**Qsar Study and Molecular Docking of**

**Coumarin Derivatives of Prostate Cancer and Colon Cancer**

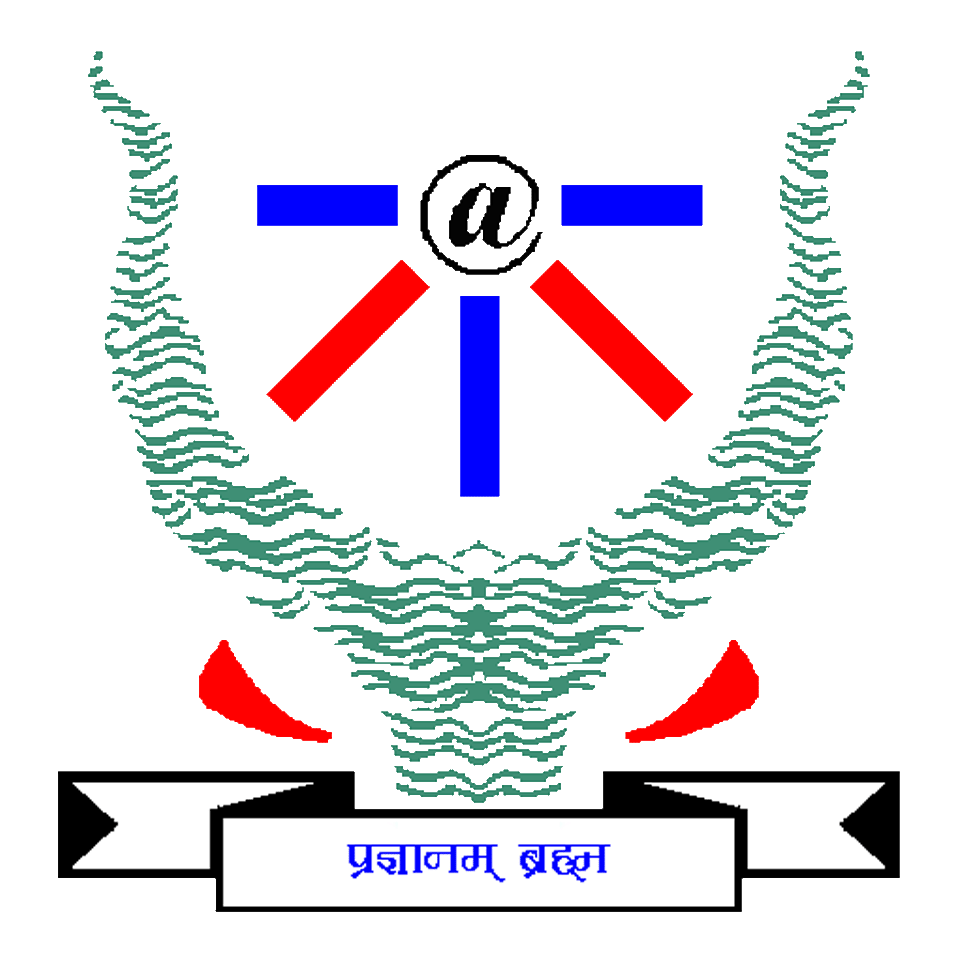
*A thesis submitted*

*In partial fulfilment of the Requirements for the Degree Of*

**MASTER OF TECHNOLOGY**

In

**BIOINFORMATICS**



Submitted by:

**Utkarsh Singh**

**MBI2019011**

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, Utkarsh Singh, Enrollment No. MBI2019011 certify that this thesis work entitled ***“Qsar Study and Molecular Docking of Coumarin Derivatives of Prostate Cancer and Colon Cancer”*i**s submitted by me in partial fulfilment of the requirement of the Degree of Master of Technology in Department of Applied Science, **Indian Institute of Information Technology, Allahabad. I** understand that plagiarism includes:

1. Reproducing someone else’s work (fully or partially) or ideas and claiming it as one’s own
2. Reproducing someone’s else work (Verbatim copying or paraphrasing) without crediting.
3. Committing literary theft (copying some unique literary construct).

I have given due credit to the original authors/sources through proper citation for all the words, ideas, diagrams, graphics, computer programs, experiments, results, websites that are not my original contribution. I have used the quotation marks to identify verbatim sentences and given credit to the original authors/sources.

I affirm that no portion of my work is plagiarized. In the event of a complaint of plagiarism, I shall be fully responsible. I understand that my supervisor may not be in a position to verify that this work is not plagiarized**.**

**Name:** UTKARSH SINGH **Date:**

**Enrollment No:** MBI2019011

**Department of Applied Sciences**

**IIIT-Allahabad (U.P.)**

**CERTIFICATE FROM SUPERVISOR**

It is certified that the work contained in the thesis titled” **Qsar Study and Molecular Docking of Coumarin Derivatives of Prostate Cancer and Colon Cancer**” by “Utkarsh Singh” has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

**Date: Dr. Nidhi Mishra**

**Place:** Allahabad Assistant Professor

IIIT Allahabad

**CERTIFICATE OF APPROVAL**

The forgoing thesis is hereby approved as a credible study in the area of Bioinformatics and its allied areas 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 is submitted.

Signature & Name of Committee Members\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

On final examination and approval of the thesis\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ORIGINALITY REPORT**

**ACKNOWLEDGEMENT**

First and foremost, I would like to express my deep and sincere attitude to my supervisor **Dr. Nidhi Mishra**, Assistant Professor, Department of Applied Sciences, IIIT Allahabad who has supported and encouraged me throughout my thesis work with her patience and knowledgeable skills. Her guidance and encouragement gave me the motivation to complete my thesis work successfully and on time.

I am also highly obliged **to Dr. Aashutosh Mishra**, Head of Department, Department of Applied Science, IIIT Allahabad and all my respected faculty members whose experience and guidance helped me come out of situations whenever I got stuck in my project work.

I am extremely thankful to **Mr. Akhilesh Maurya**, Research Scholar, Department of Applied Science, IIIT Allahabad for providing me great suggestions and advice to improve the presentation of my thesis work.

I am also thankful to God, my family and friends who stood by me as a motivation to complete my work successfully. Also, I would like to thank the non-teaching staff for their valuable cooperation. I would like to thank my family members for their blessings and inspiration. Above all, I would like to thank almighty for his blessings.

UTKARSH SINGH

MBI2019011

IIIT ALLAHABAD

**ABSTRACT**

**Ethnopharmacological relevance**: Coumarin or 1,2 benzopyrone derivatives are one of the most critical classes of benzopyrones it occurs naturally in tonka beans and cinnamon, but can also be found in certain amount in the green tea, carrots, some beers and bison grass, primarily responsible for inhibiting the growth, proliferation, and metastasis of various tumor cells with a different kind of mechanism including inhibition of carbonic anhydrase, targeting PI3K/Akt/mTOR signaling pathways, regulating the reactive oxygen species, inhibition of microtubule polymerization, inhibition of tumor multidrug resistance, promoting cell apoptosis protein activation and inhibition of tumor angiogenesis. Some of the Coumarin derivatives have been used for the treatment of prostate, colon, and breast cancer and encouraging results have been obtained. However little work has been done concerning the determination of the pharmacogenomic effect of Coumarin on Colorectal cancer (CRC) and adenocarcinoma of the prostate (CAP).

**Aim of the study:** The present study aims to predict the possible anti-cancerous effects of Coumarin compounds and their derivatives against core targets of (CRC) and (CAP). the determination of their relative efficiencies in forming a drug-target complex by comparison of their binding energies obtained through 3d-Qsar Study followed by Molecular Docking. Moreover, this study also identifies the active binding sites in the protein involved in the formation of the Coumarin-protein complex through various covalent as well as non-covalent interactions. Apart from the core set of targets, the above set of procedures was also performed with two novel proteins (5H8B and 5YC1) that were recently reported to be involved in the upregulation of (CRC) and (CAP).

**Material and methods:** To begin with, the 3D-SDF files of all-natural and synthetic Coumarin Derivatives were downloaded after the literature survey and Extract from open babel. Following that, all the proteins whose expression was aberrantly upregulated during the development of colorectal cancer (CRC) and adenocarcinoma of the prostate were searched out through a literature survey and navigation through various protein databases like PDB.  Side by side, the protein targets specific to Coumarin Derivatives were determined through open babel Consequently, dock the Coumarin derivatives with protein of the HCT-116 and DU-145 using Autodock Vina. Subsequently, the 2D protein-ligand interaction diagram of the above interactions was obtained using Discovery Studio.

**Results:** Based on the results obtained after 3d- qsar Molecule 42 and Molecule 11 is highly preferable structure derived by the 3d-qsar model concerning molecular docking, we found that out of all the Coumarin derivatives which had HCT-116 as their putative target, Molecule(16) isoxazolyl was found to be most effective binding giving the lowest binding energy(-8.7 kcal/mol with an inhibition constant of 4.54 micromolar). for Du-145 , molecule 11 gave the lowest binding energy(-8.08 kcal/mol with an inhibition constant of 1.19 micromolar

**Conclusion:** Our in-silico analysis revealed that out of all Coumarin derivatives, isoxazolyl and Mammea B/BA hydroxyl could serve as potential therapeutic agents for controlling the overexpression of our core target 5H8B and respectively. molecule 11 and molecule 18 compositions of thiazolyl-3-aryl-pyrazole-4-carbaldehydes could turn to be effective therapeutic agents for controlling the overexpression of 5YC1 respectively.

**INDEX**

**TITLE** **PAGE**

1. INTRODUCTION

2. METHODOLOGY

3. RESULTS AND DISCUSSION

4. CONCLUSION

5. REFERENCES

1. **INTRODUCTION**

Millions of people continue to fall sick with cancer disease each year all over the world, especially in developed countries. Colorectal cancer (CRC) is the third most common cause of cancer death in both men and women in the World accounts for nearly 9555027 deaths due to cancer were recorded in 2018, of which 880792 (9.2%) were related to colorectal cancer A persistent change in your bowel habits, including diarrhea or constipation or a change in the consistency of your stool. Rectal bleeding or blood in your stool. Cramps, gas, or pain like abdominal discomfort happens persistent.  further people get older the risk of colorectal cancer increases. Colorectal cancer can occur mostly in people older than 50, but colorectal cancers also occur in young adults and teenagers. The average age at the time of diagnosis for men is 68 and for women is 72 in terms of colorectal cancer.

**Stages of colon cancer:**

**Primary tumor:** Primary tumor (T) assign how large the original tumor is and whether cancer has grown into the wall of the colon and spread to the nearby areas of the colon.

**Regional lymph nodes:** Regional lymph nodes (N) assign to whether cancer cells have spread to nearby of the lymph nodes.

**Distant metastases:** Distant metastases (M) assign to whether cancer has spread from the colon to other organs of the body, for-example the liver or lungs.

**Colon cancer therapies**:

**Radiation Therapy:** Radiation is the strategic way to kill cancer cells by using ionizing radiation or photons.

**Chemotherapy:** it is highly individualized and difficult to perform which based on several factors

Prostate cancer is the world's second leading cause of cancer death in men, behind only lung cancer. Around 1 man out of 41 will die of prostate cancer. Prostate cancer is a serious disease, further, most of the men diagnosed with prostate cancer do not die cause of it. an estimated 19.3 million new cancer cases (By excluding18.1 million nonmelanoma skin cancer) and almost 10.0 million cancer deaths (9.9 million excluding nonmelanoma skin cancer) occurred in 2020 A persistent impact occurs in the human body  Painful ejaculation, Dull pain in the lower pelvic area, hips or upper thighs, Bone pain, Pain in the lower back, loss of weight, Frequent urinating, Trouble urinating, pain, burning or weak urine flow, Blood in the urine (Hematuria) and Loss of appetite, The risk of getting prostate cancer goes up. Prostate cancer is rarely found in younger men who are below age 40. Damage to the DNA and alike genetic material of prostate cells is more likely for men over the age of 55. Damaged or to form tumors the abnormal prostate cells can begin to grow out of control. There have four stages in prostate cancer.

**Prostate cancer therapies:**

**Radiation Therapy:** The radiation therapy is the strategic way to kill cancer cells by use of ionizing radiation or photons

**Chemotherapy:** chemotherapy is very difficult to perform and it is highly individualized which is based on several factors

**Hormone Therapy:**  Hormone therapy is part of the standard of care for prostate cancer and it is also called androgen deprivation therapy or ADT.

**Active Surveillance:**  the concept that low-risk prostate cancer is method which use in Active Surveillance

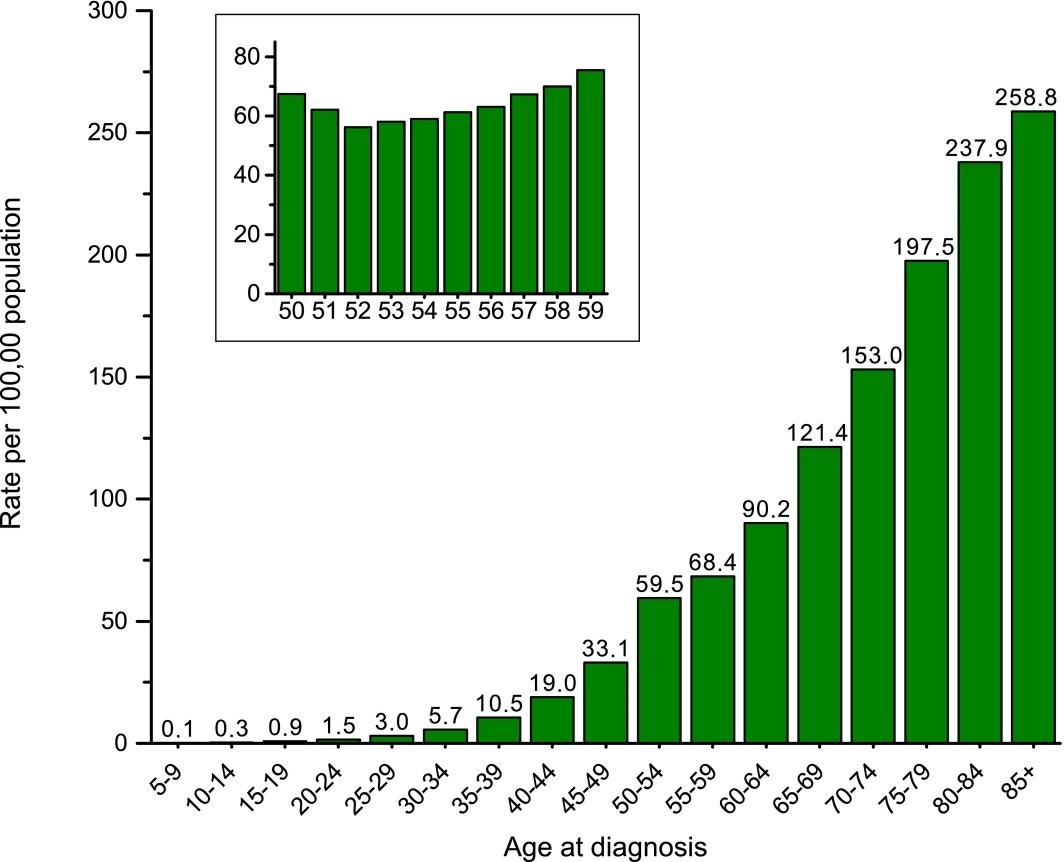
**Precision Medicine:** Precision medicine is method to treat the right patient with the right medicine at the right time by using of new diagnostic tests.

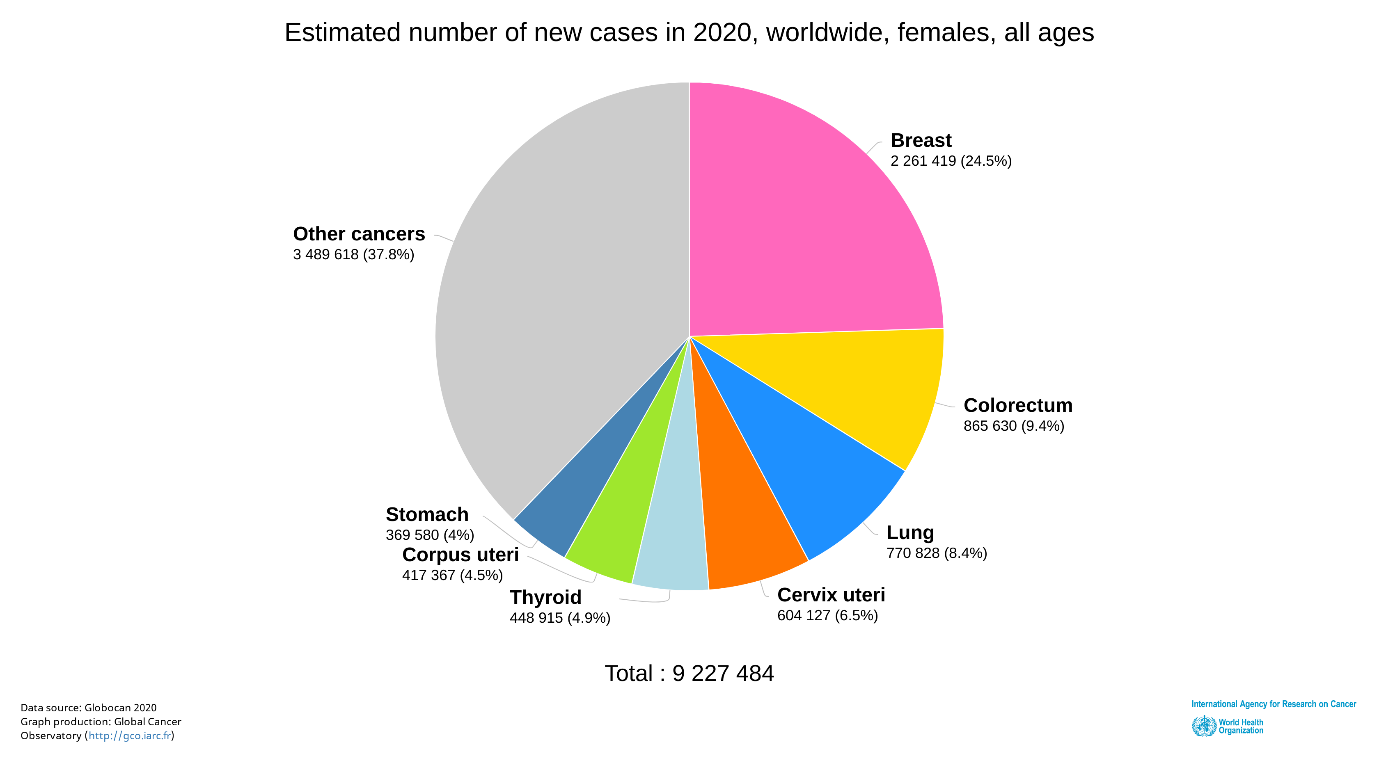
**immunotherapy:** most of the patients with anti-cancer immune responses, progressing cancer.

