Syrian Arab Republic

Minister of Higher Education And Research

Syrian Virtual University

الجمهورية العربية السورية

وزارة التعليم العالي

الجامعة الافتراضية السورية

QSAR and 3D-QSAR Principles and applications in Drug Design (antineoplastic drugs)

A Project submitted for the Master´s degree in Bioinformatic

Student: Aya Ahamd Hassan Almasri

Student ID: Aya_157254

Supervisors:

A.Prof. Majd Aljamali

Dr.Khansaa Hussein

2023

QSAR and 3D-QSAR Principles and applications in Drug Design (antineoplastic drugs)

Table of Contents

Table of Figures

Abstract:

QSAR and 3D-QSAR techniques marked a huge milestone in drug design development, especially in antineoplastic drugs.

QSAR models utilizes molecular descriptors to predict the relationships between the chemical structure and the biological activity, which aids in designing and developing potent compounds.

Classical QSAR models have limitations in drug design. Thus, 3D-QSAR methods were developed in order to provide more accurate results of the drug-target interactions.

The applications of these methods has led to the development of drug design and antineoplastic drugs with improved efficacy and reduced side effects and toxicity.

In conclusion, QSAR and 3D-QSAR play a vital role in the development of more effective and more selective drugs.

First Section: QSAR and 3D QSAR Principles

SAR: Structure-Activity Relationship:

A qualitative relationship which explores the relationship between a molecule structure and its biological activity. The SAR hypotheses assumption is that similar molecules have similar activities. Which helps us in further researches in predicting and developing more effective drugs and even less toxic side effects. Therefore, many technical and chemoinformatic approaches have been developed based on this principle, e.g. the chemoinformatic method Quantitative Structure-Activity Relationship (QSAR).

Quantitative Structure-Activity Relationship (QSAR)

QSAR is a quantitative relationship and is a cheminformatics' approach used to develop models that estimate the macroscopic characteristics of molecules based on their structural properties. It involves the use of various machine-learning techniques to approximate functions using input-output data. QSAR modelling is commonly used in drug design projects where the three-dimensional structure of the target macromolecule is unknown.

QSAR Classifications based on predictor variables: (12)

- 1D-QSAR: correlating molecular activity with molecular properties like pKa, log P, etc.
- 2D-QSAR: correlating activity with structural 2D patterns like connectivity indices, 2D-pharmacophores.
- 3D-QSAR: correlating activity with non-covalent interaction fields surrounding the molecules.
- 4D-QSAR: additionally including ensemble of ligand configurations in 3D-QSAR.
- 5D-QSAR: explicitly representing different induced-fit models in 4D-QSAR.
- 6D-QSAR: further incorporating different solvation models in 5D-QSAR.

QSAR Theory

All QSAR analyses are based on the hypothesis of linear additive contributions of the different structural properties or features of a compound to its biological activity, provided that there are no nonlinear dependences of transport or binding on certain physicochemical properties. (1)

The Free Wilson analysis and Hansch analysis, both developed in 1964, marked a milestone in the development of QSAR by correlating certain structural features or physicochemical properties with biological activities (1)

Free Wilson Analysis:

This mathematical model has been derived to describe the presence and the absence of certain structural features. It is often used to see at a glance which physicochemical properties might be important to the biological activity.

Equation 1

$$\log 1/C = \sum a_i + \mu$$

Where:

a_i:The values of biological activity group contributions of the substituents X1, X2, ... X~ in the different positions p of compound

μ : is the biological activity value of the reference compound, most often the unsubstituted parent structure of a series

Thus, Free Wilson Analysis is based on assumptions that the entire drug list should have the same parent structure, and the substitution pattern in various derivatives has to be the same, in addition that the substitutions have to contribute to the biological activity additively. (1)

Hansch Analysis (extra thermodynamic approach)

The linear free-energy-related Hansch model is a statistical method used to correlate physicochemical parameters with biological activities.
The following equation has been developed based on the concept that the transport of a drug from the site of application to its site of action depends in a nonlinear manner on the lipophilicity of the drug.
On the other hand the binding affinity to its biological part depends on the lipophilicity, the electronic properties and other linear free-energy-related properties

Equation 2

$$\log 1/C = a\ (\log P)^2 - \mathrm{b} \log P + \mathrm{c}\ \sigma + \cdots + \mathrm{k}$$

P : n-octanol/water parameter
a,b,c: regression coefficients
k: constant term

Lipophilicity Factors:
The (log P) in the previous equation represents the Partition coefficient. It refers to the whole molecules and is a linear relationship between P and drug activity

Electronic Factors:
These factors are basically represented by the Hammett constant (σ).
It refers that the distribution of electrons within a molecule depends on the nature of the electron withdrawing and donating groups found in that structure.

Here is an example that describes a quantitative relationship between the antiadrenergic activities of compounds and their lipophilicity and electronic properties

.*Equation 3*

$$\log 1/C = 1.15\ \pi - 1.464\ \sigma + 7.817$$

(n = 22, r = 0.945, s = 0.19, F = 78.6; Q2 =0.841, Spress = 0.238)
N: number of compounds
R: internal validation correlation coefficient
S: standard deviation; measure of absolute quality of model (should be < 0.3)
Q2: squared cross-validation correlation coefficient (measures for internal predictivity) .
F: Fisher value; measure for the statistical significance.

Where $\log 1/C$ is the logarithms of reciprocal values are the correct scaling, while C is the molar concentration that causes a certain biological effect.
π : Lipophilicity parameter.
σ : electronic parameter (1)

Comparison:

From the previous equations we can conclude the following results: (Table 1) Constructed by the student

Free Wilson Analysis	Hansch Analysis
• mathematical approach that analyses compounds with the same parent structure	• linear free- energy-related
• no calculations of physicochemical parameters	• correlates physicochemical properties with biological activity
• limited in application	• predicts the activity of other substituents
• Faster method	• more general
• more simple	• uses new types of descriptors

Table 1 Comparison between Free Wilson Analysis and Hansch analysis (student)

3D QSAR (CoMFA method):

The three dimensional QSAR method has been developed over the years and the most widely used method is CoMFA.

CoMFA(Comparative Molecular Field Analysis) method is a molecular field-based method formed the first real 3D-QSAR method. Where it uses partial least square analysis and commercial software.

The CoMFA method has proved its value when classical QSAR methods fail. An advantage in CoMFA is that it considers the properties of the ligands in their expected bioactive confirmations, Thus it is more suitable to describe ligand-receptor interactions. Which gives us information on regions in space whether they are favourable or unfavourable for the ligand- receptor interaction. (1)

It is a method used in drug design that helps in bioactivity and physical properties of molecules prediction, especially to derive quantitative models for enzyme inhibition constants and binding affinities.

The CoMFA process:

It starts from 3D structures and correlates biological activities with 3D-property fields. Following the steps below. (Figure 1) (1)

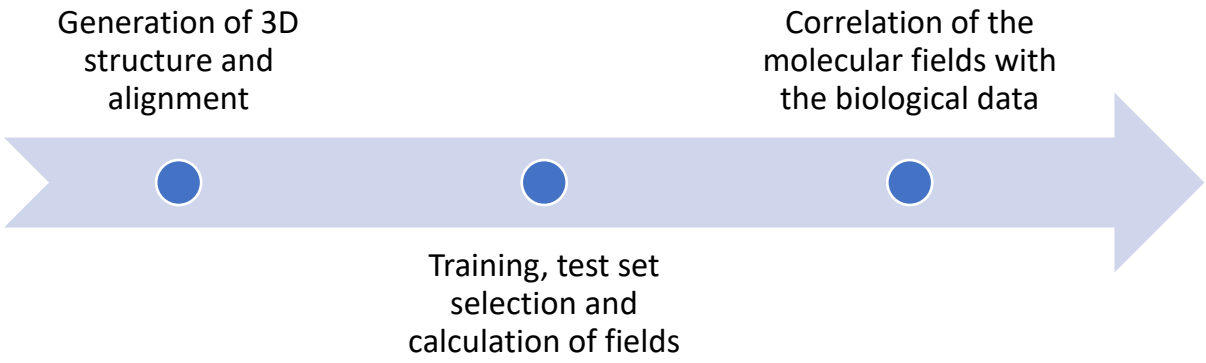


Figure 2 CoMFA process (student)

I. Generation of 3D structure and alignment

In CoMFA, first group of compounds is selected. These compounds should be chemically related and should act by the same mechanism of action; however, in contrast to classical QSAR methods, they should have a common pharmacophore, not necessarily the same molecular skeleton.

To get the common pharmacophore which refers here to 3D- structures, first the 2D and 2.5D structures of all molecules are converted to 3D structures using Standard computer programs which are CONCORD and CORNIA. These two programs create only one low-energy structure per molecule.

We can also draw the structures of the compounds in 2D ChemDraw and then be converted to 3D structures using the default conversion procedure implemented in the CS Chem 3D ultra. The generated 3D structures are then subjected to energy minimization and geometry optimization using Spartan.

Alignment of molecules is performed based on orientation rules derived from the common pharmacophore. Correct alignment is difficult for molecules that are not from a congeneric series or have a large number of rotatable bonds.

Some evidence suggests that even if molecules are studied in different geometries, their similarities and affinities can still be properly described

II. Training, Test Set Selection, and Calculation of Fields:

The data set is divided into a training set and a test set for the development and evaluation of a CoMFA model.

The training set is used to derive the CoMFA model, which involves calculating molecular fields and deriving quantitative structure-activity relationships.

The test set is used to validate the CoMFA model and assess its predictive power.

The calculation of fields involves considering the three-dimensional structures and binding modes of protein ligands to identify regions in space that are favourable or unfavourable for ligand-receptor interactions.

