Design of novel benzylidenecyclohexenones derivatives as potential anti-cancer inhibitors using 3D-QSAR, pharmacokinetics, and molecular docking studies.

Mohammed Er-rajy (errajy.mohammed@yahoo.com)
Universite Sidi Mohamed Ben Abdellah Faculte des Sciences Dhar El Mahraz-Fes
https://orcid.org/0000-0001-5460-684X
Nidal Naceiri Mrabti
Universite Sidi Mohamed Ben Abdellah Faculte des Sciences Dhar El Mahraz-Fes
Sara Zarougui
Universite Sidi Mohamed Ben Abdellah Faculte des Sciences Dhar El Mahraz-Fes
Menana Elhallaoui
Universite Sidi Mohamed Ben Abdellah Faculte des Sciences Dhar El Mahraz-Fes

Research Article

Keywords: 3D-QSAR, benzylidenecyclohexenones, molecular docking, ADMET, Drug design

Posted Date: April 13th, 2022

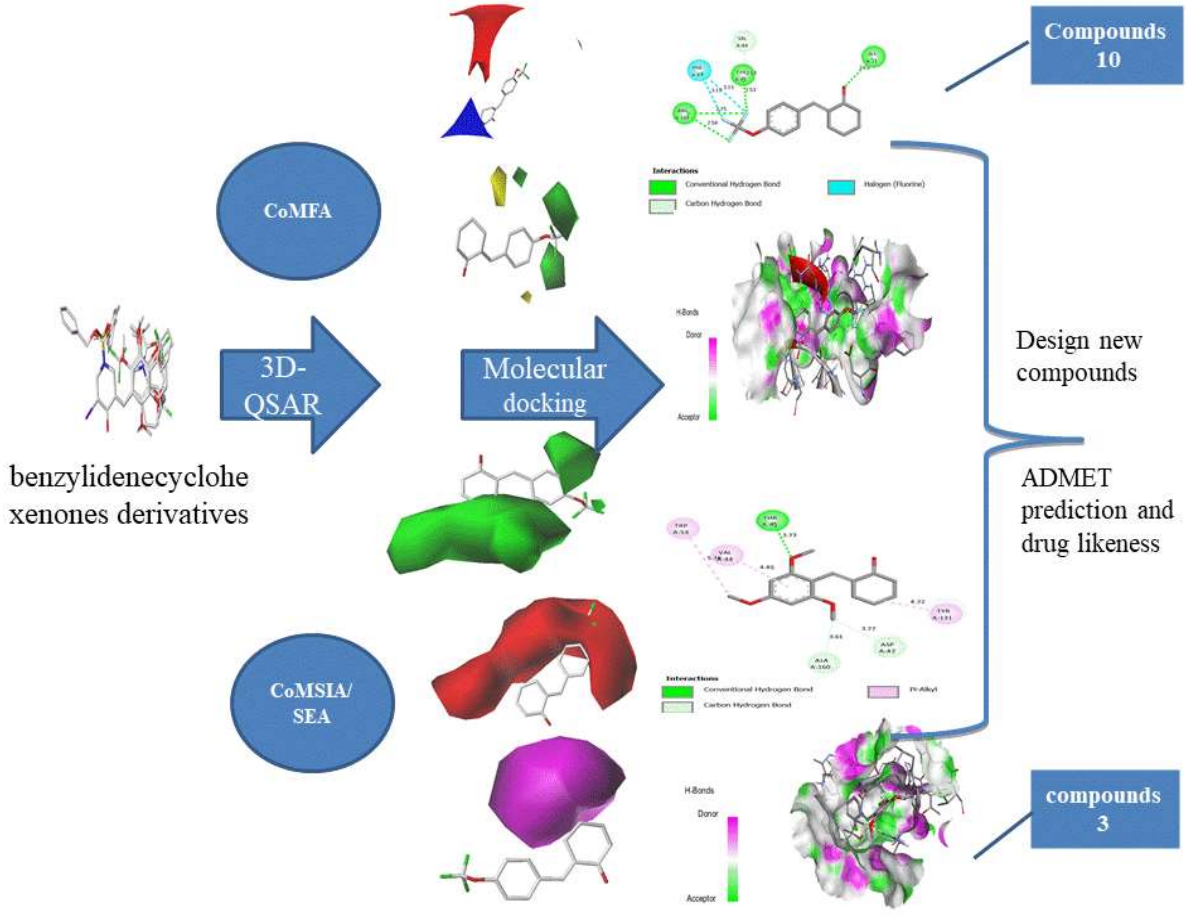
DOI: https://doi.org/10.21203/rs.3.rs-1510373/v1

Design of novel benzylidenecyclohexenones derivatives as potential anti-cancer inhibitors using 3D-QSAR, pharmacokinetics, and molecular docking studies.

Mohammed Er-rajy [1]*, Mohamed El fadili [1], Nidal Naceiri Mrabti [1], Sara Zarougui [1], Menana Elhallaoui [1].
[1] LIMAS Laboratory, Faculty of Sciences Dhar El Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco.
*errajy.mohammed@yahoo.com

Abstract

The 3D-QSAR models were established in this study based on comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA), the optimal CoMFA model established gave Q^2= 0.671, R^2= 0.925 and R^2_{pred} = 0.868, and the best CoMSIA/SEA model gave Q^2= 0.627, R^2= 0.775, and R^2_{pred} = 0.962. The predictive ability of the developed models was evaluated by external and internal validation. In this study, steric, electrostatic, and hydrogen bond acceptor fields played a key role in the anti-cancer activity. Molecular docking results theoretically revealed the importance of residues ARG164 and THR45 in the active site of the TrxR enzyme. Based on these results, we designed several new inhibitors, and their inhibitory activities were predicted by the best model (CoMFA). In addition, these new inhibitors were analyzed for their ADMET properties and their similarity to drugs. These results will be of great help for the optimization of new anti-cancer drug discovery.



Keywords: 3D-QSAR, benzylidenecyclohexenones, molecular docking, ADMET, Drug design.

Introduction

Thioredoxin reductase (TrxR) is found in all known types of living things, TrxR is the only known enzyme that can reduce thioredoxin[1]. TrxR, together with thioredoxin (Trx) and reduced nicotinamide adenine dinucleotide phosphate (NADPH) form the most important antioxidant systems, which is called the thioredoxin system. Thioredoxin is small redox protein present in almost all known organisms and is essential in mammals[2]. The thioredoxin system plays a critical role in many signaling pathways related to cell proliferation and survival[3], and also permits upregulation in a variety of tumors, including those associated with chemoresistant cancers. Quantitative structure-activity relationship (QSAR) and ADMET studies have been very important methods for predicting new drug candidates with relatively low economic investment costs[4]. The QSAR methodology allows quantitative correlation between variations in biological activity (pIC_{50}) and molecular structures or properties. Generally, the 3D-QSAR study statistically analyzes the relationship between the three-dimensional structure information of molecules and the biological activity of compounds to establish a valid mathematical model, Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Index Analysis (CoMSIA) are the two most commonly used methods to build the models. To evaluate the predictive power of the constructed 3D-QSAR models, external and internal validation were discussed to check for outliers in the training set and test set. Molecular docking studies were performed to predict the optimized binding conformation of a ligand and to understand the structural interaction between a ligand and its protein. Then, we evaluated their ability to act as a drug; each designed compound was tested by pharmacokinetic parameters (ADMET) and by the ability to act as a drug. Based on the results of this work, new compounds were proposed to develop new potential TrxR inhibitors[5].

Results and discussion

Molecular alignment

The molecular alignment method is the sensitive step used to build a weak 3D-QSAR model. The set of molecule derivatives was aligned to the common core by the simple Sybyl X-2.1 alignment technique, using the most active compound (compound N°3) as a template. According to Fig.1, all the 3D molecular structures of the training and test sets are superimposed on the common core.

