



Research Article

Development of Predictive *in Silico* Cytotoxic Activity Model to Predict the Cytotoxicity of a Diverse Set of Colchicine Binding Site Inhibitors

 Sumanta Kumar Sahu,  Krishna Kumar Ojha,  Vijay Kumar Singh

Department of Bioinformatics Central, University of South Bihar Gaya, Bihar India

Abstract

Microtubule-targeting agents often have limitations to the development of resistance. Colchicine binding site (CBS) agents have several advantages compared with other tubulin inhibitors. Numerous medications in this class are less susceptible to multidrug resistance that restricts the viability of different inhibitors. In the present study, molecules that bind to the CBS of tubulin are collected from PubMed literature against the A549 cancer cell line. Regression models were established between the descriptor and IC_{50} value of all the compounds present in the training set based on significant molecular fingerprints using multiple linear regression (MLR). Fifteen most significant descriptors selected include Burden modified eigenvalue descriptors, PaDEL-weighted path descriptor, autocorrelation descriptor, topological distance matrix descriptor, MLFER descriptor, Barysz matrix descriptor, chi path cluster descriptor, and validated using internal and external validation parameters. The selected MLR-GA model has $R^{2\text{adjusted}} = 0.7895$, $Q^2_{CV} = 0.76577$, $R^2_{pred} = 0.7419$, and $R^2_{tes} = 0.77373$. An applicability domain is also defined so that it defines the chemical space that the model can predict. The above details suggest a good predictive model for CBS inhibitors that can predict the IC_{50} value of the unknown chemical compound.

Keywords: IC50, Microtubule, Colchicine, Descriptor, Regression, QSAR

Cite This Article: Kumar Sahu S, Kumar Ojha K, Kumar Singh V. Development of Predictive *in Silico* Cytotoxic Activity Model to Predict the Cytotoxicity of a Diverse Set of Colchicine Binding Site Inhibitors. *EJMO* 2022;6(2):172–181.

In a eukaryotic cell, microtubules play an important role in the cell cytoskeleton, formed by α - and β -tubulin heterodimers in a head-to-tail manner. Moreover, Microtubules also play an essential role in cellular functions, such as cytoplasmic and intracellular transport of organelle movement in spindle formation, as well as in the maintenance of cell shape.^[1] The dynamic features of a microtubule's continuous polymerization and depolymerization determine its function.^[2] The duplicated chromosomes of a cell are divided into two identical sets during mitosis before the cell is cleaved into two new daughter cells. Cell cycle

arrest and apoptosis can be induced by disrupting the microtubule dynamic equilibrium.^[1] As a result, microtubules have emerged as an important target for anticancer therapy. Several drugs have been identified as tubulin inhibitors. Microtubule-targeting compounds can be separated into two categories based on their method of action: microtubule-stabilizing agents that increase polymerization and microtubule-destabilizing agents that inhibit tubulin polymerization.^[3]

Tubulin inhibitors can attach to several sites on the tubulin heterodimer, the most common of which are the vinca al-

Address for correspondence: Krishna Kumar Ojha, PhD. Department of Bioinformatics Central, University of South Bihar Gaya, Bihar India

Phone: +916312229530 **E-mail:** kris@cub.ac.in

Submitted Date: September 28, 2021 **Accepted Date:** April 02, 2022 **Available Online Date:** June 06, 2022

©Copyright 2022 by Eurasian Journal of Medicine and Oncology - Available online at www.ejmo.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



kaloid, taxane, and colchicine binding site (CBS).^[2] In comparison to other tubulin inhibitors, microtubule-targeting agents usually have limitations in developing resistance. CBS agents, on the other hand, have various advantages. The CBS is located between the α - and β -tubulin dimers.^[4] Colchicine itself is an FDA-approved drug for gout, but it also has implications for different inflammatory diseases such as familial Mediterranean fever pericarditis and Behcet's disease.^[5] Many drugs that bind to CBS are less sensitive to multidrug resistance (MDR), which reduces the effectiveness of other inhibitors.^[6] The overexpression of class III tubulin, which affects the conformation of the taxane binding site but does not confer resistance to drugs that bind to the CBS, is one of the major causes of drug resistance, suggesting that the CBS may circumvent resistance.^[2] Microtubules are important regulators of endothelial cells and vasculature network formation, and it has been shown that CBS is an attractive target for a tumor's established vasculature.^[7]

Publications related to CBS inhibitors have increased over the last decades.^[2,8-18] Most of the inhibitors were developed to address conventional MDR in the last few years, which gives enough data to build a cytotoxic activity model purely based on known tubulin inhibitors.^[1,4]

QSAR is the most used chemometrics approach for obtaining data that are useful for drug discovery and medicinal chemistry. It relates a molecule's bioactivity to its physical or chemical properties. Descriptors are numerical values that represent all of the physical and chemical attributes stored with the chemical structure. The IC_{50} value represents the response or dependent variable in the QSAR model, whereas these descriptors represent independent variables. In the literature, there are numerous examples of QSAR models being successfully used to screen chemicals for their biological activities.^[19-22]

Previously, many QSAR and pharmacophore models were built to predict the biological activity values of colchicine binding inhibitors and to analyze the mechanism of some particular scaffolds.^[18,23,24] Some models were developed with noncolchicine binding inhibitors and are based on a set of compounds with few scaffolds.^[18] Thus, there is a need for models based on known colchicine binding inhibitors to predict and identify novel colchicine binding inhibitor molecules, which can give a more accurate result.

Tian et al. previously built a 3D-QSAR pharmacophore model using 26 tubulin inhibitors using the HypoGen algorithm; however, there were insufficient data to build a better QSAR model.^[25] Similarly, Zhang et al. built a model employing 116 compounds (e.g., arylthioindoles,

thiazoles, benzo[b]furan, colchicine, and CA4) as tubulin inhibitors, but did not specify which binding site they were targeting, although the binding site plays an essential role.^[18] We built a single QSAR model for predicting the biological activity of heterogeneous compounds that bind to the CBS because earlier QSAR models used only a specific scaffold.

Recent QSAR studies have relied on vast, heterogeneous data sets containing a wide range of structurally varied molecules. A total of 213 chemicals that bind to tubulin's colchicine binding domain were obtained from PubMed literature, which was tested against the A549 cancer cell line. Regression models were developed using multiple linear regression (MLR) based on significant molecular descriptors. Finally, the contribution of several descriptors was studied, and the best prediction model was chosen by comparing model measures, model performance, and accuracy as per the Organization for Economic Cooperation and Development criteria (OECD).

Methods

Figure 1 depicts a flow diagram of the study's workflow. This includes a large-scale QSAR model for predicting and analyzing the inhibitors of the CBS, which was carried out according to OECD guidelines: (i) a defined endpoint for the data set; (ii) an unambiguous learning algorithm; (iii) a defined QSAR model applicability domain (AD); (iv) acceptable goodness of fit, robustness, and productivity measurements; and (v) a mechanistic interpretation of the QSAR model.

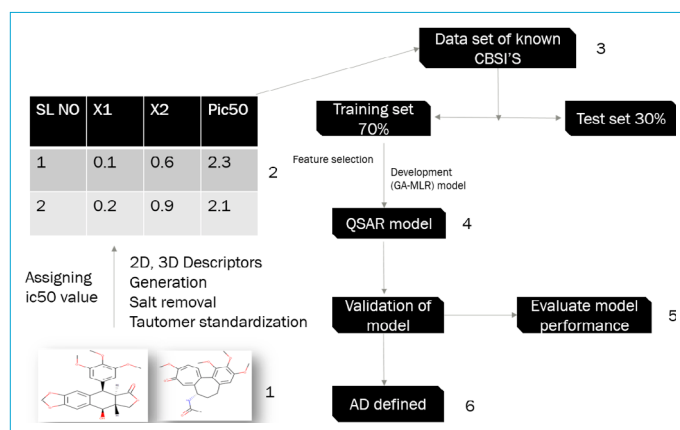


Figure 1. Workflow of QSAR modeling, in which is as follows: (1) molecular structures were collected and curated; (2) 2D and 3D optimization and descriptors generation and assigning pIC50 value to the respective structure; (3) the data set was divided into training and test set; (4) built model on training set data by deleting highly correlated data and selecting significant descriptors; (5) the performance of the model is evaluated by R^2 , Q^2 , RMSE, and other statistical parameters; (6) an applicability domain (AD) is defined.

Data Collection and Activity Evaluation

The publication related to inhibitors of CBS of tubulin increased in the last few years. This gives us the required data to achieve a predictive QSAR model. A total of 213 colchicine binding inhibitors against the A549 cancer cell line were collected from the PubMed literature of the last 10 years (Fig. 2). All of these compounds were divided into three groups, Sublib-I, Sublib-II, and Sublib-III,^[24–40] based on the types of molecules. All of the molecules with their IC_{50} values are presented in Table S1–S17, which are available in supplementary material.

Sublib-I: Combretastatin A4 (CA-4) connecting cis double bond bridge and two hydrophobic rings. Several derivatives were reported by structural modification of CA-4. All of the structures with different structural modifications are represented in the following libraries.

Diaryl-heterocyclic analogues of combretastatin A-4 (Table S1), 5-(2-chlorophenyl)-4-(4-(3,5-dimethoxyphenyl)piperazine-1-carbonyl)-2H-1,2,3-triazole (MAY) (Table S2), 3-O-acylated derivatives (Table S3), 2-morpholin-4-yl-5-nitro-benzoic acid 4-methylsulfanyl benzyl ester derivatives (Table S4), 2-aryl-4-(3,4,5-trimethoxybenzoyl)-5-substituted-1,2,3-triazoles (Table S5), pyridine-bridged CA-4 analogues (Table S6), docking-based virtual screening of CA-4 (Table S7), CA-4 is mediated by metabolic modification of the 3-hydroxy-4-methoxyphenyl's phenolic hydroxyl and ether groups (Table S8), diaryl-heterocyclic analogues of CA-4 (Table S9), 3,6-diaryl-7H-[1,2,4] triazolo [3,4-b][1,3,4]thiadiazines (Table S10).