Several strategies have been proposed for the treatment of colon cancer and prostate cancer developed but unfortunately, not a single approach so far has turned out to be competent enough that could single-handedly control the disease. Even the use of chemotherapy more often than not requires a combination of drugs which in the later run has many side effects. In the recent years, the use of plant-based products for the treatment of different types of diseases has gained paramount importance and has encouraged the researchers and oncologists all over the world to use them as a possible substitute for the complex treatment procedures.

The benzopyran-2-one, or chromen-2-one ring system, which is present in natural products because the anticoagulant warfarin, which has medicinal chemists and intrigued chemists for many years to explore the synthetic analogues of natural coumarins for their applicability as drugs or medicine, is used to make coumarin compounds. The majority of compounds based on the coumarin ring system have been produced using novel synthetic approaches. The coumarin sulfamates (COUMATES), pyranocoumarins, and furanocoumarins are intriguing derivatives that have use in photochemotherapy, anticancer, and anti-HIV therapy, as well as as central nervous system stimulants, anti-coagulants, anti-bacterial, anti-inflammatory, and dyes. Aromatase and sulfatase inhibiting effects have been discovered in several coumarins and their active metabolite 7-hydroxycoumarin analogues, which are of particular importance in breast cancer treatment. Selective coumarin oestrogen conjugates and oestrogen receptor modulators (SERMs) have also been proposed as coumarin-based anti-breast cancer medicines. There is a strong impetus to uncover possible novel medication therapies for breast cancer because it is the second biggest cause of mortality in women in the world, after lung cancer. As a result, the goal of this review is to highlight important coumarin analogues with anti-breast cancer activities, as well as their structure-activity relationships and mechanisms of action on selected receptors in breast tissues, as well as the structure-activity relationships methods that have been used to construct these pharmacologically important coumarin analogues.

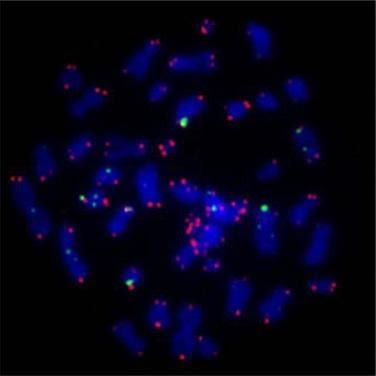
Charts to show rate of colon cancer and prostate cancer:



,

**HCT-116 Cell Information:**

The colon cancer starts in Rectum or the colon large intestine. The rectum is placed at the end of the colon. Rectum and colon are located in the lower portion of your digestive system. American Society estimates that about one out of 25 women and one out of 23 men will develop colorectal cancer during their lifetime. For transforming growth factor beta 1 (TGF beta 1) and beta 2 (TGF beta 2) expressions the HCT 116 cells response are positive. HCT116 is a human colon cancer cell line which is used in drug screenings and therapeutic research. An adultmale initiates the human colorectal carcinoma cell line HCT116 (ATCC® CCL-247TM). The cells are adherent with an epithelial morphology. The implantation into immunocompromised mice, the cells form distant metastases and primary tumors.



**Fig 1.1 NMR of HCT-116 cell line**

**Characteristics of HCT-116 cells:**

* HCT116 cells have a mutation at codon 13 of the KRAS proto-oncogene, making them good transfection targets for gene therapy research.
* In xenograft models, cells transfected with viral vectors expressing the p53 gene can metastasis and have an epithelial appearance.The cells are still stuck in the G1 phase.
* The knockout of MARCH2 inhibited HCT116 cell growth by putting a strain on the endoplasmic reticulum.5-Fu/P85 copolymer micelles were also observed to impede the growth of HCT116 colonies.

**Uses for HCT-116 Cell Line**

HCT 116 is utilised in tumorigenicity tests and other research, and it has been established that Cyclin D1 is important for lithocholic acid hydroxyamide action. With flavopiridol, docetaxel, and 5-FU, HCT116 cells can function in xenografts and restrict tumour development in vitro. The HCT116 cell line has been used in a number of biological investigations addressing colon cancer growth and inhibitors. In HCT-116 cells, there are two variations: one lacking The Insp8 gene is part of a cell's energy metabolic process, and the other with high Insp8 gene expression, which can impact the cellular phenotype such as outcome.

**HCT-116 Xenograft Model**

Validated colon cancer HCT-116 xenograft model was established by Altogen Labs (Austin,TX) and commercially available, please see the following link for HCT-116 in vivo xenograft model laboratory testing service:

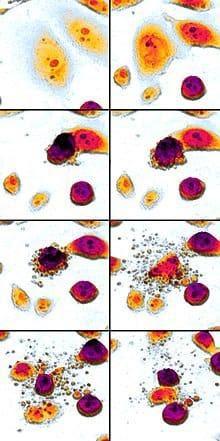
**DU-145 Cell Information**

The prostate is about a walnut in terms of shape and size. It is a small gland located in men organs and part of men reproductive system. Prostate cancer develops in the prostate gland It sits low in the pelvis, just in front of the rectum and below the bladder. Like a consequence of abnormal androgenic stimulation as the prostate cancer (PC) developed. Thereafter most of the prostate cancer cell lines express mutated forms of androgen receptor (AR) and it is androgen independent as like DU145.

A new stably transfected prostate cancer cell line expressing the AR (DU145-AR) are produced and characterized.  A lower proliferation rate than mock transfected cells showed by Untreated DU145-AR cells, although responded to testosterone treatment. PSA mRNA, undetectable in mock DU145 cells,

By testosterone in DU145-AR have presence of DU 145 and it is upregulated.  After testosterone treatment in about five percent of DU 145-AR cells showed modification of morphology and enriched of f-actin. secreted urokinase type DU145-AR plasminogen activator PA and plasminogen activator uPA protein activity were lower than in AR negative cells:

Commonly human prostate cancer cell line DU 145 uses in research. Therapeutic research performed by use of DU145, PC3, and LNCaP are considered to be the standard prostate cancer cell lines.



**Fig 1.2 NMR of DU-145 cell line**

**Characteristics of DU-145 cells:**

* The DU145 cell line was derived from primary prostate adenocarcinoma origin's central nervous system metastasis, which was removed during a parieto-occipital craniotomy.
* DU145 does not express prostate-specific antigen (PSA) and it is not hormone-sensitive.  By comparison of PC3 cells DU145 cells have moderate metastatic potential, where PC3 have high metastatic potential. DU145 cells are androgen receptor positive.
* A consequence of abnormal androgenic stimulation developed prostate cancer.
* The majority of two classical cell lines PC-3 and DU-145, from the human prostate cancer cell lines are reported to be androgen receptor (AR)-negative.
* Most of the DU145 cell lines express mutated forms of androgen receptors (AR). or they are androgen independent.

**Uses for DU-145 Cell Line**

* DU 145 is the prostate cancer cell lines which are commonly used in research.
* In restoring functional expression of the AR gene the cotreatment of AR-negative cell line DU145 with 5-Aza-CR and TSA is more effective.
* Silibinin reduced constitutive Stat3 phosphorylation in a concentration-dependent manner In a prostate cancer model DU145 cell line.
* The significant growth-inhibitory effects against two well-characterized androgen-independent prostate cancer cell lines where one is DU 145 from the silibinin’s ability to inhibit pSTAT3 translates.

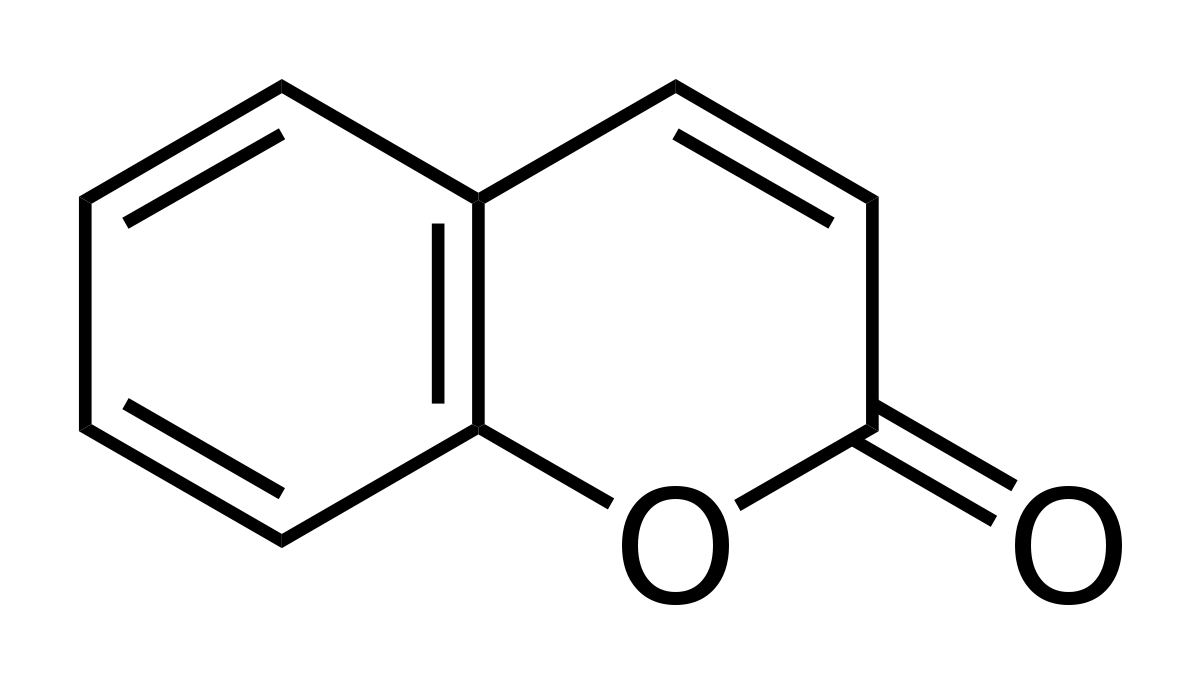
**DU-145 Xenograft Model**

Validated prostate cancer DU-145 xenograft model was established by Altogen Labs (Austin,TX) and commercially available, please see the following link for DU-145 in vivo xenograft model laboratory testing service:

**Coumarin**

Coumarin which formula C9H6O2 is an aromatic organic chemical compound 2H-1-benzopyran-2-one. lactone-like chain −(CH)=(CH)−(C=O)−O−, has replaced from its benzene molecule to forming a second six membered heterocycle with 2 adjacent hydrogen atoms which shares 2 carbons with the benzene ring. This molecule can considered as a lactone and it can also be placed in the benzopyrone chemical class.

**Basic Structure of Coumarin Derivative.**



**C9H6O2**

**fig 1.3 Coumarin chemical structure**

Coumarin may be found in a variety of herbal compounds, including celery and strawberries, as well as woodruff, sweet clover, tonka beans, and lavender oil.

Coumarin and its derivatives are abundant in nature, and many of them have important and varied biological actions. Coumarins may be found in the leaves, roots, and seeds of a wide variety of plants. Tonka bean has a high concentration of coumarin in nature. Tonka bean is known in French as coumarou, which is also the origin of the term coumarin. Coumarin is a colourless crystalline substance with a bitter taste and a sweet odour similar to vanilla. Natural and synthetic coumarin derivatives are split into numerous subclasses.Simple coumarins (e.g. coumarin and limettin), furanocoumarins (e.g. imperatorin and isopimpinellin, psoralen, angelicin), and pyranocoumarins (e.g. imperatorin and isopimpinellin, psoralen, angelicin) are the most commonly reviewed coumarins classes (e.g. xanthyletin, seselin). Murray et al classified coumarins using a biogenetic method based on the quantity of nuclear oxygen atoms present in coumarin-containing molecules.

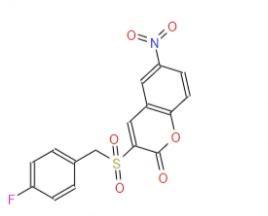
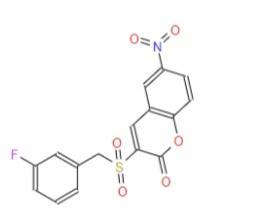
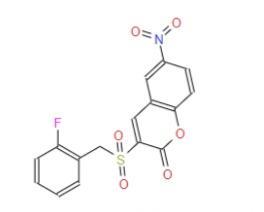
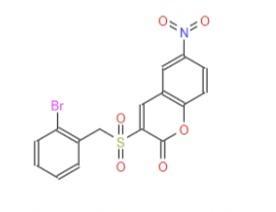
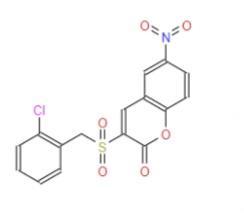
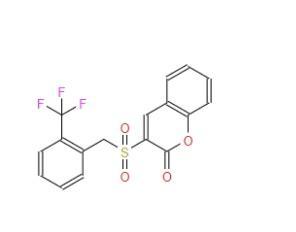
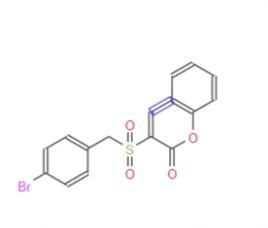
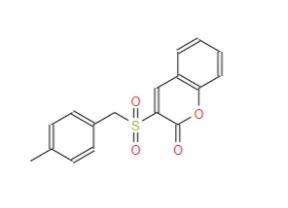
Many coumarin derivatives have been discovered to offer a variety of therapeutic effects, including anticancer, photochemotherapy, anti-HIV treatment, and usage as central nervous system stimulants, according to various studies. Antibacterial, anti-inflammatory, anti-coagulants, and dyes are all possible applications. Coumarins are lipid-lowering medicines that have a mild effect on triglyceride levels.

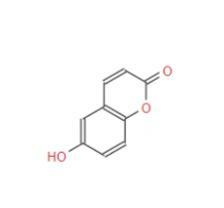
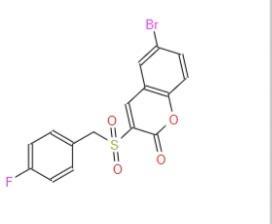
By scavenging reactive oxygen species, hydroxycoumarins can protect against free radical damage. It's a potent antioxidant that breaks apart chains. Some coumarin derivatives are employed as food adulterants and flavouring agents, however consuming too much of them can induce side effects such moderate nausea, diarrhoea, and hepatotoxicity. Coumarin-based medications are being marketed as therapeutic medications in various European countries. Coumarin derivatives are used to treat lymphoedema in Europe, however they are not licenced for therapeutic use in the United States owing to hepatotoxicity. However, because coumarins have a low estrogenic action, derivatives of them can be used as therapeutic agents to prevent menopause-related disorders including osteoporosis, an increased risk of cardiovascular disease, and cognitive deficits.

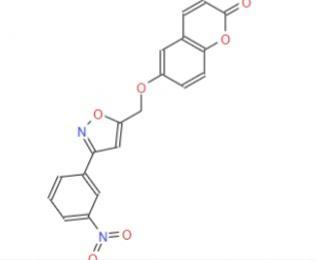
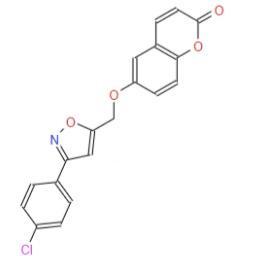
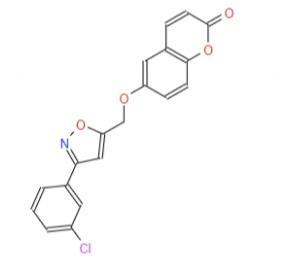
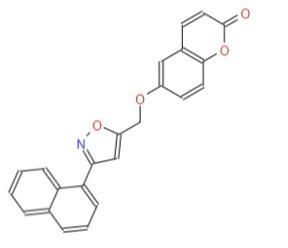
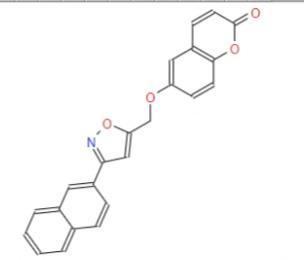
**DATASET:**

