3D QSAR of Coumarin derivatives against MCF-7 Breast cancer Cell Line

A thesis submitted

In partial fulfilment of the Requirements for the Degree Of

MASTER OF TECHNOLOGY

In

BIOINFORMATICS

Submitted by:

Atanu Mahapatra

MBI2019009

Under the supervision of:

Dr. Nidhi Mishra

Assistant Professor, IIIT Allahabad

To the:

DEPARTMENT OF APPLIED SCIENCES

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

JUNE 2021

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this thesis entitled **"3D QSAR of Coumarin derivatives against MCF-7 Breast cancer Cell Line"** submitted towards fulfilment of MASTER'S THESIS with specialization in BIOINFORMATICS at Indian Institute of Information Technology, Allahabad, is an authenticated record of my original work carried out under the guidance of **Dr. Nidhi Mishra.** Due acknowledgements have been made in the text to all other materials used. The thesis was done in full compliance with the requirements and constraints of the prescribed curriculum.

Atanu Mahapatra **Date:**
Enrollment No: MBI2019009
Department of Applied Sciences
IIIT Allahabad

CERTIFICATE FROM SUPERVISOR

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief. The master's thesis titled **"3D QSAR of Coumarin derivatives against MCF-7 Breast cancer Cell Line"** is a record of the candidate's work carried out by her under my guidance and supervision. I do hereby recommend that it should be accepted for the fulfilment of the requirement of the MASTER'S THESIS at IIIT Allahabad.

Date:

Place: Allahabad

Dr. Nidhi Mishra
Assistant Professor
IIIT Allahabad

CERTIFICATE OF APPROVAL

The forgoing thesis is hereby approved as a credible study in the field of Bioinformatics carried out and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein, but approve the thesis only for the purpose for which it has been submitted.

Signatures of the Examiner's Committee Members (On final examination and approval of the thesis).

Dr. Nidhi Mishra
Thesis Mentor
(Asstt. Professor, Dept. of Applied Sciences, IIIT Allahabad)

ACKNOWLEDGEMENT

The extensive endeavour, bliss euphoria that accompanies the successful completion of any project work would not be complete without the expression of gratitude to people who made it possible and whose constant guidance and encouragement crown all the efforts. This thesis was not only an endeavour but also a great and interesting learning experience for me and it bears the imprint of a number of people who directly or indirectly were a great source of help and encouragement.

I would like to express my sincere gratitude to my mentor Dr. Nidhi Mishra, for her continuous motivation and guidance. Her valuable suggestions, comments and support were of immense help to me. I am grateful to her for taking out time from her busy schedule and being very supportive in guiding my work.

I would also like to thank my senior, Mr. Akhilesh Kumar Maurya, for constantly guiding me throughout my work, and helping me out whenever I needed.

Special thanks to my parents and peers at IIIT-A for their efforts and cooperation. The interesting and informative discussions we had together greatly contributed to the completion of this work.

TABLE OF CONTENTS

1. Introduction

Breast cancer is one of the most frequent malignancies that women throughout the world face. Breast cancer is the second leading cause of death for women after lung cancer, with a lifetime chance of contracting it expected to be about 12%. Estrogen receptor (ER) positive, progesterone receptor (PR) positive, or human epidermal growth factor receptor 2 (HER2) positive breast cancer are the three types of breast cancer and all three types promote mammary cell proliferation. Since oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) receptors are expressed in receptor sensitive breast cancers, they are good candidates for therapy at various levels.

These are the treatments:

i) chemotherapy, hormone treatment with tamoxifen or an aromatase inhibitor including coumarin derivatives, and drug targets for human epidermal growth factor receptor 2 (HER2)

ii) surgery

iii) radiation - The lack of all three estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) receptors has resulted in the discovery of triple-negative breast cancer (TNBC).

The most aggressive type of breast cancer, triple-negative breast cancer, has little treatment options. In recent years, triple-negative breast cancer (TNBC) has become twice as common in women of African heritage, especially among younger and premenopausal women. In 2012, the number of new instances of breast cancer increased considerably, with about 1.7 million new cases diagnosed. Breast cancer claimed the lives of roughly 40,610 persons in the United States in 2017, with 63,410 instances of ductal carcinoma in situ (the earliest kind of breast cancer) and 252,500 new cases of invasive breast cancer recorded.

According to a survey, between 10% and 20% of all new instances of breast cancer will most likely be identified as triple-negative breast cancer (TNBC). The prognosis for triple-negative breast cancer (TNBC) is poor, with a significant risk of recurrence, a shorter progression-free survival duration, and a high death rate. About 15% of triple-negative breast cancer (TNBC) patients develop brain metastases, and post-mortem autopsy have revealed a 30% risk of metastatic development. In the development of novel treatments, understanding the molecular basis of triple-negative breast cancer (TNBC) is becoming more crucial. Patients with hormone receptor-positive tumours have a better prognosis, according to a new review, since they are exposed to chemotherapeutic medicines that target endocrine receptors/proteins, which are the quickest means to reduce carcinogenesis morbidity and mortality. Owing to the absence of pathway-specific target proteins and approved target therapies, triple-negative breast cancer (TNBC) does not have a target therapy. Patients with triple-negative breast cancer (TNBC) have benefitted from a limited number of standard chemotherapy regimens, but adverse effects include high cytotoxicity and chemotherapeutic resistance have severely limited chemotherapeutic treatment options. Bad prognosis and a lack of targets are the problems with targeted medication for breast cancer. To find a new medicine, specifically for advanced types of breast cancer, we need a special drug that inhibits the target protein.

Coumarin and its derivatives have appeared as one of the most functional molecular scaffolds in recent decades, with a broad range of pharmacological and therapeutic applications. Coumarin derivatives are used in a wide range of pharmacologically active molecules and share a similar structural function. Since coumarin derivatives share a similar structural function, it's simple to make coumarin derivatives that are pharmacologically active against specific proteins, and they're also a good candidate for structural modifications and derivatization. Furthermore, coumarin hybrids have been discovered to have a variety of medicinal uses, including photochemotherapy, antitumor activity, and the ability to function as central nervous system stimulants (CNS). Furthermore, through scavenging reactive oxygen molecules, hydroxycoumarins may be used to avoid free radical damage. The active metabolite, 7-

hydroxycoumarin conjugates, has also been shown to inhibit sulfatase and aromatase. Despite the fact that coumarin derivatives have a variety of biological functions, the most intriguing fact is that some of them have a strong pharmacological effect against breast cancer. Coumarin-based hybrids have been shown to suppress cell proliferation in the MCF-7 breast cancer cell line in a number of experiments.

Coumarin (2H-1-benzopyran-2-one) has a ring system and is used in a variety of drugs, including the anticoagulant warfarin. It has some very interesting pharmacological properties. To investigate the therapeutic applicability of natural coumarins and their synthetic analogues, this medicinal chemist would require decades of experience. Various coumarin ring-based molecules can be generated using synthetic methods, and these coumarin derivatives have a variety of pharmacological properties. Furanocoumarins, pyranocoumarins, and coumarin sulfamates, for example, have been used as antitumor, photochemotherapy, and anti-HIV therapy, as well as central nervous system stimulants, antibacterial, anti-inflammatory, anti-coagulants, and dyes, due to the multiple orientations of coumarin molecules. Some coumarin derivatives and their active metabolite 7-hydroxycoumarin analogues have been shown to suppress sulfatase and aromatase in breast cancer chemotherapy. SERMs based on coumarin and coumarin-estrogen conjugates have also shown promise as antibreast cancer agents. Since breast cancer is the second leading cause of death of women, after lung cancer, there is a strong incentive to develop new drugs for breast cancer care. As a result, the primary goal of this research is to look into essential coumarin derivatives with antibreast cancer properties, as well as to illustrate their mechanisms of action and quantitative structure-activity interactions on selected receptors in breast cancer cells, as well as to look into various methods for synthesising these pharmacologically significant coumarin derivatives.

2. MCF-7 Cell lines

MCF-7 (Michigan Cancer Foundation-7) is a Detroit-based institute where Herbert Soule and colleagues obtained the cancer cell line in 1973. The Barbara Ann Karmanos Cancer Institute has replaced the Michigan Cancer Foundation. Females are more likely to develop breast cancer. In basic cancer research, cancer cell lines such as the oestrogen receptor (ER) MCF7 cell line, which has been utilised in cancer science for over 45 years, are often employed. MCF-7 is a cell line that was initially identified in 1973 from the breast tissue of a 69-year-old Caucasian lady named Sister Catherine Frances Mallon. She was a nun at the Immaculate Heart of Mary convent in Monroe, Michigan. The first of two mastectomies she underwent showed that the tissue removed was benign. Five years later, another surgery revealed a malignant adenocarcinoma in a pleural effusion, from which the MCF-7 cell line was derived.

