

RESEARCH

Open Access



In-silico activity prediction and docking studies of some flavonol derivatives as anti-prostate cancer agents based on Monte Carlo optimization

Faezeh Tajjani¹, Shahin Ahmadi^{2*}, Shahram Lotfi³, Parvin Kumar⁴ and Ali Almasirad¹

Abstract

The QSAR models are employed to predict the anti-proliferative activity of 81 derivatives of flavonol against prostate cancer using the Monte Carlo algorithm based on the index of ideality of correlation (IIC) criterion. CORAL software is employed to design the QSAR models. The molecular structures of flavonols are demonstrated using the simplified molecular input line entry system (SMILES) notation. The models are developed with the hybrid optimal descriptors i.e. using both SMILES and hydrogen-suppressed molecular graph (HSG). The QSAR model developed for split 3 is selected as a prominent model ($R_{validation}^2 = 0.727$, $IIC_{validation} = 0.628$, $Q_{validation}^2 = 0.642$, and $\bar{r}_m^2 = 0.615$). The model is interpreted mechanistically by identifying the characteristics responsible for the promoter of the increase or decrease. The structural attributes as promoters of increase of pIC_{50} were aliphatic carbon atom connected to double-bond (C...=...), aliphatic oxygen atom connected to aliphatic carbon (O...C...), branching on aromatic ring (c...(...)), and aliphatic nitrogen (N...). The pIC_{50} of eight natural flavonols with pIC_{50} more than 4.0, were predicted by the best model. The molecular docking is also performed for natural flavonols on the PC-3 cell line using the protein (PDB: 3RUK).

Keywords Flavonoids compounds, Prostate cancer, QSAR, CORAL software, Molecular docking

Introduction

Flavonoids are a class of polyphenolic compounds which possess a phenyl benzopyrone structure (C6–C3–C6) and are present in all vascular plants. These are produced

as secondary plant metabolites, which are known to demonstrate broad-spectrum pharmacological activities, but the human body is unable to produce them [1–3]. These compounds according to saturation level subdivided into flavanols, flavonols, flavones, flavanones, isoflavones, flavanonols, and chalcones [4, 5].

The CYP17A1 has an important role in the biosynthesis of dehydroepiandrosterone (DHEA) as the precursor of androgens and overexpression of this enzyme can cause prostate cancer. Abiraterone as an approved anti-prostate cancer drug is a CYP17A1 inhibitor [6, 7]. Flavonols are characterized by a hydroxyl group present at C-3 of the flavone skeleton and there are some reports about the CYP17A1 inhibitory activity of flavonoids like rutin,

*Correspondence:

Shahin Ahmadi

ahmadi.chemometrics@gmail.com; sahmadi@iautmu.ac.ir

¹ Department of Medicinal Chemistry, Faculty of Pharmacy, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran

² Department of Pharmaceutical Chemistry, Faculty of Pharmaceutical Chemistry, Tehran Medical Sciences, Islamic Azad University, Tehran, Iran

³ Department of Chemistry, Payame Noor University (PNU), Tehran 19395-4697, Iran

⁴ Department of Chemistry, Kurukshetra University, Kurukshetra, Haryana 136119, India



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

morusflavone, quercetin, kaempferol and isorhamnetin [8–10].

These have also been attracted by medicinal chemists because of their effective anti-prostate cancer properties. Prostate cancer is the most common type of diagnosed cancer among males worldwide with the incidence of 28 cases per 100,000 and mortality being 7 per 100,000 [11–13]. Normal growth and maintenance of the prostate is dependent on androgen hormones that act through the androgen receptor. Activation of the androgen receptor drives the development of prostate cancer. It has been reported that the agents such as flavonols that down-regulate androgen receptors can inhibit the development of prostate cancer cells [14–16].

The influence of chemical structures of flavonols over their anticancer activities has been investigated experimentally and shown that structural modification can further increase its anti-cancer activity and ability to activate PC-3 cell apoptosis. However, the structure–activity relationship for flavonols as anti-prostate cancer agents has captured attention by quantitatively correlating the molecular structures or properties with variation in pharmacological activity [17, 18].

The anti-prostate cancer activity is expressed typically with IC_{50} (half maximal inhibitory concentration) values. Quantitative structure–activity relationships (QSARs) are a powerful tool to predict IC_{50} of flavonoids in general. Already, no study has been reported on QSAR modeling for predicting the IC_{50} of flavonols against prostate cancer.

QSAR model is a mathematical equation which is widely employed to estimate and predict pharmacological activity or physical, chemical properties/activities of chemicals using descriptors derived from chemical structure [19–22]. The CORAL (Correlation and Logic) freeware software is employed for designing the Quantitative structure–activity/activity relationships (QSPRs/QSARs) models in compliance with OECD principles [23–26]. In CORAL software, the SMILES notations of the molecular structure are used as an input file and produce the best model based on Monte Carlo optimization [27–30]. It can be applied to compute the optimal descriptor by using solely SMILES or molecular graph-based descriptor or a combination of both descriptors (so-called hybrid descriptor). A literature survey reveals that the index of ideality of correlation (IIC) parameter of CORAL software can be employed to build robust QSAR models [31–34].

Molecular docking simulation is a computational methodology that purveys automatic tools to measure the conformation of a protein–ligand complex. The aim of molecular docking is to regulate the position of the ligand in the protein. An energy-based scoring function

is commonly used in docking procedures to find the energetically most advantageous ligand conformation when attached to the target. Intermittently, the Monte Carlo computational methodologies are also applied in molecular docking simulation [35, 36].

Since ancient times various natural products have been used as traditional medicine against various human diseases. Moreover, natural products are easily applicable, cheap, accessible and acceptable treatment method with minimum cytotoxicity [37]. As a results of QSAR modeling, the pIC_{50} activity of some natural flavonols as anti-proliferative agents were predicted and reported.

The goal of this report is to devise reliable first QSAR models utilizing CORAL software to predict pIC_{50} of 81 flavonols against prostate cancer. In the development of QSAR models, a hybrid optimal descriptor, a combination of SMILES and hydrogen suppressed graph (HSG), is employed. The index of ideality of correlation (IIC) is used to improve the predictive potential of QSAR models. Further, the pIC_{50} is also calculated for a series of eight natural flavonols using the QSAR models of all splits. As mentioned above flavonols show their anti-prostate cancer activity through different mechanism of actions. However, molecular docking is also performed for eight natural flavonol derivatives in order to evaluate their potential affinity to CYP17A1 (PDB: 3RUK).

Methods

Data

Experimental data on anti-prostate cancer (PC-3) activities of 86 flavonols were taken from the four literature reports (Additional file 1: Table S1) [11, 38–40]. The numerical values of activity were converted to a negative logarithmic scale, pIC_{50} ($-\log IC_{50}$) (Molar) for QSAR modelling. The range of pIC_{50} for PC-3 cell line was from 3.39 to 6.28. The current dataset was not previously used for QSAR modeling. The molecular structures of the flavonol derivatives were sketched by BIOVIADraw 2019 and transferred to the SMILES code for modeling with the CORAL software. Three splits were made from the dataset and each split was further randomly divided into four sets i.e., training ($\approx 35\%$), invisible training ($\approx 25\%$), calibration ($\approx 15\%$), and validation ($\approx 25\%$) sets. In CORAL-based QSAR modeling, each set was assigned its specific accountability. The task of the training set (TRN) was to compute correlation weights and the task of the invisible training set (iTRN) was to control the adaptability of the data which were not employed in the training set. The assignment of the calibration set (CAL) was to detect the overtraining whereas the final estimation of the predictive potential of the designed QSAR model was assigned to the validation set (VAL) [34, 41].

Hybrid optimal descriptor

Herrin, the optimal hybrid optimal descriptor based on SMILES and HSG was employed to create QSAR models for pIC_{50} of flavonol compounds. The literature reports showed that the QSPR models produced through the 'hybrid' optimal descriptor had better statistical parameters than the model designed by individually SMILES or HSG descriptors [42, 43].

