IN SILICO APPROACH TO FIND POTENTIAL INHIBITORS OF P-GLYCOPROTEIN FROM PIPERAZINE DERIVATIVES

by

Abhijit Kapare

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Masters in Bioinformatics and Computational Biology

Spring 2013

IN SILICO APPROACH TO FIND POTENTIAL INHIBITORS OF P-GLYCOPROTEIN FROM PIPERAZINE DERIVATIVES

by

Abhijit Kapare

Approved: _____
Dr. Arun Kumar, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Dr. Errol L. Lloyd, Ph.D.
Chair of the Department of Computer and Information Sciences

Approved: _____
Babatunde Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____
James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I would like to express my sincere thanks to my thesis advisor, Dr. Arun Kumar, who provided great knowledge and guidance throughout my thesis. I would like to thank him for all the support and advice he offered me throughout my thesis at the University of Delaware.

Special thanks to Virupaksha and Alpana Bastikar for their time and help in guiding me on the QSAR analysis and providing me the VLife MDS suite trial.

My genuine gratitude goes to Michael Wang, who helped me giving inputs about the thesis and Norman Goodacre, who guided me regarding the docking analysis and the tools.

I would like to thank Dr. Cathy Wu, Dr. Li Liao and Dr. Sandeep Patel for serving on my thesis committee and providing their insightful suggestions. I also thank all the CBCB staff and faculty members.

I am grateful to my friends Sanjeev Patra, Magdalena Altyn, Liang Sun, Vikrant Borse, Abhishek Kolagunda and my lab mates Weam Elbezanti and Konul Mahmudova for their constant encouragement and support.

Finally, I thank my parents and my sisters for their love and support.

ABHIJIT KAPARE

University of Delaware, Delaware

April 2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The development of new inhibitor possessing high efficacy, low toxicity and more potency is a pivotal approach to overcome p-glycoprotein (p-gp) mediated multidrug resistance (MDR) in cancer treatment. In this study, we performed drug-designing analysis to find the lead molecules that could inhibit p-glycoprotein (p-gp) by blocking its ATPase activity to counter MDR in cancerous cells.

We selected PB28 as the primary ligand, which is a dual inhibitor of p-glycoprotein and MRP1, another protein involved in MDR. Pharmacophore analysis gave idea about the moiety required to bind p-gp. Docking analysis suggested that few molecules have better docking scores than PB28. By Quantitative Structural Analysis Relationship (QSAR) analysis using multiple variable selection by simulated annealing method, we found the most dominant descriptors in analyzing the biological activity (EC50). We validated the 2D and 3D QSAR models on the 12 set of compounds and found that the activity predicted by the two models is almost similar.

The study showed that there is a possibility to find the better potent inhibitors than PB28. New piperazine derivatives could be tested in silico by the QSAR model generated by this study for analyzing the biological activity. The lead molecules could be tested on MRP1 and then subjected for clinical trials.

Chapter 1

INTRODUCTION

MDR is one of the important causes for the failure of the chemotherapeutic drugs in cancer. MDR or multidrug resistance is the mechanism by which the cell efflux the drugs out of the tumor cells making the cells resistant to the drugs. MDR involves proteins like MDR1 or p-glycoprotein or p-gp, MRP1 or multidrug resistance associated protein 1 and BCRP i.e. breast cancer receptor protein. P-gp is the most important protein involved in MDR. P-glycoprotein was discovered in 1970's after scientists first isolated the MDR Chinese hamster cell lines [17]. Since then the efforts to inhibit the p-glycoprotein have been significantly increased [27]. P-gp has two domains, transmembrane domain, where the drug or the substrate binds and the nucleotide-binding domain, where the ATP binds [32]. P-gp has open conformation facing the inside the cell and after the substrate binding, the protein conformation changes to facing outward of the cell by the ATPase activity to convert the ATP into ADP. Then the substrate is effluxed outside the cancerous cell. Another ATP is utilized to revert the conformation of p-gp back to its original shape i.e. open conformation towards the inner side of the tumor cell [40] [31] [14] [6] [7] [30]. Scientists are unable to understand the mechanism responsible for how ATPase activity is responsible to get the substrate out of the tumor cells. PB28 is a good inhibitor of p-gp as well as the MRP1 protein. There should be more emphasis on the dual activity of PB28 and its derivatives for the treatment of MDR.

1.1 Motivation:

MDR has become a real headache for the scientists. Recently very potent teriquidar was discovered and is in the clinical phases but it binds to just p-gp. In order to bind to the second most important MDR protein, MRP1, there is a need to find the good affinity ligand for both the proteins to inhibit, as just inhibiting p-gp won't be sufficient. PB28 could inhibit both the proteins as tested in vitro. By inhibiting both the proteins, MDR could be suppressed resulting in the killing of the tumor cells by the combination of the drugs, where one drug could inhibit the MDR and the other drug would act to kill the growing tumor and we could hope to cure the cancer. If we could be able to get more potent inhibitor of p-gp by analyzing the many piperazine derivatives, we could test them in silico by docking and QSAR analysis.

1.2 Objective:

- To develop a pipeline for in silico drug designing for p-glycoprotein and its inhibitors. Our pipeline was divided into 3 modules with 1st module involves selection of target as p-gp and ligand as PB28. Then 2nd module involves, docking of both the molecules, pharmacophore analysis, library generation based on the pharmacophore, batch docking, interaction study and the 3rd module involves the SAR and QSAR analysis of the molecules.
- To propose potential lead compounds for p-glycoprotein inhibition. Four potential leads were proposed using the analysis conducted in this thesis. Those leads were selected according to the docking analysis, 2D and 3D analysis and activity analysis.

1.3 Approach:

Ligand based approach for the drug designing was taken as we selected PB28 as our ligand and we have a single protein to bind to i.e. P-glycoprotein. All 22 ligands were docked with the p-glycoprotein and binding energy scores were compared. Then their QSAR analysis was performed and the QSAR model was evaluated by testing the model with the 12 sets of compounds with unknown EC50. The model was validated since both the models gave the similar results with some difference. 4 molecules selected as lead compounds.

1.4 Thesis Contributions:

The pipeline for the drug design to find the potent p-gp inhibitors was designed. By the docking analysis, found the 8 molecules having better binding energy than PB28 suggesting we could find the better inhibitors of p-gp by in silico methods. Pharmacophore mapping analysis gave the idea of the basic moiety or the features, which are the most important for the binding to the p-gp. This knowledge could be utilized to generate the new molecules. QSAR 2D and 3D models could be used to test new molecules with unknown EC50 values to get their biological activity. Generated 4 lead compounds as per the docking analysis, 2D and 3D QSAR analysis and the biological activities, which are all better than the PB28. In the future, the in vivo analysis for the clinical trials may suggest a potential drug to inhibit p-gp.

1.5 PB28 and piperazine derivatives analysis

PB28 is a cyclohexylpiperazine derivative, having a good binding affinity with the p-gp, binding energy is -2.33 as per VLife MDS software using genetic algorithm. Pharmacophore model of PB28 was generated and p-gp binding moiety was studied in details. ChemAxon's MarvinSketch was used to modify PB28 and generate structures [25]. The edited molecules were searched for analogs using the NCBI's PubChem [3]. 8 molecules out of the 21 selected molecules showed better binding energies than the PB28 molecule as per the docking analysis. All 22 molecules were subjected to the QSAR analysis using simulated annealing by multiple selection methods. 2D QSAR score was r^2= 0.8020, q^2= 0.6116. 3D QSAR analysis gave the results as r2= 0.7977, q2= 0.6358. The generated 2D and 3D models gave almost similar pEC50 values when tested on 12 different compounds with unknown EC50 values.

1.6 Thesis Outline:

In the next chapters, the terminologies used in the thesis are explained in the background and literature review chapter. In the methodology, the modules used in the drug designing pipelines are explained. In the results and discussion section, the docking between the PB28 and p-gp and then piperazine derivatives and p-gp is discussed. Interaction between the selected molecules and pb28 is explained. Pharmacophore mapping and 2D, 3D QSAR results are discussed in details. In the conclusion section, the explanation about the 4 lead molecules and why they were selected as the lead molecules is explained along with the future work.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

Cancer is a class of diseases so it's highly unlikely to have the single cure for cancer [4]. In cancer, cells grow and divide uncontrollably and which in turn forms malignant tumors and those tumors may migrate to the other parts of the body maybe by blood flow or the lymphatic system [19] [8]. Other type of tumor is benign tumor, which doesn't migrate to the other parts of the body nor is harmful [42] [15] [11]. Cancer over expresses some genes and represses certain genes in order to survive and multiply and to protect itself from the body's immune system and certain drugs [10]. Early the detection, more are the chances of cancer getting cured by chemotherapy [16].

A single drugs journey from the synthesis to the marketing takes many years. We could synthesize the drug in silico and test it using the computational bioinformatics tools and then simulate the lead molecules in silico for toxicity and ADME properties and then clinical trial simulations. The clinical trial simulations aren't so much predictive but in the future by taking the systems biology aspect and training the model with various sets over the time period, we could be able to propose a good predictive model for the clinical trial simulations, that could even save us more time and money. The cancer treatment takes a longer time and it needs a lot of money for the treatment. Million people die because of the cancer throughout the world. The US government spends millions of dollars in a year for the treatment of the various forms of cancers.

2.1 Cancer demographics:

The figure below represents the top 5 male cancer sites per 100,000 persons in the United States in 2009 in all the races. Prostate cancer was predominantly seen in the male cases followed by the lung then colon and rectum cancer and so on.

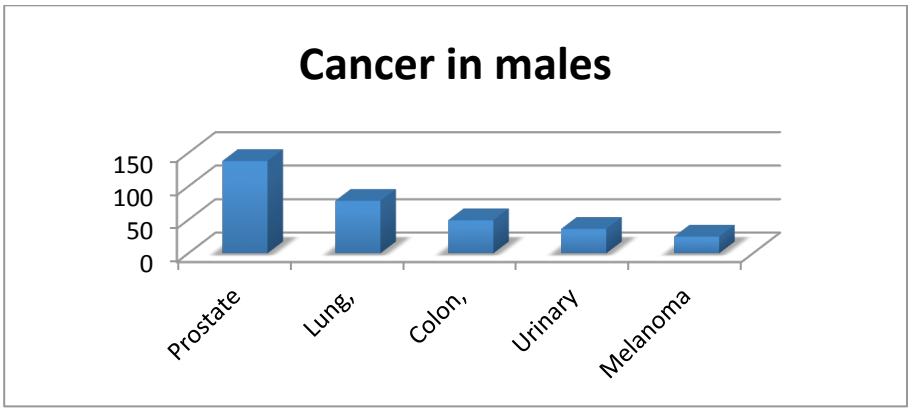


Figure 2.1: Top five cancer sites in males in United States in 2009 for all races [37].