CoMFA and CoMSIA studies

The activity values predicted by the CoMFA and CoMSIA models and their residuals for the data set are presented in the table 2.

Based on the CoMFA descriptor, 3D-QSAR model was proposed to explain and quantitatively predict the effects of steric and electrostatic fields of substituents on the anticancer activity of a series of twenty-six benzylidenecyclohexenones. The independent variables obtained from the training set were subjected to a PLS cross-validation analysis to identify the appropriate statistical parameters for each module. According to the table, the CoMFA model has a contribution of steric and electrostatic fields explaining 70.5% and 29.5% of the variance respectively, the cross-validated correlation coefficient Q^2 value of the training set and non-cross-validated correlation coefficient R^2 are 0.671 and 0.925, respectively, and F-test value is 69.778, a high predictive value for the external validation of test set (R^2_{pred} = 0.868), the optimal number of principal components using to generate the CoMFA model is 3.

In the CoMSIA study, the evaluation of the analysis of the 4 selected models with different field combinations, such as steric (S), electrostatic (E), hydrophobic (H), hydrogen bond donor (D), and hydrogen bond acceptor (A), represented in table 3. According to the table 3, Among the different combinations of fields chosen, the best model is SHA, which obtained the highest Q^2 value of 0.627 with only one component, and a value of R^2 = 0.775, F value of 65.389 and SEE value of 0.257, the

contributions of steric, hydrophobic and hydrogen bond acceptor fields were 0.159, 0.267 and 0.574 respectively, and also the CoMSIA/SEA model shows a large external prediction value of the test set ($R^2 = 0{,}962$).
The two correlation plots between the observed biological activity and the biological activity predicted by both CoMFA and CoMSIA/SEA models showed a strong correlation, as shown in the Fig.2. Therefore, the constructed CoMFA and CoMSIA/SEA models are robust and can be used to design new compounds and predict the activities of new compounds in the future.

External Validation

The 3D-QSAR models were considered as reliable predictive models, if the values of the statistical parameters of the models, such as R^2, Q^2, and R^2_{pred} are greater than 0.6, 0.5, and 0.6, respectively[6]. Both CoMFA and CoMSIA/SEA models indicate a favorable stability estimate and good predictive quality, which was confirmed by the predictive ability of the external validation. To verify these results, we tested the robustness and predictability of the best models we chose. The results obtained by external validation test calculated by the two models CoMFA and CoMSIA/SEA are presented in table 4.

According to the results in table 4, both CoMFA and CoMSIA/SEA models passed all the Golbraikh and Tropsha criteria, so both models have better predictive power for the new compounds and it meets all the validation methods. To better understand the important regions in the 3D space around the molecules, we used the CoMSIA/SEA contour maps.

Analysis of CoMFA and CoMSIA Contour Maps

To visualize the information contained in the two 3D-QSAR models, we chose the most active compound 3 of the series as a reference. The steric, electrostatic and hydrogen bond acceptor fields of the CoMSIA and CoMFA contour maps are shown in both Fig.3 and 4.

CoMFA Contour Map

In the steric contour maps, the green contour (80%) indicates the region where the bulky groups promote activity, and the yellow contour indicates (20%) the region where the bulky groups were unfavorable to activity. For the electrostatic fields, the blue contours (80%) indicate the regions where electron-donating groups increase activity, while the red contours (20%) indicate the regions where electron-drawing groups increase activity. Fig.3 shows the steric and electrostatic contours of the CoMFA model. The green and yellow contours indicate steric interactions, and the red and blue contours indicate electrostatic interactions.

The two contour maps (steric and electrostatic) from the CoMFA study can help provide information on regions that may decrease or increase the biological activity of benzylidenecyclohexenones derivatives. Three large green contour maps are located in the proximity of positions 3 and 4 in the benzene ring of compound 3, indicating that large groups at these two positions would be favorable for increasing the inhibitory activity. A small yellow region near position 1 in the benzene cycle of the same compound also suggests that a large group at this position can reduce activity.

CoMSIA Contour Map

In the CoMSIA model, three different fields called steric, electrostatic and hydrogen bond acceptor fields were analyzed. The three contour maps are shown in Fig.4. Steric (a), electrostatic (b), and hydrogen bond acceptor (c) fields.

In the steric contour map (a) of CoMSIA, we observed two broad green contours covering benzene position 3, indicating that the selection of the bulky substitution group is necessary in this region to have better activity.

In the CoMSIA electrostatic contour maps, we observed a large red contour authoring molecule 3, indicating that the addition of an electron-withdrawing group or atoms may increase the activity.
In the hydrogen bond acceptor field, the magenta and red contours identified favorable and unfavorable positions for hydrogen bond acceptors. In Fig.4(c), a magenta outline around R1 indicated that a hydrogen bond acceptor substituent at these sites would increase activity.

Applicability domain

The Williams plot is constructed by taking the normalized residual values on the Y-axis and the leverage values on the X-axis (Fig.5). The number of descriptors is 3 and the number of compounds in the training set is 21, the limits of the normalized residual r ± 2.5, so the critical leverage value (h*) is [2.5(3 +1)]/21= 0.47.

The aim of the applicability domain is to find the compounds outside the DA of the constructed CoMSIA/SEA model. From Fig.5, it can be concluded that two test set compounds (2 and 22) have leverage values above the critical value of 0.47. However, the normalized residual for each compound in the training and test set is below the limits of the normalized residuals (r = ± 2.5). It is possible that the activity value of these two compounds was not correctly predicted due to the availability of incorrect experimental data for these two molecules. For this reason, we will remove molecule 2 and 22 from the list of molecules we propose to use as novel inhibitors of TrxR enzyme activity, and also exclude these molecules from molecular modeling for the remainder of this study.

Y-Randomization

The 6 new 3D-QSAR models obtained, each new module has a new value of R^2 and Q^2.The results obtained are presented in table 5.

The Y-randomization method was performed to validate and evaluate the possibility of the fortuitous correlation in the CoMFA and CoMSIA/SEA models. Multiple random mixing of the dependent variable (pIC50) was performed, then after each mixing, a 3D-QSAR model was developed and the results obtained are presented in Table 5. The low Q^2 and R^2 values obtained after each shuffle indicate that the good results of our original CoMFA and CoMSIA/SEA models are not driven by the chance correlation of the training set.

Molecular docking

We performed molecular docking to better understand the inherent potency of anticancer activity and also to discover the types of ligand-receptor interactions. The interaction modes obtained by molecular docking for compounds 3 and 10 are shown in Fig.6.

After visual observation of the two results, there are three hydrogen bonds with the residues ARG 164, THR 45 and GLY 23 with a distance equal to 2.56, 2.53 and 2.19, respectively, and one halogen bond with the residue PHE 43, with a distance equal to 3.15 in the most active molecule number 3. In the molecule number 10 we observe a hydrogen bond with the residues THR 45 and three of alkyl bonds with the residues TRP 53, VAL 44 and TYP 131. The three fluorine of the molecule number 3, establishes two hydrogen bonds which allows stabilizing this molecule.

Design of new compounds

Based on the proposed CoMFA models, new compounds are designed using compound 3 as a model. The four new compounds were minimized and aligned with the database. According to the equation obtained by 3D-QSAR (CoMFA) method, the predicted anti-cancer activity values of the proposed new compounds can be calculated. The obtained result of predicted pIC_{50} activity values and structure of the new compounds (A1, A2, A3 and A4) present in table 6.

According to the table 6 the four new designed molecules have good predicted activities. To ensure viable drug design, the four predicted compounds were evaluated for their synthetic accessibility and were examined for their ADMET properties in silico.

Lipinski's rule and ADMET prediction

The main objective of this study is to predict the drug-like properties of the best synthetic molecules and four molecules that we have proposed as new drugs (A1, A2, A3, A4), and then to study their toxicity on the organism if they are used in pharmaceutical applications, in order to identify the molecules with drug-like properties.

