PROJECT REPORT

**Qsar Study and Molecular Docking of**

**Coumarin Derivatives Of Prostate Cancer And Colon Cancer**

****

**Group Members**

Utkarsh Singh - MBI2019011

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, ALLAHABAD

**UNDER THE SUPERVISION OF**

**NIDHI MISHRA**

# **TABLE OF CONTENTS**

1. Literature Review……………………………………………..03

2. Introduction…………………………………………………...04

3. Materials and methods………………………………………..07

4. Conclusion……………………………………………………21

5. References…………………………………………………… 23

# 

# **ABSTRACT**

Quantitative structure-activity relationships (QSAR) have been applied for decades in the development of relationships between physicochemical properties of chemical substances and their biological activities to obtain a reliable statistical model for prediction of the activities of new chemical entities. The fundamental principle underlying the formalism is that the difference in structural properties is responsible for the variations in biological activities of the compounds. In the classical QSAR studies. Protein–ligand docking aims to predict and rank the structure(s) arising from the association between a given ligand and a target protein of known 3D structure. Despite the breathtaking advances in the field over the last decades and the widespread application of docking methods, several downsides still exist. In particular, protein flexibility—a critical aspect for a thorough understanding of the principles that guide ligand binding in proteins—is a major hurdle in current protein–ligand docking efforts that needs to be more efficiently accounted for. In this review the key concepts of protein–ligand docking methods are outlined, with major emphasis being given to the general strengths and weaknesses that presently characterize this methodology. Despite the size of the field, the principal types of search algorithms and scoring functions are reviewed and the most popular docking tools are briefly depicted. Recent advances that aim to address some of the traditional limitations associated with molecular docking are also described

**1. LITERATURE REVIEW:**

Cancer is one of the most serious health problems in the world. Over the past three decades, the number of people with different types of cancer has increased, and the problem is expected to worsen in the next few years if new methods of treatment are not discovered

In the modern drug design and discovery, quantitative structure-activity relationship (QSAR) methodology generally plays a crucial role . QSAR studies are performed by correlation structural descriptors from a set of chemicals to their known biological activities and predict the activity values of non examined compounds lying in the applicability domain of the model . 3D-QSAR studies are successfully applied to guide the development and design of new novel molecules with an improved activity Steric and electrostatic fields have been used as an independent variable against the biological activity of compounds . One of the types of 3D-QSAR technique is the CoMFA method.The correlation between the 3D structures of the studied molecules and their activities can be better interpreted relying on the calculation of hydrophobic, H-bond donor, and H-bond acceptor

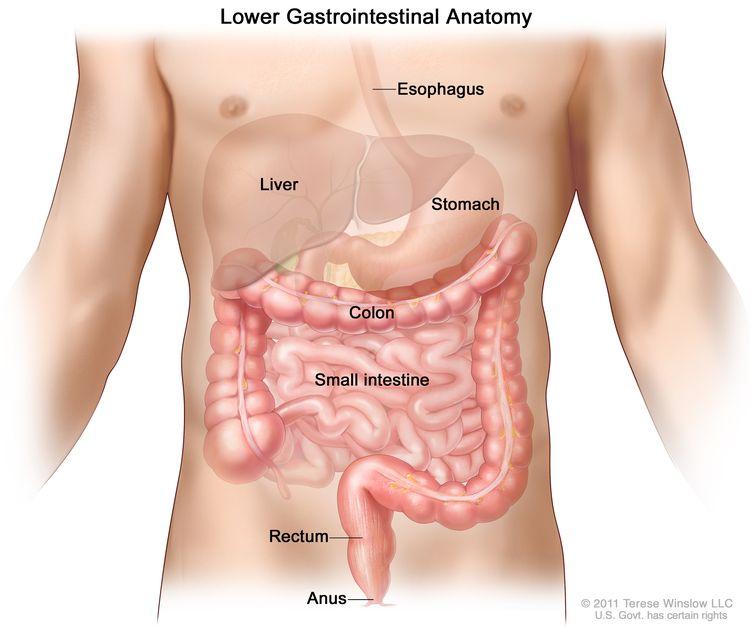
Pharmaceutical research has successfully incorporated a wealth of molecular modeling methods, within a variety of drug discovery programs, to study complex biological and chemical systems. The integration of computational and experimental strategies has been of great value in the identification and development of novel promising compounds. Broadly used in modern drug design, molecular docking methods explore the ligand conformations adopted within the binding sites of macromolecular targets. This approach also estimates the ligand-receptor binding free energy by evaluating critical phenomena involved in the intermolecular recognition process.Molecular docking is a key tool in structural molecular biology and computer-assisted drug design. The goal of ligand—protein docking is to predict the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure. Successful docking methods search high-dimensional spaces effectively and use a scoring function that correctly ranks candidate dockings. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization.

# **2. INTRODUCTION**

Millions of people continue to fall sick with cancer disease each year all over the word especially in developed countries. So, our purpose and challenges in this article was to find a new chemotherapeutic agent that can be used as a cure for cancer and less toxic at the same time.

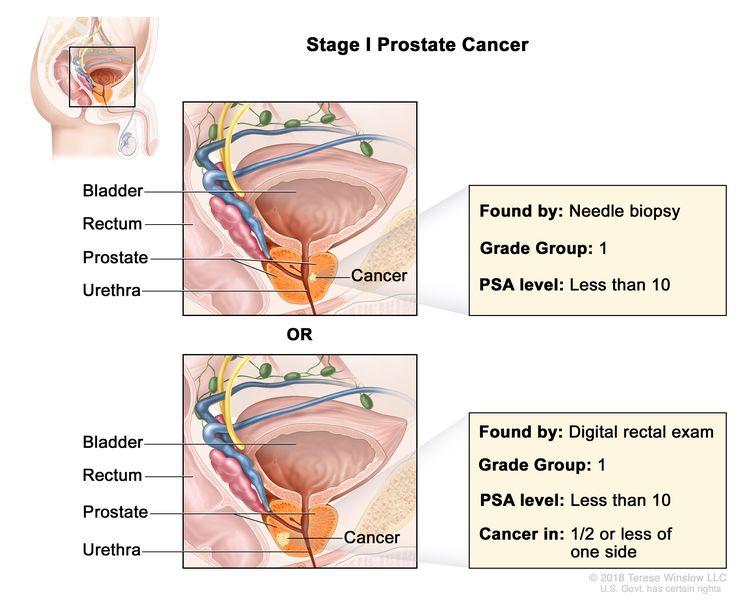
**2.1 Colon Cancer :-**

* Colon cancer is a type of cancer that begins in the large intestine (colon). The colon is the final part of the digestive tract.
* Colon cancer typically affects older adults, though it can happen at any age.
* The American Cancer Society (ACS) predict that 101,420 people in the U.S. will receive a new diagnosis of colon cancer.
* Colon cancer describes co-occurring colon cancer and rectal cancer is also common.
* Rectal cancer originates in the rectum, which is the final several inches of the large intestine, closest to the anus.



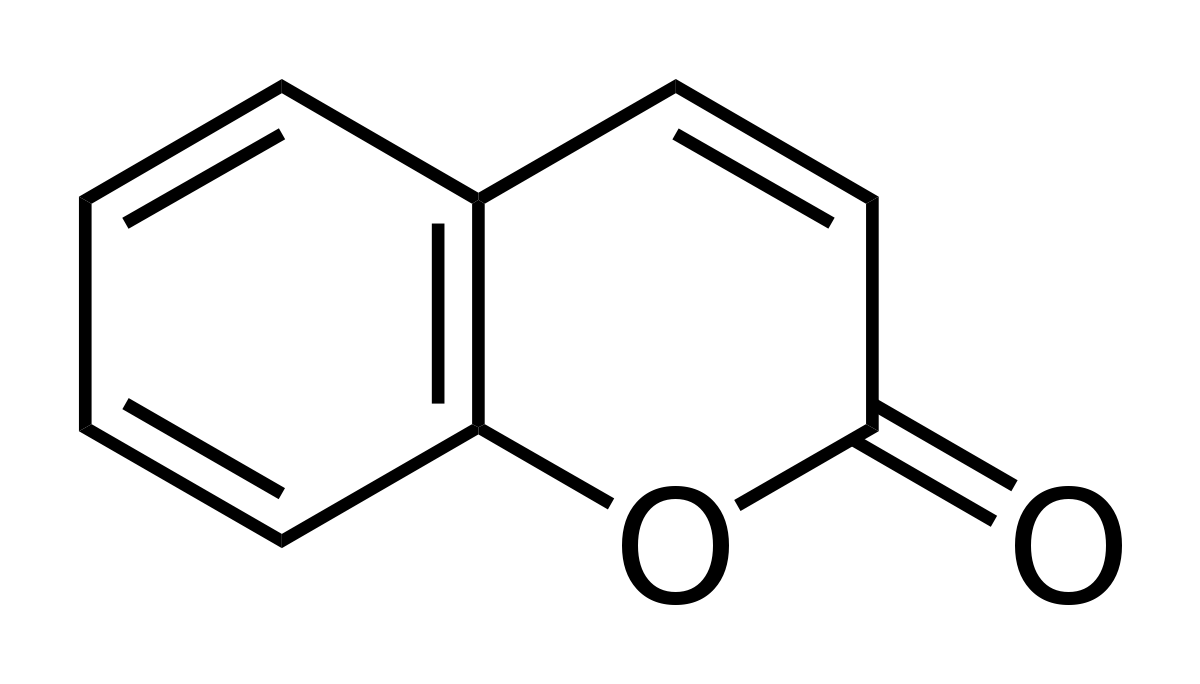
**2.2 Prostate Cancer :-**

* A cancer in a man's prostate, a small walnut-sized gland that produces seminal fluid.
* prostate cancer begins when cells in the prostate gland start to grow out of control.
* The prostate is a gland found only in males. it makes some of the fluid that is part of semen.
* The American Cancer Society (ACS) predict that there will be around 174,650 new diagnoses of prostate cancer and around 31,620 deaths from this type of cancer.
* Around 1 in 9 males will receive a diagnosis of prostate cancer at some point in their life. However, only 1 in 41 of these will die as a result of it.
* This is because treatment is effective, especially in the early stages. Routine screening enables doctors to detect many cases of prostate cancer before they spread.



Here there are four kinds of coumarin compounds which show activity against colon cancer cell (HCT-116) line and prostate cancer cell line (DU-145) we take datasets of 50 50 coumarin compounds for both cancer cell lines to align.

**2.3 Basic Structure of Coumarin Derivative**

  **C9H6O2**

**Coumarin**

* Coumarin and its derivatives are one of the most critical class of benzopyrones that are widely utilized as pharmaceutical compounds.
* Due to their activities as anticoagulants , anti-asthmatics, anti-allergic , antidepressants , antitumor and so on.

**2.4 IC50 :-** The half maximal inhibitory concentration (IC50) is a measure of the effectiveness of a substance in inhibiting a specific biological or biochemical function. According to the FDA, IC50 represents the concentration of a drug that is required for 50% inhibition in vitro.

