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

Leila Dinparast1, Siavoush Dastmalchi2,3,4,*

1Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.
2School of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran.
3Pharmaceutical Analysis Research Center, Tabriz University of Medical Sciences, Tabriz, Iran.
4Faculty of Pharmacy, Near East University, POBOX: 99138, Nicosia, North Cyprus, Mersin 10, Turkey.

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. 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. 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. Most of the clinically used drugs have limited effectiveness and selectivity, accompanied with serous toxicity, and unacceptable side effects. Moreover, the most common tumors show resistance against the significant number of commercially available anticancer drugs. 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.

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, anti-inflammatory, antioxidant, antiviral, antimicrobial, antidepressants, and anti-HIV effects. Also, they are promising compounds due to their low toxicity, little drug resistance, less side effects, high bioavailability, and ease of chemical synthesis. Several studies have shown that coumarins are potential anticancer agents having growth suppressive effects on many types of cancers such as ovarian, breast, skin, prostate, liver, and pancreatic.

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. 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. 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
cost spent for drug design and discovery are significantly decreased by using various QSAR techniques.\(^2\)

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.\(^3\) 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.,\(^4\) 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 GuassView 5.0 software.\(^5\) The energy minimization of the structures were conducted initially using the empirical method (i.e., MM+).\(^6\) followed by semi-empirical technique AM1 using the Polak-Ribiere algorithm included in HyperChem 7.5 software.\(^7\) The molecular descriptors for the fully optimized molecular structures were calculated using Dragon (version 3.0) program.\(^8\) Log P and log D were calculated by ACD/Labs 6.0 program\(^9\) 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.\(^10\) These pretreatments on the descriptors were performed using R 3.2.3 software.\(^11\)

**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.\(^12,13\) 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. \(r^2\), \(r^2_{\text{Adj}}\), \(Q^2\), \(r^2_{\text{RMSE}}\), and \(\Delta r^2_{\text{RMSE}}\) with their default values set to > 0.6, > 0.6, > 0.6, > 0.5, and < 0.2, respectively.\(^14\) 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:

\[
Q^2_{\text{(test)}} = 1 - \frac{\text{PRESS}}{\sum_{i=1}^{N} (y_{\text{obs,i}} - y_{\text{pred,i}})^2}
\]

\[
\text{PRESS} = \sum_{i=1}^{N} (y_{\text{pred,i}} - y_{\text{obs,i}})^2
\]

\[
\text{APE} = \left(\frac{\text{pIC}_{50(\text{pred})} - \text{pIC}_{50(\text{obs})}}{\text{pIC}_{50(\text{obs})}}\right) \times 100
\]

\[
\text{MAPE} = \frac{\sum_{i=1}^{N} \text{APE}}{N}
\]

\[
\text{SDEP} = \sqrt{\frac{\sum_{i=1}^{N} (y_{\text{obs,i}} - y_{\text{pred,i}})^2}{N}}
\]

Where, \(y_{\text{obs,i}}\), \(\text{pIC}_{50(\text{obs})}\), \(y_{\text{med,i}}\), and \(\text{pIC}_{50(\text{med})}\) are the experimental and predicted activities of an individual compound in the test set, respectively. N is the number of molecules and \(y_{\text{med}}\) 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):

\[
F = \sum_{i=0}^{l} \frac{P_{c,i} - P_{i}}{P_{\text{Max}} - P_{\text{Min}}}
\]

