"3D QSAR studies, virtual screening and machine learning of novel Protein Kinase C derivatives to obtain new inhibitors for Cancer"

A DISSERTATION

SUBMITTED IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGY

IN

BIOINFORMATICS

Submitted by:

Rohan Ajit Singh

2K16/BIO/04

Under the supervision of

Prof. Pravir Kumar

Department of Biotechnology

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

JULY, 2018

Delhi Technological University

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, Rohan Ajit Singh, 2K16/BIO/04 student of M.Tech Bioinformatics, hereby declare that the project Dissertation titled **"3D QSAR studies, virtual screening and machine learning of novel Protein Kinase C derivatives to obtain new inhibitors for Cancer"** which is submitted by me to the department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without paper citation. The work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi

Date:

ROHAN AJIT SINGH

Department of Biotechnology
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled **"3D QSAR studies, virtual screening and machine learning of novel Protein Kinase C derivatives to obtain new inhibitors for Cancer"** which is submitted by Rohan Ajit Singh, 2K16/BIO/04, Department of Biotechnology, Delhi Technological University, Delhi in partial fulfillment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part of full for any Degree or Diploma to this University or elsewhere.

Place: Delhi

Date:

Prof. PRAVIR KUMAR
SUPERVISOR

ACKNOWLEDGEMENT

At the time of submission of my M.Tech Dissertation, I would first like to thank GOD for giving me patience, strength, capability, and willpower to complete my work. Apart from our efforts, the success of this project depends largely on the encouragement and guidelines of many others. I, therefore, take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

My initial thank is addressed to my mentor Prof. Pravir Kumar, Professor, Departmentof Biotechnology, Delhi Technological University, who gave me this opportunity towork in a project under him. It was his enigmatic supervision, constantencouragement and expert guidance which have enabled me to complete this work.I humbly seize this opportunity to express my gratitude to him.

I would also like to extend my sincere gratitude to Professor Jai Gopal Sharma for providing basic infrastructure and facilities.

I would like to extend my sincere gratitude to Dr. Rashmi Ambasta for her keen observation and continuous advice. Her thoughtful inputs on the project issues have been invaluable for the productive progress of the project. I extend my thanks to technical staff Mr. Jitender Singh and Mr. C.B. Singh who had been an aid whenever required. Lastly, I wish to extend my thanks to my family and friends who have supported me through the entire process.

ROHAN AJIT SINGH
2K16/BIO/04

ABSTRACT

Protein kinase C (PKC) family is a group of kinases that have always been the focus of pharmaceutical industries for the last two decades. These kinases are capable of modulating important cellular functions such as differentiation, proliferation and even cell survival. The over expression of novel class of PKC has been reported in many benign cancers which further leads to malignant if left unchecked. Therefore inhibiting n-PKC by effective compounds is necessity. In this study 3D-QSAR modeling was performed on a series of novel class of Protein Kinase C derivatives acting as inhibitors in various cancers. The compounds were collected from two datasets with the same scaffold, and utilized as a template for a new model to screen the ZINC database and Binding database of commercially available derivatives. The datasets were divided into training and test sets. As the first step, comparative analysis was conducted out by Swiss ADME and top features were selected to create similar inhibitors by Marvin JS package and molecular property and bioactivity scores were calculated. The constructed compounds were used as test set in machine learning to create various models. In parallel docking studies were applied to a set of known n-PKC inhibitors and constructed inhibitors. The validity and the prediction capacity of the resulting models were further evaluated. It is crucial for developing targeted therapy by use of specific inhibitors that have more rapid action than current available treatments.

Keywords: Cancer, Protein Kinase C, Binding Energy, nPKC inhibitors, SwissADME, 3D-QSAR

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

PKCδ	Protein kinase C delta type
PKCε	Protein kinas C epsilon type
PKCη	Protein kinase C eta type
PKCθ	Protein kinase C theta type
PDB	Protein Data Bank
WEKA	Waikato Environment for Knowledge Analysis
QSAR	Quantitative Structure Activity Relationship
DAG	Diacylglycerol
PLC	Phospolipase C
IP3	Inositol triphosphate
MAPK	Mitogen Activated Protein Kinases
EGF	Epidermal Growth Factor
CDK	Cyclin Dependent Kinase
MASPIN	Mammary Serine Protease Inhibitor
SNCG	Synuclein Gamma
HDAC	Histone Deacetylases
HAT	Histone Acetylase
ROC	Receiver Operating Curve
MR	Molecular Refractivity
PSA	Polar Surface Area
ESOL	Estimating Solubility

1. INTRODUCTION

1.1 PKC Family

The Protein Kinase C (PKC) family is a group of protein kinases that have always been the focus of pharma industries for the past two decades. The interest in PKC studies began pacing when it was discovered that the isozymes of PKC are having a direct role in various human diseases, including cancer [1], autoimmune disease [2], diabetes [3, 4], heart failure [5, 6], Parkinson's disease and Alzheimer's disease [7]. There are different protein kinases in the human genome. Some may phosphorylate few substrates, whereas other phosphorylates many substrates and play a role in regulation of cellular responses. PKC phosphorylate serine and threonine residues. The PKCs were identified by Nishizuka and collaborators over three decades ago. After 5 years, it was found that PKC can be activated by phorbol ester. This led to study PKC as a target for drug development into cancer and other diseases.

1.2 PKC Signaling pathway

PKC members play a major role in cellular responses such as proliferation of cell, expression of gene and protein secretion. The structure includes two domains, a C- terminal which acts as a kinase domain connected to an N - terminal region which is a regulatory domain by a hinge region. The Kinase activity of PKC is inhibited at some intervals by an auto inhibitory pseudo substrate which directly binds to C- terminal. The classification of PKC into 3 broad groups is based on the differences in its regulatory domain. Conventional PKC (cPKC) consist three isoforms namely α, β and γ which have two functional regulatory domains C1 and C2, both of these play a major role in the activation of cPKC. Diacylglycerol (DAG) and a phospholipid binds to C1 and calcium binds to C2 domain. Novel PKC consist of four isoforms ε, η, δ and θ. This class also has two domains C1 and C2 but there is no need of calcium for activating C2 domain. The third class of PKCs is atypical enzymes, which consist of two isoforms. They don't have a C2 domain and even the C1 domain is non-functional, therefore it doesn't require any second messanger for its activation. Many growth factors in the cell lead to the activation of phospholipase (PLC) γ or PLCβ, which further on helps in cleaving phosphotidylinositol 4,5- bisphosphate which in turns generate DAG on inositol triphosphate (IP_3). IP3 later on helps in release of Ca2+ ions. DAG helps in recruitment of novel and classical PKCs towards the plasma membrane, which leads to change in conformation and results in activation. Soon after this activation PKC is constitutively phosphorylated at various sites. Classical and novel

PKCs can be activated by other agents such as phorbol esters and bryostatins that mimics the effect of DAG.

1.3 PKC in Breast Cancer

Cancer a collection of diseases is the stage when cells of the body start to split up and proliferate without any rule and start broadcasting to other nearby tissues. Normally the cells divide and grow when they are needed and as they grow old and start losing their function which leads to their damage, they eventually die, which allows new healthy cells to take their place. This normal process is not accompanied by the cancerous cells as when the cells get older and start abnormal functioning instead of dying these cells survive and even new cell are formed when not needed. Cancerous cells ignore the signals of feedback inhibition to stop dividing and apoptosis. The extra load of cells further leads to tumor growth. Mainly cancers form various types of solid tumors, which are lumps of tissues. These tumors are further categorized into two main types malignant and benign. Malignant are the ones which can be spread into other nearby tissues. At the time of their growth some cells are broken off from the solid lump and travel through the blood and lymph to invade into other tissues. Benign tumors, on the other side because of their large size do not invade other tissues and once removed from the affected site they don't grow back. Cancers are caused by changes in genes that are the main components for proper functioning of the cells. These changes are inherited from the parents to the offspring's.

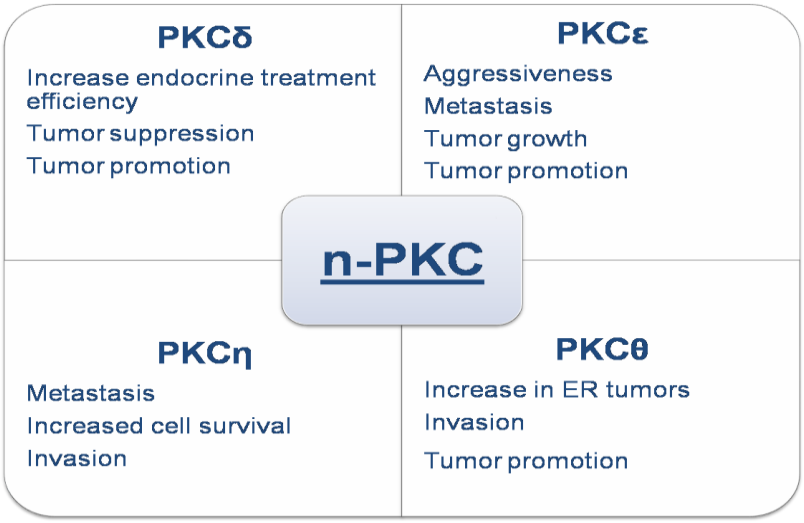

Figure 1: Functions of different classes of novel class of PKC in promotion of tumors, which leads to metastasis and tumor growth.

During involution and growth cycles, numerous changes take place in mammary epithelial cells [8]. Mainly activation of apoptotic and mitogenic signals are the changes in cells. Isoforms of PKC are overexpressed in breast tumor cell lines and tissues of malignant breast which control the pathways leading to cancer [9].

PKCδ role is still ambiguous and a lot of research is to be done to have information regarding the expression patterns of it in tumors. The Protumorigenic role of PKCδ has been seen in mammary cells via induction of anchorage dependent growth and survival [10]. Depletion of PKCδ is sufficient to initiate apoptosis in murine mammary tumor cells. [11]. Many studies have shown antiproliferative action of PKCδ, a dominant- negative mutant of PKCδ impairs arrest in G1 phase of cell cycle in breast cancer cells [12]. It is also linked with resistance to tamofexin and antiestrogen in breast cancer [13]. The expression level of PKCδ may help in the future for a marker in endocrine therapy to check the responsiveness of the patients.

PKCε is proposed to be a marker of aggressiveness of breast cancer. PKCε expression is correlated with ER negativity, HER2 expression which in turn leads to poor survival rate of breast cancer patients [14]. PKCε down regulation in MDA-MB-231 cancer cells have also reduced the chances of development of metastasis and tumor growth [14]. PKCε enhances the metastatic activity through parathyroid hormone protein in breast cancer cells, which further leads to activation of MAPK signalling cascade and transcription of genes [15]. Level of Akt is increased when PKCε is overexpressed and regulation of Akt by PKCε is done through integrin's interactions [16]. Rho GTPase could function as PKCε effectors as Rho GTPase are activated by PKC when it mediates invasion of breast cancer cells [17].

PKCη is involved in differentiation of the mammary gland. It is up regulated during pregnant state transition and decrease during lactation period [18]. Expression of PKCη is regulated by estradiol in T47-D and MCF-7 cell lines [19]. Many cell cycle proteins are modulated by PKCη. Moreover, activation of apoptotic caspases and release of cytochrome c is inhibited by PKCη [20]. High levels of PKCη were seen in metastatic breast tumor and were correlated with the condition of positive lymph node status [21]. Thus, PKCη could be represented as a good target for anti-cancerous treatments.

PKCθ have an important role in breast cancer, it is involved in depression of c-Rel and Akt activation. c-Rel activation induces the genes such as c-Myc and cyclin D1, which promote a rare transformation in the phenotype of mammary gland cells and further leads to the formation

of an invasive phenotype [22]. Increased PKCθ levels are also observed in ER- breast tumors [22].

1.4 Role of PKC in Cell Cycle

PKCs are generally activated by phorbol esters which is a tumor promoting agent and transduce mitogenic signals [23, 24]. PKC enzyme regulation in cell cycle is very complex in nature as different isozymes of this class possess opposing effects in different cells at a same time.

PKCδ inhibits G1 phase of cell cycle in response to agonist and activators like inositol hexaphosphate (IP6) and testosterone [25, 26, and 27]. Indirectly targeting cyclin E, cyclin D1 or cyclin A can hamper the G1-S phase [28, 29]. PKCδ down regulates the expression of cyclin D1 in colon tumor cells [30] and in smooth muscle cells as well [31]. With the further findings, it was seen that the loss of activity of PKCδ results in increased cyclin D1 in colon tumor cells [32]. PKCδ can also inhibit mitosis in murine fibroblasts [33]. It can also act as better positive regulators well [34, 35]. PKCδ increases the expression of cyclin A, cyclin D1 and cyclin E which enhances the rate of G1-S transition. [36, 37]. The regulation of the opposing effects of PKCδ can be done by phosphorylation of Tyr155 [38].

PKCε effects on proliferative responses can be predominantly seen in G1/S phase [39]. Loss of its activity in some cells such as NSCLC causes reduction in activation level of cdk2 complexes [40]. It induces transcription of cyclin D1 which helps in up regulating other cyclins proteins like cyclin D1 and cyclin E. [41]. PKCε is down regulated during the differentiation [42] and it helps in promoting adipogenic commitment of preadipocytes [43]. During the myogenic differentiation, up regulation of cyclin D causes enhanced expression of PKCε [44]. It also plays a major role in the activation of AP1 in T lymphoma cells, which further activates T cells [45]. In tissues like epidermis [46] and squamous epithelia [47], PKCη has shown association with mitotic cells. This isozyme up regulates level cyclin D and E in MCF-7 breast cancer cells [48]. It also plays a major role in T cell activation [49]. PKCθ positively regulates cells like breast cancer cells and gastrointestinal stromal cells by repressing the expression of certain cell cycle proteins [50]. It is also involved in activation of T cell and is triggered by CD 28 and TCR [51].

1.5 Inhibitors of PKC

1.5.1 BALANOL

It has a fungal origin and is produced naturally by Verticillium balanoides [52]. It is used for the inhibition of various serine/thr kinases like PKA and PKC by binding to the catalytic domain and compete with ATP as it has an affinity of about 3000 times that of ATP [53]. It was discovered in late 20th century in quest of new potent inhibitors for serine/thr kinases. The molecular structure of balanol has 3 main regions, namely 4-hydroxy benzoyl moieties, hexahyroazepane and benzophenone. The moieties are connected through an amide and ester linkage to each other. It was seen that balanol was unable to inhibit other tyrosine kinases such as EGF kinases and Src kinases.

1.5.2 7-HYDROXYSTAUROSPORINE

7-Hydroxystaurospoeine has an antineoplastic activity and is synthetically derived from staurosporine. It inhibits a large number of phosphokinases such as Ca^{2+} dependent PKC, serine/thr kinase AKT and cyclin dependent kinases. Almost 22 drugs in different studies around the globe are being studied on different tumor cell lines such as fallopian cancer, epithelial cancer, chronic lymphocytic leukaemia in which mostly will move to phase 3 trials.

1.5.3 SOTRASTAURIN

Sotrastaurin is orally available PKC inhibitor with immunosuppressive activities. It inhibits PKCθ and PKCβ which further delays activation of T-cell and B-cell.

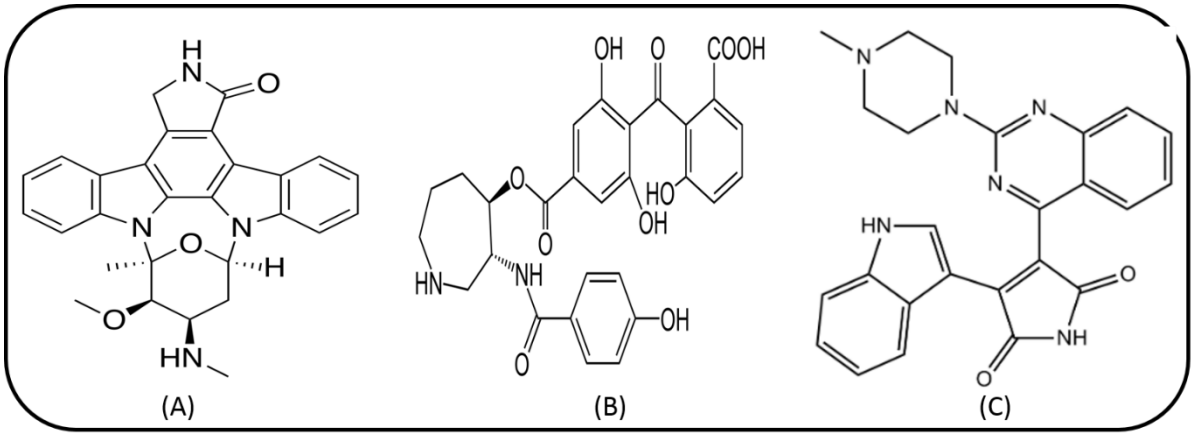

Figure 2: Most widely used inhibitors of novel class of PKC, (A) 7-Hydroxystaurosporine, (B) Balanol and (C) Sotrastaurin which have a role in regulating cancer.

1.6 Role of Interactors in Regulation of Novel Class of PKC

n-PKC in association with different regulatory proteins plays a vital role in many cellular functions by up regulation or down regulation of other proteins in the signalling pathway. The figure represents different substrates of n-PKC that were obtained from SignaLink 2.0 and Signor. Experiments conducted so far have proven that all four PKCs of novel class down regulates the activity of nitric oxide synthase 3 by phosphorylation at Thr495. Bisindolylmaleimide 1 a highly selective inhibitor down regulates PKCδ and PKCε by chemical inhibition, whereas protein phosphatase 2Ac down regulates activity of PKCδ by dephosphorylation at Ser645 and Ser664. On the other side DAG a powerful secondary messanger that plays a vital role in cell signalling helps in the activation of n-PKC by binding to regulatory domain.

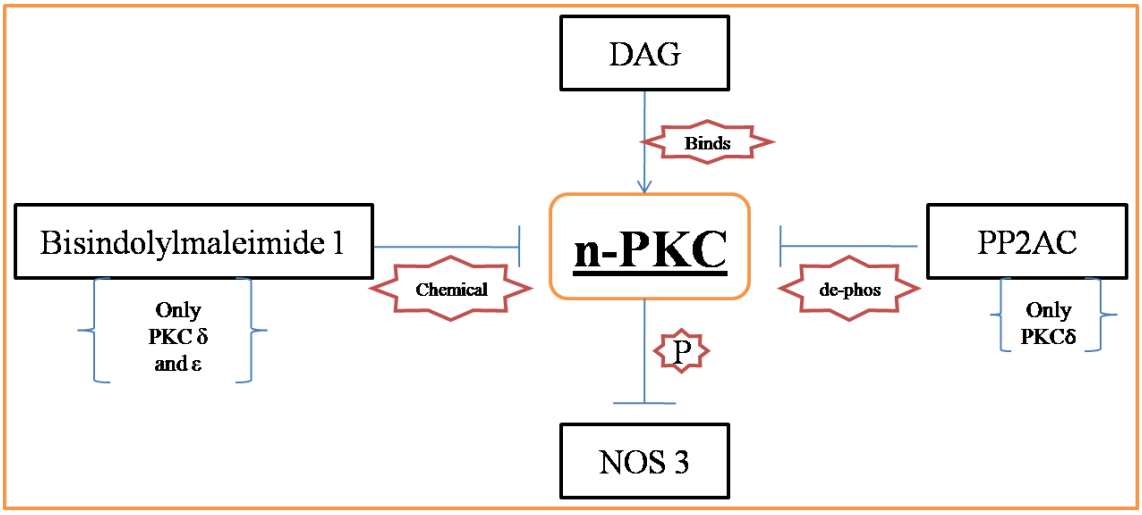

Figure 3: n-PKC acts as a substrate for DAG which helps in up regulation whereas bisindolylmaleimide 1 and PP2AC are used in down regulation. PP2AC is selective for PKCδ and bisindolylmaleimide 1 has selective binding for PKCδ and PKCε respectively.

2. REVIEW OF LITERATURE

2.1 Molecular Perspective of Cancer

In the past many researchers have debated over the information obtained from various sources about the role of different genes and protein in metastasis. The role of these mutated genes was an outbreak in cancer research and since then a large community of researches all over the world are engaged in finding the proper treatment for this fatal disease. Some related mutations like Chromosomal translocation disorders such as Abl gene in blood cancer, single point mutations such as in the case of colon cancers and deletion are caused due to the genetic changes in genome leading to generation of oncogenes. Nearly 65% of the tumors are caused due to mutation and abnormalities in the guardian of the genome that is p53 gene. Mutations in these proteins hamper the basic functioning and causes disturbance in its functioning. Normal p53 protein plays a major role in maintaining cell death related processes such as senescence and helps in the division of cells along with cytoskeleton stability. The mutation are generally caused in its DNA binding domain. Cancer cells generally stall in G1 and G2 phase for a bit longer time than normal as p53 interacts with CDK1-P2 allowing cancerous cell to attain proper morphology. Hypo methylation also leads to rapid increase in movement in DNA by insertion and deletion of many genes which at last causes chromosomal instability. LINE family consist of L1 which is observed in many cancers. Hypo methylation in regulatory sites of specific promoters activate many kind of oncogenes such as MASPIN gene and SNCG in breast cancer. Many genes involved in vitamin response, repair, apoptosis, and cell cycle are influenced by inactivation caused by hyper methylation which gives an insight of using these hyper methylated promoters as biomarkers in the early prediction of cancer.

Deacetylation by HDAC is seen in many tumors such as prostate and ovarian. Sirtuin is one of the main HDAC enzyme and its cooperation with DNMT1 have affected the methylation pattern. HDAC can be regulated by a small non coding RNAs such as miRNA which help in regulating the growth of cell. Even Histone acetylation disturbance in various cancer types caused due to deletion of HAT and related genes leads to array of events that help in the direct proliferation of cancer cells. Methyltransferase and demethylase enzymes being responsible for histone methylation pattern causes mutations in UTX demethylase leading to renal carcinoma.

2.2 PKC Modulators

Participation of isozymes of PKC in cancer promotion or in its antagonizing directly supports the notion for PKC as a potential target in anticancerous therapies. Many modulators of PKCs are currently in market as well as in last phases of clinical trials. Strategies used for drug development for PKC inhibitors generally include ATP binding pocket inhibitors, various kinase inhibitors. The first in this series was staurosporine which compete with ATP to show its antiproliferative action. Many other new compounds such as Enzastaurin along with Midostaurin were developed using staurosporine as a base compound and now are widely being used in clinical trials. The first PKC inhibitor to be evaluated in cancer clinical trials was Midostaurin as it exhibited selectivity towards ATP binding pockets, since then it has shown a promising effect in anticancerous activity and even caused delay in the development of metastasis in lungs. It is also being used to potentiate anti-tumorigenic activity of doxorubicin which is a cytotoxin used in clinics. Moreover, it has shown biological activities in lymphatic disorders like acute myeloid leukaemia. Enzastaurin is a serine/thr kinase inhibitor and at very low concentration inhibits classical PKC but it inhibits novel PKCs at higher concentration. Its anti- tumor role is linked to PI3K pathway. Presently its role is being confirmed in patients having solid tumors as its dosage can be tampered accordingly from 15 to 745 mg/day.

Bryostatins an activator of DAG and phorbol ester induced PKC isozymes which help in down regulating of cells, which in turn causes suppression of certain responses leading to cancer. Certain prototypes are being developed in which bryostatin 1 was the first. Its exposure for short duration may result in nPKC activation and nuclear localization. In fact, long duration exposure shows an antagonistic effect by decreasing PKC activity caused due to depletion of PKC present in the membrane. The effects of this compound are promising for phase I trails only and now this is being used with other agonist such as vincristine, gemcitabine and cisplatin for proper administration and localization in tumor cell lines.

Another compound which was initially created as a modulator is Ingenol-3-angelate which was extracted from Euphorbia peplus a plant species found in central Europe. Its structure is similar to phorbol ester and modulates PKC delta in myeloid cancerous cells and models of colon cancers. It has shown antiproliferative role in G1 phase of cell cycle and in apoptosis. This is being used in solid tumors and currently in phase III for treating actinic keratosis. It has emerged as an immunostimulator agent that not only treat tumors, but also generates CD8 T cells which shows anticancerous properties on distant as well as nearby local tissues. Based on

this study only some therapies have been demonstrated so far, but in future will allow more insights into treatment.

ole

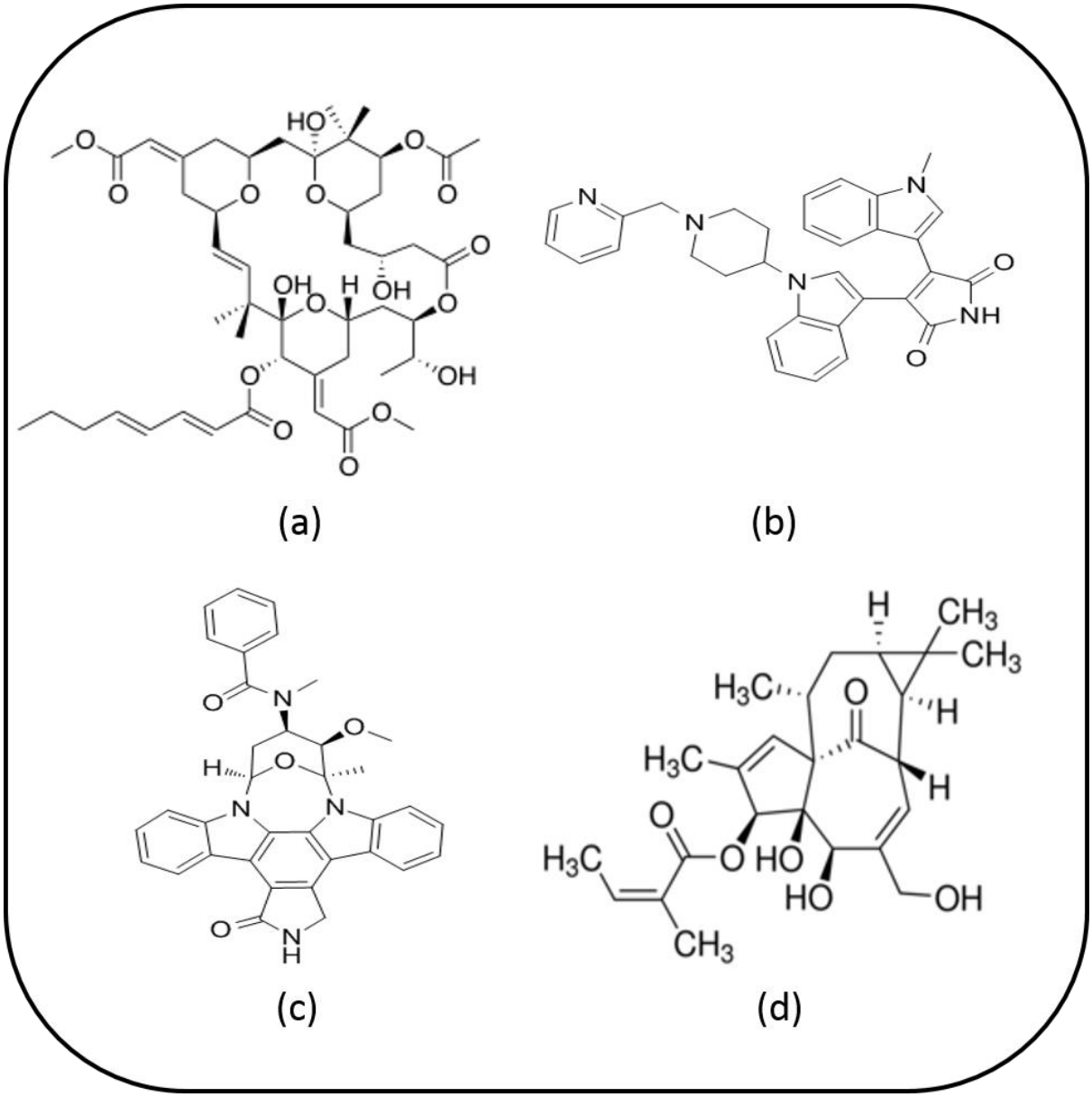

Figure 4: nPKC modulators having role in inhibiting cancer. (a) Bryostatin I, (b) Enzastaurin, (c) Midostaurin, and (d) Ingenol-3-angelate.

3. METHODOLOGY

3.1 DATA PREPARATION

- Selection of different class of Protein kinase C with the help of various literatures. All the three subfamilies: novel, conventional and atypical was analyzed and studied from various sources.
- Obtained 3D structures of novel PKC from RCSB PDB [54]. Thereafter PDBSum was used for acquiring information of active sites [55].

  From the three classes, our study was based on a novel class of PKC.As n-PKC is involved in cell survival, migration, and angiogenesis and had a potential to be a therapeutic target.

  i. PDB structure of PKCδ was obtained from SwissModel [56].
  ii. PDB structure of PKCε was also obtained by doing homology modelling from Swiss Model [56].
  iii. PDB structure of intact PKCη named as 3TXO was obtained from RCSB PDB [57].
  iv. PDB structure of PKCθ isozymes named as 4Q9Z was obtained from the RCSB PDB [58].
- Identification of bound Ligands and analyzing them.

  Ligands were identified with the help of databases like ChEMBL [59], Zinc and Binding database [60]. These databases gave compounds that may have drug like effect on our target protein.
- Collection of similar targets of n-PKC (ε, δ, θ, η) from Binding Database. Finally 2644 similar ligands were obtained from binding database

3.2 DATA PREPROCESSING

- Compiled the data together and after removing the duplicates left with 1911 unique compounds.
- Similar targets were arranged on the basis of different parameters (IC 50, Smiles format).

3.3 ADME ANALYSIS

- Insert Smiles format of similar targets in Swiss ADME Tool.
- Calculation of ADME properties (nearly 45 attributes were calculated) and arranging targets on the basis of those.

- Selection of compounds those had similar ADME properties to the known inhibitors inhibitors (balanol, sotrastaurin and 7-hydroxystaurosporin) of n-PKC class.

3.4 QSAR STUDIES

- Selection of top compounds from ADME and with the help of their known properties constructed 20 new structures. Different properties along with important molecular descriptors for the study were calculated by ACD ChemSketch [61], Molinspiration [62].
- The new compounds created were added with extra benzene rings, halogens, phosphates and methyl groups respectively

3.5 MACHINE LEARNING

- With the help of Weka tool different models such as J48, multilayer perceptron and Ibk models were created.
- The data obtained from binding database was filtered and on the basis of IC 50 value (nM) whole set of data was classified as inhibitors and non-inhibitors which was later used as a training dataset.
- Similar attributes were calculated for the 20 constructed compound to make it compatible to be used as a test set. After classification it was seen that 16 compounds were having inhibition property.
- Model validation and on the basis of models top 8 were inhibitors selected for docking studies.

3.6 VIRTUAL SCREENING

- The docking protocol prescribed by SwissDock was used during the entire work.
- Docking studies of common known inhibitors reported for all four classes were docked with their respective target proteins by SwissDock. Parallel studies were made using top 8 compounds from weka analysis were also docked separately with each of the 4 target proteins.
- Full fitness value of the docked structure was noted down and along with the images of bound ligand with the target were saved for the results. With the help of UCSF chimera all the ligand and target structures of protein were saved in a separate file in pdb format. Comparing and analyzing the binding sites.

3.7 Tools and Databases Used During the Study

- **ZINC**- Freely available database for obtaining compounds for virtual screening. This database contains over 30 million compounds which can be directly used for docking study [63]. Physical properties like molecular weight, net charge, rotatable bonds, polar surface area, hydrogen donors, hydrogen acceptors, etc. are annotated with the searched molecule. There is a vendor list for each molecule which can be purchased. By alterations of different compounds many combination can also be formed by this database. ZINC follows some criteria to evaluate properties for a searched compound such as H-bond is greater than six, rotatable bonds more than 15. Atoms other than P, Cl, C, O, H and N are removed.
- **BindingDB**- Chemical database for obtaining similar targets to known protein. Binding DB mainly focuses on protein targets that can act as a drug like small molecule. It contains about 1 million ligands binding data for 7500 targets of protein. This helps in drug discovery studies and development of Quantitative structure-activity relationship along with validation of free energy methods and molecular docking. The data is collected by the measurement techniques such as NMR, calorimetry.
- **UCSF Chimera**- It is capable of analysis and visualization of structures. UCSF Chimera is protein visualization tool which also gives information about other attributes such as density maps, sequence alignments, conformational ensembles, trajectories and docking results [60]. This tool is supported by National Institute of Health and mainly developed by Resource for Biocomputing, Informatics and Visualization. The protein structures can be obtained in many formats along with pdb. It includes various tools such as structure comparison, structure editing, amber, ensemble analysis, aligning chain sequences. The interaction between the ligand and selected protein is easily recognizable that plays a clear insight in understanding the proper binding pattern just by viewing the structure.
- **SwissADME**- It is used to calculate physiochemical properties. This is a user friendly tool in which just by inserting the smiles format of a particular molecule, it calculates various descriptors to predict pharmacokinetic properties and parameters for ADME. It helps to analyze whether a selected molecule behaves as drug like or is near to those properties. It mainly gives account of forty one descriptors such as Molecular weight, number of rotatable bonds, total polar surface area, number of H-bond donors and acceptors, logarithm of many values, violations of Lipinski rule. [64]

- **Molinspiration**- Tool used to calculate molecular properties. Molinspiration is a package of cheminformatics tool which helps in manipulating of molecules by inserting smiles format. It also offers easy to use property calculator which helps in classifying of inhibition property of that molecule whether it is a kinase, enzyme or protease inhibitor. The tool also calculates useful properties such as polar surface area and logP which further helps in characterization.
- **ACD ChemSketch-** Used to construct new chemical compound structure. This is used to draw new structures including features like 3D structure viewing, assigning functionality, calculating molecular descriptors such as molar refractivity, molecular weight, reaction scheme, etc. It contains an in built dictionary which is used search new compounds and generate IUPAC names.
- **Protein Data Bank** (PDB) - It contain whole information of nucleic acids and proteins. PDB consists of all the basic information about the 3D shapes of nucleic acids and proteins which make it easier for a student or researcher to understand the structure of a given protein. At present it contains 43911 distinct protein sequences, 9993 nucleic acid containing structures and 38176 structures of human sequences. The structures released in PDB are the outcome of various protein purification techniques such as X-ray crystallography, Nuclear magnetic resonance and electron microscopy. Data preparation tools such for data extraction (pdb_extract, SF-Tool), data format conversion (PDBML2CIF, Point Suite and Maxit) can be found within PDB. There are many visualization options within this like pathway view, pose view, protein feature view and human gene view.
- **Swiss Similarity**- Used for ligand based virtual screening. It is a web tool that is freely available for rapid virtual screening of different libraries of small ligands [65]. Bioactive molecules and drugs are screened and then predictions are carried out by various approaches such as non superimpostional and superimpostional similarity. It stores large data of about 7 lakh bioactive compounds, 7 thousand drugs and 2 million commercial compounds.
- **WEKA**-Machine learning software. It was developed by University of Waikato, New Zealand and is written in Java. Weka is a collection of different algorithms used for analysing data and supports numerous tasks such as clustering, classification, feature selection and data pre-processing [66]. The data is used to train a model with different number of attributes and based on the training set other dataset known as test data is

being classified. More the number of attributes and data better is the model created. Several panels such as pre-process panel uses filtering algorithm for pre-processing excel data sheets, classify panel for regression which on prediction creates receiver operating curves (ROC), associate panel is used for the identification of any relation present between the attributes, cluster panel is used for clustering with the help of say k-means algorithm.

- **SwissDock-** Docking software for protein and ligand. It is a web server used for docking studies. The target protein and ligands are prepared accordingly by selecting the chains of protein to be docked. The backhand process of docking is hidden which makes it a user friendly tool. It is important to have all H-bonds added within the ligand, otherwise the ligand is not compatible for docking study. SwissDock usually try to form different clusters based on the query and each cluster have their own maximum fitness values which allows the user to choose the best cluster [67]. The docking results can be viewed with the help of UCSF chimera.
- **Open Babel –** An inter conversion tool for chemical structures. There is a problem when dealing with a chemical file, to obtain the correct format Open Babel is best as it supports 111 formats in which 82 are readable. Common formats used in chemistry are mopac, Gaussian; in cheminformatics mol, smiles, mol2; in reaction formats mdl, etc. These formats are then further used for various studies such as virtual screening and structural associated studies. Even coordinates are generated such as 3D coordinates which allow easy conversion to 3D formats such as SDF [68].
- **ChEMBL-** Chemical database for molecules having drug like similarity. It is curated manually and maintained by European Bioinformatics Institute (EBI). This database is usually useful as it categorise compounds on the basis of IC50, KC 50 values which helps in designing inhibitors for a particular target protein. Drug discovery is based on screening libraries in ChEMBL. It contains about 6 lakh products curated from around 30 thousand literature and journals. At present it contain 11 thousand targets, records of 21 lakh compound and activities of 1.5 crore compounds.
- **UniProt** – It is a repository of universal protein data. UniProt is an accessible protein database that includes information of sequence and function of various protein which are derived from whole genome sequencing throughout the world. It is funded by European bioinformatics institute located in Hinxton, UK. All the biological functions of various proteins are curated from literature sources and then compiled so as user gets

the compact and accurate result for the query search. It provides four main core databases Uniparc, UniRef, TrEMBL and Swiss-Prot which cover certain areas such as subcellular location, pattern of expression, catalytic activity, protein-protein interaction, etc.

- **SwissModel** - Tool used for homology modelling. SwissModel is used to create templates based on the sequence of proteins. It is generally gives top 50 templates on the basis of similarity and other criteria which can be taken directly to form a model of the unknown protein structure. This tool helps in creating predicted model which can solve out the problem of structure. The obtained pdb structure can be further analysed with the ligands by virtual screening.

4. RESULT

4.1 PDB Structures by Homology Modelling

nPKC class target protein structure with bound ligands which were taken as a template during whole study for virtual screening and modelling are shown below. Functional specificity and different biological functions of proteins are based on its structure. The residues responsible for proper functioning of protein are arranged in precise geometry. The residues in the active site of the protein provides protein with its activity and rest of the protein is required to maintain spatial position of whole protein. Homology modelling is knowledge based structure prediction. In this the unknown structure are aligned with known template sequence. For successful modelling similarity between the two sequences should be more than 30 %.

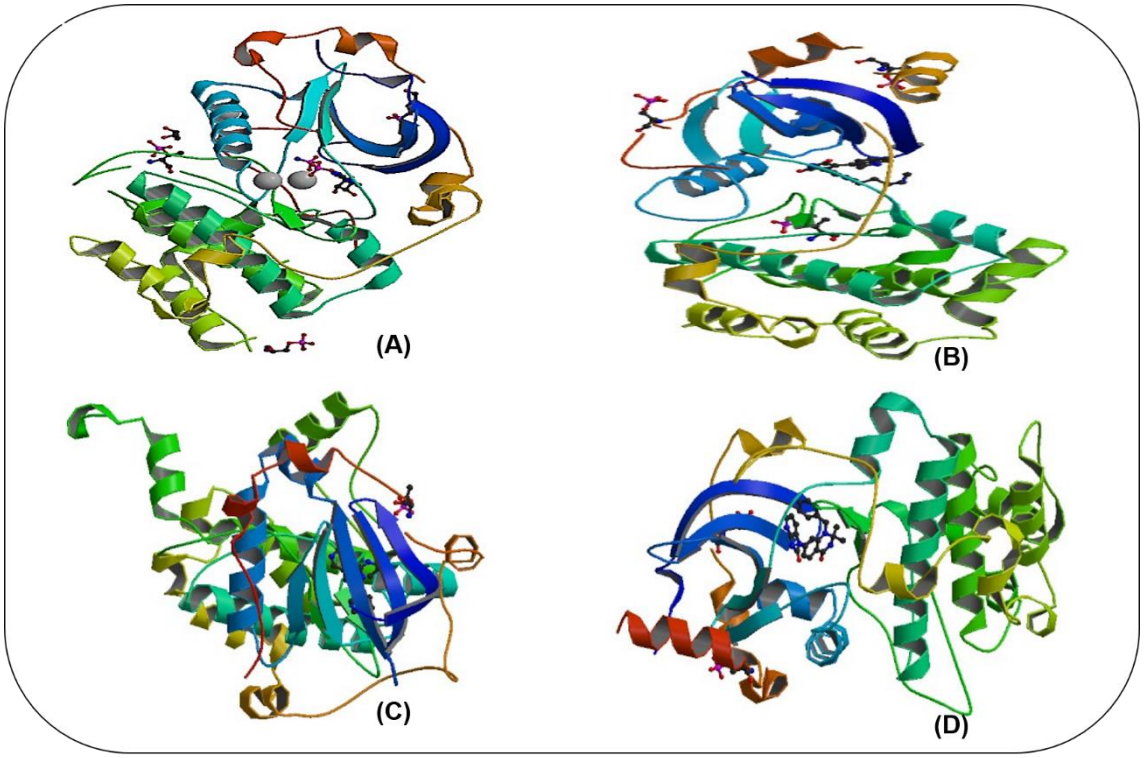

Figure 5: (A)-PKC delta model created by most similar template from SwissModel. (B)PKC ε model created by most similar template from SwissModel homology modelling (C)- PKC eta kinase in complex with a naphthyridine (PDB – 3TXO). The method used for obtaining crystal structure was X- ray diffraction at 1.5 angstrom. (D) Structure of PKC θ with 2,2-dimethyl-7-(2-oxidanylidene-3-H-imidazo[4,5]pyridin-1-yl)-1-(phenylmethyl)-3-H-quinazolin-4-one (PDB 5F9E). The method used for obtaining the structure was X- ray diffraction at 2 angstrom.

The three dimensional structures of proteins were predicted using homology modelling where data for experimental structures were missing. Out of four novel PKCs, two of them were modelled to obtain proper three dimensional structure that can be further used for virtual screening. Using all the required information from template sequences the model of protein was constructed. SWISS-MODEL tool from expasy server was used for this purpose.

4.2 Open Targets Platform

The Open Targets Platform is a freely accessible data tool which compiles the information from 17 sources and gives an integrated information about the different targets and their association with the diseases. The platform generally provides a target centered workflow for identification of the disease associated with specified targets or a disease centered workflow for identification of targets that may have association with the disease. It was seen that PKCs association score with cancer was 1. Even diseases like Down syndrome and tuberculosis have some link with PKCs. The diseases linked to PKC with their association score and therapeutic areas are listed in the **Table 1.**

Disease	**Association Score**	**Therapeutic Area**
Cancer	**1**	**Neoplasm**
Leukemia	1	Hematological System Disease
Prostate Carcinoma	0.981	Reproductive system disease
Colorectal Cancer	0.941	Digestive System Disease
T Cell Lymphoblastic Leukemia	0.927	Hematological System Disease
Bile Duct Carcinoma	0.776	Endocrine System Disease
Melanoma	0.721	Skin Disease
Lung Adeno Carcinoma	0.664	Respiratory System Disease
Congenital Heart Disease	0.615	Cardiovascular Disease
Peritoneal Neoplasm	0.287	Neoplasm
Medulloblastoma	0.275	Nervous System Disease
Renal carcinoma	0.262	Kidney Disease
Glioma	0.225	Nervous System Disease
Pincoblastoma	0.2	Nervous System Disease

Adenocarcinoma	0.093	Digestive System Disease
Bipolar Depression	0.056	Nervous System Disease
Retinoblastoma	0.046	Neoplasm
Xeroderma Pigmentosum	0.045	Neoplasm
T Cell Leukemia	0.038	Hematological System Disease
Tuberculosis	0.036	Respiratory System Disease
Down Syndrome	0.035	Genetic Disorder
Diphtheria	0.031	Infectious Disease
Aneamia	0.025	Phenotype

Table 1: PKCs association score in different disease along with their therapeutic area.

4.3 Virtual Screening

Virtual screening and docking studies were conducted to examine the binding pattern of our receptor to the constructed ligands. These methods are mostly computer based and are used to identify new and unique ligands from vast libraries on the basis of known structures. Large libraries contain thousands of organic molecules which are docked into the selected receptor and on the basis of affinity, molecules are sorted. Once these structures are predicted they are sent for experimental analysis and later on are compared with structure obtained from X-Ray crystallographers and NMR spectroscopy for final validation. Based on the n-PKC binding site characteristics, screening of compounds were carried out in binding database (2644 compounds found) and they were sorted according to their IC 50 values. Finally the top compounds were selected randomly from the list and on the basis of their properties and chemical structures other 20 new structures were drawn. These structures were drawn with the help of Molinspiration and ACD ChemSketch by allowing addition of different functional groups in carbon backbone. Genuine ligands were generated computationally which were further analysed more accurately by sending them for machine learning.

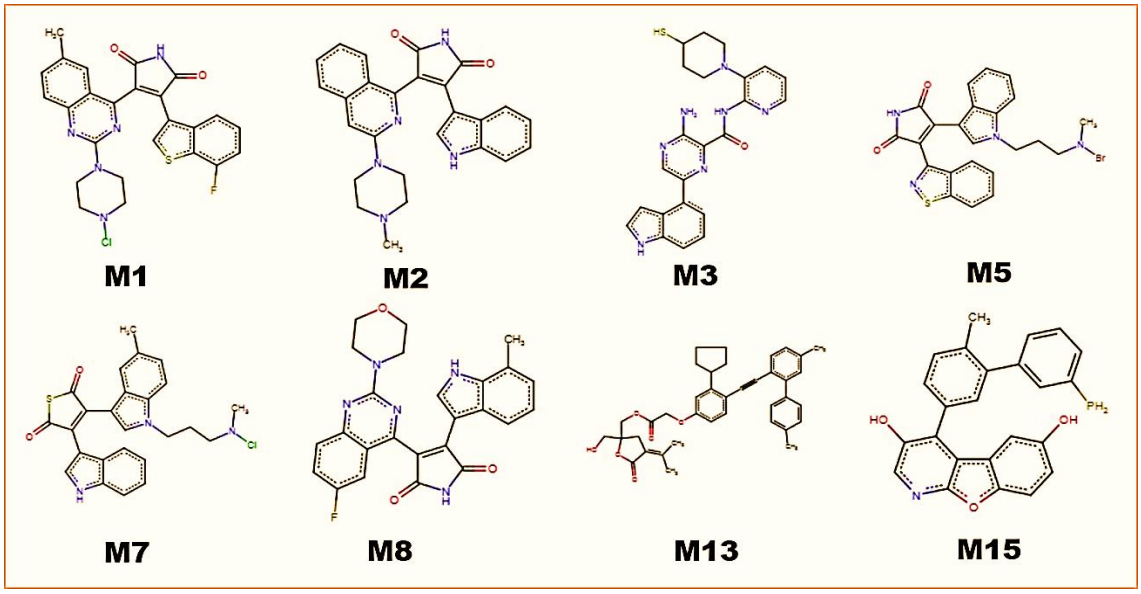

Figure 6: Top 8 constructed compounds having similar properties to compounds obtained from binding database.

4.4 Machine Learning Analysis

As it is based on artificial intelligence that trains the system automatically and further improves from the past experiences without being programmed again. In our study the learning process began with various attributes such as physicochemical properties, lipophilicity, water solubility, pharmacokinetics, drug likeness and medicinal chemistry. All of these contained 45 total attributes which trained the data accordingly and created a robust training set (1911 compounds). To obtain accurate results the same attributes were calculated for the test set (20 constructed compounds). One should note that more the number of attributes more are the chances of better training of data. For proper model creation the quality of training set must be of good quality having many distinct parameters. Binding database was used for training data and then rearranged on the basis of physicochemical properties and IC50 values. Inhibitors were classified as having IC50 < 1000nM and the rest having IC 50 > 1000 were classified as non-inhibitors. For model creation the descriptors of test and training data sets was calculated by SwissADME.

Weka was used to carry out machine learning as it is a user friendly tool and it includes different machine learning algorithms that can be applied to different datasets for data mining. Three models were generated and each gave an accuracy of more than 80 %. On the basis of the output result 8 out of 20 compounds in test set were classified as Inhibitors.

4.4.1 Models

1. **Decision trees- J48**

   For building decision tree C4.5 algorithm and a classifier J48 is implemented. The default settings for the parameter are J48 –C 0.25 –M 2, where C is confidence value and M is minimum number of instances in two popular branches of a particular decision tree. Decision tree learning go from first item that is a observation to the last possible item to search for a good result. It is a predictive modelling approach that is widely used in statistics and machine learning approaches.

2. **Multilayer Perceptron**

   It is based on neural network approach where there is one input layer for each attribute and one output layer for each class. It can have many hidden layers based on the problem, for eg, one hidden layer is suitable for problems having single convex region. Its performance is near to 79 % which is far better than other methods. It works by creating multiple epochs through the given data. Training is continued till the training time exceeds the given limit, if not so then the accuracy is hampered.

3. **k-nearest neighbors method**

   It is a lazy classifier and is used to select appropriate values from the classificated k-based methods. It is sometimes used for distance weighting also. IBk consist of IB1 classifier which is a instance based learner and predicts a unique class of single training instance nearest to given class for every test instance.

In each model accuracy that is ratio of predictions made by different models is mainly based on correctly identified problems calculated during the working of various algorithm based on machine learning. Many areas under the curve such as ROC curves gives an account of specificity and sensitivity respectively.

	J 48 Tree	Multilayer Perceptron	IBk Model
Accuracy	83.25%	81.11%	81.89%
PRC Area	0.832	0.837	0.825
ROC Area	0.822	0.814	0.802
Kappa statistics	0.5359	0.4798	0.4983
Mean absolute error	0.2144	0.2153	0.1923
Root mean squared error	0.3684	0.3849	0.4048
Relative absolute error	55.62%	55.78%	49.82%
TP Rate	0.833	0.811	0.819
FP Rate	0.331	0.36	0.353
Precision	0.825	0.802	0.811
F-Measure	0.826	0.805	0.812

Table 2: Accuracy along with other useful parameters calculated for various models by Weka

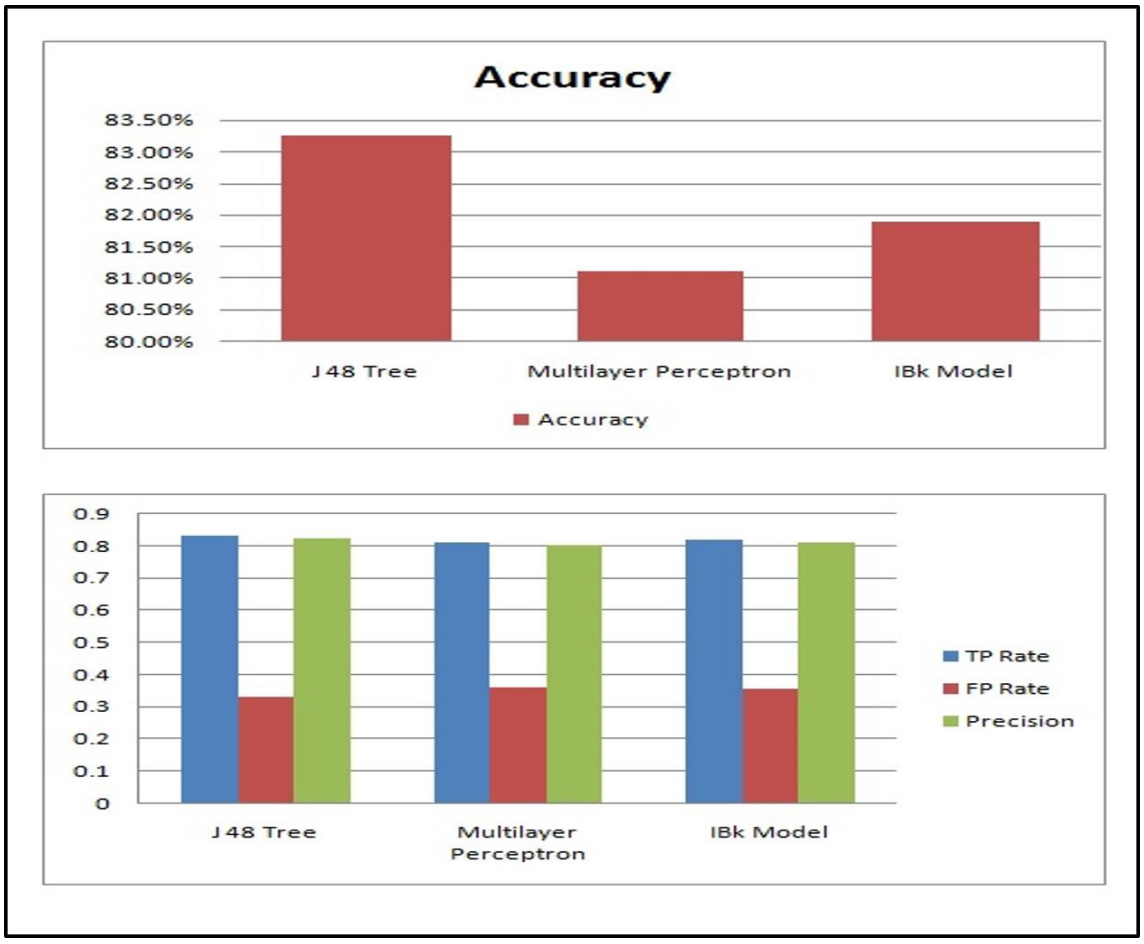

Figure 7: J48 Tree model with highest accuracy of 83.25% was the best suited model for selection of top 8 compounds.

4.5 Molecular Docking with Known Inhibitors and Top 8 Constructed Ligands

Molecular docking is the most used tool for discovering new drugs. Docking studies are conducted between protein-protein and ligand-protein to predict the possible binding sites between the two structures. Docking tools generally search for high dimensional space and rank the molecules according to different scoring functions. Nowadays docking is being used for large libraries of compounds which proposes a new hypotheses on inhibition property of ligands. The most important step while staring docking is to correctly set up the target protein and ligand structures. As in SwissDock all the hydrogen's were added and 3D coordinates were generated separately for each of the target and ligand.

Top 8 constructed compounds were docked with all four PKCs and their 3D structures were downloaded in mol2 format along with H-bonds. Meanwhile known inhibitors were searched for n-PKCs and by analysing all it was found out that 7-hydroxystaurin, Sotrastaurin, and Balanol were common for all n-PKCs. These 3 inhibitors were docked with all four PKCs. The crystal structure of PKCδ, PKCε, PKCη and PKCθ in association of different ligands were used as receptors.

Full fitness values comparison of top 8 constructed compounds and 3 known inhibitors has shown that compound M3, M8 and M15 out of the 8 has higher binding affinity with all the four classes of n-PKC. This shows that they can be used as selective inhibitors for novel class of PKCs. The known inhibitor sotrastaurin which acts as a common inhibitor in all forms of n-PKC has higher fitness values then other known inhibitors, but our constructed compounds have even higher fitness value than it. The fitness values of top 8 compounds computed by SwissDock is shown in **Table 3** while the full fitness graph with all isozymes of n-PKC is shown below in **Figure 8, 9 and 10** respectively.

	M1	M2	M3	M5	M7	M8	M13	M15
Delta	-2290.8	-2237	-2162.7	-2263.7	-2273.5	-2274.9	-2259.1	-2283.7
Epsilon	-2669.2	-2618.7	-2574.7	-2639.5	-2656.1	-2659.4	-2630.9	-2664.6
Eta	-2177.3	-2127.2	-2050.4	-2150	-2160.9	-2162.3	-2145.9	-2170.4
Theta	-4179.6	-4129.1	-4227.3	-4322.5	-4339.3	-4339.1	-4319.2	-4341.5

.

Table 3: Full fitness values (Kcal/mol) of top 8 constructed compounds with novel class of PKCs

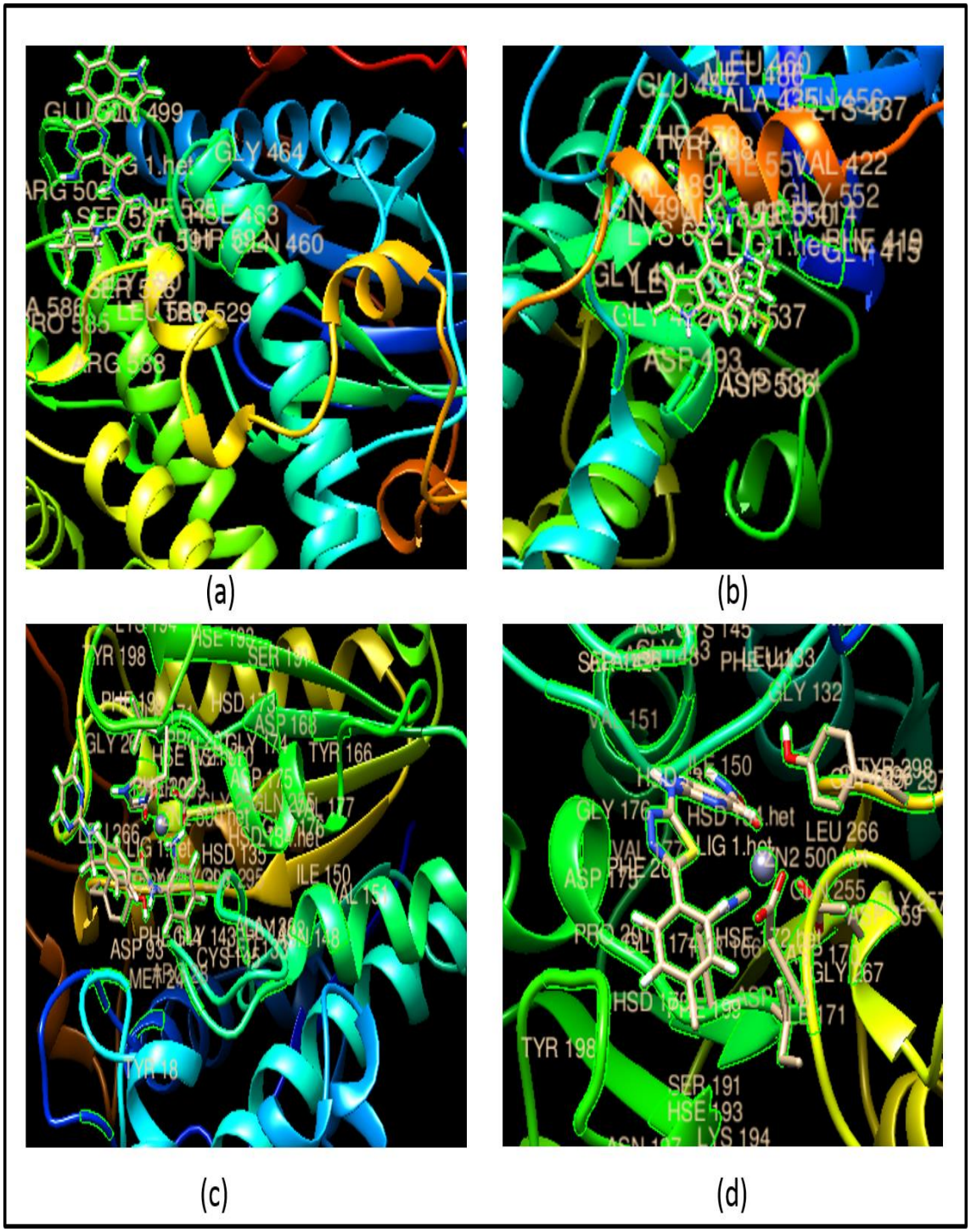

Figure 8: Docking pose of M3 with all four nPKCs. (a) Residues involved in receptor binding with the ligand PKC δ were GLU 460, GLY 464 and THR 529, (b) PHE 419, VAL 422 and ALA 435 were the main contributors in binding with PKC ε, (c) Mainly aromatic residues such as TYR and PHE played a major role in binding of ligand with PKC η, (d) PKC θ with M3 was due to GLY 176, PHE 201.

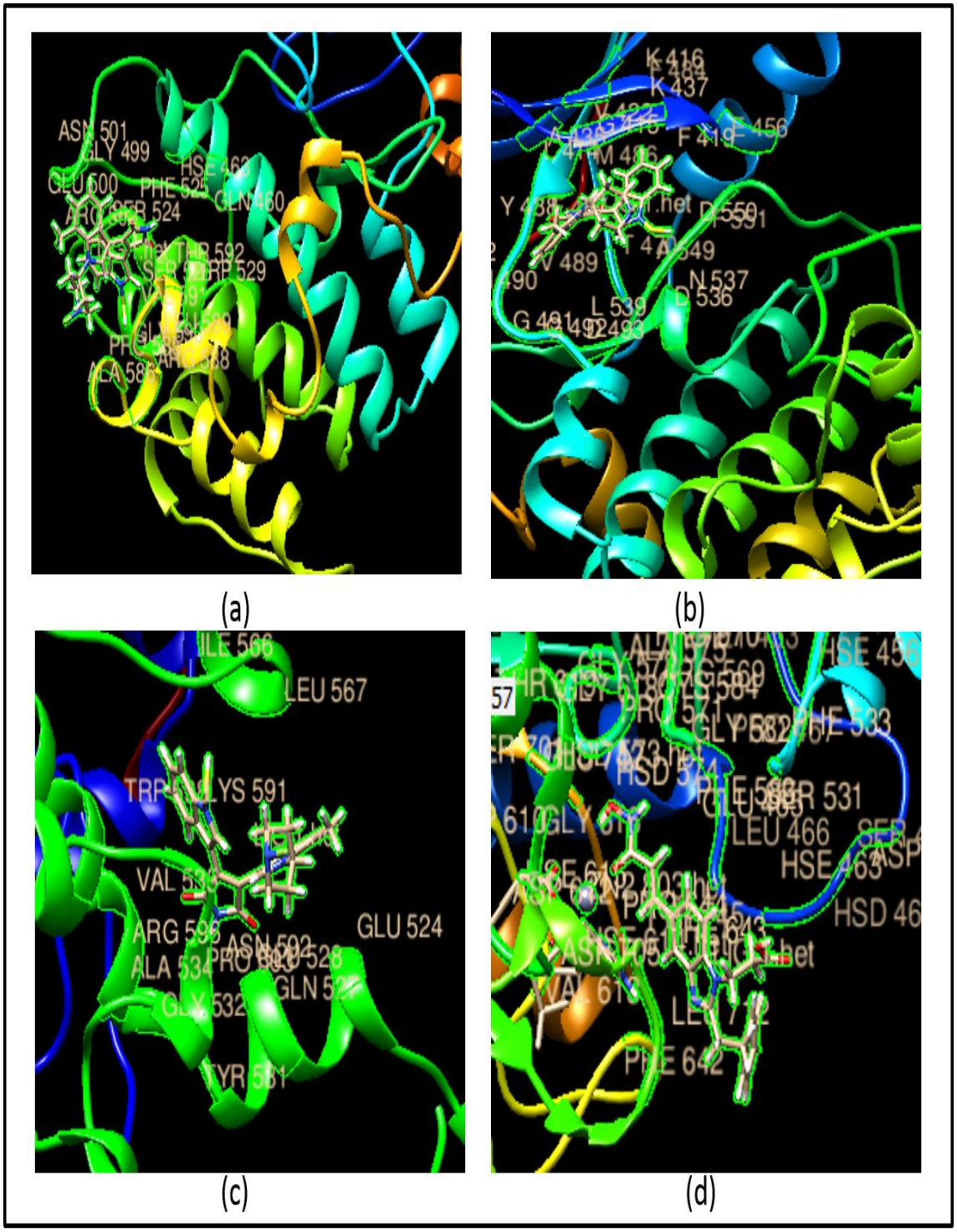

Figure 9: Docking pose of M8 with all four nPKCs. (a) LYS 499, GLU 500 and ASN 501 were the main contributors in binding with PKC δ, (b) Residues involved in receptor binding with the ligand PKC ε were TYR 488 and VAL 489, (c) Mainly residues such as LYS 591 and VAL 536 played major role in ligand binding with PKC η, (d) PKC θ binding with M3 was primarily due to LYS 611, VAL 613 and PHE 642.

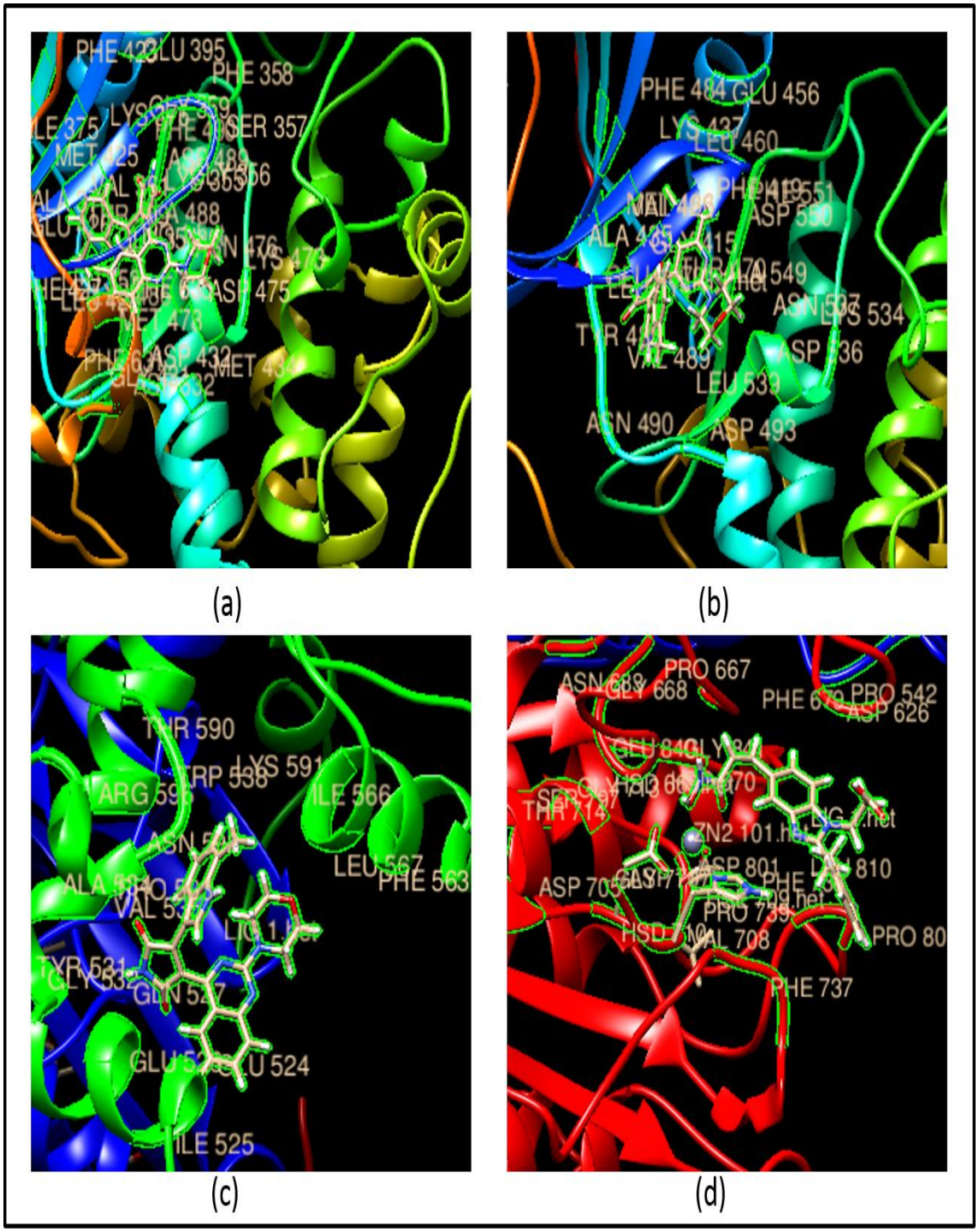

Figure 10: Docking pose of M15 with isozymes of nPKCs. (a) GLU 426, ASN 476 ALA 488 are main contributors in binding with PKC δ, (b) Top 3 residues that were involved in receptor binding with the PKC ε were TYR 431, THR 470 and VAL 489, (c) Residues such as GLU 524 and TYR 531 played significant role in ligand binding with PKC η, (d) PKC θ binding with M3 was primarily due to ASP 801 and GLY 847.

4.6 ADME Analysis

4.6.1 Physicochemical Properties

These simpler molecular descriptors like molecular refractivity (MR), molecular weight (MW), polar surface area (PSA) were calculated by using OpenBabel. PSA was calculated by the fragmental technique known as topological derived polar surface area which considers phosphorus and sulphur as polar atoms. All the other physicochemical parameters were computed using SwissADME which were later on used to select the best compounds.

Molecule	Formula	MW	Heavy atoms	Aromatic heavy atoms	R bonds	HBA	HBD	MR	TPSA	iLogP
M1	C24H18ClFN6O2	476.89	34	19	3	6	2	137.59	94.22	2.71
M2	C26H23N5O2	437.49	33	19	3	4	2	139.85	81.33	2.43
M3	C23H23N7OS	445.54	32	21	5	4	3	133.12	151.62	2
M5	C25H20FN3O2S	445.51	32	18	3	4	1	129.89	83.4	2.82
M7	C23H19BrN4O2S	495.39	31	18	6	4	1	131.55	95.47	3.11
M8	C26H23FN4O2S	474.55	34	18	3	4	1	142.01	98.07	3.08
M13	C25H22ClN3O2S	463.98	32	18	6	3	1	133	83.4	3.14
M15	C25H20FN5O3	457.46	34	19	3	6	2	132.03	100.21	2.53

Table 4: Physicochemical properties of top 8 constructed compounds

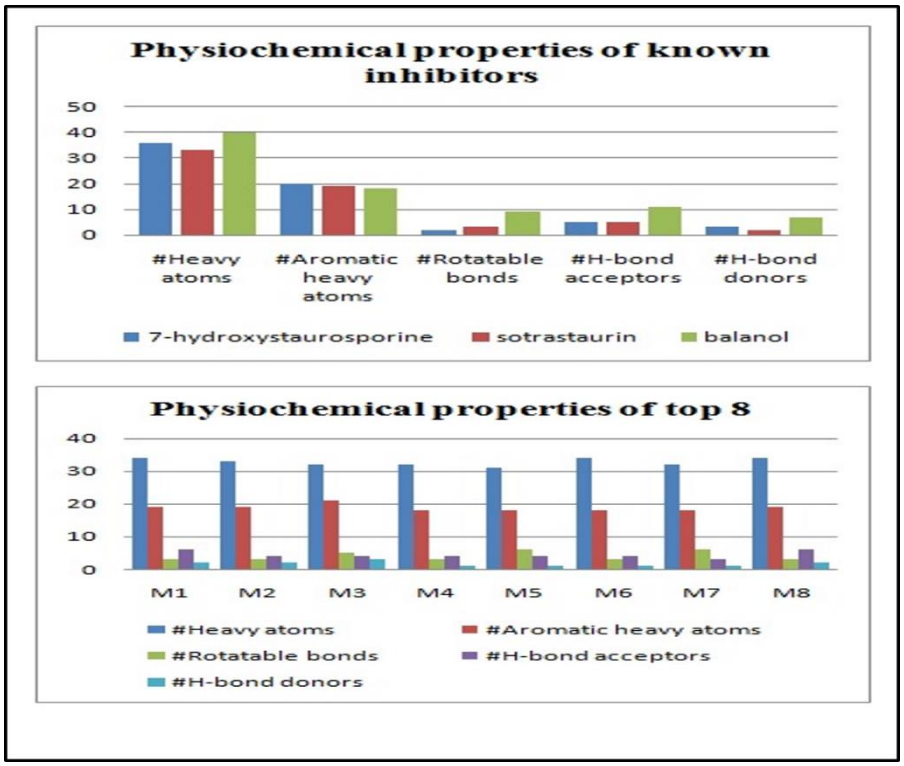

Figure 11: Comparison of physicochemical properties of known inhibitors along with the top 8 constructed compounds.

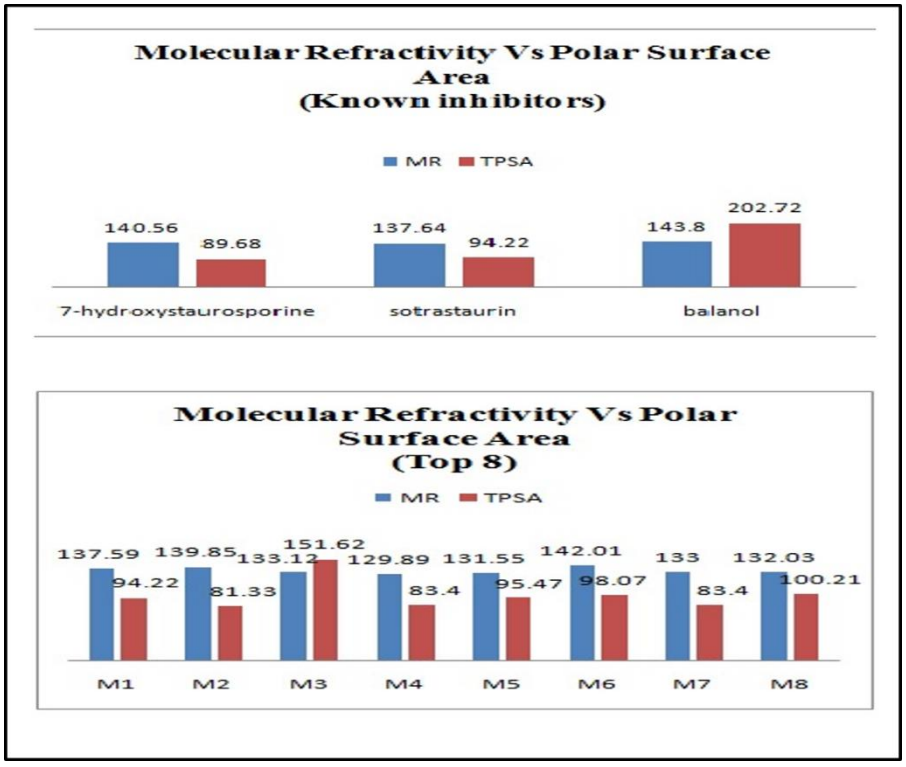

Figure 12: Graphical representation of molecular refractivity and polar surface area of top 8 compounds along with the known inhibitors.

4.6.2 Drug likeness

Drug-likeness generally assess different molecule qualitatively to validate them to become an oral drug based on its bioavailability and focuses to select best molecule. These kind of studies are routinely used for the purpose of extracting filtered information from large chemical libraries which helps in selection of most compatible molecules of pharmacokinetics profile. All the filters used in this section are being used by big pharma industries which allows to choose a better candidate which can be used as a drug. The Lipinski filter uses the famous ‘rule of 5’ and all of our constructed molecules followed this rule. Other filters are being used by Amgen (Ghose), Pharmacia (Egan) and Bayer (Muegge). These kind of estimations allows one to select best fitting method according to the demand of the project. The Bioavailability score by Abbot mainly predicts probability of all compounds to have threshold oral bioavailability in rat models.

	Lipinski violations	Ghose violations	Muegge violations	Leadlikeness violations	Synthetic Accessibility	Bioavailability Score
M1	0	1	1	1	3.61	0.55
M2	0	1	0	1	3.69	0.55
M3	0	1	1	1	3.91	0.55
M4	0	0	0	2	3.68	0.55
M5	0	2	0	2	3.7	0.55
M6	0	1	0	2	3.71	0.55
M7	0	2	0	2	3.73	0.55
M8	0	1	0	1	3.69	0.55

Table 5: Qualitative analysis of factors like Lipinski and bioavailability score of top constructed compounds.

Known Inhibitors	Lipinski violations	Ghose violations	Muegge violations	Leadlikeness violations	Synthetic Accessibility	Bioavailability Score
7hydroxy..	0	2	1	1	5.09	0.55
Sotrastaurin	1	0	1	1	3.68	0.55
Balanol	1	1	0	2	4.43	0.17

Table 6: Analysis of known inhibitors for evaluating its drug likeness that makes it a better candidate to be used as a drug.

4.6.3 Water Solubility

Soluble molecule facilitates numerous activities related to drug development which helps in formulation and handling of that molecule. Oral administration as well as solubility influence absorption which makes them important parameters while drug handling. Even drug administered should have high solubility in water so as to deliver small quantity of active ingredient as a proper dosage. SwissADME generates three parameters for water solubility.

ESOL (Estimating Solubility): Aqueous solubility is the main physical properties for many chemist which deals with medicines. Solubility affects the distribution of many biological compounds which directly has an effect on their efficacy. Many methods such as Monte Carlo simulations by Duffy and use of descriptors by Wegner are being used for predicting solubility solely from the molecular structure. Many research have stated the $LogP_{octanol}$ has a correlation with the solubility.

Ali Solubility: It uses a General Solubility Equation of the QSPR model which is solely based on melting point of chemical substance. It predicts the solubility of compounds that are nonionizable in nature. It cannot be applied on virtual compounds because this method is based on experimental descriptors like melting point.

	ESOL Solubility	ESOL Class	Ali Solubility	Ali Class	Silicos-IT Solubility	Silicos-IT class	GI absorption
M1	1.16E-05	Moderately soluble	2.02E-05	Moderately soluble	6.85E-09	Poorly soluble	High
M2	1.64E-05	Moderately soluble	2.76E-05	Moderately soluble	8.53E-09	Poorly soluble	High
M3	1.76E-05	Moderately soluble	9.45E-07	Poorly soluble	4.72E-08	Poorly soluble	Low
M4	2.31E-06	Moderately soluble	1.15E-06	Moderately soluble	3.30E-08	Poorly soluble	High
M5	3.85E-06	Moderately soluble	2.38E-06	Moderately soluble	8.45E-09	Poorly soluble	High
M6	2.19E-06	Moderately soluble	9.31E-07	Poorly soluble	8.61E-08	Poorly soluble	High
M7	1.46E-06	Moderately soluble	3.91E-07	Poorly soluble	2.34E-09	Poorly soluble	High
M8	1.91E-05	Moderately soluble	2.16E-05	Moderately soluble	4.37E-09	Poorly soluble	High

Table 7: Water solubility of top 8 constructed molecules based on ESOL model and Ali model.

	ESOL Solubility	ESOL Class	Ali Solubility	Ali Class	Silicos-IT Solubility	Silicos-IT class	GI absorption
7-hydroxy.	1.56E-05	Moderately soluble	5.94E-05	Moderately soluble	1.70E-07	Poorly soluble	High
Sotrastaurin	3.65E-05	Moderately soluble	5.64E-05	Moderately soluble	2.02E-08	Poorly soluble	High
Balanol	3.91E-04	Soluble	3.21E-05	Moderately soluble	6.85E-06	Moderately soluble	Low

Table 8: Water solubility of known inhibitors of novel class of PKC isozymes based on ESOL model and Ali model.

5. DISCUSSION

A good ligand based method for drug designing along with the machine learning approach proves to be an effective method in detecting the best lead for drug discovery. Initially the compounds related to our target of interest were collected. The selected compounds after removing discrepancies were left out with 1912 unique compounds which were treated as a training set for machine learning. These compounds were analysed according to 45 descriptors and based on their similarity, 20 new compounds were created by altering their chemical structure. These compounds were then used as test set for running machine learning on this data to identify the compounds which have the property to act as an inhibitor. Three models were created which were based on the physic chemical properties of balanol, sotrastaurin, and 7-hyroxystaurosporine. The accuracy of these 3 models were calculated out to be 82 % which is high as compared to previous models generated by other models and can be considered to create a specific inhibitor for novel class of PKC. The top 8 constructed compounds were classified as potential inhibitors based on their ADME properties.

All 4 isozymes of nPKC were not having well defined co-crystallised structure. PKC and PKC were searched for similar templates. Based on the top 50 templates the one having maximum sequence identity along with sequence similarity score were treated as best template and was taken to create a model of these proteins. These models were further used as target receptors for docking studies.

Thereafter molecular docking results of full fitness with potential inhibitors were analysed comparatively. All the selected compounds had a central Zn^{2+} atom at the binding site that was required for proper binding with the receptor. Out of 8 compounds only three M3, M6 and M8 were having greater binding energy then the nPKC isozymes. One of the compound on further analysis showed some distortion in binding site hence they were discarded. Hence M6 and M8 were selected as the most potent inhibitors in the study. Based on all these results these two compounds can be tested through in-vivo and in-vitro assays for conforming the inhibitory effect on nPKC for treating cancer.

6. CONCLUSION

We carried out our study by selecting the novel compounds by selecting 2644 compounds initially which were further used to create machine learning models through weka. 20 new compounds were constructed by the help of ADM Sketch by altering side chains and rings of different compounds and later on they were used as a test data set. By machine learning and virtual screening studies top 8 compounds were classified as inhibitors by three models. These 8 compounds were further analysed by calculating different descriptors. These compounds were finally docked with the target protein so as to confirm their higher potential to work as drug for treating cancer in future. Thus on the basis of in-silico studies 2 main compounds M6 and M8 proved to be the better candidates as their fitness energies were better than the know inhibitors of nPKCs and therefore have the potential to be used for in-vitro analysis to narrow the gap between the treatment of cancer by these selective ligands.

7. REFERENCES

1. E.M. Griner, M.G. Kazanietz, “Protein kinase C and other diacylglycerol effectors in cancer”, Nature Review Cancer, vol. 7, pp. 281- 294, 2007.
2. N. Isakov, A. Altman, “Protein kinase C (theta) in T cell activation”, Annu Rev Immunology, vol. 20, pp. 761-794, 2002.
3. D. Koya, G.L. King, “Protein kinase C activation and the development of diabetic complications”, Diabetes, vol. 47, pp. 859-866, 1998.
4. S.R. Sampson, D.R. Cooper, “Specific protein kinase C isoforms as transducers and modulators of insulin signalling”, Mol Genet Metabolism, vol. 89, pp. 32-47.
5. K. Inagaki, L. Chen and F. Ikeno, “Inhibition of delta-protein kinase C protects against reperfusion injury of the ischemic heart *in vivo*”, Circulation, vol. 108, pp. 2304-2307, 2003.
6. G.R. Budas, E.N. Churchill, R.D. Mochly, “Cardio protective mechanisms of PKC isozyme-selective activators and inhibitors in the treatment of ischemia-reperfusion injury”, Pharmacol Res, vol. 55, pp. 523-536, 2007.
7. M.K. Sun, D.L. Alkon, “Protein kinase C pharmacology: Perspectives on therapeutic potentials as antidementic and cognitive agents”, Recent Pat CNS Drug Discovery, vol. 1, pp. 147-156, 2006.
8. M.E. Lukashev and Z. Werb, “ECM signalling: orchestrating cell behaviour and misbehaviour”, Trends Cell Biology, vol. 8, pp. 437–441, 1998.
9. Y. Tanaka, M.V. Gavrielides, Y. Mitsuuchi, T. Fujii, and M.G. Kazanietz, “Protein kinase C promotes apoptosis in LNCaP prostate cancer cells through activation of p38 MAPK and inhibition of the Akt survival pathway”, Journal of Biol. Chemistry, vol. 278, pp. 33753–33762, 2003.
10. J.F. Liu, M. Crepin, J.M. Liu, D. Barritault and D. Ledoux, “FGF-2 and TPA induce matrix metalloproteinase-9 secretion in MCF-7 cells through PKC activation of the Ras/ERK pathway”, Biochem. Biophys. Res. Communication, vol. 293, pp. 1174–1182, 2002.
11. G.K. Lonne, K.C. Masoumi, J. Lennartsson and C Larsson, “Protein kinase C delta supports survival of MDA-MB-231 breast cancer cells by suppressing the ERK1/2 pathway”, Journal of Biol. Chem., vol. 284, pp. 33456–33465, 2009.

12. T. Fujii, A.M. Nakamura, G. Yokoyama, M. Yamaguchi, K. Tayama, “Antineoplaston induces G (1) arrest by PKCalpha and MAPK pathway in SKBR-3 breast cancer cells”, Oncology Report, vol. 14, pp. 489–494, 2005.
13. S.M. Nabha, S. Glaros, M. Hong, A.E. Lykkesfeldt, R. Schiff, “Upregulation of PKC-delta contributes to antiestrogen resistance in mammary tumor cells. Oncogene, vol. 24, pp. 3166–3176, 2005.
14. R.K. Lindemann, M. Braig, P. Ballschmieter, T.A. Guise, A. Nordheim, “Protein kinase Calpha regulates Ets1 transcriptional activity in invasive breast cancer cells”, International Journal of Oncology, vol. 22, pp. 799–805, 2003.
15. Q. Pan, L.W. Bao, L.W. Kleer, C.G. Sabel, M.S Griffith, “Protein kinase C epsilon is a predictive biomarker of aggressive breast cancer and a validated target for RNA interference anticancer therapy”, Cancer Res., vol. 65, pp. 8366–8371, 2005.
16. E. Toton, E. Ignatowicz, K. Skrzeczkowska, M. Rybczynska, “Protein kinase C epsilon as a cancer marker and target for anticancer therapy”, Pharmacology Reports, vol. 63, pp. 19–29, 2011.
17. Q. Pan, L.W. Bao, T.N.Teknos, S.D. Merajver, “Targeted disruption of protein kinase C epsilon reduces cell invasion and motility through inactivation of RhoA and RhoC GTPases in head and neck squamous cell carcinoma”, Cancer Res., vol. 66, pp. 9379–9384, 2006.
18. PA. Masso, G. Verstovsek, K. Darcy, C. Tagliarino, “Protein kinase C eta up regulation and secretion during postnatal rat mammary gland differentiation”, Eur. Journal of Cell Biology, vol. 77, pp. 48–59, 1998.
19. G. Karp, A. Maissel, E. Livneh, “Hormonal regulation of PKC: estrogen up-regulates PKC eta expression in estrogen-responsive breast cancer cells”. Cancer Lett., vol. 246, pp. 173–181, 2007.
20. N. Rotem, G. Oberkovitz, S. Ghanem, E. Livneh, “PKC eta confers protection against apoptosis by inhibiting the pro-apoptotic JNK activity in MCF-7 cells”, Exp. Cell Res, vol. 315, pp. 2616–2623, 2009.
21. P.A. Masso, J.S. Winston, S. Edge, K.M. Darcy, H. Asch, “Altered expression and localization of PKC eta in human breast tumors”, Breast Cancer Res. Treat., vol. 68, pp. 211–223, 2001.

22. K. Belguise, G.E. Sonenshein, “PKC theta promotes c-Rel driven mammary tumorigenesis in mice and humans by repressing estrogen receptor alpha synthesis”, Journal of Clinical Investigation, vol. 117, pp. 4009–4021, 200.
23. U. Kikkawa, Y. Takai, Y. Tanaka, R. Miyake and Y. Nishizuka, “Protein kinase C as a possible receptor protein of tumor promoting phorbol esters”, Journal of Biol. Chem., vol. 258, pp. 11442– 11445, 1983.
24. M. Castagna, Y. Takai, K. Kaibuchi, K. Sano, U. Kikkawa and Y. Nishizuka, “Direct activation of calcium-activated, phospholipid-dependent protein kinase by tumor promoting phorbol esters”, Journal of Biol. Chem., vol 257, pp. 7847– 7851, 1982.
25. A.W. Ashton, G. Watanabe, C. Albanese, E.O. Harrington, J.A. Ware and R.G. Pestell, “PKC delta inhibition of S-phase transition in capillary endothelial cells involves the cyclin-dependent kinase inhibitor p27 (Kip1)”, J. Biol. Chem., vol. 274, pp. 20805–20811, 1999.
26. SR. Cerda, R. Mustafi, H. Little, G. Cohen, S. Khare, C. Moore, “Protein kinase C delta inhibits Caco-2 cell proliferation by selective changes in cell cycle and cell death regulators”, Oncogene, vol. 25, pp. 3123–3138, 2006.
27. D.K. Bowles, K.K. Maddali, V.C. Dhulipala and D.H Korzick, “PKC delta mediates anti-proliferative, proapoptic effects of testosterone on coronary smooth muscle”, Journal of Cell Physiology, vol. 293, pp. 805– 813, 2007.
28. J.A. Vrana, A.M. Saunders, S.P. Chellappan and S. Grant, “Divergent effects of bryostatin1 and phorbol myristate acetate on cell cycle arrest and maturation in human myelomonocytic leukemia cells (U937)”, Differentiation, vol. 63, 33–42, 1998.
29. M. Nakagawa, J.L. Oliva, D. Kothapalli, A. Fournier, R.K. Assoian, and M.G. Kazanietz, “Phorbol ester induced G1 phase arrest selectively mediated by protein kinase C delta-dependent induction of p21”, Journal of Biol. Chem., vol. 280, pp. 33926–33934, 2005.
30. M.A. Pysz, O.V. Leontieva, N.W. Bateman, J.M. Uronis, K.J. Curry, D.W. Threadgill, “PKCα tumor suppression in the intestine is associated with transcriptional and translational inhibition of cyclin D1”, Exp. Cell Res., vol. 315, pp. 1415– 1428, 2009.
31. S. Fukumoto, Y. Nishizawa, M. Hosoi, H. Koyama, K. Yamakawa, S. Ohno, “Protein kinase C delta inhibits the proliferation of vascular smooth muscle cells by suppressing G1 cyclin expression”, Journal of Biol. Chem., vol. 272, pp. 13816–13822, 1997.

32. K. Kitamura, K. Mizuno, A. Etoh, Y. Akita,A. Miyamoto, K. Nakayama, “The second phase activation of protein kinase C delta at late G1 is required for DNA synthesis in serum-induced cell cycle progression”, Genes Cells, vol. 8, pp. 311–324, 2003.
33. G. Czifra, I.B. Toth, R. Marincsak, I. Juhasz, I. Kovacs, “Insulin like growth factor I coupled mitogenic signaling in primary cultured human skeletal muscle cells and in C2 myoblasts”, Cell. Signal, vol. 18, pp. 1461–1472, 2006.
34. C.H. Cho, C.S. Lee, M. Chang, I.H. Jang, S.J. Kim, I. Hwang, “Localization of VEGFR- 2 and PLD2 in endothelial caveolae is involved in VEGF-induced phosphorylation of MEK and ERK”, Journal of Heart Circ. Physiology, vol. 286, pp. 1881–1888, 2004.
35. J.L. Walker, P. Castagnino, B.M. Chung, M.G. Kazanietz and R.K. Assoian, “Post-transcriptional destabilization of p21 cip1 by protein kinase C in fibroblasts”, Journal of Biol. Chem., vol. 281, pp. 38127–38132, 2006.
36. V.C. Grossoni, K.B. Falbo, M.G. Kazanietz and A.J. Urtreger, “Protein kinase C delta enhances proliferation and survival of murine mammary cells”, Mol. Carcinogenic, vol. 46, pp. 381–390, 2007.
37. SF. Steinberg, “Distinctive activation mechanisms and functions for protein kinase C delta”, Biochem. Journal, vol. 384, pp. 449–459, 2004.
38. E. Balciunaite and A. Kazlauskas, “Early phosphor inositide-3-kinase activity is required for late activation of protein kinase C epsilon in platelet derived growth factor stimulated cells: evidence for signalling across a large temporal gap”, Biochemistry Journal, vol. 358, pp. 281–285, 2001.
39. K.M. Bae, H. Wang, G. Jiang, M.G. Chen, L. Lu and L. Xiao, “Protein kinase C epsilon is over expressed in primary human non-small cell lung cancers and functionally required for proliferation of non-small cell lung cancer cells in a p21/Cip1-dependent manner”, Cancer Res., vol. 67, pp. 6053–6063, 2007.
40. J.W. Soh and I.B. Weinstein, “Roles of specific isoforms of protein kinase C in the transcriptional control of cyclinD1 and related genes, Journal of Biol. Chem., vol. 278, pp. 34709– 34716, 2003.
41. L.C. Yang and D.D. Bikle, “Role of protein kinase C alpha in calcium induced keratinocyte differentiation: defective regulation in squamous cell carcinoma”, Journal of Cellular Physiology, vol. 195, pp. 249–259, 2003.

42. P.R. Webb, C. Doyle and N.G. Anderson, “Protein kinase C epsilon promotes adipogenic commitment and is essential for terminal differentiation of 3T3-F442A pre adipocytes”, Cell Mol. Life Sciences, vol. 60, pp. 1504–1512, 2003.
43. G.C. Gaboardi, G. Ramazzotti, A. Bavelloni, M. Piazzi, R. Fiume, A.M. Billi, “A role for PKCε during C2 C12 myogenic differentiation”, Cell. Signalling, vol. 22, pp. 629–635, 2010.
44. E.M. Genot, P.J. Parker and D.A. Cantrell, “Analysis of the role of protein kinase Cα, ε, and ζ in T cell activation”, Journal Biol. Chem., vol. 270, pp. 9833–9839, 1995.
45. D. Breitkreutz, D. Braiman, L. Wiksman, N. Daum, M.F. Denning and T. Tennenbaum, “Protein kinase C family: on the cross roads of cell signalling in skin and tumor epithelium”, Journal of Cancer Res.Clinical Oncology, vol. 133, pp. 793–808, 2007.
46. M. Kashiwagi, M. Ohba, K. Chida and T. Kuroki, “Protein kinase C eta (PKC eta) its involvement in keratinocyte differentiation”, J. Biochem., vol. 132, pp. 853–857, 2002.
47. G. Fu, J. Hu, N. Niederberger, V. Rybakin, J. Casas, P.P. Yachi, “Protein kinase C eta is required for T cell activation and homeostatic proliferation”, Science Signal. vol. 4, pp. 84-89, 2011.
48. E. Fima, M. Shtutman, P. Libros, A. Missel, G. Shahaf, G. Kahana, “PKC eta enhances cell cycle progression, the expression of G1 cyclin and p21 in MCF-7 cells”, Oncogene, vol. 20, 6794–6804, 2001.
49. W.B. Ou, M.J. Zhu, G.D. Demetri, D.M. Fletcher and J.A. Fletcher, “Protein kinase C [theta] regulates KIT expression and proliferation in gastrointestinal stromal tumors”, Oncogene, vol. 27, pp. 5624– 5634, 2008.
50. A. Altman, N. Isakov and G. Baier, “Protein kinase C theta: a new essential super star on the T-cell stage”, Immunology Today, vol. 21, pp. 567–573, 2000.
51. T. So and M. Croft, “Regulation of the PKC theta NFkappa B axis in T lymphocytes by the tumor necrosis factor receptor family member OX40”, Frontier Immunology, vol. 3, pp. 133-139, 2012.
52. K. Koide, M.E. Bunnage, G.L. Paloma, J.K. Kanter, S.S. Taylor, L.L. Runton and K.C. Nicolaou, “Molecular design and biological activity of potent and selective protein kinase inhibitors related to balanol”, Chem. Biology, vol. 2, pp. 601–608, 1995.
53. P. Kulanthaivel, Y.F. Hallock, C. Boros, S.M. Hamilton, W.P. Janzen, L.M. Ballas, C.R. Loomis, J.B. Juang, B. Katz, J.R. Steiner and J. Clardy,“ Balanol: A novel and potent inhibitor of protein kinase C from the fungus Verticillium balanoides” Journal of Chem. Soc., vol. 115, pp. 6452–6453, 1993.

54. H.M. Berman, “The Protein Data Bank,” Nucleic Acids Research, vol. 28, pp. 235-242,Jan. 2000.
55. R.A. Laskowski, “Enhancing the functional annotation of PDB structures in PDBSum using key figures extracted from the literature”, Bioinformatics, vol. 23, pp. 1824–1827, 2007.
56. M. Biasini, “SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information”, Nucleic Acids Research, vol. 42, pp. 252-258, 2014.
57. J.P. Evenou, P. Floersheim, C. Gaul, L. Monovich, G. Rummel, W. Schuler, W. Stark, A. Strauss, A. Matt, E. Vangrevelinghe, J. Wagner, N. Soldermann, “ 2,6-Naphthyridines as potent and selective inhibitors of the novel protein kinase C isozymes”, Bio organism medicinal Chemistry Letter, vol. 21, pp. 7367-7372, 2011
58. D.M. George, E.C. Breinlinger, M. Friedman, Y. Zhang, J. Wang, M. Argiriadi, M. Barth, Q. Lang, J.J. Edmunds, “ Discovery of selective and orally bioavailable protein kinase C theta (PKCθ) inhibitors from a fragment hit”, vol. 58, Journal of Med. Chem., pp. 222-236, 2015.
59. P. Bento, A. Gaulton, A. Hersey, L.J. Bellis, J. Chambers, M. Davies, F.A. Krüger, Y. Light, L. Mak, S. Mcglinchey, M. Nowotka, G. Papadatos, R. Santos and J.P. Overington, “The ChEMBL bioactivity database: an update,” Nucleic Acids Research, vol. 42, pp. 1083–1090, 2013.
60. T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson, “BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities,” Nucleic Acids Research, vol. 35, pp. 198–201, 2007.
61. ACD/ChemSketch Freeware, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2014.
62. “Molinspiration cheminformatics,” Choice Reviews Online, vol. 43, 2006.
63. J.J. Irwin, T. Sterling, M.M. Mysinger, E.S. Bolstad, R.G. Coleman, “ZINC: A Free Tool to Discover Chemistry for Biology”, Journal of Chemical Information and Modelling, vol. 52(7), pp. 1757-1768, 2012.
64. A. Diana, O. Michielin, V. Zoete, “SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules”, Science Report, vol. 7, 42717-42727, 2017.

65. V. Zoete, A. Daina, C. Bovigny, O. Michielin, “SwissSimilarity: A Web Tool for Low to Ultra High Throughput Ligand-Based Virtual Screening”, “Journal of Chemical information and modelling, vol. 56, pp. 1399-1404, 2016.
66. E. Frank, M.A Hall, H. Witten, “The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Fourth Edition, 2016.
67. A. Grosdidier, V. Zoete, O. Michielin, “SwissDock, a protein-small molecule docking web service based on EA Dock DSS”, Nucleic Acids Research, vol. 39, pp. 270-277, 2011.
68. N.M. Boyle, M. Banck, C.A. James, G.R. Hutchinson, “Open Babel: An open chemical toolbox”, Journal of Cheminformatics, vol. 3, pp. 1758-1769, 2011.