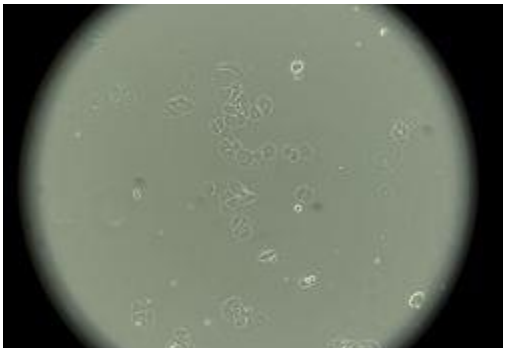


Figure 1: MCF-7 cell

Radiotherapy and hormonotherapy were used to cure her breast cancer. It was impossible for cancer researchers to obtain a mammary cell line that could last more than a few months before MCF-7 cell lines, but now MCF-7 cell lines will last longer than a few months. The patient, sister Frances Mallon died in 1970. Her cells were the source of much of current knowledge about breast cancer.

2.1. Characteristics of MCF-7 cell lines

The ability to process estradiol through cytoplasmic estrogen receptors and the ability to shape domes were both retained in MCF-7 cell lines. It's a primary tumour, also known as invasive breast ductal carcinoma, that develops from pleural effusion. Estrogen receptors can be found in the cell. Estrogen-induced proliferation. Progesterone receptors can be found on cells. It is impossible for a cell to have Amplification of the ERBB2 gene is not permitted.

MCF-7 cells maintain estrogen sensitivity while still being cytokeratin sensitive. Desmin, endothelin, GAP, and vimentin are all non-receptive to them. MCF-7 cells form in monolayers and resemble epithelial cells. The cell line will shape domes and process estradiol via cytoplasmic estrogen receptors when developed in vitro. Anti-estrogens can inhibit the proliferation of MCF-7 cancer cells via modulating insulin-like growth factor proteins, which can be utilized to treat them. MCF-7 cells are easy to propagate, but experts say they are a slow-growing cell. It normally takes 30-40 hours for MCF-7 cells to double.

2.2 Uses for MCF-7 Cell Lines

MCF-7 cell lines are epithelial cancer cells originating from breast adenocarcinoma that have been extensively studied. It resembles separated mammary epithelial cells. This cell lines can be used for detecting PI3K and MAPK involvement also it can use for easy detection of ERK and Akt phosphorylation. MCF-7 cells are employed in vitro for breast cancer research because they preserve many of the desirable properties of mammary epithelial cells. Estrogen is processed in the form of estradiol in MCF-7 cells via oestrogen receptors in the cytoplasm. MCF-7 cells express progesterone receptors but do not express HER2. MCF-7 cells are the most commonly employed in detecting oestrogen receptor (ER)-positive breast cancer cells, and numerous subclones have been developed to show distinct forms of ER-positive tumours with varying receptor expression.

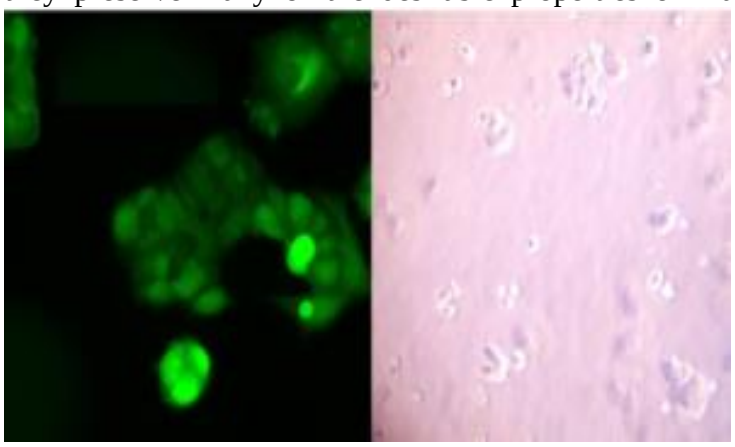


Figure 2: MCF-7 cell

2.3 Stability of cell line

The MCF-7 line is not genetically identical to the original isolated clone. MCF-7 cell lines used to have 85 chromosomes, but that number has since been decreased by 16 chromosomes. The modal chromosome number of the MCF-7 cell line is now 82.

2.4. MCF-7 Molecular Profile

These cells are E2-sensitive to estrogen and have high oestrogen receptor (ER) transcript levels but low oestrogen receptor (ER) levels. MCF7 cells increase oestrogen receptor expression in the absence of estrogens (ER). An autocrine substance that activated the insulin-like growth factor receptor was shown to be required for MCF7 cells' response to E2 (IGF-IR). The proliferation of breast cancer cells can be controlled by the progesterone receptor (PR) and oestrogen receptor (ER), as well as plasma membrane-associated growth factor receptors.

3. Coumarin

Coumarin (2H-1-benzopyran-2-one) is an aromatic organic chemical compound with formula $C_9H_6O_2$. This molecule has a benzene molecule with two adjacent hydrogen atoms replaced by a lactone-like chain −(CH)=(CH)−(C=O)−O−, forming a second six membered heterocycle that shares two carbons with the benzene ring. This molecule can be placed in the benzopyrone chemical class and considered as a lactone. Coumarin can most often found in various herbal compounds such as sweet clover, lavender oil, woodruff, and tonka beans as well as in various edible plants like strawberries and celery. Coumarin is a colorless crystalline solid with a sweet odor resembling the scent of vanilla and a bitter taste.

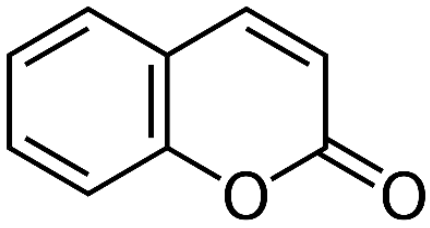


Figure 3: Coumarin

Coumarin and its derivatives are abundant in nature, and many of them have beneficial and various biological activities. Coumarins can be found in the leaves, roots, and seeds of a wide variety of plants. Tonka bean has a high concentration of coumarin in nature. French name for tonka bean is coumarou, this word is the source of the name cumarin. Real and synthetic coumarin derivatives are classified into many subclasses. The most often reviewed coumarins are simple coumarins (such as coumarin and limettin), furanocoumarins (such as imperatorin and isopimpinellin, psoralen, and angelicin), and pyranocoumarins (such as imperatorin and isopimpinellin, psoralen, and angelicin) (e.g. xanthyletin, seselin). Murray et al categorised coumarins based on the quantity of nuclear oxygen atoms discovered in coumarin-containing substances, using a biogenetic method.

Various study found that many coumarin derivatives have numerous therapeutic benefits including antitumor, photochemotherapy, anti-HIV therapy and also use as central nervous system stimulants. Antibacterial, anti-inflammatory, anti-coagulants, and dyes are all possible applications. Coumarins are lipid-lowering drugs with a minor triglyceride-lowering action. Hydroxycoumarins is a powerful chain-breaking antioxidant that protects against free radical damage by scavenging reactive oxygen species. Some coumarin derivatives are employed as food adulterants and flavouring agents, but excessive quantities can induce hepatotoxicity and side effects including moderate nausea and diarrhoea. Coumarin-based medicines are now marketed as therapeutic treatments in many European countries. Coumarin derivatives are used to treat lymphoedema in Europe, however they are not authorised for therapeutic use in the United States owing to hepatotoxicity. However, new research has discovered that coumarins exhibit modest estrogenic action, suggesting that their derivatives might be used as therapeutic agents to treat menopause-related illnesses including osteoporosis, cardiovascular disease, and cognitive deficits.

The addition of a catechol to the structure of coumarin has been reported to boost the cytotoxic activity in tumour cell lines in several studies. In a well-established model of multistage carcinogenesis, naturally occurring coumarins were used to create mouse skin tumours and successfully treat the tumours. The pharmacological and biochemical characteristics of basic coumarin structure can be influenced by substitutions of the fundamental chemical structure. These substitute coumarin has therapeutic applications, and can reduce the toxicity level. The substituted benzopyranobenzothiazinones a coumarin derivate expressed estrogenic activity against MCF-7 breast cancer cell lines.