Sublib-II: Podophyllotoxin is a five-ring system isolated from the roots and rhizomes. Various podophyllotoxin analogues have recently been developed to find novel anti-cancer drugs with improved therapeutic efficacy and overcome medication resistance by structural alteration. These are represented in the following libraries.

Table S11 represents the podophyllotoxin derivatives having C- and E-ring modification and Table S12 represents B-, C-, and E-ring modified structure.

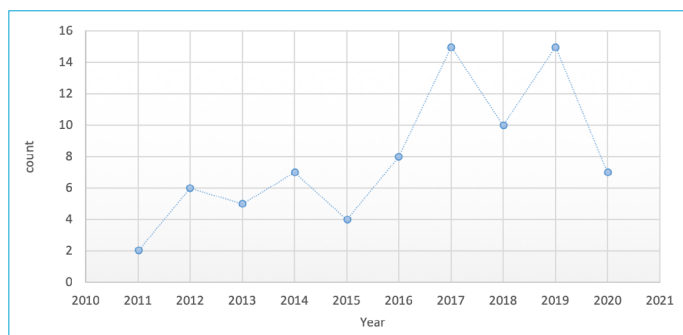


Figure 2. Graph represents an increment of publication related to colchicine binding inhibitors from the Pubmed literature of the last 10 years.

Sublib-III: Other CBS inhibitors (represented in the following libraries).

benzo[c]acridinediones (Table S13), N-aryl-6-methoxy-1,2,3,4-tetrahydroquinoline derivatives (Table S14), 2-phenylquinoline-4-carboxamide derivatives (Table S15), 1-indolyl acetate-5-nitroimidazole (Table S16), [1,2,4]triazolo[1,5-a]pyrimidines (Table S17).

Building of Molecular Structures and Descriptor Calculation

All the collected molecules are 2D represented using MarvinSketch (<https://chemaxon.com/products/marvin>) software and are converted to their 3D representation and optimized geometrically using ChemSketch 3D viewer by acdlab (<https://www.acdlabs.com/>). Each of these compounds had associated in vitro cytotoxic activity values (IC_{50} values) against the A549 cell line.

Here, we generate all the 2D and 3D descriptors using PaDEL-Descriptor software including atom-type electrotopological state descriptors, 2D autocorrelations, WHIM, Petitjean shape index, and count of chemical substructures identified by Laggner. File formats were converted to SD format and saved and exported to PaDEL-Descriptor software V2.20, which is an open Source Software available for the calculation of molecular descriptors and fingerprints. A set of 1400 molecular descriptors was calculated for all the 213 molecules.

Data Preprocessing and Selection of Significant Descriptors

The data set is preprocessed, which is the deletion of missing values. Here, our response value is IC_{50} , but as the unit of all IC_{50} values is in the micrometer range, we convert IC_{50} to pIC_{50} ($-\log IC_{50}$) because higher values indicate exponentially greater potency, so pIC_{50} becomes our Y (dependent variable). We use all the other features as our inputs x (independent variables). The calculated molecular descriptors (x) were normalized by a technique that preserves range (maximum and minimum) which scales the data, so every feature sits in the range 0–1.

The selection of relevant descriptors is most important because the process reduces the number of input descriptors when developing a predictive model which reduces the computational cost of modeling and improves the performance of the model. In the first phase, any parameter that was not calculated (missing value) for any number of compounds in the data set was rejected. Some of the descriptors were discarded because all of the compounds had a zero value (zero tests). To reduce the influence of collinearity and eliminate redundancy, a correlation matrix was created with a cutoff value of 0.9, and variables that showed exact linear dependencies between subsets of

the variables and multicollinearity were removed from the analysis (high multiple correlations between subsets of the variables). A systematic search was performed in the order of missing value test, zero tests, multi co-linearity, and descriptors were discarded by applying a genetic algorithm (GA) to select the best significant descriptors toward the biological activity value.^[41] There are numerous examples in the literature where MLR-GA methods were used successfully for the selection of a significant set of descriptors as powerful search techniques based on the evolution of biological systems for QSAR modeling.^[42-44] GA is a natural variable selection method that belongs to the class of evolutionary algorithm techniques inspired by natural evolution. Here, we set the parameters: number of iterations to 100, equation length (number of descriptors) to 15, mutation probability to 0.3, the initial number of equations generated to 500, and the number of top equations selected in each equation to 30 based on mean absolute error-based criteria.

Building and Evaluating the QSAR Regression Model

The development of a model MLR is an approach that demonstrates a direct link between a dependent variable Y (IC50) and an independent variable x (descriptors). The model is fitted to minimize the sum of square discrepancies between experimental and predicted values of s biological activities. The dependent variable (IC50) Y in regression analysis is dependent on x (descriptors); however, MLR examination expands this concept to include more than one variable, and the regression equation comes in the form:

$$Y = a + c_1x_1(\pm b_1) + c_2x_2(\pm b_2) + c_3x_3(\pm b_3),$$

where Y is the dependent variable, bs are standard errors (SEs) of constant, cs are regression coefficients for corresponding x values (descriptors), and a is a regression constant. In the current work, the models were built with the simple MLR method with the selected variables from GA using DTC-QSAR_v1.0.5 (<https://dtclab.webs.com/software-tools>).

The established QSAR models are judged by the statistical parameters N (number of substances in regression), K (number of descriptors), and R^2 (squared correlation coefficient); Q^2 (cross-validated correlation coefficient), $\text{pred } R^2$ (R^2 for external test set), and F-test (Fischer's value) for statistical significance. The difference in the experimental activity of the data set determined by the regression equation is measured by the regression coefficient R^2 . A QSAR model is thought to be predictive if the accompanying conditions are fulfilled: $R^2 > 0.6$, $Q^2 > 0.6$, and $\text{pred } R^2 > 0.5$. The F-test reflects the proportion of variation clarified by the model, which varies due to the regression error. The F-test has high

estimations, indicating that the model is statistically significant. The model's overall strength is demonstrated by the low standard error of projected R^2 , Q^2 , and $\text{pred } R^2$.

Internal and External Validation of the QSAR Model

The cross-validation technique was used to internally validate the modeled QSAR equation. This method provides more information regarding the predicted reliability of the QSAR equation. In this study, the leave-one-out cross-validation technique was used, and the cross-validated Q^2_{cv} was evaluated using the expression:

$$Q^2_{cv} = 1 - \left[\frac{(Y - Y_{\text{pred}})^2}{(Y - Y_{\text{tr}})^2} \right],$$

where Y_{tr} is the average observed concentration of the training set, Y is the observed concentration, and Y_{pred} is the predicted concentration. The squared correlation coefficient (R^2) was calculated to compare the predicted value obtained by the QSAR equation with the observed concentrations from the experiment, but the issue with R^2 is that it always increases as we add more features to the model, even if they are unrelated to the response. Therefore, choosing the model with the highest R^2 is not a reliable approach for choosing the best model. R^2 adjusted for this is also defined, and unlike R^2 , the error predicted by adjusted R^2 will begin to increase as model complexity becomes very high. R^2 adjusted is defined as

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(N - 1)}{N - P - 1}.$$

Because of its diagnostic means of evaluating the model's predictive power, the cross-validated Q^2_{cv} value is often smaller than the R^2 value of the QSAR model.^[45]

The best combination of training and test data (i.e., Compd ID, descriptor matrix, and response) was fed into the MLR-plusValidation 1.3 (<https://dtclab.webs.com/software-tools>) program, which uses the cross-validation method (leave-one-out) and test set validation based on model acceptable criteria to validate the model internally and externally. Golbraikh and Tropsha provided the following statistical properties of the test set for a better QSAR model with strong predictive power^[46]:

- I. $R^2_{\text{pred}} > 0.6$,
- II. $(r^2 - r^{20})/r^2 < 0.1$,
- III. $0.85 < k < 1.15$ or $0.85 < k' < 1.15$.

Here, R^2 represents the squared correlation coefficient between observed and predicted activities, r^{20} represents the squared correlation coefficient between predicted and observed activities, and k and k' represent the regression slopes passing through the origin.

$$R_{pred}^2 = 1 - \frac{\sum(Y_{pred_{test}} - Y_{obs_{test}})^2}{\sum(Y_{obs_{test}} - \bar{Y}_{training})^2}$$

Here $Y_{pred_{test}}$ and $Y_{obs_{test}}$ are the predicted and observed activities of test set compounds and $Y_{training}$ is the average values of the training set observed activity compounds.

Development of AD

The AD of the proposed model is defined as the area of chemical space compound structure and response where the model can predict accurately. One of the most significant goals of QSAR modeling is to anticipate the activity of new chemical compounds that fall inside the AD of the created model. The accuracy of any QSAR model is dependent on the accuracy of these novel compounds' predictions. The AD of a QSAR model specified the chemical structure space defined by the properties of the molecules in the training set. In this case, the AD is defined by Roy and Kar's basic standardization approach.^[47] The descriptors of the training set should ideally follow a normal distribution pattern. Therefore, according to this, 99.7% of the population set remains within 3 standard deviations (SD), and if the associated standardized descriptors of the compound are larger than ± 3 SD, the compound may be an outlier in the training set or outside AD in the test set.