The QSAR model employed to predict pIC_{50} of flavonol derivatives is demonstrated in the following equation:

$$pIC_{50} = C_0 + C_1 \times \text{Hybrid DCW}(T^*, N^*). \quad (1)$$

Here, C_0 is the regression coefficient and C_1 is the slope computed by the least-squares method; DCW (descriptor

of correlation weights) is computed with correlation weights of molecular features extracted from HSG and SMILES notations. The following equation is employed to compute DCW:

$$DCW(T^*, N^*) = \sum CW(A_K), \quad (2)$$

where A_K is an attribute of SMILES or HSG, the T^* and N^* define the threshold value and number of epochs of the Monte Carlo optimization, respectively.

$$\text{Hybrid DCW}(T^*, N^*) = \text{SMILES DCW}(T, N^*) + \text{Graph DCW}(T^*, N^*). \quad (3)$$

The DCW of HSG and SMILES employed here are illustrated as Eqs. (4) and (5):

$$\begin{aligned} \text{SMILES DCW}(T, N) = & \sum CW(S_k) + \sum CW(SS_k) + CW(\text{BOND}) + CW(\text{NOSP}) + CW(\text{HARD}) + CW(\text{PAIR}) \\ & + CW(C_{\max}) + CW(N_{\max}) + CW(O_{\max}) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{HSG DCW}(T, N) = & \sum CW(e1_k) + \sum CW(e2_k) \\ & + \sum CW(e1_k + e2_k) + \sum CW(|e1_k - e2_k|) \\ & + \sum CW(pt2_k) + \sum CW(pt3_k) \\ & + \sum CW(pt2_k + pt3_k) + \sum CW(|pt2_k - pt3_k|) \\ & + \sum CW(S2_k) + \sum CW(S3_k) + \sum CW(S2_k + S3_k) \\ & + \sum CW(|S2_k - S3_k|) + CW(C5) + CW(C6) \end{aligned} \quad (5)$$

Table 1 The detailed description of SMILES attributes and graph invariants for constructed models of pIC_{50}

	ID	Definition
SMILES attribute	Sk	SMILES atom, <i>i.e.</i> , one symbol (e.g. 'C', 'N', '=', etc.) or a group of symbols that cannot be examined separately (e.g., 'Cl', 'Br', 'Si', etc.)
	SSk	a mixture of two SMILES-atoms
	BOND	Presence or absence of chemical bonds: double (=), triple (#), and stereochemical (@) or @@
	PAIR	Association two of BOND, NOSP, and HALO
	HARD	Association of BOND, NOSP, and HALO in the united structural code
	NOSP	Presence or absence of different chemical elements: nitrogen (N), oxygen (O), sulfur (S), and phosphorus (P);
	C_{\max}	Maximum number of rings
	N_{\max}	Maximum number of nitrogen atoms in a molecule
	O_{\max}	Maximum number of oxygen atoms in a molecule structure
Graph invariant	e2k	Morgan extended connectivity of first order
	e3k	Morgan extended connectivity of second-order
	pt2k	Number of paths of lengths 2 and 3 starting from a given vertex in the graph
	pt3k	Number of paths of length 3 starting from a given vertex in the graph
	$S2_k$	Valence shells of the second orders
	$S3_k$	Valence shells of the third orders
	C5 and C6	Codes of rings (five-member and six-member rings, with the data on the presence or absence of heteroatoms, aromaticity, and the total number of given rings in the molecule)

The SMILES attributes and HSG invariant applied in Eqs. (4) and (5) are depicted in Table 1.

A flowchart of a Monte Carlo optimization cycle is presented by Sokolovic et al. [44]. At first cycle, the CW(x) of features is randomly generated and then optimized based on the proposed objective function. Herein, two kinds of target functions consisting of the balance of correlation without IIC (TF1) and the balance of correlation with IIC (TF2) are studied.

The following mathematical equation is employed to compute the TF₁ and TF₂:

$$TF_1 = R_{TRN} + R_{iTRN} - |R_{TRN} - R_{iTRN}| \times Const \quad (6)$$

$$TF_2 = TF_1 + IIC_{CAL} \times Const \quad (7)$$

The $R_{training}$ and $R_{invTraining}$ are the correlation coefficients for the training and invisible training sets, respectively. The empirical constant (Const) is usually fixed [45, 46].

The IIC_{CAL} is calculated with data on the calibration (CAL) set as the following:

$$IIC = R_{CAL} \times \frac{\min(-MAE_{CAL}, +MAE_{CAL})}{\max(-MAE_{CAL}, +MAE_{CAL})} \quad (8)$$

R_{CAL} is the correlation coefficient for the calibration set. The negative and positive mean absolute errors are shown with ^-MAE and ^+MAE , which are computed using the following equations:

$$^-MAE_{CAL} = -\frac{1}{N^-} \sum_{y=1}^{N^-} |\Delta_k| \quad \Delta_k < 0, \quad N^- \text{ is the number of } \Delta_k < 0 \quad (9)$$

$$^+MAE_{CAL} = +\frac{1}{N^+} \sum_{y=1}^{N^+} |\Delta_k| \quad \Delta_k \geq 0, \quad N^+ \text{ is the number of } \Delta_k \geq 0 \quad (10)$$

$$\Delta_k = \text{Observed}_k - \text{Calculated}_k \quad (11)$$

The 'k' is the index (1, 2,...N). The observed_k and calculated_k are related to numerical values of the endpoint.

This IIC is obtained by using the correlation coefficient between the observed and predicted values of the endpoint for the calibration set, taking into account the positive and negative dispersions between the observed and calculated values [47].

Applicability domain

The applicability domain (AD) is another key guideline that should be included in a built QSPR/QSAR model. It was defined by the OECD as "the response and chemical structure space in which the model produces predictions with a specified reliability" [48, 49]. The CORAL-based QSAR model computes AD based on the dispersion of SMILES features in the training and calibration sets [50]. The AD is defined as 'Defect_{A_k}', which was computed with the following equation:

$$\text{Defect}_{A_k} = \frac{|P_{TRN}(A_k) - P_{CAL}(A_k)|}{N_{TRN}(A_k) + N_{CAL}(A_k)} \quad \text{If } A_k > 0$$

$$\text{Defect}_{A_k} = 1 \quad \text{If } A_k = 0 \quad (12)$$

$P_{TRN}(A_k)$ and $P_{CAL}(A_k)$ are the probability of an attribute 'A_k' in the training and the calibration sets; $N_{TRN}(A_k)$ and $N_{CAL}(A_k)$ are the number of times of A_k in the training and calibration sets, respectively.

The statistical defect is computed using the following equation:

$$\text{Defect}_{Molecule} = \sum_{k=1}^{N_A} \text{Defect}_{A_k} \quad (13)$$

N_A is the number of active SMILES attributes for the given compounds.

In CORAL, a substance is an outlier if inequality 14 is fulfilled:

$$\text{Defect}_{molecule} > 2 \times \overline{\text{Defect}_{TRN}} \quad (14)$$

$\overline{\text{Defect}_{TRN}}$ is an average of statistical defect for the dataset of the training set.

Validation of the model

It is most important to validate the predictive potential of a constructed QSAR model. In the present manuscript,

Table 2 The mathematical equation of different statistical benchmark of the predictive potential for CORAL models

Criterion of the predictive potential	Comments	Refs.
$R^2 = 1 - \frac{\sum (Y_{obs} - Y_{prd})^2}{\sum (Y_{obs} - \bar{Y}_{train})^2}$	Y_{obs} is the observed endpoint for the training set, and Y_{prd} is the predicted endpoint values for the training set of compounds \bar{Y}_{train} is the mean observed endpoint of the training set	
$Q^2 = 1 - \frac{\sum (Y_{prd(train)} - Y_{obs(train)})^2}{\sum (Y_{obs(train)} - \bar{Y}_{train})^2}$	$Y_{obs(train)}$ is the observed endpoint, and $Y_{prd(train)}$ is the predicted response of the training set compounds	
$CCC = \frac{2 \sum (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2 + n(\bar{x} - \bar{y})^2}$	n is the number of compounds, and x_i and y_i denote the mean of observed and predicted values, respectively	[51]
$c_{R_p^2} = R \sqrt{R^2 - R_r^2}$	R^2 is squared correlation coefficient of models and R_r^2 is squared mean correlation coefficient of randomized models	[52]
$\bar{r}_m^2 = \frac{r_m^2 + r_0^2}{2}$	r^2 is the squared correlation coefficient value between observed and predicted endpoint values, and r_0^2 and r_m^2 are the respective squared correlation coefficients when the regression line is passed through the origin by interchanging the axes	[53]
$\Delta r_m^2 = r_m^2 - r_0^2 $	For the acceptable prediction, the value of all Δr_m^2 metrics should preferably be lower than 0.2 provided that the value of r^2 is more than 0.5	
$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right)$		
$r_0^2 = r^2 \times \left(1 - \sqrt{r^2 - r_m^2}\right)$		
$MAE = \frac{1}{n} \times \sum Y_{obs} - Y_{prd} $		

the reliability and robustness of the QSAR models were verified using the following three methodologies: i) internal validation or cross-validation by considering the training dataset, ii) external validation by considering the prediction set and iii) data randomization or Y-scrambling.

The various standard statistical metrics such as correlation coefficient (R^2), cross-validated correlation coefficient (Q^2), concordance correlation coefficient (CCC), the IIC, Q_{F1}^2 , Q_{F2}^2 , and Q_{F3}^2 , standard error of estimation (s), mean absolute error (MAE), Fischer ratio (F), novel metrics (r_m^2) and Y-scrambling ($c_{R_p^2}$) were employed to validate the developed QSAR models. The mathematical equations of various validation metrics are shown in Table 2.

