Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/pt4wp993

QSAR STUDIES ON BICALUTAMIDE DRUG FOR PROSTATE CANCER TREATMENT

Ayesha Hassan[1*], Sadia Afsheen[2], Danish Abbas[3], Bilal Atiq[3], Atiqa Saleem[4], Dawood Zia[5], Muhammad Irfan[1], Zain Ul Abideen[1], Hamza Mobeen[1], Muhammad Abu-Bakar Sidique[1]

[1*]Department of Biosciences, COMSATS University Islamabad, Park Rd, Islamabad Capital Territory 45550, Pakistan
[2]International Center for Chemical and Biological Sciences (ICCBS), University of Karachi, 75270, Pakistan
[3]Department of D Pharmacy, Hamdard Institute of Pharmaceutical Sciences HUIC, Islamabad, Pakistan
[4]Biomedical Engineering Department HITEC University Taxila, Pakistan
[5]Department of Pharmacy, University of Lahore, Pakistan

***Corresponding Author:** Ayesha Hassan
*Email: ayeshahasan3546@gmail.com

Abstract

Prostate Cancer (PC) is a dangerous and deadliest type of cancer it is the main reason of male death globally. The development of new and stronger anti-prostate cancer composites is a constant requirement. It can also be associated with alterations in AR functions. Indeed, androgen blockade by drugs that prevent the production of androgens and/or block the action of the AR inhibits prostate cancer growth. For its treatment, a very effective drug is used named as Bicalutamide. This drug is used to block the androgen action. It works by preventing testosterone from binding to the androgen receptors in prostate cancer cells. As a means to improve the effectiveness of the Bicalutamide drug, and in order to exploit the well-established potential of the fluorine atom in enhancing the pharmacological properties and drug-like physicochemical characteristics of candidate compounds, a wide array of diverse new structures has been designed and synthesized, through the introduction of fluoro-, trifluoromethyl- and trifluoromethoxy groups in diverse positions of both aromatic rings of the parent scaffolds. We have employed 2D and 3D QSAR approaches to identify the best descriptor to design better active compounds. In the 2D QSAR method, different types of descriptors existed from which some were eliminated due to their same or zero value. Overall 192 descriptors from which 142 were eliminated and 50 were used for further analysis. In LNCAP and VCaP cell lines, the correlation coefficient (R2) values 0.99 were after pruning. In 3D QSAR, The generated model against the 22Rv1 cell line (by CoMFA and CoMSIA) gave the best results. The q2 value for CoMFA is 0.365 and for CoMSIA is 0.430 against the 22Rv1 cell line. We conclude that 2D QSAR in LNCaP and VCaP had better results compared to other cell lines whereas in 3D QSAR, dataset compounds yielded better results against the 22Rv1 cell line.

Keywords: Prostate Cancer, Bicalutamide Drug, QSAR, Androgen Receptor, Inhibition, Cancer Cell Lines

Introduction

Prostate Cancer is a serious threat to mankind and generally occurs in men aged 50 years or older and is the most widely recognized cancer that influences men other than skin disease. Prostate cancer is the second normal male prostate cancer [1]. It is the second driving reason for disease-related passing in men at around 35000 every year in Europe. The lifetime danger making up clinical disease and prostate tumor-related passing is 30%, 10% and 3% separately [2]. Testosterone hardship can cause interceded tumor cells to pass by apoptosis and a target reaction of 30-40% can be seen in a virgin illness. Androgen receptors stay vital in the advancement of prostate malignancy. The movement of prostate disease is additionally incorporated with expanded development factor generation and an adjusted reaction to development factors by prostate tumor cells [3]. Androgen activity including the amalgamation of testosterone, its vehicle to target tissues, and changed over from 5 alpha reductase to the dynamic 5 alpha dihydrotestosterone (DHT). Androgen activity in prostate tumors, as in the ordinary prostate organ and other target organs, is interceded by the androgen receptor (AR), an activated transcription factor that is an individual from the steroid/thyroid hormone receptor superfamily.

Bicalutamide is an anti-proliferative agent and is used to treat prostate cancer. This drug is purchased as casodex for the patients. It is normally utilized gathered with a gonadotropin-discharging hormone (GnRH) simple or careful expulsion of the testicles to treat prostate disease. Among the medications utilized for the treatment of PC, bicalutamiude specifically obstructs the activity of androgens while displaying fewer symptoms in examination with other AR antagonists. It works by blocking the androgenreceptor (AR), the biological target of the androgen sex hormones testoste rone and dihydrotestosterone (DHT) [4]. It does not decrease androgen levels. This medication can have some estrogen-like effects in men as well. Bicalutamide is well-absorbed. It is not affected by the food. The elimination half-life of the medication is around one week. It is thought to cross the blood-brain and affect both the body and brain of a person [5]. Due to its selectivity for the AR, casodex drug does not cooperate significantly with different steroid receptors and in this way, there's no clinically applicable off-target hormonal movement such as progestogenic, glucocorticoid , estrogenic etc.

A very sensitive human cell line is LNCaP. Cells of this cell line are generally utilized as a part of the field of oncology which is the investigation of growth. LNCaP cells are the androgen-delicate human prostate malignancy got from the left supraclavicular node. They are disciple epithelial cells developing in masses and also as single cells [5]. It developed an AI-PCa cell show that almost imitates clinical disease, LNCaP sublines have been made to give the most clinical tissue culture devices to date. LNCaP contains an anomalous androgen receptor framework with wide steroid-restricting specificity. Progestagens, estradiol and a few antiandrogens resist androgens for the official to the androgen receptor in the cells to a higher degree than in other androgen delicate frameworks. Ideal development of LNCaP cells is seen after the growth of the synthetic androgen R1881 (0.1 nM). We have discovered that the androgen receptor in the LNCaP cells contains a solitary point transformation changing the feeling of codon 868 (Thr to Ala) in the ligand-restricting area.

VCaP is an adherent, epithelial cell line with high Androgen receptor and Prostate-specific expression. VCaP is the only prostate cancer cell model that expresses the Androgen receptor splice variant, AR-V7, and the TMPRSS2-ERG gene fusion [6]. Another important carcinoma epithelial cell line is 22Rv1 from a xenograft that was serially proliferated in mice after castration-induced relapse and backslide of the parental, androgen-subordinate CWR22 xenograft. The cell line communicates PSA (prostate-specific antigen). Growth is ineffectively invigorated by lysates and also by dihydrotestosterone which are immunoreactive with AR neutralizer [7.] It has been determined that administration of NFkappaB ligand RANKL promoted DU145 cell disruption in bone and as a result osteolytic lesions formed [8]. Soluble factors are also made by DU-145 cells that initiate pre-osteoblast precursors and increase RANKL expression, thus facilitating prostate cancer metastasis in bone. The differences in intracellular and extracellular protein expressions between human prostate cancer lines LNCap and DU145 were examined. The proteins of the two cell lines were extracted and condensed by using protein extraction kits. The intracellular and extracellular proteins were quantitatively detected on a microplate reader by using the bicinchoninic acid (BCA) method. The proteins in cell culture fluid were qualitatively assayed by SELDI-TOF-MS. The results showed that the intracellular protein contents of LNCap cells were extremely higher than those of DU145 cells [9]. After serum-free culture, both intracellular and extracellular protein contents of LNCap and DU145 were decreased to some extent. The intracellular proteins were decreased by 5% in LNCap and by 36% in DU145 respectively, while the extracellular proteins were decreased by 89% in LNCap and 96% in DU145 respectively.

Talking about its treatment strategies, the Quantitative Structure-Movement Connections (QSAR) technique involves some statistical models that relate an arrangement of biological features from a progression of analogues with molecular properties, this is known as molecular descriptors. Examinations look to see how molecular properties beforehand recognized may impact some biological activities [10]. Both 2D and 3D QSAR studies have concentrated on the advancement of ideal QSAR models through factor determination. This suggests just a subset of accessible descriptors of chemical structures, which are the most important and statistically significant regarding connection with biological activity, is chosen. The ideal choice of factors was accomplished by various search and relationship techniques, for example, MLR, PLS examination and so on. All the more particularly, these strategies utilize either summed-up annealing, genetic algo, or evolutionary algorithms, as the stochastic optimization tool. It has been exhibited that these algos joined with different chemometric tools have viably enhanced the QSAR models contrasted with those without variable determination. In 3D QSAR the properties of molecules are calculated one by one by computer-aided programs.

2. Materials and Methods

There are different tools which are used for the 3D QSAR technique.

2.1 ChemDraw

ChemDraw is an editor tool for molecules or chemical structures. It takes chemical drawing to the next level through its features. We use this tool to form different chemical structures through different formulas. We can also clean up the structure. Other features include Chemical structure conversion and chemical name-to-structure conversion. Its different features accelerate the research even faster and enable new and growing areas of scientific research [11]. It can support different file formats like MOL, SDF SKC etc.

2.2 PyMol

PyMol is the open source software which is used to visualize different molecules. This molecular visualization system can form excellent 3D pictures of little particles and biological molecules, similar to proteins. As indicated by the creator, right around a fourth of every distributed picture of 3D protein structures in the literature was made utilizing PyMol. This visualization software is widely used in structural biology. The Py part of PYmol stands for the programming language Python. In PyMOL, we have distinct molecule editing properties such as bond rotation other than that we have an intuitive molecular relaxation etc [12]. A similar protein structure also provides various modes like Standard cartoon, surface etc. Different features of this software help us a lot in distinct fields of biology in structural biology, we can easily visualise our 3D structure in a better way and check its mutation if any.

2.3 Maestro Software

Maestro is an exceptionally helpful software in the field of bioinformatics that incorporates different actions like displaying structures, and imagining the consequences of computations on these structures. It stands for Maestro stands for Managed Automation Environment for Simulation, Test, and Real-time Operations. It is the graphical UI (GUI) for all the computational projects like LigPrep (that is for the preparation of ligands), Field-based QSAR, and so on. This software consists of various features like we can form the structure of different molecules and then handling or operating that structure for assembling or organizing. It can also store all the information of these structures and visualizing the consequences of results on these structures.

The principle Maestro work process includes performing activities on the showed structures, to change the impression and content of the structures. Sometimes we only need to change just a piece of structure and maybe visualize only that part. So for this, maestro allows us to just select particular residues or atoms from the structure on which we want to work. In Maestro, we are continually working on a project and in this project, we gathered different chemical structures and all their information. In the project, we have specified an entry to each molecule or structure and its related data. The structures and information are sorted out into sections and each entry contains different properties of molecules [13]. All the projects are shown in the project table where all the entries are displayed and its related data is given. We can import the structures or data in formats like SYBYL mol2, mdl SD, sdf etc.

Furthermore, Maestro adds two properties to the structures on import: the full path to the file from which the structures were imported (Source Path), and the file name without the path (Source File). By using this Project Table, we can continue our work and proceed through MOE. Molecular docking is also done in maestro which was used by us for the preparation of receptor-based models. Protein can be prepared in it and the water molecules are removed for the more pure form. Ligands were prepared in it as well.