The fields are calculated using methods such as comparative molecular field analysis (CoMFA) or comparative molecular similarity indices analysis (CoMSIA), which utilize various functions and algorithms to assess the similarity and interactions of atoms and groups within the molecules

III. Correlation of the molecular fields with the biological data

The best correlation method is the Partial least square analysis (PLS). This method extracts latent variables from the Y and X blocks to achieve maximum correlation

between these variables. (Figure 3)

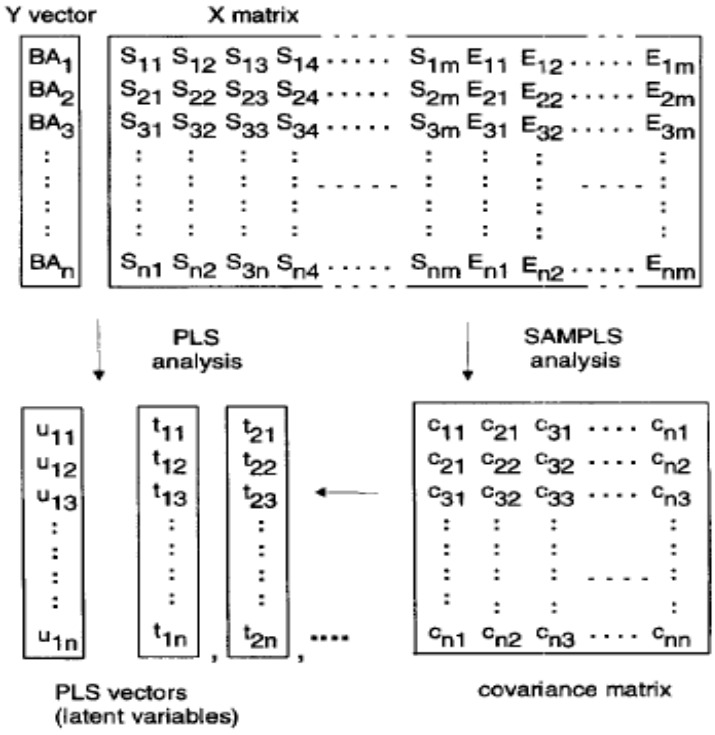


Figure 3 (1)

The optimum number of latent variables is determined through cross-validation, where models are derived by excluding compounds from the data set.

The PLS analysis results in a regression equation with thousands of coefficients, allowing predictions for compounds outside the training set.

- Statistical measures commonly used to assess the predictive power of CoMFA models include:

1. Squared Cross-Validated Correlation Coefficient (Q2): Q2 is calculated by comparing the predicted values of the model with the experimental values using cross-validation. It provides an indication of how well the model predicts the activity of compounds in the training set.
2. Standard Deviation of Predictions (spress): This measure is also calculated from the predicted values using cross-validation. It represents the spread or variability of the predicted values around the experimental values.
3. Squared Correlation Coefficient (r2): Similar to Q2, r2 measures the goodness of fit between the predicted and experimental values. It indicates the proportion of the variance in the activity that can be explained by the model.

4. Standard Deviation (s): This measure represents the spread or variability of the experimental values. It provides an indication of the overall variability in the dataset.
5. SpRES (Standardized Prediction Residual Sum of Squares): SpRES is a measure of the prediction error of the model. It represents the sum of the squared differences between the predicted and experimental values, standardized by the standard deviation of the experimental values.
6. Optimum Number of Latent Variables: The lowest SpRES value is often used as a criterion for determining the optimum number of latent variables in the model.

Model Validation

Internal Model Validation:

- The developed models underwent internal validation using the leave-one-out (LOO) cross-validation technique. In this technique, one compound is removed from the dataset in each cycle, and the model is built using the remaining compounds. The model is then used to predict the activity of the removed compound. This process is repeated until all compounds have been eliminated once.
- To assess the quality of the models, the cross-validated squared correlation coefficient, R2cv (Q2), was calculated. This coefficient measures the correlation between the observed and predicted activities of the training set compounds.
- The overall significance of the regression coefficients was evaluated by calculating the variance ratio, F value. The F value represents the ratio of the regression mean square to the deviations mean square and helps ascertain the significance of the regression coefficients in the model.
- These validation techniques provide statistical measures to assess the accuracy and reliability of the developed QSAR models. (3)

External Model Validation:

External validation was employed in order to determine the predictive capacity of the developed model as judged by its application for the prediction of test set activity values and calculation of predictive R2(R2pred) value. (3)

Y- Randomization test

- The robustness of the developed QSAR model was assessed using the Y-randomization technique. This technique involves permutating the response values (activity) while keeping the descriptor matrix unchanged. By doing this, the relationship between the descriptors and the activity is disrupted, and any correlation observed is likely due to chance.
- to evaluate the impact of randomization, the squared mean correlation coefficient of the randomized model (Rr2) is compared to the squared correlation coefficient of the non-random model (R2) using the R2p parameter. Ideally, in a randomized model, the average value of Rr2 should be zero, indicating no meaningful relationship between the descriptors and the activity. The value of R2p should be equivalent to the R2 value of the developed QSAR model.

- To account for any discrepancies between the randomized and non-randomized models, a correction factor, cRp2, is calculated. This correction penalizes the developed models for the difference in squared correlation coefficients between the randomized and non-randomized model (3)

QSAR Workflow

The following QSAR Workflow summarizes the process for all classifications: (Figure 4)

Figure 4 QSAR Workflow (student)

QSAR Sofwares:

Among the notable QSAR software packages available are OASIS, DRAGON, ODESSA, MOE (Molecular Operating Environment), ChemDraw, and ChemAxon

- OASIS, with its user-friendly interface, offers a range of molecular descriptors and statistical analysis tools to build QSAR models. It provides options for data preprocessing, variable selection, and model validation.
- DRAGON is another widely used software known for its extensive selection of molecular descriptors and its ability to handle large datasets. It offers a variety of structural and quantum chemical descriptors, allowing for comprehensive analysis of chemical properties.
- ODESSA is a powerful software designed for modeling and predicting activity landscapes. It employs advanced machine learning algorithms and feature selection techniques to generate robust QSAR models.
- MOE is a comprehensive software package that offers various tools for molecular modeling, including QSAR analysis. ChemDraw and ChemAxon are popular software tools known for their user-friendly interfaces and extensive libraries of chemical structures and molecular descriptors.

Overall, these QSAR software packages facilitate efficient and accurate modeling and prediction of compound activities, aiding in the rational design of novel drugs

Second Section : Applications

QSAR METHODS have proved its value in drug design. As we know pharmacophore descriptors play a huge role in building the cornerstone of a molecule. QSAR methods are important for lead in the prediction of biological activities and guide the design of new compounds.

Absorption and distribution of drugs:

Absorption and distribution of drugs often exhibit nonlinear relationships with lipophilicity, as observed in many experimental data.

The following figure shows the permeation of barbiturates and carbamates through different phases and membranes. The permeations of these compounds follows a nonlinear dependence on their lipophilicity. It also shows that similar dependencies on lipophilicity are observed for the absorption rates of carbamate in the gut and other absorption processes. (Figure 5a)

The neurotoxic activities of a series of homologous alcohol also show a nonlinear lipophilicity dependence. The blood-placenta barrier shows a similar but slightly less pronounced lipophilicity dependence for a group chemically different drugs. (Figure 5b) (2)

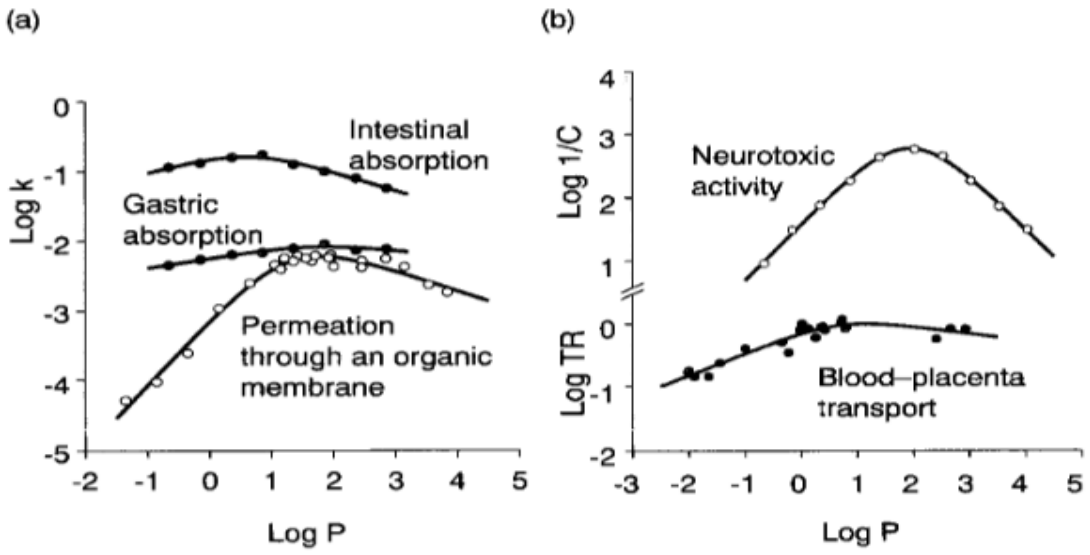


Figure 5 (1)

Lipophilic drugs are easily absorbed and distributed, but they may have low bioavailability because of the direct elimination in the liver or undesired side effects in the CNS. This is why it is recommended in drug design to make drugs only as lipophilic as necessary. The blood-brain barrier and the blood-placenta barrier penetration are also influenced by lipophilicity.

QSAR methods play a crucial role in understanding drug absorption and distribution by correlating experimental data with appropriate models, such as lipophilicity parameters and steric or hydrophobic fields, ionization.

These methods help in:

- Identifying the relationship between lipophilicity and drug absorption.
- Identifying the structural features of drugs that contribute to their distribution properties, allowing for the optimization of drug candidates with desired distribution profiles.
- Providing insights into how different molecular properties influence drug distribution in various tissues and organs by analysing the (SAR) of drugs.
- Predicting the distribution of drugs in different physiological compartments, which helps the design of drug delivery systems and optimizing drug dosing regimens.

Enzyme Inhibitors

Enzyme inhibitors are molecules that bind to enzymes and interfere with their activity, either by blocking the active site or altering enzyme's conformation.