Mechanistic Interpretation of Feature Importance

Feature importance analysis can assist in identifying features that are crucial for bioactivity. The dependency between the descriptors and the pIC_{50} values is defined by the mutual information (MI) technique using the scikit-learn python mutual information library.^[48,49] It is equal to zero if the two variables are independent and a high value means high dependency.

Results

In silico QSAR analysis selecting descriptors based on the GA, a multilinear regression model was developed containing 15 optimum descriptors. The final selected MLR-GA model is:

$pIC_{50} = 11.95784(\pm 0.98072) - 1.5729(\pm 1.25867) SpMin8_Bhi - 1.09622(\pm 0.49337) VR1_DzZ - 8.7476(\pm 1.48314) ATS6v + 5.09357(\pm 1.40762) SPC-5 - 2.90358(\pm 0.48462) MATS2c + 4.17124(\pm 1.12727) SpMax4_Bhv - 4.23574(\pm 0.61635) ATSC3e - 1.10742(\pm 0.35732) VE3_D + 1.42696(\pm 1.71113) SpMin2_Bhs + 4.88608(\pm 1.01668) MLFER_L - 4.23956(\pm 0.76338) SpMin5_Bhi + 7.67107(\pm 0.70273) WTPT-4 - 4.5866(\pm 0.42955) AATS3e - 1.95703(\pm 0.6097) SpMin3_Bhp - 0.41569(\pm 1.62157) SpMin2_Bhm$

($N_{train} = 153$, $SEE = 0.582$, $R^2_{train} = 0.77372$, $R^2_{adjusted} = 0.7895$, $PRESS = 40.30803$, $F = 27.12717$, $Q^2_{cv} = 0.71863$, average $rm^2(LOO) = 0.62706$, $\Delta rm^2(LOO) = 0.14376$, $N_{test} = 57$, R^2_{pred}

$= 0.7419$, $R^2_{test} = 0.71883$).

From the above model, it can be deduced that the 15 most significant descriptors contained Burden modified eigenvalue descriptor, PaDEL-weighted path descriptor, autocorrelation descriptor, topological distance matrix descriptor, MLFER descriptor, Barysz matrix descriptor, and chi path cluster descriptor. The details are presented in Table 1. The MI between the descriptors and pIC_{50} defines that all the values are greater than zero. Higher the value is more dependent that particular feature to pIC_{50} is.

The values of $R^2_{train} = 0.77372$ and $R^2_{test} = 0.71883$ confirm the good extrapolation between the training and test sets of data. Furthermore, the QSAR model is reliable because of the small variation between R^2 and Q^2_{cv} (<0.5%). Figures 3–6 show a plot of predicted IC_{50} values versus experimental IC_{50} values, as well as the residual between experimental and predicted values. All descriptors with their decreasing order of ranking for dependency toward pIC_{50} are shown in Figure 7, which indicates that descriptor WTPT-4 shows a higher dependency for pIC_{50} .

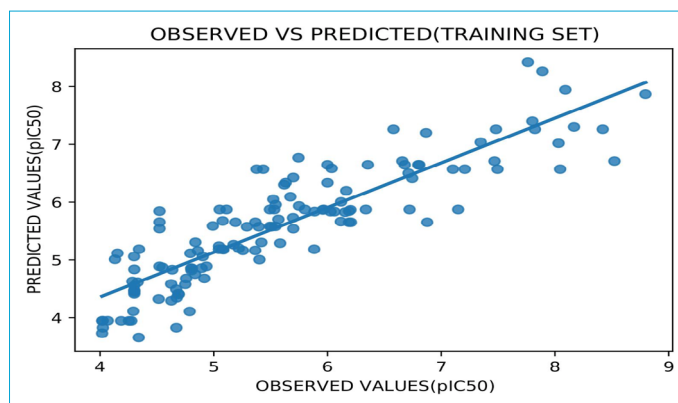


Figure 3. Represents the difference between observed and predicted IC_{50} values of train set compounds.

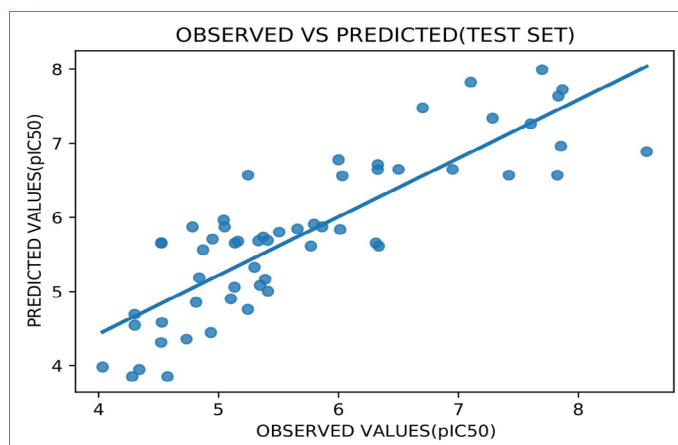
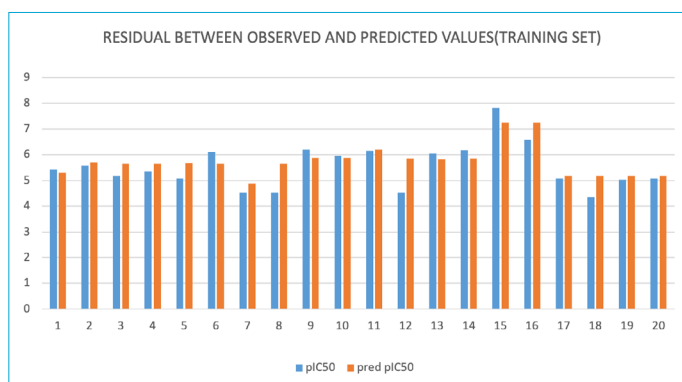
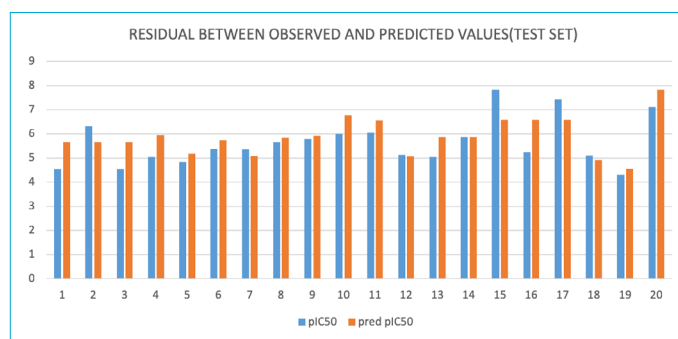


Figure 4. Represents the difference between observed and predicted IC_{50} values of test set compounds.

Table 1. provides a detailed description of the descriptors used in the model. The details about descriptors are taken from the padel descriptor list

Descriptor name	Details of descriptors	Descriptor type	Contribution
SpMin8_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 8 / weighted by relative first ionization potential	Burden modified eigenvalues	Negative
VR1_DzZ	Smallest absolute eigenvalue of Burden modified matrix - n 3 / weighted by relative polarizabilities	Barysz matrix	Negative
ATS6v(2D autocorrelations)	Smallest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative mass	Autocorrelation	Negative
SPC-5(2D)	Simple path cluster, order 5	Chi path cluster	Positive
MATS2c	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative first ionization potential	Autocorrelation	Negative
SpMax4_Bhv	Largest absolute eigenvalue of Burden modified matrix - n 4 / weighted by relative van der Waals volumes	Burden modified eigenvalues	Positive
ATSC3e(2D autocorrelations)	Centred Broto-Moreau autocorrelation of lag 3 weighted by Sanderson electronegativity	Autocorrelation Descriptor	Negative
VE3_D(2D matrix-based descriptors)	logarithmic coefficient sum of the last eigenvector from topological distance matrix	Detour matrix	Negative
SpMin2_Bhs	Topological Distance Matrix Descriptor	Burden modified eigenvalues	Positive
MLFER_L	Solute gas-hexadecane partition coefficient	Molecular linear free energy relation	Positive
SpMin5_Bhi	Smallest absolute eigenvalue of Burden modified matrix - n 5 / weighted by relative first ionization potential	Burden modified eigenvalues	Negative
WTPT-4	Sum of path lengths starting from oxygens	Weighted path	Positive
AATS3e	Average Broto-Moreau autocorrelation - lag 3 / weighted by Sanderson electronegativities	Autocorrelation	Negative
SpMin3_Bhp	Smallest absolute eigenvalue of Burden modified matrix - n 3 / weighted by relative polarizabilities	Burden modified eigenvalues	Negative
SpMin2_Bhm	Smallest absolute eigenvalue of Burden modified matrix - n 2 / weighted by relative mass	Burden modified eigenvalues	Negative

**Figure 5.** The difference between experimental and predicted activity values of twenty numbers of compounds in the training set.**Figure 6.** The difference between experimental and predicted activity values of twenty numbers of compounds in the test set.

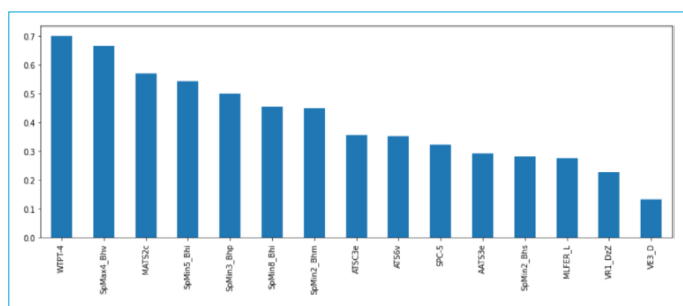


Figure 7. Shows the descriptor dependency towards the dependent variable that is pIC50. Higher the value shows higher dependency.