2.4 Molecular Operating Environment Tool

The MOE tool helps us in the analysis of different molecules and also we can calculate distinct molecular features. QSAR's graphical interface is a door into the QSAR framework and enables the user to choose which descriptors to ascertain. The interface filters MOE and consequently recognizes QSAR modules. Through MOE, users can rapidly and effectively construct descriptors and include them in the framework; the inherent descriptor's source code is circulated with QSAR and is utilized as a template. When the computation of our descriptors is done then we further move towards the MOE QSAR model building step where we evaluate our model. After that, we do the model refinement where some more descriptors are calculated. The well-known technique named SAR (Structure-activity relationship) is essential in drug designing. QSAR methodology shows that our structure is associated with its biological activity and that as an outcome model can be demonstrated as a function of computational physiochemical properties. Then that model could be utilized for the prediction of its activity [14]. In the graphical interface of QSAR, we calculate the descriptors of all our molecules. There are different types of descriptors like 2D or 3D. Now after calculating these descriptors, we gathered them in a Molecular Database of the MOE system and we can view it by using Database Viewer in MOE.

2.5 Sanjeev’s Lab Tool

Sanjeev’s Calculator is used to convert the IC50 values of the derivatives into pIC50 values in micromolar units. IC50 represents the concentration at which a substance exerts half of its maximal inhibitory effect. This value is typically used to characterize an antagonist of a biological process (ex. phosphorylation). In pharmacology, it is an important measure of potency for a given agent. Traditionally, this value is expressed as a molar concentration.

$$pIC50= -logIC50$$

2.6 2D Molecular Descriptors

2D sub-atomic descriptors are considered to be arithmetic properties that are computed from association portrayal of particles for example components, formal bonds and charges, but not the nuclear directions. 2D descriptors, in this way, are not subject to the compliance of particles and are almost appropriate for extensive database anticipates

2.6.1 Physical Properties

The accompanying physical properties can be figured from the association table with no reliance on the compliance of a particle:

Table2.1: Physical Properties Descriptors

Code	Description
Apol	The sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities.
Bpol	The sum of the absolute value of the difference between atomic polarizabilities of all bonded atoms in the molecule (including implicit hydrogens) with polarizabilities.
FCharge	The total charge of the molecule
Mr	It's Molecular refractivity. It is calculated from an 11-descriptor linear model with r2 = 0.997, RMSE = 0.168.
SMR	Molecular refractivity which includes implicit hydrogen. This is the atomic contribution model. The model was skilled in about 7000 structures. Results may vary from mr descriptor.
Weight	The molecular weight includes implicit hydrogens with atomic weights.
logP(o/w)	It is a Log of the octanol/water partition. It is calculated from a linear type atom model with r2 = 0.931, and RMSE=0.393 on almost 1,847 molecules.
SlogP	It is a Log of the octanol/water partition coefficient. It is the property that is the atomic contribution model which calculates the logP of a given structure.
vdw_vol	It's van der Waals volume that is calculated using the connection table approximation.
Density	Its molecular mass density.
vdw_area	It is an area of van der Waals surface that is calculated using a connection table approximation.

2.6.2 Subdivided Surface Areas

The Subdivided Surface Areas are descriptors in light of an available van der Waals surface region estimation for every particle, vi alongside some other nuclear property, pi. The vi is figured utilizing an association table estimation. Every descriptor in an arrangement is characterized to be the total of the vi over all iotas, I with the end goal that pi is in a predefined extent (a, b].

Table2.2: Surface Areas Descriptors

Code	Description
SlogP_VSA0	It is the sum of vi where Li <= -0.4.
SlogP_VSA1	Its sum of vi where Li is in (-0.4,-0.2].
SlogP_VSA2	It is the sum of vi where Li is in (-0.2,0].
SlogP_VSA3	It is the sum of vi where Li is in (0,0.1].
SlogP_VSA4	Its sum of vi where Li is in (0.1,0.15].
SlogP_VSA5	Its sum of vi where Li is in (0.15,0.20].
SlogP_VSA6	It is the sum of vi where Li is in (0.20,0.25].
SlogP_VSA7	It is the sum of vi where Li is in (0.25,0.30].
SlogP_VSA8	It is the sum of vi where Li is in (0.30,0.40].
SlogP_VSA9	It is the sum of vi where Li > 0.40.
SMR_VSA0	It is the sum of vi where Ri is in [0,0.11].
SMR_VSA1	It is the sum of vi where Ri is in (0.11,0.26].

SMR_VSA2	It is the sum of vi where Ri is in (0.26,0.35].
SMR_VSA3	It is the sum of vi where Ri is in (0.35,0.39].
SMR_VSA4	It is the sum of vi where Ri is in (0.39,0.44].
SMR_VSA5	It is the sum of vi where Ri is in (0.44,0.485].
SMR_VSA6	It is the sum of vi where Ri is in (0.485,0.56].
SMR_VSA7	It is the sum of vi where Ri > 0.56.

2.6.3 Particle Counts and Bond Counts

The atom check and bond count descriptors are elements of the tallies of molecules and bonds (subdivided by different criteria).

Table 2.3: Particle Counts and Bond Counts Descriptors

Code	Description
a_aro	No. of aromatic atoms.
a_count	No. of atoms which includes implicit hydrogens. It is calculated as the sum of (1 + hi) above all non-trivial atomss i.
a_heavy	No. r of heavy atoms.
a_ICM	It is atom information content. It is the entropy of element distribution in a molecule. Let nibe the no. of existences of atomic no. i in the molecule. Let pi = ni / n where n is the sum of ni. The value of a_ICM is negative of the sum of pi log pi.
a_IC	Its total atom information content.
a_Nh	No. of hydrogen atoms. It is calculated as the sum of hi above all non-trivial atoms plus no. of non-trivial hydrogen atoms.
a_Nb	No. of boron atoms.
a_Nc	No. of carbon atoms.
a_Nn	No. of nitrogen atoms.
a_No	No. of oxygen atoms.
a_Nf	No. of fluorine atoms.
a_Np	No. of phosphorus atoms.
a_Ns	No. of sulfur atoms.
a_nCl	No. of chlorine atoms.
a_nBr	No. of bromine atoms.
a_Ni	No. of iodine atoms.
b_1rotN	No. of rotatable single bonds. A bond is rotatable if it’s not in the ring and also the atom of the bond is where (di+hi) < 2.
b_1rotR	It's a fraction of rotatable single bonds. Here b_1rotN is divided by the b_count.
b_ar	No. of aromatic bonds.
b_count	No. of bonds. It is calculated as the sum of (di/2 + hi) above all non-trivial atoms ‘i’.
b_double	No. of double bonds.
b_heavy	No. of bonds among heavy atoms
b_rotN	No. of rotatable bonds. A bond is rotatable if it is not in a ring, and also the atom of the bond is where (di+hi) < 2.

b_rotR	It's b_rotN divided by the b_count.
b_single	No. of single bonds. Aromatic bonds here are not measured to be single bonds.
b_triple	No. of triple bonds. Aromatic bonds here are not measured to be triple bonds.
VAdjMa	Vertex adjacency information. 1 + log2 m where m is the no. of heavy-heavy bonds. If m is zero, then zero is given back.
VAdjEq	Vertex adjacency information. -(1-f)log2(1-f) - f log2 f where f = (n2 - m) / n2, n is the no. of heavy atoms and m is the no. of heavy-heavy bonds.

2.6.4 Kier and Hall connectivity

Table 2.4: Kier and Hall connectivity descriptors

Code	Description
chi0	Its atomic connectivity index. It is calculated as the sum of 1/sqrt(di) above all heavy atoms i with di > 0.
chi0_C	Carbon connectivity index. It is calculated as the sum of 1/sqrt(di) above all carbon atoms i with di > 0.
chi1	It is calculated as the sum of 1/sqrt(didj) above all bonds among heavy atoms i & jwhere i < j.
chi1_C	Carbon connectivity index (order 1). This is calculated as the sum of 1/sqrt(didj) over all bonds between carbon atoms i and j where i < j.
chi0v	Atomic valence connectivity index (order 0). This is calculated as the sum of 1/sqrt(vi) overall heavy atoms i with vi > 0.
chi0v_C	Its carbon valence connectivity index. It is calculated as the sum of 1/sqrt(vi) above all carbon atoms i.
chi1v	It is the atomic valence connectivity index.
chi1v_C	It is carbon valence connectivity.
Kier1	It is the first kappa shape index which is (n-1)2 / m2
Kier2	It is the second kappa shape index which is (n-1)2 / m2
Kier3	It is the third kappa shape index which is (n-1) (n-3)2 / p32 for odd. (n-3) (n-2)2 / p32 for even n.
KierA1	Its first alpha improved shape index.
KierA2	Its second alpha improved shape index.
KierA3	Its third alpha improved shape index.
KierFlex	Its Kier molecular flexibility index is (KierA1*KierA2 / n).
Zagreb	It is the sum of di2above all heavy atoms.

2.6.5 Adjacency and Distance Matrix Descriptors

Following are adjacency matrices of heavy atoms.

Table 2.2: Adjacency and Distance Matrix Descriptors

Code	Description
BalabanJ	It is Balaban's connectivity topological index.
Diameter	It is the largest value in the distance matrix.
Petitjean	(Diameter - radius) divided by diameter.
Radius	If ri is the largest matrix entry in a row of distance matrix D. The radius is well-defined as the smallest of ri

VDistEq	If m is the sum of distance matrix entries at that time VdistEq is defined to be the sum of log2 m - pi log2 pi divided by m where pi is no. of distance matrix.
VDistMa	If m is the sum of the distance matrix entries at that time VDistMa is well-defined to be the sum of log2 m - Dij log2 Dij divided by m above all i & j.
WeinerPath	Wiener path no.
WeinerPol	Wiener polarity no.

2.6.6 Pharmacophore Feature Descriptors

The Pharmacophore type descriptors are considered as bits of a particle and allot a sort to every molecule utilizing a control-based framework. This means hydrogens are stifled among the figuring. The list of capabilities is Acceptor, Donor, Polar which is both Donor and Acceptor, Negative (corrosive), Positive (base), Hydrophobic and others as well. For instance, - COOH will be composed in its deprotonated shape paying little respect to how the structure is put away.