4. Quantitative Structure Activity Relationship

In chemical, biological, and technical sciences, QSAR models are fundamentally regression or classification models. QSAR regression models, like other regression models, include a set of predictor variables (x) and a response variable (y) (Y).

The predictor variables (x) in QSAR are physico-chemical characteristics or theoretical molecular descriptors of chemicals, while the response variable (Y) is the biological behaviour of a chemical. QSAR models were used to uncover the first correct connection between chemical structures and biological behaviour of molecules in a data set of compounds. QSAR models can be used to predict the biological activity of novel compounds.

The concentration of a drug required to trigger a certain biological reaction is known as biological activity. When quantitatively representing physicochemical characteristics or chemical structures, a mathematical connection known as the quantitative structure-activity relationship can be constructed. If the mathematical expression is rigorously validated using experimental data, this model may be used to anticipate the modelled reaction of different chemical structures.

The following mathematical model can be used to define a QSAR:

Biological Activity = *f* (physiochemical properties and/or structural properties) + error

Even with a proper model, empirical variability and model error (bias) might be detected.

4.1 Molecular Descriptors

The molecular descriptors are chemical knowledge about a molecule represented by a logic and mathematical method that converts it into a useful number or the outcome of a standardised experiment. Descriptors can be classified as 1D (one Dimension), 2D (two Dimension) and 3D (three Dimension).

4.1.1 1D Descriptors

1D (one Dimension) descriptors are

- The total number of atoms in a molecule
- The number of C, H, O, S, N, F, Cl, Br, I, and P atoms in molecules, both pure and relative
- The number of bonds between molecules, double, triple, and aromatic bonds, for example, are all types of bonds.
- Number of benzene rings divided by the number of atoms in a molecule (number of benzene rings divided by the number of atoms in a molecule).
- The total atomic weight of molecules and their molecular weight.
- Number of rotatable bonds, ignoring all terminal hydrogen atoms.
- Number of hydrogen bond acceptors (Hbond acceptor)
- Number of hydrogen bond donors (Hbond donor)

4.1.2 2D Descriptors

2D descriptors are

- Topological Descriptors- Topological descriptors based on molecular graph representation are commonly used in QSPR and QSAR studies because they help to distinguish molecules based on their size, degree of branching, stability, and overall form.
- Topological Polar Surface Area (TPSA)- topological polar surface area also known as the polar surface area (PSA) of a molecule. It is the surface sum over all polar atoms or molecules, primarily oxygen (O) and nitrogen(N), also hydrogen atoms (H) attached with them.

4.1.3 3D Descriptors

3D descriptors are

- Steric parameters are
  - Length-to-breadth ratio (L/B) of molecules.
  - Molecular thickness
  - Ovality- ratio of the actual surface area and minimum surface of molecules
  - Molecular volume
  - Sterimol parameters
  - Taft steric parameter E_s
- Quantum chemical descriptors are
  - Atomic charges of molecules
  - LUMO - Lowest occupied molecular orbital energy
  - HOMO - Highest occupied molecular orbital energy
  - DIPOLE - Components of dipole moment along inertia axes
  - Hf - Heat of formation
  - Mean Polarizability, $\alpha = 1/3(\alpha_{xx}+\alpha_{yy}+\alpha_{zz})$
  - EA – Electron Affinity
  - IP – Ionization Potential
  - ΔE – Energy of Protonation
  - Electrostatic Potential, $V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')dr'}{|r'-r|}$

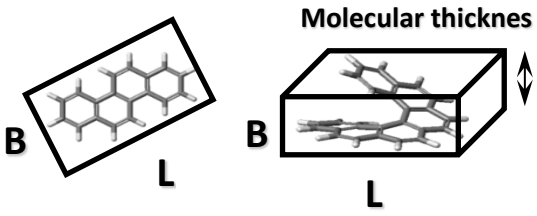


Figure 4: Molecule

4.2 3D-QSAR

3-D QSAR is a method of determining the three-dimensional structures of a group of tiny molecules whose biological activities are known using a force field. This is a set for training. Either experimental data or molecular superimposition tools must be used to align the training set. The results of the experiments are based on ligand-protein crystallography. Instead of using experimental constants, 3-D QSAR software employs calculated potentials, such as the Lennard-Jones potential, and it is concerned with the entire molecule rather than a single molecule. Cramer et al. developed the first 3-D QSAR technique, Comparative Molecular Field Analysis (CoMFA). It examined the molecule's shape, as well as the steric and electrostatic fields. To link them, the partial least squares regression (PLS) method is utilised.

4.2.1 Comparative molecular field analysis (CoMFA)

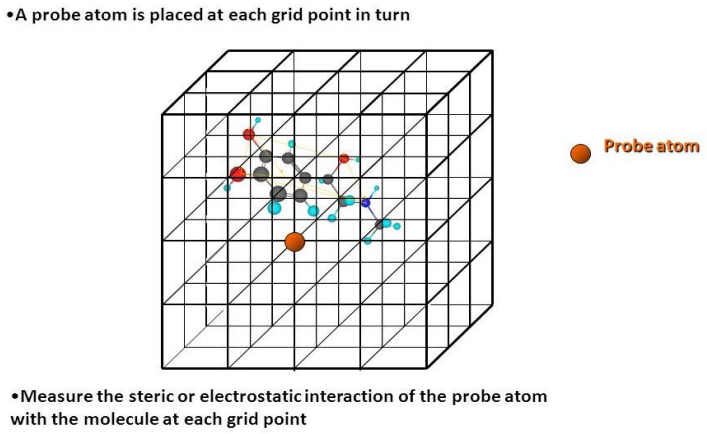


Figure 5: Prob atom is placed in a grid

Comparative molecular field analysis is a ligand-based approach that is based on molecular fields. This approach establishes a quantitative link between molecular structures and their reaction characteristics. This approach focuses on ligand properties such as steric and electrostatic interactions, yielding favourable and unfavourable receptor–ligand interactions. Since comparative molecular field analysis is a descriptor-based technique, all of the aligned ligands are placed in an energy grid box, and a suitable probe is placed at each lattice point, the energy is calculated. The measured energy corresponds to electrostatic force Coulombic and steric force van der Waals at each unit fraction of the grid box. These calculated values are descriptors, which are used in the construction of models. The biochemical behaviour of molecules is then combined with these descriptors, resulting in a rigorous linear regression approach such as partial least squares (PLS). The PLS findings are used to determine a molecule's favourable and unfavourable electrostatic and steric ability, as well as to link it to biological activity.

4.2.2 Methodology of CoMFA

Quantitative structure-activity relationships (QSAR) have been applied from long ago in the development of relationships between physicochemical properties and biological activities of chemical substances to obtain a reliable statistical model for activities of new chemical entities prediction. Affinities of ligands to their binding sites, inhibition constants, other biological endpoints, with atomic group or molecular properties and rate constants are studies in classical QSAR. The difference in structural properties is responsible for the variations in biological activities of the compounds is the fundamental principal which underlying the formalism. For QSAR models validation, the strategies which adopted are various:

1. Cross-validation or internal validation is used to measure data as it is being extracted. The further data retrieval perturbs the original model, the more stable the model is. A calculation of model robustness is cross validation.
2. External validation by dividing the available data collection into a training set for model creation and a forecast set for model predictivity testing;
3. external validation by blind measuring the model on new external data.
4. To ensure that the modelling descriptors and the solution test Y-scrambling or data randomization do not have a chance association.

Statistical methods, input data precision, descriptor selection, and, most importantly, validation of the established model are all important factors in the performance of a QSAR model. By the use of validation process and for a specific purpose the reliability and relevance of a procedure are established. Validation of QSAR models should be based primarily on the models' applicability domain, robustness, and prediction efficiency.

There are a variety of validation methodologies that can be troublesome, such as determining whether the training and test sets were chosen to optimise the predictive ability of the model being released in external validation, and leave one-out cross-validation contributes to an overestimation of predictive capacity.

Setting training set scale, methods of selecting training set compounds, and the effect of variable selection for training set models for deciding the accuracy of prediction are all factors that must be considered when validating QSAR models. The creation of novel validation parameters is also critical for assessing the consistency of a QSAR model.