Discussion

The important step in building a QSAR model is the selection of significant descriptors. The highest-performing model was not necessarily the best. The chemical space is diverse, and the goal of a good machine learning model is to generalize well from the training data. Here, the performance of the model may be less compared with previously built models but it covers all of the chemical spaces without any loss of data. This allows to make predictions accurately the model has never seen. A total 1400 number of descriptors were generated and only 15 optimum descriptors were selected. The whole data set was split randomly into training (70%) and test set (30%). Further, all the calculations are done with training set compounds only. A systematic search performed in the order of missing value test, zero tests, multi co-linearity, and descriptors was discarded by applying a GA to select the best significant descriptors toward the biological activity value.

Previously, a QSAR model was built for CBS using structural diversity of the selected 116 compounds as tubulin inhibitors by Zhang and group using Discovery Studio, who concluded that a new molecule should have strong van der Waals attraction, which affects the activity value. Similarly, the addition of polar groups to the model increases the activity of colchicine. Also, variations of the steric and electrostatic nature of compounds can lead to an increase or decrease in the activity.^[18]

Here, the model was built using freely available tools, built using 213 molecules of heterocyclic scaffolds taken from PubMed literature from the last 10 years reported as CBS inhibitors. Feature importance analysis by mutual information was used to evaluate the relative importance and contribution of each descriptor to the model. The first descriptor in the model, which has the highest contribution is WTPT-4 defined as the sum of path lengths starting from oxygen. It indicates that the presence of oxygen with different path lengths plays a major role in anticancer activity.^[50]

SpMin3_Bhp, SpMin2_Bhm, SpMin5_Bhi, and SpMin8_Bhi

have a negative contribution to the model and are the descriptors of the absolute eigenvalue of Burden modified matrix of $n = 1-8$, weighted relative polarizabilities, relative mass, and relative first ionization potential. The first ionization potential is an atomic property that represents the outermost electronic state. It is an estimation of the energy needed to remove one valence electron from a neutral atom. It indicates that the presence of a less electronegative atom or group increases the anticancer activity, as a less electronegative atom or group requires less ionization potential.^[51]

SpMax4_Bhv is the largest absolute eigenvalue of Burden modified matrix $n = 4$, weighted by relative van der Waals volumes. SpMin2_Bhs is the topological distance matrix descriptor. Increasing the relative van der Waals volumes affect the anticancer activity.^[52]

Similarly, VR1_DzZ, ATS6v, and MATS2c have a negative contribution to the model and are the descriptors of the absolute eigenvalue of Burden modified matrix of $n = 1-8$, weighted by relative polarizabilities, relative mass, and relative first ionization potential. The descriptors are the index of molecular branching with the smallest value corresponding to the chain graph and the highest value to the most branch graph.^[52] As the descriptors are negatively related to the anticancer activity, it may be concluded that decreasing the complexity of the molecule may increase the anticancer activity.^[53,54] SPC-5 is defined as a simple path cluster of order 5 that has a positive impact on the model.^[55]

VE3_D is described as the logarithmic coefficient sum of the last eigenvector from the topological distance matrix. This descriptor encodes electronic, topological, and other geometrical aspects of the compound. The presence of the descriptor in the QSAR model indicates the role of steric and electronic interaction influencing anticancer activity. It utilizes the molecular linear free energy relation.^[56]

AATS3e is negatively correlated to the model defined as average Broto-Moreau autocorrelation-lag3 weighted by Sanderson electronegativity. It is distributed along with the topological structure and calculated at a given spatial lag usually ranging from 1 to 8. It may be concluded that surrounding electronegativity decreases the anticancer property.^[53] MLFER_L is defined as the overall or summation of solute hydrogen bond acidity that has a positive contribution to the model.^[57]

The model with selected descriptors passed the acceptability criteria proposed by Golbraikh and Tropsha (Table 2). Regression statistics, that is the p-value of the model, suggests that the coefficient of the descriptors used in this model are statistically significant at a confidence interval of 95%.

Table 2. represents the acceptability criteria of a predictive model proposed by Golbraikh and Tropsha.

Parameter	Model score	Threshold score
Q^2_{cv}	0.71863	(Threshold value $Q^2 > 0.5$)
R^2_{train}	0.77372	(Threshold value $R^2 > 0.6$)
R^2_{test}	0.72318	(Threshold value $R^2 > 0.6$)
$ r^2_0 - r^2_o $	0.05088	(Threshold value $ r^2_0 - r^2_o < 0.3$)
k	0.98571	(Threshold value: $[0.85 < k < 1.15]$ and)
$[(r^2 - r^2_o)/r^2]$	0.00602	$((r^2 - r^2_o)/r^2) < 0.1$
k	1.00378	$[0.85 < k < 1.15]$
$[(r^2 - r^2_o)/r^2]$	0.07638	$((r^2 - r^2_o)/r^2) < 0.1$

A standardized approach of the AD defined all the compounds of the training set present within the AD except 6 compounds (3, 56, 27, 36, 69, and 75). However, the difference between observed and predicted values is low. As a result, these compounds can be considered influential in fitting the model performance not necessarily outliers to be deleted from the training data set. Similarly, compounds 21, 39, and 5 show outside the AD, but maximum compounds of the test set present within the AD show the confidence within the defined AD.

Conclusion

In this study, we built a QSAR model for the prediction of IC_{50} of the unknown chemical compound for the CBS of tubulin on the training set compounds by using the MLR-GA approach. A set of 213 compounds collected from PubMed literature that binds to CBS of tubulin is used for the model building. From the set of 1400 descriptors, 15 optimum descriptors were selected as highly contributing to the biological activity value. Cross-validation of the model (LOO), Tropsha's metrics, and rm^2 metrics validate the internal and external predictabilities of the model developed using training and test sets. The selected MLR-GA model has $R^2_{train} = 0.77372$, $R^2_{adjusted} = 0.7895$, and $R^2_{pred} = 0.7419$. Also, the evaluation of AD shows that the model is reliable to make predictions within the chemical space for which it is developed.

Disclosures

Ethics Committee Approval: The study was approved by the Local Ethics Committee.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Authorship Contributions: Concept – K.K.O., V.K.S.; Data collection &/or processing – S.K.S.; Analysis and/or interpretation – S.K.S., V.K.S.; Writing – S.K.S., K.K.O.; Critical review – K.K.O., V.K.S., S.K.S.

References

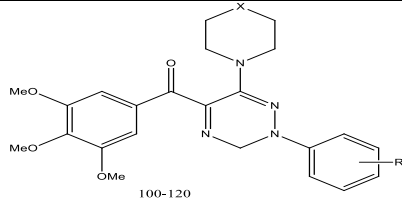
- Nogales E. Structural insights into microtubule function. *Annu Rev Biochem* 2000;69:277–302. [CrossRef]
- Arnst KE, Wang Y, Hwang DJ, Xue Y, Costello T, Hamilton D, et al. A potent, metabolically stable tubulin inhibitor targets the colchicine binding site and overcomes taxane resistance. *Cancer Res* 2018;78:265–77. [CrossRef]
- Field JJ, Díaz JF, Miller JH. The binding sites of microtubule-stabilizing agents. *Chem Biol* 2013;20:301–15. [CrossRef]
- McLoughlin EC, O'Boyle NM. Colchicine-binding site inhibitors from chemistry to clinic: a review. *Pharmaceuticals (Basel)* 2020;13:8.
- Anzengruber F, Graf V, Hafner J, Meienberger N, Guenova E, Dummer R. Efficacy and safety of colchicine in inflammatory skin diseases: a retrospective, monocentric study in a large tertiary center. *J Dermatolog Treat* 2021;32:104–9. [CrossRef]
- Kavallaris M. Microtubules and resistance to tubulin-binding agents. *Nat Rev Cancer* 2010;10:194–204.
- Ren J, Wu L, Xin WQ, Chen X, Hu K. Synthesis and biological evaluation of novel 4β-(1,3,4-oxadiazole-2- amino)-podophyllotoxin derivatives. *Bioorganic Med Chem Lett* 2012;22:4778–82. [CrossRef]
- Manzoor S, Bilal A, Khan S, Ullah R, Iftikhar S, Emwas AH, et al. Identification and characterization of SSE15206, a microtubule depolymerizing agent that overcomes multidrug resistance. *Sci Rep* 2018;8:3305.
- Wang Z, Chen J, Wang J, Ahn S, Li CM, Lu Y, et al. Novel tubulin polymerization inhibitors overcome multidrug resistance and reduce melanoma lung metastasis. *Pharm Res* 2012;29:3040–52. [CrossRef]
- Eissa IH, Dahab MA, Ibrahim MK, Alsaif NA, Alanazi AZ, Eissa SI, et al. Design and discovery of new antiproliferative 1,2,4-triazin-3(2H)-ones as tubulin polymerization inhibitors targeting colchicine binding site. *Bioorg Chem* 2021;112:104965.
- Leoni LM, Hamel E, Genini D, Shih H, Carrera CJ, Cottam HB, et al. Indanocine, a microtubule-binding indanone and a selective inducer of apoptosis in multidrug-resistant cancer cells. *J Natl Cancer Inst* 2000;92:217–24. [CrossRef]
- Mahal K, Resch M, Ficner R, Schobert R, Biersack B, Mueller T. Effects of the tumor-vasculature-disrupting agent verubulin and two heteroaryl analogues on cancer cells, endothelial cells, and blood vessels. *ChemMedChem* 2014;9:847–54.
- Wu X, Wang Q, Li W. Recent advances in heterocyclic tubulin inhibitors targeting the colchicine binding site. *Anti-Cancer Agents in Medicinal Chemistry*. Available at: <https://www.ingentaconnect.com/contentone/ben/acmc/2016/00000016/00000010/art00013>. Accessed May 20, 2022.
- Ansari M, Shokrzadeh M, Karima S, Rajaei S, Fallah M, Ghassemi-Barghi N, et al. New thiazole-2(3H)-thiones containing