**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.\(^45\)\(^46\)\(^47\) 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$^1$ | R$^2$ | pIC$_{50}$ (Exp.) | pIC$_{50}$ (Pred.) | APE (Eq. 6) | APE (Eq. 7) |
|----------|-------|-------|------------------|------------------|-------------|-------------|
| 1        | OCH$_3$ | OH    | 6.406            | 6.340            | 1.021       | 0.431       |
| 2'       | H      | O-C(O)CH$_3$ | 7.833            | 8.054            | 4.818       | 2.529       |
| 3        | H      | O-C(O)CH$_3$Cl | 7.504            | 7.866            | 1.657       | 0.724       |
| 4        | H      | O-C(O)CH$_3$Cl | 8.167            | 8.032            | 3.726       | 4.489       |
| 5        | H      | O-C(O)CH$_3$Cl | 8.119            | 7.817            | 3.254       | 3.073       |
| 6        | H      | O-C(O)CH$_3$Br | 7.963            | 7.703            | 0.051       | 0.556       |
| 7'       | H      | O-C(O)CH$_3$CH$_3$ | 7.889            | 7.615            | 9.305       | 9.618       |
| 8        | H      | O-C(O)CH$_3$CH$_3$ | 7.456            | 7.452            | 0.056       | 0.875       |
| 9        | H      | O-C(O)CH$_3$CH$_3$ | 7.221            | 7.893            | 0.057       | 0.344       |
| 10       | H      | O-C(O)CH$_3$CH$_3$ | 7.676            | 7.680            | 0.076       | 0.857       |
| 11       | H      | O-C(O)CH$_3$CH$_3$ | 7.845            | 7.849            | 0.092       | 0.027       |
| 12       | H      | O-C(O)CH$_3$CH$_3$ | 7.407            | 7.684            | 0.775       | 1.126       |
| 13'      | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.168            | 7.690            | 0.114       | 2.433       |
| 14       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.087            | 7.306            | 0.198       | 2.809       |
| 15       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 6.612            | 6.560            | 5.313       | 4.060       |
| 16       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.565            | 7.557            | 5.067       | 3.592       |
| 17       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.442            | 7.428            | 0.713       | 2.529       |
| 18'      | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.140            | 7.209            | 1.088       | 1.399       |
| 19'      | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.629            | 7.509            | 9.202       | 7.212       |
| 20       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 8.056            | 7.628            | 0.486       | 5.691       |
| 21       | H      | O-C(O)CH$_3$C(O)(O)CH$_3$ | 7.991            | 7.562            | 12.410      | 8.225       |
| 22       | H      | NH$_3$ | 8.481            | 8.491            | 2.825       | 2.086       |
| 23       | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 7.529            | 7.611            | 3.482       | 6.674       |
| 24       | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 8.018            | 7.280            | 7.279       | 2.640       |
| 25'      | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 7.318            | 7.196            | 0.958       | 0.034       |
| 26       | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 6.543            | 6.574            | 1.569       | 5.705       |
| 27'      | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 7.564            | 7.703            | 1.659       | 1.358       |
| 28'      | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 7.676            | 7.596            | 1.452       | 0.591       |
| 29'      | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 7.582            | 7.541            | 1.033       | 3.408       |
| 30       | H      | HN-C(O)CH$_3$C(O)(O)CH$_3$ | 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            |

Advanced Pharmaceutical Bulletin, 2020, Volume 10, Issue 2 | 273
Table 3. The statistical parameters of developed models

| Models | N  | \( r^2 \) | \( r^2_{\text{Adj impost}} \) | \( q^2_{\text{LOO}} \) | \( r^2 - q^2_{\text{LOO}} \) | SEE | SDEP | \( \overline{PM}^2 \) | \( \Delta PM^2 \) | \( P \text{ value} \) | MAPE (Train set) | MAPE (Test set) |
|--------|----|----------|----------------|----------------|----------------|-----|-----|------------|------------|----------------|----------------|----------------|
| Eq. 6  | 31 | 0.669   | 0.634           | 0.483          | 0.206          | 0.670 | 0.376 | 0.437     | 0.378     | 0.994          | \( 8.08 \times 10^{-4} \) | 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          | \( 1.89 \times 10^{-4} \) | 0.152           | 0.298          |

The predictive power of the developed models was evaluated using internal and external validation measurements. For this purpose, the squared correlation coefficient \((r^2)\), leave one out cross-validated correlation coefficient \((q^2_{\text{LOO}})\), the squared adjusted correlation coefficient \((r^2_{\text{Adjusted}})\), the standard error of estimate (SEE), the SDEP, \(\overline{PM}^2\), and \(\Delta PM^2\) were calculated for the train set and \(q^2_{\text{LOO}}\) was computed for the test set (Table 3). The squared correlation coefficient is the parameter fitted on the whole train set and the QSR models with \( r^2 > 0.6 \) are considered reliable.\(^{57}\) As seen in Table 3, \( r^2 \) values of 0.689 and 0.749 were obtained for equations 6 and 7, respectively. The \( q^2_{\text{LOO}} \) and \( r^2 - q^2_{\text{LOO}} \) are other measurement criteria for evaluating the performance of QSR models, which should be higher than 0.5 and 0.3,
respectively. 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 $R^2$ and $\Delta R^2$ were determined to further assess the predictive ability of the QSAR models. $R^2$ metric which was introduced by Roy and Roy determines the proximity between the observed and predicted activities for the data set. It has been suggested that for the models with reliable predictive power, the values of $R^2$ and $\Delta R^2$ should be more than 0.5 and lower than 0.2, respectively. The obtained $R^2$ and $\Delta R^2$ values for equations 6 and 7 are 0.378, 0.417 and 0.094, 0.138, respectively. $\Delta R^2$ for both of models are in the acceptable range but $R^2$ 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.
Dinparast and Dastmalchi