R^2 statistic is a metric to evaluate the goodness of fit of a regression analysis. It measures the variation of experimental data with the predicted ones. The range of R^2 is between 0 (no correlation) and 1 (perfect fit). R^2 cross-validated (Q^2) is used for internal validation. The concordance correlation coefficient (CCC) is calculated to measure both precision and accuracy detecting how far each observation deviate from the best-fit. The CCC is calculated to detect both precision and accuracy distance of the observations from the fitting line and the degree of deviation of the regression line from that passing through the origin, respectively [51]. A lower value of MAE and s is desirable for good internal/external predictivity. Roy et al. [54] introduced a new metric r_m^2 that penalizes the r^2 value of a model when there is large deviation between r^2 and r_0^2 values (Table 2). For a reliable QSAR model, the r_m^2 and Δr_m^2 should be greater than 0.5 and smaller than 0.2,

respectively. Y-scrambling or Y-randomization is an assessment to ensure the developed QSAR model is not due to chance, thereby giving an idea of model robustness [52]. For a robust QSAR model, Todeschini $c_{R_p^2}$ parameter [55] is also calculated which should be more than 0.5. One of the important statistical parameters to judge different QSAR models is \bar{r}_m^2 for test set. Here, this parameter is used to select best model between six proposed models.

Model interpretation

A straightforward process for the structural interpretation of QSPR/QSAR models is provided by the CORAL application. Three types of attributes may be identified by computing the correlation weights across several iterations of the Monte Carlo optimization algorithm. The positive numerical value of CWs in every iteration is considered for endpoint increase, the attributes with a negative value of CWs in every iteration is a notation for endpoint decrease. The unstable numerical value in the different runs is not considered for predicting the promoter of the increase/decrease endpoint [19, 56].

Molecular docking

Molecular docking is a method commonly employed in drug discovery and development to identify protein–ligand binding configurations. This approach involves the docking of a molecule with a specific macromolecule and then computing the binding free energy between the ligand and receptor [35]. The structure was sketched in ChemDraw 16.0, and the energy was minimised in Chem3D using the MM2 technique [57]. The crystallographic structure of Human cytochrome P450 CYP17A1

in complex with abiraterone was obtained from the Protein Data Bank (PDB: 3RUK) and used for molecular docking [58]. AutoDock Vina was employed for docking studies (Molecular Graphics Lab, CA, USA) [59]. The value of exhaustiveness was 8 and the dimensions of the grid box were 20.0, 20.0, and 20.0 Å in size. The findings and illustration were examined visually using Discovery Studio visualizer 2021.

Results and discussion

QSPR modelling for pIC_{50}

Three types of outliers affect the model quality in QSPR/QSAR study. The first is the outliers in the dependent variable y , the second is the outliers in the direction of the independent variable X , and the third type of outliers indicates a different relationship between X and y . [60]. Here, based on several preliminary QSAR models,

six compounds (compounds No. 31, 32, 36, 37, 67, and 80) identified as outliers, these molecules showed a large absolute error (>3 s). These compounds fall in first type of outlier. The structure of these compounds is similar to the main body of the samples. So, they were removed from the data set before further data processing.

In this study, the balance of correlation approach was employed to generate QSAR models. A total of six QSAR models was generated utilizing two kinds of target functions i.e. TF_1 ($W_{\text{IIC}}=0.0$) and TF_2 ($W_{\text{IIC}}=0.2$). To obtain the preferable threshold value (T^*) and the number of epochs (N^*), the range of 1–10 for threshold and 1 to 50 for epoch were employed. In the case of TF_1 , the value of T^* and N^* were 1 and 10 for split 1; 1 and 3 for split 2; 1 and 7 for split 3, respectively. However, in the case of TF_2 , the value of optimum (T^* , N^*) for splits 1, 2, and 3 were (1, 10), (1, 10), and (1, 7), respectively.

Table 3 The statistical characteristics of CORAL models for pIC_{50} generated with TF_1 and TF_2

Split	Target function	Set	n	R^2	CCC	IIC	Q^2	s	MAE	Y-rand	cR_p^2	\bar{r}_m^2	Δr_m^2	F
1	TF1	TRN	29	0.813	0.897	0.551	0.779	0.234	0.189	0.033	0.797	–	–	117
		iTRN	20	0.813	0.866	0.607	0.773	0.275	0.229	0.050	0.788	–	–	78
		CAL	12	0.575	0.706	0.378	0.228	0.437	0.334	0.162	0.487	–	–	14
		VAL	19	0.679	0.765	0.402	0.549	0.364	0.259	–	–	0.510	0.279	36
	TF2	TRN	29	0.723	0.839	0.794	0.674	0.284	0.214	0.032	0.707	–	–	70
		iTRN	20	0.804	0.806	0.462	0.759	0.338	0.285	0.050	0.778	–	–	74
		CAL	12	0.688	0.802	0.828	0.552	0.313	0.267	0.064	0.655	–	–	22
		VAL	19	0.729	0.789	0.489	0.666	0.272	0.207	–	–	0.555	0.253	46
2	TF1	TRN	30	0.881	0.937	0.821	0.864	0.185	0.133	0.033	0.865	–	–	207
		iTRN	19	0.854	0.896	0.492	0.820	0.226	0.165	0.079	0.813	–	–	99
		CAL	11	0.502	0.653	0.279	0.029	0.546	0.378	0.087	0.456	–	–	9
		VAL	20	0.652	0.759	0.642	0.563	0.399	0.321	–	–	0.494	0.287	34
	TF2	TRN	30	0.781	0.877	0.773	0.740	0.251	0.175	0.027	0.767	–	–	100
		iTRN	19	0.805	0.831	0.865	0.752	0.278	0.216	0.049	0.780	–	–	70
		CAL	11	0.939	0.863	0.968	0.923	0.319	0.263	0.040	0.919	–	–	140
		VAL	20	0.775	0.826	0.611	0.713	0.334	0.269	–	–	0.613	0.218	62
3	TF1	TRN	30	0.736	0.848	0.751	0.674	0.288	0.204	0.043	0.714	–	–	78
		iTRN	18	0.754	0.853	0.856	0.704	0.277	0.231	0.077	0.714	–	–	49
		CAL	12	0.418	0.571	0.476	0.137	0.440	0.336	0.076	0.378	–	–	7
		VAL	20	0.651	0.798	0.651	0.457	0.302	0.231	–	–	0.518	0.190	34
	TF2	TRN	30	0.782	0.878	0.774	0.735	0.261	0.194	0.020	0.772	–	–	101
		iTRN	18	0.850	0.909	0.857	0.812	0.223	0.189	0.048	0.825	–	–	90
		CAL	12	0.777	0.807	0.881	0.697	0.300	0.246	0.100	0.725	–	–	35
		VAL	20	0.727	0.847	0.628	0.642	0.266	0.209	–	–	0.615	0.165	48

The mathematical relationship for the developed QSAR model of pIC_{50} using TF_1 and TF_2 for three splits are displayed below:

The Monte Carlo optimization with target function TF_1

$$\begin{aligned} \text{Split1 } pIC_{50} = & - 8.4912(\pm 0.2835) \\ & + 0.0978(\pm 0.0021) \\ & \times DCW(1, 10) \end{aligned} \quad (15)$$

$$\begin{aligned} \text{Split2 } pIC_{50} = & - 16.266(\pm 0.2769) \\ & + 0.1309(\pm 0.0017) \\ & \times DCW(1, 3) \end{aligned} \quad (16)$$

$$\begin{aligned} \text{Split3 } pIC_{50} = & - 4.2842(\pm 0.2158) \\ & + 0.0626(\pm 0.0015) \\ & \times DCW(1, 7) \end{aligned} \quad (17)$$

The Monte Carlo optimization with target function TF_2

$$\begin{aligned} \text{Split1 } pIC_{50} = & - 3.1689(\pm 0.2140) \\ & + 0.0272(\pm 0.0007) \\ & \times DCW(1, 10) \end{aligned} \quad (18)$$

$$\begin{aligned} \text{Split2 } pIC_{50} = & - 9.6171(\pm 0.3420) \\ & + 0.0758(\pm 0.0017) \\ & \times DCW(1, 10) \end{aligned} \quad (19)$$

$$\begin{aligned} \text{Split3 } pIC_{50} = & - 7.0645(\pm 0.3206) \\ & + 0.0482(\pm 0.0013) \\ & \times DCW(1, 7) \end{aligned} \quad (20)$$

The statistical results of designed QSAR models for three splits utilizing TF_1 and TF_2 are presented in Table 3. As can be seen, all developed QSAR models were acceptable statistically and agreed with the requirements of various validation criteria.

According to the results presented in Table 3, it was found that the models constructed using TF_2 (with IIC) had better statistical results than the models constructed using TF_1 (without IIC). The results of calibration and validation sets were better for the models constructed by using TF_2 , but the inferior quality of the model for the training sets was obtained. Hence, it can be expressed that the models designed with the IIC are more statistically considerable and robust for the present dataset. Based on validation metric study of QSPR/QSAR models by Ojha et al., the \bar{r}_m^2 value of models is used to judge the quality of the predictions by different models. The QSAR model developed by TF_2 for split 3 was selected as a prominent model with highest \bar{r}_m^2 ($\bar{r}_m^2 = 0.615$).

The plot of observed pIC_{50} versus predicted pIC_{50} for three models designed with TF_2 is displayed in Fig. 1. In the QSAR model generated by utilizing the Monte Carlo method, the outliers were introduced by the statistical defects. So, in the present QSAR model created by TF_2 , the number of outliers was found six for all splits. Table 4 displays flavonols IDs, SMILES codes, and descriptor of correlation weights (DCWs) with their experimental and predicted pIC_{50} .