- Molecular Modeling: 3D-QSAR was used to conduct molecular modelling experiments on 50 compounds. The GAFF was used to conduct energy minimization after the 3D structures of the molecules were constructed (General AMBER Force Field)

- Molecular Alignment: Using the most active compound as the reference structure 36, the data set was aligned on the common core using a rigid alignment technique. The most important routine for creating a stable 3D-QSAR model is molecule structural alignment. The alignment has a major impact on the estimation accuracy and statistical efficiency of 3D-QSAR models. Describe the proposed alignment and the alignment's traditional substructure.
- CoMFA Studies: CoMFA models were produced by automatically importing aligned molecules into a 3D cubic lattice with a grid spacing of 2. Using an sp3 or c3 carbon probe atom with a Van der Waals radius of 1.52 and positive charge +1 at each lattice point and a distance-dependent dielectric, the General AMBER Force Field creates electrostatic Coulombic potential fields and steric Lennard-Jones (6-12) potential field energies. At speed up the measurement of potentials and reduce noise, the column filtering value has been set to 2.0 Kcal/mol as a standard parameter.
- Partial Least Square Analysis: The CoMFA models were produced and internally validated using PLS regression analysis and cross-validation tests. PLS is the most widely used method for determining multilinear relationships between dependent and independent variables. The cross-validation test will then be used to assess the models' self-consistency and it generated by PLS. The explanatory characteristics of the CoMFA descriptors are utilised as independent variables, while the IC50 (target properties) is used as a dependent variable. The ideal number of components and cross-validated correction coefficient were found in the first run of PLS utilising the full cross-validated leave-one-out LOO method analysis (Q2). To speed up the cross-validation measurement, the sample-distance partial least square (SAMPLS) approach was utilised. PLS was re-run using the non-cross-validation approach for the second time, with the best number of components calculated. The models in this stage were retrieved using statistical indicators such as the squared correlation coefficient R2 and the standard error of estimate.
- Model Validation: Validation of 3D-QSAR models should include generating correct predictions for data sets that were not utilised in the model creation. They should also be statistically meaningful and stable, with their application boundaries specified. A total of 40 trining compounds and ten measuring compounds were used in this analysis. Using CoMFA models developed by the training set, the test data are then matched using the same methods before being used to determine their predictive anticancer behaviours. The equation is used to determine the coefficient of determination for the data set (R2test). Yipred test and YiObs test represent the observed and expected values of the test set substances, respectively, and indicate the mean activity value of the training set in the latter equation.
- Y-Randomization Test: Y-randomization method is use to validated to generate model. The IC50 are randomly shuffled many times which is the activities of the studied molecules, and a new QSAR model is developed after every iteration. Q2 and R2 values in the latest QSAR versions are predicted to be lower than in the original one. For eliminate the possibility of chance correlation this technique is performed. whenever Q2 and R2 by obtained higher values, which describe that because of structural chance correlation and redundancy, an acceptable 3D-QSAR cannot be produced for this data collection.

5. Molecular Docking

Molecular docking is a method to predicts the preferred orientation of one molecule to another molecule and when they bound to each other form a stable complex molecule. These knowledges of the preferred orientation may be used to predict the binding strength between two molecules or binding affinity between two molecules using scoring functions. Biologically molecules such as proteins, peptides, nucleic acids, carbohydrates and lipids play a central role in signal transduction. The relative orientation of the two binding molecules may affect the type of signal produced. Therefore, docking is use to predict both the strength of the bond and type of the signal produced by the molecules.

Molecular docking is a structure-based drug design approach that predicts the binding-conformation of small molecules including ligands to the appropriate target binding site, such as protein molecules. Characterization of molecule binding behaviour is critical in drug development to understand basic biochemical processes.

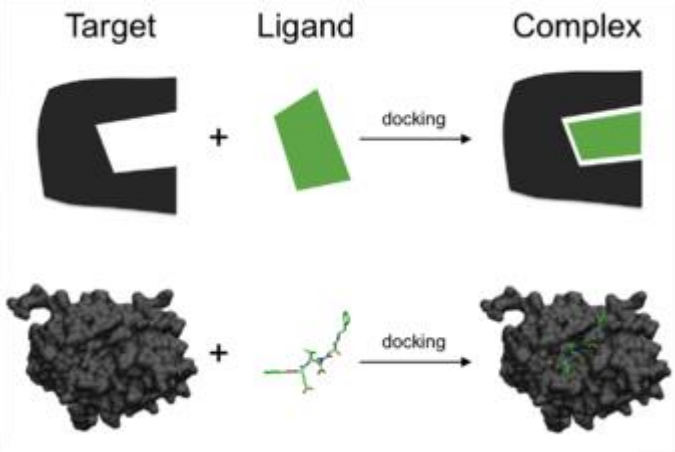


Figure 6: A stable complex is formed by docking a small molecule ligand (green) to a protein target (black).

Molecular docking is a lock-and-key problem in which one molecule tries to figure out which lock would open up by finding the correct orientation of the key. Where key hole is present on the surface of the lock that determine the turning direction of the key after it is inserted in lock. The protein can be thought of as a lock, and the ligand as a key in this case. The aim of molecular docking, which is an optimization problem, is to find the best-fit orientation of a ligand that binds to a specific protein. However, since both the ligand and the protein are versatile, it's better to think of it as a hand-in-glove situation rather than a lock-and-key situation. Throughout the docking procedure to achieve an overall best-fit the ligand and the protein adjust their conformation and this type of conformational adjustment of molecules gives best binding fit, this is known as induced-fit. Molecular docking is concerned with simulating the molecular recognition process computationally. To obtain an optimal conformation for both the ligand and the protein, as well as their relative orientation between protein molecules and ligand molecules.

5.1 Docking approaches

Most commonly two docking approaches are popular within the molecular docking community. One approach uses the matching technique which describes the protein molecule and the ligand molecule as complementary surfaces. The second approach based on simulation of the actual docking process in which the ligand molecules and protein molecules pairwise interaction and calculate binding energies. Both approaches have their significant advantages as well as some disadvantages.

5.1.1 Shape complementarity

This shape complementarity methods match the geometric shape of molecules. A series of geometrical features distinguishes protein and ligand molecules, allowing them to dock. Complementary surface descriptors on the molecular surface can be one of these geometrical features. The molecular surface of the receptor is defined as solvent-accessible surface area in this docking process and the ligand's molecular surface is described as its matching surface description. The surface matched between the two surfaces of molecules. The shape matching description help finding the complementary pose of docking the target the ligand molecules. Another way is to describe the hydrophobic features of the protein molecules using the turns in main-chain atoms. Also, there is another approach is to use the Fourier shape descriptor technique, where the shape is based on complementarity approaches, they are typically fast and robust, they cannot build the model of movements of the molecules and while recent advances in docking methods enable these to investigate ligand flexibility, it is difficult to accurately predict dynamic changes in the conformations of ligand molecules or protein molecules. Shape complementarity methods are useful for quickly scanning through thousands of ligands in seconds and this docking method can actually figure out whether they can bind at the protein's

active site or not, and this method can be used for protein-protein interactions. They are also much more pharmacophore-based approaches, since this docking method use geometric descriptions of the ligands molecules to find optimal binding.

5.1.1 Simulation

Simulation of docking is much more complicated process. In this method, the protein molecules and the ligand molecules are basically separated by some distance and various conformation the ligand molecules find its right position into the protein's active site after a certain number of conformations in its conformational space. These moves incorporate with rigid body transformations occurs such as translations and rotations, as well as internal changes in the ligand's structure including torsion angle rotations. All these moves are inducing a system's gross energetic expense of the ligand in the conformation space. Hence, after each step, the total energy of the system is determined. The flexibility of the ligand molecule is easily integrated in this docking simulation process, while complementarity techniques require ingenious methods to integrate ligand flexibility. It also models more accurately than the previous approach, since shape complementary techniques are more of an approximation in fact.

Simulation method needs powerful computer, it is computationally expensive, which is main drawback of this simulation method. Also, it needs to explore a huge energy landscape. It is a grid-based technique. To get proper bond between two molecules needs optimization methods. Now days speed of computer increased, it helps docking simulation more practical such that the free energy of the overall system is minimized.