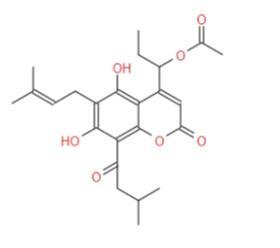
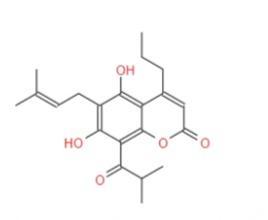
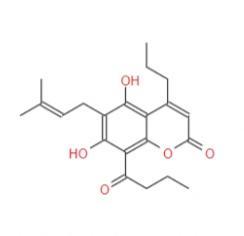
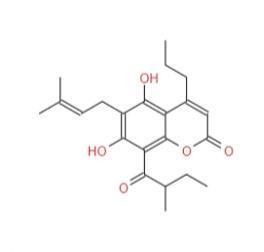
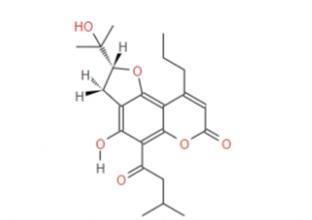
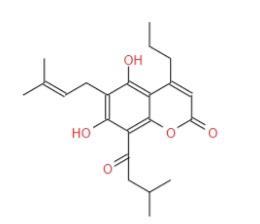
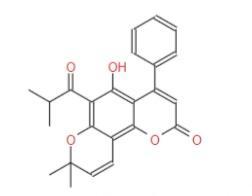
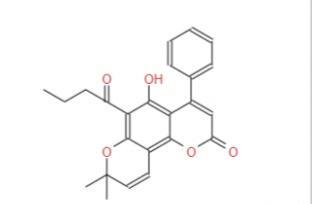
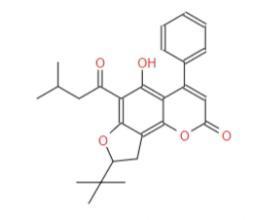
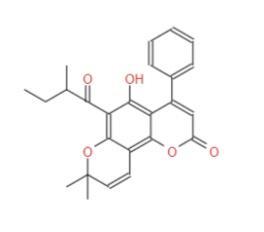
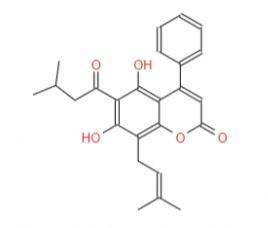
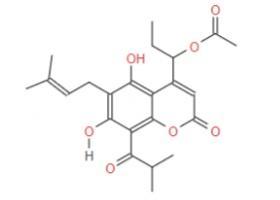
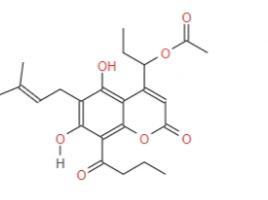
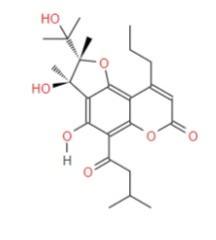
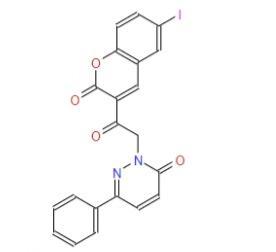
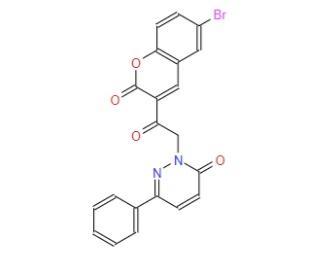
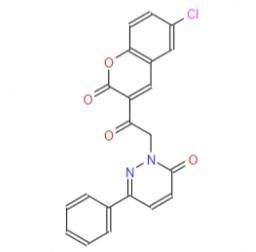
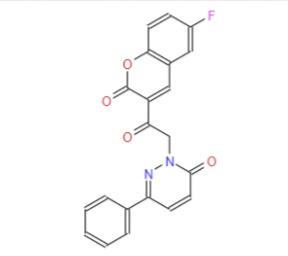
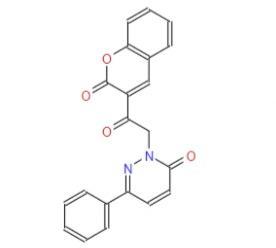
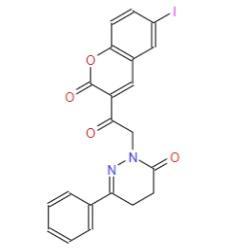
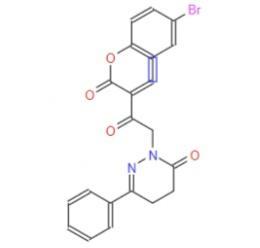
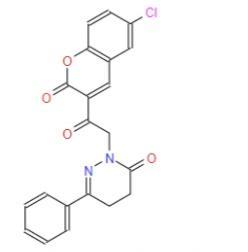
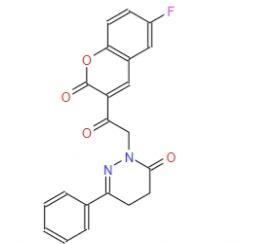
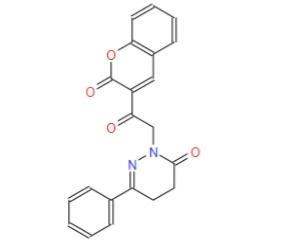
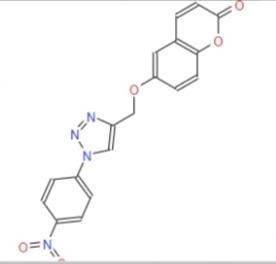
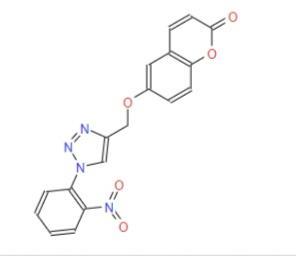
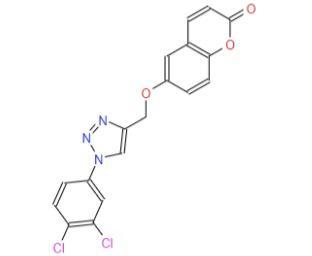
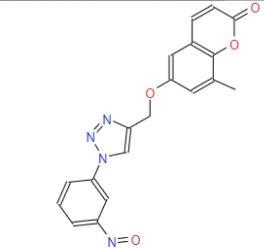
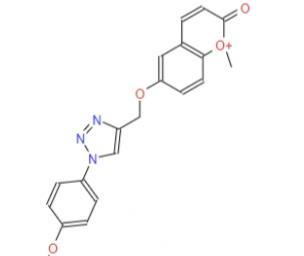
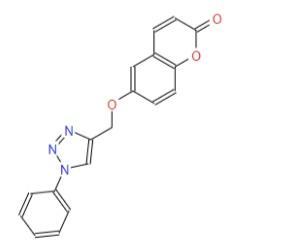
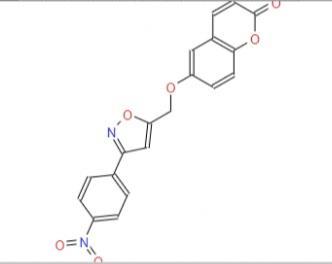
50 compounds that were previously assessed for their Hct-116 inhibition 21 compounds that were previously assessed for their Du-145 inhibition activities were used as the data set These compounds introducing a substituted Coumarin group at the R position of the parent scaffold resulted in significantly improved activities Coumarin.

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

****

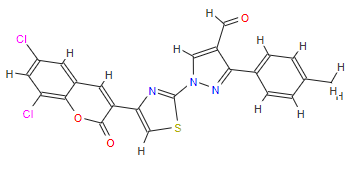
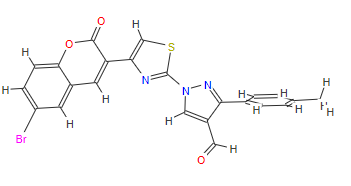
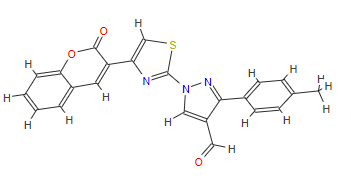
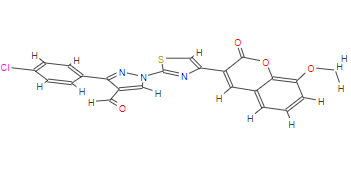
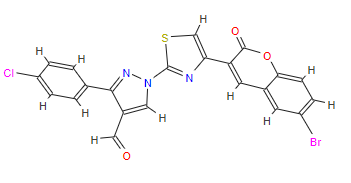
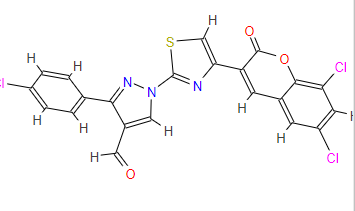
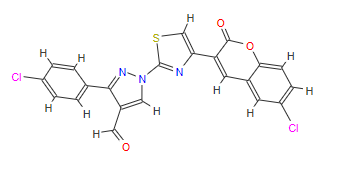
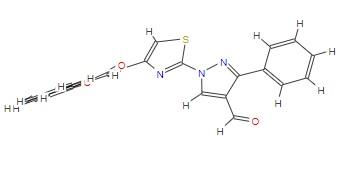
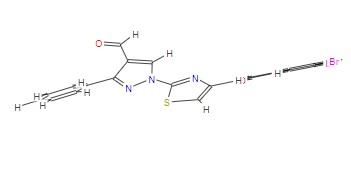
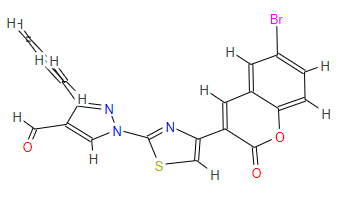
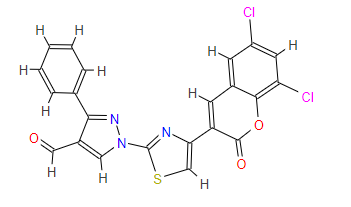
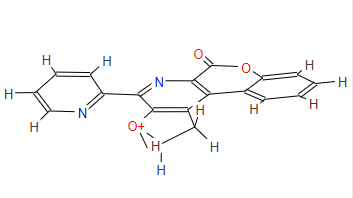
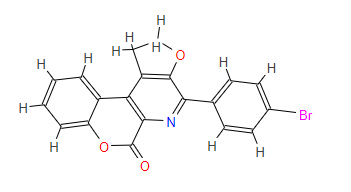
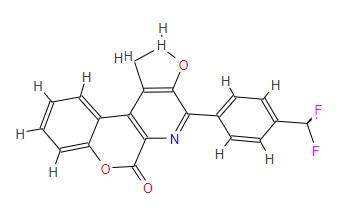
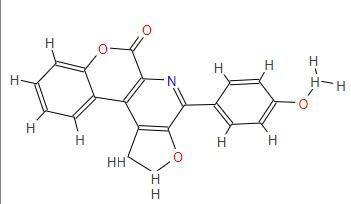
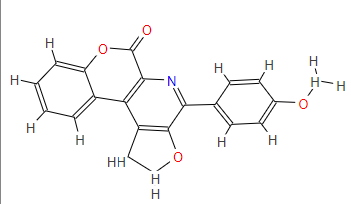
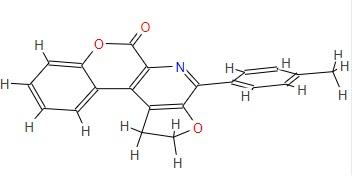
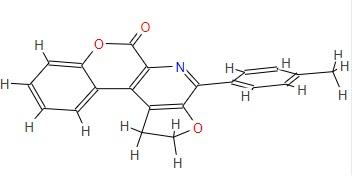
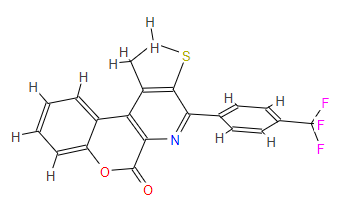
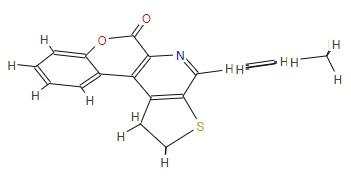
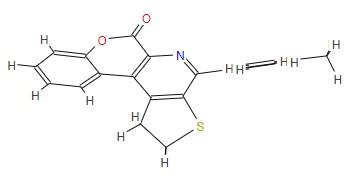
****

****

****

**(fig 1.4 1-50 coumarin compounds)**

**Coumarin compounds for Du-145:**

****

**(fig 1.5 1-21 Coumarin derivatives)**

**IC50:** Half-maximal inhibitory concentration is informative measure of a drug's efficacy. For inhibition of a biological process by half IC50 indicates how much drug is needed to providing a measure of potency of an antagonist drug in pharmacological research. it is the most widely used. a pharmacological compound are based on assays which is utilize by whole cell system is approach to determine IC50. it provides outstanding potency information, may not differentiate a compound's ability to inhibit specific interactions and results can depend on the experimental cell line used. IC50 values of individual ligand-receptor pairings.

1. **METHODOLOGY**

**QSAR**

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 principle underlying the formalism. For QSAR models validation, the strategies which are adopted are various.

1. Cross-validation or internal validation is used to measure data as it is being extracted. The model is resilient, with a larger q2 indicating that data extraction has less of an impact on the original model. A measure of model robustness is cross validation.

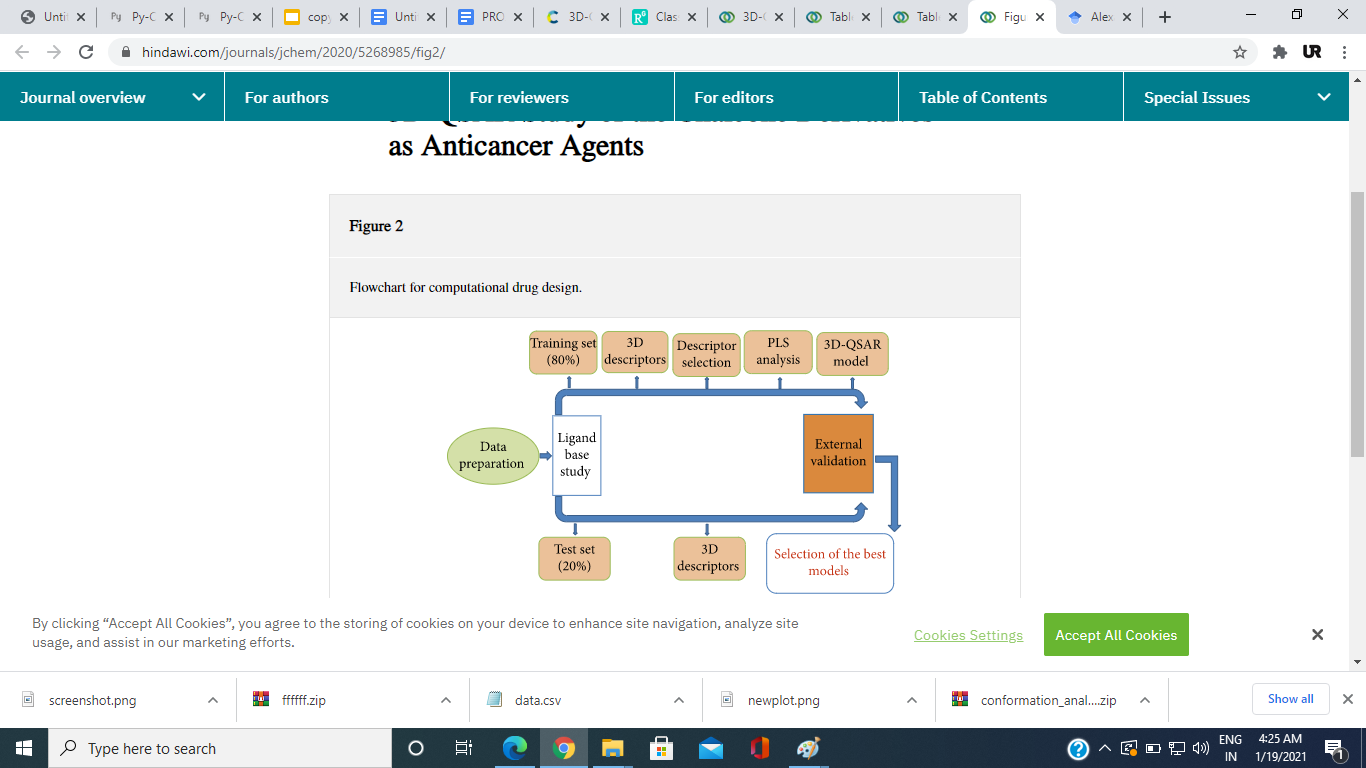
2. external validation by dividing the available data set into a training set for model creation and a prediction set for model predictive testing; 3. external validation by blind measuring the model on new external data.

4. To ensure that the modelling descriptors and the response measure Y-scrambling or data randomization do not have a chance correlation.

Statistical tools, input data fidelity, descriptor selection, and, most significantly, validation of the produced model are all critical factors in the success of a QSAR model. The reliability and relevance of a procedure are established through the use of validation processes and for a specific purpose. Validation of QSAR models should be based primarily on the models' applicability domain, robustness, and prediction performance.

There are a variety of validation approaches that might be problematic, such as determining if the training and test sets were chosen to optimise the predictive ability of the model being published in external validation, and leaving one-out cross-validation leads to an overestimation of predictive capacity.

In the validation of QSAR models, several variables such as training set size, methods of selecting training set compounds, and the influence of variable selection for training set models on determining the quality of prediction require consideration. The creation of new validation parameters is also critical for determining the quality of QSAR models.



**Fig 2.1 QSAR**

**3D-QSAR**

three-dimensional

QSAR is a method that employs force field computations and necessitates the usage of three-dimensional structures of a specified set of biological activities training set. It is concerned with the whole molecule rather than a single substituent, and it employs calculated potentials, experimental constants, and is concerned with the whole molecule rather than a single substituent, such as the Lennard-Jones potential of small molecules.

The experimental data, which is based on ligand-protein molecule superimposition software or crystallography, must be overlaid or aligned on the training set. Cramer et al. developed the first 3-D QSAR, known as Comparative Molecular Field Analysis (CoMFA). The means of partial least squares regression are used in 3d qsar to investigate the electrostatic and steric fields form of the molecule.

**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)**

|  |  |  |
| --- | --- | --- |
| Sr | Molecule | IC50 for HCT-116 |
| 1. | 0 | 83.9 |
| 2. | 1 | 65.8 |
| 3. | 2 | 33.5 |
| 4. | 3 | 56.6 |
| 5 | 4 | 44.7 |
| 6 | 5 | 30.6 |
| 7 | 6 | 38.8 |
| 8 | 7 | 32.6 |
| 9 | 8 | 53.7 |
| 10 | 9 | 41.8 |
| 11 | 11 | 46.3 |
| 12 | 12 | 42.1 |
| 13 | 13 | 50.6 |
| 14 | 14 | 54.7 |
| 15 | 15 | 46.2 |
| 16 | 16 | 55.9 |
| 17 | 17 | 51.5 |
| 18 | 18 | 42.1 |
| 19 | 19 | 55.8 |
| 20 | 20 | 48.2 |
| 21 | 21 | 42.2 |
| 22 | 22 | 51.5 |
| 23 | 23 | 37.8 |
| 24 | 24 | 26.9 |
| 25 | 25 | 57.3 |
| 26 | 26 | 55.8 |
| 27 | 27 | 44.3 |
| 28 | 28 | 1.93 |
| 29 | 29 | 2.28 |
| 30 | 30 | 3.88 |
| 31 | 31 | 3.13 |
| 32 | 32 | 5.55 |
| 33 | 33 | 11.98 |
| 34 | 34 | 9.98 |
| 35 | 35 | 13.33 |
| 36 | 36 | 9.55 |
| 37 | 37 | 16.66 |
| 38 | 38 | 76.7 |
| 39 | 39 | 11.5 |
| 40 | 40 | 12.0 |
| 41 | 41 | 25.9 |
| 42 | 42 | 32.2 |
| 43 | 43 | 75.8 |
| 44 | 44 | 32.0 |
| 45 | 45 | 30.8 |
| 46 | 46 | 16.1 |
| 47 | 47 | 46.4 |
| 48 | 48 | 12.9 |
| 49 | 49 | 14.5 |
| 50 | 50 | 15.9 |
|  |  |  |

**Table 2.1**

|  |  |  |
| --- | --- | --- |
| Sr | Molecule | IC50 for  DU-145 |
| 1. | 0 | 53.39 |
| 2. | 1 | 35.59 |
| 3. | 2 | 27.92 |
| 4. | 3 | 23.79 |
| 5 | 4 | 20.88 |
| 6 | 5 | 31.83 |
| 7 | 6 | 34.81 |
| 8 | 7 | 25.3 |
| 9 | 8 | 41.05 |
| 10 | 9 | 35.01 |
| 11 | 10 | 27.97 |
| 12 | 11 | 11.91 |
| 13 | 12 | 14.86 |
| 14 | 13 | 30.9 |
| 15 | 14 | 22.32 |
| 16 | 15 | 38.18 |
| 17 | 16 | 50.23 |
| 18 | 17 | 20.86 |
| 19 | 18 | 14.71 |
| 20 | 19 | 10.81 |
| 21 | 20 | 31.42 |

