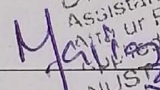


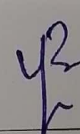
THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MBA/MS/MPhil Project/Thesis written by Ms. Areeba Rehman (00000320793), of Atta-ur-Rahman School of Applied Biosciences (ASAB) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulation, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed by GEC members of the scholar have also been incorporated in the said thesis.

Signature: 
Dr. Maria Shabbir
Assistant Professor
Atta-ur-Rahman School of
Applied Bioscience (ASAB)
NUST Islamabad

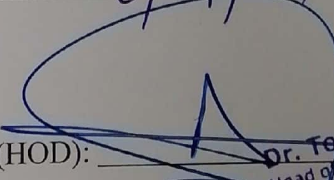
Name of Supervisor: Dr. Maria Shabbir

Date: 6-9-21

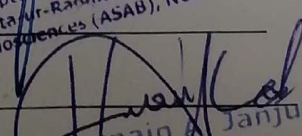
Signature: 
Dr. Yasmin Badshah
Assistant Professor
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST Islamabad

Name of Co-Supervisor: Yasmin Badshah

Date: 6/9/21

Signature (HOD): 
Dr. Fougeer Ahmed
Head of Department (HOD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Date: _____

Signature (Dean/Principal): 

Date: _____

Dr. Hussain Janjua
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST, Islamabad

National University of Sciences & Technology**MS THESIS WORK**

We hereby recommend that the dissertation prepared under our supervision by:
(Student Name & Regn No.) Areeba Rehman & Reg No. 00000320793

Titled: Elucidating the role of missense SNP of Protein Kinase C epsilon in HCV-induced Hepatocellular Carcinoma be accepted in partial fulfillment of the requirements for the award of MS Degree in Healthcare Biotechnology degree with (A grade).

Examination Committee Members

1. Name: Dr. Rumeza Hanif

Signature: [Signature]

2. Name: Dr. Sobia Manzoor

Signature: [Signature]

3. Name: Dr. Erum Dilshad

Signature: [Signature]

Co-Supervisor's name: Mrs. Yasmin Badshah

Signature: [Signature]

Supervisor's name: Dr. Maria Shabbir

Signature: [Signature]

Date: 12-8-21

[Signature]
Dr. Touqeer Ahmed
 Head of Department (HOD)
 Dept of Healthcare Biotechnology
 Atta-ur-Rahman School of Applied
 Biosciences (ASAB), NUST Islamabad

Head of Department

12/8/2021
 Date

COUNTERSIGNED

Date: _____

[Signature]
Dean/Principal
 Atta-ur-Rahman School of
 Applied Biosciences (ASAB)
 NUST, Islamabad

Certificate for Plagiarism

It is certified that MS Thesis Titled "Elucidating the role of missense SNP of Protein Kinase C epsilon in HCV-induced Hepatocellular Carcinoma" by Areeba Rehman has been examined by me. I undertake the follows:

- a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
- b. The work presented is original and own work of the author (i.e., there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
- c. There is no fabrication of data or results which have been compiled / analyzed.
- d. There is no falsification by manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Supervisor name: Dr. Maria Shabbir

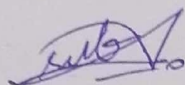
Signature: _____

Date: _____

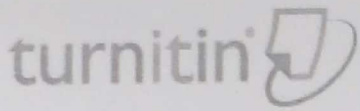
Maria Shabbir
Assistant Professor
Atta ul Ullah School of
Applied Bioscience (ASAB)
NUST Islamabad

Declaration

I hereby declared that except where Specific reference is made to the work of others, the content of this dissertation is original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other University. This dissertation is the result of my own work and includes the outcome of the work done.



Areeba Rehman



Digital Receipt

Maria Shabbir
Assistant Professor
School of
Health Bioscience (ASAR)
Islamabad

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

- Submission author: Areeba Rehman
- Assignment title: Elucidating the role of missense SNP of Protein Kinase C eps...
- Submission title: Elucidating the role of missense SNP of Protein Kinase C eps...
- File name: Areeba_thesis_for_plagiarism_checking.docx
- File size: 98.92K
- Page count: 48
- Word count: 12,582
- Character count: 66,200
- Submission date: 13-Sep-2021 04:41AM (UTC-0700)
- Submission ID: 1647341394

Elucidating the role of missense SNP of Protein Kinase C epsilon in HCV-induced
Hepatocellular Carcinoma

Abstract

Hepatocellular carcinoma is a major health concern for the past few decades because of its chemoresistance aggressive nature and misdiagnosis. Identification of pathogenic SNP in PKCε can be a potential biomarker for early diagnosis and treatment of HCC. In this study we identify a PKCε variant associated potential pathogenic SNPs of PKCε and selected a high risk variant for HCV-induced HCC. The association of viral pathogenicity was also checked with the identified potentially pathogenic variant of PKCε. The *in silico* structural and functional analysis of the selected SNP was also carried out using different online tools and molecular dynamic simulation. Lastly, the analysis of molecular interaction of PKCε with a potential inhibitor, bosutinib was performed to estimate its therapeutic potential.

**Elucidating the role of missense SNP of Protein Kinase C epsilon in HCV-
induced Hepatocellular Carcinoma**



By

Areeba Rehman

00000320793

Supervisor

Dr. Maria Shabbir

Co-Supervisor

Mrs. Yasmin Badshah

**Atta-ur-Rehman School of Applied Biosciences
National University of Science and Technology
Islamabad, Pakistan**

2021

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MBA/MS/MPhil Project/Thesis written by Ms.

Areeba Rehman (00000320793), of Atta-ur-Rahman School of Applied Biosciences (ASAB) has been vetted by undersigned, found complete in all respects as per NUST Statutes/Regulation, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: _____

Date: _____

Signature: _____

Name of Co-Supervisor: _____

Date: _____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal): _____

Date: _____

Certificate for Plagiarism

It is certified that MS Thesis Titled “Elucidating the role of missense SNP of Protein Kinase C epsilon in HCV-induced Hepatocellular Carcinoma” by Areeba Rehman has been examined by me. I undertake the follows:

- a. Thesis has significant new work/knowledge as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
- b. The work presented is original and own work of the author (i.e., there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
- c. There is no fabrication of data or results which have been compiled / analyzed.
- d. There is no falsification by manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- e. The thesis has been checked using TURNITIN (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Name & Signature of Supervisor

Supervisor name: _____

Signature: _____

Date: _____

Declaration

I hereby declared that except where Specific reference is made to the work of others, the content of this dissertation is original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other University. This dissertation is the result of my own work and includes the outcome of the work done.

Areeba Rehman

Table of Contents

Abstract	1
Chapter 1	2
Introduction	2
Chapter 2	5
Literature Review	5
I. PKC ϵ ; A unique member of PKC family	7
II. PKC ϵ in cell cycle: Effects of its upregulation and downregulation.....	12
III. PKC ϵ dysregulation in different cancer types	13
IV. Molecular Targets and interactions of PKC ϵ	17
V. PKC ϵ Inhibitors	20
Chapter 3	22
Material and Methods	22
Chapter 4	
Results.....	33
Chapter 5	
Discussion	61
Conclusion	67
References	68

List of Tables

Table 1. Types of PKC ϵ variants and their numbers.....	42
Table 2. Potentially harmful SNPs filtered through SIFT, PolyPhen, REVEL, MetaLR and Mutation Assessor.....	43
Table 3. Screened SNPs through CADD and their classification in all the applied tools.....	44
Table 4. Pharmacokinetic properties of Betulin obtained from admetSAR.....	48
Table 5. Physio-chemical properties of Betulin obtained from admetSAR.....	49
Table 6. Regression properties of Betulin obtained from admetSAR.....	49
Table 7. Different binding modes along with vina scores and cavity sizes of model 5 of PKC ϵ and Betulin	50
Table 8. Statistical analysis of PRKCE SNP data obtained from PCR.....	58
Table 9. Statistical analysis of PCR data of PRKCE SNP with respect to gender.....	59
Table 10. Statistical analysis of PCR data of PRKCE SNP with respect to gender.....	59

List of Figures

Figure 1. Structure of PKC ϵ	12
Figure 2. Involvement of PKC ϵ in various carcinogenic processes.....	16
Figure 3. A general picture of PKC ϵ interaction with several adaptors.....	20
Figure 4. Flow chart of in silico methodology followed in this study	22
Figure 5. Flow chart of wet lab methodology followed in this study	22
Figure 6. Predicted model 2 of PKC epsilon	35
Figure 7. Predicted model 5 of PKC epsilon.	36
Figure 8. Superimposed structures of mutated and wild type PKC ϵ	37
Figure 9. Phylogenetic tree of PKC protein family	38
Figure 10. Subcellular localization of PKC ϵ	38
Figure 11. Change in the amino acid residue from glutamic acid to lysine.....	47
Figure 12. Chemical structure of Betulin obtained from PubChem	48
Figure 13. Surface view of PKC ϵ and Betulin.....	51
Figure 14. LigPlot visualization of docked Betulin	52
Figure 15. 3D visualization of docked PKC ϵ and Betulin on PyMOL.....	53

List of Abbreviations

AA	Arachidonic Acid
ADMET	Absorption Distribution Metabolism Excretion and Toxicity
AGC	cAMP-dependent, cGMP-dependent and protein kinase C
ALT	Alanine Transaminase
AML	Acute Myeloid Leukemia
aPKC	Atypical Protein Kinase C
ARMS	Amplification-refractory Mutation System
ATF2	Activating Transcription Factor 2
BID	BH3-Interacting Domain
Cdk	Cyclin-Dependent Kinase
cDNA	Complementary DNA
CG-NAP	Centrosome- and Golgi-localized protein kinase N Associated Protein
CI	Confidence Interval
cPKC	Conventional Protein Kinase C
CT	Cycle Threshold
DAG	Diacylglycerol
dNTP	Deoxynucleoside triphosphate
E2F	E2 factor
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial–Mesenchymal Transition
ENH	Enigma Homolog

ERK	Extracellular signal-regulated Kinase
FASTA	FAST-All
GTP	Guanosine Triphosphate
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HOPE	Have Our Protein Explained
iCCA	Intrahepatic Cholangiocarcinoma
IKK	I κ B kinase
I-TASSER	Iterative Threading ASSEmby Refinement
MAPK	Mitogen Activated Protein Kinase
MD	Molecular Dynamics
MMP	Matrix Metalloproteases
mTORC2	Mammalian Target of Rapamycin Complex 2
NEMO	NF-kappa-B essential modulator
NF-kB	Nuclear Factor Kappa Light Chain Enhancer of Activated B cells
nPKC	Novel Protein Kinase C
NSCLC	Non-small Cell Lung Cancer
OR	Odd Ratio
P53	Protein 53
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PE	Phorbole Ester
PI3K	Phosphoinositide 3-kinase

PKC	Protein Kinase C
PKC ϵ	Protein Kinase C epsilon
PKC δ	Protein Kinase C delta
PLC	Phospholipase C
PS	Phosphatidylserine
qRT-PCR	Quantitative Reverse Transcriptase Polymerase Chain Reaction
RACK	Receptor for Activated C Kinase
RCC	Renal Cell Carcinoma
Rg	Radius of Gyration
RI	Reliability Index
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
RNA	Ribonucleic Acid
RO5	Rule of 5
RR	Relative Risk
RT	Reverse Transcriptase
SDF	Standard Data Format
SDS	Sodium Dodecyl Sulphate
SFTP	Secure File Transfer Protocol
siRNA	Small interfering RNA
SMILES	Simplified Molecular Input Line Entry System
SNP	Single Nucleotide Polymorphism
STAT3	Signal Transducer and Activator of Transcription 3

TACE	TNF α -Converting Enzyme
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TGF- β	Transforming Growth Factor Beta 1
TNFR1	Tumor Necrosis Factor Receptor 1
TNF α	Tumor Necrosis Factor alpha
TRAIL	TNF Related Apoptosis Inducing Ligand
TRAMP	Thrombospondin Related Apical Membrane Protein
UTR	Untranslated region
UVR	Ultraviolet Radiation
XIAP	X-linked Inhibitor of Apoptosis Protein
ZF	Zinc Finger

ACKNOWLEDGEMENTS

In the name of Allah, the Most Merciful, the Most Kind. All praises and thanks to Allah Almighty Who has bestowed His countless blessings and gifts upon me and without His mercy I would not have been able to complete this thesis.

I would like to express my sincere gratitude to my supervisor and co-supervisor Dr. Maria Shabbir and Mrs. Yasmin Badshah for their guidance and motivation that made this project possible.

I would also like to thank my friends and colleagues for their constant support, encouragement, and help throughout the project.

Lastly, a special thanks to my parents as this journey would not have been possible without their love support and guidance.

Thank you.

Areeba Rehman

MS HCB

ASAB, NUST

August 12, 2021

Abstract

Hepatocellular carcinoma is a major health concern for the past few decades because of its chemoresistance aggressive nature and misdiagnosis. Identification of pathogenic SNP in PKC ϵ can be a potential biomarker for early diagnosis and treatment of HCC. In this study we identify potential pathogenic SNPs of PKC ϵ and isolated a high-risk variant for HCV-induced HCC. Wet lab validation through tetra ARMS PCR of that SNP carried out afterwards. The association of viral pathogenicity was also checked with the identified potentially pathogenic variant of PKC ϵ . The *in silico* structural and functional analysis of the selected SNP was also carried out using different online tools and molecular dynamic simulations. Lastly, the analysis of molecular interaction of PKC ϵ with a potential inhibitor, Betulin, was performed to estimate its therapeutic potential.

CHAPTER 1

Introduction

Single nucleotide polymorphisms (SNPs) are the kind of polymorphism which signify the variation of a single base pair in a particular genetic sequence. SNPs can be present in different regions of genes and have been frequently held responsible for the susceptibility of developing many diseases. Most significant among those diseases are the complex disorders such as cancers which have a wide range of factors including SNPs, aiding in their development (Shastry, 2007).

The location of SNPs in different regions of genes determines their effects on the expression of genes and their role in cancer susceptibility. SNPs in both the coding (exons) and non-coding (introns) sequence of genes play their role in this regard. SNPs in the promoter regions, 3' UTR, 5' UTR and other non-coding regions can affect gene expression in multiple ways such as by enhancing or decreasing promoter activity, histone modifications, DNA methylation and mRNA splicing. Exonic SNPs on the other hand can either cause change in the encoded amino acid or not depending upon the location and type of the changed bases in a codon. Change in the first two bases of the codon often leads to the alteration of amino acid encoded by it, thus named as non-synonymous or missense SNPs. However, due to the degeneracy of genetic code, change in the third base of a codon does not lead to amino acid substitution at most instances. Non-synonymous or missense SNPs can cause increased susceptibility to cancer by making changes in the structure and function of proteins (Deng, Zhou, Fan, & Yuan, 2017).

The protein PKC ϵ has been positively associated with cancers in several studies. However, those studies usually indicate its involvement in the carcinogenic process at functional level. Any genetic association, such as that of the presence of SNPs in the Protein Kinase C Epsilon (PRKCE) Coding gene, between PKC ϵ and cancer susceptibility and progression is yet to be found. PKC ϵ is

a member of large family of serine threonine kinases. It is activated and relocated to specific compartments in the cell depending on the secondary messengers activated and the adaptor proteins activated in response to the extracellular stimulus activation of different receptors. Receptors like tyrosine kinase receptors, tyrosine kinase-coupled receptors and G-protein-coupled receptors can lead to PKC ϵ induction which in turn is involved in various physiological process such as immune system activation, inflammatory responses, , exocrine, endocrine and nervous system stimulation (Akita, 2002).

PKC ϵ is a phorbol ester dependent and calcium independent protein kinase. Numerous studies have implicated PKC ϵ in various carcinogenic processes. The transforming and survival potential of PKC ϵ has been observed in neoplastic growth has been observed in multiple animal model studies. The role of PKC ϵ in endothelial to mesenchymal transformation (EMT) may contribute to the invasive, migratory and metastatic properties of the cancer cells (Jain & Basu, 2014). In liver cancer, progressively elevated PKC ϵ levels were observed in the hyperplastic cells of rat liver in comparison to the normal liver cells which might have been associated with their malignant development (Perletti et al., 1996). Besides that, well defined role of PKC ϵ in liver cancer has not been documented yet; however, seeing its involvement in numerous other cancers, its role in liver cancer has also been hypothesized.

Liver cancers have been classified into various types and subtypes depending on the tissues of their origin. The six main type of liver cancer include hepatocellular carcinoma, intrahepatic cholangiocarcinoma, bile duct cystadenocarcinoma, combined hepatocellular and cholangiocarcinoma, hepatoblastoma and undifferentiated carcinoma. Hepatocellular carcinoma is the most frequently occurring type of liver cancer especially in developing countries like Pakistan. The risk factors associated with it include environmental exposures and hepatitis B and

C viruses (Srivatanakul, Sriplung, & Deerasamee, 2004). Any genetic association specifically that of PKC ϵ SNPs with the increased susceptibility of developing hepatocellular carcinoma is yet to be found. Therefore, the main focus of this study is to identify the potentially damaging SNPs present in PKC ϵ and investigate and develop their association with liver cancer. This will help us ascertain significant information about PKC ϵ which would most likely enable the establishment of a new prognostic marker for liver cancer and a potential therapeutic target.

For a potential therapeutic agent, we analyze the interaction of Betulin, a pentacyclic triterpenoid, with PKC ϵ . Betulin is commonly found in birch trees (*Betula* spp.). It has been previously studied for its anti-neoplastic activity in different cancer cell lines. Its potential as an anti-inflammatory agent and anticancer agent has been found proving high efficacy and improved safety in different cancers.

CHAPTER 2

Literature Review

Cancer development and progression is a complex biological phenomenon manifested by the alteration in the normal morphology and physiology of cells due to changes in the signaling pathways which regulate the survival, growth and differentiation of cells. The classification of cancer into different types is typically done on the basis of their primary tissue or organ of origin such as liver cancer. Usually, tumors originating in liver are heterogeneous in nature because of their cellular morphology resembling different types of hepatic cells. However, some major subtypes of primary liver cancer depending on the type of cells which undergo neoplastic transformation are; hepatocellular carcinoma (HCC), intrahepatic cholangiocarcinoma (iCCA) and mixed hepatocellular cholangiocarcinoma (HCC-CCA). Both HCC and iCCA develop from the cells that originate from the common progenitor cells; hepatoblast which mature into hepatocytes and cholangiocytes (Sia, Villanueva, Friedman, & Llovet, 2017). HCC is one of the most frequently occurring types of cancer worldwide and the common risk factors associated with it are viral hepatitis, drugs and alcohol abuse, or non-alcoholic fatty liver disease which may lead to the genetic alterations leading to neoplastic development of cells. Besides that, the susceptibility to develop liver cancer may increase due to the presence of single or multiple gene disorders in individuals such as hemochromatosis and hypothyroidism respectively (Dragani, 2010).

Whatever the cause maybe, it inevitably affects the genomic integrity of cells. As a result, several modulators of cellular functions are either up regulated or down regulated which hampers the normal functionality of cells leading to the neoplastic development. The risk of neoplastic development can also be markedly increased due to the presence of Single Nucleotide Polymorphisms (SNPs) inside the coding or regulatory sequence of adaptor proteins that are

responsible for the survival of cells. One such protein that has been of great interest in cancer studies is Protein Kinase C (PKC) which is a serine threonine kinase and has significant oncogenic properties. Studies have implicated the isoform of this family of enzymes in progression of different types of cancers. PKC has a number of isoforms which are categorized into three groups: the 'conventional' or 'classical' PKCs (cPKCs), 'typical' or 'novel' PKCs (nPKCs) and 'atypical' PKCs (aPKCs), based on their structural and functional similarities (Newton, 1995). The general observation about PKC is that the enzymes of this family are involved in essential processes of cell survival, proliferation, differentiation and motility (Isakov, 2018). However, some of the isozymes may play completely opposite roles; for example, PKC ϵ and PKC δ that belong to the same subfamily of nPKCs have opposing effects on cellular growth, differentiation and apoptosis. PKC δ is pro-apoptotic while the raised levels of PKC ϵ have been associated with many types of cancers (Duquesnes, Lezoualc'h, & Crozatier, 2011). Protein kinases have been frequently identified as prominent mediators of oncogenesis. The analysis across various cancer types for the presence of mutations in different protein kinases has provided significant evidence which gives insight about their roles in oncogenic processes (Greenman et al., 2007; Newton, 2010). Isoforms of PKC especially PKC ϵ is one of the proteins with greatest oncogenic potential as several researches have demonstrated the increased levels of PKC ϵ in multiple cancers and have associated it with aggressive phenotype of most solid tumors (Isakov, 2018). Although evidence points towards a significant role of PKC ϵ in oncogenesis but so far the data available only relates this with the aberrant regulation of its upstream and downstream elements and very little is known about the relation of mutations and polymorphisms inside the Protein Kinase C Epsilon Coding (PRKCE) gene with the susceptibility of cancer development and progression (Gorin & Pan, 2009; Jain & Basu, 2014).

I. PKC ϵ ; A unique member of PKC family

A. Discovery and structure

The first PKCs to be identified were cPKCs which were purified from the bovine brain using cDNA libraries in 1986 (P. J. Parker et al., 1986). The following year, using the probes from cDNA of each of the cPKC of rat brain, three additional PKC isotypes δ , ϵ , and ζ were isolated under low stringency conditions (Ono et al., 1987). Human PKC ϵ was first isolated in early nineties when two independent cDNA libraries of brain were screened for the clones E1 and E2. These clones were then subjected to sequence comparison and it was concluded that they were the same and their differences were merely artifacts of cloning. Thus, a single protein named PKC ϵ was made from these two clones which had 90-92% nucleotide and 98-99% amino acid similarity with the previously isolated rat, mouse, and rabbit PKC ϵ sequences (Basta et al., 1992).

The chromosomal location of PRKCE is 2p21 comprising of 15 exons which make a 737 amino acid protein. The basic structure of PKC ϵ can be divided into three main components, the C-terminal catalytic/kinase domain, the N-terminal regulatory region, and hinge region connecting the two. The protein structure is composed of conserved domains, which are present in almost all of the other members of PKC family, interspersed with variable regions. The regulatory region of PKC ϵ , like other PKCs, contains a pseudosubstrate segment which acts as a molecular switch and performs an autoinhibitory function by positioning itself inside the substrate binding cavity. This positioning is important for the regulation of function of the protein and is determined by the presence or absence of molecules that specifically target PKC ϵ and bind with it. This protein segment is composed of amino acid residues which resemble the substrate of PKC ϵ except for the presence of an Ala residue in the phospho-acceptor site.

Despite the structural similarity with substrate, due to the low affinity of this pseudo-substrate it cannot effectively inhibit the function of protein in cells (Newton, 2018).

One of the conserved domains of PKC ϵ , like most other isozymes of PKC family, is C1 located in the regulatory region, it has a high affinity for diacylglycerol (DAG) and molecules similar to it such as phorbol-esters (Newton, 2018). An unzipped β -sheet present in this domain makes a groove where these molecules bind (Duquesnes et al., 2011). Studies have demonstrated that DAG competes with the tumor-promoting phorbol-esters (TPA) or PMA (another name for TPA) to bind to PKC which leads to conclusion that they both have the same binding site on PKC proteins. The C1 domain contains tandem repeats named C1a and C1b which bind to these molecules except for the fact that only one domain accomplishes full length binding. These repeats are the conserved pattern of His and Cys residues which make two structures called zinc fingers, for interacting with the zinc ions. These structures are also responsible for localizing the protein to the Golgi apparatus (MELLOR & PARKER, 1998). The binding affinity of C1b of nPKCs with DAG is double that of cPKCs which aids in their Ca⁺² independent induction in response to elevated DAG levels. This increased affinity is due to the presence of Trp instead of Tyr in the nPKCs C1b ligand-binding loop. It has been observed that the presence of Trp gives a more closed and restricted structure to the binding pocket which allows for a stronger interaction of protein with small molecules such as DAG. The larger phorbol-esters; however, are not affected by the dynamics of ligand-binding loops (Newton, 2018).

The C2 domain of nPKCs also located in the regulatory region is not like the C2 domain of cPKCs. It lacks the acidic residue, Asp, which is required for interacting with the Ca⁺² ions and the folding of its β -strand is circularly permuted resulting in the different positions of C-

and N- termini in the fold (MELLOR & PARKER, 1998; Newton, 2018). The activation of nPKCs through the binding of their C2 domain with phospholipids, *in vitro*, has been reported. This domain also plays significant role in the interaction and translocation of the protein to the plasma membrane and other cellular compartments (MELLOR & PARKER, 1998). The loops in this domain joining $\beta 1$ – $\beta 2$ and $\beta 5$ – $\beta 6$ strands are important for binding of the protein to the anionic membranes, specifically the residues in loop 1 and 3 which are necessary for C2 binding with vesicles containing phosphatidic acid (Duquesnes et al., 2011). RACK (Receptors for Activated C Kinase) β -COP is a PKC ϵ specific protein responsible for its vesicular trafficking and subcellular translocation. A sequence of eight amino acids inside C2 domain, homologous to the β -COP facilitates their interaction (Duquesnes et al., 2011; MELLOR & PARKER, 1998). The presence of C2 domain is necessary for the protein-protein interactions, as the inhibition of this domain has shown to inhibit cellular translocation and thus, the function of PKC ϵ (Duquesnes et al., 2011).

The kinase domain of the catalytic region of PKC ϵ is highly conserved in the protein family. It contains all the components of a kinase required for the catalytic process. It contains an activation loop, a highly conserved sequence surrounding Ser/Thr residues, the phosphorylation of which is necessary to re-align the active site residues (Duquesnes et al., 2011). Only a couple of inhibitors have been developed for the cellular studies of PKC family isozymes, due to the highly similar active site structure of PKCs and other enzymes of its superfamily, AGC (Newton, 2018).

The V5 region or the C-terminal tail of PKC ϵ contains the turn and hydrophobic motifs, two highly conserved phosphorylation sites significant for the protein stability and protection from degradation (Newton, 2010). A conserved motif PXXP, that allows for priming

phosphorylation of the protein is also in this region (Newton, 2018). The 7-21 highly variable residues at the extreme carboxyl terminal have been used to develop protein specific antibodies (Duquesnes et al., 2011). The V5 domain of PKCs is essential for their tethering and maturation in addition to providing surface for docking of regulatory proteins, structuring core, and ATP and substrate binding (Newton, 2018). Using the chimeras of PKC ϵ and PKC δ having swapped their V5 region, the significance of this domain in localizing these proteins to their respective cellular compartments was demonstrated (Duquesnes et al., 2011).

The hinge region acts a site for caspase dependent cleavage of PKC ϵ . It also allows the association of protein with the plasma membrane and components of cytoskeleton (Lehel et al., 1995). It contains three phosphorylation sites, the Ser residues through which PKC ϵ interacts with the 14-3-3 proteins and is activated by them (Duquesnes et al., 2011).

B. Cellular localization

The activation and translocation of nPKCs is mainly done by DAGs which interacts with C1 domain. Phorbol-esters also interact with C1 domain but the selectivity of binding for both DAG and TPA may vary at C1a and C1b regions (Akita, 2002; Duquesnes et al., 2011). PKC ϵ can also be activated by phosphatidylinositol 3,4,5-triphosphate (PIP3) and fatty acids produced in response to different cellular stimuli; for example, PDGF (platelet derived growth factor) bradykinin, to generate response and help in the processes like cell adhesion, motility, growth and apoptosis (Akita, 2002). The abundance of DAG on Golgi membranes primarily localizes PKC ϵ towards them (Newton, 2018). Generally, for membrane interactions, all PKCs need an acidic lipid, phosphatidylserine (PS) which is present only on the cytoplasmic surface of the membranes. The DAGs bind to C1 domain and PS binds to C2 domain of the PKCs mediating their membrane translocation (Newton, 1995). Increased translocation and

constitutive presences of PKC ϵ was observed on the mitochondrial membranes useful for cardio-protection (Duquesnes et al., 2011).

The responsiveness of PKCs to secondary messengers depends on their phosphorylation at three specific conserved locations: the activation loop Thr-566, C-terminal hydrophobic site Ser-729, and an auto-phosphorylation site The-710. The immature PKC ϵ which is either non-phosphorylated or hypo-phosphorylated, directly binds to anchoring proteins like centrosome and Golgi localized PKN- associated protein (CG-NAP) through its catalytic domain followed by the phosphorylation of Ser-729 and Thr-566 at centrosome/Golgi surface. These phosphorylations are important for the enzymatic activity and stable binding of PKC ϵ to its secondary messengers. The type of secondary messenger binding the C1 domain of PKC ϵ determines to some extent the subcellular localization of the protein. DAG and tridecanoic acids binding localize PKC ϵ at the cytoskeleton and/or plasma membrane whereas the binding of linoleic acids and arachidonic acid (AA) translocates it to Golgi network. Similarly, the adaptor proteins of PKC ϵ also affect its localization such as β -COP translocates it to the Golgi apparatus.

An actin binding binding motif inside the C1 domain of PKC ϵ in the middle of the first and second Cys-rich regions, is responsible for the protein's interaction with the actin filaments when stimulated by DAG and AA. The actin filaments not only serve as the anchor protein for PKC ϵ but also maintains its conformation in a catalytically active state (Akita, 2002).

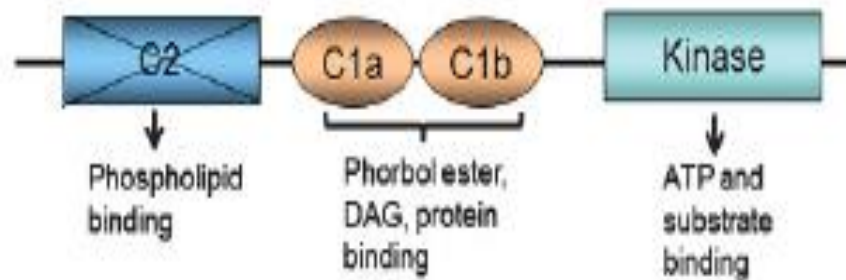


Figure 1: Structure of PKCε (Garg et al., 2014).

II. PKCε in cell cycle: Effects of its upregulation and downregulation

The role of PKCε has been observed at different stages of cell cycle. PKCε interaction with 14-3-3, the ubiquitous regulatory proteins, is important for the final stage of cell cycle when in the separation of two daughter cells takes place. The enzyme PKCε contains specific phosphorylation sites within its hinge region through which it interacts with 14-3-3. This interaction maybe direct such as at Ser368 or indirect: at Ser350 via p38 MAP kinase and Ser346 via GSK3. The importance of PKCε interaction with 14-3-3 was signified when the process of cytokinesis was inhibited due to the mutations preventing it (Duquesnes et al., 2011).

The up regulation of PKCε in the NIH3T3 cells of mouse was observed to increase the saturation density of cells while decreasing their doubling time of cell population. Similarly, PKCε overexpression in the rat colonic epithelial cells resulted in the uncontrolled growth of cells leading to neoplastic transformation (J. D. Black, 2000). In Non-small-cell lung carcinoma (NSCLC) the inhibition of PKCε caused depression in the transition of cells from G1-S phase in p21^{Cip1} (inhibitor of cyclin-dependent kinase (Cdk)) dependent manner (Bae et al., 2007; Gorin & Pan, 2009). In the presence of PMA and elevated levels of some PKC isozymes in late G1 phase of cell cycle, the cell does not proceed to the synthesis phase due to the inhibition of E2F to release from its Rb complex and bind the E2F binding sequences, a necessary step for progression of cell from

G1-S phase. Using the reporter gene for E2F, it was proved that PKC ϵ overexpression in rat 3Y1 fibroblasts, enhanced E2F activity rather than inhibiting it, which was done by some other isozymes of PKC family (Livneh & Fishman, 1997). This establishes the fact that PKC ϵ is important for the proliferation of cells (A. R. Black & Black, 2013).

Usually, the levels of PKC ϵ are lower during the cellular differentiation than the proliferation process. However, the role of PKC ϵ in cell differentiation and cell cycle arrest was supported by the studies in malignant colon, breast and pancreatic cancers where its expression was observed to be lower than the normal tissue (Ali, Ali, El-Rayes, Philip, & Sarkar, 2009).

A contrasting effect of PKC ϵ on cyclin D1, a tumor promoter, has been reported in two different studies. The NIH3T3 cells with constitutive expression of mutant PKC ϵ showed increased expression of cyclin D1 along with cyclin E, c-fos, c-jun and elevated growth rate of cells (Soh & Weinstein, 2003). Whereas the overexpression of PKC ϵ in transformed R6 fibroblasts (rat embryo fibroblast cell line) has demonstrated the reduced induction of cyclin D. Cyclin A level was also decreased and a marked reduction in c-fos expression was also seen while the expression of c-jun and c-myc remained unchanged. This discrepancy is thought to be due to the translational or post-translational effects on cyclin D1; however the results indicate that PKC ϵ alone when overexpressed might bypass the normal physiological pathways of cell to contribute in the oncogenic process (Han, Cacace, Sgambato, & Bernard Weinstein, 1995).

III. PKC ϵ dysregulation in different cancer types

The involvement of PKC ϵ with cancer development and progression has been investigated and evidenced through many studies. PKC ϵ was found to aid cancer cell proliferation, differentiation, survival and motility in both cancer patients and the *in vivo* and *in vitro* studies. The overexpression of PKC ϵ has been seen in a large number of NSCLC cancers compared to the

normal epithelium of lungs. The human NSCLC anchorage dependent growth also depends on PKC ϵ as its depletion impairs their growth and proliferation. PKC ϵ has also shown to protect lung cancer cells from apoptosis induced by tumor necrosis factor (TNF). The nude mice depleted of PKC ϵ using small hairpin RNA could not grow a xenograft. Another inhibitor of PKC ϵ , eV1-2 peptide, in NSCLC tumor producing nude mice inhibited the tumor growth. Targeted disruption of PKC ϵ inactivated Rac1 which reduced the motility of lung cancer cells resulting in the impairment of their metastasis. In addition, the depletion of PKC ϵ also demonstrated downregulation of many matrix metalloproteases (MMPs) which caused the prevention of metastasis and inability of NSCLC to colonize lungs (Garg et al., 2014). Evidence of implication of PKC ϵ in carcinogenesis was first found in small cell lung carcinoma in which its abnormally high levels were observed (Isakov, 2018).

Similarly, in prostate cancer the elevated levels of PKC ϵ have been frequently observed such as in early prostate adenocarcinomas (Jain & Basu, 2014). TRAMP mice (a model with spontaneous development of prostate cancer) have shown markedly increased levels of PKC ϵ than the control mice. PKC ϵ overexpression has demonstrated androgen independent advancement of prostate cancer in the LNCaP cells (androgen dependent in normal conditions) of both castrated and intact athymic nude mice. The PMA or TNF α induced apoptotic signals are down modulated in prostate cancer cells by PKC ϵ and the assembly of B1-integrins and actin filaments containing matrix adhesions is also promoted by PKC ϵ . The involvement of PKC ϵ in integrin signaling links it to Akt survival pathway in recurrent cells of prostate cancer (Garg et al., 2014).

Both intrinsic and extrinsic pathways of apoptosis were found to be negatively regulated by PKC ϵ . Protection from TRAIL induced apoptosis is conferred by PKC ϵ in several cancer types such as melanoma, erythroleukemia, acute myeloid leukemia (AML), glioma and breast cancer. In

MCF-7 breast cancer cells this was done by preventing the caspase 8 and 9 activation and cytochrome c release. In melanoma cells, PKC ϵ prevented apoptosis by negatively regulating activating transcription factor 2 (ATF2) which facilitates apoptosis by increasing the membrane permeability (Isakov, 2018).

In the invasive ductal breast cancer patients nearly 75% of primary tumors contain upregulated PKC ϵ . The inhibition of cell death by PKC ϵ in breast cancer is carried out by its prevention of Bax activation and translocation to the mitochondria. In the breast cancer cells of BLAB/c mice the overexpression of PKC ϵ resulted in tumor growth and spontaneous metastasis to lungs. Similarly, the increased PKC ϵ levels in glioma and renal cell carcinoma (RCC) also increased their metastatic abilities (Ali et al., 2009). The actin binding motif of PKC ϵ which enables the protein to interact with the cytoskeleton, has been implicated in the invasive process of cancer. Removal of this motif from the PKC ϵ enzyme conferred loss of metastatic and invasive potential to tumors (Garg et al., 2014). The causative role of PKC ϵ in breast cancer invasive, aggressive and motile phenotype was established when MDA-MB231 cells (a metastatic breast cancer cell line) showed marked decrease in their anchorage-independent colony formation, proliferation, motility and invasion after RNAi inhibition of PKC ϵ (Gorin & Pan, 2009).

With the expression of PKC ϵ , polarized morphology of cells and acquisition of invasive properties was observed. PKC ϵ overexpressing NIH3T3 cells when inoculated in mice, they led to the metastasis of liver and invasion of other nearby tissues by tumor (Gorin & Pan, 2009). The invasiveness driven by PKC ϵ in the breast and head and neck cancer models was partially due to small Rho GTPases such as RhoA and/or RhoC which when activated help in the migration of cells. In human glioma cells PKC ϵ was shown to mediate cell adhesion and motility via β -integrin chain interaction and extracellular signal-regulated kinase (Erk) phosphorylation. The

development of metastatic squamous cell carcinomas was also observed in the skin of PKC ϵ overexpressing transgenic mice. Besides that, the susceptibility to develop skin cancer after exposure to UV was increased in mice skin with high levels of PKC ϵ expression (Garg et al., 2014). The oncogenic ability of elevated PKC ϵ in the fibroblast, prostatic and colonic epithelial cell lines is mediated by Ras signaling through Raf-1 activation (Akita, 2002).

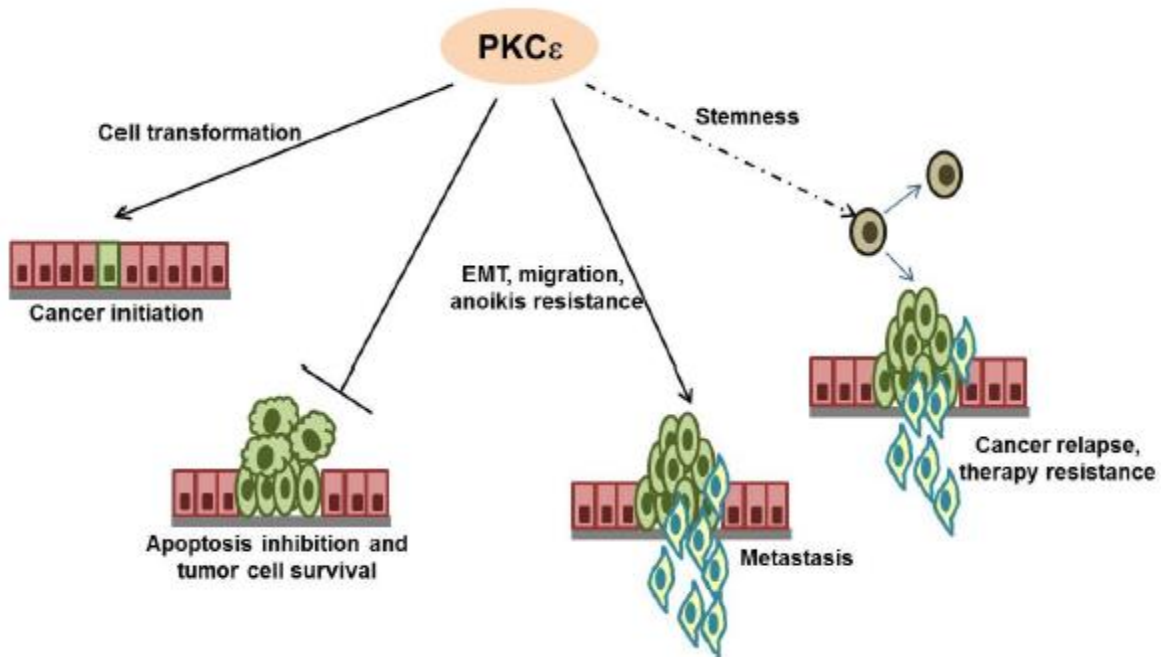


Figure 2. involvement of PKC ϵ in various carcinogenic processes

A. PKC ϵ ; a prognostic marker

PKC ϵ can be a potential prognostic biomarker for multiple types of cancers as its dysregulation has been associated with them. More than one studies have correlated the levels of with the grade and progression of cancer such as the low expression level of PKC ϵ was found in early prostate adenocarcinoma specimens. The abnormal expression of PKC ϵ was also found in other tissue samples like urogenital tract, which correlated directly with tumor stage (Gorin & Pan, 2009). Besides, much of the biomarker potential of PKC ϵ can be present

in the genetic alterations like SNPs yet to be associated with its dysregulation and mechanisms yet to be delineated that might prove to be essential for devising an effective anti-cancer strategy.

IV. Molecular Targets and interactions of PKC ϵ

A. PKC ϵ interaction with P53, a tumor suppressor

PKC ϵ prevents apoptosis by several methods, as mentioned previously, one of which is Akt survival pathway. Akt is involved in the regulation of P53 which indicates the effect of varying levels of PKC ϵ on P53 as well. Their relationship was observed when the PKC ϵ conferred resistance to TRAIL induced apoptosis in PKC ϵ overexpressing MCF-7 cells. The levels of P53 were markedly decreased along with the increase in human homolog of murine double minute 2 (Hdm2) and phospho-Hdm2, the negative regulators of P53, levels. On the other hand, P53 depletion lowered BH3 interacting-domain death agonist (BID) mRNA levels and PKC ϵ depletion raised the levels of P53 while AKT depletion increased both P53 and BID (Shankar, Sivaprasad, & Basu, 2008). This study establishes an inverse between the two proteins. However, another study performed to investigate the effects of individual PKC isozyme on P53 mediated apoptosis using yeast cells, concluded a positive relation of PKC ϵ and P53 which means PKC ϵ can aid in the process of apoptosis (Coutinho et al., 2011). This conclusion goes in opposition to all the previously established data regarding the pro-survival and anti-apoptotic properties of PKC ϵ . However, this discrepancy can be attributed to the complex regulation of PKC isozymes and P53 pathways in the higher organisms, that is, mammals as compared to yeast which is a much simpler organism. Besides, in the breast cancer cells and mammary epithelial cells activated glucocorticoid receptor (GR) mediated inhibition of apoptosis was shown to be PKC ϵ dependent. The increased expression of PKC ϵ at both

mRNA and protein level along with diminished activation of caspase-3 and -7 was recorded in these cells and the down modulation of PKC ϵ using siRNA restored the cells' potential to undergo programmed cell death (Aziz, Shen, & Maki, 2012).

B. PKC ϵ and a counteracting kinase

A prominent counteracting kinase of PKC ϵ from the same class, nPKCs, is PKCd. Upon activation the function of this enzyme has frequently been reported as opposite to those of PKC ϵ . While PKC ϵ is believed to be pro survival, PKCd facilitates the process of apoptosis (Basu & Sivaprasad, 2007). With multiple differences in their structure and pro-apoptotic properties of PKCd, both PKC ϵ and PKCd regulate vital physiologic processes such as cardioprotection, cell survival and proliferation (Duquesnes et al., 2011).

C. PKC ϵ interaction and function along-with other adaptor proteins

PKC ϵ overexpression has been observed with neoplastic development and its role has been demonstrated in various oncogenic processes such as transformation, proliferation, survival, endothelial to mesenchymal transformation (EMT) and cell to cell contacts. To carry out these processes it has to interact with many other adaptor proteins which link various components of complex biological pathways underlying these processes.

In rat colonic epithelial cells, the cellular transformation was carried out by overexpressed PKC ϵ in Raf-1/mitogen activated protein kinase (MAPK) dependent manner. In the prostate specific mouse models Akt and signal transducer and activator of transcription 3 (STAT3) phosphorylation was increased. UVR-induced phosphorylation of phosphoinositide 3-kinase (PI3K), Erk, and STAT3 was increased by PKC ϵ in the murine model skin cancer. This shows that the pathways linked to the adaptor proteins MAPK/ERK, PI3K/Akt and Stat3 are likely to be involved in the transformation process (Jain & Basu, 2014).

The relationship of PKC ϵ with survival protein Akt has been mentioned repeatedly before. In addition to that another pathway frequently implicated with cancer, found downstream of PKC ϵ in rat fibroblasts, is Nuclear factor κ B (NF- κ B). PKC ϵ also facilitated the assembly of TNF receptor 1 (TNFR1) signaling complex for activation of NF- κ B following the stimulation by TNF α in cells of prostate cancer. In glioblastoma cells the mechanism of activation of NF- κ B by PKC ϵ following epidermal growth factor (EGF) signal on EGF receptor (EGFR) has been discussed in detail. EGF stimulation activated Phospholipase C γ -1 which monoubiquitinated PKC ϵ preparing it for NF- κ B essential modulator (NEMO) docking followed by membrane recruitment of inhibitor of κ B kinase (IKK) complex leading to IKK β phosphorylation by PKC ϵ . The activation of NF- κ B resulted in pyruvate kinase M 2 (PKM2) induction promoted process of glycolysis thus helping in the formation of glioblastoma multiforme. Other targets regulated by PKC ϵ for promoting cell survival include ATF2 and Bcl-2 family proteins: X-linked inhibitor of apoptosis protein (XIAP), Bcl-X_l, BID, Bax and Bad (Jain & Basu, 2014).

The role of PKC ϵ in the metastatic process has been observed and associated with aggressive phenotype of several types of cancer notably; breast cancer, prostate cancer, NSCLC and RCC. Consequently, its role EMT has also been speculated and evidence collected through various studies. PKC ϵ has been observed to elevate the production of TGF β , and Snail, Slug and Twist, the positive regulator and transcription factors, respectively, of EMT. TNF α -converting enzyme (TACE) which is an MMP that aids in the shedding ectodomain of multiple secreted proteins, has also shown to be induced by PKC ϵ to produce growth factors and cytokines. The PKC ϵ induction of Akt2 to facilitate in EMT has also been speculated. Besides

that, EMT may also be promoted by PKC ϵ induced Rho GTPases, TGF β and STAT3 (Jain & Basu, 2014).

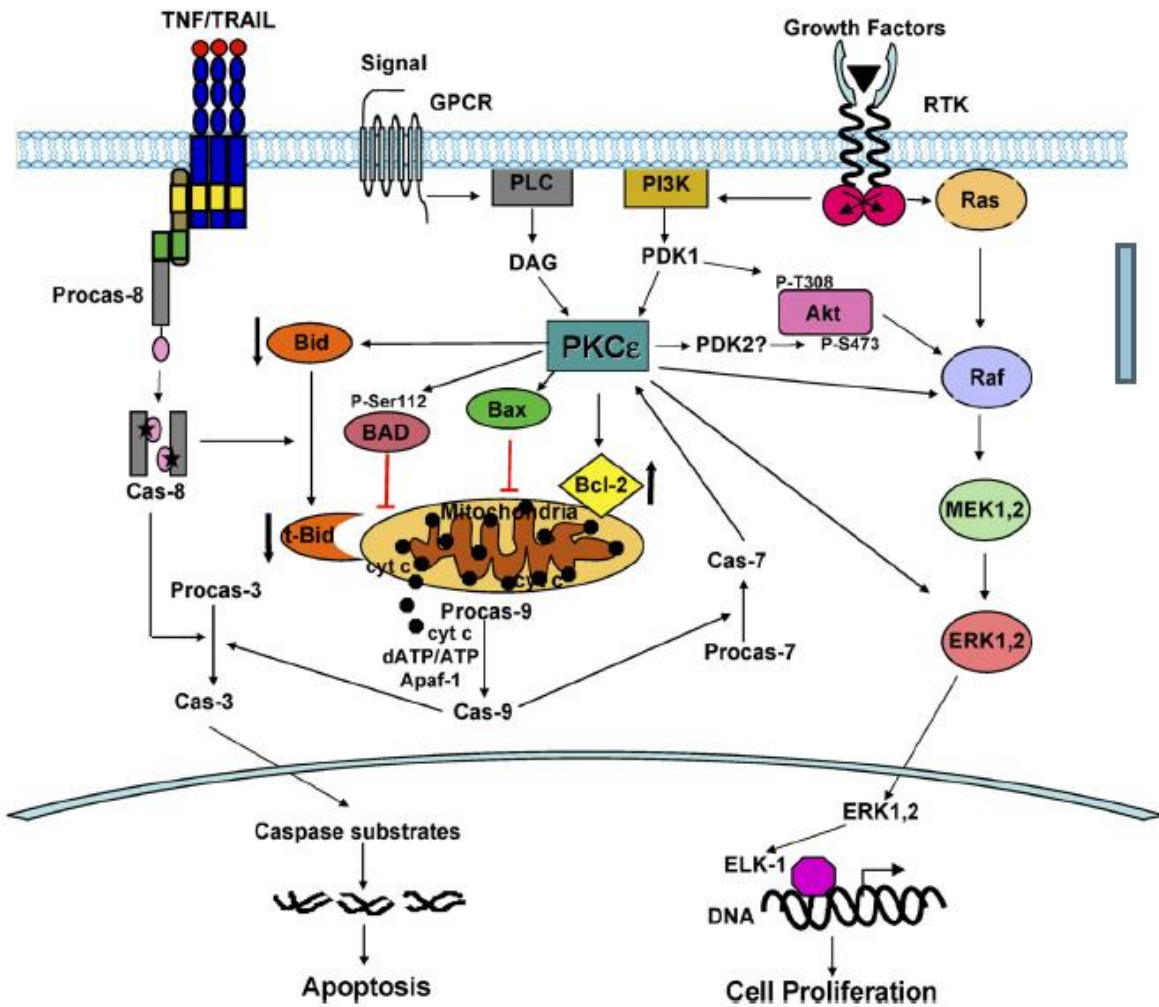


Figure 3. A general picture of PKC ϵ interaction with several adaptor proteins to aid in the process of cellular proliferation and promoting neoplastic development (Basu & Sivaprasad, 2007).

V. PKC ϵ Inhibitors

Due to the high homology in structures of PKC family isoforms, it has been really difficult to design an isozyme specific inhibitor. Only one PKC ϵ specific inhibitor, a peptide “PKC ϵ - V1-2”, can be found to date which is has been used extensively in PKC ϵ related researches (Kim et

al., 2003; Pravdic et al., 2009). Some non-specific inhibitor like GF-109203X, siRNAs against specific isozymes and short peptides derived from PKC such as C2 domain derivate peptide, have been used in different studies to inhibit and identify the roles of these isozymes in specific processes (Brandman, Disatnik, Churchill, & Mochly-Rosen, 2007; Jerdeva et al., 2005; Uchiyama et al., 2009). Furthermore, PKC ϵ phosphorothionate oligonucleotides have been used in addition to PKC binding protein, enigma homolog (ENH) which interacts with both N-type Ca²⁺ channels and PKC ϵ making a PKC ϵ -ENH-Ca²⁺ complex to study PKC ϵ specific roles (Maeno-Hikichi et al., 2003; Mehta, Radominska-Pandya, Kapoor, Dave, & Atkins, 2002). DAG-lactones have also been investigated for their PKC ϵ specific binding and inhibition (Ann et al., 2015).

CHAPTER 3

Material and Methods

In silico tools and methods were employed to predict the structure of normal PKC ϵ and to identify the potential damaging missense SNPs in the protein which can be further investigated and validated as biomarkers for hepatocellular carcinoma through wet lab procedures.

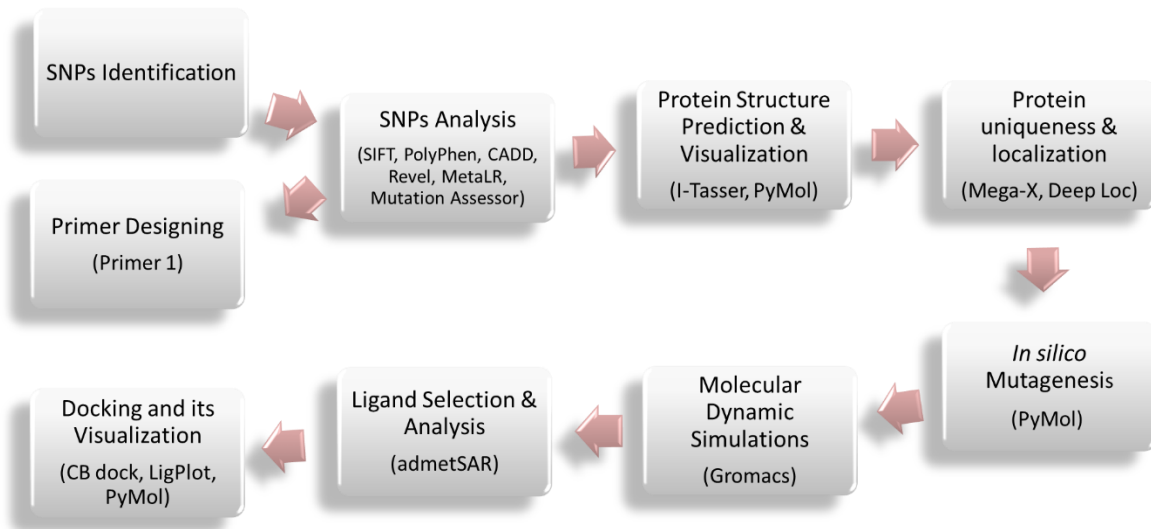


Figure 4. Flow chart of *in silico* methodology followed in this study.

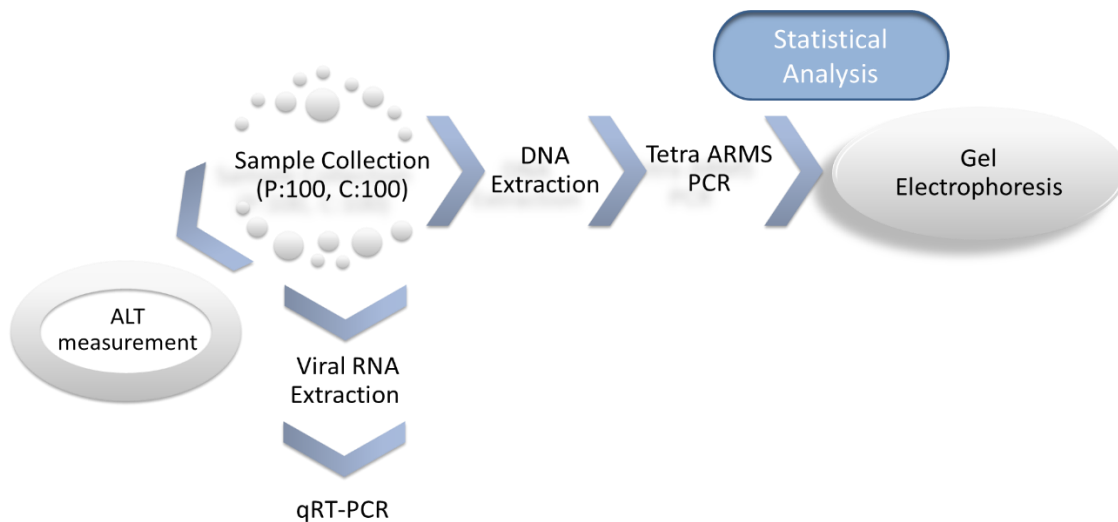


Figure 5. Flow chart of wet lab methodology followed in this study.

A. Protein Structure Prediction and Visualization

The protein sequence of PRKCE gene with transcript ID: PRKCE-201 ENST00000306156.8 was obtained from Ensembl database in FASTA format. This sequence was then submitted to I-TASSER (Iterative Threading ASSEMBly Refinement) which is an online tool for prediction of protein structures based on the threading approach of protein modelling and generates each predicted protein model with a confidence score ranging from -5 to 2 (Jianyi Yang & Zhang, 2015; Zhang, Freddolino, & Zhang, 2017). The predicted models were then visualized with the help of PyMOL molecular visualization system (Schrödinger, 2010). In addition, the predicted models by I-TASSER were cross-checked using InterPro database and other literature sources available, regarding the structural features of already studied and determined similar proteins (Blum et al., 2021).

B. Protein Uniqueness and Localization

The uniqueness of PKC ϵ in its family of proteins was found by constructing a phylogenetic tree of this protein using MEGA-X software. The amino acid sequences of all the proteins were of PKC family were retrieved in FASTA format and submitted to this software which constructed a distance based phylogenetic tree for all of these proteins.

For finding the subcellular localization, the amino acid sequence of PKC ϵ was submitted to DeepLoc, an online tool. The tool generated a tree with probability scores of localization of PKC ϵ into different compartments within the cell.

C. SNPs Identification

For the identification of potential damaging SNPs, data about all the missense SNPs of PKC ϵ was taken from the ENST00000306156.8 transcript id of PKC ϵ found in the

ensemble database. The data was then sorted and subjected to online tools which include SIFT, PolyPhen, CADD, REVEL, MetaLR, and Mutation Assessor. Each of these tools employs an algorithm which scores and categorizes the SNPs into different levels of harmfulness. SIFT scores from 0 – 1 with zero scoring SNP being the deleterious and one being the tolerated (Sim et al., 2012). PolyPhen scores also range from 0 – 1 but classifies SNPs having score 0.9 and above as probably damaging, 0.4 – 0.8 as possibly damaging and below 0.4 as benign SNPs (Adzhubei, Jordan, & Sunyaev, 2013). CADD uses the terms: likely benign and likely deleterious for scores from 0 – 30 and 31 – 35 respectively (Rentzsch, Witten, Cooper, Shendure, & Kircher, 2019), while REVEL categorizes them as likely disease causing with score 0 – 0.4 and likely benign with score 0.5 – 0.9 (Ioannidis et al., 2016). MetaLR also places SNPs in two groups: damaging (0.0 – 0.4) and tolerated (0.4 – 0.9) (Dong et al., 2015) but Mutation Assessor places them in four groups of high (0.9), medium (0.5 – 0.9), low (0.2 – 0.4), and neutral (0.0 – 0.1) SNPs (Reva, Antipin, & Sander, 2011). The data from these tools was analyzed and compared to see which potentially harmful SNPs were present among all of them.

D. SNPs Analysis

Online tools for analyzing the selected potentially harmful SNPs were also applied. These tools include I-Mutant, MutPred and HOPE. After entering the protein sequence and specifying the SNPs, I-Mutant gives its results in terms of DDG Value (DG (New Protein) – DG (Wild Type) in Kcal/mole) of the SNP and RI (Reliability Index) of the results. According to the tool's binary classification DDG values smaller than -0.5 indicates large decrease of stability while the value greater than 0.5 shows large increase of stability in the protein structure (Capriotti, Fariselli, & Casadio, 2005).

MutPred on the other hand generates the output as a general probability score which if greater than 0.5, indicates pathogenicity of the protein, in addition to the probability scores and P-values for the gain or loss of certain structural and functional properties of the proteins. The P-value of 0.05 or smaller points to the significance of predicted change, rejecting the null hypothesis (Pejaver et al., 2020).

HOPE analyses and evaluates the SNPs with reference to the information available about the protein on other databases such as UniProt and generates a detailed report enlisting all the possible effects the respective SNP might have on the proteins structure and conserved properties (Venselaar, Te Beek, Kuipers, Hekkelman, & Vriend, 2010).

E. *In silico* Mutagenesis

The selected SNP was induced in the predicted wild structure of PKC ϵ . The mutation of the selected SNP was induced in the best predicted wild type structure of PKC ϵ , using PyMol. Amino acid sequence of the wild PKC ϵ structure was made visible by using the selection tools in lower right corner of PyMOL window. Then the residue to be mutated was selected and colored differently through the option given in object control panel. After that, from the Wizard option in the external graphical user interface, ‘Mutagenesis of protein’ was selected, and the selected residue was replaced with the mutated residue by selecting ‘no mutation’ options that appears on the object control panel, and the change was applied. Finally, the session was saved to be used in further *in silico* processes.

F. Molecular Dynamic Simulations

Molecular dynamic simulations of the wild protein structure (fig7.) and mutagenesis induced structure were carried out using an open source GROMACS software package on supercomputer. The interface used for giving the commands and running simulations is

PuTTY while WinSCP was used for the transfer of files between the supercomputer and personal computer using ssh file transfer protocol (SFTP).

G. Docking and its Visualization

The molecule selected for docking is Betulin which is a pentacyclic triterpenoid. This molecule has a lot of exploratory potential as a therapeutic agent against HCC as it has previously shown promising antineoplastic activity. Furthermore, its properties have been explored using an online tool, admetSAR (Cheng et al., 2012). For that canonical smile for the molecule Betulin was taken from PubChem and submitted to admetSAR which generated a list of predictions of different parameters of the candidate drug.

The best protein structure obtained from I-TASSER was docked with the Betulin (CID: 72326) which has proven anti-inflammatory and antineoplastic properties. The structure of Betulin was obtained from PubChem and its docking with the protein was carried out using an online tool named CB Dock (Liu et al., 2020). Docked structure obtained from CB Dock was first prepared on PyMOL for visualization and then opened in LigPlus to see detailed molecular interactions between the ligand and protein (Schrödinger, 2010; Wallace, Laskowski, & Thornton, 1995).

H. Primer Designing

Primer 1is an online tool which was used to design primers for tetra ARMS PCR of the screened potentially damaging SNP (Collins & Ke, 2012).

I. Sample Collection

For sample collection first approval of the institution review board was acquired. Then written and oral consent was taken from the patients whose samples were taken. Hundred samples from the patients of HCV induced HCC and equal number of control samples were

collected for carrying out the wet lab validation procedures and conforming our hypothesis about the presence of the missense SNPs in PRKCE gene of those patients.

J. DNA Extraction

Manual procedure for DNA extraction was employed. Extraction of DNA was done from nearly two hundred samples half of which were control samples and the rest were samples from of hepatitis C virus (HCV) induced HCC patients. For this, buffer solutions were prepared first which were named solution A, solution B, solution C and solution D. For the preparation of 500ml of solution A; 54.72g of sucrose (0.32mM), 0.238g of MgCl₂ (5mM) and 6.057g of Tris-base (10 mM) were added in nearly 450ml of distilled water. The pH of solution was then set to 7.5 – 8 using HCL (conc.) and NaOH (40% w/v solution). The volume was then brought up to a little less than 500ml using distilled water and the solution was autoclaved. 5ml of Triton X-100 (pure) was added after the solution was autoclaved because it is sensitive to heat.

Similarly, 500ml of solution B was also prepared. 11.685g of NaCl (400mM), 0.29g of EDTA (2mM) and 6.057g of Tris-base (10 mM) were added in nearly 450ml of distilled water. pH of the solution was again set to 7.5 – 8 in the same manner and the volume was brought up to 500ml before it was autoclaved. Solution C consisted of phenol only which is at risk of oxidation when exposed to light therefore, it was wrapped properly in the aluminum foil and carefully kept away from the light. 50ml of solution D was prepared with chloroform (24ml) and isoamyl alcohol (1ml).

DNA extraction is a two-day procedure. It was started by taking 750µl of blood and equal volume of solution A in an Eppendorf tube. Eppendorf was inverted a few times to mix the blood with the solution A and then kept at the room temperature for 10 minutes. Then, it

was centrifuged at 13000rpm for 1 minute and the supernatant was discarded. This starting step was carried out once again to get a clean pellet. After that, 400µl of solution B was added in the Eppendorf tube and mixed with the pipette/tapped/vortexed to dissolve the pellet in the solution. Then, it was centrifuged again at 13000rpm for 1 minute and supernatant was discarded followed by the addition of 400µl of solution B, 12µl of 20% SDS solution and 5µl of Proteinase K. The sample tubes were then kept at 37degree Celsius to incubate overnight.

Extraction day 2 was started by adding 250µl of solution C and 250µl of solution D in the sample tubes that were incubated overnight. The tubes were then centrifuged for 10 minutes at 13000rpm. The upper aqueous layer of the two layers formed after centrifugation was carefully separated into another Eppendorf, to which 55µl of sodium acetate and chilled iso-propanol were added. After addition of iso-propanol, the tube was inverted several times to precipitate the DNA. It was again centrifuged for 10 minutes at 13000 rpm and the supernatant was discarded. To resuspend the DNA pellet, 200µl of chilled ethanol was added and again centrifuged at 13000 rpm for 8 minutes. The supernatant was discarded, and the tubes were inverted in the laminar flow until ethanol was completely dried.

K. ARMS PCR

The type of PCR which is used to identify presence of point mutations or SNPs in the genomic DNA is Amplification Refractory Mutation System Polymerase Chain Reaction (ARMS PCR). For this kind of PCR specific tetra primers, two internal primers and two external primers, against the selected SNP which was selected after going through all the above mentioned *in silico* processes, were used. For each sample 25µl of reaction mixture

was prepared. This mixture consisted of 2 μ l of PCR buffer, 4.5 μ l of MgCl₂, 2 μ l dNTPs, 0.4 μ l of Taq Polymerase, 1 μ l of each of the four primers, 1.5 μ l of the DNA sample and 10.6 μ l of PCR water. After preparation, the reaction mixture was vortexed/centrifuged for a few seconds to mix it well and remove any bubble formed. PCR tubes containing reaction mixture were then placed in the Thermocycler. During the optimization process of the primers, multiple reactions were carried out together using the Gradient Thermocycler which was already programmed with a range of temperatures. Each PCR was composed of 3 steps and 35 cycles. Thermocycling conditions and steps of the method stored in the machine are mentioned below.

Step 1: In stage1 the DNA template denaturation was allowed to occur at 95 degrees Celsius for 5 minutes. It was done to break the double stranded DNA into single strand. In stage 2 the denaturation of the daughter DNA took place. This stage had 35 cycles and each cycle took 30 seconds.

Step 2: The annealing temperature of 50 °C was provided and 30 seconds were allocated for this step too.

Step 3: Annealing of the primers was followed by their extension for 30 seconds at 72 degrees Celsius. The final extension was allowed to occur at 72 degrees Celsius for 7 minutes.

After completion of PCR, the tubes were stored in the freezer until running the PCR product on gel electrophoresis.

L. Gel Electrophoresis

For visualization of genomic DNA 1% (w/v) Agarose gel was used. It was prepared by adding 1 gram agarose in 100ml of 1X TBE or TAE buffer. This mixture was then

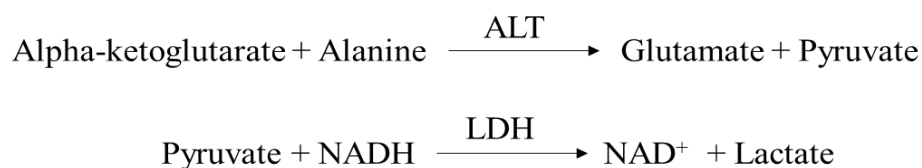
microwaved until the agarose completely dissolved in the buffer. 3 – 5 μ l of Ethridium Bromide was added in the mixture after cooling it down a little, for staining the DNA. Mixture was then poured in the gel casting tray was kept at room temperature for solidification. After the gel was set, it was placed in buffer tank of electrophoresis apparatus. The 5 – 7 μ l of DNA sample combined with 1 – 2 μ l of loading dye and were loaded in each well. Electrophoresis was performed at 120 V and 300 A for 20 – 30 minutes in 1X TBE or TAE buffer. DNA bands were analyzed by placing the gel in UV transilluminator or gel doc.

For visualization of PCR product 2% (w/v) agarose gel was prepared. The preparation procedure was same as mentioned above except this time 2g of Agarose was dissolved in 100ml of buffer. When the gel was prepared, 10 – 12 μ l of the PCR product from individual tube was combined with 1 – 2 μ l of loading dye and then loaded in a well. Electrophoresis was performed at 80 V for 20 minutes in 1X TBE or TAE buffer. The gel was again analyzed by placing the gel in UV transilluminator or gel doc.

M. Measurement of ALT levels

ALT levels of the control and disease samples both were measured to get an idea about the change in liver function of the diseased individuals. Kit method was used for this purpose. The kit contained two solutions named R1 and R2. R1 contained Tris-buffer and Alanine amino acid while R2 contained Alpha ketoglutarate as a substrate, Lactate dehydrogenase enzyme and NADH cofactor. Blood samples were first centrifuged at 3000rpm for about 2 minutes to separate the serum from them. The working reagent was prepared by adding R1 and R2 in 4:1. For 1 sample 400 μ l of R1 and 100 μ l of R2 were

taken and 50µl of sample was added in it. The reactions which take place after addition of samples are as follows:



Lactate is the colorless end product, and its absorbance was measured at 340nm using a (UV-visible) spectrophotometer (Microlab 300 – Semi-automated chemistry analyzer).

The probe of spectrophotometer dipped in the reaction solution takes up nearly 400µl of it and checks its absorbance/optical density and gives the concentration of ALT present.

The formula for concentration is:

$$\text{concentration} = \Delta\text{Abs}/\text{min} * \text{kit factor}$$

Kit factor is 1746 for any absorbance to be checked at room temperature using 340nm wavelength. Normal ALT values lie below 40-41 U/l.

N. Viral RNA Extraction

As the samples used were of HCV induced HCC patients, therefore the viral load of these HCV induced HCC blood samples was also measured to compare for comparison and association of SNPs with the severity of disease. For that, RNA of hepatitis C virus was extracted from the blood serum of the patients using kit method. For viral RNA extraction from one serum sample, 200µl of serum was taken in an eppendorf and 500µl of VNE buffer was added. This mixture was vortexed for 5-7 seconds and 500µl of 75% ethanol was added and again vortexed for 5-7 seconds. It was then poured in spin column and centrifuged at 8000rpm for 1minute. The spin column was then taken out and the collection tube was discarded. RNA was retained on the filter of spin column which was then placed

in a new collection tube and 500µl of wash buffer 1 was added into the spin column. The spin column was again centrifuged at 8000rpm for 1 minute and the collection tube was discarded. Second washing was done with 750µl of wash buffer 2 two times. In second washing, both times spin column was centrifuged at 14000rpm for 1 minute and collection tube was discarded. After washing the filter with trapped RNA was placed in a new collection tube and centrifuged for 1 minute at 14000rpm for complete drying of wash buffers from the filter and the collection tube discarded. Lastly, the filter was placed in a new Eppendorf and 50µl of RNase free water was added in it. The eppendorf with filter was centrifuged at 8000rpm for 1 minute leading to the dissolution of RNA in the RNase free water. The dissolved RNA was collected in the eppendorf below. The filter was discarded and eppendorf with RNA was saved at -20 °C for qRT-PCR.

O. Quantitative Reverse Transcriptase Polymerase Chain Reaction (qRT-PCR)

This kind of PCR is used for the quantification of RNA. The RNA in sample is first transcribed by reverse transcriptase into cDNA which is then amplified and quantified at the same time through real time PCR. For real time quantification of the reaction SYBR® Green fluorescent dye is used which does not bind to ssDNA and only binds in the minor groove of dsDNA. When excited by the light source, the dye produces signal which is almost a hundred times stronger when it bound to the DNA molecule than in the unbound state. The change in fluorescence over the time is recorded against the cycle number and an amplification plot is generated which shows the progress of PCR. 44µl of reaction mixture which contained 43µl master mix and 1.2µl RT enzyme from Favorgen kit, was used for 6µl of each sample. At least two standards with known concentrations were used. The standards are used to determine the upper and lower limit of the concentrations

obtained from the samples. The sample concentrations are compared with the standard curve and a linear graph is plotted which gives the starting quantity/copy number of the template molecule on x-axis against cycle threshold (CT) on y-axis. The sample which smaller CT value has more viral load than the one with larger CT value. Viral load is generated at the end of the reaction.

CHAPTER 4

Results

1. Predicted PKC ