEP,  $\overline{r}m^2$ , and  $\Delta\overline{r}m^2$  were calculated for the train set and Q<sup>2</sup><sub>(test)</sub> was computed for the test set (Table 3). The squared correlation coefficient is the parameter fitted on the whole train set and the QSAR models with r<sup>2</sup> > 0.6 are considered reliable.<sup>57</sup> As seen in Table 3, r<sup>2</sup> values of 0.689 and 0.749 were obtained for equations 6 and 7, respectively. The q<sup>2</sup><sub>(LOO)</sub> and r<sup>2</sup> - q<sup>2</sup><sub>(LOO)</sub> are other measurement criteria for evaluating the performance of QSAR models, which should be higher than 0.5 and 0.3,

**Table 4.** Correlation matrix of selected descriptors

	EEig02x	LP1	RDF030u	Mor04p
EEig02x	1			
LP1	0.24176	1		
RDF030u	0.164324	-0.06386	1	
Mor04p	0.001153	0.122393	0.571913	1

respectively.<sup>58-60</sup> The calculated values of these parameters for equations 6 and 7 are 0.483, 0.206 and 0.530, 0.210, respectively. The generated QSAR models are the result of the GA-MLR methodology based on a uni-objective (i.e., F) optimization function. The other two metrics  $\overline{rm}^2$  and  $\Delta\overline{rm}^2$  were determined to further assess the predictive ability of the QSAR models.  $\overline{rm}^2$  metric which was introduced by Roy and Roy determines the proximity between the observed and predicted activities for the data set.<sup>32</sup> It has been suggested that for the models with reliable predictive power, the values of  $\overline{rm}^2$  and  $\Delta\overline{rm}^2$  should be more than 0.5 and lower than 0.2, respectively.<sup>61,62</sup> The obtained  $\overline{rm}^2$  and  $\Delta\overline{rm}^2$  values for equations 6 and 7 are 0.378, 0.417 and 0.094, 0.138, respectively.  $\Delta\overline{rm}^2$  for both of models are in the acceptable range but  $\overline{rm}^2$  values are lower than 0.5. As previously noted, the obtaining of not satisfying values is possible because in the applied GA-MLR tool was optimized based on only F function. The relatively small values of SDEPs (0.437 and 0.416) show the narrow distribution of error and indicate good performances of the proposed models for all the compounds in the train set. An important criterion for the external validation is  $Q^2_{(test)}$  calculated for the test (unseen) set. Its value, greater than 0.5 indicates the validity of the model. In this study,  $Q^2_{(test)}$  for the two developed models with three- and four-parameters are 0.670 and 0.691, respectively. These results demonstrate that both models have good predictive power and are reliable for the prediction of the antiproliferative activities of coumarin analogs. Furthermore, Eq. 7 has significantly higher prediction ability in comparison to Eq. 6 with *P*-value of close to zero. Figure 1 represents the correlation between the experimental and predicted  $pIC_{50}$  values according to the equations 6 and 7 for the studied coumarin compounds (total data). The resulted correlation coefficients of 0.688 and 0.717 between observed and calculated activities using Eq. 6 and Eq. 7, respectively, demonstrate the reliability of the proposed models for predictive purposes.

### Conclusion

In the present study, QSAR analysis was performed using GA-MLR method to construct models for predicting the antiproliferative activities of coumarin derivatives as potential anticancer compounds. The internal and external validation methods were used to investigate the predictive performance of the two developed MLR models. The calculated validation parameters showed that both of the models could predict biological activities of coumarins well. Based on the obtained results, the predictive power and the performance of the model with four descriptors (Eq. 7) is higher than the model with three descriptors (Eq. 6) owing to the inclusion of one more significant variable (Mor04p) in the model. Our findings could be helpful in estimating the activity as well as in designing, synthesizing, and developing the novel anticancer drugs based on coumarin scaffold.

### Ethical Issues

Not applicable.

### Conflict of Interest

Authors declare no conflict of interest in this study.

### Acknowledgments

The authors would like to acknowledge the Ministry of Health and Medical Education, and also, Biotechnology Research Center at Tabriz University of Medical Sciences for the financial support.

### References

1. World Health Organization. Cancer: Key facts. Available from: <https://www.who.int/news-room/fact-sheets/detail/cancer>. 12 September 2018.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61(2):69-90. doi: 10.3322/caac.20107
3. Arias E, Heron M, Tejada-Vera B. United States life tables eliminating certain causes of death, 1999-2001. *Natl Vital Stat Rep* 2013;61(9):1-128.

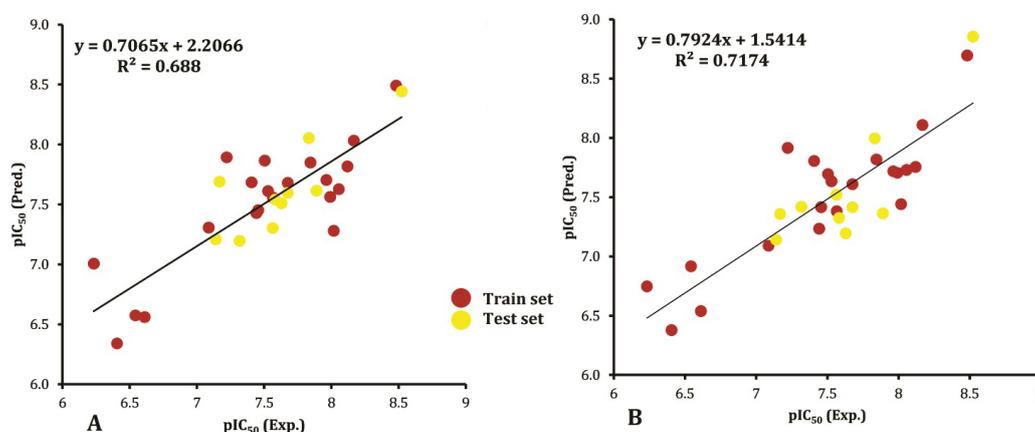


Figure 1. The correlation between the experimental and predicted  $pIC_{50}$  values for coumarin compounds (total data) according to Eq. 6 (A) and Eq. 7 (B).

4. Rahmani-Nezhad S, Safavi M, Pordeli M, Ardestani SK, Khosravani L, Pourshojaei Y, et al. Synthesis, in vitro cytotoxicity and apoptosis inducing study of 2-aryl-3-nitro-2H-chromene derivatives as potent anti-breast cancer agents. *Eur J Med Chem* 2014;86:562-9. doi: 10.1016/j.ejmech.2014.09.017
5. Gottesman MM. Mechanisms of cancer drug resistance. *Annu Rev Med* 2002;53:615-27. doi: 10.1146/annurev.med.53.082901.103929
6. Temraz S, Mukherji D, Alameddine R, Shamseddine A. Methods of overcoming treatment resistance in colorectal cancer. *Crit Rev Oncol Hematol* 2014;89(2):217-30. doi: 10.1016/j.critrevonc.2013.08.015
7. Weigt S, Huebler N, Strecker R, Braunbeck T, Broschard TH. Developmental effects of coumarin and the anticoagulant coumarin derivative warfarin on zebrafish (*Danio rerio*) embryos. *Reprod Toxicol* 2012;33(2):133-41. doi: 10.1016/j.reprotox.2011.07.001
8. Fylaktakidou KC, Hadjipavlou-Litina DJ, Litinas KE, Nicolaides DN. Natural and synthetic coumarin derivatives with anti-inflammatory/antioxidant activities. *Curr Pharm Des* 2004;10(30):3813-33. doi: 10.2174/1381612043382710
9. Bansal Y, Sethi P, Bansal G. Coumarin: a potential nucleus for anti-inflammatory molecules. *Med Chem Res* 2013;22(7):3049-60. doi: 10.1007/s00044-012-0321-6
10. Iranshahi M, Askari M, Sahebkar A, Adjipavlou-Litina D. Evaluation of antioxidant, anti-inflammatory and lipoxigenase inhibitory activities of the prenylated coumarin umbelliprenin. *Daru* 2009;17(2):99-103.
11. Hassan MZ, Osman H, Ali MA, Ahsan MJ. Therapeutic potential of coumarins as antiviral agents. *Eur J Med Chem* 2016;123:236-55. doi: 10.1016/j.ejmech.2016.07.056
12. Ostrov DA, Hernández Prada JA, Corsino PE, Finton KA, Le N, Rowe TC. Discovery of novel DNA gyrase inhibitors by high-throughput virtual screening. *Antimicrob Agents Chemother* 2007;51(10):3688-98. doi: 10.1128/aac.00392-07
13. Chougala BM, Shastri SL, Holiyachi M, Shastri LA, More SS, Ramesh KV. Synthesis, anti-microbial and anti-cancer evaluation study of 3-(3-benzofuranyl)-coumarin derivatives. *Med Chem Res* 2015;24(12):4128-38. doi: 10.1007/s00044-015-1449-y
14. Sashidhara KV, Kumar A, Chatterjee M, Rao KB, Singh S, Verma AK, et al. Discovery and synthesis of novel 3-phenylcoumarin derivatives as antidepressant agents. *Bioorg Med Chem Lett* 2011;21(7):1937-41. doi: 10.1016/j.bmcl.2011.02.040
15. Wattenberg LW, Lam LK, Fladmoe AV. Inhibition of chemical carcinogen-induced neoplasia by coumarins and alpha-angelicalactone. *Cancer Res* 1979;39(5):1651-4.
16. Kashman Y, Gustafson KR, Fuller RW, Cardellina JH 2nd, McMahon JB, Currens MJ, et al. The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical rainforest tree, *Calophyllum lanigerum*. *J Med Chem* 1992;35(15):2735-43. doi: 10.1021/jm00093a004
17. Hamdy AM, Khaddour Z, Al-Masoudi NA, Rahman Q, Hering-Junghans C, Villingner A, et al. Synthesis of arylated coumarins by Suzuki-Miyaura cross-coupling. Reactions and anti-HIV activity. *Bioorg Med Chem* 2016;24(21):5115-26. doi: 10.1016/j.bmc.2016.08.029
18. Peng XM, Damu GL, Zhou C. Current developments of coumarin compounds in medicinal chemistry. *Curr Pharm Des* 2013;19(21):3884-930. doi: 10.2174/1381612811319210013
19. Singh RK, Lange TS, Kim KK, Brard L. A coumarin derivative (RKS262) inhibits cell-cycle progression, causes pro-apoptotic signaling and cytotoxicity in ovarian cancer cells. *Invest New Drugs* 2011;29(1):63-72. doi: 10.1007/s10637-009-9335-4
20. Musa MA, Cooperwood JS, Khan MO. A review of coumarin derivatives in pharmacotherapy of breast cancer. *Curr Med Chem* 2008;15(26):2664-79. doi: 10.2174/092986708786242877
21. Thomas N, Zachariah SM, Ramani P. 4-Aryl-4H-chromene-3-carbonitrile derivatives: synthesis and preliminary anti-breast cancer studies. *J Heterocycl Chem* 2016;53(6):1778-82. doi: 10.1002/jhet.2483
22. Bhattacharyya SS, Paul S, Mandal SK, Banerjee A, Boujedaini N, Khuda-Bukhsh AR. A synthetic coumarin (4-methyl-7-hydroxy coumarin) has anti-cancer potentials against DMBA-induced skin cancer in mice. *Eur J Pharmacol* 2009;614(1-3):128-36. doi: 10.1016/j.ejphar.2009.04.015
23. Harada K, Kubo H, Tomigahara Y, Nishioka K, Takahashi J, Momose M, et al. Coumarins as novel 17beta-hydroxysteroid dehydrogenase type 3 inhibitors for potential treatment of prostate cancer. *Bioorg Med Chem Lett* 2010;20(1):272-5. doi: 10.1016/j.bmcl.2009.10.111
24. Zhang L, Jiang G, Yao F, He Y, Liang G, Zhang Y, et al. Growth inhibition and apoptosis induced by osthole, a natural coumarin, in hepatocellular carcinoma. *PLoS One* 2012;7(5):e37865. doi: 10.1371/journal.pone.0037865
25. Neelgundmath M, Dinesh KR, Mohan CD, Li F, Dai X, Siveen KS, et al. Novel synthetic coumarins that targets NF-kappaB in Hepatocellular carcinoma. *Bioorg Med Chem Lett* 2015;25(4):893-7. doi: 10.1016/j.bmcl.2014.12.065
26. Jun M, Bacay AF, Moyer J, Webb A, Carrico-Moniz D. Synthesis and biological evaluation of isoprenylated coumarins as potential anti-pancreatic cancer agents. *Bioorg Med Chem Lett* 2014;24(19):4654-8. doi: 10.1016/j.bmcl.2014.08.038
27. Farley CM, Dibwe DF, Ueda JY, Hall EA, Awale S, Magolan J. Evaluation of synthetic coumarins for antiausterity cytotoxicity against pancreatic cancers. *Bioorg Med Chem Lett* 2016;26(5):1471-4. doi: 10.1016/j.bmcl.2016.01.054
28. Kapetanovic IM. Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chem Biol Interact* 2008;171(2):165-76. doi: 10.1016/j.cbi.2006.12.006
29. Karcher W, Devillers J. *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*. Springer Science & Business Media; 1990.
30. Cohen NC. *Guidebook on Molecular Modeling in Drug Design*. Academic Press; 1996.
31. Dastmalchi S, Hamzeh-Mivehroud M, Sokouti B. *Quantitative Structure-Activity Relationship: A Practical Approach*. 1st ed. Boca Raton: CRC Press; 2018.
32. Roy K. On some aspects of validation of predictive quantitative structure-activity relationship models. *Expert Opin Drug Discov* 2007;2(12):1567-77. doi: 10.1517/17460441.2.12.1567
33. Cao D, Liu Y, Yan W, Wang C, Bai P, Wang T, et al. Design,

- synthesis, and evaluation of in vitro and in vivo anticancer activity of 4-substituted coumarins: a novel class of potent tubulin polymerization inhibitors. *J Med Chem* 2016;59(12):5721-39. doi: 10.1021/acs.jmedchem.6b00158
34. Dennington RD, Keith TA, Millam JM. GaussView 5. 0. 8. Gaussian Inc; 2008.
  35. Allinger NL. Conformational analysis. 130. Mm2. A hydrocarbon force field utilizing v1 and v2 torsional terms. *J Am Chem Soc* 1977;99(25):8127-34. doi: 10.1021/ja00467a001
  36. Dewar MJ, Thiel W. Ground states of molecules. 39. Mndo results for molecules containing hydrogen, carbon, nitrogen, and oxygen. *J Am Chem Soc* 1977;99(15):4907-17. doi: 10.1021/ja00457a005
  37. HyperChem(TM) Professional 7.51. Gainesville, Florida, USA: Hypercube, Inc; 2002. <http://www.hyper.com/>.
  38. Todeschini R, Consonni V. Dragon, rel. 1.12 for Windows.
  39. ACD/Labs, version 2018.1. Toronto, ON, Canada: Advanced Chemistry Development, Inc; 2019. <https://www.acdlabs.com/>.
  40. Todeschini R, Consonni V. Handbook of Molecular descriptor. Weinheim, Germany: Wiley-VCH; 2000.
  41. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.
  42. Kennard RW, Stone LA. Computer aided design of experiments. *Technometrics* 1969;11(1):137-48. doi: 10.1080/00401706.1969.10490666
  43. Martin TM, Harten P, Young DM, Muratov EN, Golbraikh A, Zhu H, et al. Does rational selection of training and test sets improve the outcome of QSAR modeling? *J Chem Inf Model* 2012;52(10):2570-8. doi: 10.1021/ci300338w
  44. Roy K, Mitra I. On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design. *Comb Chem High Throughput Screen* 2011;14(6):450-74. doi: 10.2174/138620711795767893
  45. Leardi R, Lupiáñez González A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. *Chemometr Intell Lab Syst* 1998;41(2):195-207. doi: 10.1016/S0169-7439(98)00051-3
  46. Daren Z. QSPR studies of PCBs by the combination of genetic algorithms and PLS analysis. *Comput Chem* 2001;25(2):197-204. doi: 10.1016/S0097-8485(00)00081-4
  47. Ghafourian T, Cronin MT. The impact of variable selection on the modelling of oestrogenicity. *SAR QSAR Environ Res* 2005;16(1-2):171-90. doi: 10.1080/10629360412331319808
  48. Hemmer MC, Steinhauer V, Gasteiger J. Deriving the 3D structure of organic molecules from their infrared spectra. *Vib Spectrosc* 1999;19(1):151-64. doi: 10.1016/S0924-2031(99)00014-4
  49. Mahani M, ShaikhGhomi H. Comparison of multiple linear regression, partial least squares and artificial neural network for quantitative structure retention relationships of some polycyclic aromatic hydrocarbons. *Anal Methods* 2012;4(10):3381-5. doi: 10.1039/C2AY25711D
  50. Masand VH, Mahajan DT, Ben Hadda T, Jawarkar RD, Alafeefy AM, Rastija V, et al. Does tautomerism influence the outcome of QSAR modeling? *Med Chem Res* 2014;23(4):1742-57. doi: 10.1007/s00044-013-0776-0
  51. Lovász L, Pelikán J. On the eigenvalues of trees. *Period Math Hung* 1973;3(1-2):175-82. doi: 10.1007/bf02018473
  52. Du H, Hu Z, Bazzoli A, Zhang Y. Prediction of inhibitory activity of epidermal growth factor receptor inhibitors using grid search-projection pursuit regression method. *PLoS One* 2011;6(7):e22367. doi: 10.1371/journal.pone.0022367
  53. Estrada E. Edge adjacency relationships and a novel topological index related to molecular volume. *J Chem Inf Comput Sci* 1995;35(1):31-3. doi: 10.1021/ci00023a004
  54. Zhan S, Huang J, Shao Q, Fan X, Guo W. Prediction of microdialysis relative recovery of flavone derivatives based on molecular descriptors. *J Braz Chem Soc* 2012;23(11):2035-42. doi: 10.1590/S0103-50532012005000074
  55. Schuur JH, Selzer P, Gasteiger J. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *J Chem Inf Comput Sci* 1996;36(2):334-44. doi: 10.1021/ci950164c
  56. Schuur J, Gasteiger J. 3d-morse code-a new method for coding the 3d structure of molecules. *Software Dev Chem* 1996;10:67-80.
  57. Verma RP, Hansch C. QSAR modeling of taxane analogues against colon cancer. *Eur J Med Chem* 2010;45(4):1470-7. doi: 10.1016/j.ejmech.2009.12.054
  58. Roy PP, Roy K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci* 2008;27(3):302-13. doi: 10.1002/qsar.200710043
  59. Veerasamy R, Rajak H, Jain A, Sivasadan S, Varghese CP, Agrawal RK. Validation of QSAR models-strategies and importance. *Int J Drug Des Discov* 2011;2(3):511-9.
  60. Ghandadi M, Shayanfar A, Hamzeh-Mivehroud M, Jouyban A. Quantitative structure activity relationship and docking studies of imidazole-based derivatives as P-glycoprotein inhibitors. *Med Chem Res* 2014;23(11):4700-12. doi: 10.1007/s00044-014-1029-6
  61. Ojha PK, Mitra I, Das RN, Roy K. Further exploring rm2 metrics for validation of QSPR models. *Chemometr Intell Lab Syst* 2011;107(1):194-205. doi: 10.1016/j.chemolab.2011.03.011
  62. Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN. Some case studies on application of "rm2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J Comput Chem* 2013;34(12):1071-82. doi: 10.1002/jcc.23231