- 4-(3,4,5-trimethoxyphenyl) moiety as anticancer agents. *Eur J Med Chem.* 2020;185:111784. [CrossRef]
15. Bai Z, Liu X, Guan Q, Ding N, Wei Q, Tong B, et al. 5-(3,4,5-trimethoxybenzoyl)-4-methyl-2-(p-tolyl) imidazol (BZML) targets tubulin and DNA to induce anticancer activity and overcome multidrug resistance in colorectal cancer cells. *Chem Biol Interact* 2020;315:108886. [CrossRef]
16. Lu Y, Chen J, Xiao M, Li W, Miller DD. An overview of tubulin inhibitors that interact with the colchicine binding site. *Pharm Res* 2012;29:2943–71. [CrossRef]
17. Wang XF, Guan F, Ohkoshi E, Guo W, Wang L, Zhu DQ, et al. Optimization of 4-(N-cycloamino)phenylquinazolines as a novel class of tubulin-polymerization inhibitors targeting the colchicine site. *J Med Chem* 2014;57:1390–402.
18. Li DD, Qin YJ, Zhang X, Yin Y, Zhu HL, Zhao LG. Combined molecular docking, 3D-QSAR, and pharmacophore model: design of novel tubulin polymerization inhibitors by binding to colchicine-binding site. *Chem Biol Drug Des* 2015;86:731–45.
19. Tyagi C, Grover S, Dhanjal JK, Goyal S, Goyal M, Grover A. Mechanistic insights into mode of action of novel natural cathepsin L inhibitors. *BMC Genomics* 2013;14:S10.
20. Li X, Fourches D. Inductive transfer learning for molecular activity prediction: next-gen QSAR models with MolPMoFit. *ChemRxiv.* 2019 Oct 16. Doi: 10.26434/chemrxiv.9978743.v1. [Epub ahead of print]. [CrossRef]
21. Idakwo G, Luttrell IV J, Chen M, Hong H, Gong P, Zhang C. A review of feature reduction methods for QSAR-based toxicity prediction. *Challenges Adv Comput Chem Phys* 2019;30:119–39. [CrossRef]
22. Khan MF, Verma G, Akhtar W, Shaquiquzzaman M, Akhter M, Rizvi MA, et al. Pharmacophore modeling, 3D-QSAR, docking study and ADME prediction of acyl 1,3,4-thiadiazole amides and sulfonamides as antitubulin agents. *Arab J Chem* 2019;12:5000–18.
23. Zhang S, Su L, Zhang X, Li C, Qin W, Zhang D, et al. Combined toxicity of nitro-substituted benzenes and zinc to photobacterium phosphoreum: evaluation and QSAR analysis. *Int J Environ Res Public Health* 2019;16:1041. [CrossRef]
24. Varnek A, Fourches D, Horvath D, Klimchuk O, Gaudin C, Vayer P, et al. ISIDA - Platform for Virtual Screening Based on Fragment and Pharmacophoric Descriptors. *Curr Comput Aided-Drug Des* 2008;4:191–8. [CrossRef]
25. Niu MM, Qin JY, Tian CP, Yan XF, Dong FG, Cheng ZQ, et al. Tubulin inhibitors: Pharmacophore modeling, virtual screening and molecular docking. *Acta Pharmacol Sin* 2014;35:967–79.
26. Liu G, Jiao Y, Huang C, Chang P. Identification of novel and potent small-molecule inhibitors of tubulin with antitumor activities by virtual screening and biological evaluations. *J Comput Aided Mol Des* 2019;33:659–64. [CrossRef]
27. Behbahani FS, Tabeshpour J, Mirzaei S, Golmakaniyoon S, Tayarani-Najaran Z, Ghasemi A, et al. Synthesis and biological evaluation of novel benzo[c]acridine-diones as potential anticancer agents and tubulin polymerization inhibitors. *Arch Pharm (Weinheim)* 2019;352:e1800307.
28. Antúnez-Mojica M, Rodríguez-Salarichs J, Redondo-Horcajo M, León A, Barasoain I, Canales Á, et al. Structural and biochemical characterization of the interaction of tubulin with potent natural analogues of podophyllotoxin. *J Nat Prod* 2016;79:2113–21. [CrossRef]
29. Ahmad FBH, Moghaddam MG, Basri M, Abdul Rahman MB. Anticancer activity of 3-O-acylated betulinic acid derivatives obtained by enzymatic synthesis. *Biosci Biotechnol Biochem* 2010;74:1025–9. [CrossRef]
30. Xu Q, Bao K, Sun M, Xu J, Wang Y, Tian H, et al. Design, synthesis and structure-Activity relationship of 3,6-diaryl-7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazines as novel tubulin inhibitors. *Sci Rep* 2017;7:11997. [CrossRef]
31. Yang F, Yu LZ, Diao PC, Jian XE, Zhou MF, Jiang CS, et al. Novel [1,2,4]triazolo[1,5-a]pyrimidine derivatives as potent antitubulin agents: Design, multicomponent synthesis and antiproliferative activities. *Bioorg Chem* 2019;92:103260.
32. Zheng S, Zhong Q, Mottamal M, Zhang Q, Zhang C, Lemelle E, et al. Design, synthesis, and biological evaluation of novel pyridine-bridged analogues of combretastatin-A4 as anticancer agents. *J Med Chem* 2014;57:3369–81. [CrossRef]
33. Wu GR, Xu B, Yang YQ, Zhang XY, Fang K, Ma T, et al. Synthesis and biological evaluation of podophyllotoxin derivatives as selective antitumor agents. *Eur J Med Chem* 2018;155:183–96.
34. Huang L, Liu M, Man S, Ma D, Feng D, Sun Z, et al. Design, synthesis and bio-evaluation of novel 2-aryl-4-(3,4,5-trimethoxy-benzoyl)-5-substituted-1,2,3-triazoles as the tubulin polymerization inhibitors. *Eur J Med Chem* 2020;186:111846.
35. Zhang X, Rakesh KP, Shantharam CS, Manukumar HM, Asiri AM, Marwani HM, et al. Podophyllotoxin derivatives as an excellent anticancer aspirant for future chemotherapy: A key current imminent needs. *Bioorganic Med Chem* 2018;26:340–55. [CrossRef]
36. Zhu L, Luo K, Li K, Jin Y, Lin J. Design, synthesis and biological evaluation of 2-phenylquinoline-4-carboxamide derivatives as a new class of tubulin polymerization inhibitors. *Bioorganic Med Chem.* 2017;25:5939–51. [CrossRef]
37. Zheng YB, Gong JH, Liu XJ, Wu SY, Li Y, Xu XD, et al. A novel nitrobenzoate microtubule inhibitor that overcomes multidrug resistance exhibits antitumor activity. *Sci Rep* 2016;6:31472.
38. Du J, Li J, Gao M, Guan Q, Liu T, Wu Y, Li Z, Zuo D, Zhang W, Wu Y. MAY, a novel tubulin inhibitor, induces cell apoptosis in A549 and A549/Taxol cells and inhibits epithelial-mesenchymal transition in A549/Taxol cells. *Chem Biol Interact.* 2020 May 25;323:109074. [CrossRef]
39. Duan YT, Sang YL, Makawana JA, Teraiya SB, Yao YF, Tang DJ, et al. Discovery and molecular modeling of novel 1-indolyl acetate - 5-Nitroimidazole targeting tubulin polymerization as