Interpretation of the QSAR model

The mechanistic interpretation of a QSAR model is the fifth principle of OECD. The mechanistic interpretation of the QSAR model provides a correlation and a

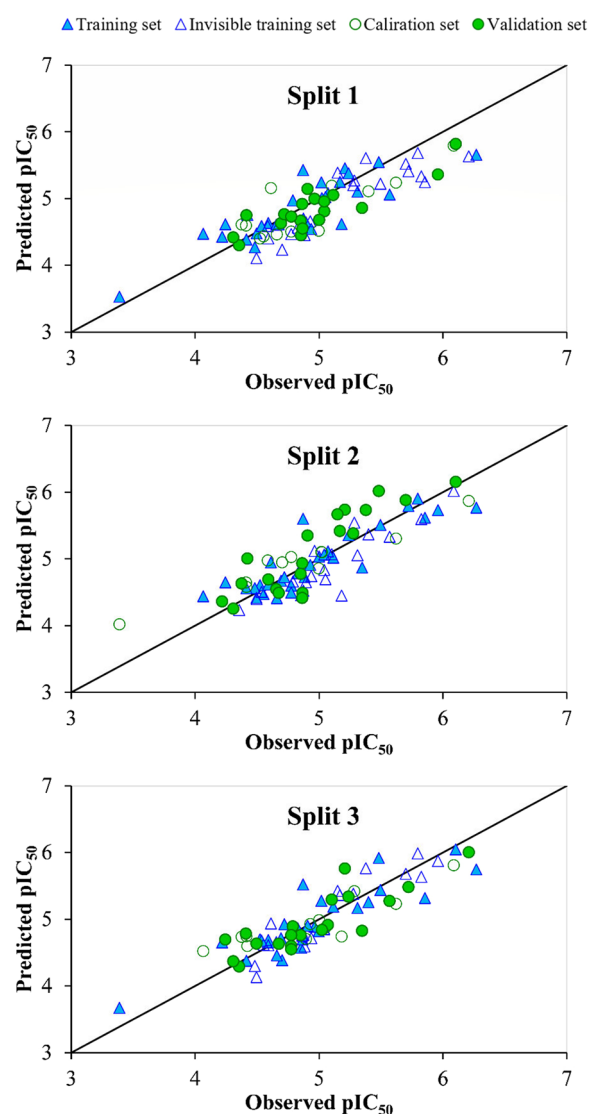


Fig. 1 Observed pIC_{50} versus predicted pIC_{50} values for three CORAL models constructed based on TF_2

Table 4 SMILES notation, the distribution of splits, DCWs, observed and predicted pIC_{50} of flavonols (+, −, #, and * show the compounds located in the training, invisible training, calibration, and validation sets respectively)

No	SMILES	Split			pIC_{50}	DCW(T, N)			Predicted pIC_{50}		
		1	2	3		Split 1	Split 2	Split 3	Split 1	Split 2	Split 3
1	<chem>COc1ccc(cc1OC)C2=C(O)C(=O)c3ccccc3O2</chem>	+	#	+	3.39	246.56	179.87	222.70	3.53	4.02	3.67
2	<chem>COC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	+	+	4.52	278.73	187.60	244.01	4.40	4.61	4.70
3	<chem>CCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	*	−	4.86	297.66	191.93	245.04	4.92	4.94	4.75
4	<chem>CCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	#	+	5.00	282.96	190.91	246.53	4.52	4.86	4.82
5	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	+	*	4.78	282.49	187.51	241.89	4.51	4.60	4.60
6	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	−	+	*	4.78	281.10	186.14	240.90	4.47	4.50	4.55
7	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	−	+	4.85	280.45	185.60	241.36	4.45	4.46	4.57
8	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	+	#	4.06	281.26	185.40	240.36	4.47	4.44	4.52
9	<chem>COc1ccc(cc1OC)C2=C(O)C(OC)C(=O)c3ccccc3O2</chem>	+	+	−	4.48	273.82	187.00	235.69	4.27	4.56	4.30
10	<chem>CCC(C)OC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	−	*	4.35	274.98	182.68	235.54	4.30	4.24	4.29
11	<chem>CCCC(C)OC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	+	−	4.55	279.85	185.82	242.37	4.43	4.47	4.62
12	<chem>CCN(CC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	+	*	4.24	286.50	188.17	243.94	4.62	4.65	4.70
13	<chem>CCN(CC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	*	+	4.22	279.62	184.40	243.02	4.43	4.37	4.65
14	<chem>CCN(CC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	−	*	4.50	281.46	184.87	242.66	4.48	4.40	4.64
15	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	−	#	+	4.59	285.56	192.48	243.07	4.59	4.98	4.66
16	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	−	*	−	4.59	278.69	188.71	242.15	4.40	4.69	4.61
17	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	−	+	−	4.88	280.53	189.18	241.78	4.45	4.73	4.59
18	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	−	+	+	5.85	309.77	200.92	256.83	5.25	5.62	5.32
19	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	−	*	5.57	302.90	197.16	255.91	5.06	5.33	5.27
20	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	−	+	5.40	304.74	197.62	255.54	5.11	5.37	5.26
21	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	#	#	5.62	309.50	196.79	255.03	5.24	5.31	5.23
22	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	+	+	5.11	302.63	193.03	254.11	5.05	5.02	5.19
23	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	−	+	5.31	304.47	193.50	253.75	5.10	5.05	5.17
24	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCOCC3)C(=O)c4ccccc4O2</chem>	*	−	+	4.96	300.56	194.37	247.94	5.00	5.12	4.89
25	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCOCC3)C(=O)c4ccccc4O2</chem>	*	−	−	5.04	293.68	190.60	247.02	4.81	4.84	4.85
26	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCOCC3)C(=O)c4ccccc4O2</chem>	*	+	*	5.35	295.52	191.07	246.66	4.86	4.87	4.83
27	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCCC3)C(=O)c4ccccc4O2</chem>	+	−	+	4.87	289.62	186.51	245.34	4.70	4.52	4.76
28	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCCC3)C(=O)c4ccccc4O2</chem>	*	+	#	4.41	291.46	186.97	244.98	4.75	4.56	4.75
29	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCN(C)CC3)C(=O)c4ccccc4O2</chem>	+	*	+	5.48	320.82	206.21	269.28	5.55	6.02	5.92
30	<chem>COc1ccc(cc1OC)C2=C(O)C(OCCCN3CCN(C)CC3)C(=O)c4ccccc4O2</chem>	*	+	−	5.96	313.95	202.44	268.36	5.36	5.73	5.87

Table 4 (continued)

No	SMILES	Split			pIC ₅₀	DCW(T, N)			Predicted pIC ₅₀		
		1	2	3		Split 1	Split 2	Split 3	Split 1	Split 2	Split 3
33	<chem>COc1ccc(cc1OC)C2=C(OCCCCCN3CCN(C)CC3)C(=O)c4cccc4O2</chem>	+	*	*	5.21	317.40	202.51	266.00	5.46	5.74	5.76
34	<chem>COc1ccc(cc1OC)C2=C(OCCCN3CCCC3)C(=O)c4cccc4O2</chem>	-	*	-	5.70	319.85	204.41	264.33	5.52	5.88	5.68
35	<chem>COc1ccc(cc1OC)C2=C(OCCCN3CCCC3)C(=O)c4cccc4O2</chem>	-	-	-	5.82	312.98	200.64	263.41	5.33	5.60	5.64
38	<chem>COc1ccc(cc1OC)C2=C(OCCCCCN3CCCC3)C(=O)c4cccc4O2</chem>	+	+	+	4.87	316.43	200.71	261.05	5.43	5.60	5.52
39	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)O</chem>	-	#	+	4.70	272.50	192.11	237.49	4.24	4.95	4.39
40	<chem>COC1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	-	+	+	4.83	284.69	189.92	247.43	4.57	4.78	4.87
41	<chem>CCOC1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	+	*	5.07	303.63	194.26	248.46	5.08	5.11	4.92
42	<chem>CCCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	+	#	5.00	288.93	193.24	249.95	4.68	5.04	4.99
43	<chem>CCCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	*	*	4.85	288.45	189.83	245.31	4.67	4.78	4.76
44	<chem>CCCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	*	+	+	4.69	287.07	188.46	244.33	4.63	4.67	4.72
45	<chem>CCCCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	*	#	4.37	286.42	187.93	244.78	4.61	4.63	4.74
46	<chem>CCCCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	+	+	+	4.59	287.22	187.73	243.78	4.64	4.62	4.69
47	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OC(C)C</chem>	+	*	#	4.42	291.72	192.86	241.93	4.76	5.01	4.60
48	<chem>CCC(C)OC1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	+	+	4.66	280.94	185.01	238.96	4.46	4.41	4.46
49	<chem>CCCC(C)OC1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	#	#	*	4.41	285.82	188.15	245.80	4.60	4.65	4.79
50	<chem>CCCN(CCCC)CCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	-	+	*	5.72	315.74	203.25	260.25	5.41	5.79	5.48
51	<chem>CCCN(CCCC)CCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	-	+	+	5.50	308.86	199.48	259.33	5.22	5.51	5.44
52	<chem>CCCN(CCCC)CCCO1=C(Oc2cc(OC)ccc2C1=O)c3ccc(OC)c(OC)c3</chem>	-	-	#	5.28	310.70	199.95	258.97	5.27	5.54	5.42
53	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCOCC4</chem>	*	*	+	4.91	305.99	197.41	248.19	5.15	5.35	4.90
54	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCOCC4</chem>	*	+	-	5.04	299.12	193.64	247.27	4.96	5.07	4.86
55	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCOCC4</chem>	+	#	*	5.02	300.96	194.11	246.91	5.01	5.10	4.84
56	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCCC4</chem>	#	+	-	4.61	306.52	192.08	248.94	5.16	4.95	4.94
57	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCCC4</chem>	+	-	*	4.79	299.65	188.31	248.02	4.97	4.66	4.89
58	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCCC4</chem>	+	-	#	5.05	301.49	188.78	247.66	5.02	4.70	4.88
59	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCN(C)CC4</chem>	*	*	+	6.10	330.85	208.01	271.96	5.82	6.16	6.05
60	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCN(C)CC4</chem>	-	#	*	6.21	323.98	204.24	271.04	5.63	5.87	6.00
61	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCN(C)CC4</chem>	-	+	-	5.80	325.82	204.71	270.68	5.68	5.91	5.99
62	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCCC4</chem>	#	-	#	6.09	329.88	206.21	267.01	5.79	6.02	5.81
63	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCN4CCCC4</chem>	-	*	-	5.38	323.01	202.44	266.09	5.61	5.73	5.77