Table 2.6: Pharmacophore Feature Descriptors

Code	Description
a_acc	No.of hydrogen bond acceptor atoms did not include acidic atoms but included atoms.
a_acid	No. of acidic atoms.
a_base	No. of basic atoms.
a_don	No. of hydrogen bond donor atoms where not include basic atoms but included atoms.
a_hyd	No. of hydrophobic atoms.
vsa_acc	The sum of VDW surface areas of clean hydrogen bond acceptors where not including acidic atoms & atoms that are together with hydrogen bond donors and acceptor.
vsa_acid	Sum of VDW surface areas, acidic atoms.
vsa_base	Sum of VDW surface areas, basic atoms.
vsa_don	Sum of VDW surface areas, pure hydrogen bond donors where not including basic atoms & atoms that are together hydrogen bond donors & acceptors.
vsa_hyd	Sum of VDW surface areas, hydrophobic atoms.
vsa_other	Sum of VDW surface areas, "other".
vsa_pol	Sum of VDW surface areas, polar atoms

2.6.7 Partial Charge Descriptors

Descriptors that rely upon the halfway charge of every particle of a synthetic structure require count of those incomplete charges. The accompanying variations are PEOE and Q. The Partial Equalization of Orbital Electronegativities (PEOE) strategy for figuring nuclear incomplete charges is a technique in which charge is exchanged among fortified molecules until balance. To ensure meeting, the measure of charge exchanged at every emphasis is checked with an exponentially scale factor. Q are stored in database with each structure. In this no partial charge has been used for atomic partial charges. This will help to have a subtle source of error. For example, QSAR models having no novel structures.

Table 2.7: Partial Charge Descriptors

Code	Description
Q_PC+ PEOE_PC+	Total +ive partial charge

Q_PC- PEOE_PC-	Total -ive partial charge.
Q_RPC+ PEOE_RPC+	Relative +ive partial charge
Q_PRC- PEOE_RPC-	Relative -ive partial charge
Q_VSA_POS PEOE_VSA_POS	Total +ive van der Waals surface area.
Q_VSA_NEG PEOE_VSA_NEG	Total -ive van der Waals surface area.
Q_VSA_PPOS PEOE_VSA_PPOS	Total +ive polar van der Waals surface area.
Q_VSA_PNEG PEOE_VSA_PNEG	Total -ive polar van der Waals surface area.
Q_VSA_HYD PEOE_VSA_HYD	Total hydrophobic van-der-Waals surface area.
Q_VSA_POL PEOE_VSA_POL	Total polar van-der-Waals surface area.
Q_VSA_FPOS PEOE_VSA_FPOS	Fractional +ive van der Waals surface area.
Q_VSA_FNEG PEOE_VSA_FNEG	Fractional -ive van der Waals surface area.
Q_VSA_FPPOS PEOE_VSA_FPPOS	Fractional +ive polar van der Waals surface area.
Q_VSA_FPNEG PEOE_VSA_FPNEG	Fractional-ive polar van-der-Waals surface area.
Q_VSA_FHYD PEOE_VSA_FHYD	Fractional hydrophobic van-der-Waals surface area.
Q_VSA_FPOL PEOE_VSA_FPOL	Fractional polar van-der-Waals surface area.
PEOE_VSA+6	The sum of vi such that qi is greater than 0.3.
PEOE_VSA+5	The sum of vi such that qi is in the range (0.25-.30).
PEOE_VSA+4	The sum of vi such that qi is in the range (0.20-0.25).
PEOE_VSA+3	The sum of vi such that qi is in range (0.15-0.20).
PEOE_VSA+2	The sum of vi such that qi is in the range (0.10-0.15)
PEOE_VSA+1	The sum of vi such that qi is in the range (0.05-0.10)
PEOE_VSA+0	The sum of vi such that qi is in the range (0.00-0.05)
PEOE_VSA-0	The sum of vi such that qi is in the range (-0.05-0.00).
PEOE_VSA-1	The sum of vi such that qi is in the range (-0.10-(-0.05))
PEOE_VSA-2	The sum of vi such that qi is in the range (-0.15-(-0.10))
PEOE_VSA-3	The sum of vi such that qi is in the range (-0.20-(-0.15))
PEOE_VSA-4	The sum of vi such that qi is in the range (-0.25-(-0.20))
PEOE_VSA-5	The sum of vi such that qi is in the range (-0.30-(-0.25))
PEOE_VSA-6	The sum of vi such that qi is less than -0.30.

2.7 Correlation Matrix

MS Excel is usually used to make different spreadsheets or databases. So, here we also used Excel to open our databases that we already created in MOE Tool and do the data analysis of all the cell lines separately through a correlation matrix. We also use the Maestro tool to open these databases as a spreadsheet in Excel in CSV or XML format.

Table 2.8: Some eliminated Descriptors

DESCRIPTION	NAME	PROPERTY
Apol	Sum of polarizabilities	The sum of the atomic polarizabilities (including implicit hydrogens) with polarizabilities taken from [CRC 1994]
A_aac	A number of hydrogen bond acceptors	hydrogen bond acceptor.
A_don	Number of hydrogen bond donor atoms	No. of hydrogen bond donor atoms where not include basic atoms but included atoms that are both hydrogen bond donors &acceptors.
A_donace		
A_heavy	Number of heavy atoms	No. of heavy atoms.
B_ar	Number of aromatic bonds	No. of aromatic bonds
A_hyd	Number of hydrophobic bonds	No. of hydrophobic bonds
B_count	Number of bonds	No. of bonds including implicit hydrogens.
B_heavy	Number of heavy bonds	No. of bonds between heavy atoms
B_rotR	Fraction of rotatable bonds	B_rotN divided by heavy
B_single	Number of single bonds	Including implicit hydrogen's). This is calculated as the sum of (di/2 + h) over all non-trivial atoms i.
Q_PC+	Total positive partial charge	The sum of positive q1.Q_PC+ is identical to PC+ which has been retained for compatibility
Q_PC-	Total negative partial charge	The sum of positive q1.Q_PC- is identical to PC- which has been retained for compatibility
Q_VSA_FHYD	Fractional hydrophobic vdw surface area	The sum of vi where |qi| is <= to 0.2 divided by the total surface area.
Q_VSA_HYD	Total hydrophobic vdw surface area	The sum of vi where |qi| <= to 0.2 divided by the total surface area.
Q_VSA_PNEG	Total polar negative vdw surface area	The sum of vi such that qi is less than -0.2 divided by the total surface area.
Q_VSA_PPOS	Total polar positive vdw surface area	The sum of the vi is such that qi > 0.2 divided by the total surface area.
Vsa_acc	VDW acceptor surface area (A**2)	The sum of VDW surface areas of uncontaminated hydrogen bond acceptors
Vsa_acid	VDW donor surface area (A**2)	The sum of VDW surface areas of uncontaminated hydrogen bond acceptors where not include basic atoms & atoms that are both hydrogen atom donors & acceptors.
Vsa_don	VDW hydrophobe surface area (A**2)	Sum of V-D-W surface areas of hydrophobic atoms.
Vsa_hyd	VDW other surface area (A**2)	Approximation to sum of V-D-W surface area of atoms.
Vsa_other	VDW polar surface area (A**2)	Approximation to the sum of V-D-W surface areas of polar atoms which are the atoms that are together hydrogen bond acceptor and donor.
Vsa_pol	No. of aromatic bonds	No. of aromatic compounds

2.8 SYBYL

For 3DQSAR, SYBYL software is the better option to be used. The SYBYL provides a complete drug and molecular design environment with comprehensive tools for molecular modeling. It includes small molecule and macromolecular modeling and simulation also which leads to identification and optimization. With the SYBYL we can perform multi-criteria drug design and predict safety issues and/or off-target pharmacology. It is ligand-based or structure-based virtual screening. We can also design a chemical library in it. Most importantly, we perform lead optimization using a variety of QSAR methods such as CoMFA.

3. Results

3.1 Sketching Structures

We have in total 85 derivatives of the bicalutamide drug on which we are working. First of all, we made all 85 structures through chemdraw which is an editor tool for molecules. These 85 structures are the derivative of bicalutamide an antiproliferative agent used to treat prostate disease. Our all structures are characterized by the position or the nearness of the trifluoromethyl group, as its presence or absence can affect the activity of the biological molecules. By improving the pharmacological properties and medication-like physiochemical attributes of these compounds, we can enhance their antiproliferative activity.

Table 3.1: Dataset of 85 Bicalutamide derivatives

Index	COMPOUND	Ar(B ring)	X	R(A ring)	Absolute IC50 (micromolar)			
					22Rv1	DU-145	LNCaP	VCaP
1	3 (Bic.)	4-F-Ph	SO2	4-CN,3 CF3	4.3	4.3	4.3	4.1
2	22c	3-CF3-Ph	S	4-CN,3 CF3	5.2	4.9	5.2	5.0
3	22d	2-CF3-Ph	S	4-CN,3 CF3	5.2	5.1	5.3	5.1
4	22h	4-OCF3-Ph	S	4-CN,3 CF3	4.3	4.1	4.3	4.3
5	22o	4-CF3-2-Pyridine	S	4-CN,3 CF3	4.2	4.0	4.0	4.0
6	23b	4-CF3-Ph	S	4-NO2, 3-CF3	4.7	4.6	4.5	4.6
7	23c	3-CF3-Ph	S	4-NO2, 3-CF3	5.2	5.0	5.1	5.0
8	23d	2-CF3-Ph	S	4-NO2, 3-CF3	5.3	5.1	5.3	5.3
9	24a	4-F-Ph	S	4-CN,2-CF3	4.5	4.3	4.5	4.3
10	24c	3-CF3-Ph	S	4-CN,2-CF3	4.8	4.6	4.9	4.6
11	24d	2-CF3-Ph	S	4-CN,2-CF3	4.9	4.7	4.9	4.7
12	25a	4-F-Ph	S	4-NO2, 2-CF3	4.7	4.5	4.9	4.6
13	25b	4-CF3-Ph	S	4-NO2, 2-CF3	4.8	4.7	5.0	4.7
14	25c	3-CF3-Ph	S	4-NO2, 2-CF3	4.9	4.8	4.9	4.7
15	25d	2-CF3-Ph	S	4-NO2, 2-CF3	4.9	4.7	4.9	4.8
16	25f	3,4-F-Ph	S	4-NO2, 2-CF3	4.0	4	4.0	4.0
17	25g	2,4-F-Ph	S	4-NO2, 2-CF3	4.8	4.6	4.9	4.7
18	25h	4-OCF3-Ph	S	4-NO2, 2-CF3	4.2	4.1	4	4.0
19	25i	3-OCF3-Ph	S	4-NO2, 2-CF3	4.8	4.6	5.0	4.7
20	25l	2-OCF3-Ph	S	4-NO2, 2-CF3	5.3	5.1	5.4	5.2
21	25o	4-CF3-2-Pyridine	S	4-NO2, 2-CF3	4.8	4.7	5.2	4.7
22	25p	5-CF3-2-Pyridine	S	4-NO2, 2-CF3	4.9	4.8	5.2	4.7
23	26c	3-CF3-Ph	S	4-CF3	4.7	4.5	4.6	4.7
24	26d	2-CF3-Ph	S	4-CF3	5.2	5.0	5.1	5.2
25	26i	3-OCF3-Ph	S	4-CF3	4.7	4.5	4.8	4.7
26	27b	4-CF3-Ph	O	4-CN, 3-CF3	5.1	4.7	5.1	5.0
27	27c	3-CF3-Ph	O	4-CN, 3-CF3	5.2	4.9	5.0	5.0
28	27d	2-CF3-Ph	O	4-CN, 3-CF3	4.9	4.7	5.0	4.7
29	27f	3,4-F-Ph	O	4-CN, 3-CF3	4.4	4.3	4.7	4.5
30	27g	2,4-F-Ph	O	4-CN, 3-CF3	4.4	4.3	4.4	4.4
31	27i	3-OCF3-Ph	O	4-CN, 3-CF3	5.0	4.7	5.0	5.0
32	27o	4-CF3-2-Pyridine	O	4-CN, 3-CF3	4	4	4	4
33	28b	4-CF3-Ph	O	4-NO2, 3-CF3	4.7	4.5	4.7	4.9
34	28c	3-CF3-Ph	O	4-NO2, 3-CF3	5.0	4.9	5.0	4.9
35	28d	2-CF3-Ph	O	4-NO2, 3-CF3	4.6	4.4	4.6	4.5
36	28e	4-CN-Ph	O	4-NO2, 3-CF3	4.5	4.4	4.7	4.5
37	28f	3,4-F-Ph	O	4-NO2, 3-CF3	4.7	4.6	4.8	4.7
38	28g	2,4-F-Ph	O	4-NO2, 3-CF3	4.6	4.6	4.9	4.6
39	28h	4-OCF3-Ph	O	4-NO2, 3-CF3	5.1	4.9	4.7	4.7
40	28i	3-OCF3-Ph	O	4-NO2, 3-CF3	5.0	4.8	5.0	5.0
41	28l	2-OCF3-Ph	O	4-NO2, 3-CF3	5.1	5.0	5.15	5.0
42	28m	4-CN,2-CF3-Ph	O	4-NO2, 3-CF3	5.2	5.1	5.2	5.0
43	28n	4-CN,3-F-Ph	O	4-NO2, 3-CF3	4.9	4.7	5.1	4.9
44	28o	4-CF3-2-Pyridine	O	4-NO2, 3-CF3	4.4	4	4.4	4
45	29b	4-CF3-Ph	O	4-CN, 2-CF3	4.5	4.4	4.5	4.6
46	29c	3-CF3-Ph	O	4-CN, 2-CF3	4.7	4.4	4.8	4.6
47	29d	2-CF3-Ph	O	4-CN, 2-CF3	4.7	4.4	4.8	4.6
48	29e	4-CN-Ph	O	4-CN, 2-CF3	4.3	4	4.4	4.2
49	29f	3,4-F-Ph	O	4-CN, 2-CF3	4.4	4.3	4.4	4.3
50	29g	2,4-F-Ph	O	4-CN, 2-CF3	4.2	4.0	4,5	4.2
51	29h	4-OCF3-Ph	O	4-CN, 2-CF3	4.5	4.4	4.5	4.6
52	29i	3-OCF3Ph	O	4-CN, 2-CF3	4.5	4.4	4.1	4.2
53	29l	2-OCF3-Ph	O	4-CN, 2-CF3	4.7	4.5	4.4	4.3
54	29m	4-CN,2-CF3-Ph	O	4-CN, 2-CF3	4.7	4.5	5.0	4.7
55	29n	4-CN,3-F-Ph	O	4-CN, 2-CF3	4.3	4.1	4.1	4.1
56	29o	4-CF3-2-Pyridine	O	4-CN, 2-CF3	4	4	4	4
57	30b	4-CF3-Ph	O	4-NO2, 2-CF3	4.7	4.6	4.8	4.7