- antiproliferative agents. *Eur J Med Chem* 2014;85:341–51.
40. Wang XF, Wang SB, Ohkoshi E, Wang LT, Hamel E, Qian K, et al. N-aryl-6-methoxy-1,2,3,4-tetrahydroquinolines: A novel class of antitumor agents targeting the colchicine site on tubulin. *Eur J Med Chem* 2013;67:196–207. [\[CrossRef\]](#)
41. Leardi R, Boggia R, Terrile M. Genetic algorithms as a strategy for feature selection. *J. Chemom* 1992;6:267–81. [\[CrossRef\]](#)
42. Ahmadi S, Habibpour E. Application of GA-MLR for QSAR modeling of the arylthioindole class of tubulin polymerization inhibitors as anticancer agents. *Anticancer Agents Med Chem* 2017;17:552–65. [\[CrossRef\]](#)
43. Habibpour E, Ahmadi S. QSAR modeling of the arylthioindole class of colchicine polymerization inhibitors as anticancer agents. *Curr Comput Aided Drug Des* 2017;13:143–59.
44. Ahmadi S, Khazaei MR, Abdolmaleki A. Quantitative structure–property relationship study on the intercalation of anti-cancer drugs with ct-DNA. *Med Chem Res* 2014;23:1148–61.
45. Adeniji SE, Uba S, Uzairu A, Arthur DE. A derived QSAR model for predicting some compounds as potent antagonist against mycobacterium tuberculosis: a theoretical approach. *Adv Prev Med* 2019;2019:5173786. [\[CrossRef\]](#)
46. Alexander DLJ, Tropsha A, Winkler DA. Beware of R2: simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models. *J Chem Inf Model* 2015;55:1316–22.
47. Roy K, Kar S, Ambure P. On a simple approach for determining applicability domain of QSAR models. *Chemom Intell Lab Syst* 2015;145:22–9.
48. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 2004;69:066138. [\[CrossRef\]](#)
49. Pedregosa F, Weiss R, Brucher M. Scikit-learn : machine learning in Python. *JMLR* 2011;12:2825–30.
50. Toppur B, Jaims KJ. Determining the best set of molecular descriptors for a Toxicity classification problem. *RAIRO - Oper Res* 2021;55:2769–83.
51. Adedirin O, Uzairu A, Shallangwa GA, Abechi SE. Optimization of the anticonvulsant activity of 2-acetamido- N -benzyl-2- (5- methylfuran-2-yl) acetamide using QSAR modeling and molecular docking techniques. *Beni-Suef Univ J Basic Appl Sci* 2018;7:430–40. [\[CrossRef\]](#)
52. Adedirin O, Uzairu A, Shallangwa A, Abechi E. A novel QSAR model for designing, evaluating, and predicting the antiMES activity of new 1H-pyrazole-5-carboxylic acid derivatives. *JOTCSA* 2018;4:437–44.
53. Alisi IO, Uzairu A, Abechi SE, Idris SO. Quantitative structure activity relationship analysis of coumarins as free radical scavengers by genetic function algorithm. *Phys Chem Res* 2018;6:208–22.
54. Ramirez-Galicia G, Garduo-Juárez R, Correa-Basurto J, Deeb O. Exploring QSARs for inhibitory effect of a set of heterocyclic thrombin inhibitors by multilinear regression refined by artificial neural network and molecular docking simulations. *J Enzyme Inhib Med Chem* 2012;27:174–86. [\[CrossRef\]](#)
55. Lee W, Park SJ, Hwang JY, Hur KH, Lee YS, Kim J, et al. QSAR model for predicting the cannabinoid receptor 1 binding affinity and dependence potential of synthetic cannabinoids. *Molecules* 2020;25:6057. [\[CrossRef\]](#)
56. Umar AB, Uzairu A, Shallangwa GA, Uba S. QSAR modelling and molecular docking studies for anti-cancer compounds against melanoma cell line SK-MEL-2. *Heliyon* 2020;6:e03640.
57. Bello AS, Uzairu A, Ibrahim MT, Gatugel YA. Quantum modeling analysis of some potent indole derivatives on NS5B Polymerase Inhibitors. *Sci World J* 2019;14:32–7.

Supplementary material

Table S1 series of thiazole-2(3H)-thiones with a 4-(3,4,5-trimethoxyphenyl) moiety as diaryl-heterocyclic analogues of combretastatin A-4 with IC₅₀ value against A549 cancer cell line.

diaryl-heterocyclic analogues of combretastatin A-4 ¹			
			
COMPOUND	X	R	IC ₅₀
100	CH ₂	H	6.45
101	NMe	H	0.622
102	NBoc	H	30
103	NCH(CH ₃) ₂		7.3
104	O	H	0.488
105	NH	H	4.31
106	NAc	H	0.133
107	NAc	2-methyl	8.35
108	NAc	3-methyl	30
109	NAc	4-methyl	0.764
110	NAc	3,4 dimethyl	2.61
111	NAc	2,5 dimethyl	30
112	NAc	2-fluro	4.65
113	NAc	3-fluro	30
114	NAc	4-fluro	30
115	NAc	4-chloro	0.628
116	NAc	4-bromo	9.06

117	NAc	4-methoxy	1.1
118	NAc	4-nitro	0.684
119	NAc	3-nitro-4-methoxy	30
120	NAc	3-amino-4-methoxy	0.648

Table S2 5-(2-chlorophenyl)-4-(4-(3,5-dimethoxyphenyl)piperazine-1-carbonyl)-2H-1,2,3 triazole with IC50 value against A549 cancer cell line.

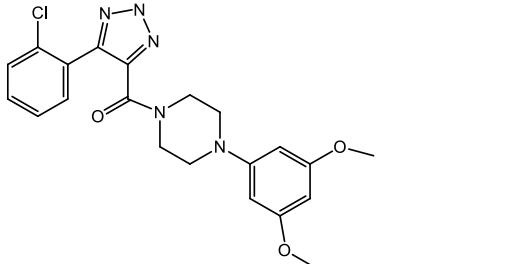
5-(2-chlorophenyl)-4-(4-(3,5-dimethoxyphenyl)piperazine-1-carbonyl)-2H-1,2,3-triazole (MAY) ²		
COMPOUND	STRUCTURE	IC50
7		0.052

Table S3 3-O-acylated derivatives with IC50 value against A549 cancer cell line.

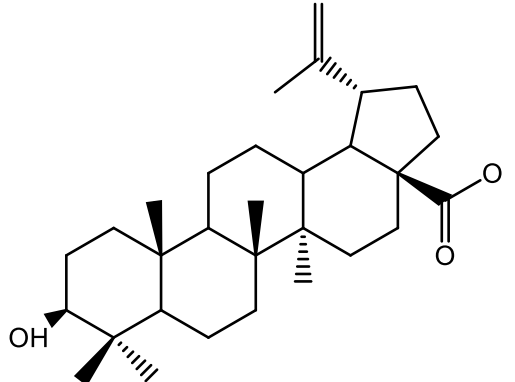
3-O-acylated derivatives ³		
COMPOUND	STRUCTURE	IC50
1		3.816

Table S4 2-morpholin-4-yl-5-nitro-benzoic acid 4-methylsulfanyl benzyl ester (IMB5046) with IC50 value against A549 cancer cell line.

2-morpholin-4-yl-5-nitro-benzoic acid 4-methylsulfanyl benzyl ester derivatives (IMB5046)

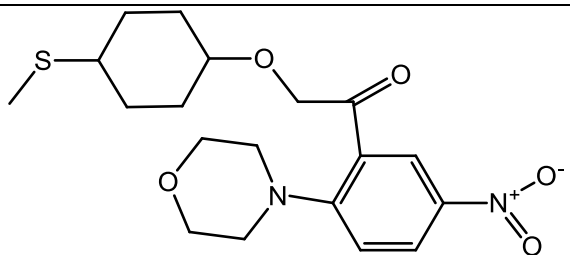
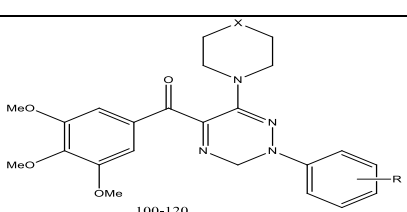
COMPOUND	STRUCTURE	IC50
13		0.199

Table S5 series of 2-aryl-4-(3,4,5-trimethoxybenzoyl)-5-substituted-1,2,3-triazoles with IC50 value against A549 cancer cell line.

2-aryl-4-(3,4,5-trimethoxybenzoyl)-5-substituted-1,2,3-triazoles ¹			
			
COMPOUND	X	R	IC50
100	CH ₂	H	6.45
101	NMe	H	0.622
102	NBoc	H	30
103	NCH(CH ₃) ₂		7.3
104	O	H	0.488
105	NH	H	4.31
106	NAc	H	0.133
107	NAc	2-methyl	8.35
108	NAc	3-methyl	30
109	NAc	4-methyl	0.764
110	NAc	3,4 dimethyl	2.61

111	NAc	2,5 dimethyl	30
112	NAc	2-fluro	4.65
113	NAc	3-fluro	30
114	NAc	4-fluro	30
115	NAc	4-chloro	0.628
116	NAc	4-bromo	9.06
117	NAc	4-methoxy	1.1
118	NAc	4-nitro	0.684
119	NAc	3-nitro-4-methoxy	30
120	NAc	3-amino-4-methoxy	0.648

Table S6 Pyridine-bridged combretastatin-A4 (CA-4) analogues with IC₅₀ value against A549 cancer cell line.

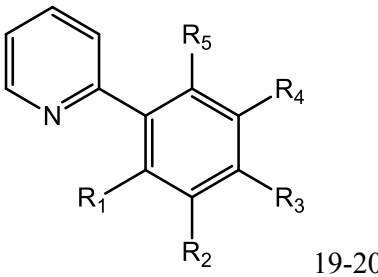
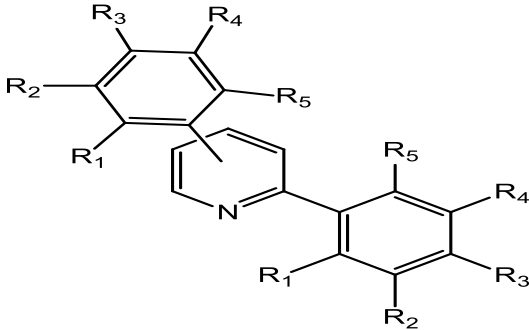
Pyridine-Bridged Analogues of Combretastatin-A4 ⁴						
 <p>19-20</p>						
COMPOUND	R ₁	R ₂	R ₃	R ₄	R ₅	IC ₅₀
19	H	OMe	OMe	OMe	H	9
20	H	H	OMe	H	H	2
21	OMe	H	OMe	H	H	0.079

Table S7 two novel and potent tubulin inhibitors reported by Liu et al with IC₅₀ value against A549 cancer cell line.