QSAR methods have been used to study enzyme inhibitors and understand their structure-activity relationships (SAR). Classical QSAR and comparative molecular field analysis (CoMFA) are employed to describe the structure-activity relationships of enzyme inhibitors. These methods have been successfully applied to different enzymes, including dihydrofolate reductase, papain, trypsin ,renin, HIV protease, and thermolysin. QSAR studies help in identifying the structural features that contribute to the inhibitory activity of enzyme inhibitors, which leads to designing and optimizing more potent and selective inhibitors. The followimg two boxes gives us a glimpse on the Classical QSAR and 3D QSAR enzyme ligands. (Table 2) (2)

Classical QSARs of enzyme ligands (substrates and inhibitors)	3D QSAR Sof enzyme ligands (substrates and inhibitors)
Hydrolases	Hydrolases
Chymotrypsin – Trypsin Cholinesterases- Papain	Acetylcholinesterase- Angiotensin-converting enzyme (ACE)
Oxidoreductases	Chymotrypsin
Alcohol dehydrogenase- Dihydrofolate reductase -Malate dehydrogenase – Xanthine oxidase	Dipeptidyl peptidase
Transferases	HIV protease
Acetyl transferases	Papain
	Renin
	Thermitase
	Thermodysin
	Oxidoreductases
	Aromatase
	Cytochrome P450
	Dihydrofolate reductase
	Monoaminoxidases
	Ligases
	Phenylethanolamine
	PJmethyltransferase
	Otherenzymes
	HIVintegrase - Topoisomerase

Table 2 Classical QSAR and 3D QSAR of enzyme ligands (1)

QSAR analyses have shown good correlation between structure derived ligand-receptor interactions and the results of QSAR analyses for enzyme inhibitors when the 3D structures of the enzymes are known.

For example: closely related QSAR models were derived for the inhibition of E.coli DHFR and Lactobacillus casei DHFR by benzyl pyrimidines, with differences in the contribution of 5-substituents of the benzyl group to biological activities where they have no effect in the case of Lactobacillus casei. The X-Ray structures of the enzymes provided an explanation for this case. It has shown that they both have the same geometry of the binding site but a rigid Leucine side chain in the L. Casei DHFR forms a much narrower cleft than a more flexible methionine side chain in E.coli DHFR.

Activity-activity relationships

QSAR is important for describing activity-activity relationships, especially in the context of combinatorial chemistry and high-throughput screening methods, where correlating in vitro and in vivo activities of a standard series of analogs is crucial for justifying the chosen in vitro test model.

Activity-activity relationships refer to the connections between various activities or effects of a drug, including therapeutic activity, side effects, and interactions with receptors or enzymes. By using QSAR methods to analyse the structure-activity relationships of drugs, researchers can identify the molecular properties that influence these activities. Understanding activity-activity relationships helps in comprehending drug mechanisms and their potential therapeutic uses. It also enables the design of more effective and targeted drugs by providing insights into selectivity and specificity. QSAR models and statistical methods are commonly employed to analyse activity-activity relationships and identify the significant variables contributing to drug activity. (2)

QSAR and 3D-QSAR applications in Drug design:

Table 3 3D QSAR applications in drug design (student)

Third Section: QSAR and 3D applications in antineoplastic drugs

Introduction

Cancer is a complex and devastating disease characterized by the uncontrolled growth and spread of abnormal cells in the body. Understanding its mechanisms of action and finding effective treatments are crucial in the fight against cancer.

The development of antineoplastic drugs, which target and inhibit the growth of cancer cells, plays a vital role in cancer treatment. These drugs are designed to disrupt specific cellular processes involved in cancer progression, such as cell division or angiogenesis. However, developing effective and selective antineoplastic drugs is a challenging task due to the complexity and heterogeneity of cancer.

Quantitative Structure-Activity Relationship (QSAR) and Three-Dimensional QSAR (3D QSAR) methods have emerged as valuable tools in drug development, particularly in the design of antineoplastic drugs. QSAR models integrate information on the molecular structure of a drug with its activity data, allowing for the identification and prediction of key structural features that contribute to its therapeutic effectiveness against cancer.

By analysing the structure-activity relationships of antineoplastic compounds, QSAR methods help researchers gain insights into the specific molecular properties responsible for their activity against cancer cells. These methods can also predict the potency, selectivity, and toxicity of potential antineoplastic drug candidates, aiding in the development of more effective and safer treatments.

Additionally, the advent of 3D QSAR approaches has further enhanced the accuracy and predictive power of drug design. By considering three-dimensional interactions between drug molecules and their targets, 3D QSAR provides a more detailed and precise understanding of the molecular mechanisms underlying the activity of antineoplastic drugs.

In conclusion, the development of antineoplastic drugs is essential in the battle against cancer. QSAR and 3D QSAR methods offer valuable tools in understanding the mechanisms of action of these drugs, predicting their activity, and designing more effective and selective treatments to combat this devastating disease.

Applications:

2D and 3D QSAR models study of novel nitrogen- mustard compounds for Osteosarcoma:

Osteosarcoma is a type of cancer that originates from mesenchymal tissues and is characterized by the production of osteoid matrixes by tumor cells.
Methotrexate, Adriamycin, cisplatin, and ifosfamide are the main chemotherapy drugs used against osteosarcoma, but they can have side effects and may lead to drug resistance.
Nitrogen-mustard anti-tumor drugs form electron-deficient dimethylimine ions in the body, which combine with electron-rich groups in biological macromolecules to destroy tumor-target DNA fragments, offering a potential solution for eliminating tumor cells; these compounds have advantages in terms of simple synthesis and low cost, making them promising for clinical use in malignant tumor drugs; a dipeptide-alkylated nitrogen-mustard compound with high anti-tumor activity has been discovered, providing new possibilities for designing chemotherapy drugs against osteosarcoma.

Dataset:
In this experiment, 22 alkylated dipeptide nitrogen mustard derivatives were used as a dataset.

1. 2D-QSAR research:
    - First data processing and structure optimization was applied to build a predictive QSAR model. The experiment involved grouping 22 compounds randomly, with 18 in the training set and four in the test set.
    - The compounds were initially constructed using ChemDraw software and then imported into HyperChem software.
    - The compounds were optimized in two steps:
      The first optimization step involved using the MM+ molecular mechanic field for rough optimization. In the second step, a more precise optimization was performed using semi-empirical methods like AM1 or PM3 in HyperChem. The molecular structure was optimized using the Polak-Ribiere algorithm until the root mean square gradient reached 0.01. Finally, the results were imported into CODESSA software to calculate five different classes of molecular descriptors: constitutional, geometrical, topological, electrostatic, and quantum chemical.
    - Linear Model development through heuristic method: Feature selection through a heuristic method was employed to build a linear model, selecting descriptors based on statistical characteristics. Ultimately, a linear model comprising six descriptors was developed.
    - Non Linear Model development using (GEP) : Gene expression programming (GEP) is an evolutionary algorithm that combines principles from genetic algorithm (GA) and genetic programming (GP).
      GEP overcomes the limitations of GA and GP, making it more efficient. Unlike GA and GP, GEP uses linear chromosomes as candidate solutions, with a head section selected from end sets and feature sets, and a tail section selected only from the end set. These chromosomes are then decoded into expression trees to obtain

mathematical equations. In this study, GEP is used with automatic problem solver (APS) to integrate descriptor values and obtain non-linear models. These models have better stability and prediction ability compared to linear models. However, a 2D-QSAR model is not enough to accurately describe the relationship between molecular three-dimensional structures and their physiological activity, so further 3D-QSAR experiments are needed.

2. 3D-QSAR research
    a. In 3D-QSAR experiments, the dataset is divided into a training set (18 compounds) and a test set (4 compounds). The training set is used to build models, while the test set is used for verification. To reduce deviation, the IC50 values are converted using -log (IC50) + 6. The compounds are optimized and modelled using SYBYL software, which utilizes the Tripos force field and Powell's gradient algorithm to minimize CoMSIA structure energy. The resulting minimal structure is used as the initial conformation.
    b. In 3D-QSAR analyses, the selection of an appropriate structure comparison method is crucial as it influences subsequent tests. This study utilizes ligand alignment to superpose the compound structures. The alignment is performed using compound I1 as the reference since it has the highest IC50 value. The superposition patterns of all compounds can be observed in the figure below. (Figure 6)

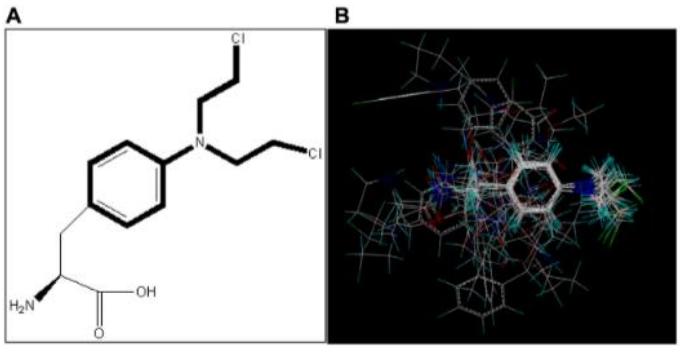


FIGURE 2
Alignment of all compounds in the dataset; compound I1 is used as a template. **(A)** Structure of compound I1 and the common substructure (shown in bold) for the alignment of all compounds. **(B)** Alignment of all the compounds.