1. Rahmani-Nezhad S, Safari 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

2. Gottesman MM. Mechanisms of cancer drug resistance. *Annu Rev Med* 2002;53:615-27. doi: 10.1146/annurev.med.53.082901.103929

3. 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

4. Weigt S, Huebler N, Strecker R, Braunbeck T, Broschard TH. Developmental effects of coumarin and the antimycobacterial coumarin derivative warfarin on zebrafish (Danio rerio) embryos. *Reprod Toxicol* 2012;33(2):133-41. doi: 10.1016/j.reprotox.2011.07.001

5. 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: 10.2174/138161204382710

6. Bansal Y, Sethi P, Bansal G. Coumarin: a potential nucleus for anti-inflammatoxy molecules. *Med Chem Res* 2013;22(7):3049-60. doi: 10.1007/s00044-012-0321-6

7. Iranshahi M, Askari M, Sabekbar A, Adjipavlou-Litina D. Evaluation of antioxidant, anti-inflammatory and lipoxygenase inhibitory activities of the prenylated coumarin umbelliprenin. *Daruu* 2009;17(2):99-103.

8. 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

9. 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 Chemotherapy* 2007;51(10):3688-98. doi: 10.1128/aac.00392-11

10. 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. *Eur J Med Chem* 2015;24(12):4128-38. doi: 10.1007/s00044-015-1449-y

11. 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

12. Watteenberg JW, Lam LK, Fladmoe AV. Inhibition of chemical carcinoigen-induced neoplasia by coumarins and alpha-angelicalactone. *Cancer Res* 1979;39(5):1651-4.

13. Kashman Y, Gustafson KR, Fuller RW, Cardella JH 2nd, McMahon JB, Currens MJ, et al. The calanolides, a novel HIV-inhibitory class of coumarin derivatives from the tropical forest tree, *Calophyllum lanigerum*. *J Med Chem* 1992;35(15):2735-43. doi: 10.1021/jm00093a004

14. Hamdy AM, Khaddour Z, Al-Masoudi NA, Rahman Q, Hering-Junghans C, Villinger 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.bmcl.2016.08.029

15. 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

16. 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

17. 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: 10.2174/0929867087862877

18. Thomas N, Zachariah SM, Ramani P. 4-Aryl-4H-chromene-3-carbonitrile derivates: synthesis and preliminary anti-breast cancer studies. *J Heterocycl Chem* 2016;53(6):1778-82. doi: 10.1002/jhet.2483

19. Bhattacharyya SS, Patil S, Mandal SK, Banerjee A, Bosjedaini N, Khuda-Bukhsh AR. A synthetic coumarin (4-methyl-7-hydroxy coumarin) has anti-cancer potential against DMBA-induced skin cancer in mice. *Eur J Pharmacol* 2009;614(1-3):129-36. doi: 10.1016/j.ejphar.2009.04.015

20. 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

21. 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):e37665. doi: 10.1371/journal.pone.0037685

22. 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

23. Jun M, Bacy AE, 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):4656-8. doi: 10.1016/j.bmcl.2014.08.038

24. Farley CM, Dilwe DF, Ueda JY, Hall EA, Awale S, Magolan R, et al. Evaluation of synthetic coumarins for antiastuseria cytotoxicity against pancreatic cancers. *Bioorg Med Chem Lett* 2016;26(5):1471-4. doi: 10.1016/j.bmcl.2016.01.054

25. 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

26. Karcher W, Devillers J. Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology. Springer Science & Business Media; 1990.

27. Cohen NC. *Guidebook on Molecular Modeling in Drug Design*. Academic Press; 1996.

28. Dastmalchi S, Hamzeh-Mivehroud M, Sokouti B. *Quantitative Structure–Activity Relationship: A Practical Approach*. 1st ed. Boca Raton: CRC Press; 2018.

29. 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

30. 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 hydrogen 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 J, 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: 10.2174/1386207117795767893

45. Leardi R, Lupiñané González A. Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemom Inform Technol 1998;41(2):195-207. doi: 10.1016/S0106-7439(98)00051-3

46. Daren Z. QSAR studies of PCBs by the combination of genetic algorithms and PLS analysis. Comput Chem 2001;25(2):197-204. doi: 10.1016/s0010-9405(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: 10.1080/1069293041233131808

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)00041-4

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, Sivadasan 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/s00444-014-1029-6

61. Olja PK, Mitra I, Das RN, Roy K. Further exploring rm2 metrics for validation of QSAR models. Chemom Inform Technol Lab Syst 2011;107(1):194-205. doi: 10.1016/j.chemolab.2011.03.011

62. Roy K, Chakraborty P, Mitra I, Olja 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