**Table 2.2**

**Molecular Modelling**

Draw Molecule

Energy Minimization

Alignment

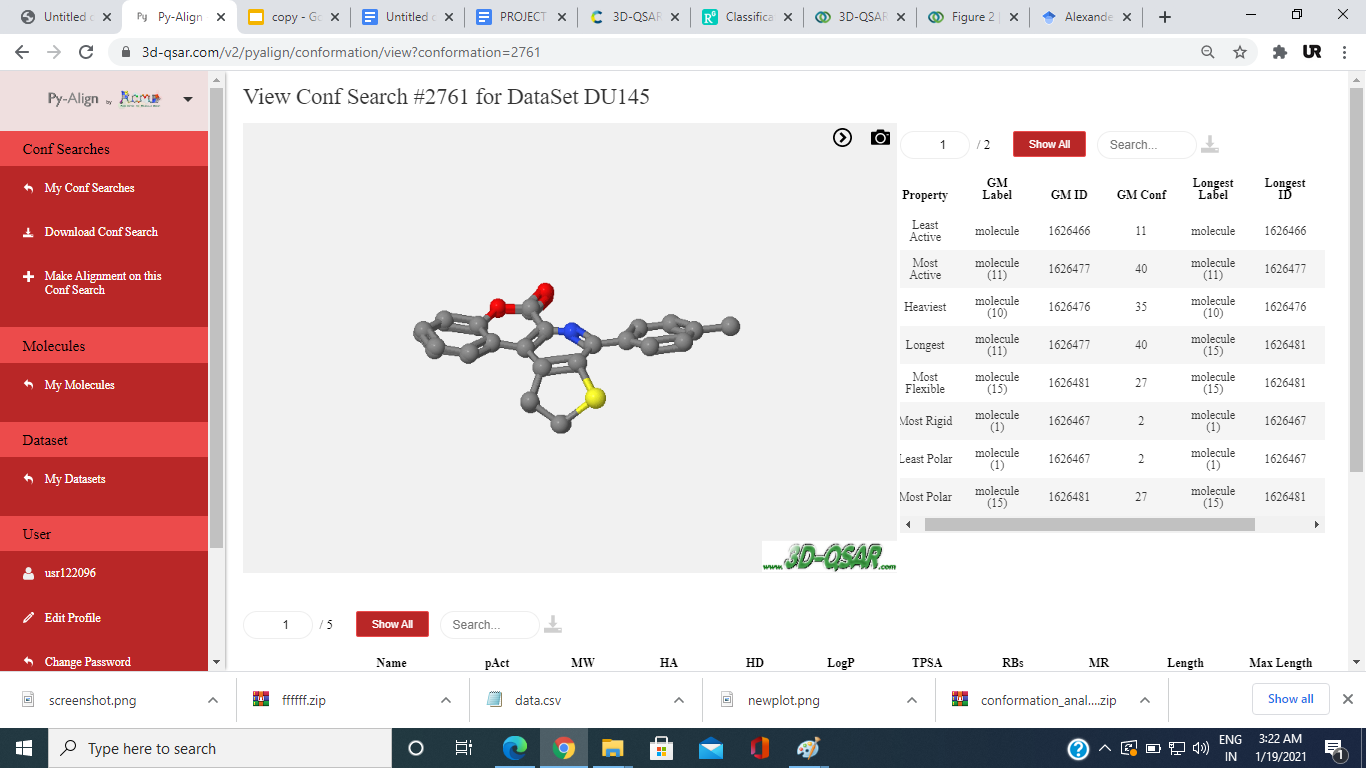
CoMFA

**Fig 2.2 molecular modelling**

**Molecular Alignment:**

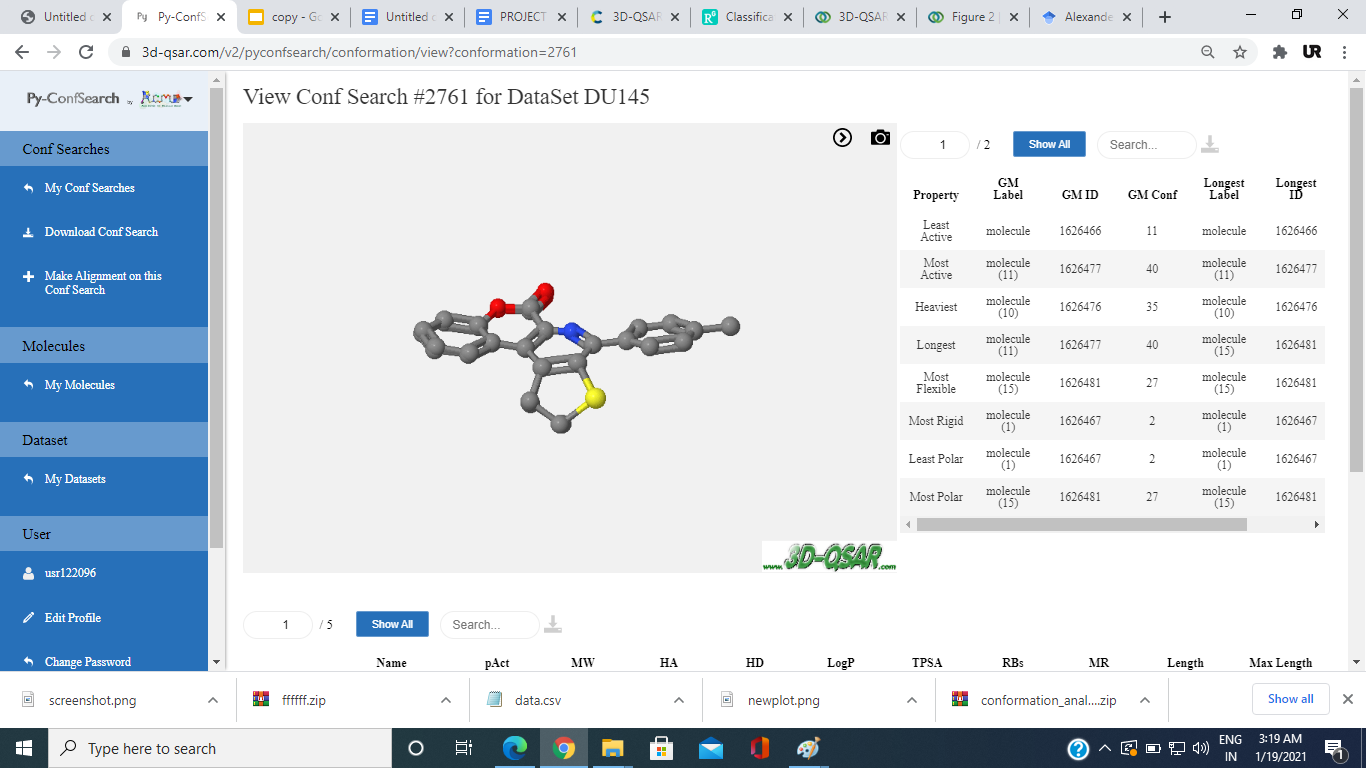
The experimental data, which is based on ligand-protein molecule superimposition software or crystallography, must be overlaid or aligned on the training set. Cramer et al. developed the first 3-D QSAR, known as Comparative Molecular Field Analysis (CoMFA). The means of partial least squares regression are used in 3d qsar to investigate the electrostatic and steric fields form of the molecule.

**For Hct-116:**

****

**Table 2.3**

**For Du-145:**



**Table 2.4**

**Alignment data**

**For HCT-116:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name Label | pAct | MW | HA | HD | LogP | TPSA | RBs | MR | Length | Max Length |
| 0 | 4.076 | 314.061 | 4 | 0 | 3.075 | 64.35 | 4 | 84.553 | 14.544 | 14.665 |
| 1 | 4.182 | 377.956 | 4 | 0 | 3.529 | 64.35 | 3 | 87.516 | 14.095 | 14.339 |
| 2 | 4.475 | 368.033 | 4 | 0 | 3.786 | 64.35 | 3 | 84.818 | 13.416 | 13.598 |
| 3 | 4.247 | 378.992 | 6 | 0 | 3.329 | 107.49 | 4 | 91.48 | 14.073 | 14.405 |
| 4 | 4.35 | 422.941 | 6 | 0 | 3.438 | 107.49 | 4 | 94.17 | 13.092 | 14.367 |
| 5 | 4.514 | 363.021 | 6 | 0 | 2.814 | 107.49 | 4 | 86.428 | 14.003 | 14.508 |
| 6 | 4.411 | 363.021 | 6 | 0 | 2.814 | 107.49 | 4 | 86.428 | 14.141 | 14.282 |
| 7 | 4.487 | 363.021 | 6 | 0 | 2.814 | 107.49 | 4 | 86.428 | 13.1 | 14.523 |
| 8 | 4.27 | 413.018 | 6 | 0 | 3.694 | 107.49 | 4 | 91.472 | 13.54 | 14.394 |
| 9 | 4.379 | 395.947 | 4 | 0 | 3.669 | 64.35 | 3 | 87.474 | 9.23 | 13.738 |
| 11 | 4.334 | 391.972 | 4 | 0 | 3.838 | 64.35 | 4 | 92.253 | 9.606 | 14.24 |
| 12 | 4.376 | 395.947 | 4 | 0 | 3.669 | 64.35 | 3 | 87.474 | 13.749 | 13.882 |
| 13 | 4.296 | 455.867 | 4 | 0 | 4.292 | 64.35 | 3 | 95.216 | 9.871 | 13.964 |
| 14 | 4.262 | 455.867 | 4 | 0 | 4.292 | 64.35 | 3 | 95.216 | 13.393 | 13.636 |
| 15 | 4.335 | 162.032 | 3 | 1 | 1.499 | 50.44 | 1 | 44.149 | 7.878 | 7.878 |
| 16 | 4.253 | 369.1 | 5 | 0 | 5.18 | 65.47 | 4 | 106.261 | 12.659 | 19.421 |
| 17 | 4.288 | 369.1 | 5 | 0 | 5.18 | 65.47 | 4 | 106.261 | 14.992 | 18.019 |
| 18 | 4.376 | 353.045 | 5 | 0 | 4.68 | 65.47 | 4 | 93.765 | 10.778 | 16.81 |
| 19 | 4.253 | 353.045 | 5 | 0 | 4.68 | 65.47 | 4 | 93.765 | 11.989 | 17.493 |
| 20 | 4.317 | 364.07 | 7 | 0 | 3.935 | 108.61 | 5 | 95.409 | 16.272 | 17.751 |
| 21 | 4.375 | 364.07 | 7 | 0 | 3.935 | 108.61 | 5 | 95.409 | 13.037 | 17.078 |
| 22 | 4.288 | 319.096 | 6 | 0 | 2.953 | 70.15 | 4 | 88.175 | 9.613 | 17.057 |
| 23 | 4.423 | 349.106 | 7 | 0 | 2.961 | 79.38 | 6 | 94.727 | 11.447 | 19.514 |
| 24 | 4.57 | 364.081 | 8 | 0 | 2.861 | 113.29 | 5 | 94.829 | 16.918 | 17.732 |
| 25 | 4.242 | 387.018 | 6 | 0 | 4.26 | 70.15 | 4 | 98.195 | 13.398 | 17.89 |
| 26 | 4.253 | 364.081 | 8 | 0 | 2.861 | 113.29 | 5 | 94.829 | 15.472 | 16.696 |
| 27 | 4.354 | 364.081 | 8 | 0 | 2.861 | 113.29 | 5 | 94.829 | 17.16 | 18.008 |
| 28 | 5.714 | 360.111 | 5 | 0 | 3.002 | 79.95 | 4 | 100.741 | 10.343 | 16.95 |
| 29 | 5.642 | 378.102 | 5 | 0 | 3.142 | 79.95 | 4 | 100.699 | 11.217 | 16.512 |
| 30 | 5.411 | 394.072 | 5 | 0 | 3.656 | 79.95 | 4 | 105.751 | 13.621 | 17.121 |
| 31 | 5.504 | 438.022 | 5 | 0 | 3.765 | 79.95 | 4 | 108.441 | 15.276 | 17.336 |
| 32 | 5.256 | 486.008 | 5 | 0 | 3.607 | 79.95 | 4 | 113.458 | 15.965 | 17.648 |
| 33 | 4.922 | 358.095 | 6 | 0 | 2.9 | 82.17 | 4 | 100.748 | 12.67 | 16.463 |
| 34 | 5.001 | 376.086 | 6 | 0 | 3.039 | 82.17 | 4 | 100.706 | 11.121 | 16.948 |
| 35 | 4.875 | 392.056 | 6 | 0 | 3.553 | 82.17 | 4 | 105.757 | 17.152 | 17.186 |
| 36 | 5.02 | 436.006 | 6 | 0 | 3.662 | 82.17 | 4 | 108.448 | 15.236 | 17.058 |
| 37 | 4.778 | 483.992 | 6 | 0 | 3.504 | 82.17 | 4 | 113.465 | 16.269 | 17.366 |
| 38 | 4.115 | 432.215 | 7 | 3 | 3.809 | 117.2 | 16 | 116.831 | 13.289 | 14.562 |
| 39 | 4.939 | 416.184 | 7 | 2 | 4.71 | 114.04 | 15 | 112.985 | 13.211 | 15.218 |
| 40 | 4.921 | 416.184 | 7 | 2 | 4.566 | 114.04 | 15 | 112.915 | 12.765 | 14.552 |
| 41 | 4.587 | 406.178 | 5 | 2 | 5.609 | 87.74 | 12 | 118.17 | 12.838 | 14.516 |
| 42 | 4.492 | 404.162 | 5 | 1 | 5.579 | 76.74 | 9 | 117.111 | 13.625 | 13.658 |
| 43 | 4.12 | 420.194 | 5 | 1 | 5.744 | 76.74 | 7 | 120.945 | 13.916 | 14.52 |
| 44 | 4.495 | 390.147 | 5 | 1 | 5.333 | 76.74 | 8 | 112.564 | 13.607 | 13.74 |
| 45 | 4.511 | 390.147 | 5 | 1 | 5.189 | 76.74 | 8 | 112.494 | 13.016 | 13.203 |
| 46 | 4.793 | 372.194 | 5 | 2 | 4.894 | 87.74 | 14 | 106.729 | 12.361 | 14.059 |
| 47 | 4.333 | 388.189 | 6 | 2 | 3.754 | 96.97 | 13 | 106.347 | 13.741 | 14.393 |
| 48 | 4.889 | 372.194 | 5 | 2 | 4.894 | 87.74 | 14 | 106.729 | 12.12 | 14.306 |
| 49 | 4.839 | 358.178 | 5 | 2 | 4.648 | 87.74 | 13 | 102.182 | 13.486 | 14.052 |
| 50 | 4.799 | 358.178 | 5 | 2 | 4.504 | 87.74 | 13 | 102.112 | 11.763 | 13.209 |

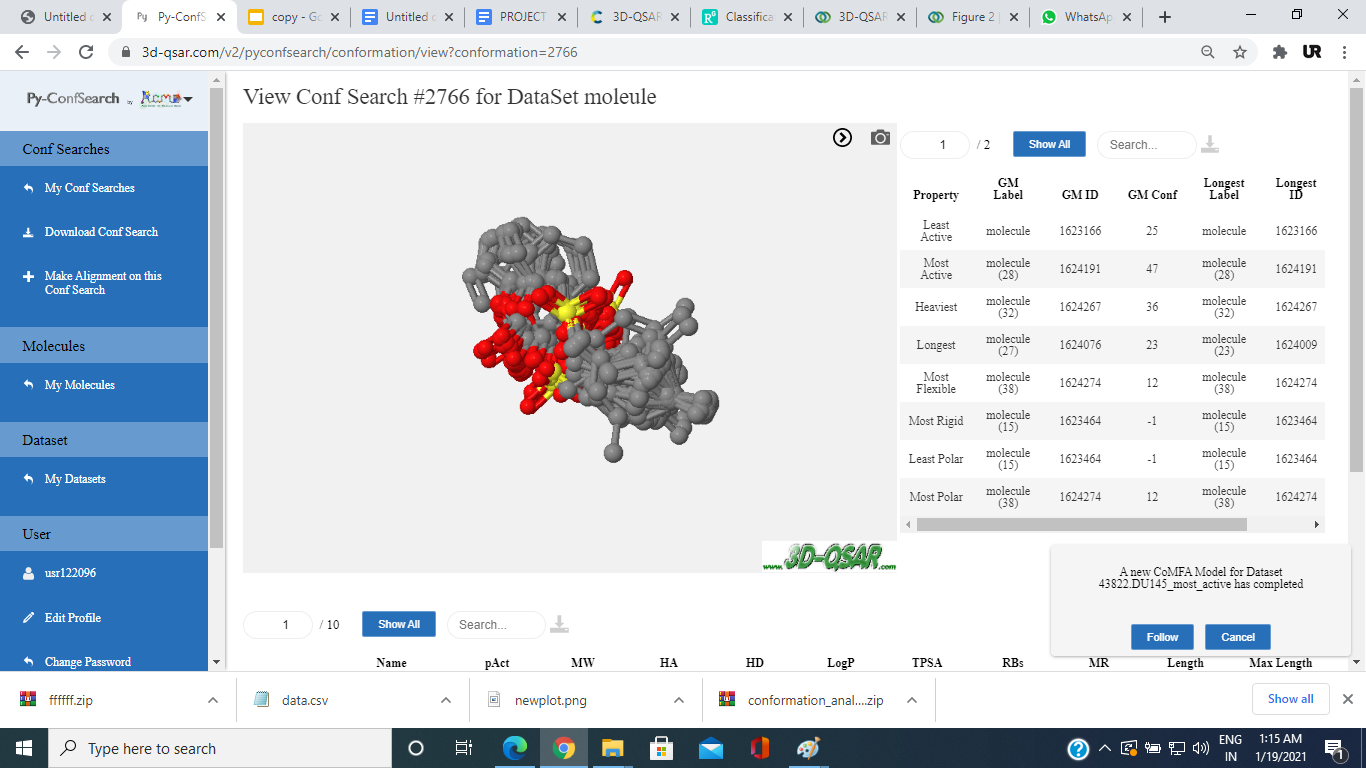
**Table 2.5**

**For DU-145:**

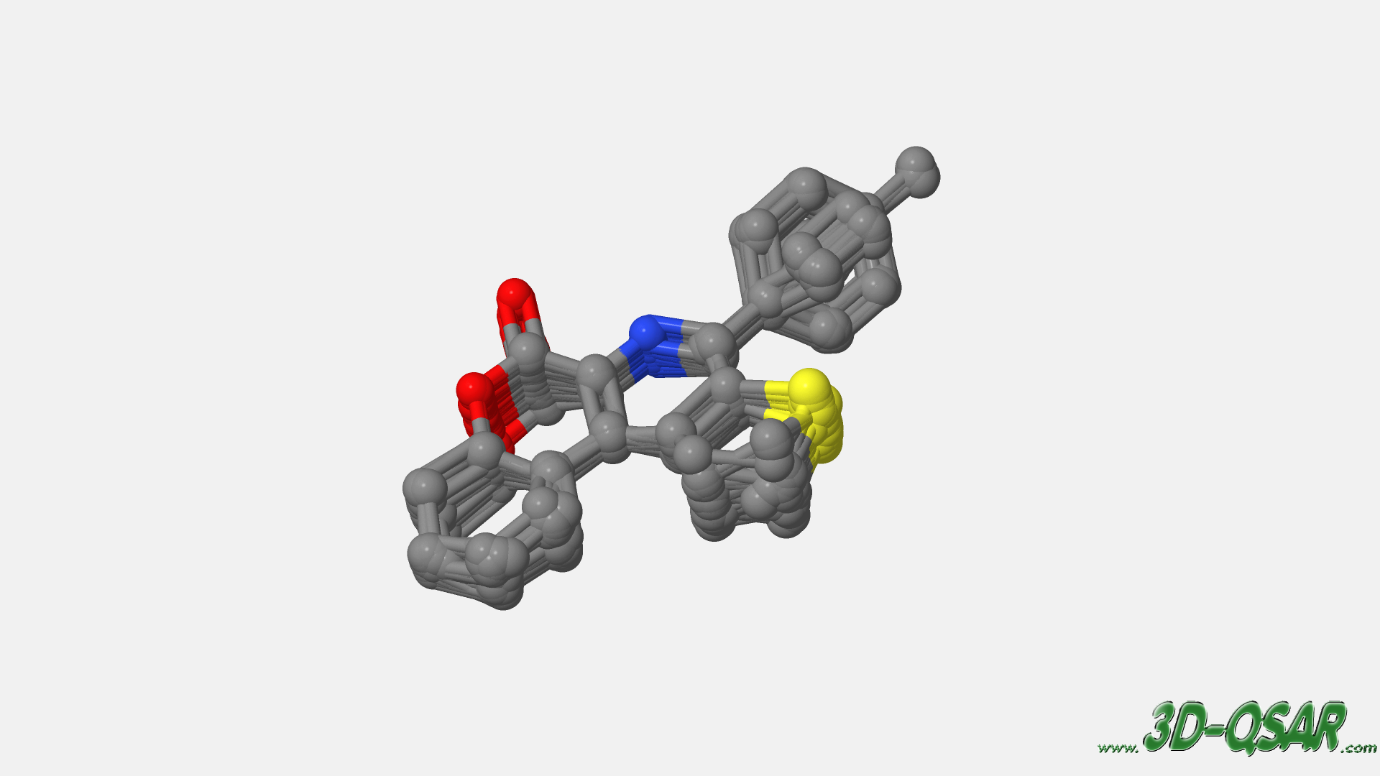
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Name Label | pAct | MW | HA | HD | LogP | TPSA | RBs | MR | Length | Max Length |
| 0 | 4.273 | 345.082 | 4 | 0 | 4.965 | 43.1 | 2 | 102.386 | 14.248 | 14.274 |
| 1 | 4.449 | 399.054 | 4 | 0 | 5.675 | 43.1 | 1 | 102.651 | 14.448 | 14.459 |
| 10 | 4.553 | 554.889 | 7 | 0 | 6.107 | 77.99 | 4 | 126.938 | 17.029 | 17.845 |
| 11 | 4.924 | 449.083 | 7 | 0 | 5.735 | 77.99 | 4 | 129.043 | 18.077 | 18.407 |
| 12 | 4.828 | 466.99 | 7 | 0 | 5.889 | 77.99 | 4 | 121.557 | 17.702 | 18.214 |
| 13 | 4.51 | 500.951 | 7 | 0 | 6.542 | 77.99 | 4 | 126.567 | 17.959 | 18.27 |
| 14 | 4.651 | 510.939 | 7 | 0 | 5.998 | 77.99 | 4 | 124.247 | 17.86 | 18.365 |
| 15 | 4.418 | 463.039 | 8 | 0 | 5.244 | 87.22 | 6 | 123.1 | 16.693 | 19.359 |
| 16 | 4.299 | 413.083 | 7 | 0 | 4.89 | 77.99 | 5 | 116.275 | 18.224 | 18.696 |
| 17 | 4.681 | 490.994 | 7 | 0 | 5.653 | 77.99 | 5 | 123.975 | 17.239 | 18.967 |
| 18 | 4.832 | 481.005 | 7 | 0 | 6.197 | 77.99 | 5 | 126.294 | 18.586 | 18.799 |
| 19 | 4.966 | 568.904 | 7 | 0 | 6.415 | 77.99 | 5 | 131.674 | 17.973 | 18.843 |
| 2 | 4.554 | 329.105 | 4 | 0 | 4.252 | 52.33 | 2 | 97.157 | 14.26 | 14.329 |
| 20 | 4.503 | 443.094 | 8 | 0 | 4.899 | 87.22 | 7 | 122.827 | 18.341 | 19.614 |
| 3 | 4.624 | 345.1 | 5 | 0 | 3.952 | 61.56 | 3 | 98.972 | 15.451 | 15.604 |
| 4 | 4.68 | 383.077 | 4 | 0 | 4.962 | 52.33 | 1 | 97.422 | 14.377 | 14.522 |
| 5 | 4.497 | 333.08 | 4 | 0 | 4.082 | 52.33 | 1 | 92.378 | 13.54 | 13.573 |
| 6 | 4.458 | 393 | 4 | 0 | 4.706 | 52.33 | 1 | 100.12 | 14.083 | 14.113 |
| 7 | 4.597 | 316.085 | 5 | 0 | 3.338 | 65.22 | 1 | 90.215 | 13.226 | 13.272 |
| 8 | 4.387 | 466.99 | 7 | 0 | 5.889 | 77.99 | 4 | 121.557 | 16.931 | 17.607 |
| 9 | 4.456 | 476.978 | 7 | 0 | 5.344 | 77.99 | 4 | 119.237 | 16.068 | 17.504 |

**Table 2.6**

**HCT-116:**



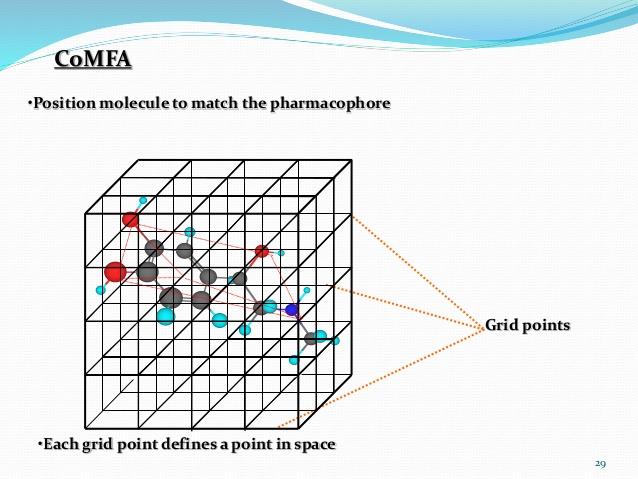
**Fig 2.3**

**DU-145:**

**Fig 2.4 (Aligned molecule structures)**

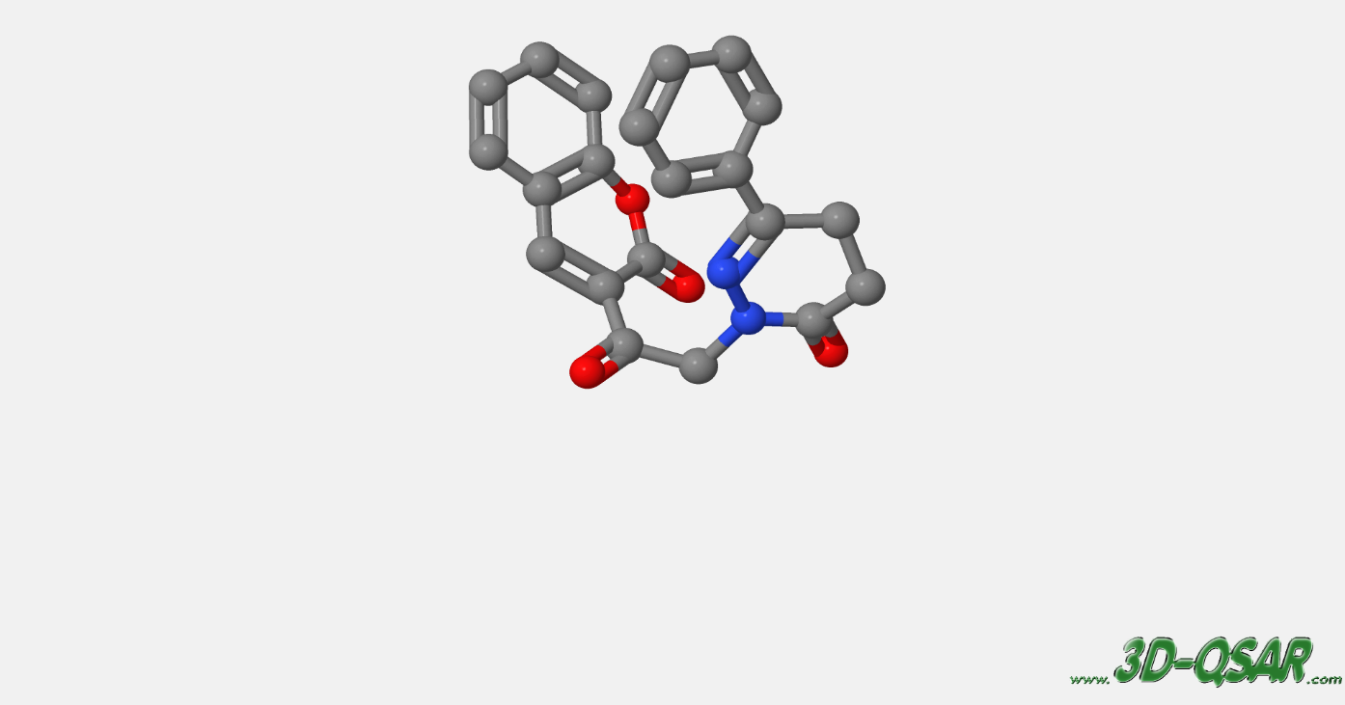
**CoMFA Studies**

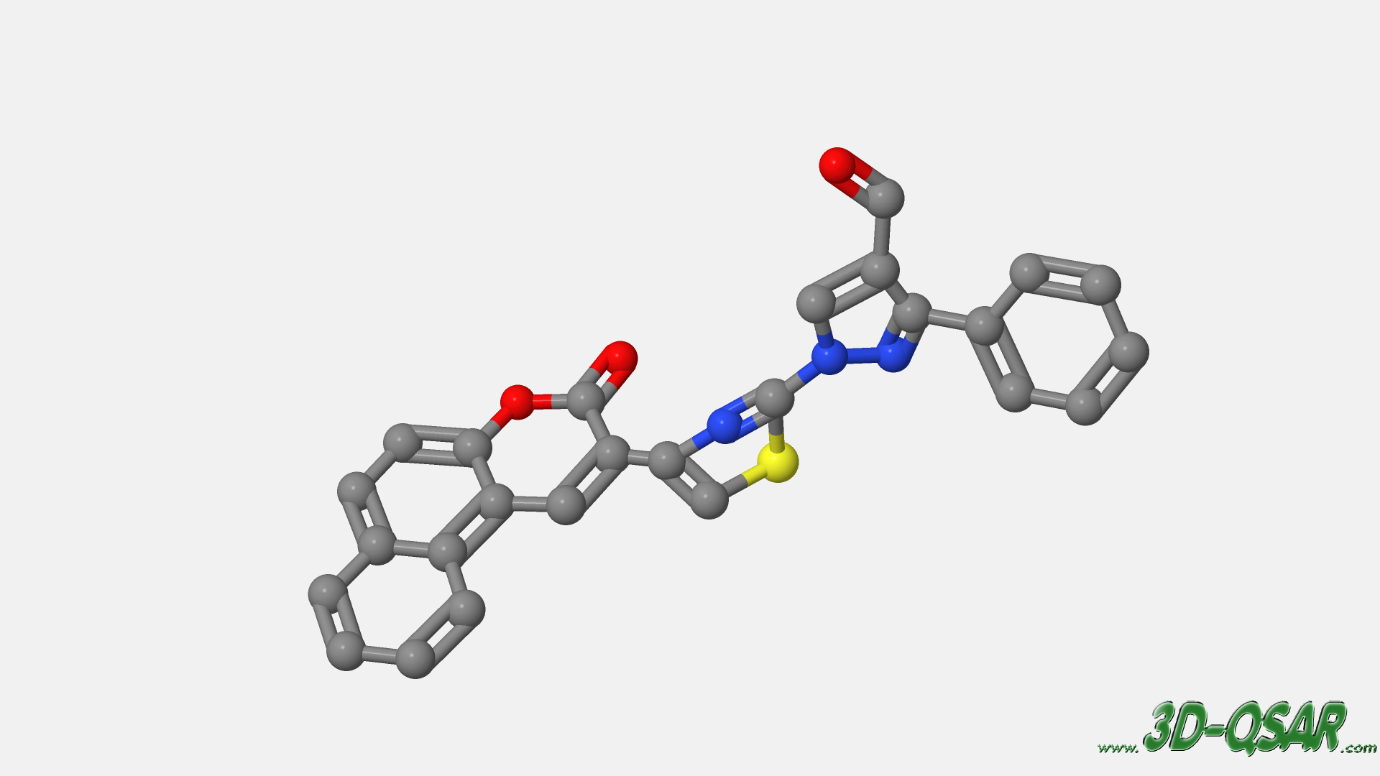
For CoMFA models generation, In a 3D cubic lattice the aligned molecules have been imported with a grid spacing of 2 Å and generated automatically. AMBER Force in General With a distance-dependent dielectric, Field uses an sp3 or c3 carbon probe atom with a Van der Waals radius of 1.52 and positive charge +1 at each lattice point. which used to generate electrostatic Coulombic potential fields and steric Lennard-Jones (6-12) potential field energies. The column filtering value has been set to 2.0 Kcal/mol as a standard setting, which is used to speed up the calculation of potentials and decrease noise.

**a prob atom is placed at each point of grid in turn.**

**Fig 2.5**

**Most Active Compound**

**HCT-116:****Fig 2.6**

**Du-145:**

**Fig 2.7**

##### **Partial Least Square Analysis:**

##### For the creation and internal validation of the CoMFA models, PLS regression analysis and cross-validation tests were utilised. PLS is a powerful method for calculating multilinear connections between dependent and independent variables. Model self-consistency can then be assessed using a cross-validation test, which is generated using PLS.

##### The explanatory characteristics of the CoMFA descriptors are employed as independent variables, while the IC50 (target properties) is employed as a dependent variable. The cross-validated correlation coefficient (Q2) and the optimum number of components were obtained from the first run of PLS utilising the full cross-validated leave-one-out LOO method analysis. To speed up the cross-validation calculation, the sample-distance partial least square (SAMPLS) approach was applied. The models were calculated in the second run of PLS using the non-cross-validation method with the optimal number of components. The models in this step were accessible using statistical metrics such as the squared correlation coefficient R2, and standard error of estimate.

**Partial Least Square Analysis** (PLS)

U=kt

Where k=constant

k=k1+k2+k3……..

j= numbers of PLS vector

BAi = a1S1i+a2S2i+a3S3i+….+amSmi + b1E1i+b2E2i+b3E3i+….+bmEmi



X1





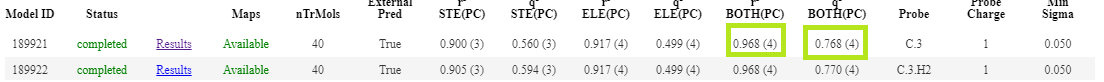
X2

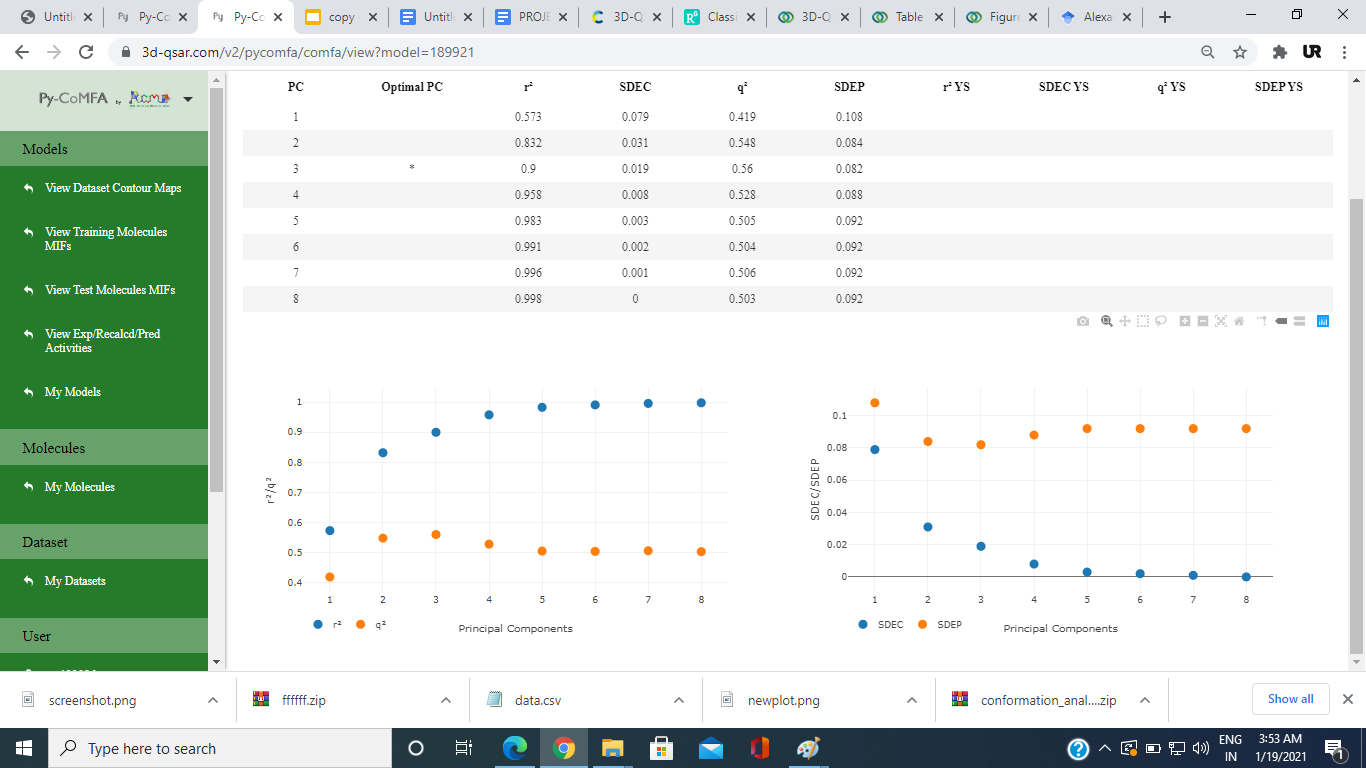


X3

**Fig 2.8**

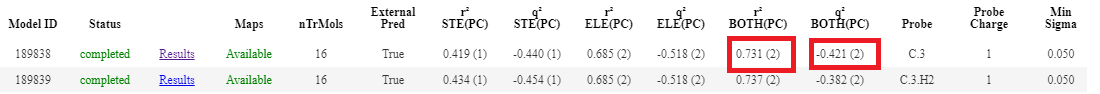
**For Hct-116:**

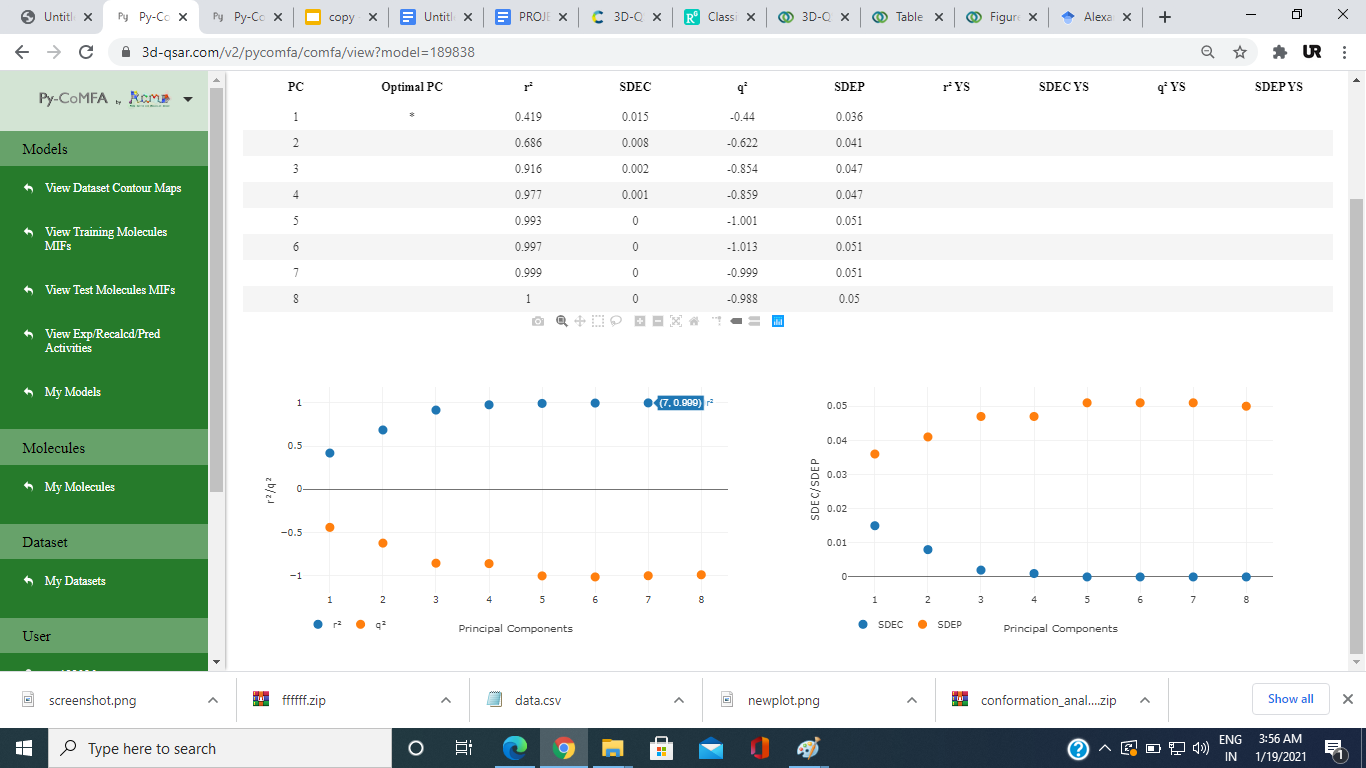
****

****

**Fig 2.9**

**For Du-145:**

****

****

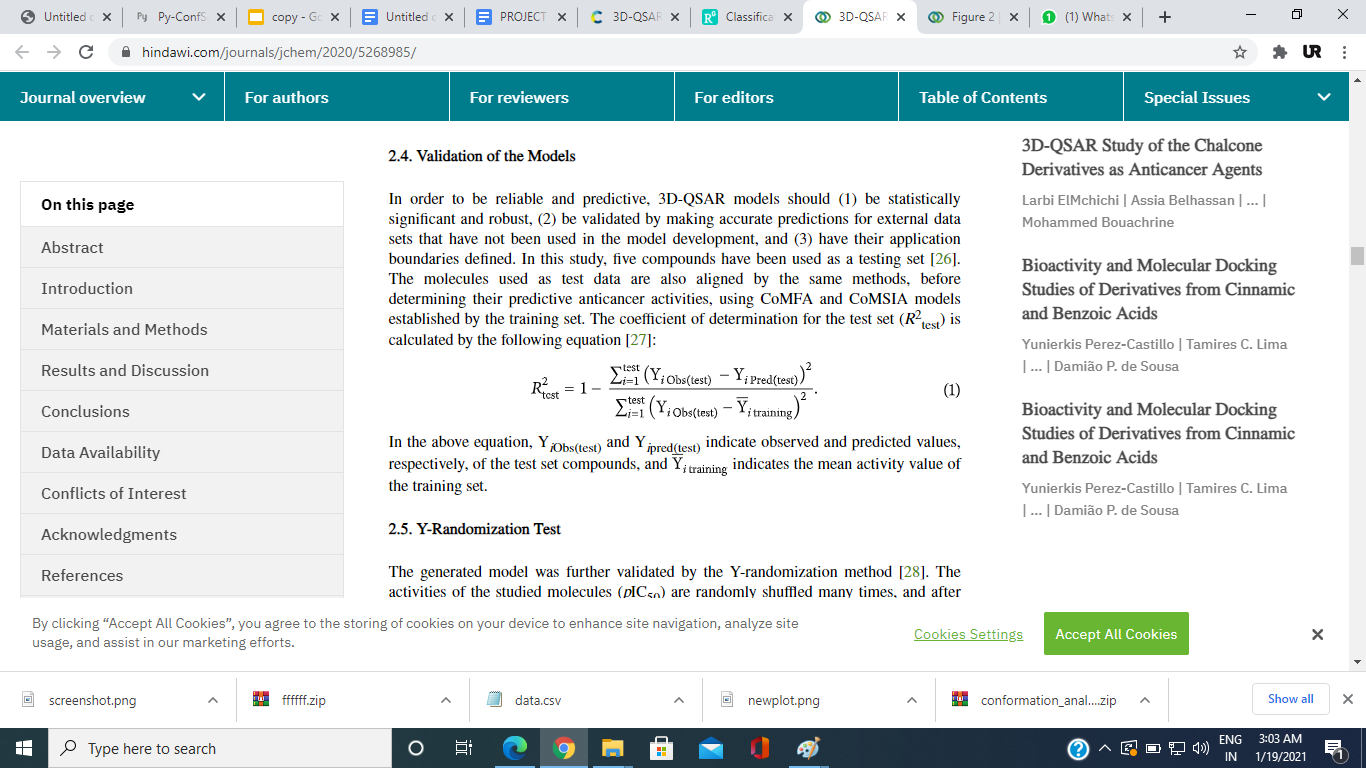
**` Fig 3.0**

##### 

##### **Validation of the Models**

##### 3D-QSAR models should be validated by providing correct predictions for external data sets that were not utilised in the model construction, as well as having their application boundaries established and being statistically significant and resilient. The testing set for this investigation consisted of 10 Hct-116 compounds and 5 Du-145 compounds. Before determining their predictive anticancer activities using CoMFA models built by the training set, the molecules chosen for test data are aligned by the same procedure that was used to align the training data. The equation R2test is used to calculate the coefficient of determination for the test set.

##### Yipred test and YiObs test show the observed and predicted values of the test set compounds, respectively, and denote the mean activity value of the training set in the above equation.



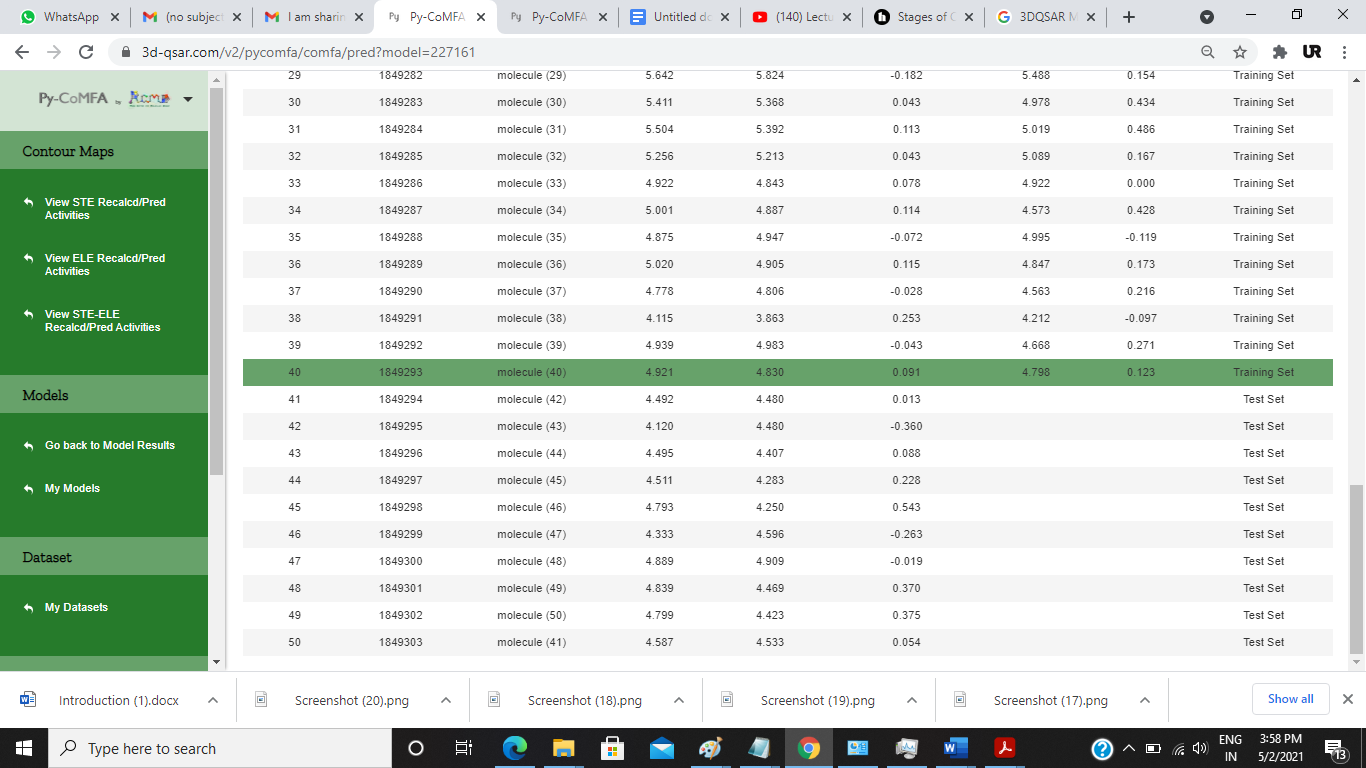
**Fig 3.1**

**Results**

**For Hct-116:**

|  |  |  |
| --- | --- | --- |
| Molecule Label | Expected pIC50 value | Predicted pIC50 value |
| 42 | 4.492 | 4.480 |
| 43 | 4.120 | 4.480 |
| 44 | 4.495 | 4.407 |
| 45 | 4.511 | 4.283 |
| 46 | 4.793 | 4.250 |
| 47 | 4.333 | 4.596 |
| 48 | 4.889 | 4.909 |
| 49 | 4.839 | 4.469 |
| 50 | 4.799 | 4.423 |
| 41 | 4.587 | 4.533 |
|  |  |  |

**Table 2.7**

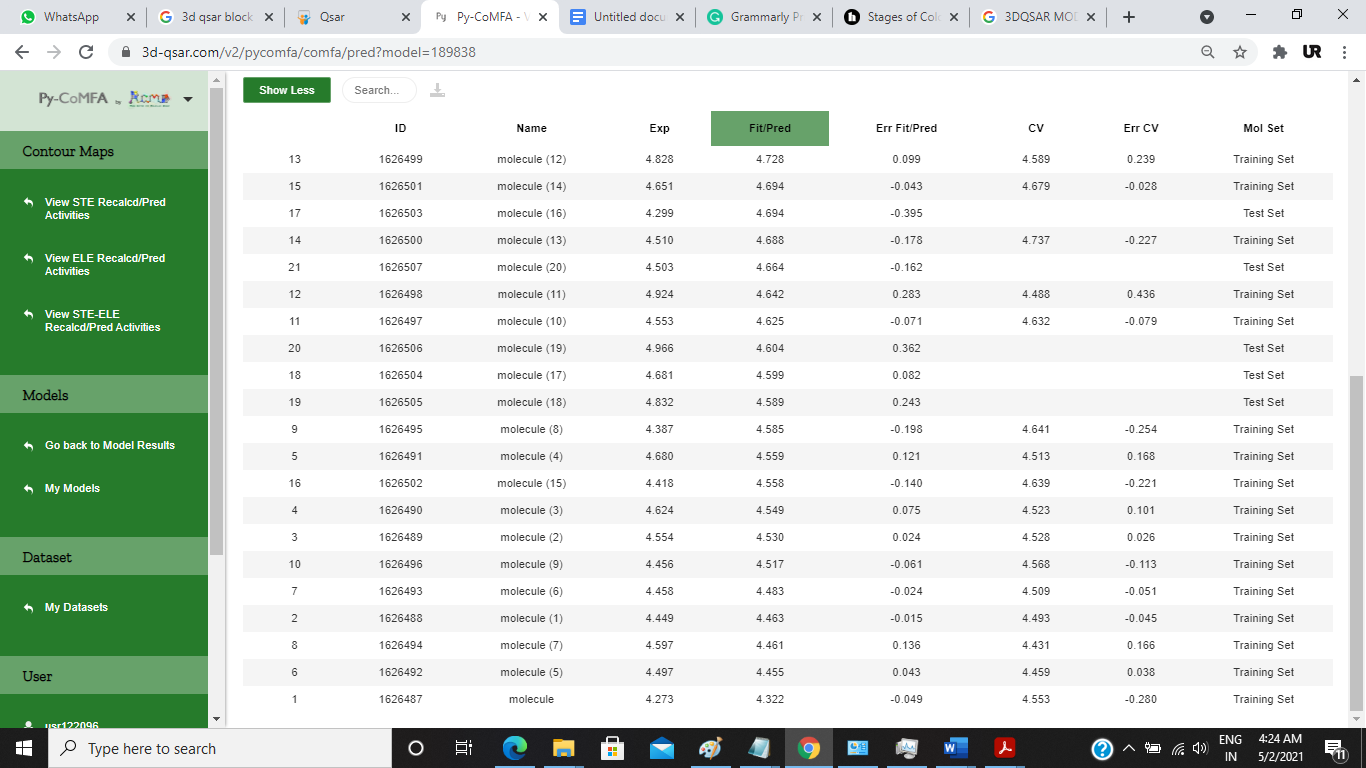
****

**Table 2.8**

**For Du-145:**

|  |  |  |
| --- | --- | --- |
| Molecule Label | Expected pIC50 value | Predicted pIC50 value |
| 16 | 4.299 | 4.694 |
| 13 | 4.510 | 4.688 |
| 20 | 4.503 | 4.664 |
| 11 | 4.924 | 4.642 |
| 10 | 4.553 | 4.625 |
| 19 | 4.966 | 4.604 |
| 17 | 4.681 | 4.599 |
| 18 | 4.832 | 4.589 |

**Table 2.9**

****

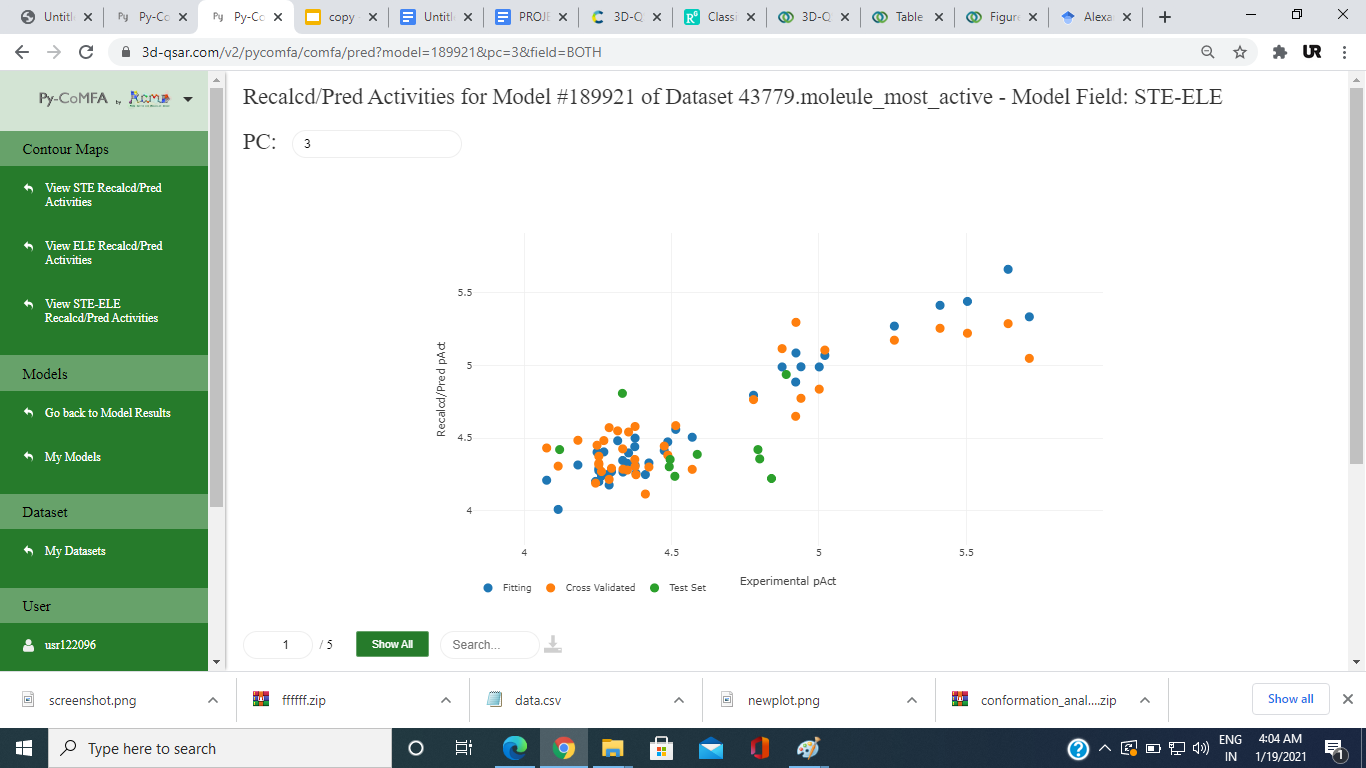
**Table 3.0**

##### **Y-Randomization Test**

To verify a model, the Y-randomization method is used. After each iteration, the IC50, which represents the activities of the investigated molecules, is randomly shuffled multiple times, and a new QSAR model is produced. Q2 and R2 values in the new QSAR models should be lower than in the old model. This strategy is used to remove the probability of coincidence. Because of structural chance correlation and redundancy, a satisfactory 3D-QSAR cannot be constructed for this data set anytime Q2 and R2 reach higher levels, indicating that an acceptable 3D-QSAR cannot be constructed for this data set.

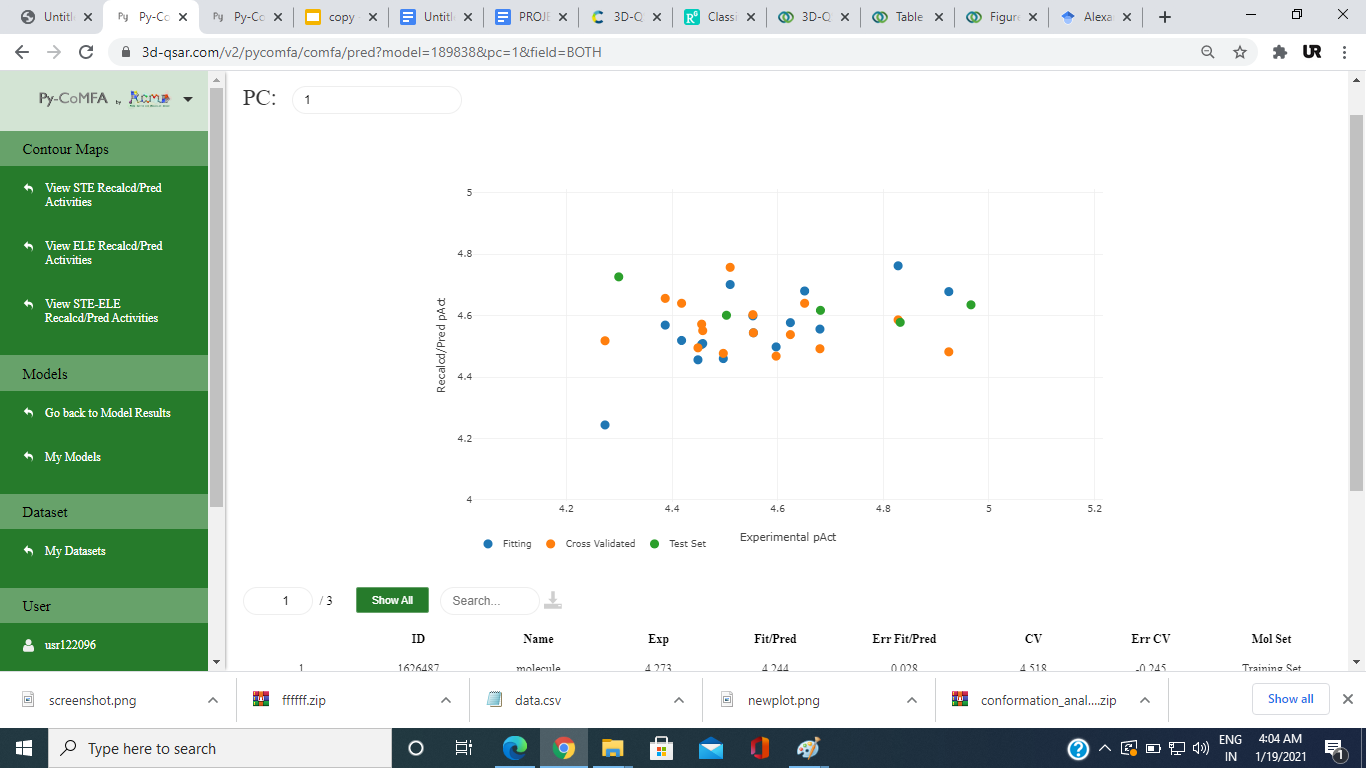
**Model Acceptability Criteria**

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

**Hct-116:**

**Fig 3.2**

**Du-145:**

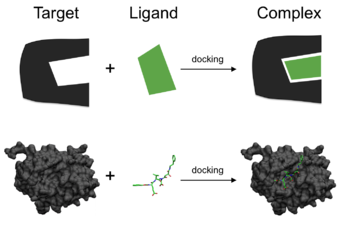
****

**Fig 3.3**

**Molecular Docking:**

Molecular Docking is used to predict the preferred orientation of one ligand when bound in an active site to form a stable complex. Docking is a method to find out binding modes of protein with ligands or inhibitors. It is able to generate a large number of possible structures. In molecular docking method we use to predict the intermolecular complex structure 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 can be thought of as a lock-and-key problem, in which one molecule is trying to figure out which lock will open up by finding the correct orientation of the key. When a key hole is present on the lock's surface, it determines which way the key will turn after it is inserted.

In Molecular docking, which can be described as an optimization problem with the main goal of finding the best-fit configurations of a ligand that binds with a particular protein, the protein can be represented as the lock and the ligand can be defined as the key. However, because both the ligand and the protein are adaptable, it's better to think of it as a hand-in-glove situation rather than a lock-and-key situation. To obtain an optimal conformation for both the ligand and the protein, as well as their relative orientation between protein molecules and ligand molecules. Molecular docking is a structure-based drug design process that predicts the binding-conformation of tiny molecules like ligands to the proper target binding site, such as protein molecules. Characterization of molecule binding behaviour is most important in drug development to explain underlying biochemical processes. The ligand and the protein modify their conformation throughout the docking process to obtain an overall best-fit, and this type of conformational change of molecules offers the best binding fit, which is referred to as induced-fit. Molecular docking is concerned with modelling the molecular recognition process simulate computationally.



**Fig 3.4**

**Shape complementarity:**

A collection of properties that make protein and ligand dockable via the solvent-accessible surface area receptor's molecular surface derived and described using shape complementarity approaches.The molecular surface of the ligand is expressed in terms of the corresponding surface description. Complementary surface descriptors on the molecular surface may be used in any of these applications. The ligand molecules and the shape matching description that are used for complementary posture of docking the target form the complementarity between the two surfaces. The main-chain atoms are turned to characterise the protein's hydrophobic uses. The application of a Fourier shape descriptor method is another approach. Whereas shape complementarity-based techniques are resilient and typically efficient, they do not develop models from dynamic and kinematic changes in the ligand-protein right configuration from docking. Pharmacophore-based techniques are also more suited to geometric descriptions of ligands, therefore geometric descriptions of ligands are used to figure out optimal binding.

* Molecules with lower binding energy are substantially more stable than those with higher binding energy.
* a molecular docking research was conducted to determine which of the coumarin derivatives has the greatest binding affinity to hct 116,
* The structure of hct-116 utilised in the research was found in the Protein Data Bank (PDB) under the code 5h8b.
* The ligand and receptor that had been created were displayed in Spartan 14 V 1.1.4 was used to convert the optimised structures of coumarin derivatives stored as SDF files to PDB files.
* Using AutoDock Vina, which is included in PyRx software, the produced ligands were docked with the constructed structures of hct 116.
* The docking data were assembled, processed, and displayed using Discovery Studio Visualizer, which included all compounds configured with regard to the targeted protein pdbqt files and their target protein pdbqt file.

**Collection of all natural and synthetic derivatives of Coumarin:**

The 3D-SDF files of all the reported coumarin both natural and synthetic were drawn and downloaded from open babel which have the ic50 values with respect to the derived cell lines after consulting several published journals.

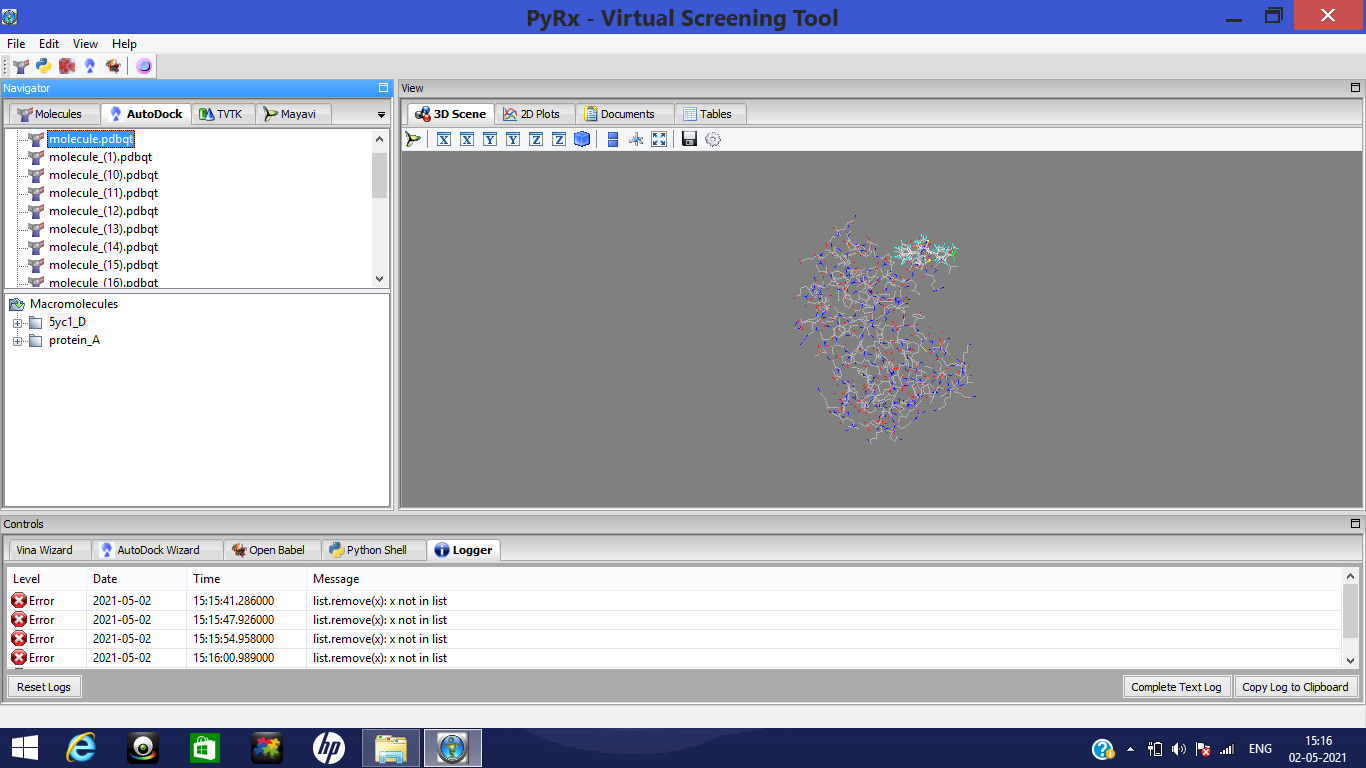
**Collection of proteins targets whose upregulation was responsible for (CRC) and (CAP)**:  All those proteins whose increased expression lead to development were discovered after going through related literature and their PDB structures were downloaded from Protein Data Bank (PDB)database.

**Molecular docking between core targets of CRC and CAP with the Coumarin derivatives using PyRx and AutoDock Vina:** CRC and CAP were docked with the corresponding to get their corresponding binding energies using autodock VINA. Along with that, the corresponding inhibition constants were also obtained. The above information regarding binding energy and inhibition constant was obtained from the dlg file after docking was completed. The number of runs was kept at 9 with a population size of 50 and 21.

**Simulation:**

The protein and ligand molecules are separated by a certain distance and conformation in this approach. All of these actions result in an overall energy cost for the ligand system in conformation space. As a result, after each movement, the system's total energy is determined. After a given number of conformations in its conformational space, the ligand molecules make their way into the active site of the protein. Docking simulation is a far more difficult procedure to simulate. Internal changes in the ligand's structure, such as torsion angle rotations, are included into these motions, as are rigid body transformations such as translations and rotations. Although form complementary approaches are more of an abstraction in reality, it more precisely represents than the prior way. The flexibility of the ligand molecule may be readily accommodated using this docking simulation approach, whereas clever approaches require employ complementarity approaches to integrate ligand flexibility.

Loading all coumarin derivatives in sdf/pdb and derived protein in pdb format.

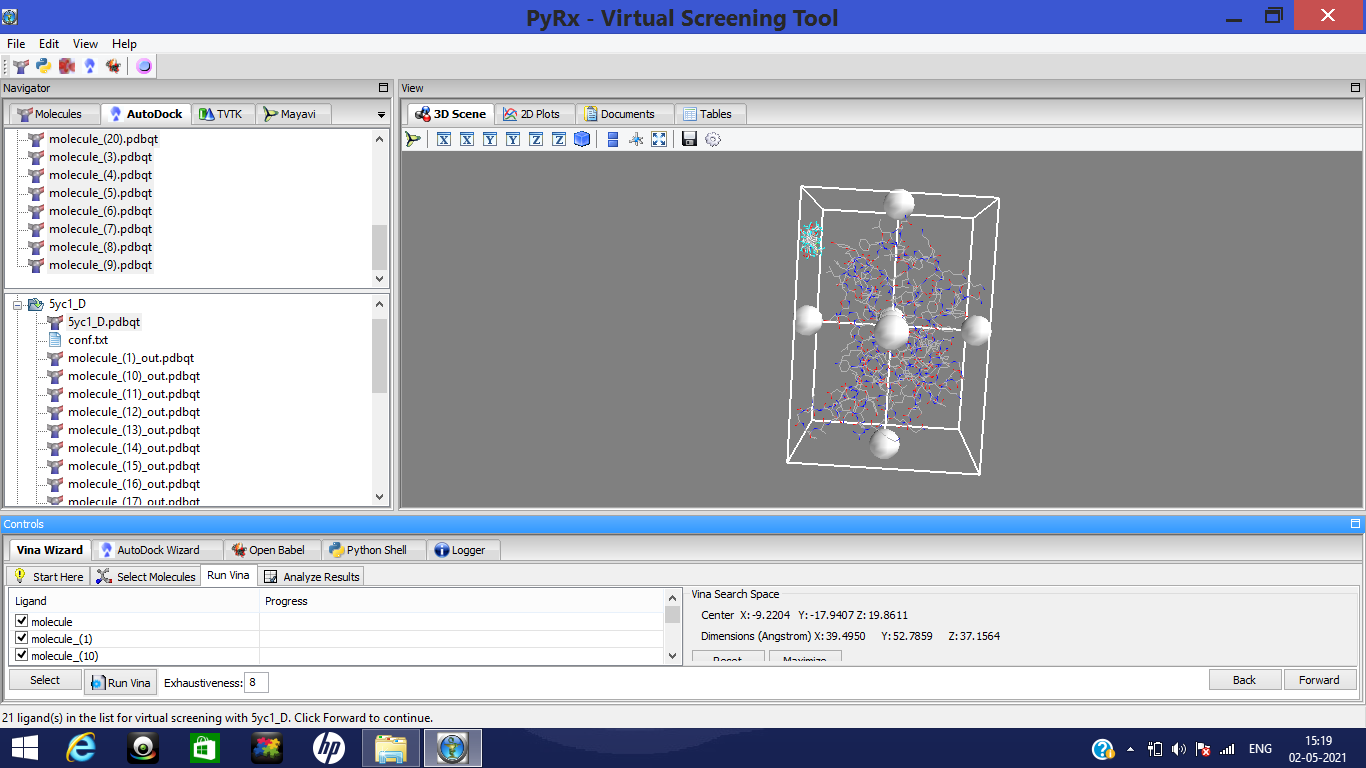


**Fig 3.5**

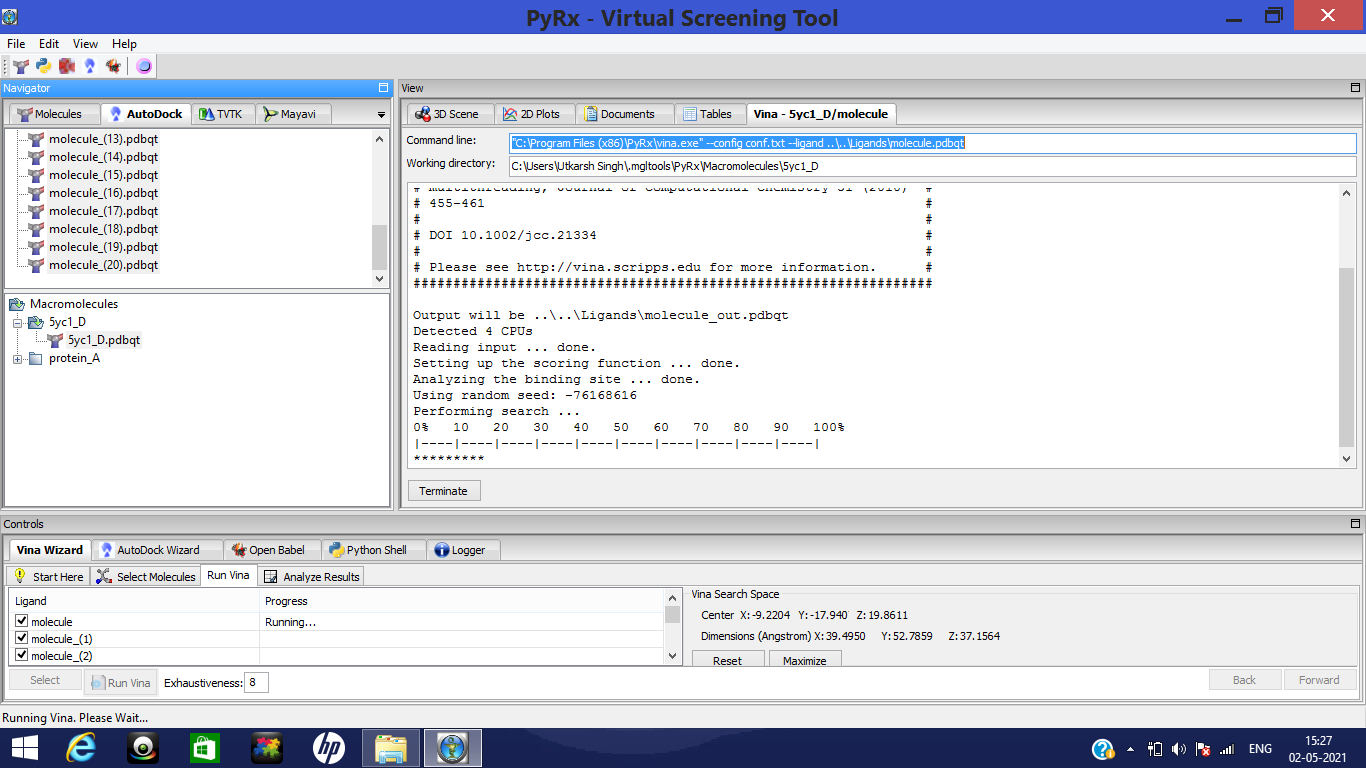
Convert ligands and protein to their pdbqt format to extract the unnecessary

Charges, bond and molecules.

cover the protein from grid box.

**Fig 3.6**

Run Autodock Vina to run the ligand in different configurations to form the most stable configuration which has lowest energy.



**Fig 3.7**

**Determination of the active binding sites of the target protein with the Coumarin derivatives using Discovery Studio Visualizer**: For this purpose, after running the auto doc vina a log file will be generated in PyRx which will show the Then the Conformation which showed the minimum binding energy and it was saved as a protein-ligand complex in pdbqt format. Following that, after opening the Discovery Studio Visualizer, the same protein-ligand complex was opened in the file section and showed a 2D diagram in order to get the view of amino acids of the proteins interacting with the ligand.

**3. RESULTS AND DISCUSSION**

**Docking of the above set of proteins with all the Coumarin derivatives through Autodock Vina:**5H8B were the targets of following Coumarin derivatives with which the docking was performed Along with 5H8B docking of the all the Coumarin derivatives was performed with and 5YC1.

|  |  |
| --- | --- |
| Sr | Coumarin derivatives  directed against the  protein (HCT-116)  (Molecule  label) |
| 1. | 0 |
| 2. | 1 |
| 3. | 2 |
| 4. | 3 |
| 5 | 4 |
| 6 | 5 |
| 7 | 6 |
| 8 | 7 |
| 9 | 8 |
| 10 | 9 |
| 11 | 1 |
| 12 | 12 |
| 13 | 13 |
| 14 | 14 |
| 15 | 15 |
| 16 | 16 |
| 17 | 17 |
| 18 | 18 |
| 19 | 19 |
| 20 | 20 |
| 21 | 21 |
| 22 | 22 |
| 23 | 23 |
| 24 | 24 |
| 25 | 25 |
| 26 | 26 |
| 27 | 27 |
| 28 | 28 |
| 29 | 29 |
| 30 | 30 |
| 31 | 31 |
| 32 | 32 |
| 33 | 33 |
| 34 | 34 |
| 35 | 35 |
| 36 | 36 |
| 37 | 37 |
| 38 | 38 |
| 39 | 39 |
| 40 | 40 |
| 41 | 41 |
| 42 | 42 |
| 43 | 43 |
| 44 | 44 |
| 45 | 45 |
| 46 | 46 |
| 47 | 47 |
| 48 | 48 |
| 49 | 49 |
| 50 | 50 |
|  |  |

**Table 31**

|  |  |
| --- | --- |
| Sr | Coumarin derivatives  directed against the  protein (DU-145) |
| 1. | Molecule |
| 2. | 1 |
| 3. | 2 |
| 4. | 3 |
| 5 | 4 |
| 6 | 5 |
| 7 | 6 |
| 8 | 7 |
| 9 | 8 |
| 10 | 9 |
| 11 | 11 |
| 12 | 12 |
| 13 | 13 |
| 14 | 14 |
| 15 | 15 |
| 16 | 16 |
| 17 | 17 |
| 18 | 18 |
| 19 | 19 |
| 20 | 20 |
| 21 | 21 |

**Table 3.2**

Following docking results were obtained after docking of**:** 5H8B and 5YC1 with their respective coumarin. Vorinostate was used as a reference inhibitor molecule for both HCT-116 and DU-145.

**For HCT-116: 5H8B**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr | Molecule | Binding Energy (in kcal/mol) |  |
| 1. | 16 | -8.3 |  |
| 2. | 36 | -8.0 |  |
| 3. | 35 | -8.0 |  |
| 4. | 29 | -8.0 |  |
| 5 | 31 | -7.9 |  |
| 6 | 32 | -7.9 |  |
| 7 | 35 | -7.8 |  |
| 8 | 17 | -7.8 |  |
| 9 | 34 | -7.7 |  |
| 10 | 33 | -7.4 | 0.0 |

**Table 3.3**

**For DU-145: 5YC1**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr | Molecule Label | Binding Energy (in kcal/mol) |  |
| 1. | 11 | -8.7 |  |
| 2. | 18 | -8.3 |  |
| 3. | 10 | -8.2 |  |
| 4. | 8 | -8.2 |  |
| 5 | 19 | -8.1 |  |
| 6 | 16 | -8.1 |  |
| 7 | 13 | -8.1 |  |
| 8 | 7 | -8.1 |  |
| 9 | 20 | -8.0 |  |

**Table 3.4**

Order of binding of Coumarin derivatives with respect to binding energy:

(i)5H8B: molecule (16) < molecule (36)< molecule (35)< molecule (29)

(ii)5YC1: molecule (11) < molecule (18)< molecule (10)< molecule(8)

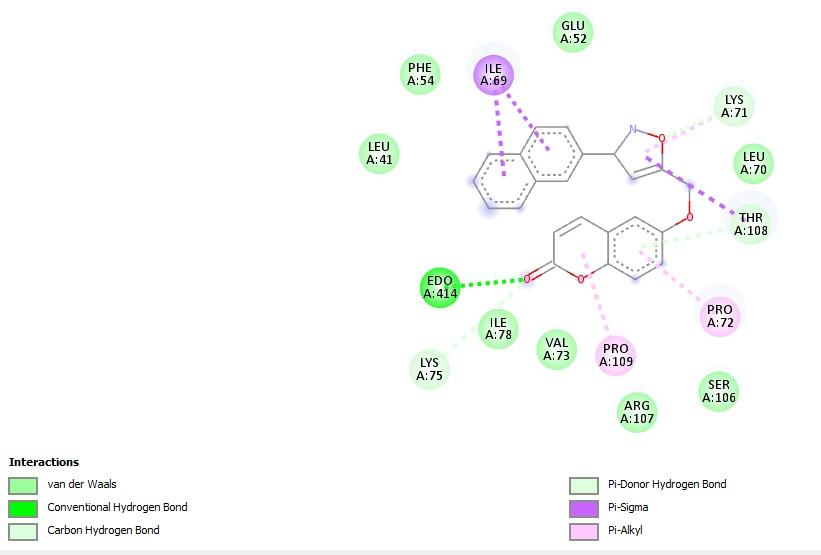
The docking results suggested that on the basis of binding energy molecule (16) showed the highest binding affinity to **5H8B**.

The docking results suggested that on the basis of binding energy molecule (11) showed the highest binding affinity to **5YC1**.

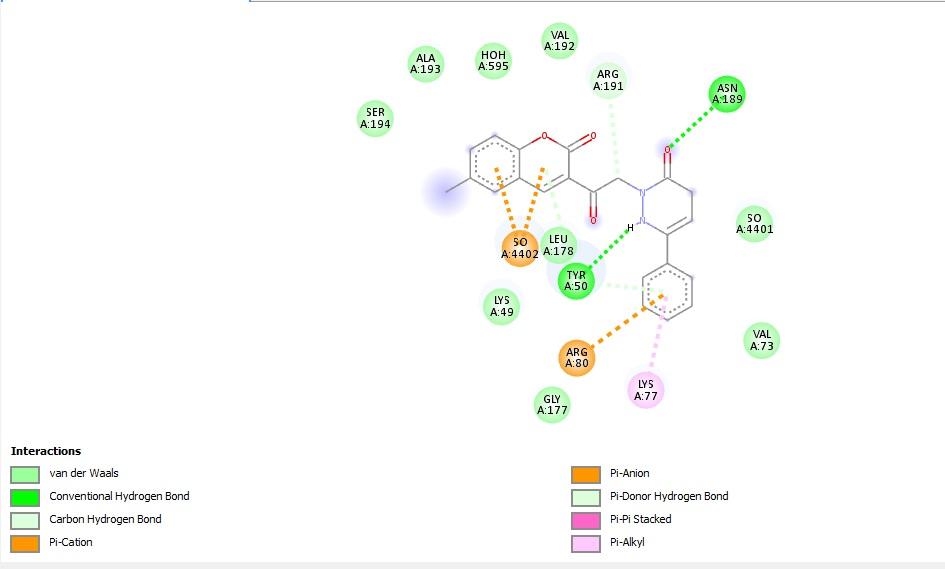
**Amino acid binding site visualization through Discovery Studio**

2D protein-ligand interaction diagrams was obtained for and with all the Comarine Derivatives with which they were docked. Following is the 2D interaction diagram of hct-116 with molecular and (i.e. the Coumarin derivatives with which they showed the best docking confirmation) respectively.

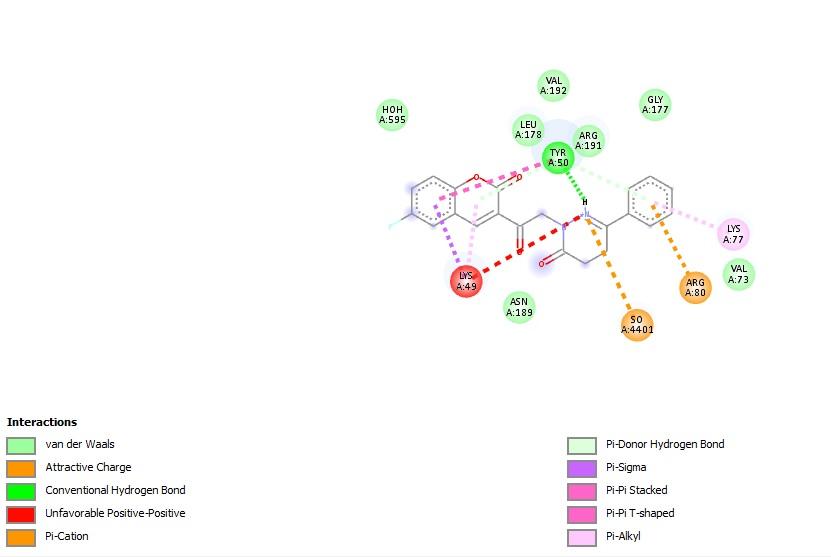
**For HCT-116**



**Fig 3.8 Molecule(16) isoxazolyl**

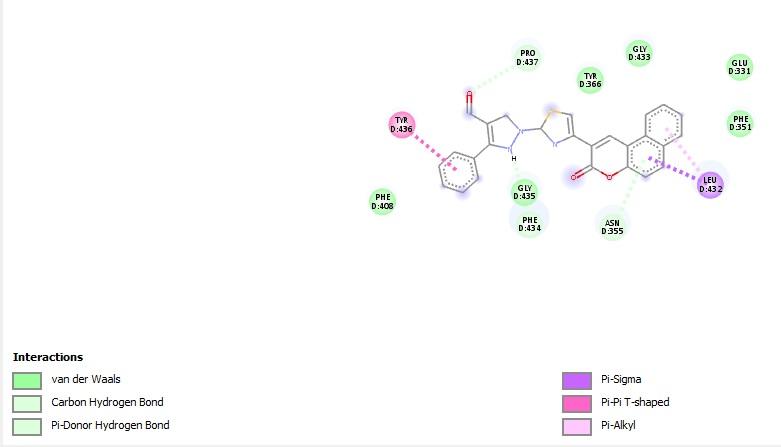


**Fig 3.9 Molecule(35)**

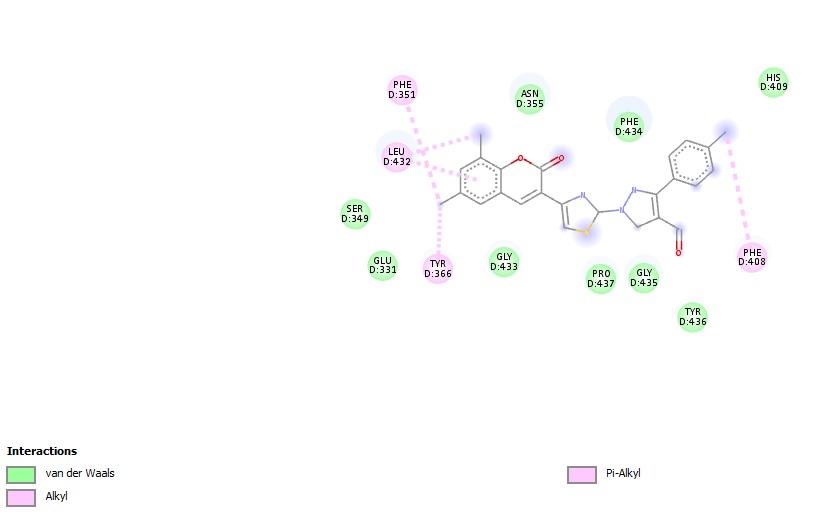


**Fig 4.0 Molecule(29)**

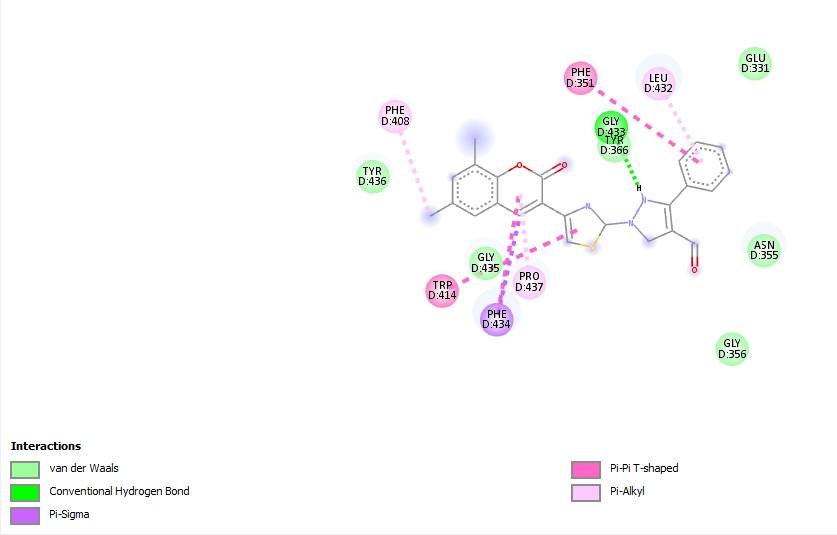
**For DU-145**



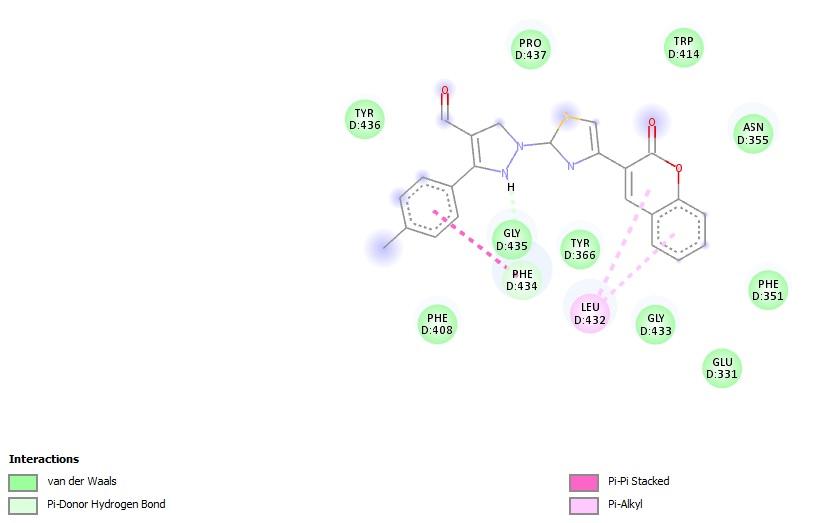
**Fig 4.1 Molecule(11) thiazolyl-3-aryl-pyrazole-4-carbaldehydes**



**Fig 4.2 Molecule(18)**



**Fig 4.3 Molecule(8)**



**Fig 4.4 Molecule(16)**

# 4. CONCLUSION

##### **3.1 Statistical Results:**

The acquired independent variables from the training set were submitted to cross-validated PLS analysis analysis to determine 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 to achieve (Hct-116)R2 = 0.968 and (D-145)R2=0.731 for the CoMFA model.

**3.2 Molecular Docking Result:**

**For Hct-116:**

* The model with 2D and 3D descriptors is superior in quality and has a good connection with anti-cancer activity.
* A better model for predicting the anti-cancer actions of these drugs is created by combining 2D and 3D descriptors.
* The validation parameters recommended value for a well-known acceptable QSAR model was satisfied by the QSAR produced model, which fulfilled the minimal requirements for validation parameters.
* Nearly all coumarin compounds have the ability to inhibit hct 116 and Du-145, according to molecular docking study.
* However, chemical molecules (16) (isoxazolyl) and molecule (11) (thiazolyl-3-aryl-pyrazole-4-carbaldehydes) show significant bind scores for Hct-116 and Du-145, respectively.
* Molecular docking studies are a useful tool for designing structural bases, whereas. A helpful strategy for ligand base design is provided by the QSAR produced model.

Based on the results of the in-silico research completed thus far, Molecule(16) may prove to be the best inhibitor for regulating the overexpression of Histone deacetylase 5H8B, while Molecule(11) may prove to be the best inhibitor for regulating the overexpression of 5YC1.

**Data Availability:**

The manuscript have all the required data.

**5. REFERNCES**

* 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
* <https://en.wikipedia.org/wiki/Coumarin#:~:text=Coumarin%20is%20a%20colorless%20crystalline,a%20chemical%20defense%20against%20predators>.

* http://www.xenograft.net/du145-xenograft-model/
* <http://www.xenograft.net/hct116-xenograft-model/>