The results obtained in table 7, indicate that all the compounds have verified all the rules of Lipinski, Veber, and Egan, which means that all the compounds have good oral bioavailability. The compounds are also evaluated for their synthetic accessibility, from the results obtained; we concluded that all the compounds have easy to synthesize characteristics. Therefore, it can be said that all compounds tend to have low attrition during clinical trials and have a better chance of reaching the market.

ADMET prediction and drug likeness

Some studies show that most of the molecules fail in clinical trials due to toxicity or poor pharmacokinetics[7]. After checking the drug similarity of all the compounds, they were subjected to ADMET prediction, the following table 8 presents the different pharmacokinetic (absorption, distribution, metabolism, excretion) and pharmacodynamics (drug efficacy and toxicity) properties of the five compounds.

Water solubility is given in log (mol/L) (insoluble $\leq 10 <$ sparingly soluble $\leq 6 <$ moderately soluble $\leq 4 <$ soluble $\leq 2 <$ very soluble $< 0 <$ very soluble)[8], this means that all compounds are very soluble in water; intestinal absorbance values are greater than 90% means that all compounds have good absorbance; Caco-2 permeability is commonly used to predict absorption of orally administered drugs, a compound considered extremely Caco-2 permeable should give expected values > 0.90[9]. It is evident from the Caco-2 values in table 8 that the five selected compounds can be considered highly Caco-2 permeable.

The volume of distribution at steady state (VDss)[10] is a parameter to characterize the distribution of drugs in various tissues in vivo. When the value of log (VDss) is less than -0.15, the volume of distribution is considered relatively low, and if the value of log (VDss) is greater than 0.45, the volume of distribution is considered relatively high. Compounds 3, A1, A2, and A3 are valued between -0.15 and 0.45 which means that the four compounds have a medium distribution of the drug in various tissues in vivo, but compound A4 has a log (VDss) value of -0.28, which means that compound A4 considered relatively low to the distribution of drugs in various tissues in vivo.

Blood-brain barrier (BBB) permeability is a very important property in the pharmaceutical field because it determines whether or not a compound can cross the BBB and thus exert its therapeutic effect on the brain, the standard value of blood-brain barrier (BBB) permeability is good if its value is greater than 0.3 and bad if it is less than -1. From the table 8, it is clear that compound A2 has a good BBB permeability, but the other molecules have an average BBB permeability.

For the CNS permeability index, compounds with values (LogPS) greater than -2 are considered able to penetrate the CNS, and unable to penetrate the CNS if the LogPS values are less than -3[11], according to table 8, we conclude that all compounds can penetrate the CNS.

The cytochrome P450 isoenzymes[12] are important for drug metabolism in the liver. CYP2D6 and CYP3A4 are two isoforms of cytochrome P450, which is an enzyme responsible for crucial detoxification in the human body and responsible for altering drug pharmacokinetics. According to the results obtained, all the tested molecules become substrates of CYP3A4, but no compound becomes a substrate of CYP2D6. At the level of inhibitors of CYP1A2 and CYP2C19, almost all the tested

compounds become inhibitors of CYP1A2, and CYP2C19, but the only compound that becomes an inhibitor of CYP1A2 is the compound A1.

The total clearance of the drug gives information about the half-life of the drug, a low value of clearance mean that the half-life of the compounds is large, according to the results obtained, all compounds with a low clearance value which means that the half-life of the drug is high for these compounds.

The Ames toxicity test and the skin sensitization test are commonly used tools to determine the toxicity of compounds; most compounds are non-toxic, with the exception of compound A1.

In conclusion, based on the drug and ADMET similarity studies, we propose compounds number 3, A2 and A3, which have good absorption, distribution and metabolism properties, low total clearance and no toxic properties, as TrxR inhibitors.

Conclusion

This work aims to study the structural requirements of a series of benzylidenecyclohexenone derivatives acting as anticancer agents. Molecular modeling is a common method used for drug discovery, with a low economic investment cost. The excellent predictive power of the CoMFA and CoMSIA/SEA models observed for a set of test compounds indicates a significant statistical quality of the models obtained, and the significance was confirmed by external and internal validation methods, showing that these models can be successfully used to predict the activity of newly designed molecules. The correlation between the different fields and activity is well evidenced by the contour maps of the CoMFA, CoMSIA/SEA, and molecular docking models, which led to the design of four new compounds (A1-4) that may be useful for the subsequent design of new TrxR inhibitors. ADMET prediction showed that the four new compounds could represent good drug candidates for anti-cancer treatment.

Material and methods

Database and biological activity

3D-QSAR models were performed on a set of 26 benzylidenecyclohexenone derivatives were taken from the literature[13]. The biological anti-cancer activities were expressed as effective concentration IC_{50} (μM). These values were converted to pIC_{50} values ($6-\log_{10}(IC_{50})$) to construct the 3D-QSAR models (Table 1). To make the 3D-QSAR models, we divided a dataset containing 26 compounds randomly into a training set (21 compounds) to build the models and a test set (5 compounds) to test the performance of the constructed models (Table 1).

Molecular modeling and alignment

Molecular alignment is a very important step[14], and the most sensitive for CoMFA and CoMSIA analysis[15]. In this work, we make a simple alignment that is available on the SYBYL-X 2.1 software.

Each molecular structure, built with the sketched module, was minimized under the standard Tripos force field[16], using the appropriate Gasteiger-Huckel atomic partial charges on the SYBYL-X 2.1 platform[17]. In addition, 0.001 kcal/ (mol Å) was set as a convergence criterion for the Powell gradient algorithm and 100 iterations to obtain a stable conformation[18].

Generation of 3D-QSAR models

The CoMFA and CoMSIA field studies were performed using Sybyl X-2.1 software (Tripos, Inc., USA). CoMFA descriptors were calculated for each network with a grid spacing of 1 Å and extending to 4 Å units in all three dimensions (3D) within the defined area. Steric and electrostatic fields were calculated using the default Tripos force field and a Van Der Waals potential, as well as coulombic terms, an sp^3 hybridized carbon atom with a +1e charge was used as the probe atom, the energy values of steric and electrostatic fields were also truncated at 30 kcal/mol[19].

For CoMSIA, a distance-dependent Gaussian-like physicochemical property was also adopted to avoid any singularity at the atomic position. Similar standard parameters and no random limits constructed for the CoMFA field were applied to the CoMSIA field calculation, including steric (S), electrostatic (E), and hydrophobic (H) effects. The attenuation factor and the filtering and attenuation factor of the column were set to the default value of 0.3, and 2.0 kcal/ mol respectively[20].

Partial least squares (PLS) analysis

The PLS method[21] was used to evaluate a linear correlation between the electrostatic, steric, hydrophobic, hydrogen bond donor (D), and hydrogen bond acceptor (A) properties, each with TrxR inhibitory activity. In the PLS analysis, the leave-one-out (LOO) cross-validation method[22] is used to determine the optimal number of components (NOC), which was then used to derive the final CoMFA and CoMSIA models. To assess the significance of the models, we relied on the following statistical parameters: the cross-validation correlation coefficient (Q^2), the coefficient of determination (R^2), the F-value (Fischer test) and the standard error of estimation (SEE)[23]. Finally, the different CoMFA and CoMSIA models were generated using a non-validated PLS analysis. To choose a good module, we need the largest possible statistical parameters Q^2 and R^2 and the smallest possible SEE.