**3. MATERIALS AND METHODS:**

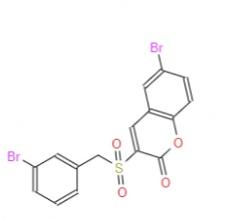
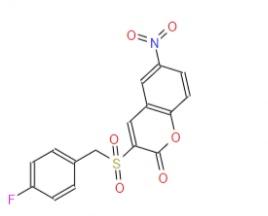
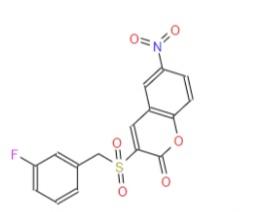
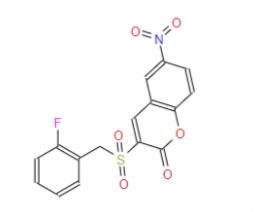
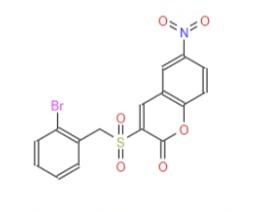
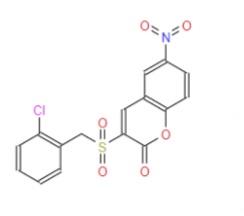
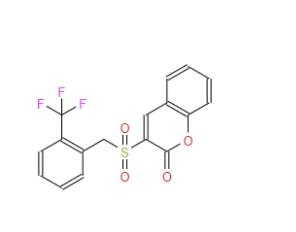
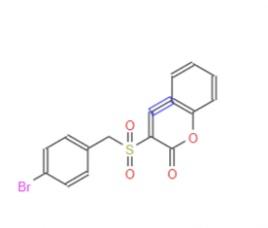
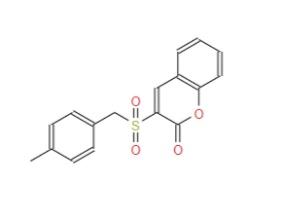
**3.1 DATASET:**

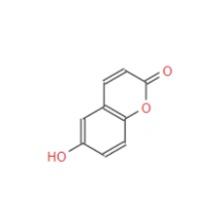
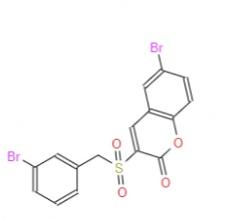
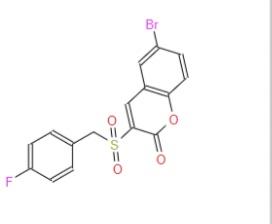
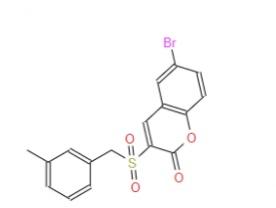
52 compounds that were previously assessed for their Hct-116 inhibition

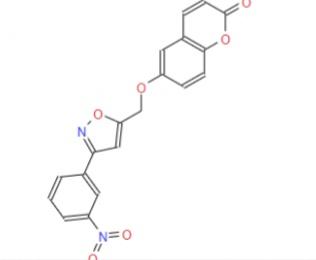
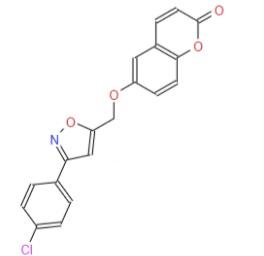
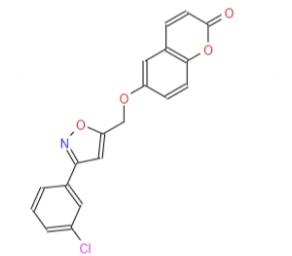
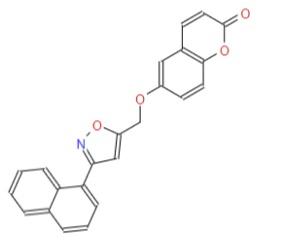
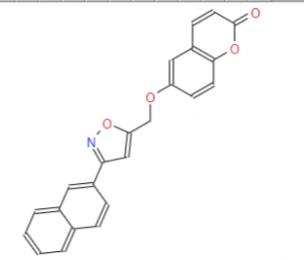
21 compounds that were previously assessed for their Du-145 inhibition

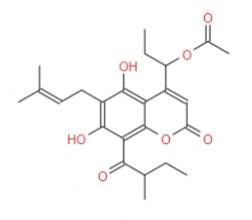
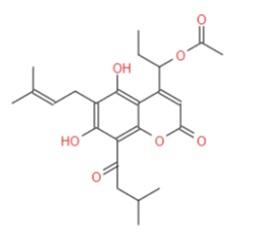
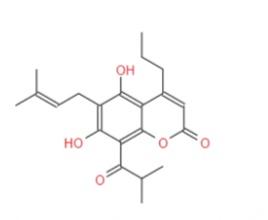
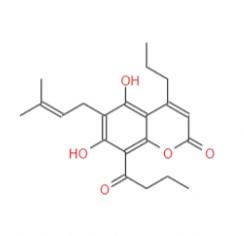
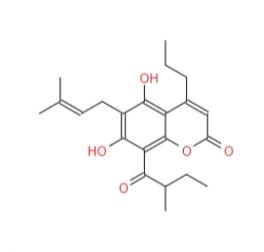
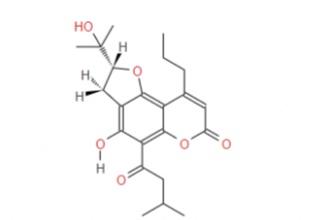
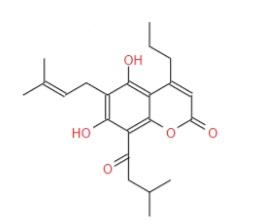
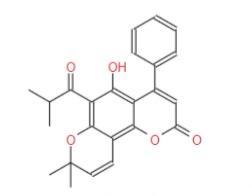
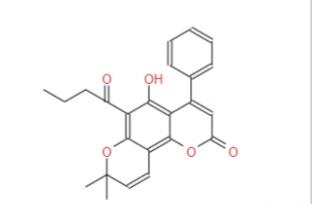
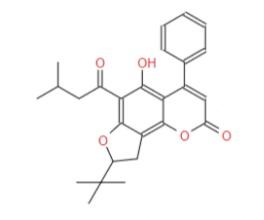
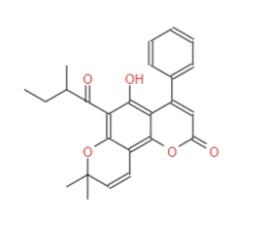
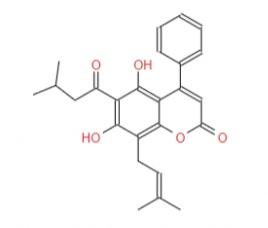
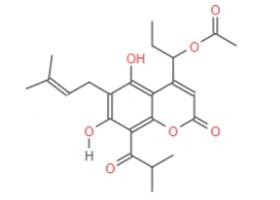
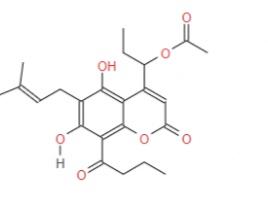
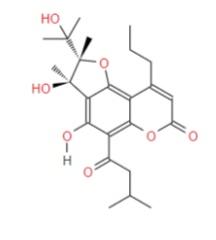
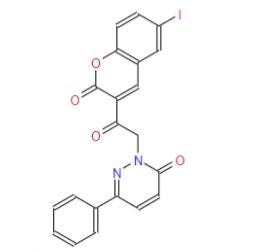
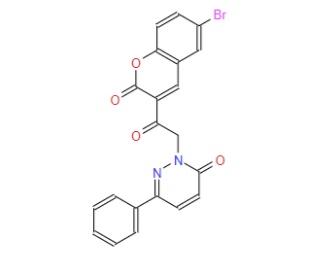
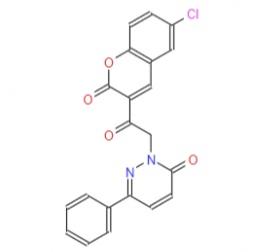
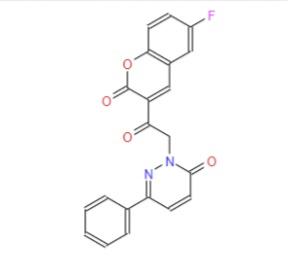
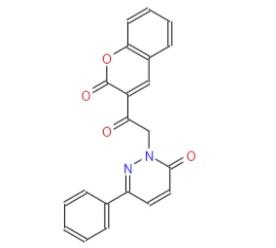
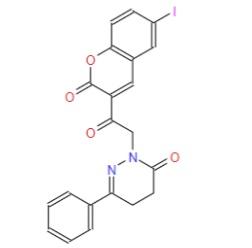
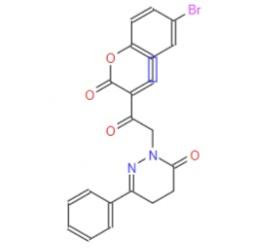
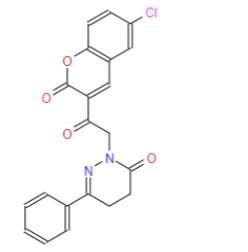
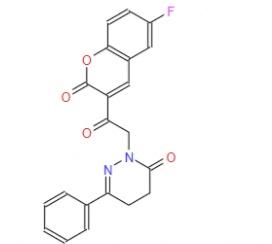
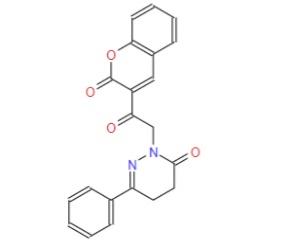
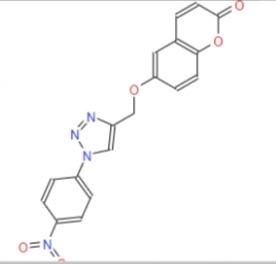
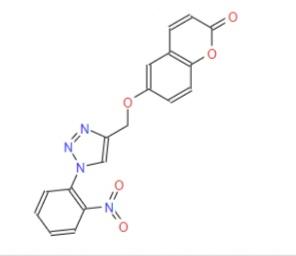
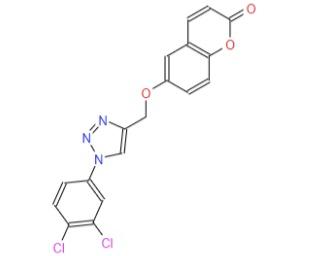
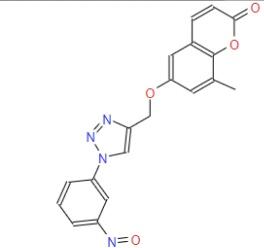
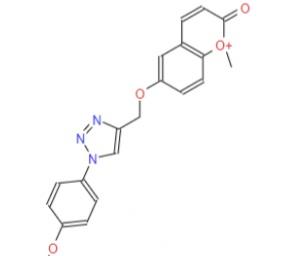
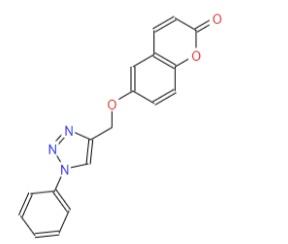
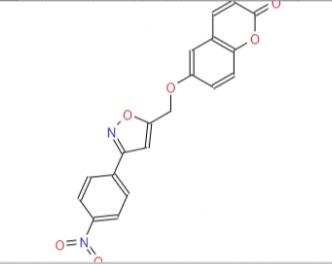
activities were used as the data set These compounds introducing a substituted comarine group at the R position of the parent scaffold resulted in significantly improved activities