In the figure below, top five cancer sites in females per 100,000 females in the United States in 2009 for all races are shown. We can see that the breast cancer in females followed by the lung cancer dominates the list and lung cancer was 2nd in male list too.

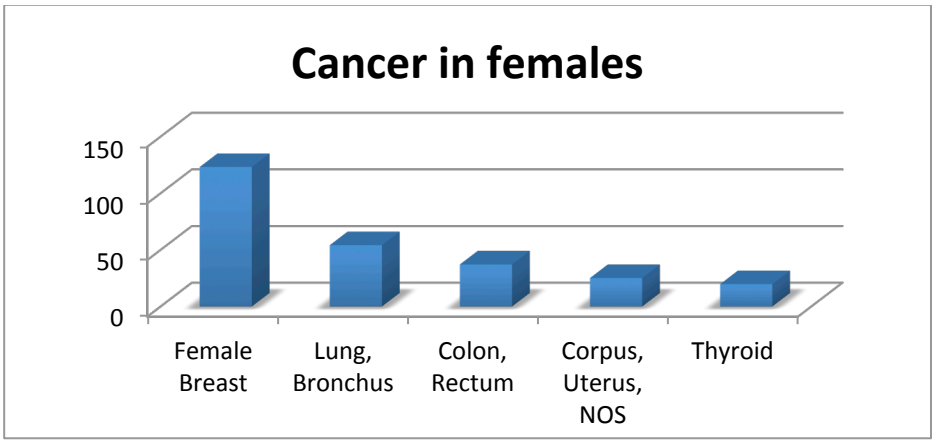


Figure 2.2: Top five cancer sites in female in United States in 2009 for all races [37]

The budget spent by the US government on the cancer is as explained in the following figure in millions of USD spent in the years 2008-10. MDR is involved in all above mentioned cancers like breast, prostrate, lung and colorectal cancers. Drug designing is important to overcome MDR in these cancers to make potent inhibitors.

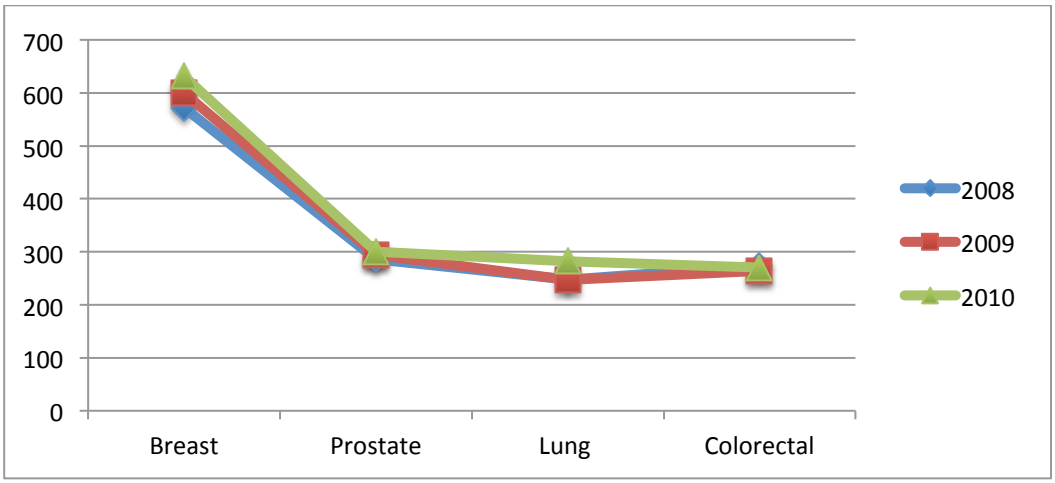


Figure 2.3: Top four type of cancer budget spent in millions USD from 2008-10 [5]

2.2 MDR

MDR or multidrug resistance in cancer is a mechanism by the cancerous cells where cancer cells become resistant to the chemotherapy drug. MDR can be seen in variety of cancers including breast, lung, ovarian, etc. A tumor may have some drug sensitive and some resistant cells. Which on later development, the whole tumor might become drug resistant. MDR in a cell efflux the drugs outside of the cancerous cell so that the tumor cells won't die due to the toxic effects of the drug. The most commonly found proteins involved in MDR are P-gp or MDR1 and multidrug resistance-associated protein or MRP1 and breast cancer resistant protein or BCRP.

2.3 P-glycoprotein

P-glycoprotein is a very important target for anti cancer therapeutics belonging to the ABC or ATP binding cassette superfamily of proteins. P-gp gets overexpressed in mammalian cell lines and human cancer cells [35]. It operates an ATP-driven pump to efflux the drugs outside of the cancer cells. P-gp genes has been sequenced and studied from human, hamsters and mice and its homologs have been identified in other species as well like drosophila, etc. Out of two isoforms of p-gp, MDR1, a class I isoform is involved in the transportation of the drug while MDR2, a class II isoform is involved in the export of phosphatidylcholine into the bile. MDR1 is a product of ABCB1 gene. Both isoforms share 78% amino acid sequence identity with each other.

2.3.1 P-gp presence in human tissues:

P-gp is expressed in large amounts in 1. Epithelial cell linings in adrenal glands, small intestines, pancreatic ducts as well as the 2. Endothelial cells of the blood brain barrier, blood testis barrier and the blood mammary tissue barrier. As p-gp is present in these many blood barrier membranes, the primary objective of the p-gp is to efflux the toxic xenobiotics outside the cells by the ‘pumping’ mechanism [34]. P-gp is overexpressed during the pregnancy in the endometrium of the secretory epithelial cells and placenta likely to provide the protection for the fetus. P-gp reduces the bioavailability and absorption of the drugs [20]. That is why the efforts to make a better anti cancer agents are proving faulty. There are substrates and the inhibitors of the p-gp. Substrates gets bind to the p-gp and are efflux out of the cancer cells [23].

2.3.2 P glycoprotein structure:

P-gp structure weight is 146992.00. Dr. Mi Sun Jin, et al. on 25th of October 2012, deposited 4F4C structure in PDB having 3.4 Aº resolution and published a paper in nature about their work. The first structure of P-gp, Dr. Aller S.G et al deposited PDB: 3G60 in March 2009 with the resolution of 4.4 Aº resolution, which was selected as the molecule of the month by the PDB. The in silico study to target p-gp was explored more after solving the p-gp structure. The better inhibitor for p-gp was important so scientists started simulating the inhibitory aspect of the lead molecules to inhibit p-gp in the cancerous cells. Two Membrane bound domains are present in p-gp and both are made up of 6 transmembrane (TM) helices and two cytoplasmic nucleotide-binding domains (NBD's). We can see the two chains A and B in the following figure.

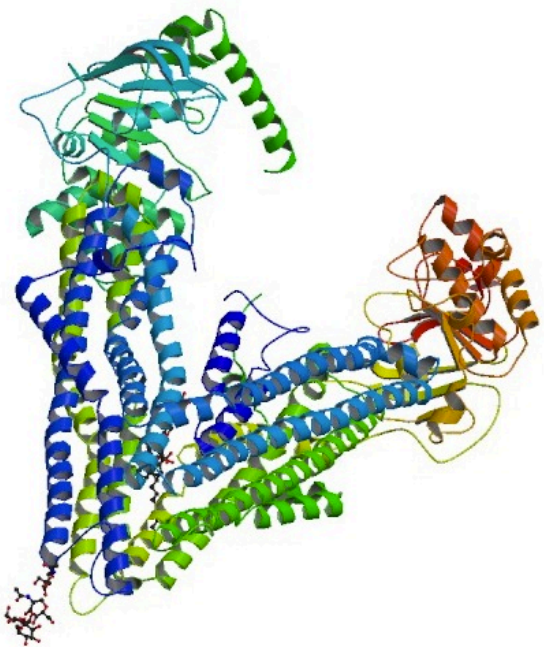


Figure 2.4: P-glycoprotein structure PDB id: 4F4C [26]

2.3.3 P-glycoprotein Activity

The drugs binds on to the interface between TMD's and the ATP binding sites are present on the interface of NBD's. The drug when binds to the substrate cavity in the p-gp, stimulates the ATPase activity and the ATP is converted in to ADP and that energy is utilized to efflux the substrate drug out of the cell. Hydrolysis of second ATP converts the p-gp back to the inward facing (open conformation in the intracellular region).

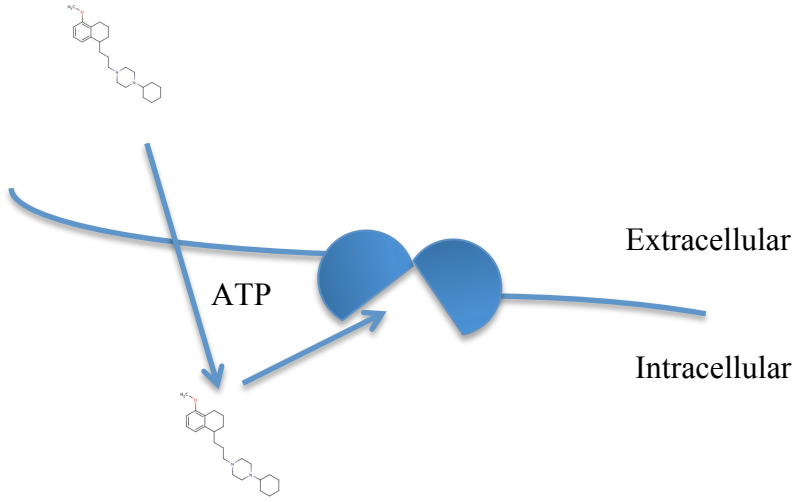


Figure 2.5: P-gp with inward facing configuration and drug entering in the cell

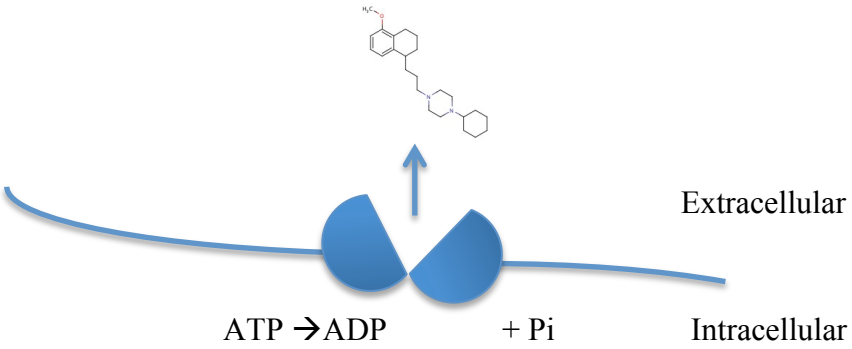


Figure 2.6: Outward facing p-glycoprotein conformation with PB28 being effluxed out of the cell by conversion of ATP to ADP by p-glycoprotein.