Figure 6 (10)

- CoMSIA is a powerful 3D-QSAR research tool that utilizes Gaussian functions to calculate various molecular fields. This method avoids significant changes in potential energy and abnormal atomic positions near the molecular surface. Unlike CoMFA, CoMSIA does not require the definition of an energy cutoff value. The correlation isosurface diagram of CoMSIA provides a comprehensive understanding of the contribution of different molecular fields to molecular activity.
- In this study, a CoMSIA analysis is conducted using the SYBYL software package, considering five molecular fields: spatial, electrostatic, hydrophobic, hydrogen

bond donor, and hydrogen bond acceptor. The analysis is performed on a 3D cubic lattice with a grid spacing of 2 Å and extending 4 Å units beyond the aligned molecules. The CoMSIA method also utilizes partial least squares (PLS) analysis to correlate the molecular fields with the experimental values. The analysis is carried out in two stages: leave-one-out cross-validation analysis and non-cross-validation analysis. Several statistical parameters, including the optimal group score, cross-validation correlation coefficient, non-cross validation correlation coefficient, estimated standard error, and F-value, are used to evaluate the non-cross validation results of the final PLS regression model for CoMSIA. The following table represents the statistical results of the optimal CoMSIA model. (10)(Table 4)

Model	q^2	ONC	r^2	SEE	F
CoMSIA	0.532	5	0.997	0.016	1601.378
Name	**S**	**E**	**H**	**D**	**A**
Contribution (%)	5.4	27	22.7	26.3	18.5

Table 4 The statistical results of the optimal CoMSIA model (10)

- To assess the stability and predictive ability of the 3D-QSAR model, external validation methods are employed. In this study, external validation is chosen to verify the model. The value of R2ext obtained was 0.987, which is greater than 0.5. This indicates that the constructed model is highly stable and possesses good statistical prediction ability. Figure (7) also demonstrates a strong correlation between the predicted values and the experimental values.

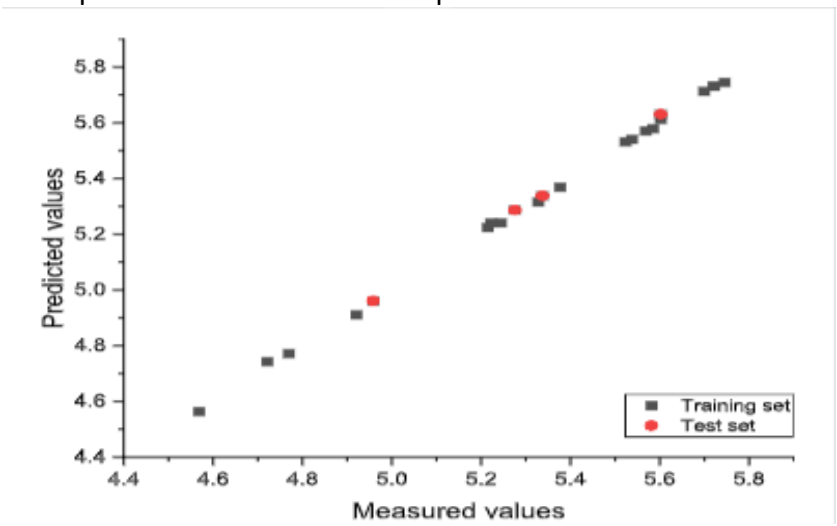
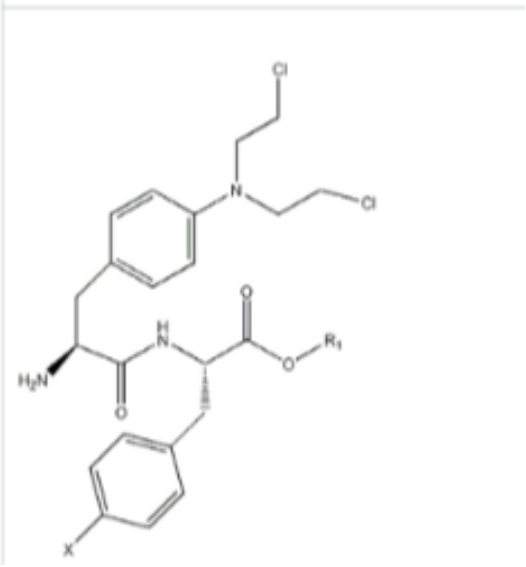


Figure 7 CoMSIA model-predicted activity values compared with the *experimental values.* (10)

Conclusion and Results:

- In this study, both linear and non-linear 2D-QSAR models were established using the heuristic method and GEP algorithm. It was observed that the non-linear model exhibited better stability and prediction ability. However, the 2D-QSAR model had a limitation in accurately describing the influence of spatial structural changes on anti-tumor activity. To overcome this limitation, a 3D-QSAR model was constructed using the CoMSIA method. This 3D-QSAR model had higher q2 and r2 values and a lower estimated standard error compared to the 2D-QSAR model. Furthermore, the 3D-QSAR model captured the changes in spatial structure and anti-tumor activity more intuitively.
- Using a combination of the molecular descriptor from the 2D-QSAR model and the molecular force field from the 3D-QSAR model, 200 new nitrogen compounds were created. Compound I1.10 exhibited the highest drug activity among these compounds.
- To further assess the effectiveness of these compounds on osteosarcoma-related receptor targets, small-molecule docking experiments were performed. The docking results for compound I1.10 were found to be satisfactory. (Figure 8) (10)



Compound I1.10: where

R_1= -CH_2CH_3, X= $N(CHCHCl)_2$

IC_{50}=2.5 µM

Figure 8 (10)

2D-QSAR and 3D-QSAR Analyses for EGFR Inhibitors

Epidermal growth factor receptor (EGFR) belongs to the receptor tyrosine kinases family and is activated by ligands such as epidermal growth factor (EGF) and transforming growth factor alpha (TGF-α). Abnormal activation of EGFR can lead to cancer, and it is considered a potential target for cancer treatment. Existing EGFR inhibitors have limitations in terms of selectivity, toxicity, and side effects, necessitating the design and synthesis of new inhibitors.

In this study, a 2D-QSAR model is used to identify EGFR inhibitors, while a 3D-QSAR model predicts their activity. Molecular docking is then employed to investigate binding sites.

Materials and methods:

1. The CfsSubsetEval method combined with the Greedy Stepwise algorithm was used to search for the optimal feature subset from a large number of combinations, aiming to find a useful subset for predicting compounds accurately.
2. Support Vector Machine (SVM) was employed for classification and sensitivity analysis due to its high performance.
3. Topomer CoMFA, which combines the topomer technique and CoMFA technology, was used to overcome the alignment problem in drug design. Partial least squares (PLS) regression is employed to build the topomer CoMFA model, and leave-one-out(LOO) cross-validation is used to evaluate the model.
4. Data preparation involved collecting 100 inhibitors from the literature and 185 noninhibitors from the DUD database.
    - For 2D-QSAR study:: The dataset was randomly divided to three training sets which accounted for 75%, 70% and 50% of the whole dataset.
    - For 3D-QSAR study, the 100 inhibitors were randomly divided into a training set (77 molecules) and an independent test set (23 molecules).
5. Molecular descriptors were calculated to represent the compounds' physicochemical and geometric properties.
6. In this study, the prediction ability of the 2D-QSAR model was evaluated using both a tenfold cross-validation test and an independent set test. In the tenfold cross-validation test, the data set was divided into ten subsets, with nine subsets used as the training set and one subset used for prediction. This process was repeated for each subset, and the correct rate was obtained for each trial. The accuracy of the algorithm was estimated by taking the average correct rate from the ten trials.
7. Sensitivity, specificity, overall accuracy were employed to evaluate the 2D prediction model.
8. q2, r2, and MAE were used to evaluate the topomer CoMFA model.
9. Steric and electrostatic field analysis was performed to design novel EGFR drugs.
10. Molecular docking was conducted using SYBYL X-2.0 and the crystal structure of EGFR

Results:

1. Feature subset containing nine molecular descriptors was selected using CFS and GS algorithms.
2. Sensitivity analysis was conducted on these descriptors to assess their impact on the activity of EGFR inhibitors.
3. The SVM classifier method was then used to build a 2D-QSAR prediction model using the optimal feature subset.
4. Through a tenfold cross-validation test, the prediction accuracies of the models, representing 75%, 70%, and 50% of the whole data set, were found to be 98.13%, 98.99%, and 91.24% respectively.
5. These models demonstrated high sensitivity, specificity, and overall accuracy exceeding 90%. The model based on the 70% data set was chosen as it provided

higher prediction accuracy. An independent test set was also used to validate the classifier's reliability, resulting in a prediction accuracy of 97.67%.

For 3D-QSAR Prediction model:

1. The 3D-QSAR prediction model was built using the topomer CoMFA technique, which combines the topomer method and CoMFA technology .
2. Two topomer CoMFA models were generated by fragmenting EGFR inhibitors into R1 and R2 groups. The model with higher q2 and r2 values was selected for analysis and prediction . Model number 2 was selected according to the figure below.(Figure 11)

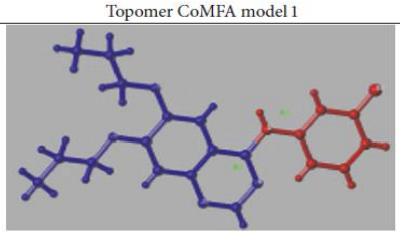
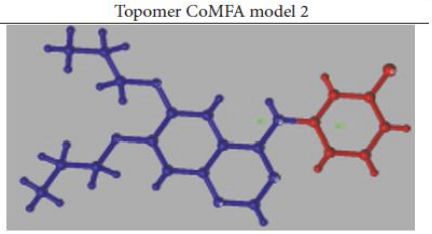
Dataset	Topomer CoMFA model 1	Topomer CoMFA model 2
Cutting model		
q^2	0.483	0.565
r^2	0.773	0.888

Figure 9 (5)

3. The model achieved a cross-validated correlation coefficient (2) of 0.565 and a non-cross-validated correlation coefficient (2) of 0.888, indicating its ability to predict the activity of EGFR inhibitors .
4. The mean absolute error (MAE) for the training set was 0.308 log units, and for the test set, it was 0.526 log units .
5. The steric and electrostatic fields of the R1 and R2 groups were calculated using the topomer CoMFA model, providing valuable insights for the design of novel EGFR drugs .
6. The 3D-QSAR model, along with molecular docking, contributes to a better understanding of the interaction between EGFR inhibitors and EGFR .