External validation of the CoMFA and CoMSIA models

To evaluate the predictive ability of the 3D-QSAR models, the biological activities of the external test set (5 compounds) were predicted by the 3D-QSAR models, and then the predictive correlation coefficient (R^2pred), which is based on the test set molecules, was calculated using the following equation $R^2pred = (SD-PRESS)/SD$ [24](SD: The sum of squared deviations between the pIC50 of the test set compounds and mean pIC50 of the training set compounds, and PRESS: The sum of squared deviations between observed(pIC_{50} obs) and predicted activity(pIC_{50} predict) of the test set compounds). For this validation method which is based on Golbraikh and Tropsha criteria[25], Golbraikh and Tropsha proposed several parameters to check the external predictability of the QSAR model. In addition, Kunal Roy[26] proposes other additional statistical parameters for external validation, such as r_m^2 and $r_m'^2$, which are calculated as follows:

$$r_m^2 = r^2\left(1 - \sqrt{(r^2 - r_0^2)}\right)$$

$$r_m'^2 = r^2\left(1 - \sqrt{(r^2 - r_0'^2)}\right)$$

$r^2$and r_0^2 are the squared correlation coefficient values between the observed and predicted values of the test set compounds with and without intercept, respectively and $r_0'^2$, is squared correlation coefficient values for the plot of observed versus predicted activity for the test set at zero intercept. For a model with good external predictability, the values of r_m^2, and $r_m'^2$ should be greater than 0.5[27].

Applicability domain

A 3D-QSAR model has been developed on a limited number of compounds, which do not occupy the entire chemical space[28]. The applicability domain (AD) of a QSAR model is defined as a region of the modeled response (pIC_{50}) and model variables (molecule properties) from the test set and the training set[29]. There are different approaches to determine the DA of QSAR models; the most commonly used being the leverage method. The leverage value (hi) $h_i = X_i^T(X^TX)^{-1}$ (i = 1. . . n) where X_i is the row vector of descriptor values of a compound i, and X is the descriptor matrix, X^T is the transpose of X. The critical value h* is usually set to 3(d+1)/n, where n is the number of chemical compounds in the training set and d is the number of descriptors[29]. A chemical compound is considered out of the applicability domain when the amount of leverage (hi) of this chemical compound is higher than the critical value h*[28]. Conversely, a chemical compound is considered to

be within the applicability domain, when the (h_i) is less than the h* value[30]. The AD analysis was performed using MATLAB software version 2011[31].

Y-Randomization

The pIC_{50} randomization test is typically used to confirm the strength and robustness of the generated models[32]. The dependent variable pIC_{50} is randomly shuffled while the matrix of the independent variable (descriptors) remains unchanged. The process of calculating the 3D-QSAR models is repeated several times, which allows obtaining new models for every iteration. The low values of Q^2 and R^2 indicate the efficiency and reliability of the optimal models[33].

Molecular Docking

Molecular docking is a technique used to model the interactions of a ligand with its protein. This is used by software such as Discovery 2020[34] to remove water molecules and visualize the ligand-protein interaction, Auto Dock Tools[35], and Vina to perform molecular docking. The TrxR-1 enzyme (coded in PDB 3EAN) is downloaded as a PDB (Protein Data Bank)[36] from the protein site (https://www.rcsb.org/) with a resolution of 2.75 Å. The three-dimensional configuration encompasses the active site (X= 29.07Å, Y= 17.21Å, Z= 125.65Å), with a grid size of 40*40*40 points in the x, y, and z directions[37]. After the preparation of the ligand and receptor, we performed the molecular docking on compound 3 which is considered the most active, and compound 10 which is considered the least active (anti-cancer).

In silico pharmacokinetics ADMET and drug likeness prediction

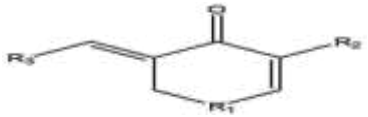
Drug similarity can be defined as a qualitative concept used in drug design to determine how "drug-like" a substance is compared to other substances[38]. When estimating the drug similarity of the design compound (A1 to 4), it is necessary to consider the compliance of the physical and chemical properties of the compound with filter variants, such as Lipinski's five rules (RO-5)[39], Ghose's filter[40], Veber's rule[41], Egan's rule[42], and Muegge's Closely[43] and synthetic accessibility rule[44], to calculate these properties, by using the swissadmet server[45]. Since the experimental evaluation of ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties[46] of small molecules is time-consuming and expensive, computational approaches can rapidly optimize the pharmacokinetic properties and toxicity of drug candidates. We used the pkCSM webserver[47] to calculate the ADMET properties of the 4 novel compounds, such as absorption in the human gut, VDss (volume of distribution at steady-state), penetration into the blood-brain barrier (BBB)[48], and central nervous system (CNS)[49], metabolism, total drug clearance, and toxicity levels of the novel molecules[50].

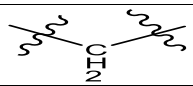
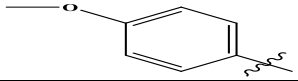
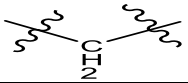
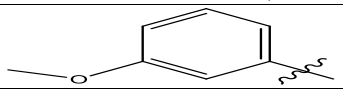
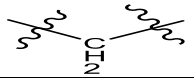
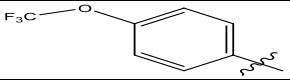
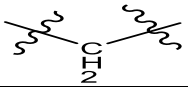
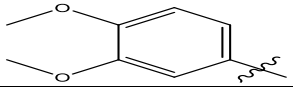
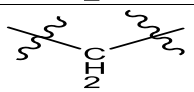
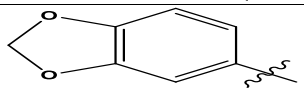
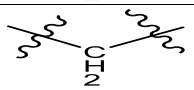
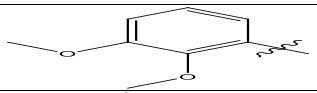
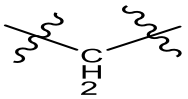
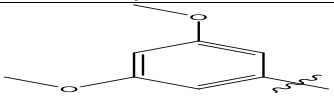
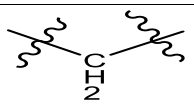
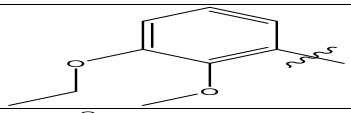
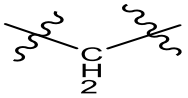
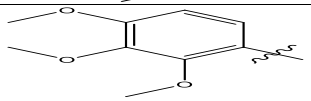
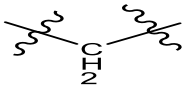
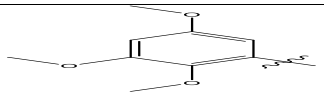
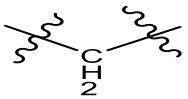
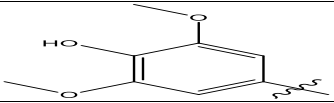
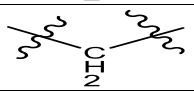
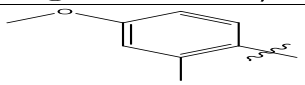
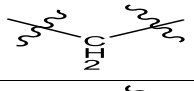
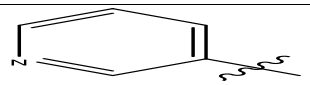
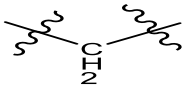
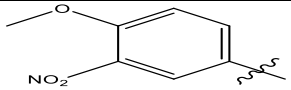
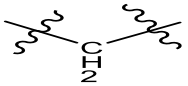
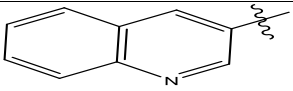
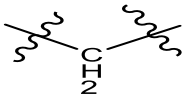
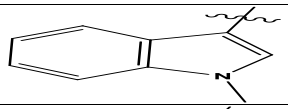
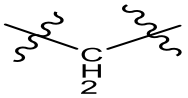
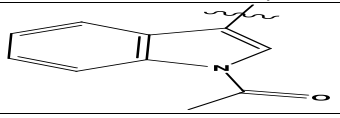
Compliance with ethical standards

Conflict of interest No conflict of interest exits in the submission of this paper, and paper is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part.