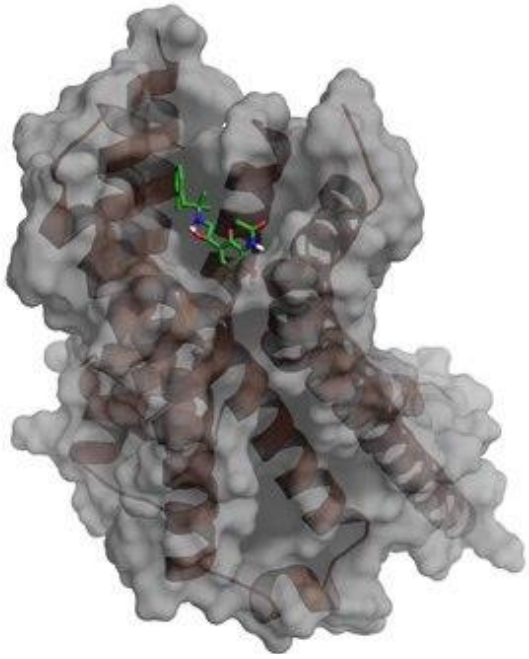


Figure 7: Docking of two molecules

6. Molecular Modelling: Dataset

We have 50 Molecules

40 Training Set for Developing QSAR Model

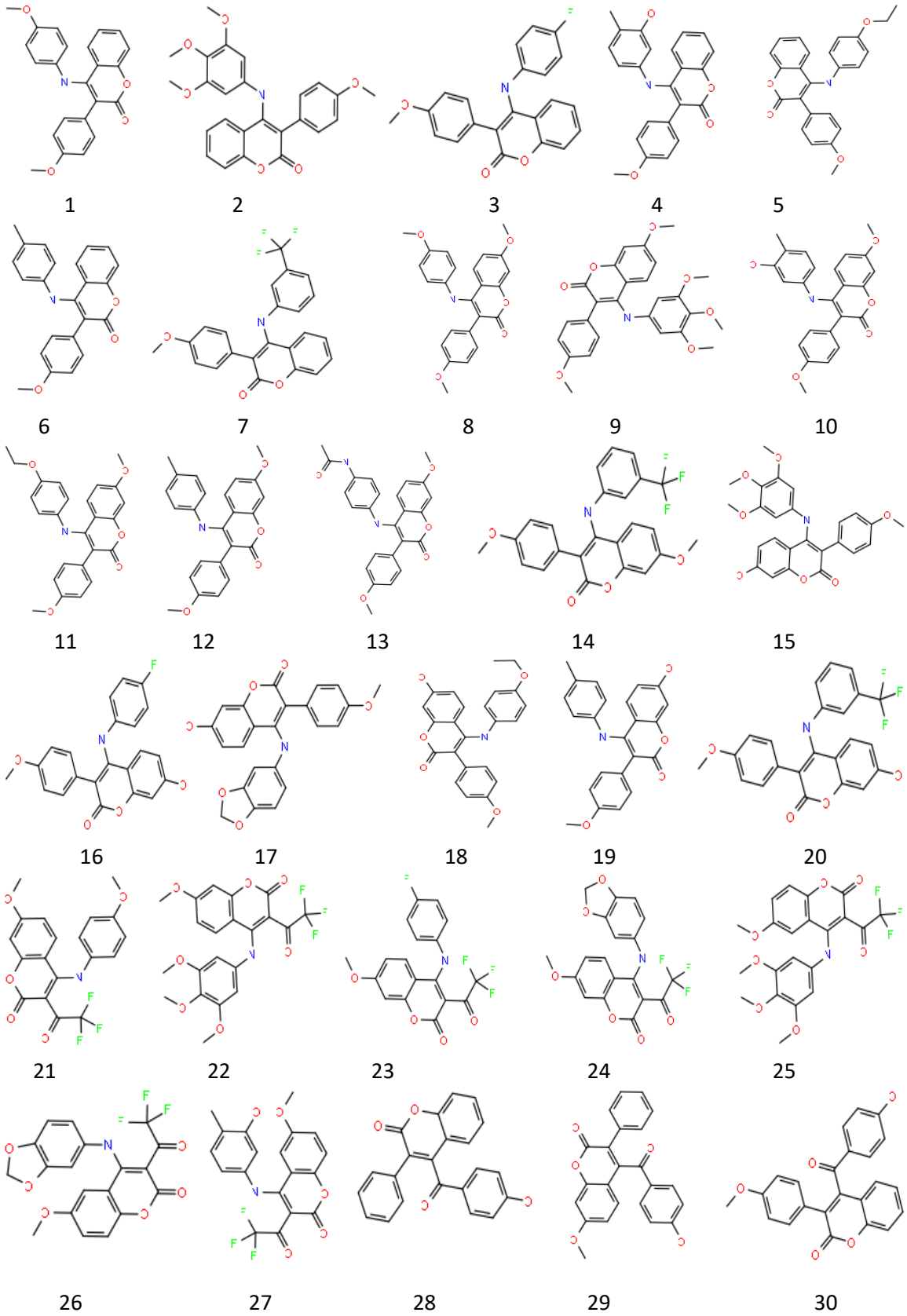
10 Test Set for Model Validation

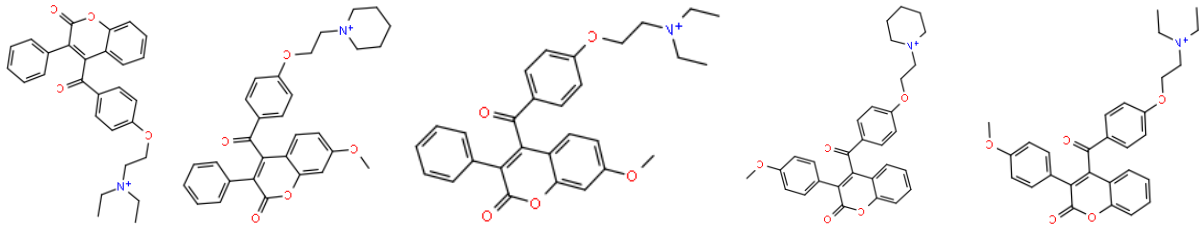
6.1 Methodology

1. Draw Molecules
2. Energy Minimization: All Molecules are Minimizing using Steepest decent method with a convergence criterion 0.001kcal/mol and atomic charge were calculated MMFF94 force field
3. Alignment: All the molecules are aligned to core part and high active molecule as a template using
4. Build 3dQSAR Model: CoMFA studies using Partial Least Square Method

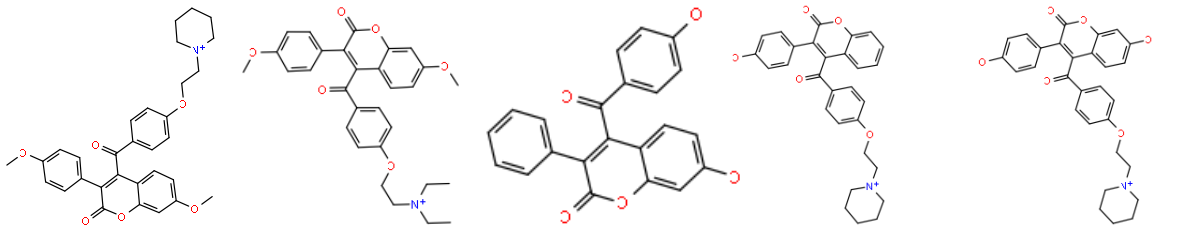
6.2 Structure of all Coumarin

Drawing all coumarin derivatives

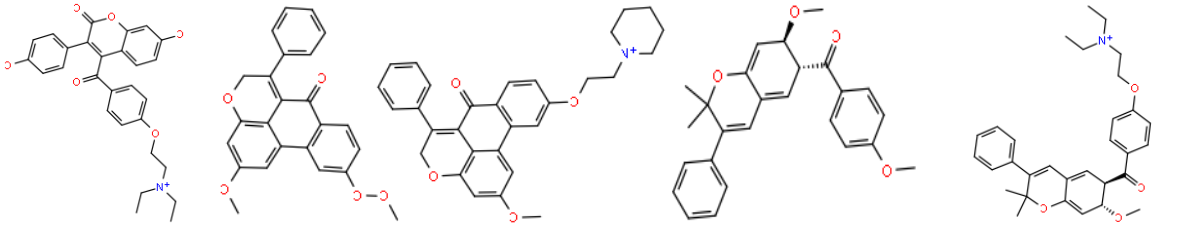




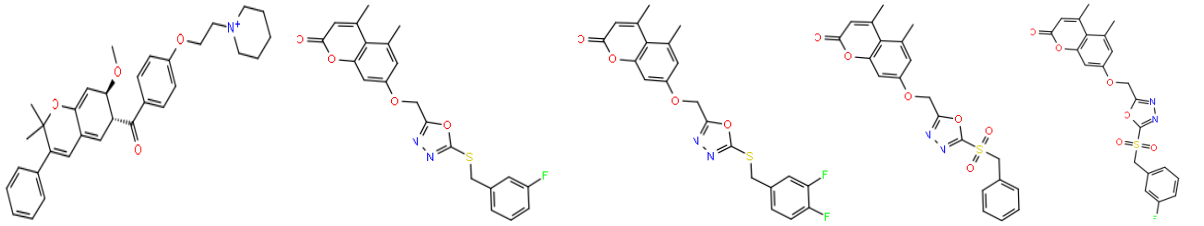
31 32 33 34 35



36 37 38 39 40



41 42 43 44 45



46 47 48 49 50

6.3 Training and Test Dataset

Approximate IC50 values of the synthesized compounds against MCF-7

Label	IC50(μM)	pIC50	Set
1	34.58	4.46	Training Set
2	82.34	4.08	Test Set
3	38.45	4.42	Training Set
4	12.08	4.92	Test Set
5	30.66	4.51	Training Set
6	65.31	4.19	Training Set
7	21.62	4.67	Training Set
8	39.25	4.41	Test Set
9	89.45	4.05	Test Set