**2.1.1Coumarin compounds for Hct-116 cell line:-**

****

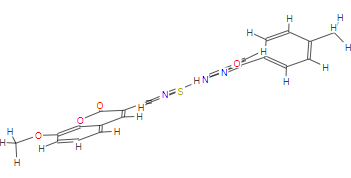
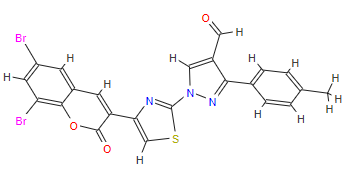
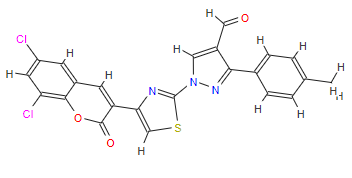
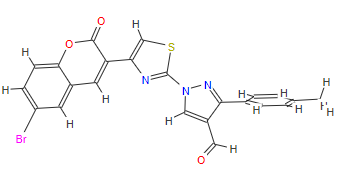
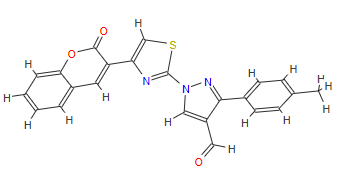
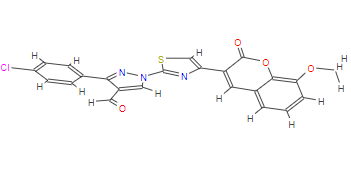
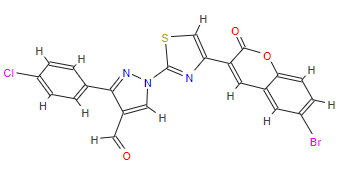
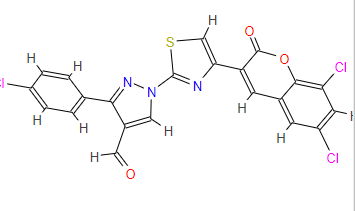
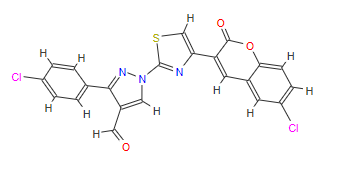
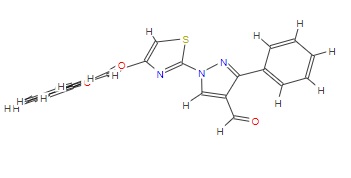
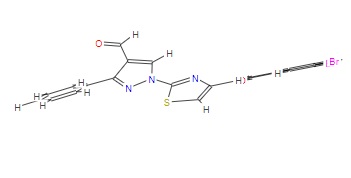
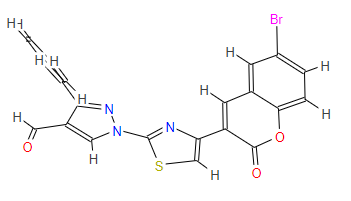
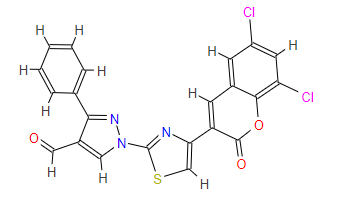
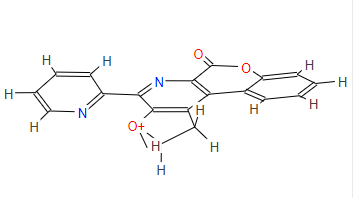
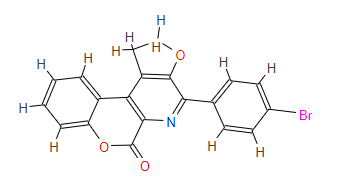
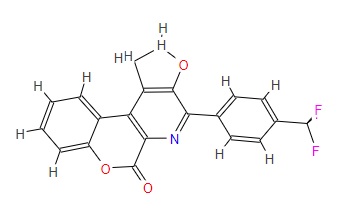
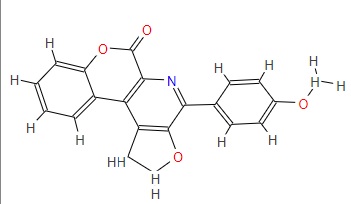
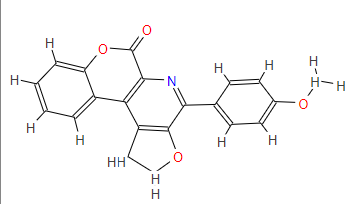
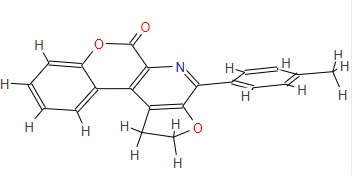
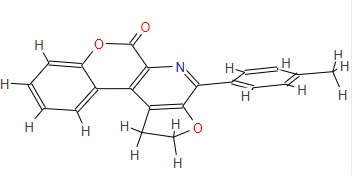
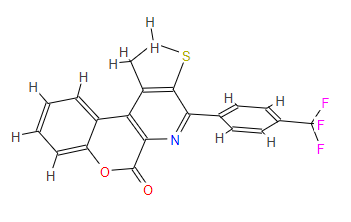
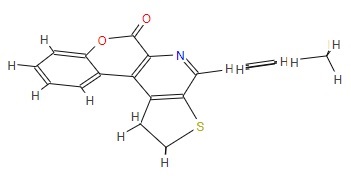
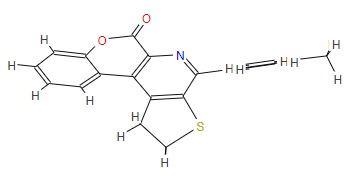
****

****

****

**Fig 1.1 ( 1-52 comarine compounds)**

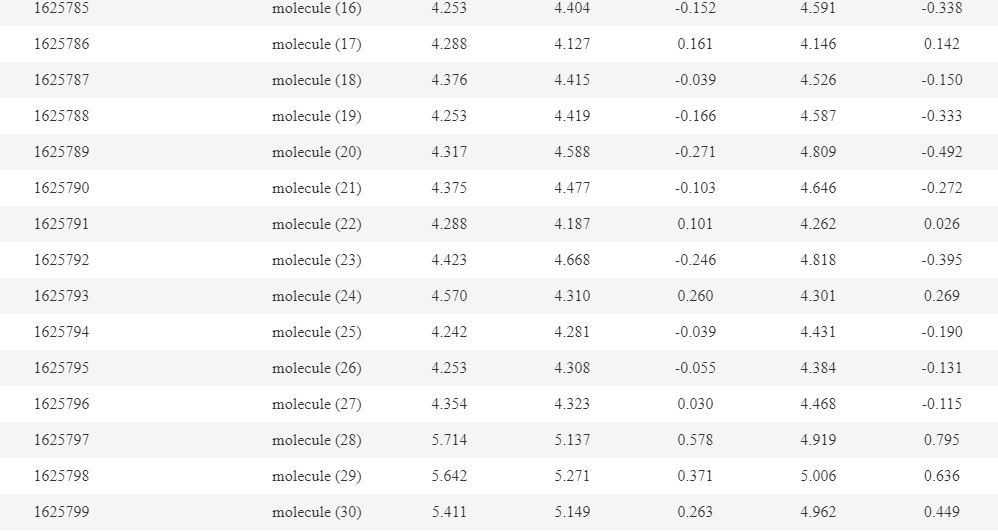
**2.1.2 Coumarin compounds for Du-145 :-**

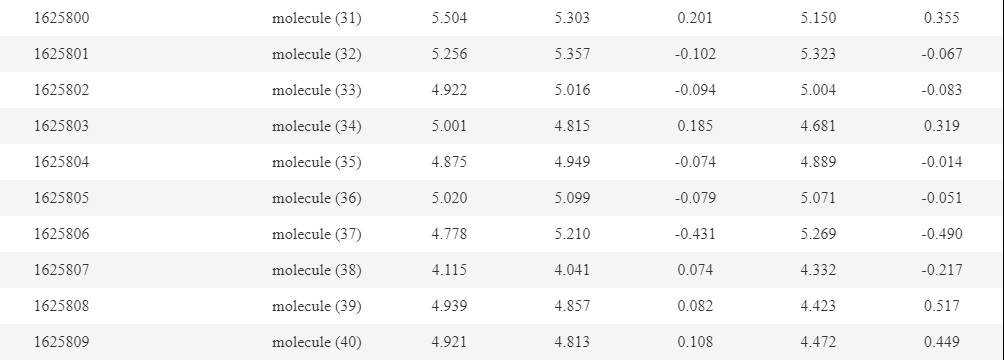
****

**Fig 1.2 ( 1-23comarine compounds)**

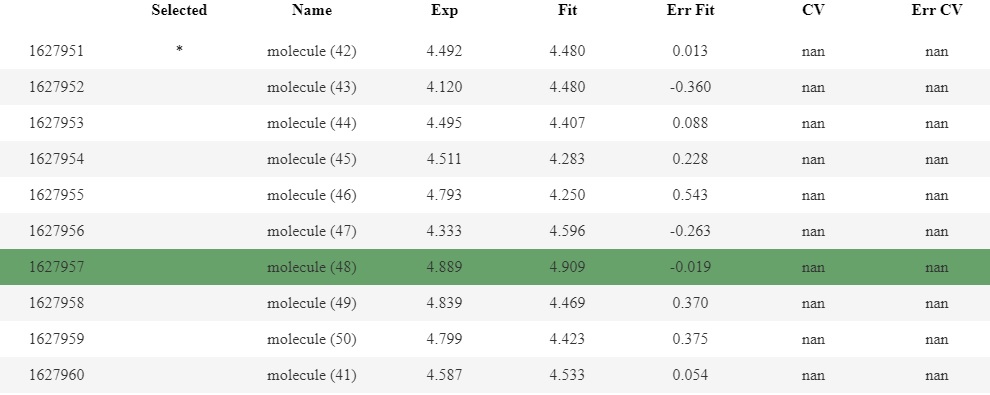
**2.1.3 Hct-116**

**Training Set:**

****

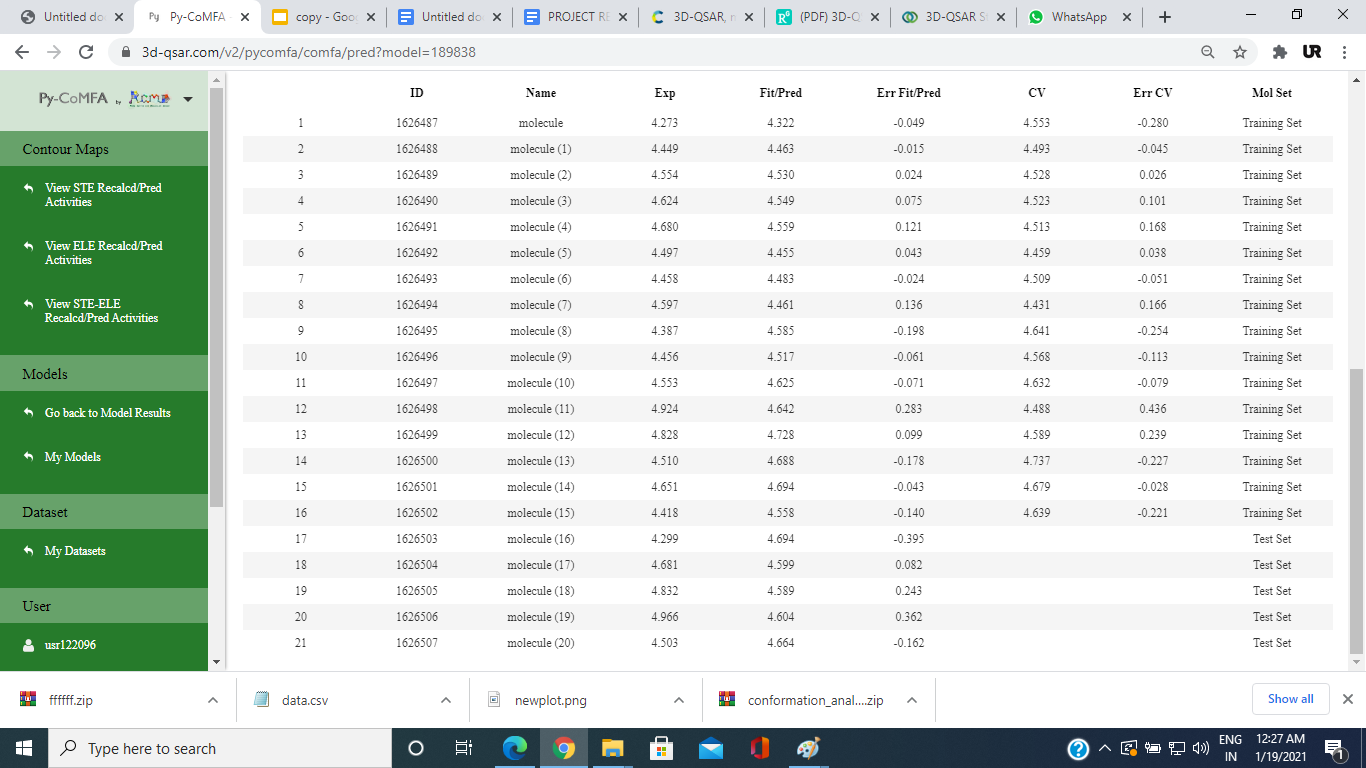
****

**Test Set:**

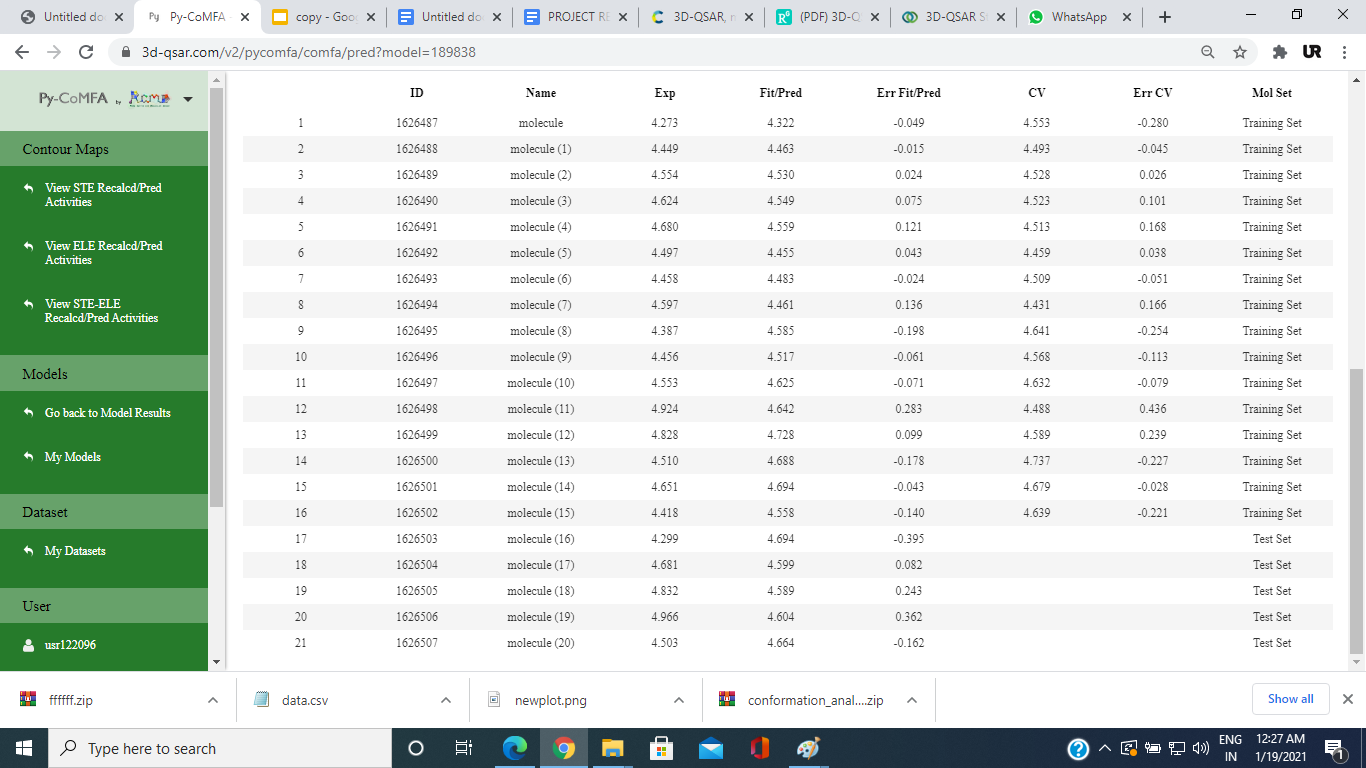
****

**2.1.4 Du-145:**

**Training Set:**

****

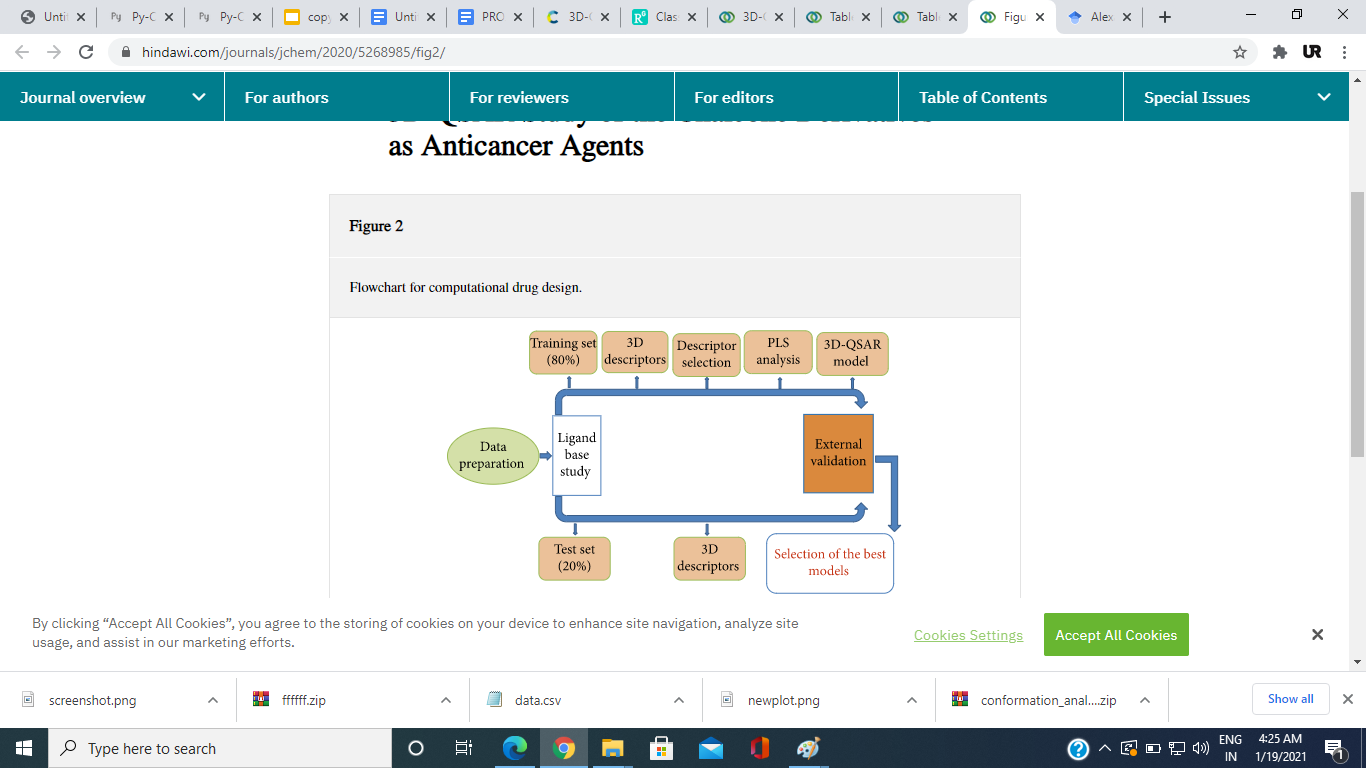
**Test Set:**

****

# **2.2 METHODS :**

**2.2.1 QSAR**

Quantitative structure-activity relationships (QSAR) have been applied for decades in the development of relationships between physicochemical properties of chemical substances and their biological activities to obtain a reliable statistical model for prediction of the activities of new chemical entities. The fundamental principle underlying the formalism is that the difference in structural properties is responsible for the variations in biological activities of the compounds. In the classical QSAR studies, affinities of ligands to their binding sites, inhibition constants, rate constants, and other biological endpoints, with atomic, group or molecular properties.



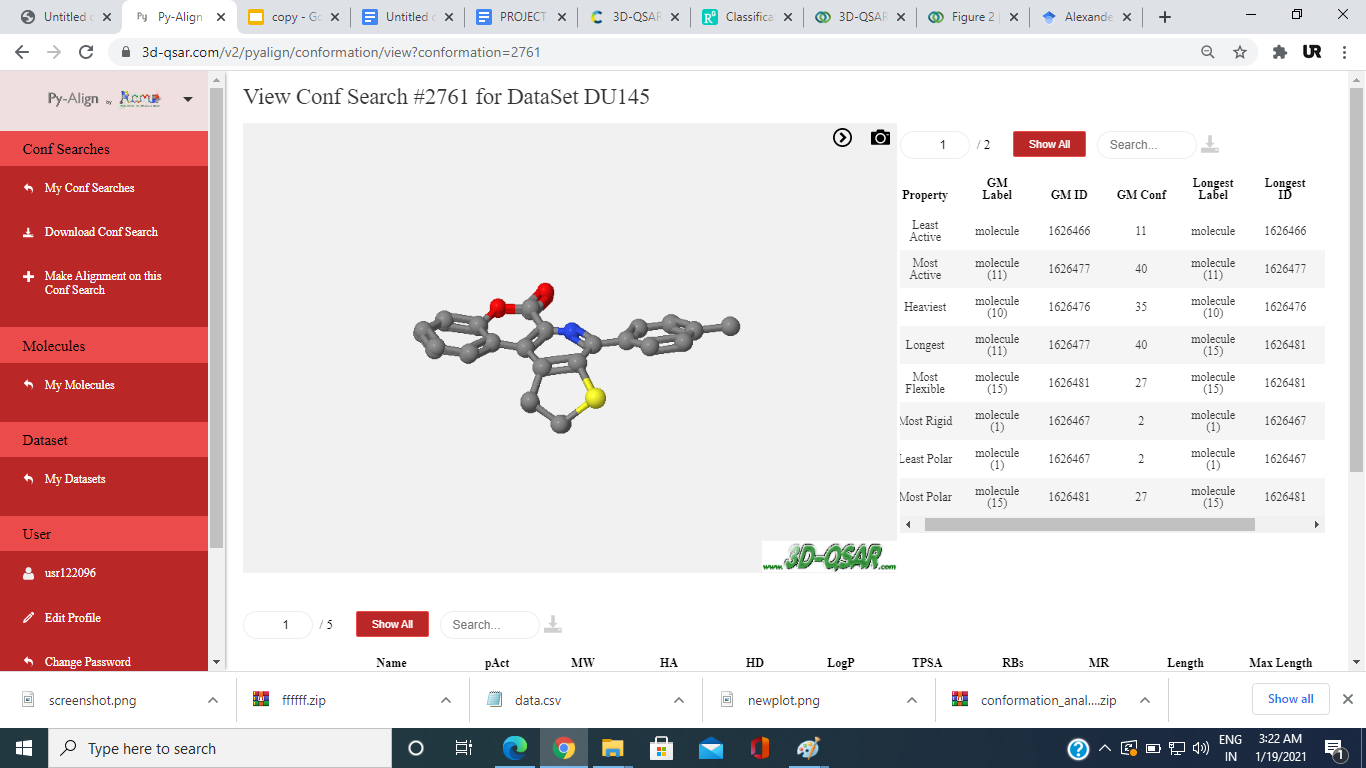
**2.2.1a Molecular Modeling**

The molecular modeling studies of 52 compounds FOR Hct-116 and 23 compounds for Du-145 have been performed utilizing 3D-QSAR After the construction of the 3D structures of the molecules, energy minimization has been performed using the GAFF (**General AMBER Force Field**)

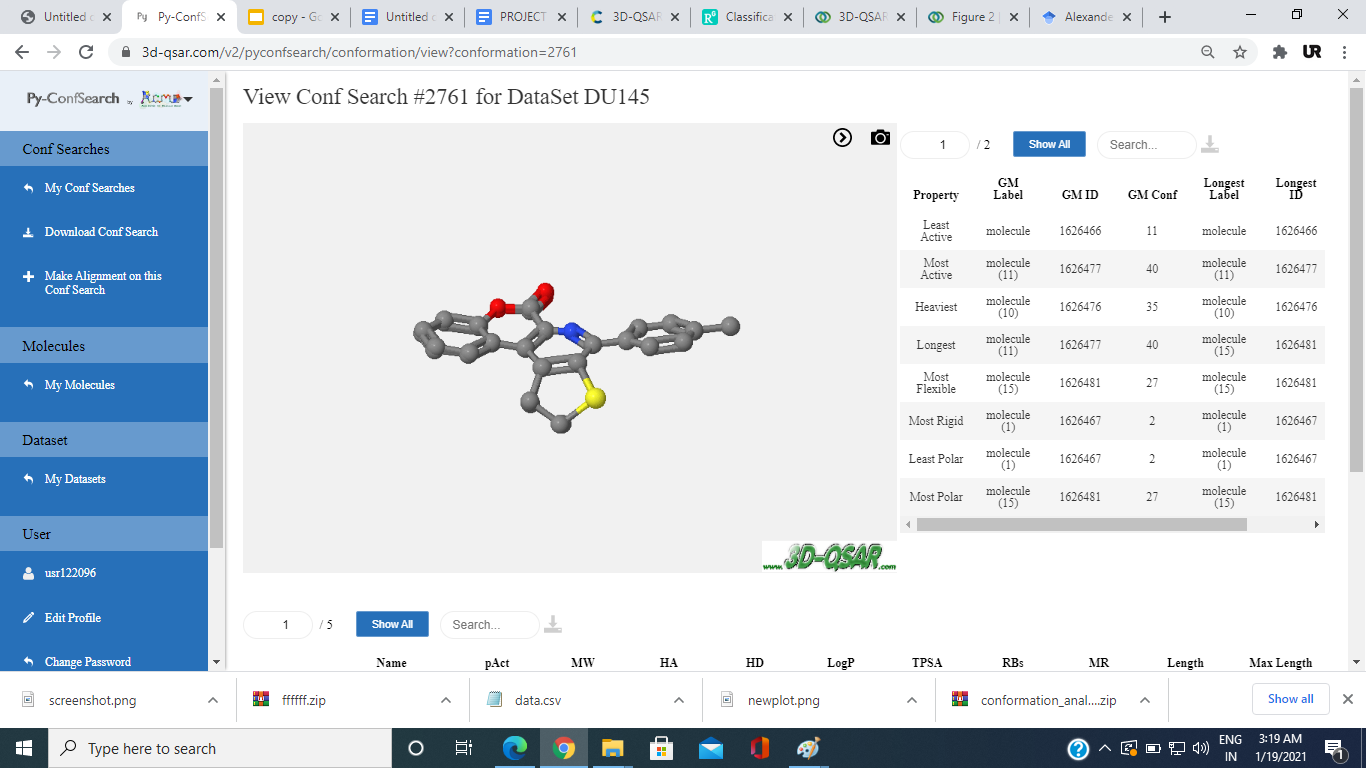
**2.2.1b Molecular Alignment:**

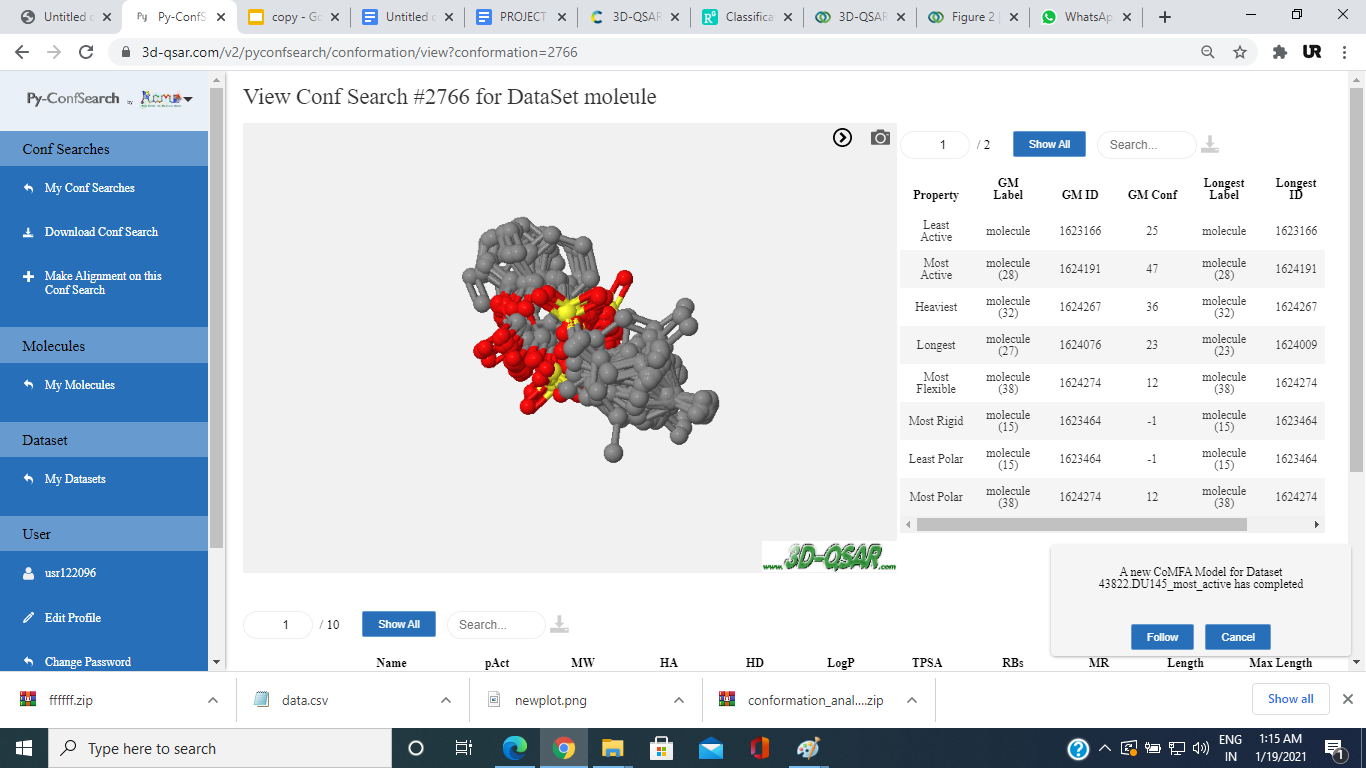
The data set has been aligned on the common core by distill rigid alignment technique using the most active compound (Molecule 11) for Hct-116 and ()as the template (Figure 1.1). The structural alignment of the molecules is the most critical routine in building a reliable 3D-QSAR model. The accuracy of the prediction and the statistical quality of 3D-QSAR models depend strongly on the alignment fig1.1 Describing the proposed alignment and common substructure for the alignment.

**For Hct-116:**

****

**For Du-145:**





**Fig 1.3 (Align molecules structures)**

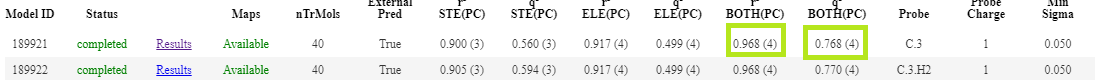
##### **2.2.1c CoMFA Studies:**

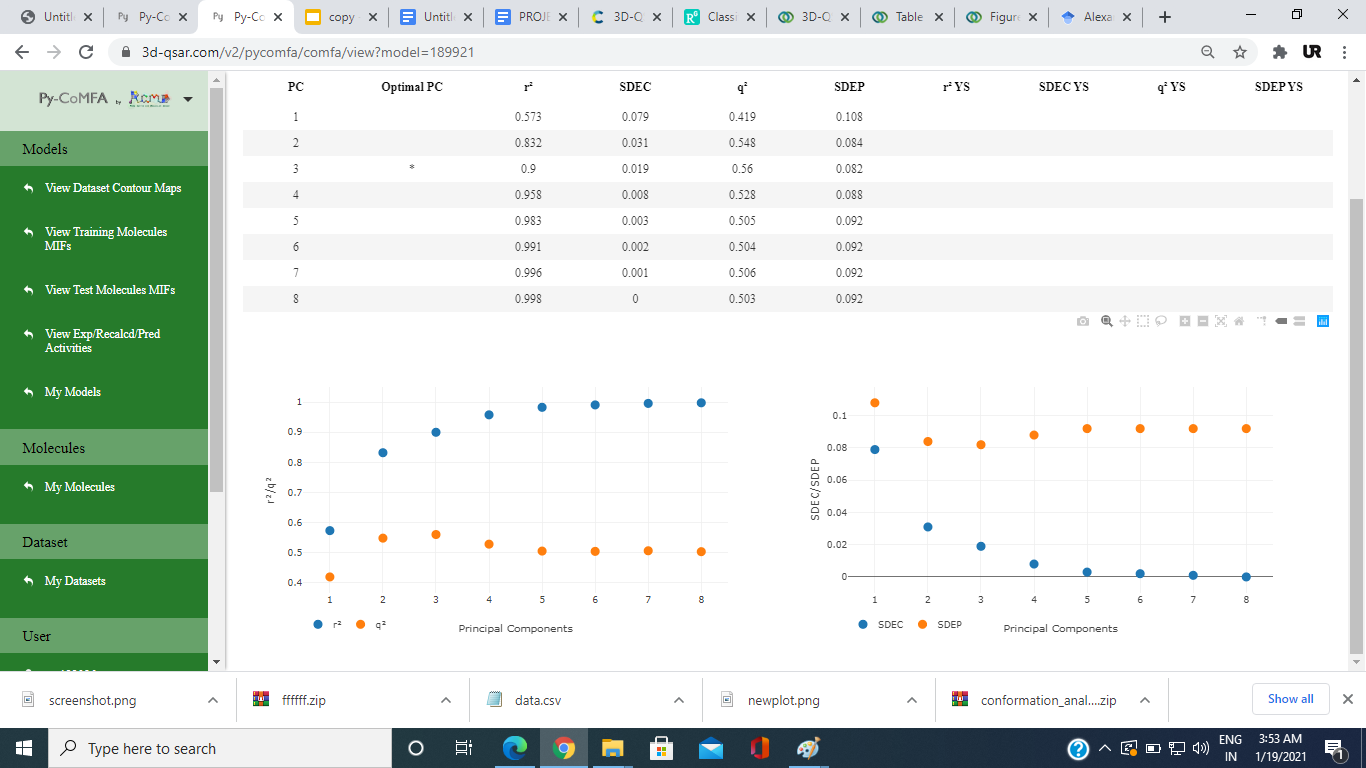
In the generation of CoMFA models , the aligned molecules have been imported in a 3D cubic lattice generated automatically with a grid spacing of 2 Å. In the CoMFA studies, General AMBER Force Field takes a sp3 carbon probe atom with a Van der Waals radius of 1.52 Å and positive charge +1 to generate steric (Lennard-Jones 6-12 potential) field energies and electrostatic (Coulombic potential) fields with a distance-dependent dielectric at each lattice point. To reduce noise and speed up the calculation of potentials, the column filtering value has been set to 2.0 Kcal/mol (default parameters).

##### **2.2.1d Partial Least Square Analysis:**

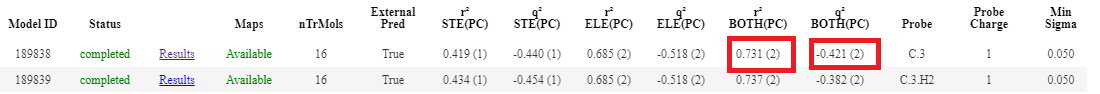
PLS regression analysis , together with cross-validation test, has been used for the generation and internal validation of the CoMFA models. PLS is a powerful tool to derive multilinear relationships between independent and dependent variables. The self-consistency of models generated by PLS can be then evaluated by cross-validation test. In the current study, CoMFA descriptors (explanatory properties) are used as independent variables, and IC50 (target properties) is used as a dependent variable. The first run of PLS analysis has been performed using the full cross-validated leave-one-out (LOO) method, which gave the cross-validated correction coefficient (*Q*2) and the optimum number of components. In this step, the sample-distance partial least square (SAMPLS) method has been used to increase the speed of cross-validation calculation . In the second run of PLS, the models have been calculated by the non-cross-validation method with the optimum number of components. The models in this step have been accessed by statistical indices such as squared correlation coefficient (*R*2), standard error of estimation

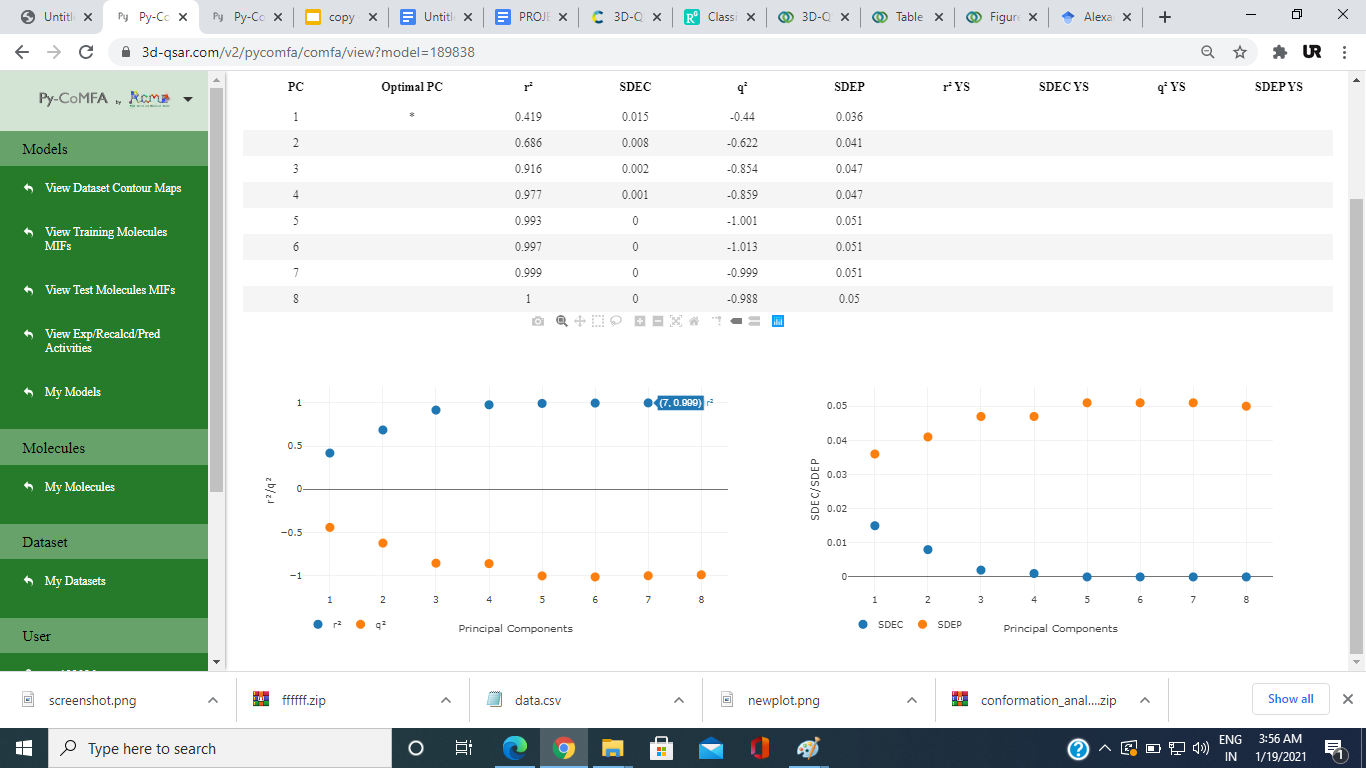
**For Hct-116:**

****

****

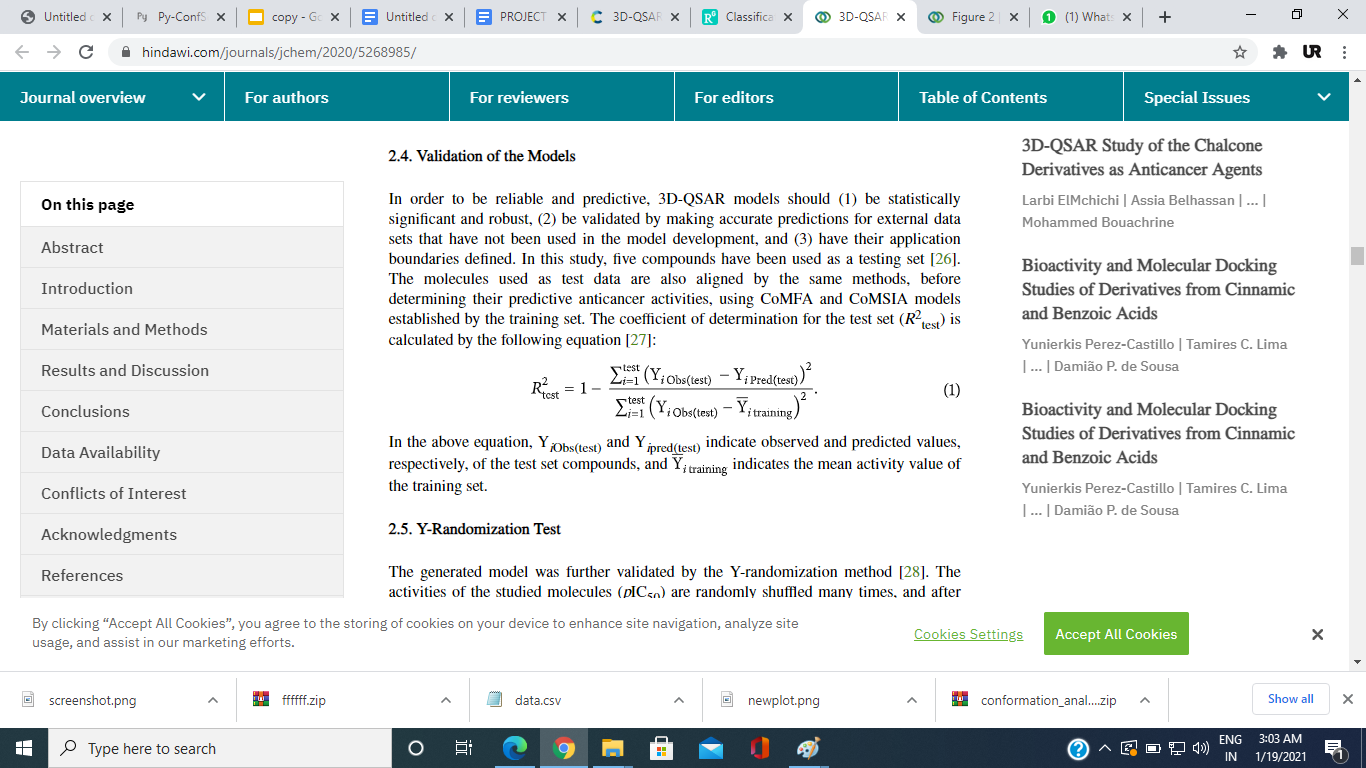
**For Du-145:**

****

****

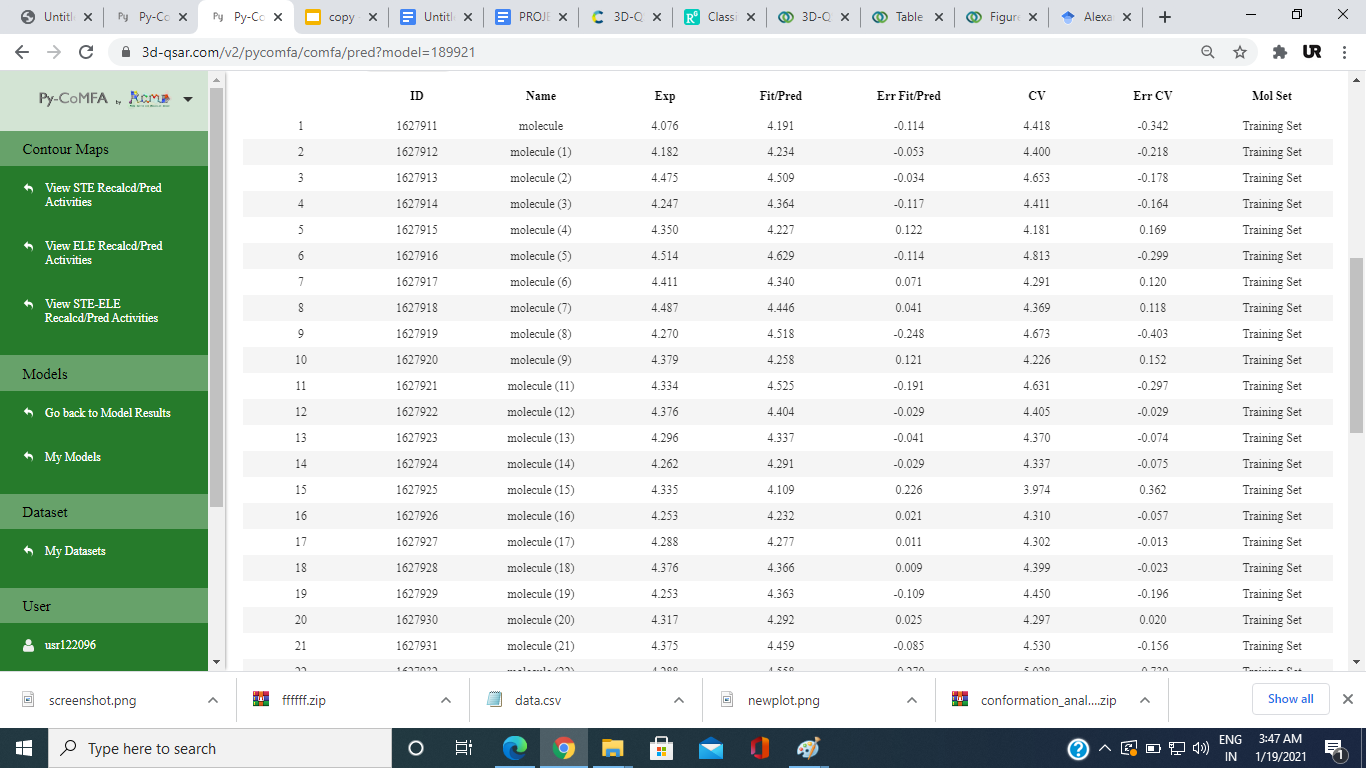
##### **2.2.1e Validation of the Models**