Table 4 (continued)

No	SMILES	Split			pIC ₅₀	DCW(T, N)			Predicted pIC ₅₀		
		1	2	3		Split 1	Split 2	Split 3	Split 1	Split 2	Split 3
64	<chem>COc1ccc2C(=O)C(=C(Oc2c1)c3ccc(OC)c(OC)c3)OCCCCN4CCCC4</chem>	+	+	+	6.27	324.85	202.91	265.73	5.66	5.77	5.75
65	<chem>COc1cc(cc(OC)c1OC)C2=C(O)C(=O)c3ccccc3O2</chem>	-	+	-	4.49	267.80	185.10	232.23	4.11	4.42	4.13
66	<chem>COC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	*	+	+	4.72	292.01	189.19	248.67	4.77	4.73	4.93
68	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	-	+	#	4.93	288.14	191.58	248.76	4.66	4.91	4.93
69	<chem>CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	-	#	4.89	287.66	188.18	244.11	4.65	4.65	4.71
70	<chem>CCCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	*	+	4.65	286.28	186.81	243.13	4.61	4.55	4.66
71	<chem>CCCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	-	+	4.54	285.63	186.27	243.58	4.59	4.51	4.68
72	<chem>CCCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	*	*	4.67	286.43	186.07	242.58	4.61	4.49	4.63
73	<chem>COc1cc(cc(OC)c1OC)C2=C(OC(C)C)C(=O)c3ccc-cc3O2</chem>	+	#	+	4.41	278.13	187.26	237.30	4.39	4.58	4.38
74	<chem>CCC(OC)OC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	*	*	*	4.31	279.29	182.94	237.15	4.42	4.25	4.37
75	<chem>CCCC(OC)OC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	*	*	+	4.87	284.16	186.08	243.98	4.55	4.49	4.70
76	<chem>CCN(CC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	-	*	+	4.86	284.80	185.07	245.25	4.57	4.42	4.76
77	<chem>CCN(CC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	-	#	5.18	286.64	185.54	244.88	4.62	4.45	4.74
78	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	*	#	*	4.77	290.74	193.15	245.29	4.73	5.03	4.76
79	<chem>CCCN(CCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	-	-	4.94	283.86	189.38	244.37	4.54	4.74	4.72
81	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	-	*	-	5.15	314.95	201.59	259.05	5.39	5.67	5.43
82	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	-	*	-	5.28	308.07	197.83	258.13	5.20	5.38	5.38
83	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	*	-	5.17	309.91	198.29	257.77	5.25	5.42	5.36
84	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	+	*	5.24	314.68	197.46	257.26	5.38	5.36	5.34
85	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	#	-	*	5.10	307.80	193.70	256.34	5.19	5.07	5.29
86	<chem>CCCCN(CCCC)CCCCOC1=C(Oc2ccccc2C1=O)c3cc(OC)c(OC)c(OC)c3</chem>	+	#	+	5.02	309.64	194.16	255.97	5.24	5.11	5.28

relationship between the chemical structure of the compounds and their property/activity. It also enunciates the molecular features which are responsible for the increase/decrease of endpoints that can be computed from QSAR models. Information on the mechanistic interpretation of flavonols as a promoter of pIC₅₀ increase/decrease may aid in the design and development of new flavonol derivatives.

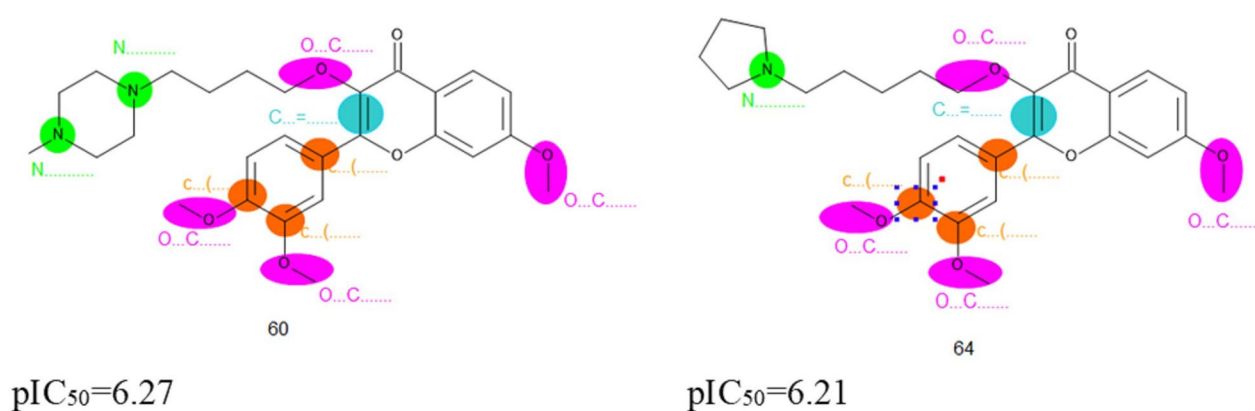
In CORAL, correlation weights (CWs) of structural attributes (SA_k) are calculated in three or more runs and

the mechanistic interpretation is achieved by analysis of CWs. If in all probes of the optimization, the numerical value of CW of structural attributes is found greater than zero, then these attributes are considered as a promoter of increase. Whereas, if the numerical value of CW of structural attributes is found smaller than zero, then these attributes are defined as the promoter of decrease [61, 62].

The list of attributes and their correlation weights for three runs of all splits computed with TF₂ is presented

Table 5 Important features interpretation for increasing of pIC_{50} values of three splits

No	Attributes	Split	CW (M_{Fk}) in run 1	CW (M_{Fk}) in run 2	CW (M_{Fk}) in run 3	N_T^a	N_{IT}^b	N_C^c	Defect (M_{Fk})	Interpretation
1	C...=...	1	1.080	1.495	0.261	29	20	12	0.000	Aliphatic carbon atom connected to double-bound
		2	0.810	1.103	0.681	29	21	12	0.000	
		3	1.978	1.006	0.076	28	20	12	0.000	
2	O...C...	1	1.216	1.517	0.893	29	20	12	0.000	Aliphatic oxygen atom connected to aliphatic carbon
		2	0.014	0.254	0.380	29	21	12	0.000	
		3	0.150	0.019	1.054	28	20	12	0.000	
3	c...(...	1	0.329	0.003	0.560	29	20	12	0.000	Branching on ring
		2	0.271	0.054	0.161	29	21	12	0.000	
		3	0.263	0.073	0.644	28	20	12	0.000	
4	N...	1	2.197	3.253	3.130	19	10	6	0.0062	Aliphatic nitrogen
		2	0.050	0.207	0.170	18	11	7	0.0015	
		3	0.728	2.431	2.322	16	15	6	0.0032	
		2	-0.718	-0.708	-0.854	22	16	10	0.0023	
		3	-0.285	-0.784	-1.012	22	16	9	0.0012	

^a No. of attributes in the training set^b No. of attributes in the invisible training set^c No. of attributes in the calibration set**Fig. 2** Good fingerprints obtained from Monte Carlo optimization method

in Table 5. The most significant structural attributes as the promoter of pIC_{50} increase were distinguished and extracted. The structural attributes as promoters of increase of pIC_{50} were aliphatic carbon atom connected to double-bound (C...=..., aliphatic oxygen atom connected to aliphatic carbon (O...C...), branching on aromatic ring (c...()), and aliphatic nitrogen (N...). The good fingerprints obtained from Monte Carlo optimization method are indicated in Fig. 2. These attributes for two compounds with the highest pIC_{50} are shown in Fig. 2 (compound no. 60 and 64).