58	30c	3-CF3-Ph	O	4-NO2, 2-CF3	4.8	4.7	4.8	4.7
59	31c	3-CF3-Ph	O	4-CF3	4.5	4.5	4.7	4.7
60	31d	2-CF3-Ph	O	4-CF3	5.1	4.5	5.1	4.9
61	31i	3-OCF3-Ph	O	4-CF3	4.8	4.5	5.6	5.3
62	32b	4-CF3-Ph	SO2	4-CN, 3-CF3	4.6	4.4	4.6	4.5
63	32c	3-CF3-Ph	SO2	4-CN, 3-CF3	4.6	4.4	4.8	4.4
64	32d	2-CF3-Ph	SO2	4-CN, 3-CF3	4.3	4.2	4.3	4.2
65	32h	4-OCF3-Ph	SO2	4-CN, 3-CF3	4.7	4.5	4.4	4.4
66	32o	4-CF3-2-Pyridine	SO2	4-CN, 3-CF3	4	4	4	4
67	33b	4-CF3-Ph	SO2	4NO2, 3-CF3	4.7	4.6	4.7	4.5
68	33c	3-CF3-Ph	SO2	4NO2, 3-CF3	4.7	4.5	4.7	4.5
69	33d	2-CF3-Ph	SO2	4NO2, 3-CF3	4.7	4.4	4.5	4.5
70	34a	4-F-Ph	SO2	4-CN, 2-CF3	4.1	4	4.0	4.1
71	34b	4-CF3-Ph	SO2	4-CN, 2-CF3	4.5	4.3	4.4	4.3
72	34c	3-CF3-Ph	SO2	4-CN, 2-CF3	4.4	4.3	4.4	4.3
73	34d	2-CF3-Ph	SO2	4-CN, 2-CF3	4.0	4	4.1	4
74	35a	4-F-Ph	SO2	4-NO2, 2-CF3	4.3	4.2	4.4	4.2
75	35b	4-CF3-Ph	SO2	4-NO2, 2-CF3	4.5	4.3	4.5	4.3
76	35c	3-CF3-Ph	SO2	4-NO2, 2-CF3	4.7	4.4	4.7	4.4
77	35d	2-CF3-Ph	SO2	4-NO2, 2-CF3	4.5	4.3	4.6	4.4
78	35f	3,4-F-Ph	SO2	4-NO2, 2-CF3	4.4	4.3	4.7	4.4
79	35g	2,4-F-Ph	SO2	4-NO2, 2-CF3	4.5	4.3	4.7	4.2
80	35h	4-OCF3-Ph	SO2	4-NO2, 2-CF3	4	4	4	4
81	35i	3-OCF3-Ph	SO2	4-NO2, 2-CF3	4.7	4.5	4.9	4.6
82	35l	2-OCF3-Ph	SO2	4-NO2, 2-CF3	4.5	4.4	4.7	4.2
83	35o	4-CF3-2-Pyridine	SO2	4-NO2, 2-CF3	4	4	4	4
84	35p	5-CF3-2-Pyridine	SO2	4-NO2, 2-CF3	4.3	4.2	4.3	4.2
85	52	2-CF3-Ph	S	4-NO2, 3-CF3	5.8	5.8	6.1	5.5

3.2 W741L Mutation

Through PyMOL we superimposed the two structures wild type and mutated type of bicalutamide drug and enzalutamide drug and checked where the W741L mutation occur. The Green Structure is 1z95 is the Androgen Receptor LBD (Ligand-binding Domain) of bicalutamide and the Pink Structure is 3v49 is the Structure of Enzalutamide arlbd with activator peptide and sarm inhibitor 1.

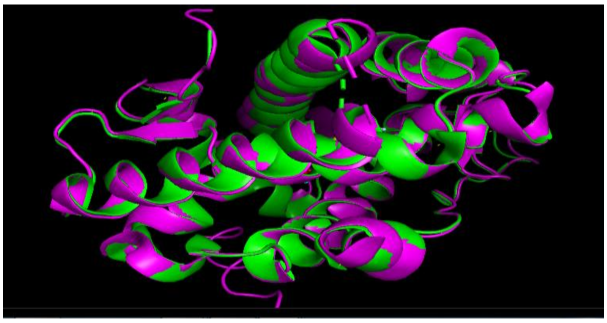


Figure 3.1: Superimposition of 1z95 (bicalutamide) and 3v49 (enzalutamide) drug

In the W741L mutation, the tryptophan is mutated with leucine in 741 position. The highlighted yellow portion is W (Tryptophan) and the red portion is L (Leucine).

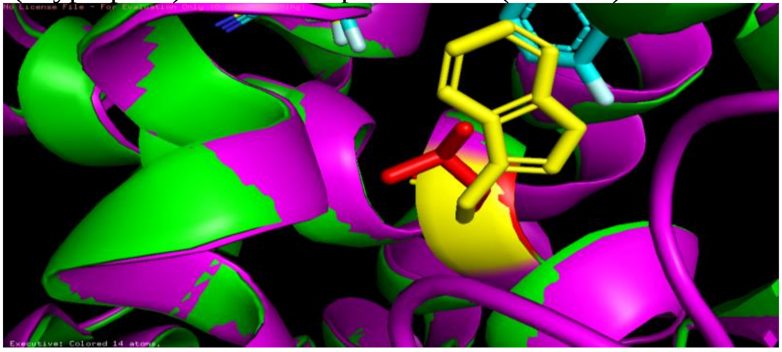

No License File - For Evaluation Only
Executive: Colored 14 atoms.

Figure 3.2: Close Visualization of W741L

3.3 Structure Minimization

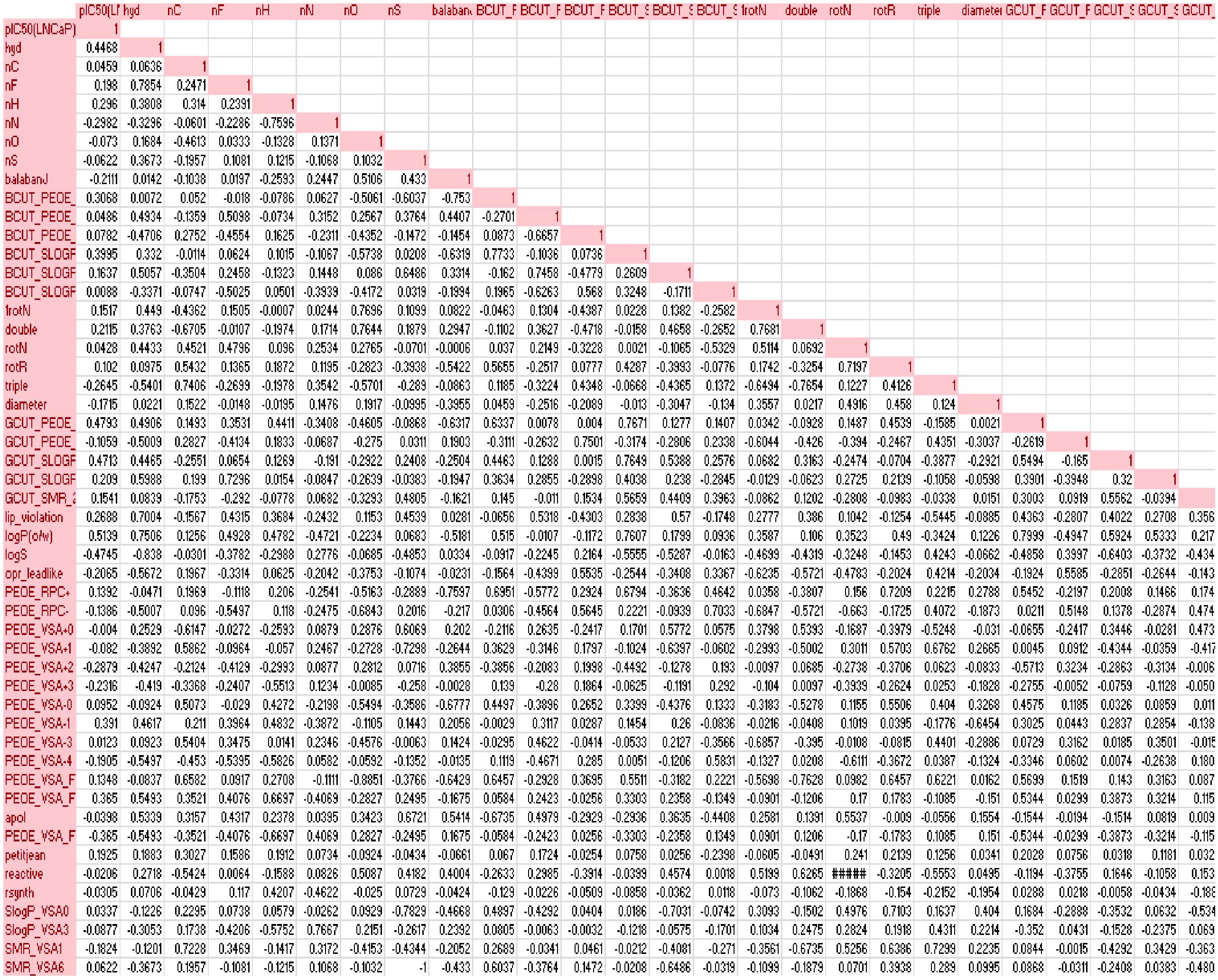
By using Maestro software, we form the structure of different molecules and then operate that structure for assembling or organizing and do the structure minimization by adding hydrogens. We also store all the information of these structures and visualizing the consequences of results on these structures.

3.4 Database Construction

After that, we form 4 databases of four human cell lines i.e.: 22Rv1, DU-145, LNCaP and VCaP of Prostate Cancer on MOE Software. In these databases, we first change the IC50 (Inhibition Constant) values of all the cell lines into pIC50 which is a negative log of IC50 in the micromolar unit. After converting the IC50 values, we calculated the 2D descriptors of all the cell lines separately. We also calculate the RMSE and R2 values of all the descriptors before pruning. Now we open these databases in MS Excel in CSV or xlsx format and perform data analysis. In MS Excel, we eliminate all the descriptors containing zero values and form the correlation matrix in Excel.

3.5 Pruning of Descriptors

Now we set a threshold value that is 0.8 and eliminate all those values which are greater than 0.8 because these values show the highest correlation. That means the two descriptors having the highest correlation values are strongly dependent upon each other and we want all independent values. So this step of eliminating the descriptors based on threshold value is called Pruning. Now we form a correlation matrix of all the cell lines i.e.: 4 in MS Excel.



	pIC50(LI	hyd	nC	nF	nH	nN	nO	nS	balaban	BCUT_F	BCUT_F	BCUT_F	BCUT_S	BCUT_S	BCUT_S	frotN	double	rotN	rotR	triple	diameter	GCUT_F	GCUT_F	GCUT_S	GCUT_S	GCUT_
pIC50(LNCaP)	1																									
hyd	0.4468	1																								
nC	0.0459	0.0636	1																							
nF	0.198	0.7854	0.2471	1																						
nH	0.296	0.3808	0.314	0.2391	1																					
nN	-0.2982	-0.3296	-0.0601	-0.2286	-0.7596	1																				
nO	-0.073	0.1684	-0.4613	0.0333	-0.1328	0.1371	1																			
nS	-0.0622	0.3673	-0.1957	0.1081	0.1215	-0.1068	0.1032	1																		
balabanJ	-0.2111	0.0142	-0.1038	0.0197	-0.2593	0.2447	0.5106	0.433	1																	
BCUT_PEOE_	0.3068	0.0072	0.052	-0.018	-0.0786	0.0627	-0.5061	-0.6037	-0.753	1																
BCUT_PEOE_	0.0486	0.4934	-0.1359	0.5098	-0.0734	0.3152	0.2567	0.3764	0.4407	-0.2701	1															
BCUT_PEOE_	0.0782	-0.4706	0.2752	-0.4554	0.1625	-0.2311	-0.4352	-0.1472	-0.1454	0.0873	-0.6657	1														
BCUT_SLOGF	0.3995	0.332	-0.0114	0.0624	0.1015	-0.1067	-0.5738	0.0208	-0.6319	0.7733	-0.1036	0.0736	1													
BCUT_SLOGF	0.1637	0.5057	-0.3504	0.2458	-0.1323	0.1448	0.086	0.6486	0.3314	-0.162	0.7458	-0.4779	0.2609	1												
BCUT_SLOGF	0.0088	-0.3371	-0.0747	-0.5025	0.0501	-0.3939	-0.4172	0.0319	-0.1994	0.1965	-0.6263	0.568	0.3248	-0.1711	1											
frotN	0.1517	0.449	-0.4362	0.1505	-0.0007	0.0244	0.7696	0.1099	0.0822	-0.0463	0.1304	-0.4387	0.0228	0.1382	-0.2582	1										
double	0.2115	0.3763	-0.6705	-0.0107	-0.1974	0.1714	0.7644	0.1879	0.2947	-0.1102	0.3627	-0.4718	-0.0158	0.4658	-0.2652	0.7681	1									
rotN	0.0428	0.4433	0.4521	0.4796	0.096	0.2534	0.2765	-0.0701	-0.0006	0.037	0.2149	-0.3228	0.0021	-0.1065	-0.5329	0.5114	0.0692	1								
rotR	0.102	0.0975	0.5432	0.1365	0.1872	0.1195	-0.2823	-0.3938	-0.5422	0.5655	-0.2517	0.0777	0.4287	-0.3993	-0.0776	0.1742	-0.3254	0.7197	1							
triple	-0.2645	-0.5401	0.7406	-0.2699	-0.1978	0.3542	-0.5701	-0.289	-0.0863	0.1185	-0.3224	0.4348	-0.0668	-0.4365	0.1372	-0.6494	-0.7654	0.1227	0.4126	1						
diameter	-0.1715	0.0221	0.1522	-0.0148	-0.0195	0.1476	0.1917	-0.0995	-0.3955	0.0459	-0.2516	-0.2089	-0.013	-0.3047	-0.134	0.3557	0.0217	0.4916	0.458	0.124	1					
GCUT_PEOE_	0.4793	0.4906	0.1493	0.3531	0.4411	-0.3408	-0.4605	-0.0868	-0.6317	0.6337	0.0078	0.004	0.7671	0.1277	0.1407	0.0342	-0.0928	0.1487	0.4539	-0.1585	0.0021	1				
GCUT_PEOE_	-0.1059	-0.5009	0.2827	-0.4134	0.1833	-0.0687	-0.275	0.0311	0.1903	-0.3111	-0.2632	0.7501	-0.3174	-0.2806	0.2338	-0.6044	-0.426	-0.394	-0.2467	0.4351	-0.3037	-0.2619	1			
GCUT_SLOGF	0.4713	0.4465	-0.2551	0.0654	0.1269	-0.191	-0.2922	0.2408	-0.2504	0.4463	0.1288	0.0015	0.7649	0.5388	0.2576	0.0682	0.3163	-0.2474	-0.0704	-0.3877	-0.2921	0.5494	-0.165	1		
GCUT_SLOGF	0.209	0.5988	0.199	0.7296	0.0154	-0.0847	-0.2639	-0.0383	-0.1947	0.3634	0.2855	-0.2898	0.4038	0.238	-0.2845	-0.0129	-0.0623	0.2725	0.2139	-0.1058	-0.0598	0.3901	-0.3948	0.32	1	
GCUT_SMR_	0.1541	0.0839	-0.1753	-0.292	-0.0778	0.0682	-0.3293	0.4805	-0.1621	0.145	-0.011	0.1534	0.5659	0.4409	0.3963	-0.0862	0.1202	-0.2808	-0.0983	-0.0338	0.0151	0.3003	0.0919	0.5562	-0.0394	
lip_violation	0.2688	0.7004	-0.1567	0.4315	0.3684	-0.2432	0.1153	0.4539	0.0281	-0.0656	0.5318	-0.4303	0.2838	0.57	-0.1748	0.2777	0.386	0.1042	-0.1254	-0.5445	-0.0885	0.4363	-0.2807	0.4022	0.2708	0.356
logP(o/w)	0.5139	0.7506	0.1256	0.4928	0.4782	-0.4721	-0.2234	0.0683	-0.5181	0.515	-0.0107	-0.1172	0.7607	0.1799	0.0936	0.3587	0.106	0.3523	0.49	-0.3424	0.1226	0.7999	-0.4947	0.5924	0.5333	0.217
logS	-0.4745	-0.838	-0.0301	-0.3782	-0.2988	0.2776	-0.0685	-0.4853	0.0334	-0.0917	-0.2245	0.2164	-0.5555	-0.5287	-0.0163	-0.4699	-0.4319	-0.3248	-0.1453	0.4243	-0.0662	-0.4858	0.3997	-0.6403	-0.3732	-0.434
opr_leadlike	-0.2065	-0.5672	0.1967	-0.3314	0.0625	-0.2042	-0.3753	-0.1074	-0.0231	-0.1564	-0.4399	0.5535	-0.2544	-0.3408	0.3367	-0.6235	-0.5721	-0.4783	-0.2024	0.4214	-0.2034	-0.1924	0.5585	-0.2851	-0.2644	-0.143
PEOE_RPC+	0.1392	-0.0471	0.1969	-0.1118	0.206	-0.2541	-0.5163	-0.2889	-0.7597	0.6951	-0.5772	0.2924	0.6794	-0.3636	0.4642	0.0358	-0.3807	0.156	0.7209	0.2215	0.2788	0.5452	-0.2197	0.2008	0.1466	0.174
PEOE_RPC-	-0.1386	-0.5007	0.096	-0.5497	0.118	-0.2475	-0.6843	0.2016	-0.217	0.0306	-0.4564	0.5645	0.2221	-0.0939	0.7033	-0.6847	-0.5721	-0.663	-0.1725	0.4072	-0.1873	0.0211	0.5148	0.1378	-0.2874	0.474
PEOE_VSA+0	-0.004	0.2529	-0.6147	-0.0272	-0.2593	0.0879	0.2876	0.6069	0.202	-0.2116	0.2635	-0.2417	0.1701	0.5772	0.0575	0.3798	0.5393	-0.1687	-0.3979	-0.5248	-0.031	-0.0655	-0.2417	0.3446	-0.0281	0.473
PEOE_VSA+1	-0.082	-0.3892	0.5862	-0.0964	-0.057	0.2467	-0.2728	-0.7298	-0.2644	0.3629	-0.3146	0.1797	-0.1024	-0.6397	-0.0602	-0.2993	-0.5002	0.3011	0.5703	0.6762	0.2665	0.0045	0.0912	-0.4344	-0.0359	-0.417
PEOE_VSA+2	-0.2879	-0.4247	-0.2124	-0.4129	-0.2993	0.0877	0.2812	0.0716	0.3855	-0.3856	-0.2083	0.1998	-0.4492	-0.1278	0.193	-0.0097	0.0685	-0.2738	-0.3706	0.0623	-0.0833	-0.5713	0.3234	-0.2863	-0.3134	-0.006
PEOE_VSA+3	-0.2316	-0.419	-0.3368	-0.2407	-0.5513	0.1234	-0.0085	-0.258	-0.0028	0.139	-0.28	0.1864	-0.0625	-0.1191	0.292	-0.104	0.0097	-0.3939	-0.2624	0.0253	-0.1828	-0.2755	-0.0052	-0.0759	-0.1128	-0.050
PEOE_VSA-0	0.0952	-0.0924	0.5073	-0.029	0.4272	-0.2198	-0.5494	-0.3586	-0.6777	0.4497	-0.3896	0.2652	0.3399	-0.4376	0.1333	-0.3183	-0.5278	0.1155	0.5506	0.404	0.3268	0.4575	0.1185	0.0326	0.0859	0.011
PEOE_VSA-1	0.391	0.4617	0.211	0.3964	0.4832	-0.3872	-0.1105	0.1443	0.2056	-0.0029	0.3117	0.0287	0.1454	0.26	-0.0836	-0.0216	-0.0408	0.1019	0.0395	-0.1776	-0.6454	0.3025	0.0443	0.2837	0.2854	-0.138
PEOE_VSA-3	0.0123	0.0923	0.5404	0.3475	0.0141	0.2346	-0.4576	-0.0063	0.1424	-0.0295	0.4622	-0.0414	-0.0533	0.2127	-0.3566	-0.6857	-0.395	-0.0108	-0.0815	0.4401	-0.2886	0.0729	0.3162	0.0185	0.3501	-0.015
PEOE_VSA-4	-0.1905	-0.5497	-0.453	-0.5395	-0.5826	0.0582	-0.0592	-0.1352	-0.0135	0.1119	-0.4671	0.285	0.0051	-0.1206	0.5831	-0.1327	0.0208	-0.6111	-0.3672	0.0387	-0.1324	-0.3346	0.0602	0.0074	-0.2638	0.180
PEOE_VSA_F	0.1348	-0.0837	0.6582	0.0917	0.2708	-0.1111	-0.8851	-0.3766	-0.6429	0.6457	-0.2928	0.3695	0.5511	-0.3182	0.2221	-0.5698	-0.7628	0.0982	0.6457	0.6221	0.0162	0.5699	0.1519	0.143	0.3163	0.087
PEOE_VSA_F	0.365	0.5493	0.3521	0.4076	0.6697	-0.4069	-0.2827	0.2495	-0.1675	0.0584	0.2423	-0.0256	0.3303	0.2358	-0.1349	-0.0901	-0.1206	0.17	0.1783	-0.1085	-0.151	0.5344	0.0299	0.3873	0.3214	0.115
apol	-0.0398	0.5339	0.3157	0.4317	0.2378	0.0395	0.3423	0.6721	0.5414	-0.6735	0.4979	-0.2929	-0.2936	0.3635	-0.4408	0.2581	0.1391	0.5537	-0.009	-0.0556	0.1554	-0.1544	-0.0194	-0.1514	0.0819	0.009
PEOE_VSA_F	-0.365	-0.5493	-0.3521	-0.4076	-0.6697	0.4069	0.2827	-0.2495	0.1675	-0.0584	-0.2423	0.0256	-0.3303	-0.2358	0.1349	0.0901	0.1206	-0.17	-0.1783	0.1085	0.151	-0.5344	-0.0299	-0.3873	-0.3214	-0.115
petitjean	0.1925	0.1883	0.3027	0.1586	0.1912	0.0734	-0.0924	-0.0434	-0.0661	0.067	0.1724	-0.0254	0.0758	0.0256	-0.2398	-0.0605	-0.0491	0.241	0.2139	0.1256	0.0341	0.2028	0.0756	0.0318	0.1181	0.032
reactive	-0.0206	0.2718	-0.5424	0.0064	-0.1588	0.0826	0.5087	0.4182	0.4004	-0.2633	0.2985	-0.3914	-0.0399	0.4574	0.0018	0.5199	0.6265	#####	-0.3205	-0.5553	0.0495	-0.1194	-0.3755	0.1646	-0.1058	0.153
rsynth	-0.0305	0.0706	-0.0429	0.117	0.4207	-0.4622	-0.025	0.0729	-0.0424	-0.129	-0.0226	-0.0509	-0.0858	-0.0362	0.0118	-0.073	-0.1062	-0.1868	-0.154	-0.2152	-0.1954	0.0288	0.0218	-0.0058	-0.0434	-0.188
SlogP_VSA0	0.0337	-0.1226	0.2295	0.0738	0.0579	-0.0262	0.0929	-0.7829	-0.4668	0.4897	-0.4292	0.0404	0.0186	-0.7031	-0.0742	0.3093	-0.1502	0.4976	0.7103	0.1637	0.404	0.1684	-0.2888	-0.3532	0.0632	-0.534
SlogP_VSA3	-0.0877	-0.3053	0.1738	-0.4206	-0.5752	0.7667	0.2151	-0.2617	0.2392	0.0805	-0.0063	-0.0032	-0.1218	-0.0575	-0.1701	0.1034	0.2475	0.2824	0.1918	0.4311	0.2214	-0.352	0.0431	-0.1528	-0.2375	0.069
SMR_VSA1	-0.1824	-0.1201	0.7228	0.3469	-0.1417	0.3172	-0.4153	-0.4344	-0.2052	0.2689	-0.0341	0.0461	-0.0212	-0.4081	-0.271	-0.3561	-0.6735	0.5256	0.6386	0.7299	0.2235	0.0844	-0.0015	-0.4292	0.3429	-0.363
SMR_VSA6	0.0622	-0.3673	0.1957	-0.1081	-0.1215	0.1068	-0.1032	-1	-0.433	0.6037	-0.3764	0.1472	-0.0208	-0.6486	-0.0319	-0.1099	-0.1879	0.0701	0.3938	0.289	0.0995	0.0868	-0.0311	-0.2408	0.0383	-0.480

Figure 3.3: Correlation Matrix of LNCaP cell line

3.6 Calculating the RMSE and R2 Value Before and After Pruning

After Pruning, we again calculate the RMSE or R2 value of all the cell lines in MOE Software and then compare the values before and after pruning.

Table 3.2: RMSE and R2 value of four cell lines

Cell Lines		RMSE	R2
22Rv1	Before Pruning	25.8132	0.567917
	After Pruning	0.292571	0.647506
DU-145	Before Pruning	26.9971	0.527709
	After Pruning	0.352569	0.545097
LNCaP	Before Pruning	0.00284395	0.9999
	After Pruning	0.054005	0.9984
VCaP	Before Pruning	0.00284395	0.9999
	After Pruning	0.00213235	0.9999

RMSE value is a root-mean-square error (RMSE) (or sometimes root-mean-squared error) and is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values observed and R2 is a correlation coefficient or cross-validation coefficient of our model. Here are the results of RMSE and R2 values before and after pruning of different descriptors. The RMSE value of the first two cell lines that is 22Rv1 and DU-145 decreases after pruning and the R2 value increases after pruning whereas the RMSE value of the other two cell lines that is LNCaP and VCaP increases and the R2 value remains the same.

3.7 3D QSAR

3.7.1 Training and Test Set

Several reported methods were present to divide the compounds into test and training sets for further processing e.g., Random selection, Hierarchical clustering, Kennard stone algorithm etc. The nth selection method also known as the Activity Sorting Method is used for dividing the compounds into test sets and training datasets. First of all, we divide our 85 derivatives into training and test sets. 66 derivatives in the training set and 19 derivatives in the test set. After that, we changed the format from .sdf to .mol and put all the derivatives in MOE for energy minimization.

3.7.2 Energy Minimization

Now as indicated by some computational model, the inter-atomic force of each atom is calculated. The position on PES (potential energy surface) is a stationary point and the net force is adequately near zero. All the obtained structures are needed to minimize its physical significance and to stabilize. After energy minimization, these structures are optimized and are useful in many experimental investigations in the field of bioinformatics. OMEGA and MOE both are for energy minimization separately.

3.7.2.1 MOE

MOE tool can be used for the energy minimization of all the derivatives. After this step, the format should be changed from mol to mol2 to support the SYBYL software for further investigations. After the changing of the format, the following steps are done for 3DQSAR.

3.7.2.2 Omega

Omega was used using the Linux operating system. All the derivatives were converted to mol2 format and the energy was minimized using commands for the test and training test when combined.

3.7.3 Constructing databases after minimization

After the minimization of all the derivatives, in Sybyl software, the spreadsheet was created having all derivatives and a database named db-prostate. Sybyl software has been noted to be suitable for this QSAR technique and for creating a database in it.

3.7.4 Template structure

The derivative having the highest PIC50 and with the lowest IC50 value here, it has been noted that the derivative named 52 has these properties. This has been selected as a common substructure. It has the following common structure shown in the figure in sybyl.

3.7.5 Ligand-Based Model

To develop a 3D QSAR model another step is alignment, probably the most crucial requirement for 3D QSAR study. There are different methods of compound alignment such as Structure based and Ligand based. In this study, substructure-based molecular studies are used. Before alignment, a database of our compounds (test and training) was made. Alignment is done based on the lowest energy conformation or the "bioactive conformation". Alignment of the database of test and training set compounds was done based on the common substructure shown and the template compound. All the compounds were fitted to the template molecule which is 52. Template compound is selected based on the highest activity which in in this study is molecule 52 exhibiting greater pIC50.

3.7.6 Alignment of database

In this, we aligned our database and selected the derivative having the highest PIC50 and the lowest IC50 value. The aligned database has been named align-db-popstarate. We opened and aligned all other structures in that template and then created the aligned database based on the common substructure.

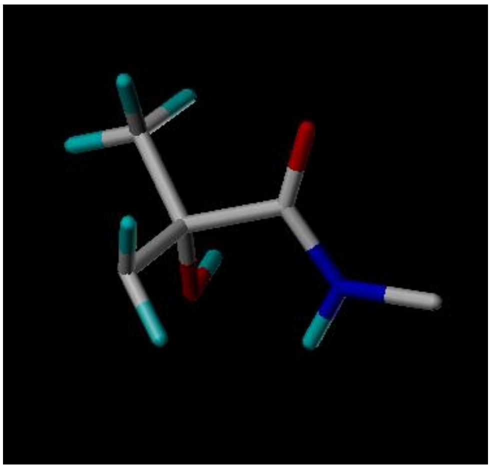


Fig 3.4: Common Substructure of all derivatives

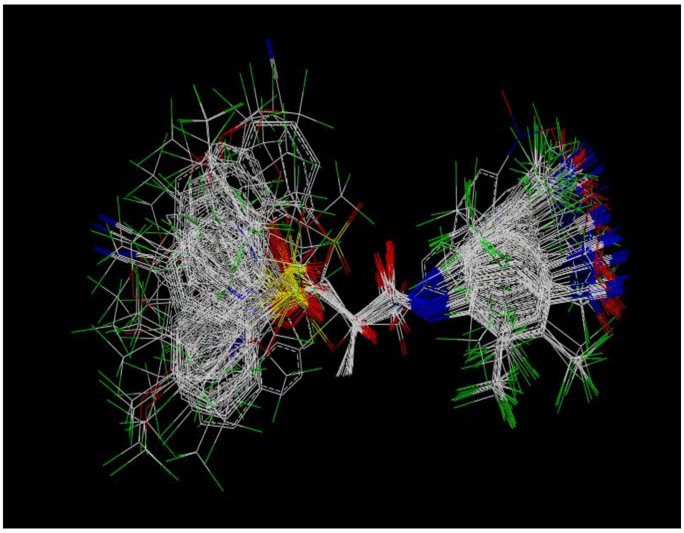


Fig 3.5: Ligand Based Alignment of Datasets minimized with Omega Tool

3.7.7 Assigning Charges

Alignment of inhibitors was done based on the lowest energy conformation for ligand-based model generation. After database alignment, the next step is model building. For designing ligand-based models different types of charges such as Gasteiger Huckle, Gasteiger Marsilli, MMFF94 Partial Charges and Pullman charges were assigned to select the model with the best and most accurate results because different charges yield different results for developed models. Certain parameters which were used to generate a ligand-based and receptor-based model are as follows: "N =optimal number of components; q 2 (LOO) = Standard cross-validated correlation coefficient; r 2 (NoV)=determination coefficient; SEE = non-cross-validated standard error; F = Fischer's F-value; pred r 2 =Predictive r 2; S = steric, E = electrostatic, H = hydrophobic field; D= donor ; A= acceptor; r 2 bs = r 2 obtained after bootstrapping for 100 runs; SDbs = bootstrapping standard deviation; r 2 CV(mean) = mean r 2 of crossvalidation in 10 groups". The next step after charge assignment and molecular alignment is model generation.

3.7.8 Model Generation

3.7.8.1 CoMFA and CoMSIA

This is the last step of model generation. Our model is generated by CoMFA and CoMSIA. Now if we talk about CoMFA which is Comparative Molecular Field Analysis, we calculated two interaction values in CoMFA which are steric and electrostatic. Steric Interaction is calculated based on Lennard-Jones potential whereas electrostatic interaction is calculated based on Coulombic potential. In SYBYL software, the grid spacing is generated automatically for the calculation of CoMFA fields. Now the next one is CoMSIA which is Comparative molecular similarity index analysis, here we not only calculate the values of steric and electrostatic fields but also computed the values of some other force fields which are hydrophobic, hydrogen-bond donor and hydrogen-bond acceptor fields. All numerical calculations were executed in a similar way as for CoMFA analysis. CoMSIA does not need any strict cutoffs due to which many important data points may be eliminated and has a more natural display of contour maps. CoMSIA models also remain unchanged when the orientation of molecules is altered and are resistant to small changes in regions. This is the reason why CoMSIA models are believed to be superior to CoMFA models.

3.7.8.2 PLS and Predictive r 2

The PLS analysis that is Partial Least Squares is also known as regression analysis. This is very helpful for analyzing the QSAR model. Here the pIC50 values and CoMFA and CoMSIA descriptors were utilized as independent variables.

This methodology gives highly stable, precise and analytical models for descriptors. To select the model which probably has the maximum predictive values, we have performed the cross-validation analysis. The non-cross-validated analysis is performed when the optimum number of components is determined. The predicted values of the PRESS, F-value (Fischer's Test Value), r 2 cross-validated and standard error of estimate values were calculated by SYBYL The predictive power of the 3D QSAR models, built based on the training set is observed by our 19 test set compounds. The predicted activities of the test set through CoMFA and CoMSIA models are obtained by the command called Predict. The bootstrapping analysis is also performed for 10 runs to evaluate the effectiveness of the derived models. Predictive r 2 value helps to express the ability of two models (CoMFA and CoMSIA models). The cross-validated value is somehow equal to the predictive r2 value. It is based on the test set molecules. The predictive r 2 value is defined as:

$$\text{r2pred}=\frac{(SD-PRESS)}{SD}$$

In this equation, PRESS is the sum of the square deviation between the predicted and observed activities of the test molecules and the sum of the square deviation between the mean activity of the training set and the biological activities of the test set is shown by SD.

3.7.8.3 External and Internal Validation of 3D QSAR Models

A model is being evaluated either by external validation or internal validation. The most frequently used statistical measures which are used to evaluate the importance of the developed model are as follows: The minimum suggested values for a significant QSAR model are:-

r2: (coefficient of determination) > 0.7
q2: (leave one out, cross-validated) > 0.5
r2 (no validation) > 0.5
Predictive r2: (r2 for external test set) > 0.5
SEE: (standard error of estimate) smaller is better
F-test: (F-test for statistical significance of the model) higher is better

3.7.8.3.1 Partial least squares analysis

Internal validation is done by the PLS method which is used to associate the inhibitory pIC50 values as dependent variables to the CoMFA and CoMSIA descriptor fields as independent variables.

3.7.8.3.2 Leave One Out

The value of standard q 2 was obtained by using the leave-one-out (LOO) method in which the resultant model is built by removing one compound and this new model is then used for predicting the activity of the excluded compound. The process is repetitive until all the compounds have been removed one by one from the complete dataset.

3.7.8.3.3 Non-Cross Validation

The optimal number of components, variance r2, standard error of estimate (S) and F ratio were obtained by a non-cross-validation analysis. The non-cross-validated analysis is also used to make predictions of the pIC50 values of the training set by the test set.

3.7.8.3.4 Bootstrapping

For internal validation of a generated 3D QSAR model, bootstrapping is also used. Here compounds are randomly chosen from the whole dataset and random sub-samples from the whole dataset are repetitively evaluated. The features of the samples that are not included are predicted from the randomly chosen compounds. The greater value of bootstrap validation tells that the model is robust.

3.7.8.3.5 Cross-Validation

Cross-validation is another method for internally evaluating a 3D QSAR model. It gives the value of r 2 which is generally less than the overall r2. It is also helpful in validating the predictive power of a model. Here the subsets of the whole dataset are repeatedly treated with regression. A molecule is removed only once per turn and predicted values of that removed molecule are being utilized to calculate the value of r2.

3.7.8.4 External Validation

For externally evaluating the 3D QSAR-developed model, the predictive ability of the model is determined by a test set consisting of 19 compounds. External validation is necessary because ambiguity is still there even if the results of internal validation are of high quality. Many authors also recommended that the predictive ability of the generated model be evaluated by relating the actual and predicted pIC50 values of the test set structures. External validation is done by the same method by aligning the test set compounds with the training dataset and then their inhibitory activities are predicted by the generated 3D QSAR model. A good external predictability value for predicted r2 should be > 0.6. Finally, the newly designed compounds were also tested and included in the predictive r2 calculations.

Table 3.3: 3D QSAR model of LNCaPcell line results with MMFF94Charges

MMFF94Charges	LigandBased Model (LNCaP)	
	85 Compounds	
Parameters	COMFA	COMSIA
N	1	1
q2(loo)	0.154	0.160
r2(nov)	0.859	0.782
SEE	0.126	0.157
F	73.265	43.025
Pred-r2	0.373	0.345
Steric (S)	0.379	0.109
Electrostatic (E)	0.621	0.348
Hydrophobic (H)	-	0.226
Donor (D)	-	0.137
Acceptor (A)	-	0.180
r2bs	0.926	0.871
SDbs	0.024	0.035
r2cv	0.163	0.153

Table 3.4: 3D QSAR model of DU-145cell line results with MMFF94Charges

MMFF94Charges	LigandBased Model (DU-145)	
	85 Compounds	
Parameters	COMFA	COMSIA
N	4	5
q2(loo)	0.282	0.260
r2(nov)	0.921	0.860
SEE	0.076	0.102
F	138.970	73.516
Pred-r2	0.558	0.501
Steric (S)	0.429	0.109
Electrostatic (E)	0.571	0.352
Hydrophobic (H)	-	0.226
Donor (D)	-	0.157
Acceptor (A)	-	0.156
r2bs	0.956	0.917
SDbs	0.014	0.022
r2cv	0.280	0. 268

Table 3.5: 3D QSAR model of VCaP cell line results with MMFF94Charges

MMFF94Charges	LigandBased Model (VCaP)	
	85 Compounds	
Parameters	COMFA	COMSIA
N	1	1
q2(loo)	0.281	0.223
r2(nov)	0.929	0.820
SEE	0.085	0.158
F	157.477	72.180
Pred-r2	0.569	0.820
Steric (S)	0.384	0.093
Electrostatic (E)	0.616	0.349
Hydrophobic (H)	-	0.246
Donor (D)	-	0.120
Acceptor (A)	-	0.92
r2bs	0.956	0.885
SDbs	0.017	0.026
r2cv	0.274	0.209

Table 3.6: 3D QSAR model of 22Rv1 cell line results with MMFF94Charges

MMFF94Charges	Ligand-Based Model (22Rv1)	
	85 Compounds	
Parameters	COMFA	COMSIA
N	4	5
q2(loo)	0.365	0.430
r2(nov)	0.956	0.872
SEE	0.062	0.105
F	215.115	81.788
Pred-r2	0.561	0.550
Steric (S)	0.398	0.120
Electrostatic (E)	0.602	0.358
Hydrophobic (H)	-	0.247
Donor (D)	-	0.121
Acceptor (A)	-	0.153
r2bs	0.955	0.917
SDbs	0.014	0.021
r2cv	0.355	0.346

Out of 85 derivatives 19 are in the test set i.e: 22h,22o,25f,25h,25l,26c,26i,27o,28e,28m,29e, 29i,29o,32o,34d,35o,52,25p and 35h and remaining 66 are in the training set. By applying distinct charges, we check the outcomes of our model. The better model generated by CoMFA and CoMSIA was attained through MMFF94 Charges. Different parameters were used for the evaluation of our 3D QSAR model like Cross Validated Coefficient, Standard Error Estimate and F-Statistic Values. Our results of the CoMSIA and CoMFA models are based on Ligand Models. The ligand-based modelling of cell line 22Rv1 is the best among all the cell lines i.e.: DU-145, LNCaP and VCaP. The values of q2 (r2 cv)= 0.355 for the CoMFA model and q 2 (r 2 cv)= 0.346 for the CoMSIA model.

3.7.8.5.1CoMFA Model

PLS analysis was done, and the result of the best cell line among all 4 is listed in Table 3.6. Our results show that the model generated by CoMFA has cross cross-validated value is 0.355 and a non-cross-validated value is 0.956. The value of F is 215.115 and the value of the standard estimate error is 0.062. Now the steric field value is 0.349 whereas the electrostatic field value is 0.062. This is the result of our model generated by CoMFA through our 85 derivatives.

3.7.8.5.2 CoMSIA Model

Similar to the CoMFA model, PLS analysis for the CoMSIA model was also done for the training set. The results of the best cell line among all 4 are also listed in Table 3.6. Our results show that the model generated by CoMSIA has cross cross-validated value is 0.346 and a non-cross-validated value is 0.8725. The value of F is 81.788 and the value of standard estimate error is 0.105. Now the steric field value is 0.120 and the electrostatic field value is 0.358. Furthermore, In CoMSIA, we have some other fields also so the value of Hydrophobic is 0.247, the value of hydrogen bond donor is 0.121 and the value of hydrogen bond acceptor field is 0.153. All the fields for CoMFA and CoMSIA models add up to 1 separately. These results indicate that the CoMSIA and CoMFA models are reliable. The predicted inhibitory activities are also listed in Table 3.6. All the results demonstrate that the CoMSIA model is also fairly predictive.

4. Discussions and conclusion

Prostate cancer is the cause of men's death worldwide and it is most frequently spreading in Europe as well. It is mostly among 67-74 aged men. It depends on a number of factors and has different stages accordingly. Treatment can be surgery, chemotherapy or hormonal therapy as well. Prostrate cell growth is strongly dependent on androgens, which are present in males as a hormone. Stopping the androgen function or say that inhibiting it would be very beneficial for the patient to survive [15]. Anti-androgen drugs have been designed and more efficient drugs are still being designed to resolve this issue. Flutamide, nilytamide, biaculamide, enzalutamide are all non-steroidal drugs which have been approved for the treatment of prostate cancer. The main issue with the androgen receptor here is a mutation, which causes the drug to function as an agonists [16]. Among the drugs, Bicalutamide and enzalutamide have been selectively used to block the androgen receptor. The most sensitive cell lines for prostate cancer to androgen hormones are 22RV1, DU145, LNCaP, and VCaP. The most sensitive of these four is LNCaP according to the research work [17]. We have used these four cell lines along with the inhibition values of Bicalutamide derivatives.

Bicalutamide derivatives have been used computationally to build models to analyze the statistical values. Eighty-five derivatives have been derived with the QSAR technique through 2D-QSAR and 3D-QSAR [18]. First, the 2D-QSAR database was made in which a total of 192 descriptors have been calculated in MOE software. After the calculation, many of the descriptors had the value 0 which was removed from the database and many of them had the same value which was also has been removed. The results before pruning and after pruning for cell line LNCaP and VCaP were not changed and were the same as 0.99 which was a good value. Most of the good results are better at 0.9 value for r2 [19].

The most critical step in the QSAR was energy minimization various software and outputs have been observed for it. It was done through MOE software and OMEGA on the Linux operating system through commands. The results with the minimization after in MOE showed value low value of q square through PLS (partial least square) in the software sybyl for which we had to minimize the energy in OMEGA as well. Again the values and PLS were done in which the value was no more than 0.3. The values with the minimization with MOE were still better than OMEGA. OMEGA energy minimization did not show any improvement and this might be the reason for OMEGA failure [20]. To make the value of q square, molecular docking studies were done using the maestro software. In molecular docking studies, the basic purpose of it is to predict the appropriate binding site of the ligand and bond all the derivatives [21] which with do the confirmation of the structure which might be the possibility to improve our model in sybyl software. The result for PLS and q square were still not more than 0.36 for the 22RV1 cell line using MMFF charges. From this, we concluded that the minimization from different software and even molecular docking had no major effect on the model building of the 3DQSAR technique. MOE-minimized derivatives were used because they had better results than omega-minimized derivatives and from maestro as well. Using MOE minimized molecules, the 22RV1 cell line was predicted to be better than the other 3 cell lines i.e. DU145, LNCaP, VCaP.

According to the research work of the European Journal of Medicinal Chemistry, Bicalutamide and enzalutamide have been proven to be better than other non-steroidal inhibition compounds for androgen receptors [22]. These 2D descriptors have given better results as compared with the 3D-QSAR model. However, some of the results for androgen receptor inhibition had better results with the 3D-QSAR technique [23].

2D-QSAR (r2) results were mostly noted to be greater than 3D-QSAR (q2) [24, 25] which may be because of the descriptors defined in the software. In 2DQSAR, sub-divided surface area descriptors, pharmacophore descriptor features and adjacency and distance matrix etc are favorable descriptors for structures of derivatives due to which results have increased in r2 [26]. The values of descriptors which were zero or the same in the column were eliminated due to which the best results of 2D QSAR are obtained as compared to 3D QSAR.

Bibliography:

1. Obeagu, E. I. (2023). Prevention and Early detection of Prostate Cancer. *Int. J. Curr. Res. Med. Sci*, *9*(7), 20-24.
2. Khodabandeh, M. (2024). Prostate Cancer and Quality of Life in the Elderly: A Literature Review. *Translational Research in Urology*, *6*(2), 76-83.
3. Hashemi, M., Zandieh, M. A., Talebi, Y., Rahmanian, P., Shafiee, S. S., Nejad, M. M., ... & Taheriazam, A. (2023). Paclitaxel and docetaxel resistance in prostate cancer: Molecular mechanisms and possible therapeutic strategies. *Biomedicine & Pharmacotherapy*, *160*, 114392.
4. Fleshner NE, Lucia MS, Egerdie B, et al. Dutasteride in localised prostate cancer management: The REDEEM randomised, double-blind, placebo-controlled trial. *Lancet.* 2012;379:1103-1111.
5. Akaza H, Hinotsu S, Usami M, et al. Combined androgen blockade with bicalutamide for advanced prostate cancer: Long-term follow-up of a phase 3, double-blind, randomized study for survival. *Cancer*. 2009;115:3437-3445.
6. Horoszewicz JS, Leong SS, Kawinski E, Karr JP, Rosenthal H, Chu TM, Mirand EA, Murphy GP (April 1983). "LNCaP model of human prostatic carcinoma". *Cancer Res*. 43(4): 1809–18
7. Woods-Burnham L[1], Basu A[1], Cajigas-Du Ross CK[1], Love A[1], Yates C[2], De Leon M[1], Roy S[3], Casiano CA[1,4]. The 22Rv1 prostate cancer cell line carries mixed genetic ancestry: Implications for prostate cancer health disparities research using pre-clinical models. 2017 Dec;77(16):1601-1608
8. Korenchuk, S; Lehr, JE; MClean, L; Lee, YG; Whitney, S; Vessella, R; Lin, DL; Pienta, KJ (2001). "VCaP, a cell-based model system of human prostate cancer". *In vivo (Athens, Greece)*. 15 (2): 163–8
9. Alimirah F, Chen J, Basrawala Z, Xin H, Choubey D (April 2006). "DU-145 and PC-3 human prostate cancer cell lines express androgen receptor: implications for the androgen receptor functions and regulation"
10. Gallagher, R., Fleshner, N. (1998) Prostate cancer: 3. Individual risk factors. Canadian Medical Association. 159(7): 807-813
11. Tan, M. H., Li, J., Xu, H. E., Melcher, K., & Yong, E. L. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacologica Sinica*, *36*(1), 3-23.
12. Clark, L., Combs, G., Turnbull, B., Slate, E., Chalker, D., Chow, J. (1996) Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skim. Journal of the American Medical Association. 267: 1957-1963

13. Redman, M., Tangen, C., Goodman, P., Lucia, M., Coltman, C., Thompson, I. (2008) Finsteride does not increase the risk of high-grade prostate cancer: a bias-adjusted modeling approach. Cancer Prevention and Research. 1(3): 174-181
14. Muratov, E. N., Varlamova, E. V., Artemenko, A. G., Polishchuk, P. G., & Kuz'min, V. E. (2012). Existing and developing approaches for QSAR analysis of mixtures. *Molecular informatics*, *31*(3-4), 202-221.
15. Mendelsohn, L. D. (2004). ChemDraw 8 ultra, windows and macintosh versions. *Journal of chemical information and computer sciences*, *44*(6), 2225-2226.
16. Real-Time PyMOL Visualization for Rosetta and PyRosetta Evan H. Baugh, Sergey Lyskov, Brian D. Weitzner, Jeffrey J. Gray. Published: August 16, 2011.
17. Human Androgen Receptor Inhibitors: Computational 3D QSAR Studies to Design Lead Compounds for Treatment of Prostate Cancer. Yayın tarihi 30 Eylül, 2013 © TurkJBiochem.com [Published online 30 September, 2013].
18. Youssef AM, Neeland EG, Villanueva EB, White MS, El-Ashmawy IM, et al. (2010) Synthesis and biological evaluation of novel pyrazole compounds. Bioorg Med Chem 18:5685-5696.
19. Structural basis for antagonism and resistance of bicalutamide in prostate cancer. Casey E. Bohl, Wenqing Gao, Duane D. Miller, Charles E. Bell, and James T. DaltonPNAS April 26, 2005. 102 (17) 6201-6206;Edited by John Kuriyan, University of California, Berkeley, CA, and approved March 16, 2005.
20. Bhandari, S. V., Nagras, O. G., Kuthe, P. V., Sarkate, A. P., Waghamare, K. S., Pansare, D. N., ... & Belwate, M. C. (2023). Design, synthesis, molecular docking and antioxidant evaluation of benzimidazole-1, 3, 4 oxadiazole derivatives. *Journal of Molecular Structure*, *1276*, 134747.
21. Molecular modeling of heme proteins using MOE: Bio-inorganic and structure-function activity for undergraduates*S, Gigi B. Ray, J. Whitney Cook 03 November 2006.
22. Cher, ML, Bova, GS, Moore, DH, Small, EJ, Carroll, PR, Pin, SS, Epstein, JI, Isaacs, WB & Jensen, RH (1996). Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. Cancer Res 56: 3091–3102.
23. Hadzi-Djokic, J. (2024). Hormone Therapy for Advanced Prostate Cancer. In *Prostate Cancer: Advancements in the Pathogenesis, Diagnosis and Personalized Therapy* (pp. 295-324). Cham: Springer Nature Switzerland.
24. Vyas, V. K., Ghate, M., & Katariya, H. (2011). 2D and 3D-QSAR study on 4-anilinoquinozaline derivatives as potent apoptosis inducer and efficacious anticancer agent. *Organic and medicinal chemistry letters*, *1*, 1-11.
25. Veras, L. D. S., Arakawa, M., Funatsu, K., & Takahata, Y. (2010). 2D and 3D QSAR studies of the receptor binding affinity of progestins. *Journal of the Brazilian Chemical Society*, *21*, 872-881.
26. Jain, S. V., Ghate, M., Bhadoriya, K. S., Bari, S. B., Chaudhari, A., & Borse, J. S. (2012). 2D, 3D-QSAR and docking studies of 1, 2, 3-thiadiazole thioacetanilides analogues as potent HIV-1 non-nucleoside reverse transcriptase inhibitors. *Organic and medicinal chemistry letters*, *2*, 1-13.