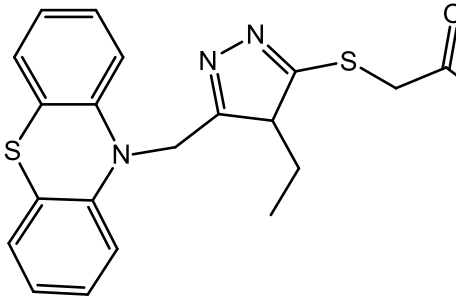
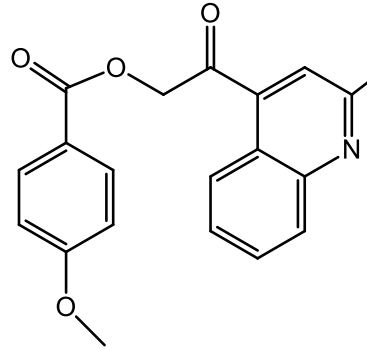
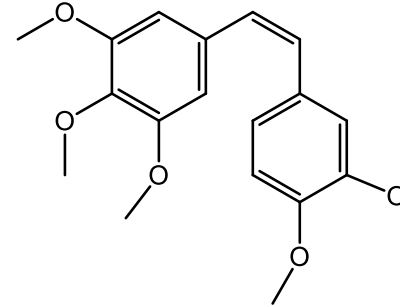
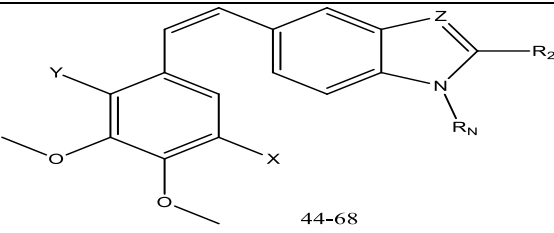
CA-4 analogus ⁵		
 22	 23	 24
COMPOUND	IC50	
22	1.68	
23	2.33	
24	1.78	

Table S8 combretastatin A-4 is mediated by modification of the 3-hydroxy-4-methoxyphenyl's phenolic hydroxyl and ether groups (B ring) with IC50 value against A549 cancer cell line.

CA-4 is mediated by modification of the 3-hydroxy-4-methoxyphenyl's phenolic hydroxyl and ether groups ⁶						
 44-68						
COMPOUND	R _N	X	Y	Z	R ₂	IC50
44	Et	OMe	H	CH	H	0.919
45	Et	OMe	H	C-CHO	H	10
46	Me	OMe	H	CH	H	0.47

47	Me	H	OMe	CH	H	1.7
48	Me	OMe	H	C-CHO	H	0.136
49	Me	H	OMe	CH	H	0.462
50	Me	H	OMe	C-CHO	H	0.045
51	Me	OMe	H	C-CH ₂ OH	H	3.9
52	Me	H	OMe	C-CH ₂ OH	H	0.014
53	Me	OMe	H	C-CH=NOH	H	0.76
54	Me	H	OMe	C-CH=NOH	H	0.0027
55	Me	OMe	H	C-CN	H	0.7
56	Me	H	OMe	C-CN	H	0.471
57	Me	OMe	H	C-COOH	H	3
58	Me	H	OMe	C-COOH	H	0.2092
59	Me	OMe	H	C-COOMe	H	0.519
60	Me	OMe	H	C-CON(Me)(CH ₂) ₂ COOMe	H	0.316
61	Me	OMe	H	C-CONH(CH ₂) ₂ COOEt	H	0.156
62	Me	OMe	H	C-CON(Me)CH ₂ (COOMe) ₂	H	0.112
63	Me	OMe	H	C-COON(Me) (CH ₂) ₂ COOH	H	1
64	Me	OMe	H	C-CONH(CH ₂) ₂ COOH	H	0.442
65	Me	OMe	H	C-Br	H	0.194
66	Me	OMe	H	C-Br	Br	0.034
67	Me	OMe	H	N	H	0.22
68	Me	OMe	H	NMe	H	0.003

Table S9 Diaryl-heterocyclic analogues of combretastatin A-4 with anticancer activity, a series of thiazole-2(3H)-thiones containing the 4-(3,4,5-trimethoxyphenyl) moiety with IC₅₀ value against A549 cancer cell line.

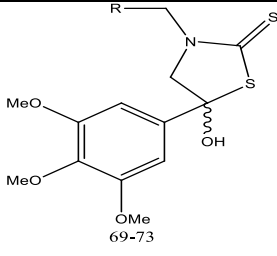
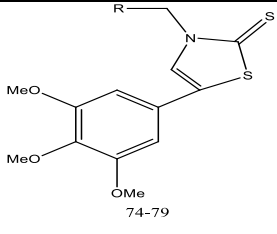
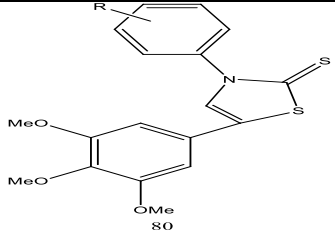
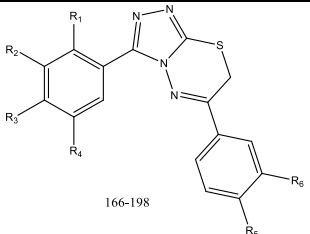
3-Substituted-N-methylindole derivatives ⁷		
		
COMPOUND	R	IC50
69	H	10.2
70	Ph	12.8
71	2-Cl-Ph	14.6
72	4-Cl-Ph	11.6
73	4-F-Ph	21.4
74	H	17.5
75	Ph	12.1
76	2-Cl-Ph	13.8
77	4-Cl-Ph	6.7
78	4-F-Ph	15.8
79	4-OH-Ph	15.6
80	4-Cl	15.4

Table S10 3,6-diaryl-7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazines with IC50 value against A549 cancer cell line.

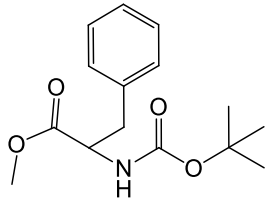
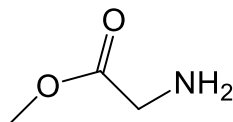
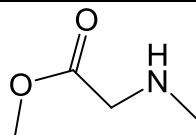
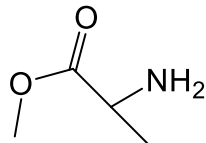
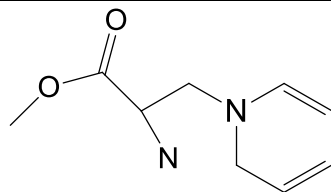
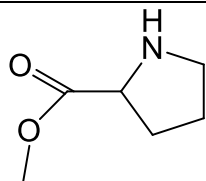
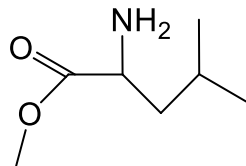
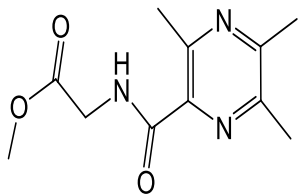
3,6-diaryl- 7H-[1,2,4]triazolo[3,4-b][1,3,4]thiadiazines ⁸


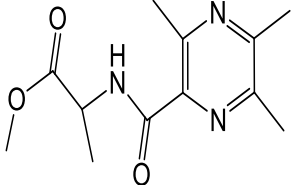
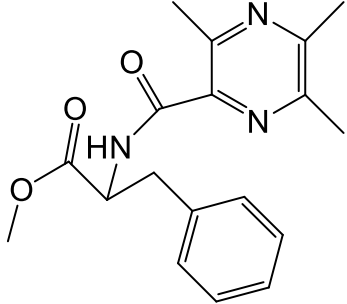
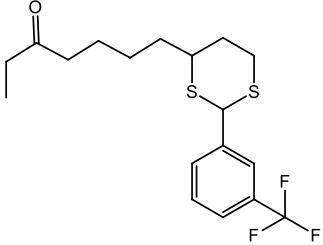
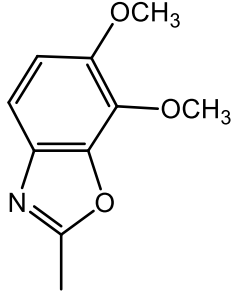
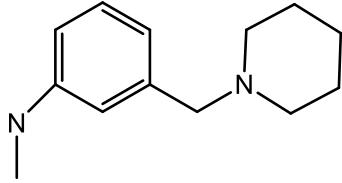
COMPOUND	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	IC50
166	OCH ₃	OCH ₃	OCH ₃	H	F	H	21.2
167	OCH ₃	OCH ₃	OCH ₃	H	Cl	H	23.2
168	OCH ₃	OCH ₃	OCH ₃	H	Br	H	7.37
169	OCH ₃	OCH ₃	OCH ₃	H	CH ₃	H	0.073
170	OCH ₃	OCH ₃	OCH ₃	H	CF ₃	H	85.4
171	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	H	50.2
172	OCH ₃	OCH ₃	OCH ₃	H	SCH ₃	H	12.5
173	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	H	16.3
174	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	NO ₂	16.5
175	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	NH ₂	1.58
176	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	OBn	8.9
177	OCH ₃	OCH ₃	OCH ₃	H	OCH ₃	OH	7.73
178	OCH ₃	OCH ₃	OCH ₃	H	F	F	78.1
179	H	OCH ₃	OCH ₃	OCH ₃	Cl	H	1.37
180	H	OCH ₃	OCH ₃	OCH ₃	CH ₃	H	0.019
181	H	OCH ₃	OCH ₃	OCH ₃	CF ₃	H	0.46
182	H	OCH ₃	OCH ₃	OCH ₃	SCH ₃	H	3.21
183	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	F	8.88
184	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	NO ₂	1.08
185	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	NH ₂	2.94
186	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	OBn	0.071
187	H	OCH ₃	OCH ₃	OCH ₃	OCH ₃	OH	0.19
188	H	OCH ₂ O	OCH ₂ O	H	OCH ₃	F	15.7
189	H	OCH ₂ O	OCH ₂ O	H	OCH ₃	NO ₂	0.062

190	H	OCH ₂ O	OCH ₂ O	H	OCH ₃	NH ₂	16.3
191	H	OCH ₂ O	OCH ₂ O	H	OCH ₃	OBn	0.079
192	H	OCH ₂ O	OCH ₂ O	H	OCH ₃	OH	3.68
193	H	OCH ₃	OCH ₃	H	OCH ₃	NO ₂	0.032
194	H	OCH ₃	OCH ₃	H	OCH ₃	NH ₂	4.2
195	H	H	H	H	OCH ₃	NO ₂	0.015
196	H	OCH ₃	H	H	OCH ₃	NH ₂	5.68
197	H	H	OCH ₃	H	OCH ₃	NO ₂	0.038
198	H	H	OCH ₃	H	OCH ₃	NH ₂	0.009

Table S11 Podophyllotoxin derivatives with IC50 value against A549 cancer cell line.