A series of natural flavonols with unknown pIC_{50} was selected and their pIC_{50} was calculated from the QSAR models of best split (split 3). Names, chemical structure

and corresponding predicted pIC_{50} of selected natural flavonol derivatives with pIC_{50} more than 4, are depicted in Table 6. These compounds were also considered for molecular docking studies.

Molecular docking studies

The docking for abiraterone was performed into the active site of Human Cytochrome P450 CYP17A1 (PDB: 3RUK) to validate the binding energy of ligand–protein interactions. The validation results showed a binding energy of -10.3 kcal/mol for abiraterone and a root-mean-square deviation (RMSD) value 1.172 Å (Fig. 3). The active pocket consisted of amino acid residues such as Val366, Val483, Val482, Ala367, Glu305, Gly301,

Table 6 The chemical structure of some natural flavonols with predicted pIC_{50} using leading model (split 3), docking scores (Kcal mol^{-1}) and amino acid interacted with 3RUK

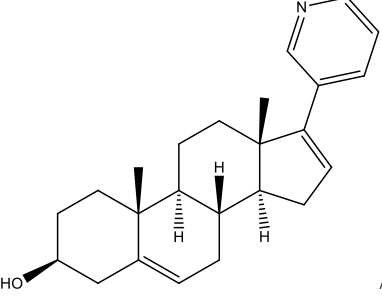
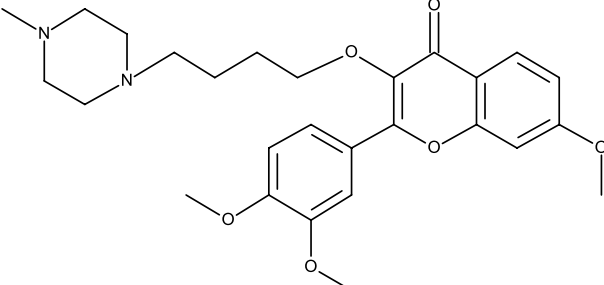
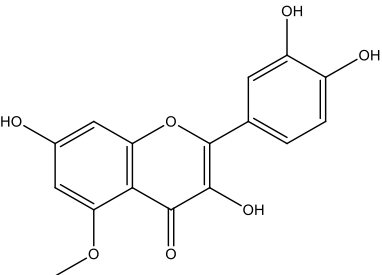
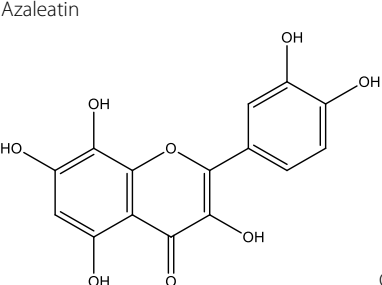
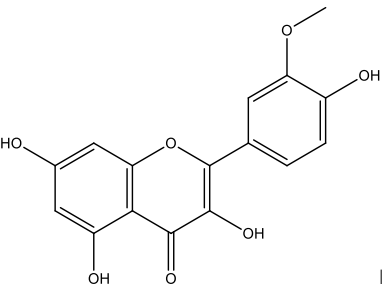
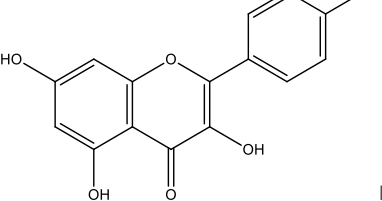
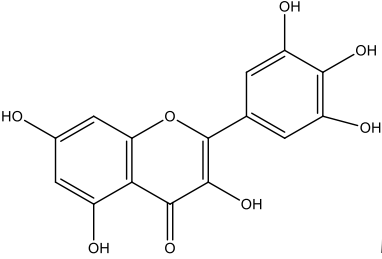
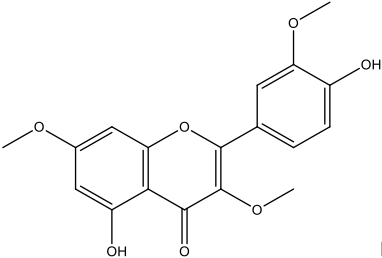
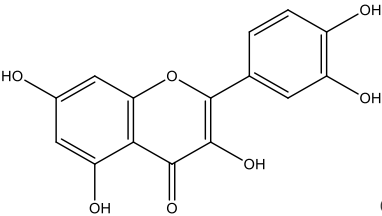
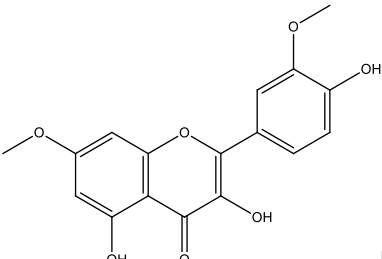
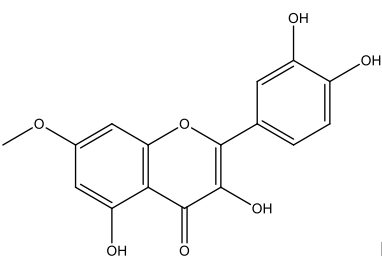
Structure	pIC_{50} (prd)	Free binding energy (kcal/mol)	Amino acid interacted
	-	- 10.3	Ala113, Gly301, Arg239, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371, Ile205, Phe114, Thr306, Cys442, Val366, Ala367, Leu209, Tyr201
 Abiraterone	6.03	- 8.1	Ala113, Gly301, Asp298, Gly444, Val483, Val366, Ala302, Ala367, Ile299, Ile443, Ile371, Glu305, Phe435, Cys442, Gly303, Ile371, Pro434, Leu447, Ile112, Phe114, Thr306
Compound no. 60 with highest Activity			
	4.36	- 8.1	Ile205, Val482, Asp298, Ala302, Ala113, Asn202, Gly297
 Azaleatin	4.59	- 8.5	Asn202, Arg239 Ile205, Ala105, Ala302, Ala113
	4.68	- 8.0	Ala113, Gly301, Asp298, Gly297, Arg239, Val236, Ala105, Ile205, Tyr201, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371
 Isorhamnetin			

Table 6 (continued)

Structure	pIC ₅₀ (prd)	Free binding energy (kcal/mol)	Amino acid interacted
 Myricetin	4.21	- 8.2	Ala113, Gly301, Asp298, Gly297, Arg239, Phe114, Phe300, Ile205, Tyr201, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371
 Pachypodol	4.35	- 7.9	Ala113, Ala105, Gly301, Arg239, Ile205, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371, Val236, Thr306
 Quercetin	4.25	- 8.4	Ala113, Ala105, Gly301, Arg239, Ile205, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371, Val236, Thr306, Phe114, Tyr201
 Rhamnazin	4.66	-8.3	Ala113, Ala105, Gly301, Arg239, Ile205, Asn202, Ile206, Glu305, Ala302, Val482, Val483, Ile371, Val236, Tyr201, Asp298, Gly297, Ala367, Ala105, Val366
 Rhamnetin	4.45	- 8.2	Ala113, Gly301, Arg239, Ile205, Asn202, Ile206, Ala302, Val482, Val483, Ile371, Asp298, Gly297, Ala367, Tyr201

Leu209, Asn220, Tyr201, Ile206, Ile205, Arg239, Phe114, Ala302, Ile371, Ala113, Thr306, and Cys442, which play fundamental roles by hydrophobic interactions and forming H-bond (Fig. 4).

In addition, the docking studies for eight natural flavonols with predicted pIC₅₀ more than 4.0 based on the best model (split 3), were conducted along with compound number 60, which has high experimental activity.

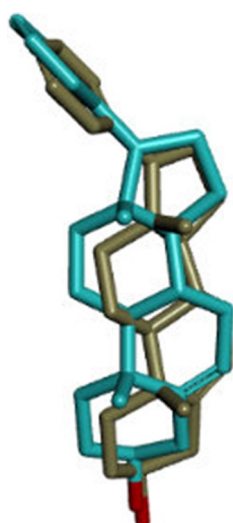


Fig. 3 Superposition of the abiraterone output docked ligand (blue) and the co-crystallized ligand (green) of 3RUKA

Natural flavonols azaleatin, gossypetin, isorhamnetin, myricetin, pachypodol, quercetin, rhamnazin, and rhamnetin exhibited binding energies of -8.1 , -8.5 , -8.0 , -8.2 , -7.9 , -8.4 , -8.3 , and -8.2 kcal/mol, respectively (Table 6). The docking outcomes matched the calculated pIC_{50} of flavonols. The superimposition image of the optimum binding pose for each suggested flavonol is displayed in Fig. 5. Figure 6 shows the 3D docking mode and 2D schematic depiction of interactions for some natural flavonols and the active ligand. The oxygen atom

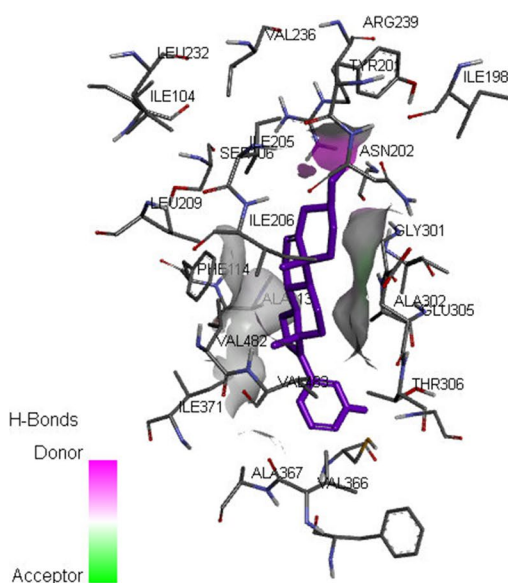


Fig. 4 3D docking mode and 2D schematic interaction diagram for the best pose of abiraterone redocked into 3RUK crystal structure



Fig. 5 Superimposed poses of docked molecules and the co-crystallized abiraterone (violet) into the active site of 3RUKA

was involved in hydrogen bond interactions with the active site amino acid residues, and so the oxygen of flavonols was particularly significant for the anti-prostate cancer effect of flavonols. The positive contribution of oxygen atom on pIC_{50} of flavonol derivatives was seen in the mechanistic interpretation of the above-mentioned QSAR models. So, the present QSAR models are acceptable for a wide range of flavonols derivatives.