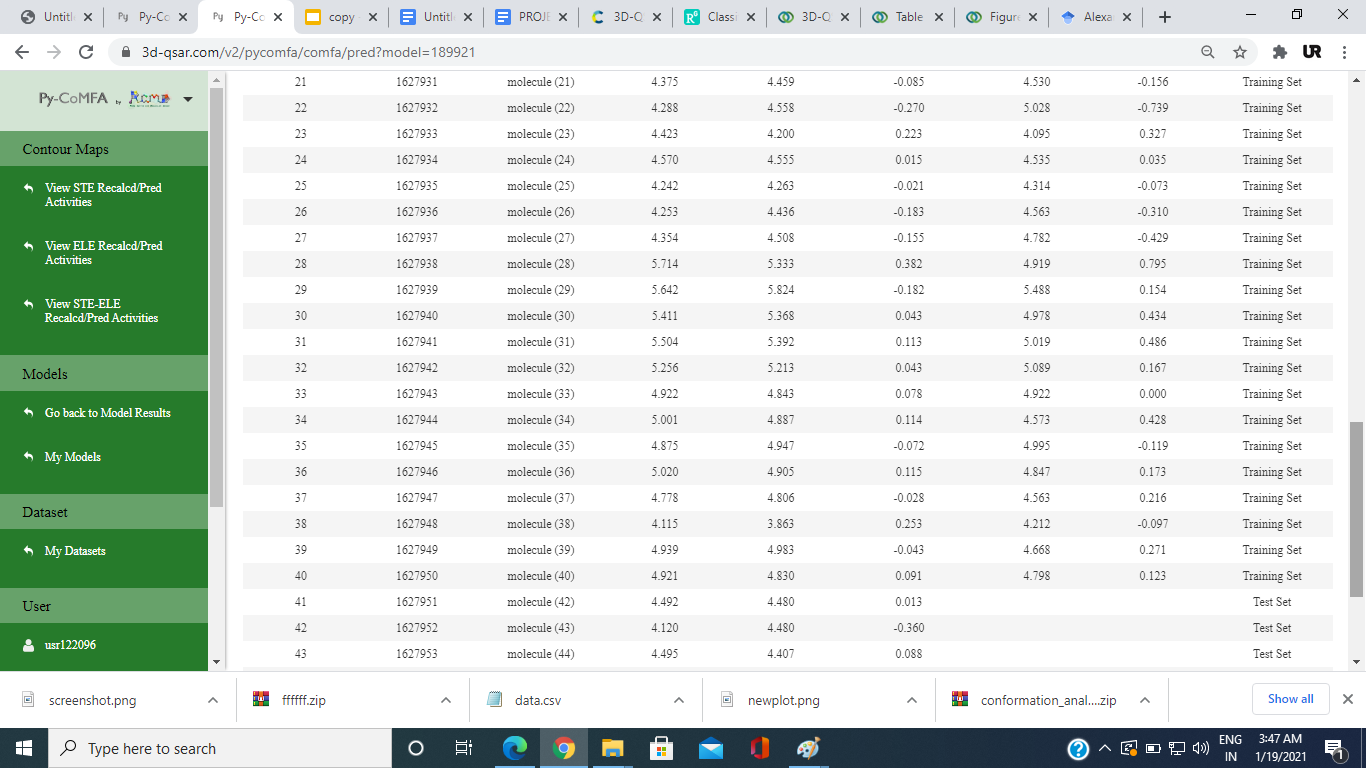
3D-QSAR models should be statistically significant and robust, be validated by making accurate predictions for external data sets that have not been used in the model development, and have their application boundaries defined. In this study, 10 compounds for Hct-116 and 5 compounds for Du-145 have been used as a testing set. The molecules used as test data are also aligned by the same methods, before determining their predictive anticancer activities, using CoMFA models established by the training set. The coefficient of determination for the test set (*R*2test) is calculated by the equation

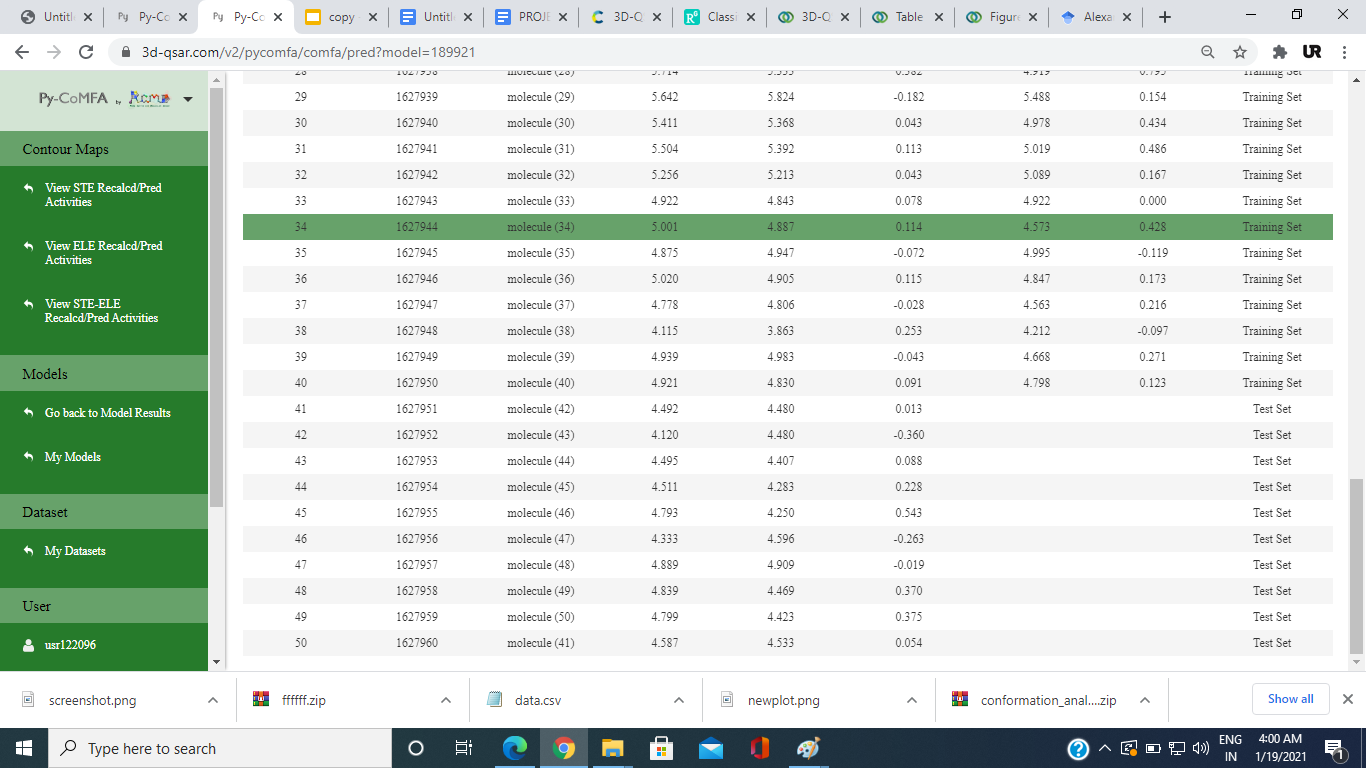


In the above equation, Y*i*Obs(test) and Y*i*pred(test) indicate observed and predicted values, respectively, of the test set compounds, and indicate the mean activity value of the training set.