2.3.4 Inhibitors and substrates of p-gp:

Drug information could be accessed online from NCI, National Cancer Institute. The drugs clinical trials phase, use, side effects, company name, use could be seen from NCI website. P-gp can successfully bind to many up to 100 drug compounds like anticancer agents such as epirubicin, teniposide, paclitaxel, vincristine, imitinib, irinotecan, antidepressant like venlafaxine or HIV protease inhibitors like sequinavir, amrinavir, etc. While inhibitors of p-gp could be like antibiotics such as erythromycin, antiarrhymic agents like verapamil, anticancer drugs like doxorubicin, cyclic peptides like cyclosporine A, steroids like progesterone. Hence, there are a wide variety of the drugs that are either substrates or the inhibitor of p-gp. Inhibitors like teriquidar, PB28, binds to the p-gp and the p-gp's structure conformation mimic the conformation like when it is attached to the substrate but it couldn't efflux the drug outside the cell.

2.3.5 Generations of p-gp inhibitors:

In 1968, MDR chinese hamster cell lines were isolated for the first time. Since then scientists are working on that p-gp to find a potent inhibitor. Inhibitors could be classified into four generations. First generation in 1981, verapamil was discovered to have the reversal MDR effect. Second generation was Dexverapamil, which had lesser side effects, compared to the first generation inhibitors. Third generation inhibitors were teriquidar like, which are very potent inhibitors of p-gp even at the nanomolar concentrations. But, these inhibitors also have their side effects but are successive and are in the clinical phases [29].

Shung Sun, et al, 2000, took 24 breast cancer patients cells and found that the tumor cells in which there is a more than 10% expression of the p-glycoproteins, those cells have the presence of MDR. While the tumor cells which showed lesser than 10% expression of p-gp, didn't show MDR effects.

2.4 Drug Designing:

It is the designing of the drugs when you have the knowledge of the target [24]. Drug designing is of two types:

1. Ligand based drug designing: A known binding chemical compound with its binding groups to the target is known and by pharmacophore mapping the minimal characteristics in that compound, required to bind to the target are known. According to that, the compound is modified to get the better binding to the target molecule. QSAR or quantitative structure and activity relationships between the molecules and their biological activity would be derived and the model could be used to analyze the biological activity of the analogs [13].

2. Structure based drug designing: Its type of drug designing where, the designing relies of the knowledge of the 3D structure of the target for the analysis maybe from the experimental studies from X-rays or NMR. If the experimental structures are not available then homology model of that target is created using certain bioinformatics tools like MODELLER, etc. [22] [39] [33].

2.5 PB28

It is an agonist for sigma-2 receptor. It is a derivative of cyclohexylpiperazine.

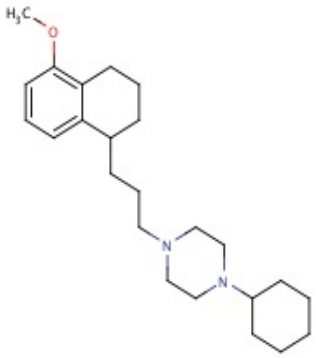


Figure 2.7: Structure of PB28

Reason to choose PB28?

MDR could be expressed by not only p-gp but also MRP1 and BCRP although p-gp is a prominent MDR agent out of all ABC transporters. Scientists are trying to look at the next generation of inhibitors like natural products and dual inhibitory ligands like PB28 and MC70, etc. These ligands have multiple inhibitory actions. For an instance, PB28 is involved in the inhibition of p-glycoprotein as well as MRP1 protein. Although it's not a potent inhibitor of p-gp as teriquidar, but it has a good affinity to bind and inhibit MRP1, unlike teriquidar. The PB28 has a piperazine scaffold and the derivatives of the PB28 could be tested and studied for their inhibitory action against p-gp and MRP1. MRP1 experimental structure is not available yet but homology modeled structure could be developed. PB28's EC50 and IC50 values are available to study the derivatives QSAR activity.

2.6 Piperazine Derivatives

In an experiment conducted by Carmen Abate et al., 2011 tested the p-gp inhibition treating with 17 ligands of piperazine derivatives including PB28. EC50 values were calculated in MDCK-MDR1 cells by calcein- AM assay [2]. Other experiment conducted by the same scientist, tested p-gp inhibition in dogs cells by performing calcein- AM assay [1]. We selected 5 compounds, which have piperazine derivatives. The EC50 values were taken from their experiments. Human and dog p-glycoprotein are homologous proteins, as per the ClustalW results [21].

2.7 Terminologies in the drug designing

- Docking: In molecular biology, docking is a method, which predicts the orientation of a drug or a small molecule on to the macromolecule or the protein.

- Interaction Analysis: Interaction study between the atoms of the ligands and the macromolecule or the protein with the bond length and the bond type analysis.

- Pharmacophore Analysis: Pharmacophore is the representation of the features of the ligand. It used to analyze the binding moiety of a compound to the macromolecule.

- QSAR: Quantitative Structure Activity Relationships as per the definition suggests, finds the relationship between the structural descriptors and the activity of that molecule on which a mathematical model could be generated to find the activity of the molecules having the same scaffold as the training set molecules. QSAR equation is as follows:

Activity = f(physiochemical properties/or structural properties) + error [36] [12] [9].

- Simulated annealing (SA) methods used to perform QSAR analysis was performed. SA follows the same principle of annealing in metallurgy where the materials are heated and cooled down slowly in a controlled manner, which increases the size of the size of the crystals and reduce the defects [18]. This overall heating and then cooling of the system affects both the temperature and thermodynamic free energy. Implementation of slow and controlled cooling in the simulated annealing is as seen as a slow decrease in the probability of accepting worst solutions while the algorithm explores the solution space.

- EC50: It is the biological activity of the molecule for a given protein. EC50 means half maximal effective concentration, which the drug induces, the response halfway between the baseline and the maximum after a specific exposure time. Using the EC50 values, we could find the drugs potency.

- Available QSAR software's: Very few software's provide the QSAR analysis like Accelrys Discovery Studio, SymBioSys eHiTS, Maestro by Schrodinger and VLife MDS by VLife Sciences.

- VLife MDS: The software has an academic license for the students on which one can perform the docking, pharmacophore and the QSAR analysis.

In the next chapter, we will see the methodologies to conduct the analysis. The drug-designing pipeline is explained in details about the data taken to conduct the analysis. The 3 modules of the pipeline are explained in the deep about the principles of those methods and how the methods are synced with one another.

Chapter 3

METHODOLOGY

Ligand based designing was chosen as the method for the design of the drugs as PB28 was the chosen ligand and p-glycoprotein was the chosen macromolecule on which the other piperazine derivatives would bind and the docking score would be calculated and then the QSAR analysis would be conducted. This chapter explains the pipeline in details about the methods used and how the pipeline flows. The drug-designing pipeline was divided into 3 modules as we took the ligand-based approach:

Module I comprises of the target selection and target optimization and existing drug selection and its optimization. We selected P-glycoprotein as out target and PB28 as the drug of preference.

Module II comprises of performing the molecular analysis of the protein and the drugs like docking, interaction, pharmacophore analysis. Then as per the results from the analysis, library of the ligands is generated and again batch docked with the protein.

Module III comprises of performing the SAR analysis by generating 2D and 3D descriptors of each ligands then QSAR analyze those descriptors from the SAR activity to find the dominant descriptors out of them. The QSAR models were tested on 12 compounds. Then, the potential lead molecules with best activity from module II and III were selected and those lead molecules were proposed for the future work.

The simulated annealing method used in the QSAR is explained in details with the parameters used in the 2D and 3D analysis. Future work is explained in the last chapter. Below is the flowchart of how the pipeline flows.