Conclusion

In the present study, a reliable QSAR model was described to predict the anti-prostate cancer activities of 81 flavonol derivatives using the Monte Carlo optimization technique of CORAL software. To date, the

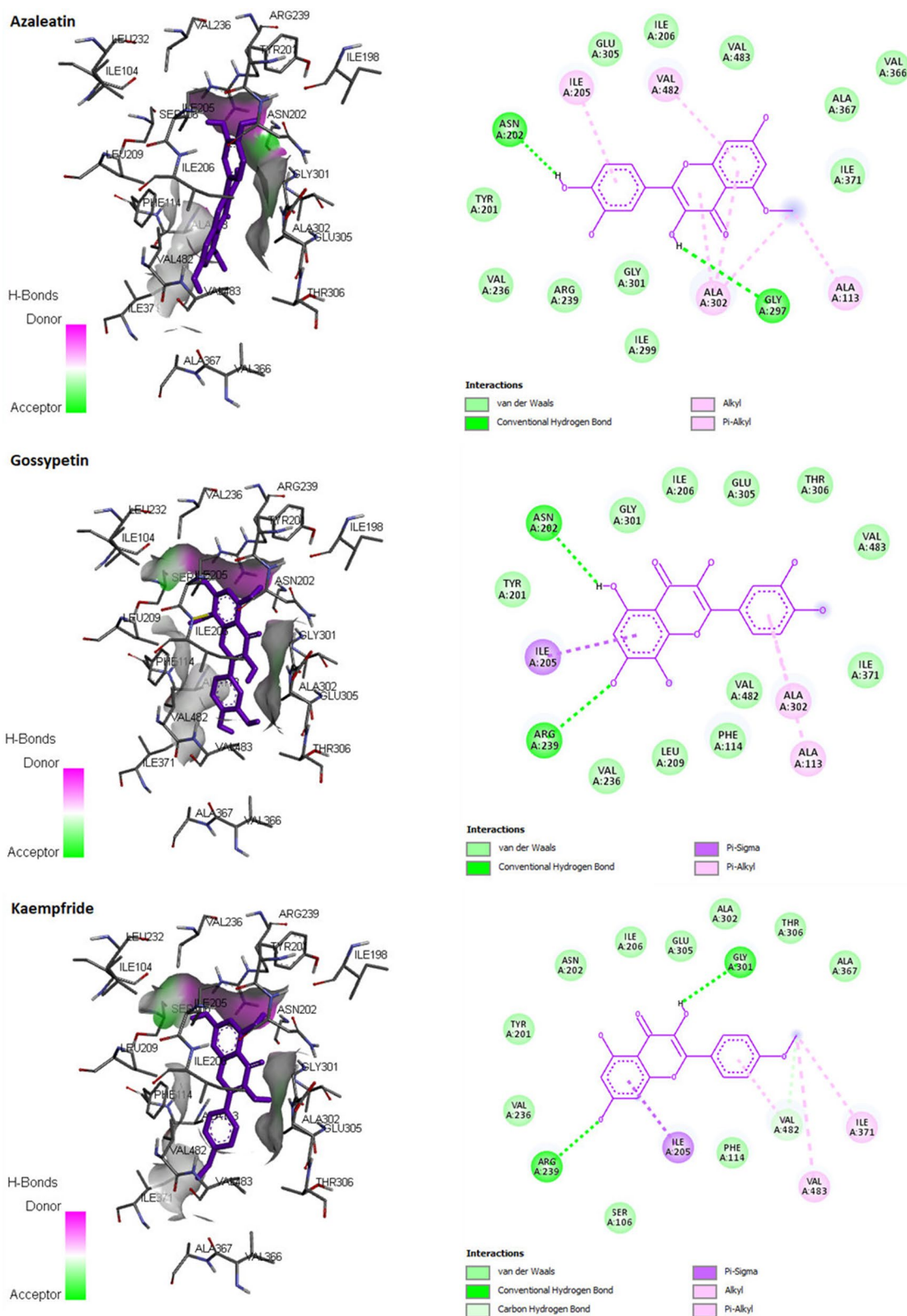


Fig. 6 3D docking mode and 2D schematic interaction diagram for the best pose of some natural flavonols against 3RUK crystal structure (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this article)

QSAR models to predict the pIC_{50} of this dataset were not previously reported. Six QSAR models were constructed utilizing the balance of correlation method with two target functions TF_1 ($WIIC=0.0$) and TF_2 ($WIIC=0.2$). The IIC was employed to improve the reliability and robustness of the models. The QSAR models developed by using TF_2 were found better than the models developed by TF_1 . The predictability and robustness of designed models were evaluated by the various statistical parameters such as R^2 , Q^2 , IIC, CCC, MAE, s , r_m^2 , Δr_m^2 , $C_{R_p}^2$, F and Y-test. Based on 'statistical defect', $d(A)$ for a SMILES attribute, the AD was also analysed and the outliers were extracted. The structural attributes as promoters of increase/decrease of pIC_{50} were identified and used to predict the pIC_{50} of natural flavonols. The mechanistic interpretation was also confirmed by molecular docking of natural flavonols into the active site of Human Cytochrome P450 CYP17A1 (PDB: 3RUK).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13065-023-00999-y>.

Additional file 1: Table S1. Chemical structures of flavonol derivatives and IC50 values against PC-3 prostate cancer cells.

Acknowledgements

The authors would like to express their deepest gratitude to Dr. Alla P. Toropova and Dr. Audrey A. Toropov for providing the CORAL software.

Author contributions

FT performed drawing of structures and building the QSAR models. SA did visualization, supervision, performed interpretation of models and molecular docking. SL wrote original draft, and PK did editing of the manuscript. AA conducted molecular docking.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

- Middleton E, Kandaswami C, Theoharides TC. The effects of plant flavonoids on mammalian cells: implications for inflammation, heart disease, and cancer. *Pharmacol Rev.* 2000;52(4):673–751.
- Panche AN, Diwan AD, Chandra SR. Flavonoids: an overview. *J Nutr Sci.* 2016; 5.
- Yan W, et al. Flavonoids potentiated anticancer activity of cisplatin in non-small cell lung cancer cells in vitro by inhibiting histone deacetylases. *Life Sci.* 2020;258: 118211.
- Liu HL, Jiang WB, Xie MX. Flavonoids: recent advances as anticancer drugs. *Recent Pat Anti-Cancer Drug Discov.* 2010;5(2):152–64.
- Ravishankar D, et al. Flavonoids as prospective compounds for anti-cancer therapy. *Int J Biochem Cell Biol.* 2013;45(12):2821–31.
- Yu T, et al. Exploring the chemical space of CYP17A1 inhibitors using cheminformatics and machine learning. *Molecules.* 2023;28(4):1679.
- Wróbel TM, et al. Non-steroidal CYP17A1 inhibitors: discovery and assessment. *J Med Chem.* 2023;66:6542.
- Fei Q, et al. Rutin inhibits androgen synthesis and metabolism in rat immature Leydig cells in vitro. *Andrologia.* 2021;53(11): e14221.
- Xin-Guang S, et al. New prenylated flavonoid glycosides derived from *Epimedium wushanense* by β -glucosidase hydrolysis and their testosterone production-promoting effects. *Chin J Nat Med.* 2022;20(9):712–20.
- Abdi SAH, et al. Morusflavone, a new therapeutic candidate for prostate cancer by CYP17A1 inhibition: exhibited by molecular docking and dynamics simulation. *Plants.* 2021;10(9):1912.
- Britton RG, et al. Synthesis and biological evaluation of novel flavonols as potential anti-prostate cancer agents. *Eur J Med Chem.* 2012;54:952–8.
- Khan I, et al. Biodegradable nanoparticulate co-delivery of flavonoid and doxorubicin: mechanistic exploration and evaluation of anticancer effect in vitro and in vivo. *Biomater Biosyst.* 2021;3: 100022.
- Le Marchand L. Cancer preventive effects of flavonoids—a review. *Biomed Pharmacother.* 2002;56(6):296–301.
- Rajamahanty S, et al. Growth inhibition of androgen-responsive prostate cancer cells with Brefeldin A targeting cell cycle and androgen receptor. *J Biomed Sci.* 2010;17(1):1–8.
- Tavsan Z, Kayali HA. Flavonoids showed anticancer effects on the ovarian cancer cells: involvement of reactive oxygen species, apoptosis, cell cycle and invasion. *Biomed Pharmacother.* 2019;116: 109004.
- Isaacs JT, Isaacs WB. Androgen receptor outwits prostate cancer drugs. *Nat Med.* 2004;10(1):26–7.
- Menezes JC, et al. Natural and synthetic flavonoids: structure–activity relationship and chemotherapeutic potential for the treatment of leukemia. *Crit Rev Food Sci Nutr.* 2016;56(sup1):S4–28.
- Chen I-L, et al. Synthesis and antiproliferative evaluation of amide-containing flavone and isoflavone derivatives. *Bioorg Med Chem.* 2008;16(16):7639–45.
- Ahmadi S, Habibpour E. Application of GA-MLR for QSAR modeling of the arylthioindole class of tubulin polymerization inhibitors as anticancer agents. *Anti-Cancer Agents Med Chem.* 2017;17(4):552–65.
- Ahmadi S, et al. Predictive QSAR modeling for the antioxidant activity of natural compounds derivatives based on Monte Carlo method. *Mol Divers.* 2021;25(1):87–97.
- Toropova AP, Toropov AA. CORAL software: prediction of carcinogenicity of drugs by means of the Monte Carlo method. *Eur J Pharm Sci.* 2014;52:21–5.
- Ghasedi N, et al. DFT based QSAR study on quinolone-triazole derivatives as antibacterial agents. *J Recept Signal Transduction.* 2022;42(4):418–28.
- Kumar P, Kumar A. Monte Carlo method based QSAR studies of Mer kinase inhibitors in compliance with OECD principles. *Drug Research.* 2018;68(04):189–95.
- Lotfi S, Ahmadi S, Kumar P. The Monte Carlo approach to model and predict the melting point of imidazolium ionic liquids using hybrid optimal descriptors. *RSC Adv.* 2021;11(54):33849–57.
- Toropova AP, et al. The system of self-consistent models for vapour pressure. *Chem Phys Lett.* 2022;790: 139354.

Received: 2 March 2023 Accepted: 30 June 2023

Published online: 26 July 2023

26. Jafari K, et al. Correlation Intensity Index (CII) as a criterion of predictive potential: applying to model thermal conductivity of metal oxide-based ethylene glycol nanofluids. *Chem Phys Lett*. 2020;754: 137614.
27. Duhan M, et al. Quantitative structure activity relationship studies of novel hydrazone derivatives as α -amylase inhibitors with index of ideality of correlation. *J Biomol Struct Dyn*. 2022;40(11):4933–53.
28. Ahmadi S, et al. CORAL: Monte Carlo based global QSAR modelling of Bruton tyrosine kinase inhibitors using hybrid descriptors. *SAR QSAR Environ Res*. 2021;32(12):1013–31.
29. Azimi A, et al. SMILES-based QSAR and molecular docking study of oseltamivir derivatives as influenza inhibitors. *Polycyclic Aromat Compd*. 2022;43:3257.
30. Hamzehali H, et al. Quantitative structure–activity relationship modeling for predication of inhibition potencies of imatinib derivatives using SMILES attributes. *Sci Rep*. 2022;12(1):1–9.
31. Kumar A, Kumar P, Singh D. QSRR modelling for the investigation of gas chromatography retention indices of flavour and fragrance compounds on Carbowax 20 M glass capillary column with the index of ideality of correlation and the consensus modelling. *Chemom Intell Lab Syst*. 2022;224: 104552.
32. Kumar P, Kumar A. Nucleobase sequence based building up of reliable QSAR models with the index of ideality correlation using Monte Carlo method. *J Biomol Struct Dyn*. 2020;38(11):3296–306.
33. Kumar A, Kumar P. Prediction of power conversion efficiency of phenothiazine-based dye-sensitized solar cells using Monte Carlo method with index of ideality of correlation. *SAR QSAR Environ Res*. 2021;32(10):817–34.
34. Toropov AA, et al. "Ideal correlations" for biological activity of peptides. *Biosystems*. 2019;181:51–7.
35. Javidfar M, Ahmadi S. QSAR modelling of larvicidal phytocompounds against *Aedes aegypti* using index of ideality of correlation. *SAR QSAR Environ Res*. 2020;31(10):717–39.
36. Ahmadi S, et al. SMILES-based QSAR and molecular docking study of xanthone derivatives as α -glucosidase inhibitors. *J Recept Signal Transduct*. 2021;42:361.
37. Dutta S, et al. Natural products: an upcoming therapeutic approach to cancer. *Food Chem Toxicol*. 2019;128:240–55.
38. Li X, et al. A new class of flavonol-based anti-prostate cancer agents: design, synthesis, and evaluation in cell models. *Bioorg Med Chem Lett*. 2016;26(17):4241–5.
39. Li X, et al. 3-O-Substituted-3', 4', 5'-trimethoxyflavonols: synthesis and cell-based evaluation as anti-prostate cancer agents. *Bioorg Med Chem*. 2017;25(17):4768–77.
40. Li X, et al. Structure–activity relationship and pharmacokinetic studies of 3-O-substituted flavonols as anti-prostate cancer agents. *Eur J Med Chem*. 2018;157:978–93.
41. Lotfi S, Ahmadi S, Kumar P. A hybrid descriptor based QSPR model to predict the thermal decomposition temperature of imidazolium ionic liquids using Monte Carlo approach. *J Mol Liq*. 2021;338: 116465.
42. Kumar A, Sindhu J, Kumar P. In-silico identification of fingerprint of pyrazolyl sulfonamide responsible for inhibition of *N*-myristoyltransferase using Monte Carlo method with index of ideality of correlation. *J Biomol Struct Dyn*. 2021;39(14):5014–25.
43. Toropova AP, et al. QSAR models for HEPT derivatives as NNRTI inhibitors based on Monte Carlo method. *Eur J Med Chem*. 2014;77:298–305.
44. Sokolović D, et al. Monte Carlo-based QSAR modeling of dimeric pyridinium compounds and drug design of new potent acetylcholine esterase inhibitors for potential therapy of myasthenia gravis. *Struct Chem*. 2016;27:1511–9.
45. Ahmadi S. Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria. *Chemosphere*. 2020;242: 125192.
46. Ghiasi T, et al. The index of ideality of correlation: QSAR studies of hepatitis C virus NS3/4A protease inhibitors using SMILES descriptors. *SAR QSAR Environ Res*. 2021;32(6):495–520.
47. Toropov AA, Toropova AP. The index of ideality of correlation: a criterion of predictive potential of QSPR/QSAR models? *Mutat Res/Genetic Toxicol Environ Mutagen*. 2017;819:31–7.
48. Ahmadi S, Khazaei MR, Abdolmaleki A. Quantitative structure–property relationship study on the intercalation of anticancer drugs with ct-DNA. *Med Chem Res*. 2014;23(3):1148–61.
49. Ahmadi S. A QSPR study of association constants of macrocycles toward sodium cation. *Macroheterocycles*. 2012;5(1):23–31.
50. Ahmadi S, Akbari A. Prediction of the adsorption coefficients of some aromatic compounds on multi-wall carbon nanotubes by the Monte Carlo method. *SAR QSAR Environ Res*. 2018;29(11):895–909.
51. Lawrence J, Lin K. Assay validation using the concordance correlation coefficient. *Biometrics*. 1992;58:599–604.
52. Rücker C, Rücker G, Meringer M. γ -Randomization and its variants in QSPR/QSAR. *J Chem Inf Model*. 2007;47(6):2345–57.
53. Ojha PK, et al. Further exploring rm2 metrics for validation of QSPR models. *Chemom Intell Lab Syst*. 2011;107(1):194–205.
54. Roy K, et al. Comparative studies on some metrics for external validation of QSPR models. *J Chem Inf Model*. 2012;52(2):396–408.
55. Todeschini R. *Milano Chemometrics*. 2010: University of Milano Bicocca, Milano, Italy.
56. da Silva Costa J, et al. Virtual screening and statistical analysis in the design of new caffeine analogues molecules with potential epithelial anticancer activity. *Curr Pharm Des*. 2018;24(5):576–94.
57. Morris GM, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*. 1998;19(14):1639–62.
58. DeVore NM, Scott EE. Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. *Nature*. 2012;482(7383):116–9.
59. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. *J Comput Chem*. 2010;31(2):455–61.
60. Cao DS, et al. A new strategy of outlier detection for QSAR/QSPR. *J Comput Chem*. 2010;31(3):592–602.
61. Ghosh K, et al. Identification of structural fingerprints for ABCG2 inhibition by using Monte Carlo optimization, Bayesian classification, and structural and physicochemical interpretation (SPCI) analysis. *SAR QSAR Environ Res*. 2020;31(6):439–55.
62. Jain S, et al. Exploration of good and bad structural fingerprints for inhibition of indoleamine-2, 3-dioxygenase enzyme in cancer immunotherapy using Monte Carlo optimization and Bayesian classification QSAR modeling. *J Biomol Struct Dyn*. 2020;38(6):1683–96.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