Tables

Table 1 Structures and biological values of benzylidenecyclohexenone derivatives.



Compound	R_1	R_2	R_3	IC_{50}	pIC_{50}
1		H		1.62	5.79
2*		H		1.06	5.97
3		H		0.59	6.23
4*		H		0.64	6.19
5		H		1.65	5.78
6*		H		0.61	6.21
7		H		0.6	6.22
8		H		0.6	6.22
9		H		0.67	6.17
10		H		5.07	5.29
11		H		1.61	5.79
12		H		1.07	5.97
13		H		1.05	5.98
14		H		0.59	6.23
15		H		0.64	6.19
16		H		3.37	5.47
17		H		0.93	6.03

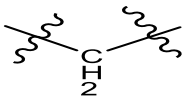
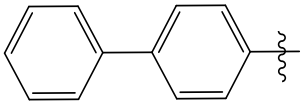
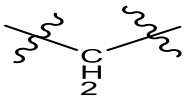
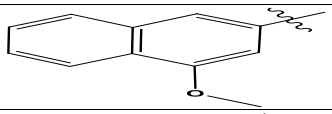
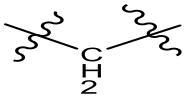
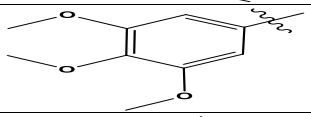
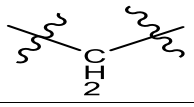
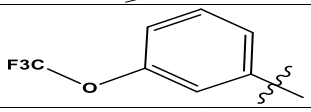
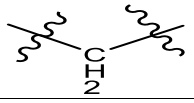
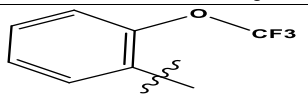
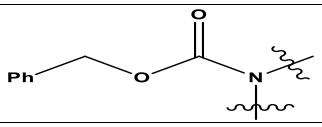
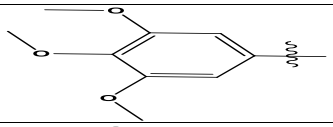
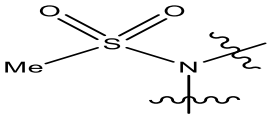
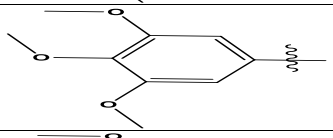
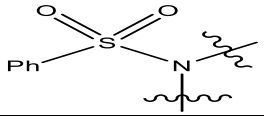
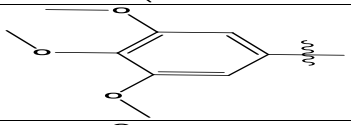
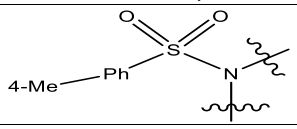
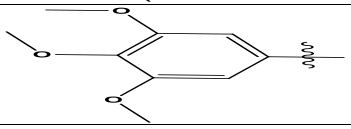
18		H		0.68	6.17
19		H		1.63	5.79
20		I		0.67	6.17
21		H		0.82	6.09
22*		H		1.01	6
23		H		24.2	4.62
24		H		15.9	4.8
25*		H		23.3	4.63
26		H		22.2	4.65

Table 2 The observed and predict pIC_{50} and their residuals of the studied compounds for CoMFA and CoMSIA analyses.

No	pIC_{50} obs	pIC_{50} predict					
		CoMFA	Residual	CoMSIA-SEA	Residual	CoMSIA-SE	Residual
1	5.79	5.913	-0.123	6.022	-0.232	5.949	-0.159
2*	5.97	5.856	0.114	5.94	0.030	5.797	0.173
3	6.23	6.207	0.023	6.071	0.159	6.139	0.091
4*	6.19	6.246	-0.056	5.947	0.243	6.269	-0.079
5	5.78	6.025	-0.245	6.012	-0.232	6.015	-0.235
6*	6.21	5.735	0.475	5.954	0.256	5.400	0.810
7	6.22	6.055	0.165	5.919	0.301	6.085	0.135
8	6.22	6.102	0.118	5.972	0.248	6.155	0.065
9	6.17	6.115	0.055	5.841	0.329	6.152	0.018
10	5.29	5.273	0.017	5.851	-0.561	5.216	0.074
11	5.79	5.958	-0.168	5.869	-0.079	5.927	-0.137
12	5.97	5.901	0.069	6.009	-0.039	5.915	0.055
13	5.98	6.038	-0.058	6.038	-0.058	5.968	0.012
14	6.23	6.089	0.141	6.146	0.084	6.165	0.065
15	6.19	6.033	0.157	5.981	0.209	6.025	0.165
16	5.47	5.784	-0.314	5.977	-0.507	5.773	-0.303
17	6.03	5.768	0.262	5.952	0.078	5.809	0.221
18	6.17	6.186	-0.016	6.084	0.086	6.209	-0.039
19	5.79	5.825	-0.035	5.92	-0.130	5.857	-0.067
20	6.17	6.321	-0.151	5.871	0.299	6.216	-0.046
21	6.09	5.983	0.107	5.998	0.092	6.002	0.088

22*	6.00	6.107	-0.107	6.029	-0.029	6.000	0.000
23	4.62	4.718	-0.098	4.889	-0.269	4.525	0.095
24	4.8	4.753	0.047	4.667	0.133	4.772	0.028
25*	4.63	4.698	-0.068	4.596	0.034	4.825	-0.195
26	4.65	4.601	0.049	4.561	0.089	4.775	-0.125

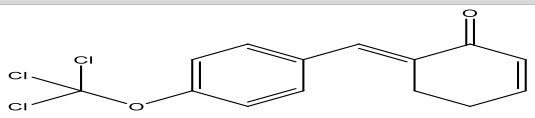
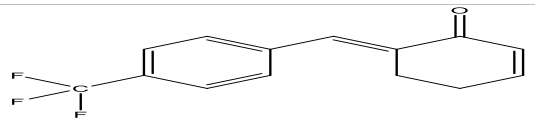
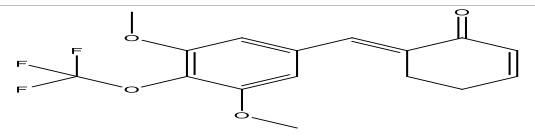
Table 3 PLS results Statistics of CoMFA and CoMSIA models.

	Q^2	R^2	SEE	F	NOC	R^2_{pred}	Fraction					
							Ster	Elec	Hyd	Donn	Accp	Hydr
CoMFA	0.671	0.925	0.157	69.778	3	0.868	0.705	0.295	-	-	-	-
CoMSIA/SE	0.609	0.937	0.144	83.807	3	0.64	0.352	0.648	-	-	-	-
CoMSIA/SEA	0.627	0.775	0.257	65.389	1	0.962	0.159	0.267	-	-	0.574	-
CoMSIA/SED	0.609	0.937	0.144	83.807	3	0.644	0.352	0.648	-	0.00	-	-
CoMSIA/SEH	0.441	0.971	0.101	132.535	3	0.884	0.258	0.433	-	-	-	0.31

Table 4 Statistical parameters for the validation of CoMFA and CoMSIA/SEA model.

Parameter	Formula	Threshold	CoMFA	CoMSIA/SEA
$Q^2_{training}$		$Q^2_{training} > 0.5$	0.671	0.627
r^2	Coefficient of determination for the plot of predicted versus observed for test set by MLR.	$r^2 > 0.6$	0.868	0.962
r_0^2	r^2 at zero intercept		0.868	0.961
$r_0'^2$	r^2 for the plot of observed versus predicted activity for the test set at zero intercept		0.848	0.956
$\lvert r_0^2 - r'^2_0 \rvert$		$\lvert r_0^2 - r'^2_0 \rvert < 0.3$	0.02	0.005
k	Slope of the plot of predicted versus observed activity for test set at zero intercept	$0.85 < k < 1.15$	1.012	1.019
$\frac{r^2 - r_0^2}{r^2}$		$\frac{r^2 - r_0^2}{r^2} < 0.1$	0.00	0.001
k'	Slope of the plot of observed versus predicted activity at zero intercept	$0.85 < k' < 1.15$	0.986	0.981
$\frac{r^2 - r_0'^2}{r^2}$		$\frac{r^2 - r_0'^2}{r^2} < 0.1$	0.23	0.006
r^2_m		$r^2_m > 0.5$	0.868	0.952
r'^2_m		$r'^2_m > 0.5$	0.850	0.956

Table 6 Structures and their predicted pIC_{50} of newly designed compound.

New compound	Structure	Pred pIC_{50}
		CoMFA
A1		6.669
A2		6.685
A3		6.594

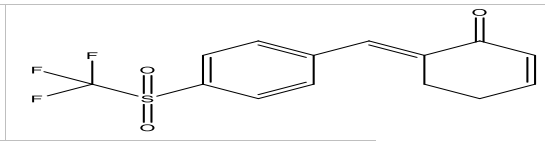
A4		6.67

Table 7 Violation of Lipinski's rule and synthetic accessibility.

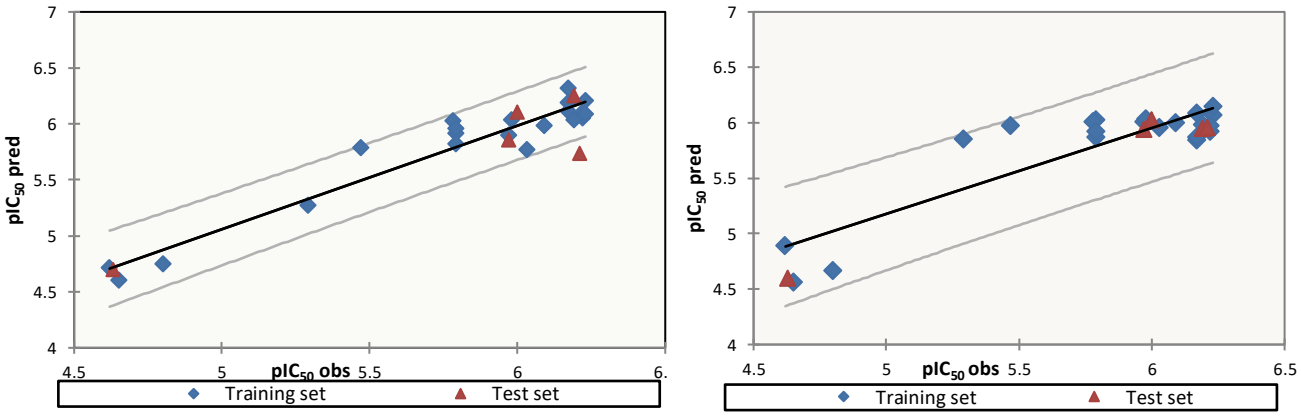
Numbers of Compounds	Property						violations			S.A
	MW	n-HA	n-HD	n-Rot	TPSA	LogP	Lipinski's	Veber	Egan	
Threshold	< 500	< 10	< 5	≤ 10	≤ 140	≤ 5	≤ 1		≤ 1	≤ 1	0 < S.A<10
3	270.25	2	0	3	26.30	3.66	0		0	0	2.50
A1	317.59	2	0	2	26.30	4.69	0		0	0	2.66
A2	252.23	1	0	1	17.07	4.01	0		0	0	2.68
A3	328.28	4	0	4	44.76	3.90	0		0	0	3.01
A4	316.30	3	0	2	59.59	3.28	0		0	0	2.80

Table 8 The results of the ADMET test.

N° of compounds	Absorption			Distribution			Metabolism					Excretion	Toxicity	
	Water solubility	Intestinal absorption (human)	Caco2 permeability	VDss (human)	BBB permeability	CNS permeability	Substrate	Inhibitor				Total Clearance	AMES toxicity	Skin Sensitization
							CYP							
							2D6	3A4	1A2	2C19	2C9			
	Numeric (log mol/L)	Numeric (%Absorbed)	Numeric (log Papp in 10^{-6} cm/s)	Numeric (Log L/kg)	Numeric (Log BB)	Numeric (Log PS)	Categorical (Yes/No)					Numeric (Log ml/min/kg)	Categorical (Yes/No)	Categorical (Yes/No)
3	-4.705	92.73	1.61	0.21	0.172	-2.234	No	Yes	Yes	Yes	No	-0.023	No	No
A1	-5.596	91.834	1.61	0.282	0.064	-2.234	No	Yes	No	Yes	Yes	-0.564	No	Yes
A2	-4.964	93.075	1.48	0.42	0.425	-2.083	No	No	Yes	Yes	No	-0.047	No	No
A3	-5.217	93.795	1.46	-0.072	-0.15	-2.588	No	Yes	Yes	Yes	No	0.011	No	No
A4	-4.405	93.594	1.42	-0.281	0.059	-2.027	No	Yes	Yes	Yes	No	0.224	No	No

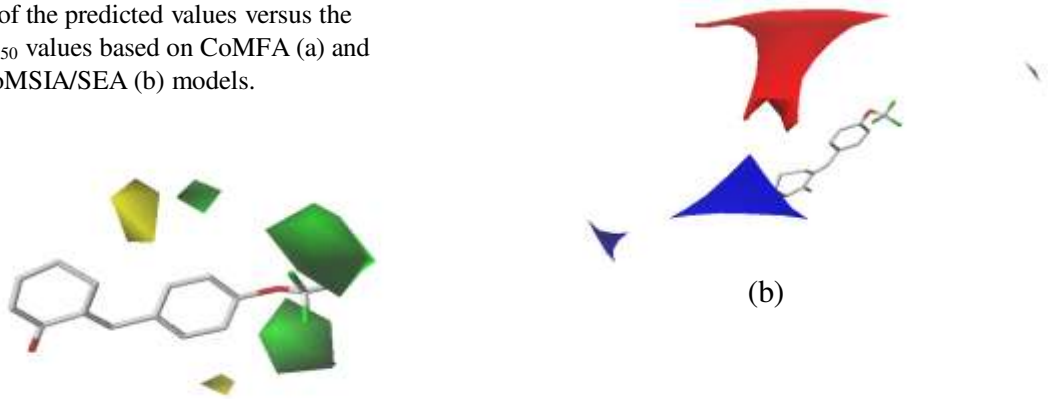
Schemes and Figures

Fig.1: 3D-QSAR structure superposition of the core and alignment of the training and testing set



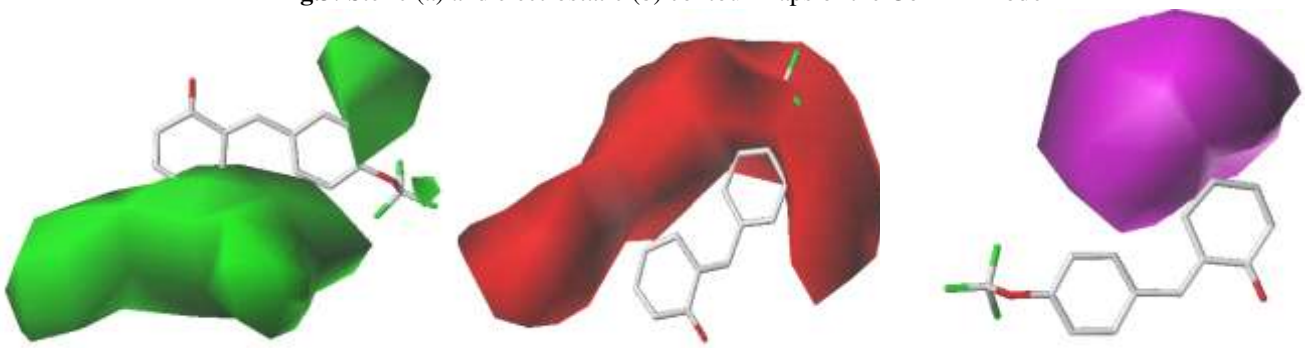
(a) (b)

Fig.2: Plot of the predicted values versus the observed pIC_{50} values based on CoMFA (a) and CoMSIA/SEA (b) models.



(a) (b)

Fig.3: Steric (a) and electrostatic (b) contour maps of the CoMFA model



(a) (b) (c)

Fig.4: (a) Steric, (b) Electrostatic and (c) Hydrogen bond acceptor Contour maps of CoMSIA analysis with 2 Å grids spacing in combination with compound 3.

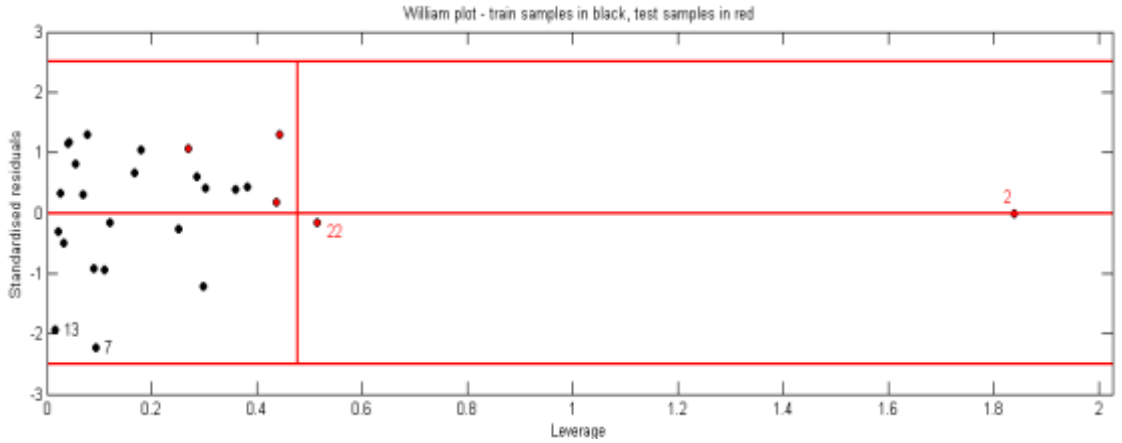


Fig.5: William plot for the developed CoMSIA/SEA model.

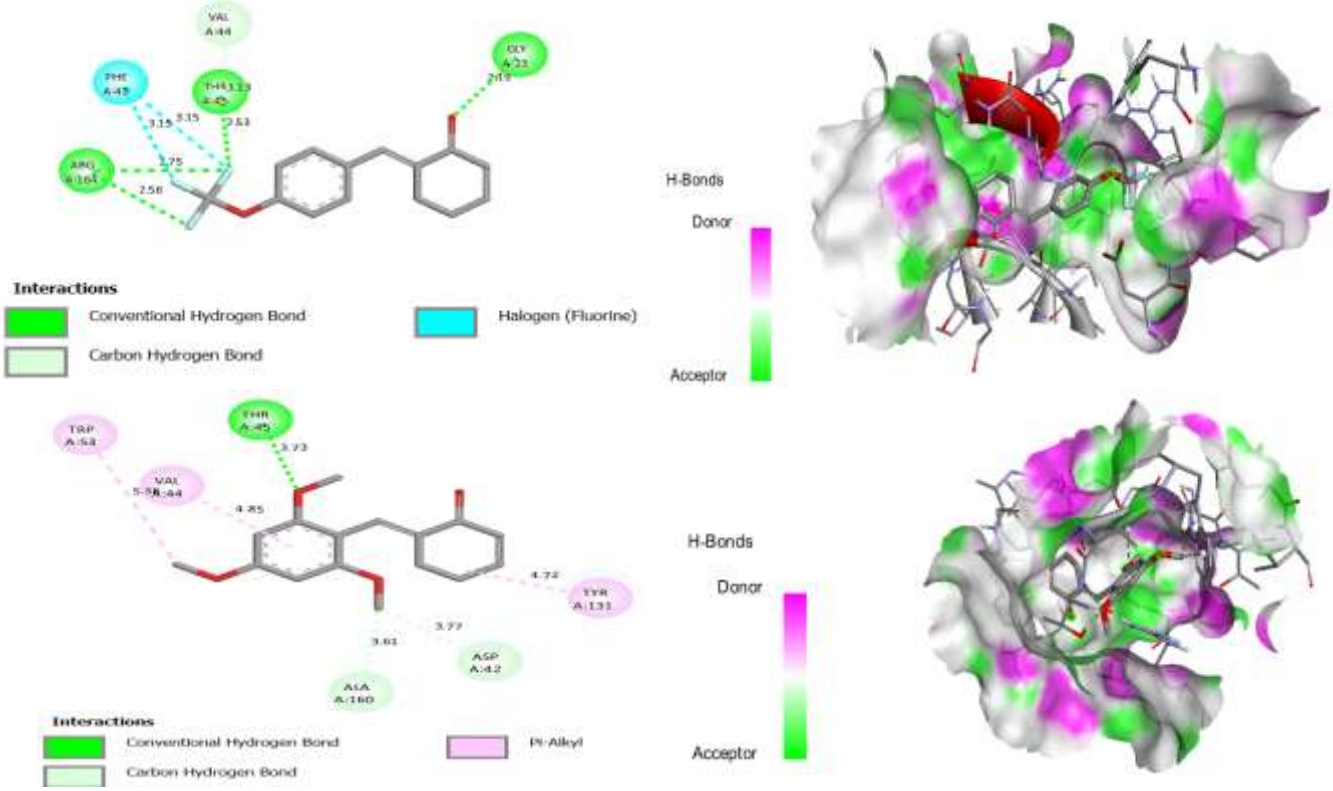


Fig.6: 2D and 3D docking poses showing interactions of compounds 3 and 10.

References

1. Arnér ESJ, Holmgren A (2000) Physiological functions of thioredoxin and thioredoxin reductase. Eur J Biochem 267:6102–6109. https://doi.org/10.1046/j.1432-1327.2000.01701.x
2. Nordberg J, Arnér ESJ (2001) Reactive oxygen species, antioxidants, and the mammalian thioredoxin system. Free Radic Biol Med 31:1287–1312. https://doi.org/10.1016/S0891-5849(01)00724-9
3. Gorrini C, Harris IS, Mak TW (2013) Modulation of oxidative stress as an anticancer strategy. Nat Rev Drug Discov 12:931–947. https://doi.org/10.1038/nrd4002
4. Grabowski H, Vernon J, DiMasi JA (2002) Returns on research and development for 1990s new drug introductions. Pharmacoeconomics 20:11–29. https://doi.org/10.2165/00019053-200220003-00002
5. Urig S, Becker K (2006) On the potential of thioredoxin reductase inhibitors for cancer therapy. Semin Cancer Biol 16:452–465. https://doi.org/10.1016/j.semcancer.2006.09.004
6. Alexander DLJ, Tropsha A, Winkler DA (2015) Beware of R2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. J Chem Inf Model 55:1316–1322. https://doi.org/10.1021/acs.jcim.5b00206
7. van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: Towards prediction paradise? Nat Rev Drug Discov 2:192–204. https://doi.org/10.1038/nrd1032
8. Aouidate A, Ghaleb A, Ghamali M, et al (2018) Investigation of indirubin derivatives: a combination of 3D-QSAR, molecular docking, and ADMET towards the design of new DRAK2 inhibitors. Struct Chem 29:1609–1622. https://doi.org/10.1007/s11224-018-1134-0
9. Press B, Di Grandi D (2008) Permeability for Intestinal Absorption: Caco-2 Assay and Related Issues. Curr Drug Metab 9:893–900
10. Rizki M, Pratama F, Palangkaraya UM, et al (2019) ADMET properties of novel 5- O - benzoylpinostrobin. JBCPP. https://doi.org/10.1515/jbcpp-2019-0251
11. Suenderhauf C, Hammann F, Huwyler J (2012) Computational prediction of blood-brain barrier permeability using decision tree induction. Molecules 17:10429–10445. https://doi.org/10.3390/molecules170910429
12. Zanger UM, Schwab M (2013) Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. Pharmacol Ther 138:103–141. https://doi.org/10.1016/j.pharmthera.2012.12.007
13. Qian J, Xu Z, Meng C, et al (2020) Design and synthesis of benzylidenecyclohexenones as TrxR inhibitors displaying high anticancer activity and inducing ROS, apoptosis, and autophagy. Eur J Med Chem 204:112610. https://doi.org/10.1016/j.ejmech.2020.112610
14. Waller CL, Oprea TI, Giolitti A, Marshall GR (1993) Three-Dimensional QSAR of Human Immunodeficiency Virus (I) Protease Inhibitors. 1. A CoMFA Study Employing Experimentally-Determined Alignment Rules. J Med Chem 36:4152–4160. https://doi.org/10.1021/jm00078a003
15. Gerhard Klebe (1976) Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules To Correlate and Predict Their Biological Activity. 303:108–116
16. Clark M, Cramer RD, Van Opdenbosch N (1989) Validation of the general purpose tripos 5.2 force field. J Comput Chem 10:982–1012. https://doi.org/10.1002/jcc.540100804
17. Tsai KC, Chen YC, Hsiao NW, et al (2010) A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. Eur J Med Chem 45:1544–1551. https://doi.org/10.1016/j.ejmech.2009.12.063
18. Powell MJD (1977) Restart procedures for the conjugate gradient method. Math Program 12:241–254. https://doi.org/10.1007/BF01593790
19. Abedi H, Ebrahimzadeh H, Ghasemi JB (2013) 3D-QSAR, CoMFA, and CoMSIA of new phenyloxazolidinones derivatives as potent HIV-1 protease inhibitors. Struct Chem 24:433–444. https://doi.org/10.1007/s11224-012-0092-1
20. Ungwitayatorn J, Samee W, Pimthon J (2004) 3D-QSAR studies on chromone derivatives as HIV-1 protease inhibitors. J Mol Struct 689:99–106. https://doi.org/10.1016/j.molstruc.2003.10.036
21. Kovalishyn V V., Kholodovych V, Tetko I V., Welsh WJ (2007) Volume learning algorithm

significantly improved PLS model for predicting the estrogenic activity of xenoestrogens. J Mol Graph Model 26:591–594. https://doi.org/10.1016/j.jmgm.2007.03.005
22. Bush BL, Nachbar RB (1993) Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. J Comput Aided Mol Des 7:587–619. https://doi.org/10.1007/BF00124364
23. Baroni M, Clementi S, Cruciani G, et al (1992) Predictive ability of regression models. Part II: Selection of the best predictive PLS model. J Chemom 6:347–356. https://doi.org/10.1002/cem.1180060605
24. Cramer RD, Patterson DE, Bunce JD (1988) Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. J Am Chem Soc 110:5959–5967. https://doi.org/10.1021/ja00226a005
25. Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20:269–276. https://doi.org/10.1016/S1093-3263(01)00123-1
26. Roy PP, Roy K (2008) On some aspects of variable selection for partial least squares regression models. QSAR Comb Sci 27:302–313. https://doi.org/10.1002/qsar.200710043
27. Roy K, Mitra I, Kar S, et al (2012) Comparative studies on some metrics for external validation of QSPR models. J Chem Inf Model 52:396–408. https://doi.org/10.1021/ci200520g
28. Gramatica P (2007) Principles of QSAR models validation: Internal and external. QSAR Comb Sci 26:694–701. https://doi.org/10.1002/qsar.200610151
29. Roy K, Kar S, Ambure P (2015) On a simple approach for determining applicability domain of QSAR models. Chemom Intell Lab Syst 145:22–29. https://doi.org/10.1016/j.chemolab.2015.04.013
30. Netzeva TI, Worth AP, Aldenberg T, et al (2005) Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. ATLA Altern to Lab Anim 33:155–173. https://doi.org/10.1177/026119290503300209
31. MATLAB - MathWorks - MATLAB & Simulink. https://www.mathworks.com/products/matlab.html. Accessed 2 Feb 2022
32. Christoph Q, Meringer M, Biometry M, et al (2007) Y-Randomization and its Variants in QSPR/QSAR. 1–42
33. Golbraikh A, Shen M, Xiao Z, et al (2003) Rational selection of training and test sets for the development of validated QSAR models. J Comput Aided Mol Des 17:241–253. https://doi.org/10.1023/A:1025386326946
34. Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes. https://discover.3ds.com/discovery-studio-visualizer-download. Accessed 11 Mar 2022
35. AutoDock. https://autodock.scripps.edu/. Accessed 11 Mar 2022
36. Protein Data Bank (2021) RCSB PDB: Homepage. In: Rcsb Pdb. https://www.rcsb.org/. Accessed 11 Mar 2022
37. Morris GM, Huey R, Olson AJ (2008) UNIT using AutoDock for ligand-receptor docking
38. Walters WP (2012) Going further than Lipinski's rule in drug design. Expert Opin Drug Discov 7:99–107. https://doi.org/10.1517/17460441.2012.648612
39. Zhang MQ, Wilkinson B (2007) Drug discovery beyond the "rule-of-five." Curr Opin Biotechnol 18:478–488. https://doi.org/10.1016/j.copbio.2007.10.005
40. Ghose AK, Viswanadhan VN, Wendoloski JJ (1999) A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. J Comb Chem 1:55–68. https://doi.org/10.1021/cc9800071
41. Veber DF, Johnson SR, Cheng HY, et al (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623. https://doi.org/10.1021/jm020017n
42. Egan WJ, Merz KM, Baldwin JJ (2000) Prediction of drug absorption using multivariate statistics. J Med Chem 43:3867–3877. https://doi.org/10.1021/jm000292e
43. Muegge I, Heald SL, Brittelli D (2001) Simple selection criteria for drug-like chemical matter. J Med Chem 44:1841–1846. https://doi.org/10.1021/jm015507e
44. Martin YC (2005) A bioavailability score. J Med Chem 48:3164–3170. https://doi.org/10.1021/jm0492002

45. SwissADME. http://www.swissadme.ch/index.php. Accessed 11 Mar 2022
46. Ferreira LLG, Andricopulo AD (2019) ADMET modeling approaches in drug discovery. Drug Discov Today 24:1157–1165. https://doi.org/10.1016/j.drudis.2019.03.015
47. Pires DEV, Blundell TL, Ascher DB (2015) pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. J Med Chem 58:4066–4072. https://doi.org/10.1021/acs.jmedchem.5b00104
48. Clark DE (2003) In silico prediction of blood-brain barrier permeation. Drug Discov Today 8:927–933. https://doi.org/10.1016/S1359-6446(03)02827-7
49. Hammarlund-Udenaes M (2000) The use of microdialysis in CNS drug delivery studies: Pharmacokinetic perspectives and results with analgesics and antiepileptics. Adv Drug Deliv Rev 45:283–294. https://doi.org/10.1016/S0169-409X(00)00109-5
50. Han Y, Zhang J, Hu CQ, et al (2019) In silico ADME and toxicity prediction of ceftazidime and its impurities. Front Pharmacol 10:1–12. https://doi.org/10.3389/fphar.2019.00434