Podophyllotoxin derivatives ^{9,10}						
COMPOUND	R ₁	R ₂	R ₃	R ₄	IC50	
25		OMe	OMe	OMe	0.0016	
26		OMe	OMe	OMe	0.0173	

27		OCH ₃	OCH ₃	OCH ₃	0.0129
28		OCH ₃	OCH ₃	OCH ₃	0.0068
29		OCH ₃	OCH ₃	OCH ₃	0.0136
30		OCH ₃	OCH ₃	OCH ₃	0.0038
31		OCH ₃	OCH ₃	OCH ₃	0.0028
32		OCH ₃	OCH ₃	OCH ₃	0.0147
33		OCH ₃	OCH ₃	OCH ₃	0.0093
34		OCH ₃	OCH ₃	OCH ₃	0.0201

35		OCH ₃	OCH ₃	OCH ₃	0.0158
36		OCH ₃	OCH ₃	OCH ₃	0.0081
37		OCH ₃	OCH ₃	OCH ₃	6.9
38		OCH ₃	OCH ₃	OCH ₃	1.8
39		OCH ₃	OH	OCH ₃	14.5

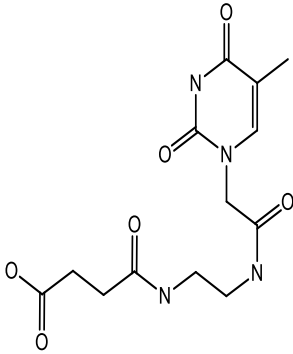
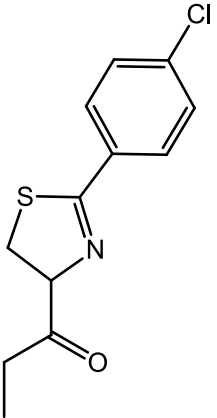
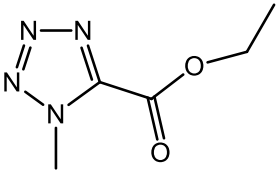
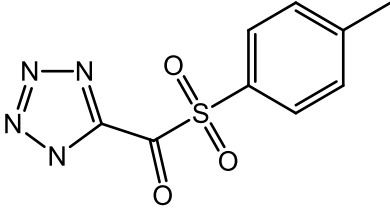
40	-	OCH ₃		OCH ₃	0.27
41		OCH ₃	OCH ₃	OCH ₃	0.18
42		OCH ₃	OCH ₃	OCH ₃	2.4
43		OCH ₃	OH	OCH ₃	3.2

Table S12 Podophyllotoxin derivatives with IC₅₀ value against A549 cancer cell line.

Podophyllotoxin derivatives ¹¹

Compound	R	R ₁	R ₂	IC ₅₀
1	H	OH	OCH ₃	3.81
2	H	OAc	OCH ₃	
3	OCH ₃	H	H	

Table S13 benzo[c]acridinediones with IC₅₀ value against A549 cancer cell line.

benzo[c] acridine-diones ¹²				
Compound	R ₁	R ₂	R ₃	IC ₅₀
125	H	H	H	100
126	H	F	H	8.23
127	H	NO ₂	H	83.2
128	H	CH ₃	H	13.46
129	H	OH	H	45.44
130	H	H	OH	9.41
131	OH	OCH ₃	H	14.5
132	OCH ₃	OH	H	8.36
133	OCH ₃	OCH ₃	H	1.31

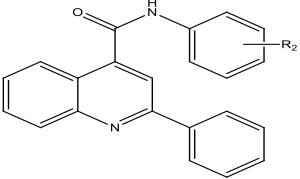
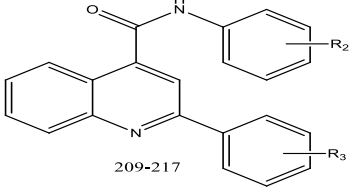
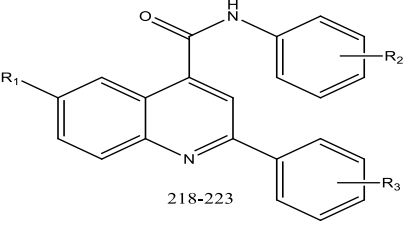
134	OCH ₃	OCH ₃	OCH ₃	100
-----	------------------	------------------	------------------	-----

Table S14 N-aryl-6-methoxy-1,2,3,4-tetrahydroquinoline derivatives with IC₅₀ value against A549 cancer cell line.

N-Aryl-6-methoxy-1,2,3,4-tetrahydroquinolines ¹³					
COMPOUND	X	Y	Z	R ₁	IC ₅₀
159	Br	-	-	NHMe	
160	Br	-	-		1.6
161	Br	-	-		1
162	Me	CH	N	-	1
163	Me	N	N	-	1
164	Cl	N	N	-	0.93
165	NHMe	N	N	-	0.21

Table S15 2-phenylquinoline-4-carboxamide derivatives with IC₅₀ value against A549 cancer cell line.

2-phenylquinoline-4- carboxamide derivatives ¹⁴

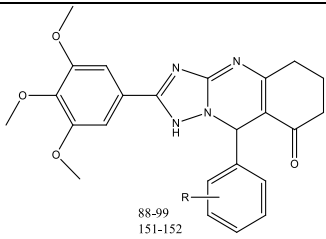
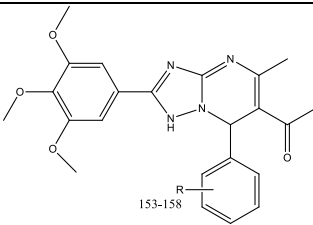
 <p>203-208</p>	 <p>209-217</p>			 <p>218-223</p>
COMPOUND	R ₁	R ₂	R ₃	IC ₅₀
203	-	4' -OCH ₃	-	20.5
204	-	4' -OCH ₂ H ₃	-	50
205	-	3'4'5' -OCH ₃	-	7.9
206	-	3' -SCH ₃	-	50
207	-	2' -CH ₃ , 4'5' -OCH ₃	-	23.7
208	-	2' -Cl , 4'5' -OCH ₃	-	50
209	-	4' -OCH ₃	4' -CH ₃	18.5
210	-	3'4'5' -OCH ₃	4' -CH ₃	5.7
211	-	2' -CH ₃ , 4' -OCH ₃	4' -CH ₃	11.6
212	-	4' -OCH ₃	3' -CH ₃	30.5
213	-	3'4'5' -CH ₃	3' -CH ₃	17.8
214	-	2' -CH ₃ , 4' -OCH ₃	3' -CH ₃	23.6
215	-	4' -OCH ₃	2' -CH ₃	50
216	-	2' -CH ₃ , 4' -OCH ₃	2' -CH ₃	45.7
217	-	2' -CH ₃ , 4' -OCH ₃	4' -CH ₃	50
218	-CH ₃	4' -OCH ₃	4' -CH ₃	46.7
219	-CH ₃	3'4'5' -CH ₃	4' -CH ₃	29.7
220	-CH ₃	2' -CH ₃ , 4' -OCH ₃	4' -CH ₃	20.3
221	-Cl	4' -OCH ₃	H	30.1
222	-Cl	3'4'5' -CH ₃	H	50

223	-F	2' -CH ₃ , 4' -OCH ₃	H	50
-----	----	--------------------------------------------	---	----

Table S16 1-indolyl acetate-5-nitroimidazole with IC50 value against A549 cancer cell line.

1-indolyl acetate e-5-Nitroimidazole ¹⁵		
<p style="text-align: center;">134-150</p>		
COMPOUND	R	IC50
135	o-cl	4.24
136	m-Cl	5.02
137	p-Cl	4.51
138	o-Br	5.11
139	m-Br	6.03
140	p-Br	5.55
141	o-NO ₂	2.12
142	m-NO ₂	2.86
143	p-NO ₂	2.01
144	o-CH ₃	4.02
145	m-CH ₃	4.32
146	p-CH ₃	3.88
147	o-CH ₃	3.11
148	m-OCH ₃	4.09
149	P-OCH ₃	3.97
150	2,4-2-Cl	2

Table S17 [1,2,4]triazolo[1,5-a]pyrimidines with IC50 value against A549 cancer cell line.

[1,2,4] triazolo[1,5-a]pyrimidine derivatives ¹⁶		
 <p>88-99 151-152</p>		 <p>153-158</p>
COMPOUND	R	IC50
88	4-CH ₃	53.14
89	4-F	92.59
90	4-NO ₂	50.57
91	3-F,4-CH ₃	94.83
92	3-F, 4-OCH ₃	85.64
93	3,5-Br ₂	52.46
94	3,4-Cl ₂	26.67
95	2-NO ₂ , 4-OCH ₃	94.47
96	4-Br	21.33
97	4-CF ₃	55.73
98	4-CH ₂ CH ₃	65.21
99	4-CN	95.67
151	3-F	45.82
152	3- NO ₂	73.93
153	4-F	5.63
154	4-NO ₂	1.02
155	3-F, 4-CH ₃	95.86
156	3-NO ₂ , 4-OCH ₃	16.08
157	3,5-Br ₂	28.28
158	3,4-Cl ₂	70.27