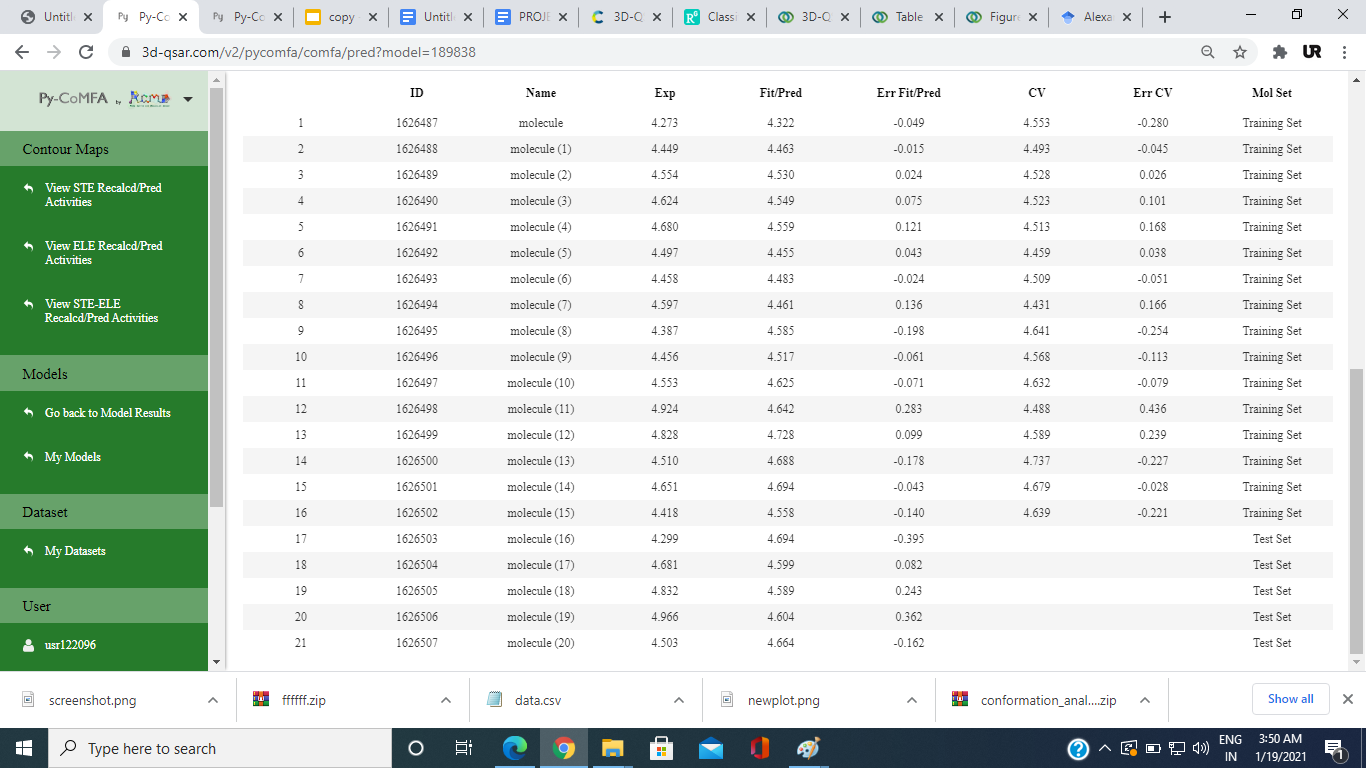
**For Hct-116:**

****

****

****

**For Du-145:**

****

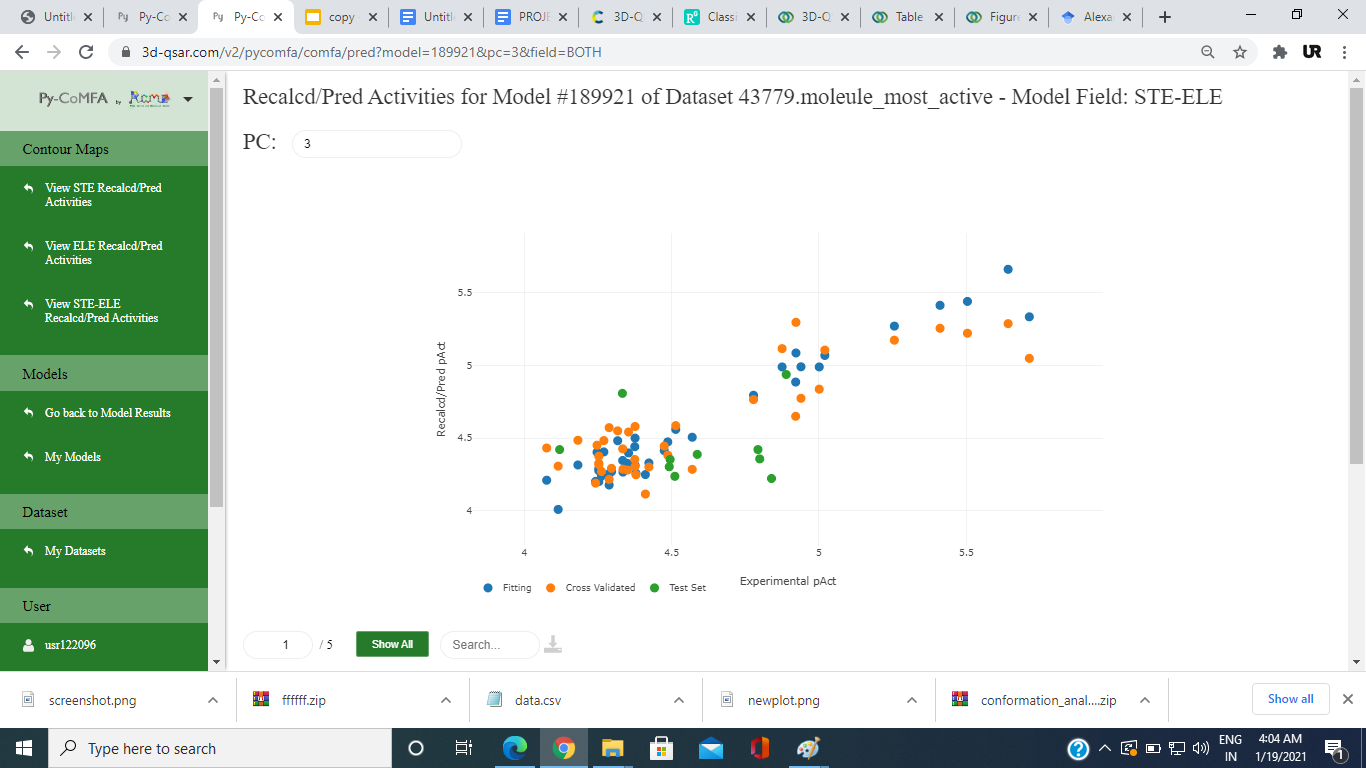
##### **.2.2.1f Y-Randomization Test**

The generated model was further validated by the Y-randomization method . The activities of the studied molecules (IC50) are randomly shuffled many times, and after every iteration, a new QSAR model is developed. The new QSAR models are expected to have lower *Q*2 and *R*2 values than those of the original model. This technique is performed to eliminate the possibility of chance correlation. If higher values of *Q*2 and *R*2 are obtained, it means that an acceptable 3D-QSAR cannot be generated for this data set because of structural redundancy and chance correlation.

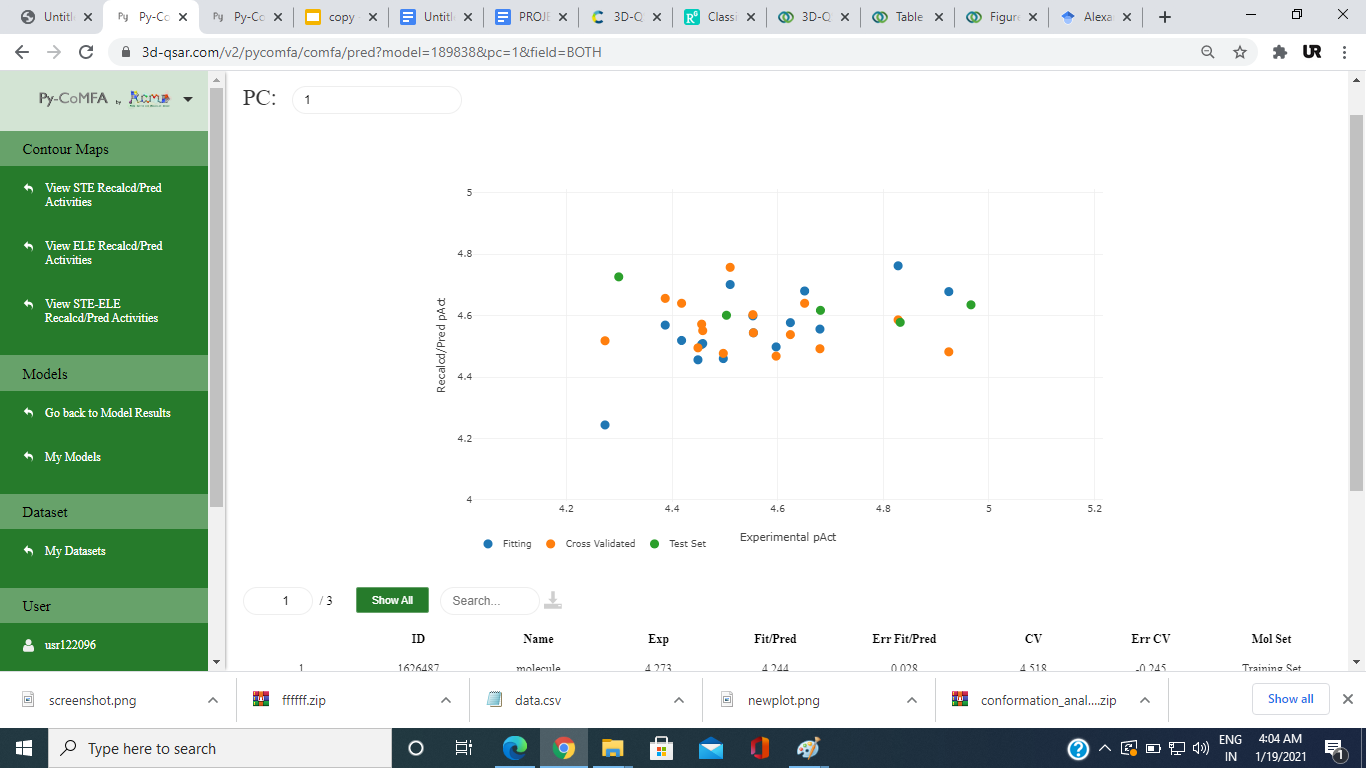
**Model Acceptability Criteria**

If methods give a *Q*2 value >0.5 and *R*2 > 0.6, models are termed as acceptable

**Hct-116:**

****

**Du-145:**

****

**2.2.2 Molecular Docking:**

It is a method which predicts the preferred orientation of one ligand when bound in an active site to form a stable complex.Docking is used for finding binding modes of protein with ligands or inhibitors. It is able to generate a large number of possible structures. In molecular docking, we attempt to predict the structure of the intermolecular complex formed between two or more molecules

* The molecules binding energy which use less energy to bind is much more stable from others molecule
* Molecular docking study was carried out in order to elucidate which of the coumarin derivatives has the best binding affinity against hct 116.
* The structure of Mtb hct 116 used in the study was obtained from a protein data bank with PDB code 5h8b.
* The prepared ligand and receptor were shown in The optimized structures of coumarin derivatives initially saved as SDF files were converted to PDB files using Spartan 14 V 1.1.4.
* The prepared ligands were docked with prepared structures of hct 116 using AutoDock Vina incorporated in PyRx software.
* The docked results were compiled, visualized, and analyzed using Discovery Studio Visualizer.

# **3. CONCLUSION**

##### **3.1 CoMFA Statistical Results:-**

In CoMFA, the obtained independent variables of the training set have been subjected to cross-validated PLS analysis to identify the value of Q2 ( .768 for 4 components) for Hct-116 and (0.421 for 2 components) for Du-145 , and then non-cross-validated PLS analysis has been performed to obtain (Hct-116)R2 = 0.968 and (D-145)R2=0.731 For the CoMFA model.

**3.2 Molecular Docking Result :**

**For Hct-116:**

|  |  |  |
| --- | --- | --- |
| **LIGAND** | **Binding affinity (BA)**  **Kcal/mol** | **Target** |
| molecule (41) | -10.38 | 5H8B |
| molecule (42) | -9.03 | 5H8B |
| molecule (45) | -9.74 | 5H8B |
| molecule (46) | -12.08 | 5H8B |
| molecule (47) | -8.53 | 5H8B |

* The model with 2D and 3D descriptors is of higher excellence and presents a satisfactory correlation with the anti-cancer activity.
* The combination of 2D and 3D descriptors produces a better model to predict the anti-cancer activities of these compounds.
* The QSAR generated model met the criteria for minimum recommended value of validation parameters for a generally acceptable QSAR model.
* The molecular docking analysis has shown that nearly all the coumarin derivatives potentially inhibit hct 116 and Du-145.
* However, compound molecules(46) have higher bind scores for Hct-116 and molecule(20) have higher bind scores for Du-145.
* The QSAR model generated provides a valuable approach for ligand base design, while the molecular docking studies provide a valuable approach for structure base design.

**Data Availability:**

All the data are included within the manuscript.

**REFERENCES**

* Klenkar, J., & Molnar, M. (2015). *Natural and synthetic coumarins as potential anticancer agents*. *7*(7), 1223–1238.
* Putri, D. E. K., Pranowo, H. D., & Haryadi, W. (2019). *Study on Anti-Tumor Activity of Novel 3-Substituted 4 Anilino-Coumarin Derivatives Using Quantitative Structure-Activity Relationship ( QSAR )*. *948*, 101–108. <https://doi.org/10.4028/www.scientific.net/MSF.948.101>
* Vlietstra, R. J., Alewijk, D. C. J. G. Van, Hermans, K. G. L., Steenbrugge, G. J. Van, & Trapman, J. (1998). *Advances in Brief Frequent Inactivation of PTEN in Prostate Cancer Cell Lines and Xenografts1*.
* Ibrahim, B., Omotayo, A., Oyebamiji, A. K., & Oyewole, O. R. (2018). *A DFT-Based QSAR and Molecular Docking Studies on Potent Anti- Colon Cancer Activity of Pyrazole Derivatives ADFT-BasedQSARandMolecularDockingStudiesonPotentAnti-ColonCancerActivityofPyrazoleDerivatives Strictly as per the compliance and regulations of :* (January).
* Dinparast, L., & Dastmalchi, S. (2020). *A QSAR Study on the 4-Substituted Coumarins as Potent Tubulin Polymerization Inhibitors*. *10*(2), 271–277. https://doi.org/10.15171/jcvtr.2015.24