Conclusion:

In this study, the researchers used both 2D-QSAR and 3D-QSAR prediction models to analyse EGFR inhibitors. The first step was to build a 2D-QSAR model that could predict whether a compound was an inhibitor or a non-inhibitor. The accuracy of the model was evaluated using both a tenfold cross-validation test and an independent set test, with accuracies of 98.99% and 97.67% respectively. Next, the researchers built a topomer CoMFA model based on EGFR inhibitors. Two models were obtained by cutting different molecular bonds. After comparing the performance of the models, model 2 was chosen due to higher $q2$ and $r2$ values. This model was then used to predict EGFR inhibitors. Finally, the researchers selected a series of similar chemical inhibitors and conducted molecular docking studies to study the interacting sites between EGFR and the inhibitors. The docking results identified Thr766 and Met769 as important interacting sites, suggesting that these residues played a crucial role in the activity of EGFR. In summary, this study utilized 2D-QSAR and 3D-QSAR models, as well as molecular docking, to analyse EGFR inhibitors. The models showed high

prediction accuracy, and the docking results provided insights into the key interacting sites involved in EGFR activity. (5)

3D-QSAR studies of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer

Polo-like kinases (PLKs) is a serine-threonine kinase that have five family members (PLK1-5). This group of enzymes play an important role in mitosis and are necessary for centrosome maturation and bipolar spindle formation.
PLK1 overexpression has been found in many types of cancers(lung cancer, prostate cancer, colon cancer) and it plays a very important role in cell proliferation. Therefore, PLK1 has been known to be a broad-spectrum anti-cancer target.
QSAR methods have been applied to identify the best PLK1 enzyme candidate inhibitors for prostate cancer treatment. However, a series of novel pteridinone derivatives were synthesized and evaluated in their biological activity. This series will go through different molecular docking methods including the 3D-QSAR (CoMFA and CoMSIA) to generate predictive and robust models to predict and study new drug candidates with a reasonably low economic impact.
The following figure is a flowchart of different studies that has been used, But in our study we will focus only on the 3D-QSAR study. (Figure 9)

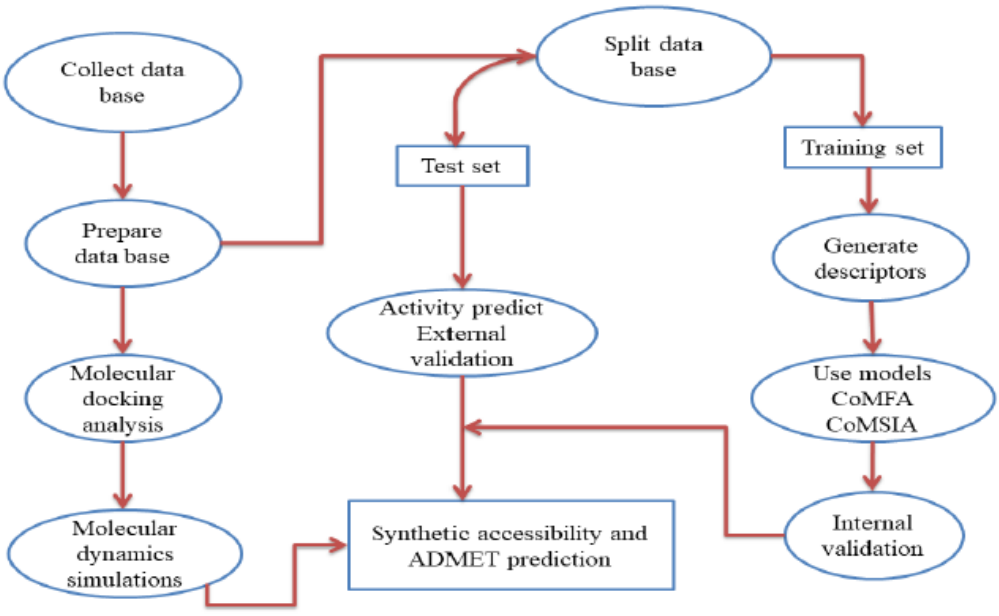


Figure 10 (6)

Materials and methods:

1. Dataset and data processing

In this study different CoMFA and CoMSIA models have been built based on a set of experimental data(28 derivatives). This dataset have been divided to 80% training set (22 derivatives) to construct y model and 20% test set (6 derivatives)to evaluate the performance of the built model. These steps have been made to evaluate their anti-cancer biological activity (IC50).

2. Molecular alignment and Generation of the Models:

Molecular alignment is an important step for generating CoMFA and CoMSIA models. Using the following steps:

a) The molecules of the new pteridinone derivatives were aligned using SYBYL-X 2.1 software and a stable configuration was obtained.
b) CoMFA and CoMSIA descriptors were computed for each lattice using steric, electrostatic, acceptor hydrogen bonding, and hydrophobic fields.
c) The partial least-squares (PLS) method was used to correlate the CoMFA and CoMSIA fields with the biological activity of pteridinone derivatives.

The predictive capabilities of the 3D-QSAR models were evaluated using an external test set and the predictive correlation coefficient (R2 pred) had to be greater than 0.6.

The QSAR model was considered good if the Q2 value was greater than 0.5. The leave-one-out (LOO) method was used to verify the predictive ability of the models.

Results:

- The models CoMFA (Q^2= 0.67, R^2 = 0.992), CoMSIA/SHE (Q^2 = 0.69, R2 = 0.974) and CoMSIA/SEAH (Q2 = 0.66, R2 = 0.975) models were used to study molecular modeling.
- The three models were satisfactory according to the results of the statistical validation (R^2 prediction value of CoMFA, CoMSIA/SHE and CoMSIA/SEAH models is 0.683, 0.758, and 0.767 respectively). We used these models to predict the activity of the molecules in the test set, and then we can use these models to predict the activity of new molecules as PLK1 inhibitors for prostate cancer treatment.
- A molecular docking study was performed to identify the type of binding between the most active ligand and the PLK1 inhibitor. The key amino acids affecting the activity of these inhibitors, such as R136, R57, and Y133, easily formed hydrogen bonds with the selected small molecules and a halogen bond with the L69 residue, so these different bonds could allow the PLK1 inhibitors to maintain stability in the binding site.
  Thus, the two selected ligands formed dynamically stable interactions with their protein during the 50 ns simulation time. Finally, an ADMET prediction of two ligands showed that only compound N° 28 (-R) could become a good drug candidate for cancer drug development. (Figure 10) (6)

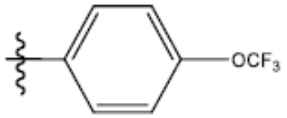
28 

Figure 11 (6)

3D-QSAR study of novel thieno[2,3-d][1,2,3] triazines targeting EGFR in lung Cancer

Lung cancer represents about one-third of cancer deaths.Non-small-cell lung cancer (NSCLC) is the major type of lung cancer affecting nearly 80–85% of patients.
The epidermal growth factor receptor (EGFR) is overexpressed in NSCLC cell lines of epithelial origin, especially H1299.Thus, lowering the EGFR concentration is a plausible target for the design and development of cytotoxic agents targeting H1299 lung cancer cells

This research conducts the design and synthesis of novel thieno[2,3-d][1,2,3]triazine derivatives as potential epidermal growth factor receptor (EGFR) inhibitors for the treatment of non-small-cell lung cancer. EGFR is a very important therapeutic target in lung cancer, and it is important to develop new inhibitors with improved efficacy. Thus, this study is based on pharmacophore modelling, 3D-QSAR analysis, and molecular modelling techniques to design and evaluate the synthesized compounds.
Studies have shown the super cytotoxic activity of the synthesized compounds compared to the approved EGFR inhibitors gefitinib and erlotinib. This paper emphasizes the potential of the designed compound 6b as a lead compound for the discovery of anti-lung cancer agents targeting EGFR.

Materials and Methods:

1. In this study, a data set consisting of 18 compounds with tetrahydrobenzothieno[2,3-d][1,2,3]triazine and dihydrocyclopentathieno[2,3-d][1,2,3]triazine scaffolds was synthesized and selected to establish a QSAR model. The inhibition potencies of these compounds, reported as IC50 values, ranged from 25 to 58 nM. The IC50 values were then converted into molar values and further converted into pIC50 values using the formula pIC50 = -Log(IC50). The 3D structures of the thienotriazine derivatives were created using the builder panel in Maestro software. These structures were subsequently optimized using the LigPrep module.

2. Pharmacophore 3D-QSAR modeling: Phase (v4.1) software was used to create pharmacophore and 3D-QSAR models for EGFR inhibitors. The ligands and their activity values (pIC50) were used to develop the pharmacophore model, with the ligands categorized as active, inactive, or moderately active. Phase (v4.0) shape

screening was employed for flexible alignment, using the highest active compound (6b) as a template and generating 100 conformers. The training and test sets were randomly divided in three trials to ensure the reliability of the QSAR model. The pharmacophore sites for thienotriazine training and testing included a predefined set of chemical features in Phase, such as hydrogen bond acceptor (A), hydrogen donor (D), hydrophobic interaction (H), and two aromatic rings (R).

3. The 3D-QSAR model was created using the partial least squares (PLS) regression statistics. The grid spacing was kept at 1Å .

- The best QSAR model in this study was ADHRR.26, which consisted of six components and utilized a PLS factor. When the model was applied to the training set ligands, it showed a correlation (R) of 0.85 with the observed inhibition of EGFR activity. The model's efficacy was further tested through external validation. A graph was plotted to compare the actual values with the predicted values.

Results:

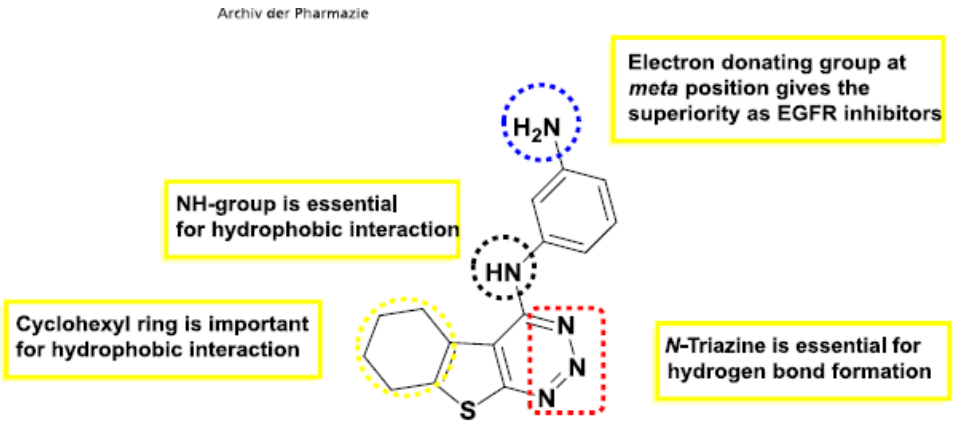


Figure 12 The observed structure–activity relationships of the newly synthesized compounds against H1299. EGFR, epidermal growth factor receptor (7)

- 3D-QSAR model prediction: The best fitted model ADHRR.26 (R^2=0.9908, Q^2= 0.8493, F= 97.10) consists of one hydrophobic interaction, one hydrogen bond donor, one hydrogen bon acceptor, and two aromatic features .
- The 3D-QSAR models in this study were analysed using contour maps. The presence of a hydrogen bond donor, specifically the amino group on C4 of the triazine ring, was found to be essential for the EGFR inhibitory activity. The hydrophobic contour highlighted the importance of hydrophobic interactions between the 4-anilino group and other bicyclic systems. The para substitution of the 4-anilino group enhanced the hydrophobic

interaction, consistent with the high potency of gefitinib. Another important position for hydrophobic interaction was the cyclohexenyl group adjacent to thiophene. The presence of electron-withdrawing groups at the meta and para positions of the anilino group attached to triazine was also necessary for activity, as seen in approved EGFR inhibitors like gefitinib and erlotinib. (7)

Conclusion

The 3D-QSAR and molecular modeling studies revealed comparable binding modes of compound 6b, gefitinib, and erlotinib in the EGFR active site **.** The pharmacokinetic features of compound 6b were predicted to have promising pharmacokinetic and physicochemical properties . The generated pharmacophore model (ADHRR) provided insights into the essential features required for EGFR inhibitory activity .

3D-QSAR study of thieno [3,2-b]pyrrole-5-carboxamide derivatives as LSD1 inhibitors

LSD1 is a a flavin adenine dinucleotide (FAD) dependent monoamine oxidase (MAO) that has regulatory effects on the physiological processes of the cell. Studies have shown that LSD1 is highly expressed in multiple types of tumor cells, including breast cancer, lung cancer, gastric cancer, acute myeloid leukemia, and other types of cancers. Thus, multiple clinical trials investigated a variety of monoamine oxidase inhibitors (MAOI) to hinder the function of this enzyme in cancer cells. However, the studied MAOIs were shown to have poor selectivity to LSD1 or poor inhibitory activity.

Thus, the purpose of this study was to analyze a group of thieno[3,2-*b*]pyrrop-5-carboxamides that have presented great potential as reversible inhibitors against LSD1, using molecular docking and 3D-QSAR modeling.

Materials and methods:

1. Fifty-five thieno[3,2-*b*]pyrrop-5-carboxamide-containing compounds were included in the data set.
2. CoMFA CoMSIA were used to establish 3D-QSAR models.
    a. Different molecular properties were used as independent variables to calculate the relationship between their structure and biological activities.
    b. pIC_{50} (log IC_{50}) were the dependent variable.
3. The data set was divided randomly through the criteria:
    a. The pIC_{50} values of the compounds in the test set should be distributed in various orders of magnitude in proportion to the whole set.

b. At the same time, the structures of the compounds in the test set should sufficiently represent the diversity of the whole dataset.

4. Forty-three compounds (78% of total compounds) were selected as the training set for the construction of CoMFA and CoMSIA models, while the remaining 12 compounds (22% of total compounds) were selected as an independent test set for validating the reliability of the model.
5. Energy minimization was performed using Powell gradient algorithm.
6. All compounds were calculated by Gasteiger–Huckel charges using the Tripos force field.
7. The reported crystal structures of LSD1 in complex with 5 ligands were obtained from Protein Data Bank.
8. Docked poses were then selected based on:
    a. The docking scores
    b. And empirical criteria Essential Chemical Interactions Described for Analogue Ligands (ECIDALs).
9. In CoMFA analysis, the steric field and electrostatic field were calculated by Lennard Jones and coulombic potential functions, respectively. The partial least square (PLS) method was used to analyze CoMFA and CoMSIA models. Cross-validation analysis was carried out with lease-one-out (LOO) to obtain the cross-validation coefficient q2 and the optimal number of components.

Results:

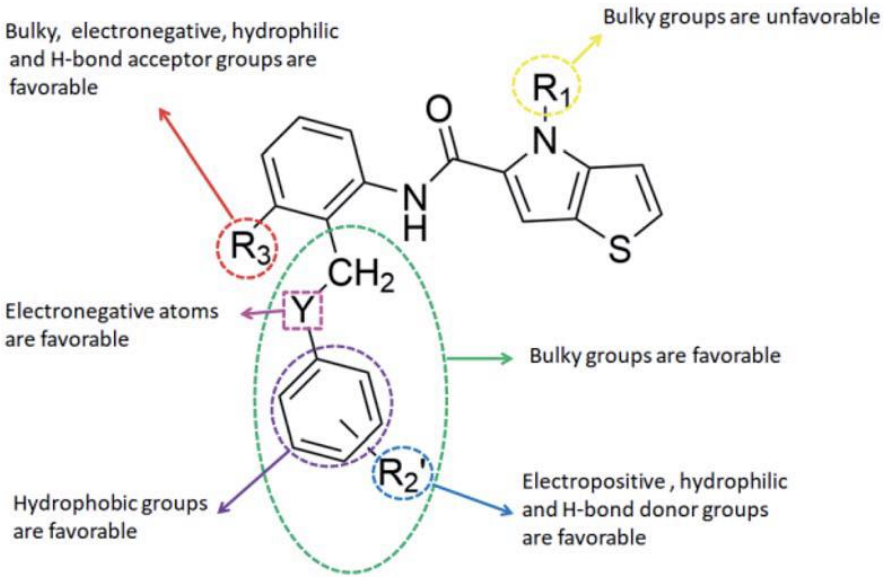


Figure 13 Structure–activity relationship (SAR) information obtained from 3D-QSAR Study (11)

1- The CoMFA (q2 = 0.783, r2 = 0.944) and CoMISA models (q2 = 0.728, r2 = 0.982) were received to perform molecular modeling study.
2- The 3D-QSAR modeled compounds revealed the following:

a. R1 substituents should not be too large, and R2 substituents should be appropriately increased in order to achieve the length of electronegativity region, and the structure of hydroquinone played an important role in improving the activity.
b. Introducing positive groups at the R2 region can be a potential modification for the structure because it allows for H-bonds to form between R2 and amino acids Asp555 and Asp556 on FAD.
c. Introducing hydrogen bond acceptor groups in R3 could improve the activity of the drug.

3- The structure–activity relationship (SAR) information revealed by the 3D-QSAR contour maps analysis was summarized in the figure above, which may be helpful in designing new LSD1 inhibitors with high activity.
4- In compound (54) some amino acids in the binding site were involved in hydrophobic interactions with the compound, while a few others caused unfavorable H-bonds affecting the compound's inhibitory effects.
5- Based on these 3D-QSAR models, 8 new small molecules were designed in silico. Further docking, molecular dynamic simulation, calculation of binding free energy, ADME and bioavailability prediction were carried out for these designed compounds and the results indicated that three of the four compounds (D4, D5 and D8) show good potential to become LSD1 inhibitors with better activity than the last compound (compound 54).
6- H-bond interactions between different amino acids and the compounds may lead to decreased inhibitory effects, while other increase the strength of the bond between the two molecules. (11)

Conclusion:

This study provides a theoretical guidance for a backbone structure for future studies to utilize in making a new LSD1 inhibitor.

Design of novel quinoline derivatives as anti- breast cancer using 3D-QSAR

Breast cancer is a cancerous development of cells in the breast and is the second leading cause of female cancer mortality. Breast cancer treatment options are surgery, chemotherapy, targeted therapy, radiotherapy, immunotherapy, and endocrine therapy. Hormone therapy is recommended for women with hormone receptor-positive malignancies and can target cancer cells in the breast and other parts of the body. Medications like Tamoxifen and Fulvestrant can inhibit the growth of breast cancer cells by reducing estrogen levels.

Drug resistance is a major challenge in cancer treatment and is one of the main causes of cancer-related deaths. Development of new quinoline derivatives can help in the discovery of novel breast cancer drugs. Computational and data science approaches, such as molecular modeling and docking, are valuable tools in drug development for breast cancer.

A 3D-QSAR model using the CoMSIA method was created to assist in the development of novel drug candidates for breast cancer.

Materials and methods:

1. The dataset concluded 16 different quinoline derivatives and it was divided randomly into a training set (12 molecules) and a test set (4 molecules)
2. The alignment process: molecular structures are produced using the sketch module on a platform like SYBYL-X and then optimized using Tripos force standard field and Gasteiger-Huckel atomic partial charges.
3. Conformation stability is achieved through iterations and a convergence threshold. The CoMFA model utilizes steric and electrostatic fields to generate descriptors, while CoMSIA adds physicochemical descriptors including hydrophobic, hydrogen bond donor (HBD), and hydrogen bond acceptor (HBA) fields.
4. Partial Least Squares (PLS) analysis is used to estimate parameter values, even with small sample sizes, by identifying correlations between CoMSIA descriptors and biological activity levels.
5. Performance evaluation is carried out using cross-validation (Q2), coefficient of determination (R2), Fischer test (F value), and standard estimation error (SEE) for model significance assessment.
6. The predictive ability of the 3D-QSAR models derived from the training set was evaluated by predicting the biological activities of four external test set molecules, using the determination coefficient of external validation (R2 pred) as a measure of prediction performance and model reliability.
7. Y-Randomization was conducted to ensure the stability of the QSAR model. The correlation coefficient of the randomized model (R2r) had to be less than that of the nonrandomized model (R2), and an additional parameter known as cR2r had to be greater than 0.5, indicating a credible model.
8. The CoMSIA model, based on electrostatics, hydrophobics, hydrogen bonds donor and acceptor (EHDA), showed the highest value of external coefficient of determination (R2 pred 0.755) among the five fields used, indicating a high level of stability and predictive quality.
9. The CoMSIA model was internally and externally tested to ensure its validity, and the ADMET characteristics of the newly developed drugs were evaluated after docking with the aromatase protein target.