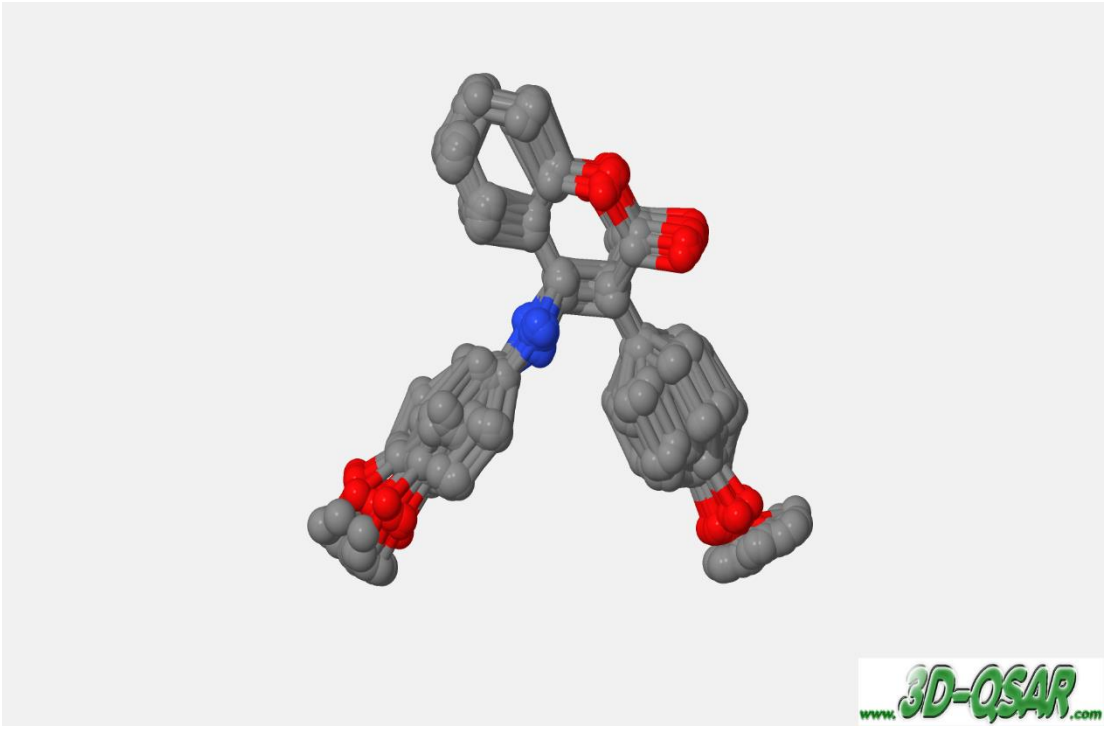
10	34.39	4.46	Training Set
11	24.31	4.61	Training Set
12	66.88	4.17	Training Set
13	85.75	4.07	Training Set
14	36.22	4.44	Training Set
15	55.98	4.25	Training Set
16	71.46	4.15	Test Set
17	89.55	4.05	Training Set
18	32.46	4.49	Training Set
19	85.47	4.07	Training Set
20	49.09	4.31	Training Set
21	55.73	4.25	Test Set
22	19.21	4.72	Training Set
23	9.74	5.01	Training Set
24	37.4	4.43	Training Set
25	15.04	4.82	Test Set
26	16.57	4.78	Training Set
27	20.14	4.70	Training Set
28	24.3	4.61	Training Set
29	17.7	4.75	Training Set
30	27.9	4.55	Training Set
31	33.7	4.47	Training Set
32	29.3	4.53	Training Set
33	32.4	4.49	Training Set
34	19.7	4.71	Training Set
35	16.3	4.79	Training Set
36	5.7	5.24	Training Set
37	12.8	4.89	Training Set
38	22.1	4.66	Training Set
39	31.7	4.50	Training Set
40	25.9	4.59	Training Set
41	27.1	4.57	Training Set
42	20.0	4.70	Training Set
43	34.5	4.46	Training Set
44	19.8	4.70	Training Set
45	6.8	5.17	Training Set
46	8.1	5.09	Training Set
47	26.04	4.58	Test Set
48	33.29	4.48	Test Set
49	38.58	4.41	Test Set
50	70.37	4.5	Test Set

6.4 Alignment

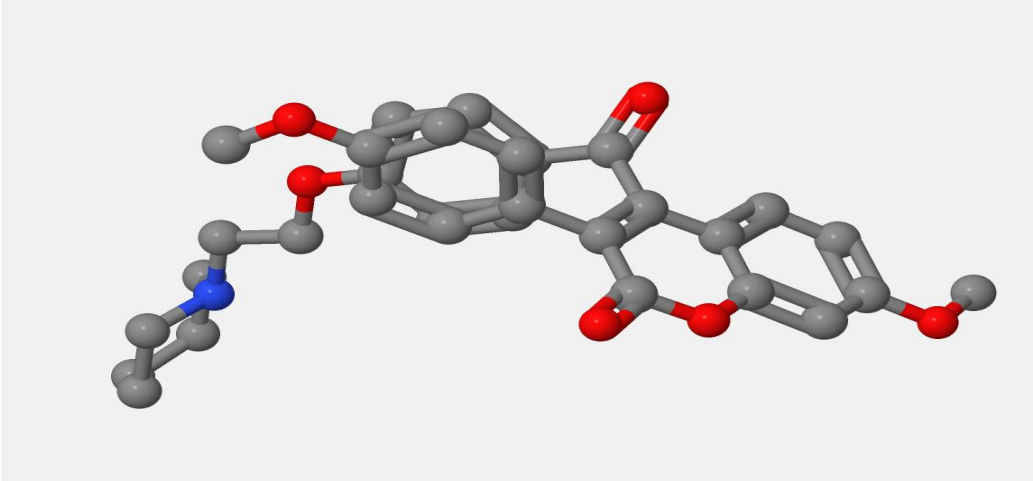
Label	pAct	MW	HA	HD	LogP	TPSA	RBs	MR	Length	Max Length
1	4.461	373.131	5	1	5.221	60.7	7	110.749	13.597	13.622
2	4.084	433.153	7	1	5.238	79.16	11	123.853	13.559	13.632
3	4.415	361.111	4	1	5.351	51.47	5	104.155	13.611	13.624
4	4.918	373.131	5	2	5.226	71.7	7	110.599	13.61	13.632
5	4.513	387.147	5	1	5.611	60.7	8	115.366	13.546	13.625
6	4.185	357.136	4	1	5.521	51.47	6	108.934	13.614	13.637
7	4.665	411.108	4	1	6.231	51.47	5	109.199	13.615	13.679
8	4.406	403.142	6	1	5.229	69.93	9	117.301	15.582	15.825
9	4.048	463.163	8	1	5.247	88.39	13	130.405	15.745	15.874
10	4.464	403.142	6	2	5.235	80.93	9	117.151	15.624	15.819
11	4.614	417.158	6	1	5.62	69.93	10	121.918	15.583	15.827
12	4.175	387.147	5	1	5.529	60.7	8	115.486	15.601	15.818
13	4.067	430.153	6	2	5.179	89.8	9	125.052	15.6	15.812
14	4.441	441.119	5	1	6.24	60.7	7	115.751	15.587	15.849
15	4.252	449.147	8	2	4.944	99.39	12	125.518	14.351	14.504
16	4.146	377.106	5	2	5.057	71.7	6	105.82	14.333	14.348
17	4.048	403.106	7	2	4.647	90.16	6	111.984	14.096	14.361
18	4.489	403.142	6	2	5.317	80.93	9	117.031	14.097	14.388
19	4.068	373.131	5	2	5.226	71.7	7	110.599	14.366	14.368
20	4.309	427.103	5	2	5.937	71.7	6	110.864	14.127	14.33
21	4.254	393.082	6	1	4.299	77.77	7	95.698	13.471	13.812
22	4.716	453.104	8	1	4.316	96.23	11	108.802	13.244	13.848
23	5.011	381.062	5	1	4.429	68.54	5	89.104	12.063	12.12
24	4.427	407.062	7	1	4.019	87	5	95.269	11.871	13.788
25	4.823	453.104	8	1	4.316	96.23	11	108.802	11.677	13.167
26	4.781	407.062	7	1	4.019	87	5	95.269	11.46	11.884
27	4.696	393.082	6	2	4.304	88.77	7	95.548	11.618	11.73
28	4.614	342.089	4	1	4.397	67.51	4	99.459	11.479	11.529
29	4.752	372.1	5	1	4.405	76.74	6	106.011	13.668	13.716
30	4.554	372.1	5	1	4.405	76.74	6	106.011	13.624	13.679
31	4.472	442.201	4	1	3.995	60.95	11	130.201	13.889	17.272
32	4.533	484.212	5	1	4.147	70.18	9	139.256	18.119	18.648
33	4.489	472.212	5	1	4.003	70.18	13	136.753	16.189	18.255
34	4.706	484.212	5	1	4.147	70.18	9	139.256	17.849	18.797
35	4.788	472.212	5	1	4.003	70.18	13	136.753	17.387	18.985
36	5.244	514.222	6	1	4.156	79.41	11	145.808	16.534	18.521
37	4.893	502.222	6	1	4.012	79.41	15	143.305	15.673	17.736
38	4.656	358.084	5	2	4.102	87.74	5	101.124	12.01	12.26
39	4.499	470.196	5	2	3.844	81.18	8	134.369	14.935	17.391
40	4.587	486.191	6	3	3.55	101.41	9	136.034	15.907	18.116
41	4.567	474.191	6	3	3.406	101.41	13	133.531	15.218	17.57
42	4.699	388.131	5	0	4.822	53.99	8	110.155	13.351	16.741
43	4.462	470.233	4	1	4.329	49.2	9	137.427	16.716	20.639