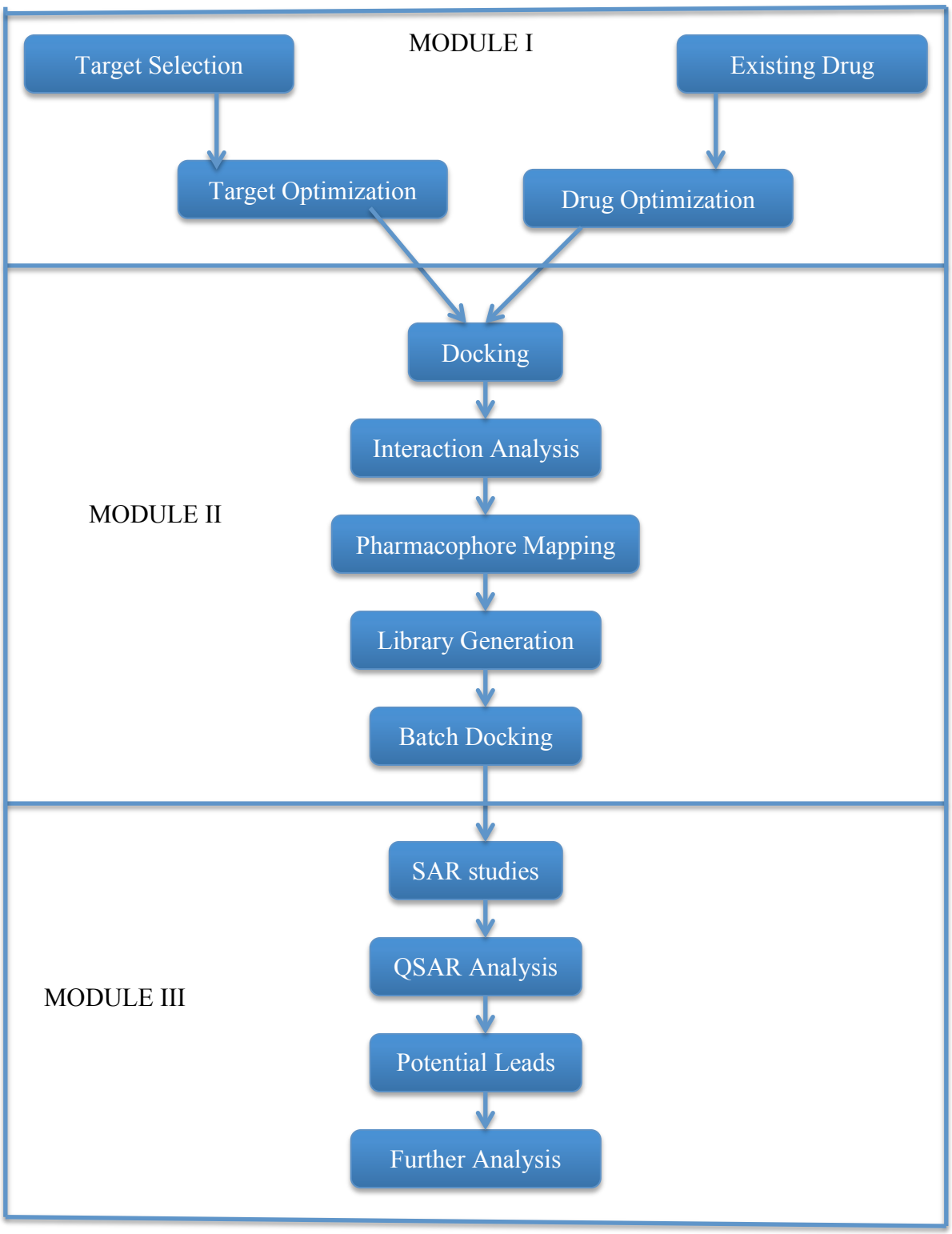


Figure 3.1: Schematic flow of drug designing Pipeline

3.1 Module I:

PB28 was optimized using VLife MDS software to make the bonds lengths and the bond angle and other force field parameters feasible to bind to the protein. P-gp was energy minimized by Swiss PDB viewer.

3.2 Module II:

Docking Analysis:

Target or in our case, p-glycoprotein and the drug, PB28 are docked together using Autodock Vina and VLife MDS software's and the binding energy was noted [28] [38]. The docking analysis of the two molecules gave the binding energy as -2.330642 as per the VLife MDS software calculations. The ligand and the target, both were flexible while docking and not rigid. Genetic algorithm was used for docking.

Interaction Study

The atoms of the ligands interacting with the amino acids of the proteins were studied. Van der Waals, hydrophobic, charge and hydrogen bond interactions were studied around the 5 Aº radiuses of the ligands atoms.

Pharmacophore Mapping

Pharmacophore mapping helped to show the hydrophobic, negative ionizable, steric or electrostatic nature of the ligands functional groups that are necessary to bind to a macromolecule or a protein [41].

Library Generation:

Piperazine was selected as the scaffold. PB28 molecule's pharmacophore analysis was important to analyze the moiety of the interaction. Piperazine derivatives were made using the ChemAxon's MarvinSketch by editing the structure of PB28. The analogues of the edited structures were found using PubChem online search. The selected 22 structures as our final set including PB28, all have piperazine scaffold. Edited 12 more test compounds to test QSAR models. Piperazine scaffold is shown in the following figure:

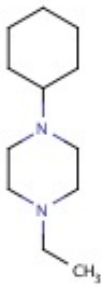


Figure 3.2: Piperazine scaffold

Batch Docking:

All the selected 22 analogs of the edited structures from PB28 were docked with p-gp using Autodock Vina and VLife MDS software. By comparison, the binding energy scores from VLife MDS were taken as they were mostly unchanged after testing one ligands docking analysis several times.

3.3 Module III:

QSAR

2D and 3D QSAR was performed. Here, the solution space is large; more than 492 descriptors of each molecule in 2D QSAR and 4056 descriptors for 3D QSAR. But, the objective function is complex. Hence, simulated annealing is the best method to find the most dominant descriptor.

Test and training set selection and r-squared and q-squared explanation:

Test set and the training sets were selected randomly. r-squared and q-squared analysis conducted to test the predictive ability of the QSAR model. r^2 if near to 1.0 means the variance in the dependent variable that is explained by the regression equation, that the actual points (here, ligands)]lie on the regression line. q^2 importance: The same data used to build the r^2, the same data is also used to evaluate it. Q squared, also known as cross-validated r- squared uses leave one out validation approach. So, q- squared predicts the values left out. Thus, q- squared is always lesser than r- squared.

2D QSAR Analysis:

In 2D QSAR, only 2D descriptor characteristics like molecular weight, slogP, hydrogen bond acceptor counts, rotating bond counts, etc. like this, 492 descriptors were analyzed. All these descriptors, which have the structural or the physicochemical significance, are used to predict the activity of that molecule.

The parameters for simulated annealing were as following:

Maximum temperature: 100ºC

Minimum temperature: 0.01ºC

Decrease in the temperature by: 10ºC

Iterations at given temperature: 5

Terms in model: 4

Perturbation limit: 1

Cross correlation limit: 1

Term selection criteria: r^2

Number of groups for cross validation: 15

Number of random iterations: 100

3D QSAR Analysis:

In 3D QSAR, only 3D descriptor characteristics like electrostatic and steric values at each point in a ligand molecule is calculated and related with the activity of that molecule by 4056 descriptors, The steric and electrostatic interference values should be lower or higher depending upon the chosen descriptors by the model as per the contribution of those descriptors in a given model.

In the next chapter, we will see the results after the above-mentioned methods are implemented. The binding energies after the docking analysis are explained in the tabular format.

Chapter 4

RESULTS AND DISCUSSIONS

The results from the methods followed as explained in the last chapter are explained in this chapter. The results are accompanied by the observations or the discussions and the interpretations.

4.1 Pharmacophore Analysis of PB28:

Color and their meanings of the pharmacophore analysis is explained below:

Table 4.1: Color and their meaning in the pharmacophore model

H bond donor: Lime Green color	H bond acceptor: Slate Blue color
Hydrophobic: Marigold color	Aliphatic: Peach color
Negative ionizable: Red color	Positive ionizable: Green color

The larger tessellated spheres are indicative of the common pharmacophores identified in the molecules, the smaller solid features are of the individual molecules. First benzene ring in piperazine is hydrophobic. Nitrogen's in the piperazine scaffold are hydrogen bond donors. Carbons between the two rings are negative ionizable.

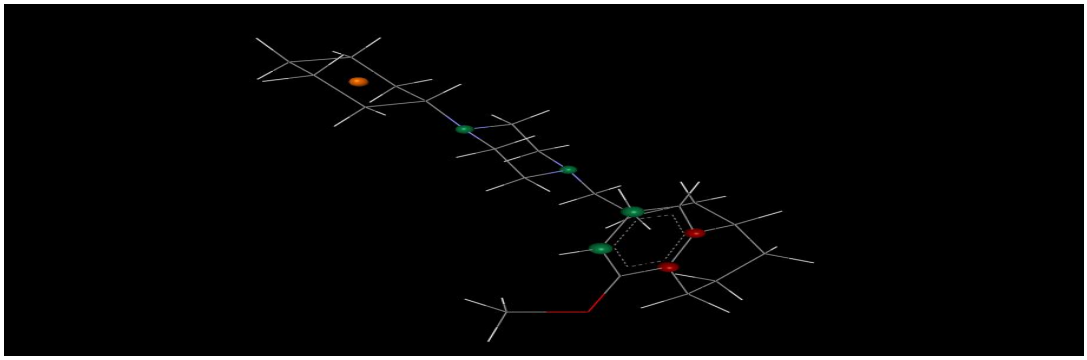


Figure 4.1: Pharmacophore analysis of PB28

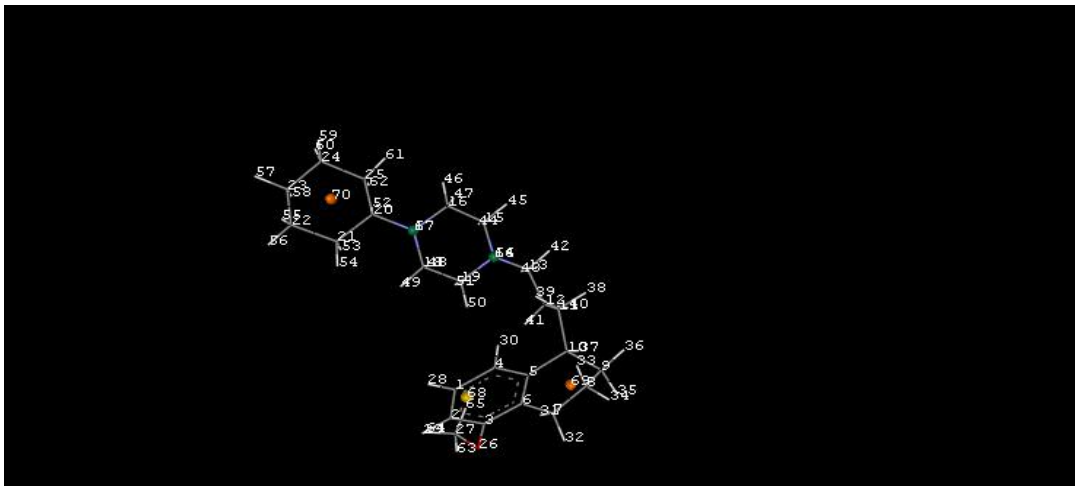


Figure 4.2: Numbering of atoms of PB28 molecule

22 total derivatives from online PubChem search were selected with their known EC50 values for conducting QSAR analysis later. The table below explains the meaning of the molecule number and its related CHEMBL id and its EC50 and pEC50 value, taken from the experiments.

Table 4.2: Molecule names with their CHEMBL id, EC50 and pEC50 values

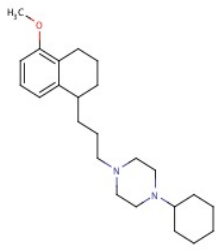
CHEMBL id	EC50(μM)	pEC50	CHEMBL id	EC50(μM)	pEC50
53325 = 25(PB28)	3.0	5.523	1672039 = 39	1.7	5.77
419822 = 22	3.8	5.42	1672040 = 40	5.2	5.284
1672030 = 30	8.8	5.056	1672041 = 41	9.1	5.041
1672031 = 31	6.6	5.18	1672042 = 42	3.6	5.444
1672032 = 32	8.1	5.092	1672043 = 43	3.4	5.469
1672033 = 33	10.0	5.0	1672054 = 54	4.4	5.357
1672034 = 34	9.8	5.009	1830685 = 85	8.6	5.066
1672035 = 35	3.2	5.495	1830687 = 87	2.7	5.569
1672036 = 36	8.8	5.056	1830689 = 89	45.0	4.347
1672037 = 37	2.4	5.62	1830691 = 91	8.6	5.066
1672038 = 38	5.3	5.276	1830692 = 92	20.0	4.699

4.2 Library Generation:

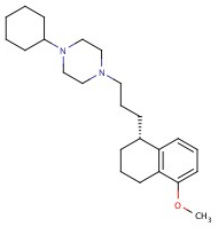
All the selected structures of derivatives of the piperazine could be seen below along with their numbers, which could be seen from the above table for their CHEMBL id's.

The Structures of the selected derivatives of piperazine are as follows:

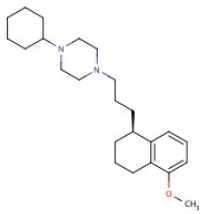
25



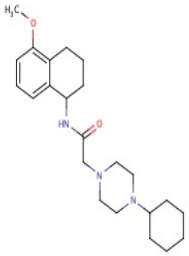
22



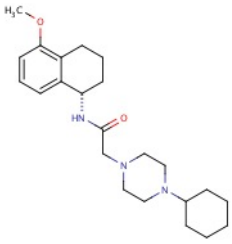
54



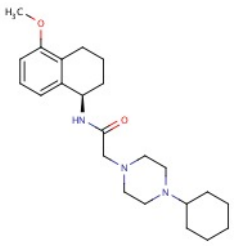
30



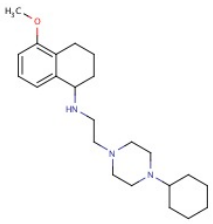
31



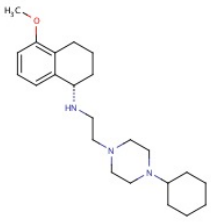
32



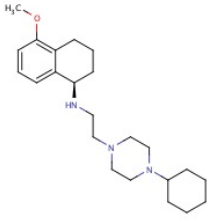
33



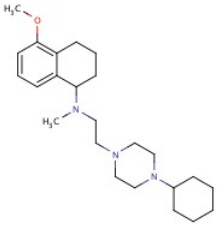
34



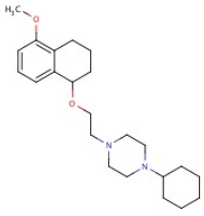
35



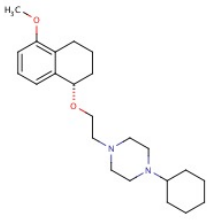
36



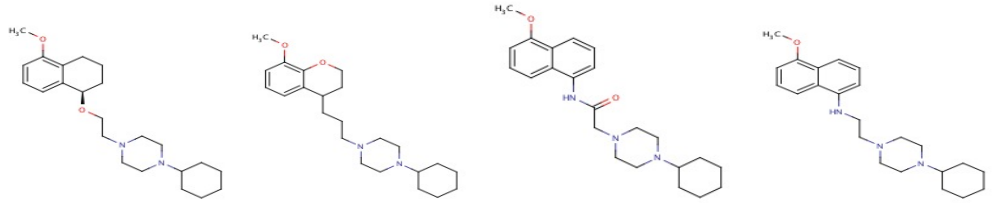
37



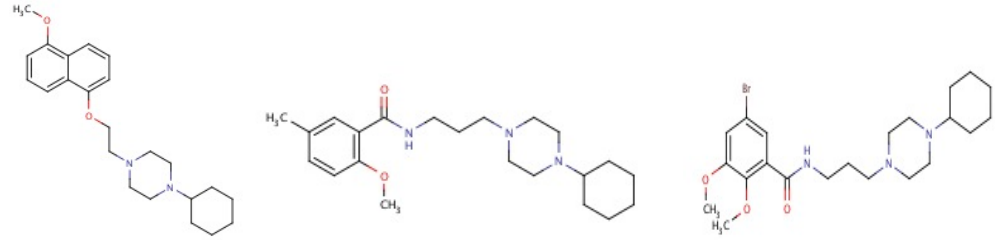
38



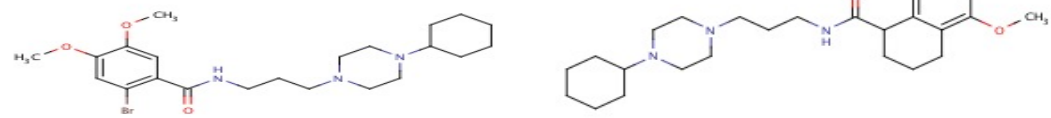
39 40 41 42



43 85 87



89 91



92

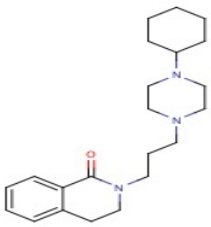


Figure 4.3: Structures of 22 selected ligands

4.3 Docking Analysis:

P-gp was docked with all the selected 22 ligands along with 12 test molecules. The molecules colored in red are the molecules having better binding energy than PB28. Those red colored molecules could make up to the list of lead molecules. Molecule number 25, PB28 was taken as a reference molecule to compare the binding energies of other molecules.

Table 4.3: Docking analysis of all 34 ligands

Molecule number	Binding Energy	Molecule Number	Binding Energy
25 (PB28)	-2.330642	85	-3.530098
22	-3.091955	87	-3.151811
30	-2.277992	89	-2.182393
31	-1.189853	91	-2.047511
32	-2.471250	92	-3.105163
33	-1.707106	1	-2.216708
34	-3.471642	2	-3.715548
35	-3.624880	3	-2.051371
36	1.182095	4	-2.608439
37	-0.897016	5	-1.465753
38	-2.568717	6	-2.139384
39	-3.154393	7	-2.462067
40	-2.689207	8	-2.836250
41	-3.520966	9	-3.866429
42	-1.899899	10	-4.148564
43	-1.660351	11	-4.656369
54	-2.907819	12	-1.495300

4.4 Interaction Summary:

Van Der Waals interactions are mostly between the amino acids and the C, H, O, N atoms of the ligands. Van der Waals interaction distance between the atoms of the ligand and the amino acids is mostly between 3.0 to 3.9 Aº. Hydrophobic interactions are only between carbon atoms of the ligands and amino acids of the p-glycoprotein and on an average they have interaction distances of about 2 to 4.9 Aº. Charge interactions are only between nitrogen atoms of the ligand and the amino acids and the interaction distance between 2 to 4 Aº.

Table 4.4: Interacting 22 ligands and amino acids of p- gp with their number and name

12- Arg	35- Tyr	258- Lys	315- Lys	327- Gln	806- Ala
16- Ser	247- Pro	259- Ser	316- Gly	328- Ala	807- Gln
23- Asp	248- Ile	260- Met	317- Leu	329- Ser	808- Gly
24- Val	249- Gln	261- Ser	318- Phe	330- Asn	809- Ile
25- Leu	250- Ala	263- Phe	319- Leu	332- Ile	810- Cys
26- Lys	251- Leu	264- Ala	320- Gly	760- Thr	811- Ser
27- Thr	252- Cys	266- Arg	321- Ile	763- Ile	812- Phe
28- Ala	253- Gly	310- Lys	322- Ser	764- Gly	813- Leu
29- Ile	254- Phe	311- Ala	323- Phe	767- Ile	814- Met
30- Lys	255- Ala	312- Gly	324- Gly	803- Leu	815- Thr
31- Thr	256- Ile	313- Val	325- Ala	804- Ala	816- Phe
32- Val	257- Ala	314- Leu	326- Met	805- Ala	872- Phe

Below is the interaction table of PB28 and other ligands having better binding activity than PB28. The Van Der Waals (VdW), hydrophobic, charge and hydrogen bond interaction is studied. VdW interaction was between C, H, O and N atoms of the ligands while hydrophobic was between only carbons of the ligands and charge interaction was only between the nitrogen's and the amino acids of p-glycoprotein.

Legends for interactions: '+'= Van Der Waals, '!'= Hydrophobic, Red color: Charge, Yellow color: Hydrogen bond interactions, 'C'= Charge interactions only

Table 4.5: Interactions between PB28 (25) and other ligands with amino acids of p-gp

	25	22	34	35	39	41	85	87	92	2	9	10	11
24								+!					
27								!					
30													
31	+!												+!
32						!							
247		+											
248		+!											
249		+!						+!					!
250		+!											
251		+!										+	
252	+	+!						!				+!	+!
253	+	+!										C	+!
254												+	+!
255	+!	+!							+!			+!	+!
256	+!	+!				!			+!			+!	+!
257	+!												+!
258	+								+!			+!	+!
259	+!								+!				+!
260	+!					!			+!				+
261													
263			+!	+!	+	+							
266			+	+	+	+							
310			+!	+!	+!								
311			+!	+!	+!								
312			+!	+!	+!								
313			+!	+!	+!								
314	+		+!	+!	+!	+!							
315	+!		+!	+!	+!	+							
316	+!	+	+	+	+	+!		+					
317	+!	+	+!	+!	+!	+!	+	+!	+!	+	+!		
318	+!	+!	+!	+!	+!	+!	+	+!	+!		+!		+
319	+!	+!	+!	+!	+	+!		+!	+!		+!		+
320	+!	+!	+!	+!	+!	+!	+	+!	+!	+	+!		

Table 4.5 continued: Interaction between the ligands and the amino acids of p-gp

	25	22	34	35	39	41	85	87	92	2	9	10	11
321	+!	+!	+!	+!	+!	+!	+!	+!	+!	+!	+!	+!	+!
322	+!	+!			+	+!	+!	+!	+!	+	+!	+!	+!
323		+!				+!		+!	+		+	+!	+!
324	+!	!				+!	+!	+!	!		+!	+!	+!
325		+!					+!	+!			+!	+!	+!
326		+!						+!				+!	+!
327		C						+!				+!	+!
328							!	+!				+!	+!
329		+!						+!				+!	+!
330								+!				+!	+!
332												!	
760							+!						
763							!						
764							!						
767							!						
805							+!				+!		
806							+!				!		
807							+!						
808							+!	+!			+!		
809							+!			!	+!		
810							+!						
811							+!	+!					
812	+!	+!	+!	+!	+!	+!	+!	+!	+!	+!	+		
813							+!	+					
815			+!	+!	+!	+!							
816	+!		!	!	+!	!		+	!				

From the tables, it could be inferred that most of the ligands binds to the 310-330 number amino acids of p-gp and many ligands are binding to the 812 i.e. Phenylalanine. The other ligands, except 25 or PB28, are attaching to more amino acids than the PB28 itself, suggesting a better binding affinity of those ligands to p-gp.

4.5 QSAR analysis

4.5.1 2D QSAR:

Test and training set randomly selected with trying different combinations and we found that taking molecules, 22, 31, 32, 39, 54, 91 and 92 as a test set, gives the best r^2 and q^2. The results were as follows by Multiple Simulated Annealing (SA):

r^2= 0.8020, q^2= 0.6116, F test= 10.1262, pred_r^2= 0.7264

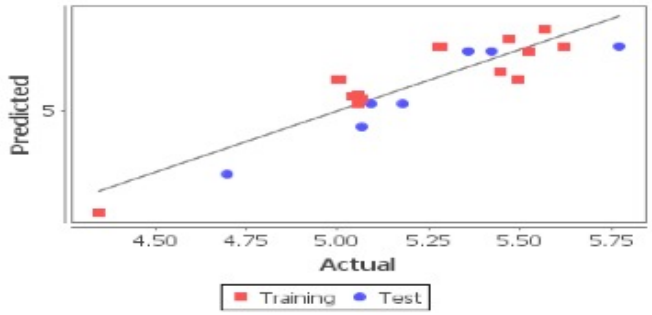


Figure 4.4: Regression analysis of 2D QSAR

Interpretation:

The training set is shown in red and the test set is shown in blue dots. There is not that much variance in the dependent variable between the test and training sets.

The dominant descriptor's, which are important in predicting the EC50, are as follow:

1. Nitrogen's count = Number of nitrogen's in a compound.
2. T_T_N_5 = any atom with any bond separated by nitrogen by 5 bonds in a molecule. (T=any)
3. T_N_O_2 = Count of nitrogen atoms with any bond separated by any oxygen atom by 5 bonds in a molecule.
4. chiV0 = It signifies atomic valence connectivity index (order = 0)

The biological activity (EC50) is dominantly dependent upon above descriptors out of the 492 calculated descriptors for 22 molecules.

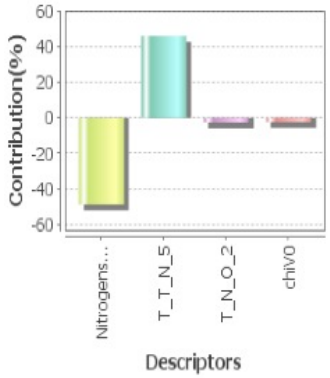


Figure 4.5: Contribution table for 2D QSAR multiple simulated annealing method

From the contribution table above, we can see that nitrogen's count is the most dominant descriptor according to the simulated annealing algorithm. Nitrogen's count should be lesser, T_T_N_5 should be higher, T_N_O_2 and chiV0 should be lesser in order to get the better biological activity.

Table 4.6: 2D QSAR model analysis of molecules with their descriptor values

	Nitrogen Count	T_T_N_5	T_N_O_5	chiV0
25	2	5	0	17.296
30	3	9	1	17.29
33	3	9	0	17.089
34	3	9	0	17.089
35	3	9	0	17.089
36	3	9	0	18.036
37	2	5	0	16.997
38	2	5	0	16.997
40	2	5	0	16.997
41	3	9	1	16.823
42	3	9	0	16.622
43	2	5	0	16.531
85	3	9	1	17.006
87	3	10	1	18.337
89	3	8	1	18.337

Discussion:

Nitrogen count isn't having a great variance but T_T_N_5 have a good variance and we can see that molecule number 87 has 10 T_T_N_5 number and greater than all of the 22 molecules means has 10 atoms in a molecule separated by nitrogen by 5 bonds.

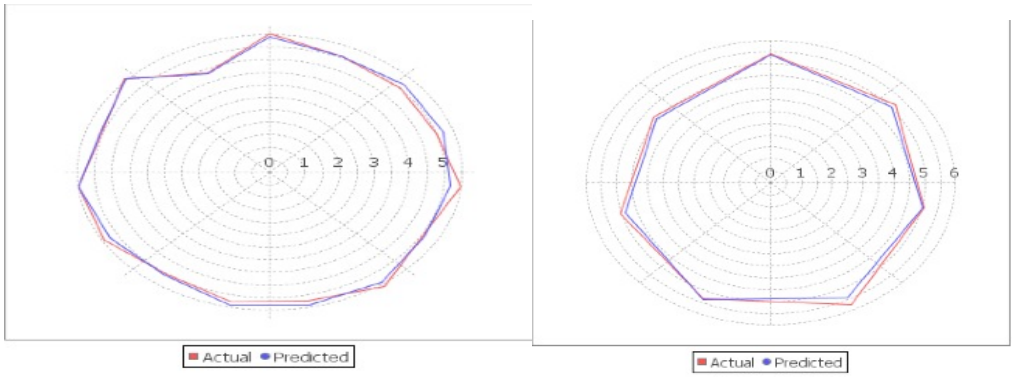


Figure 4.6: Training set pEC50 prediction (left) and test set pEC50 prediction (right)

Interpretation:

Red color indicates the actual pEC50 values while blue color indicates the predicted pEC50 values and both doesn't differ much so the model is validated. T_T_N_5 values if more than 8 then its is good. T_N_O_2 descriptor doesn't have a great variance in between the values. ChiV0 on the other hand is having a good variance but as we can see from the contribution table, the chiV0 values should not be so much lesser compared to nitrogen's count descriptor contribution. Different combination of training and test set used in order to get the better r-squared and q-squared values. Mostly the same descriptors as SA were selected in partial least square methods as model building methods with varying r-squared and q-squared. Which also validates descriptor selection by the simulated annealing method.

4.5.2 3D QSAR:

Just like 2D QSAR, we selected the variable selection method as simulated annealing (SA) and statistical model building method as multiple regression. SA parameters were same as 2D QSAR only this time term selection criteria were q^2 and number of random iterations were 500. Training and test set were randomly selected and test set is molecules 25, 34, 35, 40, 54, 89 and 91. Multiple simulated annealing results:

- r2= 0.7977, q2= 0.6358, F test= 9.8585, pred_r2= 0.6257

- Selected Descriptors: S_780, S_423, E_1189, E_5

S=steric hindrance: cost in energy due to overlapping electron clouds, which affects reactivity and conformation of the molecule.

E= Electrostatic: electrostatic force is the physical reaction that holds together the electromagnetic field created by subatomic particles, such as electrons and protons

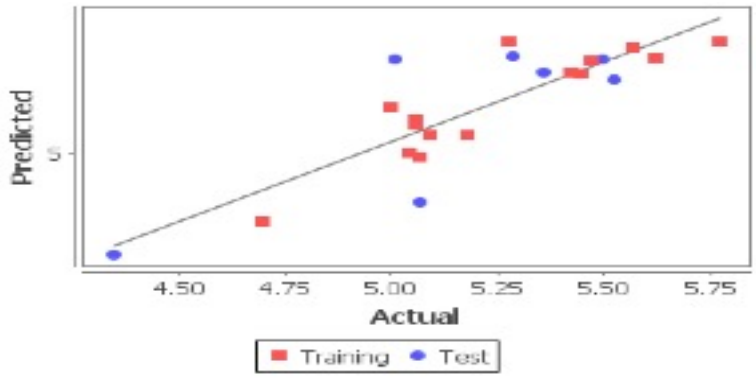


Figure 4.7: Regression analysis of 3D QSAR

Interpretation:

All points for training and test set lies on or near the regression line hence confirming that there is not that much variance between test and training set.

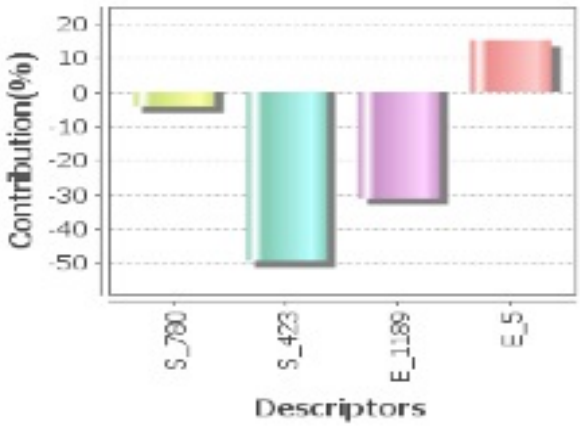


Figure 4.8: Contribution plot for 3D analysis

Observation:

From the contribution plot above we can conclude that S_780, S_423 and E_1189 should be lesser in a molecule and E_5 should be higher for better QSAR activity.

Table 4.7: 3D QSAR model analysis of training set molecules with descriptor values

	Steric-780	Steric-423	Electrostatic-1189	Electrostatic-5
37	-0.005	17.494	-1.729	-0.021
32	-0.005	30	-1.352	0.015
85	-0.004	18.128	3.433	0.029
42	-0.005	17.216	0.066	0.032
41	-0.005	30	0.561	0.08
39	-0.005	15.749	-0.816	0.042
36	-0.004	22.99	0.486	0.047
87	-0.004	12.36	-0.943	-0.031
92	-0.004	30	1.049	-0.067
33	-0.005	17.367	0.297	-0.042
43	-0.009	14.816	0.564	0.017
30	-0.005	30	-1.286	0.045
22	-0.005	12.052	1.568	0.031
38	-0.005	15.749	-0.816	0.042
31	-0.005	30	-1.352	0.015

Discussion:

Molecules 87 and 22, 43, 39 have the least S_423 values. Molecules 39, 87, etc. have the least E_1189 value. So, molecules like 87, 39 have good 3D QSAR activity.

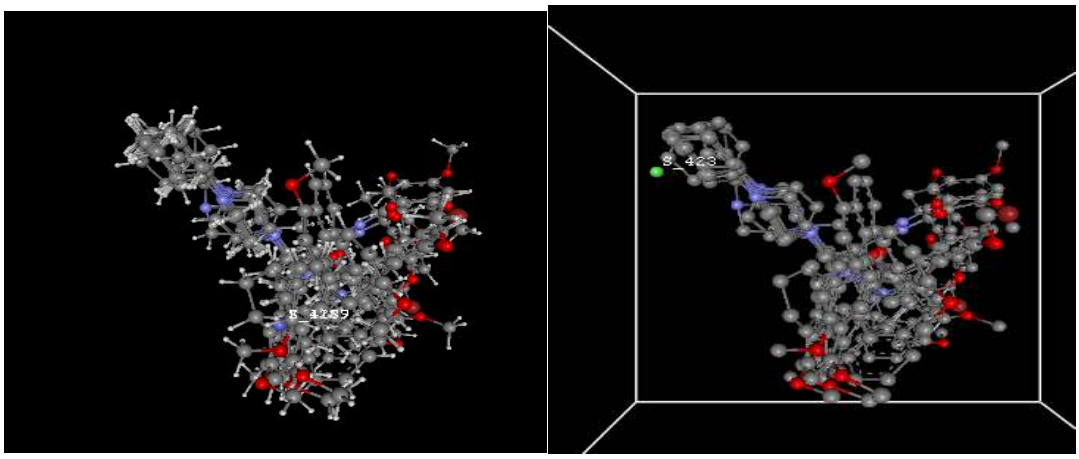


Figure 4.9: 3D points E_1189 and S_423 on the ligands

Steric-423 lies in the first ring of the piperazine scaffold and electrostatic-1189 point lies after the piperazine scaffold.

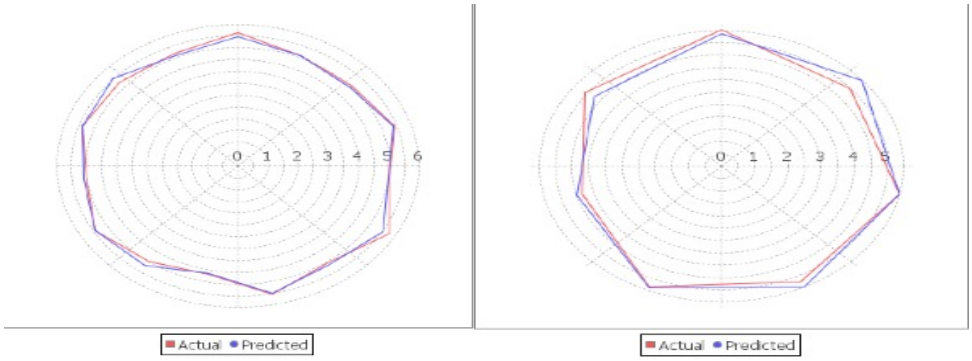


Figure 4.10: Training set pEC50 prediction (left) and test set pEC50 prediction (right)

Observation:

Red color indicates the actual pEC50 values while blue color indicates the predicted pEC50 values and both don't differ much like 2D QSAR so the model is validated.

Validation by other models:

Analyses of other variable selection methods performed in order to validate the simulated annealing model are as follows:

1. Partial least square method with forward variable selection method was used for QSAR analysis and it gave the following results:

$r^2 = 0.83$, $q^2 = 0.71$ and pred_$r^2 = 0.38$

It's pred_r-squared is lesser than simulated annealing method so we preferred simulated annealing and this method also showed the dominant descriptors as s_423, E_1189 and E_1177, out of which first two are common.

2. Multiple selection method was used with variable selection method as forward variable selection and it gave results as follows:

$r^2 = 0.7334$, $q^2 = 0.6728$ F-test = 12.8358 and pred_$r^2 = 0.6207$

It also gave the good score but its r-squared value wasn't better than the simulated annealing method. It also predicted s_1370, s_423 and s_109 descriptors as dominant descriptors, which are important in deciding the biological activity.

Validating 2D and 3D QSAR models:

From 2D and 3D prediction tables of EC50 of molecule's 1 to 12, the test molecules, both the models predict the biological activity almost similar or maybe just some differences. Extrapolation estimates the biological activity beyond the predicted EC50 and is derived on the basis of the descriptors relationship with another descriptor.

Table 4.8: 2D QSAR activity predictions of molecule number 1 to 12

	Nitrogen sCount	T_T_N_5	T_N_O_2	chiV0	Prediction	Extrapolation
1	2	5	0	16.335	5.504	-0.022
2	2	5	0	16.652	5.469	0
3	2	5	0	17.613	5.362	0
4	2	5	0	19.853	5.115	-0.168
5	3	10	0	16.445	5.894	-0.009
6	3	9	0	16.128	5.318	-0.045
7	2	5	0	17.666	5.357	0
8	2	5	0	18.627	5.25	-0.032
9	2	5	0	18.575	5.256	-0.026
10	2	5	0	16.97	5.434	0
11	2	5	0	17.022	5.428	0
12	2	5	0	17.983	5.322	0

Table 4.9: 3D QSAR activity predictions of molecule numbers 1 to 12

	S_780	S_423	E_1189	E_5	Prediction	Extrapolation
1	-0.002	-0.002	0.039	-0.009	5.841	-0.426
2	-0.009	-0.002	-0.084	-0.035	5.918	-0.397
3	-0.005	-0.241	0.52	0.053	5.963	-0.404
4	-0.014	6.786	0.412	0.028	5.84	-0.254
5	-0.001	-0.005	0.019	0.056	5.947	-0.442
6	-0.003	-0.022	-0.049	0.005	5.89	-0.413
7	-0.014	-0.004	-0.167	0.058	6.179	-0.475
8	-0.01	-0.016	-0.032	0.04	6.064	-0.407
9	-0.002	-0.012	0.29	-0.065	5.714	-0.424
10	-0.199	-0.031	0.041	0.02	9.189	-3.577
11	-0.028	30	0.286	-0.037	5.206	-0.317
12	-0.003	30	0.786	-0.009	4.781	-0.024

Predictions in pEC50 values used in the analysis as these are log normal values and could be converted to EC50 values using the equation:

pEC50 = -log(EC50)

Chapter 5

CONCLUSION AND FUTURE WORK

PB28 or molecule number 25 colored in red is placed at the top of above table as a comparison measure to compare other piperazine derivatives QSAR, docking and biological activities and we could see that molecule number 87, 10, 39 and 35 has better binding energy from docking analysis, better QSAR score and a good activities. These molecules could be selected as lead molecules. The '-' in the 2D QSAR and 3D QSAR analysis of molecule's 39 and 35 denotes that those compounds were taken as test set molecules to perform the QSAR analysis and hence, we didn't analyze their descriptor values performing their QSAR analysis.

Molecule number	Docking analysis	2D QSAR	3D QSAR	Activity (EC50)
25 (PB28)	-2.330642	Good	Good	3.0μM
87	-3.151811	Better	Better	2.7μM
10	-4.148564	Good	Better	3.0±0.6μM
39	-3.154393	-	Better	1.7μM
35	-3.624880	Good	-	3.2μM

Table 5.1: Lead molecules comparing the docking, QSAR and activity results

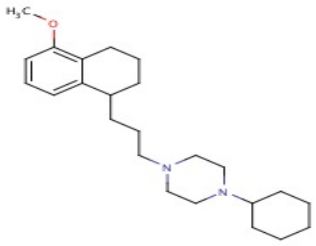


Figure 5.1: Molecule 25 (PB28)

IUPAC name: 1-cyclohexyl-4-[3-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl) propyl]piperazine

5.1 Lead Molecules

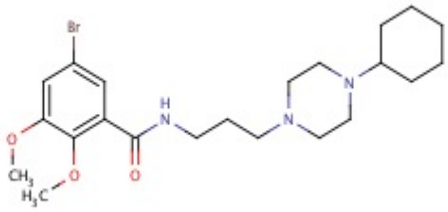


Figure 5.2: Molecule 87

IUPAC name: 5-Bromo-N-[3-(4-cyclohexyl-1-piperazinyl)propyl]-2,3-dimethoxybenzamide

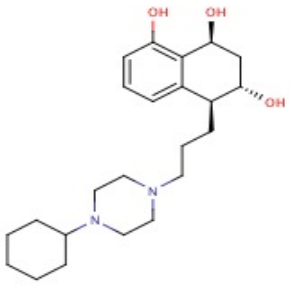


Figure 5.3: Molecule 10

IUPAC name: (1S,3S,4S)-4-[3-(4-cyclohexylpiperazin-1-yl)propyl]-1,2,3,4-tetrahydronaphthalene-1,3,8-triol

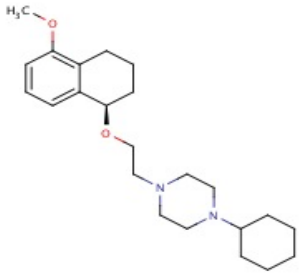


Figure 5.4: Molecule 39

IUPAC name: 1-cyclohexyl-4-[2-[(1S)-5-methoxytetralin-1-yl]oxyethyl]piperazine

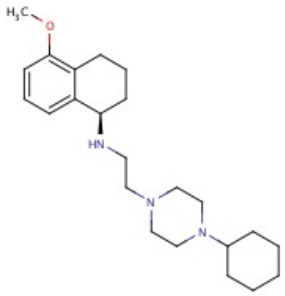


Figure 5.5: Molecule 35

IUPAC name: (1S)-N-[2-(4-Cyclohexyl-1-piperazinyl)ethyl]-5-methoxy-1,2,3,4-tetrahydro-1-naphthalenamine

Molecule number 10 is not present in PubChem. So, from chemicalize.org, the lead likeliness of this lead compound was validated with the following characteristics:

- It followed Lipinski's rule of five: yes
- Bioavailability: yes
- Lead likeness: yes

5.2 Future Aspects:

The four selected molecules, 87, 10, 39 and 35, which have better binding energy, QSAR and activities than PB28 molecule could be tested in wet lab analysis. The 2D and 3D QSAR model is well validated hence; these models could be used to test other piperazine derivatives to check their biological activity and develop more lead compounds which could be subjected to the in-vitro analysis which if would succeed could be tried in vivo in clinical phase trials and toxicological study and the best lead compound might go to the market as a final drug.

PB28 is also an inhibitor of MRP1 along with MDR1 or p-glycoprotein. The experimental MRP1 structure is not available in PDB yet but could be modeled using homology-modeling tools like MODELLER, free software or other commercial software's like VLife MDS etc. The leads could be tested on MRP1 for their binding energy analysis, QSAR analysis and biological activity (EC50 values). MODELLER was used for modeling MRP1 structure and the work could be continued further.

We can work on finding the entropy of the molecules to confirm our results. The descriptors from the QSAR study could be subjected to the brute force methods in order to find the most dominant descriptor responsible for analyzing the biological activity to confirm the results obtained by the simulated annealing results in both the 2D and 3D QSAR.

REFERENCES

[1]. Abate, C.; Ferorelli, S.; Contino, M.; Marottoli, R.; Colabufo, N. A.; Perrone, R.; Berardi, F. 2011. Arylamides hybrids of two high-affinity s2 receptor ligands as tools for the development of PET radiotracer. *Eur. J. Med. Chem.*, **46**, 4733-4741.

[2]. Abate C, Niso M, Lacivita E, Mosier PD, Toscano A, Perrone R. 2011. Analogues of sigma receptor ligand 1-cyclohexyl-4-[3-(5-methoxy-1,2,3,4-tetrahydronaphthalen-1-yl)propyl]pipe razine (PB28) with added polar functionality and reduced lipophilicity for potential use as positron emission tomography radiotracers. *J Med Chem*, **54**:1022–1032

[3]. Bolton E, Wang Y, Thiessen PA, Bryant SH (April 2008). PubChem: Integrated Platform of Small Molecules and Biological Activities. Chapter **12** IN *Annual Reports in Computational Chemistry*, Volume 4, American Chemical Society, Washington, DC.

[4]. Cancer Fact Sheet, *Agency for Toxic Substances & Disease Registry*. 30 August 2002. Retrieved 17 August 2009.

[5]. Cancer research funding, *National Cancer Institute*, reviewed 06/02/2011

[6]. Chang G (2003). "Multidrug resistance ABC transporters". *FEBS Lett.* **555** (1): 102–5.

[7]. Chang G, Roth CB (2001). "Structure of MsbA from E. coli: A homolog of the multidrug resistance ATP binding cassette
(ABC) transporters". *Science* **293** (5536): 1793–800.

[8]. Chiang AC, Massagué J (December 2008). "Molecular basis of metastasis". *The New England Journal of Medicine* **359** (26): 2814–23.

[9]. Chirico N, Gramatica P (August 2012). "Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection". *J Chem Inf Model* **52** (8): 2044–58.

[10]. Darrasse-Jeze G, Bergot AS, Durgeau A, et al (2009). Tumor emergence is sensed by self-specific CD44hi memory Tregs that create a dominant tolerogenic environment for tumors in mice. *J Clin Invest*; **119** :2648-2662

[11]. David Lowell Strayer; Raphael Rubin; Rubin, Emanuel (2008).*Rubin's pathology: clinicopathologic foundations of medicine*. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins. pp. 138–139.

[12]. Gramatica P (2007). "Principles of QSAR models validation: internal and external". *QSAR &Comb. Sci*.

[13]. Guner, Osman F. (2000). *Pharmacophore Perception, Development, and use in Drug Design*. La Jolla, Calif: International University Line.

[14]. Higgins CF, Linton KJ (October 2004). "The ATP switch model for ABC transporters". *Nat. Struct. Mol. Biol.* **11** (10): 918–26.

[15]. "How many different types of cancer are there? : *Cancer Research UK* : CancerHelp UK". Retrieved 11 May 2012.

[16]. Israel O , Kuten A (2007) . Early detection of cancer recurrence: 18F-FDG PET/CT can make a difference in diagnosis and patient care . *J Nucl Med* ; **48** (1): 28S – 35S.

[17]. Juliano R. L., Ling V. (1976) A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochim. Biophys. Acta* **455**,152–162.

[18]. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". *Science* **220** (4598): 671–680.

[19]. Klein CA (September 2008). "Cancer. The metastasis cascade". *Science* **321** (5897): 1785–7.

[20]. Kwon, H., Lionberger, R. A. & Yu, L. X (2004). Impact of P-glycoprotein-mediated intestinal efflux kinetics on oral bioavailability of P-glycoprotein substrates. *Mol. Pharm*. **1**, 455–465.

[21]. Larkin MA, Blackshields G, Brown NP, et al., (13 co-authors) (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*; **23** : 2947-2948.

[22]. Leach, Andrew R.; Harren Jhoti (2007). *Structure-based Drug Discovery*. Berlin: Springer.

[23]. Leonard, G. D., Fojo, T. & Bates, S. E (2003). The role of ABC transporters in clinical practice. *Oncologist* **8**, 411–424.

[24]. Madsen, Ulf; Krogsgaard-Larsen, Povl; Liljefors, Tommy (2002). *Textbook of Drug Design and Discovery*. Washington, DC: Taylor & Francis.

[25]. Marvin was used for drawing, displaying and characterizing chemical structures, substructures and reactions, *Marvin* 5.12.3, 2013, ChemAxon.

[26]. Mi Sun Jin, Micheal Oldham, et al (25 October 2012), Crystal structure of multiple transporter P-glycoprotein from Caenhorabditis elegans, *Nature* **490**,566–569.

[27]. Modok S, Mellor HR, Callaghan R (August 2006). "Modulation of multidrug resistance efflux pump activity to overcome chemoresistance in cancer". *Curr Opin Pharmacol* **6** (4): 350–4.

[28]. O. Trott, A. J. Olson (2010), AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* **31** 455-461.

[29]. Palmera Sousa, et al (2012), Three decades of p-gp inhibitor: skimming through several generations and scaffolds, *current medicinal chemistry*; **19**(13): 1946-2025.

[30]. Reyes CL, Chang G (2005). "Structure of the ABC transporter MsbA in complex with ADP•vanadate and lipopolysaccharide". *Science* **308**: 1028–1031.

[31]. Reyes CL, Ward A, Yu J, Chang G (February 2006). "The structures of MsbA: Insight into ABC transporter-mediated multidrug efflux". *FEBS Lett.* **580** (4): 1042–8.

[32]. Rosenberg MF, Velarde G, Ford RC, Martin C, Berridge G, Kerr ID, Callaghan R, Schmidlin A, Wooding C, Linton KJ, and Higgins CF (2001).Repacking of the transmembrane domains of P-glycoprotein during the transport ATPase cycle. *EMBO J* **20**: 5615–5625.

[33]. Schneider G, Fechner U (August 2005). "Computer-based de novo design of drug-like molecules". *Nat Rev Drug Discov* **4** (8): 649–63.

[34]. Siegsmund M, Brinkmann U, Schäffeler E, et al. (2002). Association of the P-glycoprotein transporter MDR1(C3435T) polymorphism with the susceptibility to renal epithelial tumors. *J Am Soc Nephrol* **13**:1847–1854.

[35]. Sun SS, Hsieh JF, Tsai SC, Ho YJ, Lee JK, Kao CH (2000). Expression of mediated P-glycoprotein multidrug resistance related to Tc-99m MIBI scintimammography results. *Cancer Lett*; **153**:95–100.

[36]. Tropsha A, Gramatica P, Gombar VJ (2003). "The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models". *QSAR &Comb. Sci.* **22**: 69–77.

[37]. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2009 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, *Centers for Disease Control and Prevention and National Cancer Institute*; 2013.

[38]. VLife MDS 3.5, "Molecular design suite. Vlife Sciences Technologies Pvt. Ltd.," Pune, India, 2008,*http://www.vlifesciences.com/*

[39]. Wang R,Gao Y,Lai L (2000). "LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design". *Journal of Molecular Modeling* **6** (7–8): 498–516.

[40]. Ward A, Reyes CL, Yu J, Roth CB, Chang G (November 2007)."Flexibility in the ABC transporter MsbA: Alternating access with a twist".*Proc. Natl. Acad. Sci.* U.S.A. **104** (48): 19005–10.

[41]. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA (1998). "Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)". *Pure and Applied Chemistry* **70** (5): 1129–1143.

[42]. Wilson, Kathleen Atkins; Waugh, Anne; Chambers, Graeme; Grant, Allison; Ross, Janet (2006). *Ross and Wilson anatomy and physiology in health and illness. Edinburgh*: Churchill Livingstone. pp. 53–54.

Appendix

LIST OF TOOLS AND SOFTWARES USED

Some trial versions and some full versions of the software's were used for the drug-designing pipeline.

Use and the description of those tools are as follows:

1. VLife MDS- It is a drug designing suite and is commercially available. We used VLife MDS for the 15 days trial to conduct the docking and QSAR analyses.

2. Autodock Vina: This software was used for docking study. It is freely available docking tool and is currently distributed by The Scripps Institute.

3. PubChem: It's is a database of chemical structures on NCBI website. It was used to retrieve the structures for the thesis.

4. Chemicalize.org: It stores the chemical information of many compounds. The leads compounds drug-likeliness was calculated using chemicalize.org

5. ChemAxon's MarvinSketch: It is a chemical structure-editing tool and is distributed by the University license for two years use. We used MarvinSketch to modify the ligands according to the pharmacophore information.