Results

1. Molecule 7 was selected as the template for aligning the data in 3D-QSAR.
2. The CoMSIA investigation used seven combinations of fields to develop the model, with the best model having a Q2 value of 0.610 and R2 coefficient of 0.966.

3. The CoMSIA/EHDA model showed high stability and predictive quality, validated by external validation

4. The Y-Randomization test was used to verify the robustness of the CoMSIA/EHDA model, which showed that the model is not the result of random correlation.

5. The applicability domain of the model was determined using leverage analysis, which confirmed that the predicted activity levels of all molecules were correct

6. The CoMSIA model's contour maps for electrostatics were nearly identical to the CoMFA contour maps. The presence of highly electronegative groups or atoms near the functional carbon of the hydroxyl group was found to improve activity (9)

Conclusion:

Quinoline derivatives have shown potential as antitumor agents in breast cancer therapy, as supported by the 3D-QSAR and molecular docking studies

Pharmacokinetic profiling of quinazoline-4(3H)-one analogs as EGFR inhibitors: 3D-QSAR modelling for breast cancer treatment

Epidermal growth factor receptor (EGFR) plays a crucial role in cell evolution and metastasis in breast cancer. Chemotherapy is the primary treatment for breast cancer, but it has limitations and side effects, necessitating the development of improved and safer drugs.

In-silico aided drug-design techniques, such as 3D-QSAR, are used to design novel drugs by correlating the structural features of molecules with biological activity.

Materials and methods:

- A dataset of 36 quinazoline-4(3H)-ones was collected and and 2D- structures were built. Then the structures were transformed from 2D into 3D structures using Spartan v14.0 software. Energy minimization was analysed using density functional theory calculations.

- Ligand alignment was performed using SYBYL-X software, with compound 20 as the template.(compounds were allied on the quinazolin-4-one scaffold using the distill rigid module)

- The dataset was divided into 30 training and 6 test sets, and the inhibitive activities of compounds were converted to a logarithmic scale for modelling and validation.

- CoMFA modelling uses a 3D cubic frame with a grid layout and Tripos force field to create steric and electrostatic fields for compound analysis.

- CoMSIA modelling computes steric, electrostatic, hydrophobic, hydrogen bond donor, and acceptor fields using sp3 C+ and attenuation factor, and applies the partial least square approach for modelling studies.

Results:

1. PLS regression was employed for modeling purposes, revealing the best model (CoMFA_S) with the following assessment parameters: R2 = 0.872 using three latent components, Q2 = 0.597, and a standard error of 0.154
2. The CoMSIA_SHE model demonstrated the highest Q2 and R2 values of 0.666 and 0.982, respectively, utilizing nine latent components. The model also exhibited a standard error of 0.0655765.
3. The statistical significance of the CoMSIA_SHE model has been validated by conducting an applicability domain (AD) analysis. This analysis helped identify influential and structural outliers through a William's plot (Figure 13), which displays the standardized residuals of compounds against their leverage values.

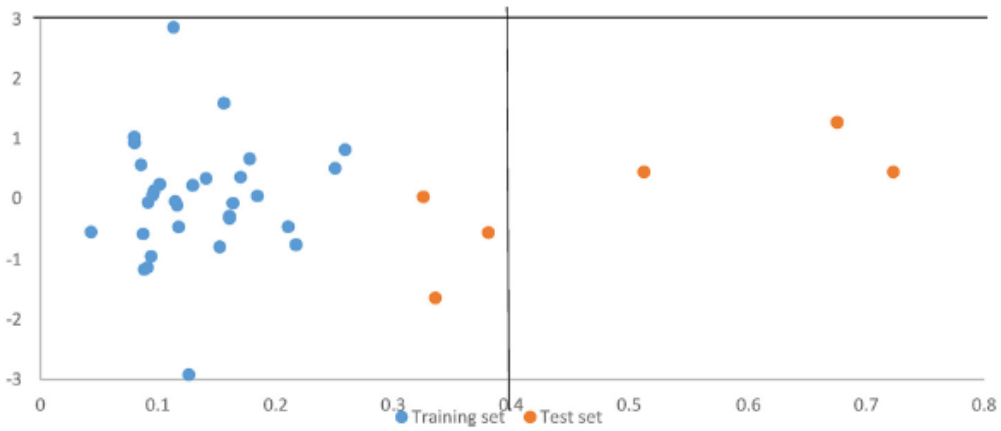


Figure 14 Williams Plot (8)

- William's plot of the model detected three influential compounds (17, 21, and 28) from the test set with leverage values greater than the threshold.

4. Several contour maps have been applied, with a result of the following:

- Introduction of bulky groups around certain regions, such as the 6,8-dibromo groups attached to the quinazoline-4-one scaffold, can increase bioactivity.

- Yellow contours in the CoMSIA_H contour map indicate that bulky groups around the nitrogen atom of the quinazoline-4-one scaffold and the ortho position of the phenyl group attached to the thiazolidinone group lead to a reduction in bioactivity.

- In the CoMSIA_E map, blue contours were observed near positions 4-6 and the C]O group of the quinazoline-4-one scaffold, as well as close to the sulfur atom, indicating that the introduction of hydrophobic groups in these regions can increase bioactivity.

- Red contours were found around the 2,3-positions of the quinazoline scaffold and close to the thiazolidinone C]O group, suggesting that electron-negative groups in these regions may reduce bioactivity

5. Five novel compounds were designed based on a template (Figure 14), aligned on a common core, and their activities were predicted, showing improved inhibitive activities, based on the CoMSIA_SHE model contour maps, in order to design new EGFR inhibitors with further techniques like molecular docking. (8)

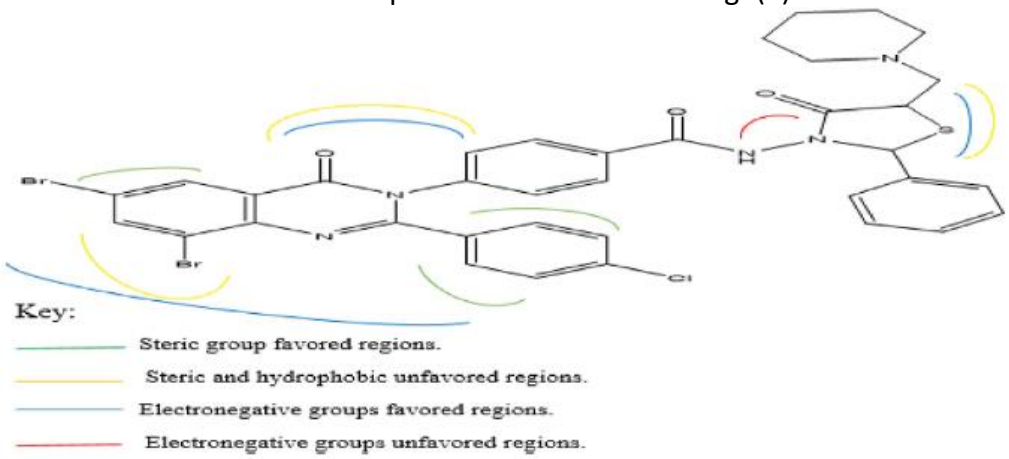


Figure 15 (8)

Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR

The enzyme Mcl-1 is involved in the anti-apoptotic process and is overexpressed in various types of cancer. The fight against cancer necessitates the development of more potent and less toxic anticancer agents. Computational techniques, such as computational aided drug design (CADD), have emerged as powerful tools in the drug discovery process. Coumarins, a class of organic compounds, have shown promising effects in various biological activities. In this study, CADD approaches were employed to identify novel coumarins as potent Mcl-1

inhibitors through 3D-QSAR analysis, ADMET prediction, molecular docking, and molecular dynamic simulations.

Materials and Methods

1. A dataset of 33 coumarin compounds with Mcl-1 inhibitory activity was collected. The IC50 values of these compounds were converted into pIC50 values. Computational studies were performed using Schrodinger suite's Maestro v12.8 software, and the compounds were energy minimized using Schrodinger's MacroModel.
2. For the 3D-QSAR analysis, the molecular alignment of the compounds was crucial. The 33 coumarin structures, along with their pIC50 values, were aligned using Maestro, with structure 4 chosen as the reference.
3. A field-based QSAR study was conducted to understand the relationship between the ligand electrostatic, hydrophobic, steric fields, hydrogen bond donors, hydrogen bond acceptors, and the biological activity. Gaussian field calculations were performed to determine the interaction energies.
4. The dataset was randomly divided into a training set (70%) and a test set (30%), with a maximum of 5 Partial Least Squares (PLS) factors utilized. The PLS method was used to correlate the Gaussian field descriptors as independent variables with the pIC50 values as the dependent variable.
5. To validate the developed model, external validation was performed. The criteria suggested by Golbraikh and Tropsha were used to assess the model's ideality, including cross-validated correlation coefficient (q2) for the training set, the correlation coefficient (r2) for the test set, and the slopes for regression through origin (k and k'). These criteria aimed to ensure the model's predictive power and accuracy.

Results:

1. The statistical parameters used to evaluate the models included r2, q2, standard deviation (SD), F-value, p-value, root-mean-square error (RMSE), and Pearson-regression.
2. The PLS model with 5 factors showed significant statistical results, with an r2=0.80, q2 = 0.81, low SD = 0.204, high F-value = 16.1, RMSE close to 0, and Pearson-R close to 1.
3. The best model (PLS factor 5) revealed the contributions of different fields, with steric, electrostatic, hydrophobic, H-bond acceptor, and H-bond donor fractions of 0.28, 0.11, 0.21, 0.18, and 0.23, respectively.
4. The plot displayed in (Figure 15) demonstrated a significant correlation between the predicted and actual activity values.

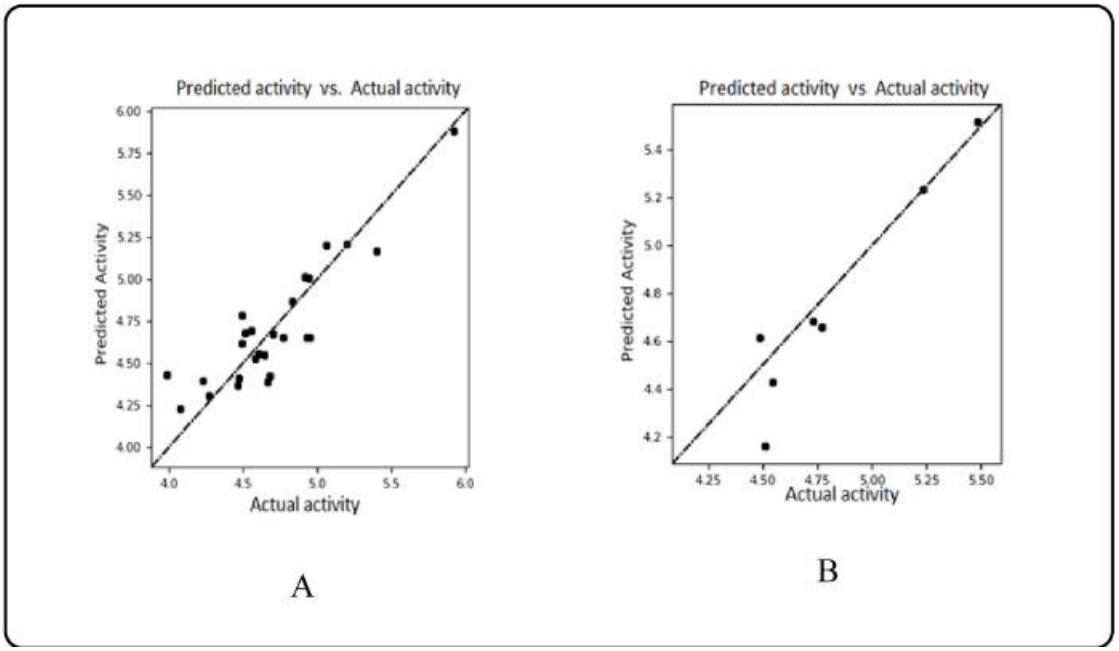


Figure 16 Relationship between experimental and predicted Mcl-1 inhibitory effect (A)Training set values, (B)Test set values. (4)

5. The 3D-QSAR model contours provided insights into the fields that impact the anticancer activity of the studied molecules. The contours revealed that electropositive and electronegative groups at specific positions, H-bond acceptors, hydrophobic groups, bulky groups, and hydrogen bond donor groups enhance the activity.
6. Contour maps analysis revealed that the Mcl-1 binding affinity could be improved through conformationally restricted substitutions at R1 and R5 in the reference compound. The compounds designed based on these parameters exhibited better predicted activity than the reference compound. (4)

Conclusion of the study:

- QSAR (Quantitative Structure-Activity Relationship) and drug design studies have been conducted using various computational techniques such as 3D-QSAR, molecular docking, and ADMET analysis. These studies have shown promising results in predicting the activity of new compounds and identifying structural variables that influence their activity
- 2D-QSAR applications are limited in drug design, the reason why we use 3D – QSAR to complete the study. But various 3D-QSAR studies have been found and proved its success in drug design with the aid of other computational techniques.
- 2D-QSAR have been used successfully in predicting some basic information's about the studied compounds.

- 3D-QSAR is a high throughput, easy and more economic method for drug design and development.
- The derived QSAR models have been validated using statistical tests such as correlation coefficient (r2) and cross-validation regression coefficient (q2) , Y-Randomization test.
- The models have demonstrated good predictive performance and significant correlation between the structural features of molecules and their inhibitory activities. Additionally, molecular docking simulations have provided insights into the binding modes of compounds and their interactions with target receptors.
- These findings have implications for the design of new drug candidates with improved efficacy and safety profiles . However, it is important to note that experimental validation of these theoretical findings is necessary to confirm their potential as drug candidates .
- The application of 3D-QSAR method has shown promising results in the treatment of various types of cancer, including prostate cancer, lung cancer, osteosarcoma, and breast cancer.
- 3D-QSAR allows for the rational design of more effective and selective compounds, leading to improved cancer treatment outcomes.
- Contour Maps identify regions where specific properties positively or negatively influence the activity, this identification has proven its success in prediction and design of new compounds with improved activity.

Summerazation of methods used in the 3D-QSAR studies (This table was constructed by the student)

Study	Data processing	Molecular generation and Alignment	Correlation	Validation
3D-QSAR studies of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer	80% training set, 20% test set	Alignment using SYBYL-X 2.1	Partial least square analysis	• External Validation • Leave- one-out method
3D-QSAR study of novel thieno[2,3-d][1,2,3] triazines targeting EGFR in lung Cancer	The IC50 values of 18 compounds were converted to pIC50 values. The 3D structures were created in Maestro software. And then optimized using LigPrep model	Phase(v4.1) software was used to create 3D-models and for flexible alignment, after that the data set was devided into three trials	Partial least square analysis	External Validation
Design of novel quinoline derivatives as anti- breast cancer using 3D-QSAR		• Alignment using SYBYL-X • Optimization using Tripos force field and Gasteiger-Huckel atomic partial charges	Partial least square analysis	• Performance evaluation: Cross-validaton, Coefficient of determination, Fischer test, standard estimation error(SEE). • External Validation • Y-randomization test
Pharmacokinetic profiling of quinazoline-4(3H)-one analogs as EGFR inhibitors: 3D-QSAR modelling for breast	2D- structures were built. Then the structures were transformed from 2D into 3D structures using Spartan v14.0 software. Energy minimization was	alignment was performed using SYBYL-X software	Partial least square analysis	Applicability domain (AD) analysis (Williams plot)

cancer treatment	analysed using density functional theory calculations.			
Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR	The IC50 values of these compounds were converted into pIC50 values. Computational studies were performed using Schrodinger suite's Maestro v12.8 software. Energy minimization using Schrodinger's MacroModel.	Alignment using Maestro	Partial least square analysis	External Validation

References

1. (1)Kubinyi, Hugo, n.d. QSAR and 3D QSAR in drug design Part 1: Methodology
2. (2)Kubinyi, Hugo, n.d. QSAR and 3D QSAR in drug design Part 2: Applications and problems
3. (3)Umma Muhammad,1 Adamu Uzairu,2 David Ebuka Arthur2, 2018. Review on: quantitative structure activity relationship (QSAR) modeling
4. (4)Esraa Albasher Osman a,1, Mohammed Abdalrahman Abdalla a,1, Mohja Omer Abdelraheem a,1, Mubarak Fadlalla Ali a,1, Shima Albasher Osman a,1, Yasmin Mohamed Tanir a,1,, Mohammed Abdelrahman b, Walaa Ibraheem a, Abdulrahim A. Alzain, 2021. Design of novel coumarins as potent Mcl-1 inhibitors for cancer treatment guided by 3D-QSAR, molecular docking and molecular dynamics.
5. (5)Manman Zhao,1 Lin Wang,2 Linfeng Zheng,3 Mengying Zhang,1, Chun Qiu,2 Yuhui Zhang,4 Dongshu Du,1,5 and Bing Niu1, 2017. 2D-QSAR and 3D-QSAR Analyses for EGFR Inhibitors.
6. (6)Mohamed El fadili 1 , Hamada Imtara 2 , Aamir Saeed 3, Abid Ur Rehman 4, Sara Zarougui 1, Shaef A. Abdullah 5, Ahmad Alahdab 6,*, Mohammad Khalid Parvez 7 and Menana Elhallaoui, M.E. 1, *, 2023. 3D-QSAR Studies,Molecular Docking,Molecular Dynamic Simulation, and ADMET Proprieties of Novel Pteridinone Derivatives as PLK1 Inhibitors for the Treatment of Prostate Cancer.
7. (7)Ranza Elrayess1, Yasmine M. Abdel Aziz1, Mohamed S. Elgawish2, Marwa Elewa1, Hosam A. Elshihawy1, Mohamed M. Said1, 2019. Pharmacophore modeling, 3D-QSAR, synthesis, and anti-lung cancer evaluation of novel thieno[2,3-d][1,2,3]triazines targeting EGFR.
8. (8)Sagiru Hamza Abdullahi, PhD *, Adamu Uzairu, PhD, Gideon Adamu Shallangwa, PhD,, Sani Uba, PhD and Abdullahi Bello Umar, PhD, 2023. Pharmacokinetic profiling of quinazoline-4(3H)-one analogs as EGFR inhibitors: 3D-QSAR modeling, molecular docking studies and the design of therapeutic agents.
9. (9)Said El Rhaboria , Abdellah El Aissouqa,*, , Samir Chtitab,* and Fouad Khalila,*, 2022. Design of novel quinoline derivatives as antibreast cancer using 3D-QSAR, molecular docking and pharmacokinetic investigation.

10. (10)Wenkun Zhuo1†, Zheng Lian1, Wenzhe Bai2†, Yanrong Chen2* and, Huanling Xia, 2023. 3D- and 2D-QSAR models' study and molecular docking of novel nitrogen-mustard compounds for osteosarcoma.
11. (11)Yongtao Xu, *abc Zihao He,abc Hongyi Liu,abc Yifan Chen,abc Yunlong Gao,abc, Songjie Zhang,abc Meiting Wang,abc Xiaoyuan Lu,a Chang Wang,a Zongya Zhao,a, Yan Liu,a Junqiang Zhao,a Yi Yua and Min Yang, 2020. 3D-QSAR, molecular docking, and molecular dynamics simulation study of thieno[3,2-b]pyrrole- 5-carboxamide derivatives as LSD1 inhibitors.
12. (12)https://www.creative-biolabs.com/