44	4.703	402.183	4	0	5.225	44.76	9	117.361	18.128	18.238
45	5.167	488.28	4	1	4.52	49.2	15	143.215	21.39	22.457
46	5.092	500.28	4	1	4.664	49.2	11	145.718	21.713	23.181
47	4.584	412.089	7	0	4.803	78.36	8	106.549	17.529	19.12
48	4.478	430.08	7	0	4.942	78.36	8	106.507	19.594	19.923
49	4.414	426.089	8	0	3.346	112.5	8	107.92	14.844	19.735
50	4.153	444.079	8	0	3.485	112.5	8	107.878	18.672	19.763

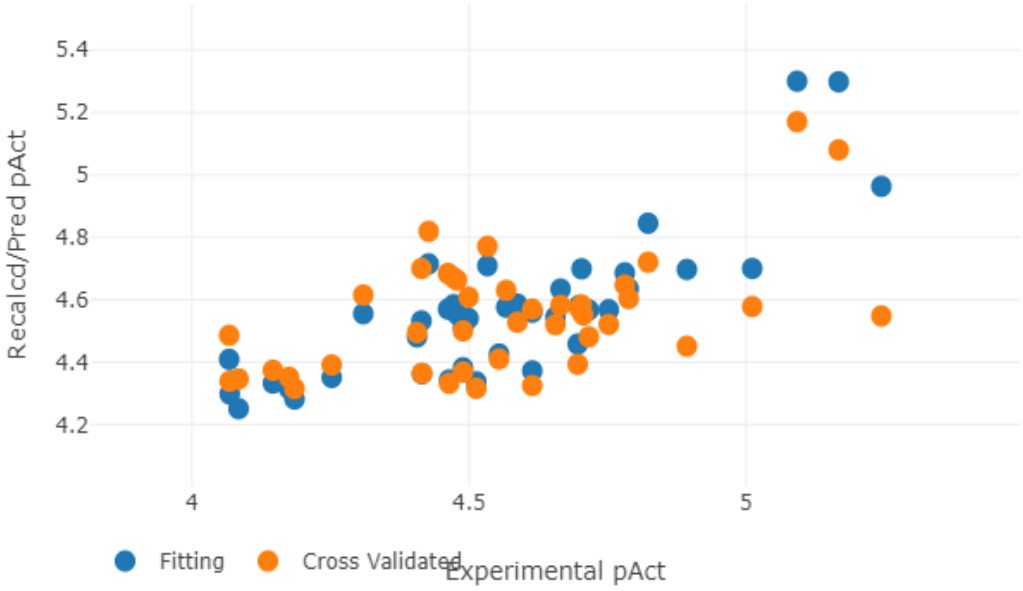
Property	Label
Least Active	13
Most Active	36
Heaviest	36
Longest	46
Most Flexible	37
Most Rigid	28
Least Polar	44
Most Polar	49
Highest MR	36
Lowest MR	23
Highest HA	25
Lowest HA	46
Highest HD	40
Lowest HD	48
Highest LogP	7
Lowest LogP	49



Most active structure



Graph of expected, predicted data and cross validation data



6.5 Result

Label	Expected pIC50 value	Predicted pIC50 value
2	4.084	4.483
8	4.406	4.494
9	4.048	4.481
16	4.146	4.491
21	4.254	4.505
25	4.823	4.494
47	4.584	4.510
48	4.478	4.504
49	4.414	4.503
50	4.153	4.508

6.6 Docking of the of estrogen receptor with all the ligand through Autodock Vina:

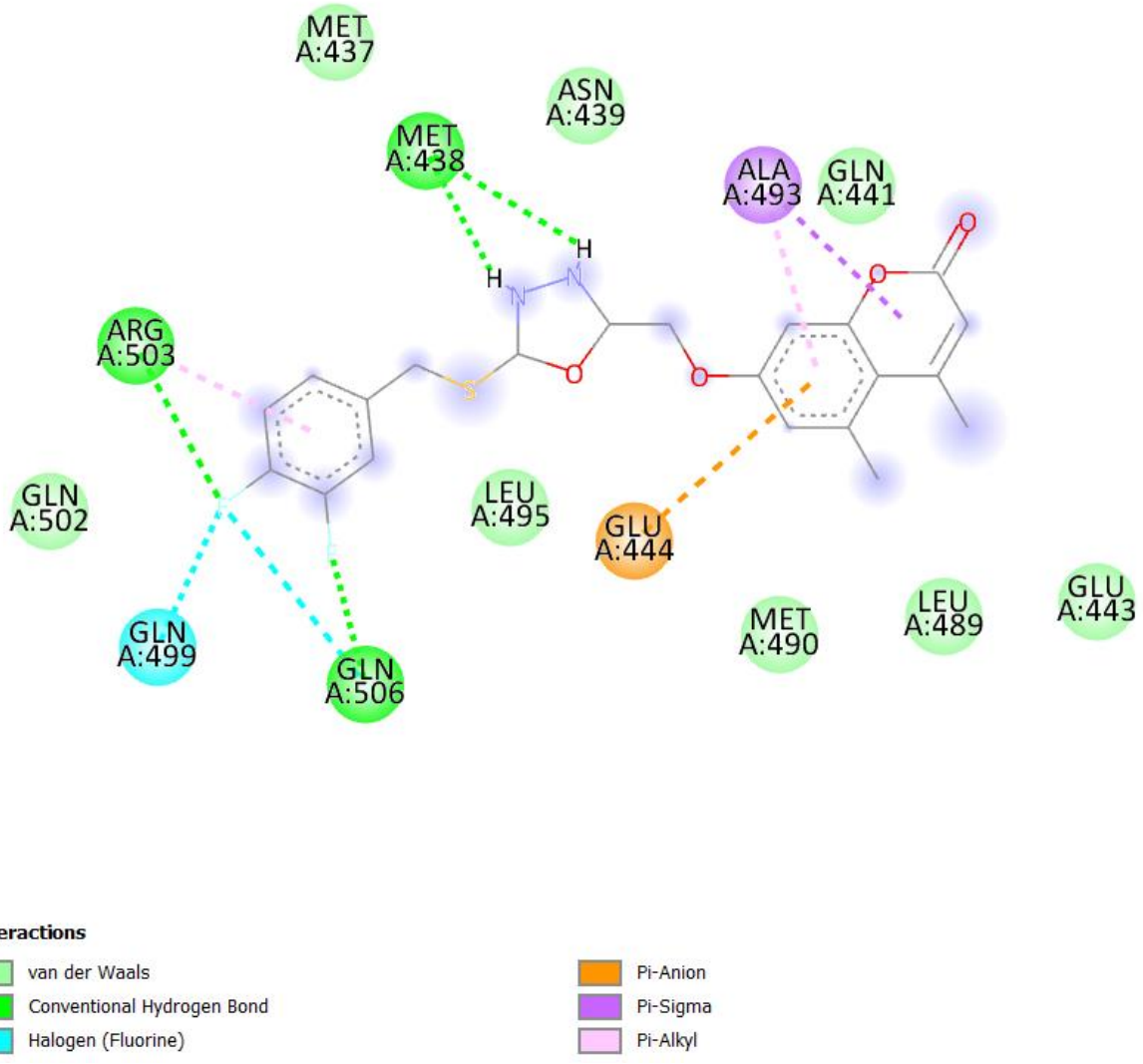
Label	Binding Affinity
1	-6.5
2	-6.3
3	-6.9
4	-6.5
5	-6.5
6	-6.4
7	-7
8	-6.5
9	-6.4
10	-6.6
11	-6.7
12	-6.9
13	-6.8
14	-7.1
15	-6.4
16	-6.6
17	-6.6
18	-6.6
19	-6.8
20	-7
21	-6.3
22	-6.2
23	-6.1
24	-6.6

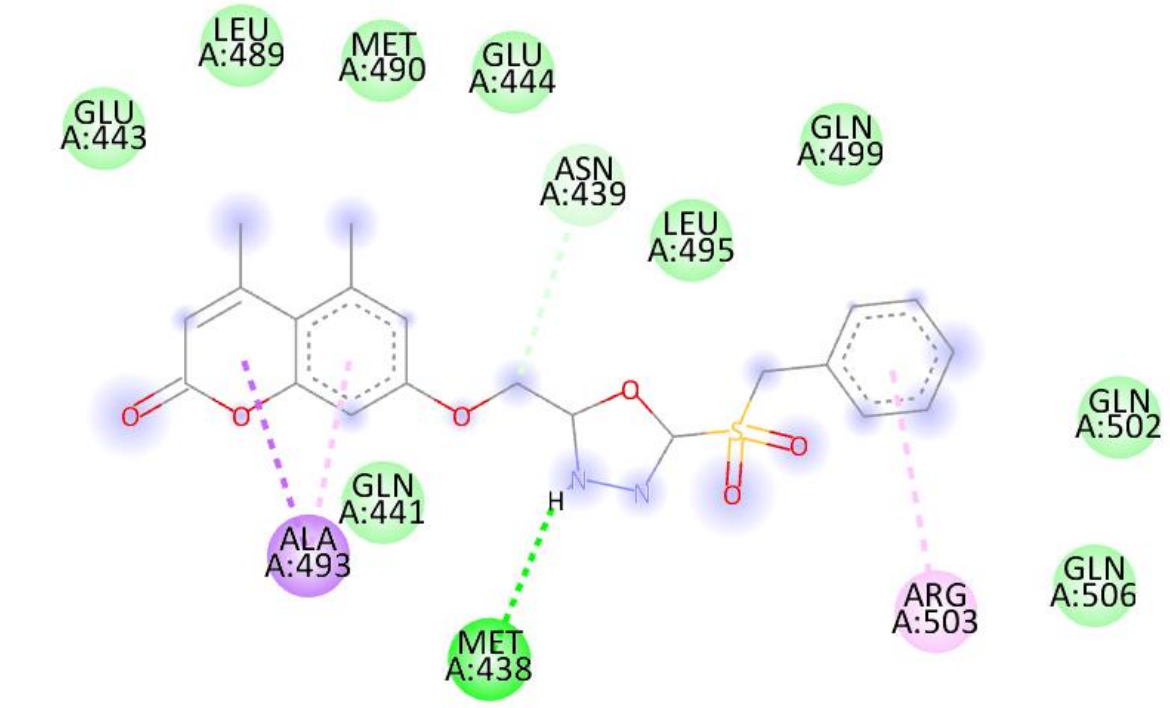
25	-5.6
26	-6.7
27	-6.3
28	-6.5
29	-6.6
30	-6.4
31	-6.5
32	-7.4
33	-6.8
34	-7.2
35	-6.5
36	-7
37	-6.3
38	-6.5
39	-7.2
40	-7.2
41	-6.8
42	-6.8
43	-6.9
44	-7.2
45	-6.2
46	-7.6
47	-7.7
48	-7.8
49	-7.8
50	-7.4

The docking results suggested that on the basis of binding energy Label 48(48th Coumarin structure) and Label 49(49th Coumarin structure) with Estrogen receptor (PDB ID- 1ERE)

6.7 Amino acid binding site visualization through Discovery Studio

2D protein-ligand interaction diagrams was obtained for Estrogen receptor (PDB ID- 1ERE) with the Label 48(48th Coumarin structure) and Label 49(49th Coumarin structure) which they were docked. Following is the 2D interaction diagram of molecule instruction.


MET
A:437
ASN
A:439
MET
A:438
ALA
A:493
GLN
A:441
H
H
N
N
O
O
O
ARG
A:503
S
GLN
A:502
LEU
A:495
GLU
A:444
GLU
A:443
LEU
A:489
MET
A:490
GLN
A:499
GLN
A:506
Interactions
van der Waals
Conventional Hydrogen Bond
Halogen (Fluorine)
Pi-Anion
Pi-Sigma
Pi-Alkyl



Interactions

- van der Waals
- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Sigma
- Pi-Alkyl

7. CONCLUSION

On the basis of the in-silico analysis performed uptil now, it can be concluded that out of all coumarin derivatives used, 48(48th Coumarin structure) and Label 49(49th Coumarin structure) could turn out to be the best inhibitor for controlling the overexpression of Estrogen receptor (PDB ID- 1ERE).

References

- **Acquired Tamoxifen Resistance in MCF-7 Breast Cancer Cells Requires Hyperactivation of eIF4F-Mediated Translation.** Fagan DH, Fettig LM, Avdulov S, Beckwith H, Peterson MS, Ho YY, Wang F, Polunovsky VA, Yee D. *Horm Cancer*. 2017 Jun 2.
- **The impact of anticancer activity upon Beta vulgaris extract mediated biosynthesized silver nanoparticles (ag-NPs) against human breast (MCF-7), lung (A549) and pharynx (Hep-2) cancer cell lines.** Venugopal K, Ahmad H, Manikandan E, Thanigai Arul K, Kavitha K, Moodley MK, Rajagopal K, Balabhaskar R, Bhaskar M. *J Photochem Photobiol B*. 2017 May 24;173:99-107.
- **Design and synthesis of ERa/ERb selective coumarin and chromene derivatives as potential anti-breast cancer and anti-osteoporotic agents** M. Kamil Hussain,a M. Imran Ansari,a N. Yadav,a Puneet K. Gupta,a A. K. Gupta,a R. Saxena,b I. Fatima,b M. Manohar,b P. Kushwaha,b V. Khedgikar,b J. Gautam,b Ruchir Kant,c P. R. Maulik,c R. Trivedi,b A. Dwivedi,b K. Ravi Kumar,d A. K. Saxenaa and K. Hajela
- **Synthesis, computational studies and antiproliferative activities of coumarin-tagged 1,3,4-oxadiazole conjugates against MDA-MB-231 and MCF-7 human breast cancer cells** Sanjeev Dhawana , Nagaraju Kerrua , Paul Awoladea , Ashona Singh-Pillaya , Sourav Taru Sahab , Mandeep Kaurb , Sreekantha B. Jonnalagaddaa , Parvesh Singha,
- **Design, synthesis and biological evaluation of novel 3-substituted 4-anilinocoumarin derivatives as antitumor agents** Guoshun Luo, Moses Muyaba, Weiting Lyu, Zhichao Tang, Ruheng Zhao, Qian Xu, Qidong You, Hua Xiang
- **Study on Anti-Tumor Activity of Novel 3-Substituted 4 Anilino-Coumarin Derivatives Using Quantitative Structure-Activity Relationship (QSAR)** DARATU E. K.Putri1, HARNO Dwi Pranowo, and WINARTO Haryadi1, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Jl. Sekip Utara, Yogyakarta 55281, Indonesia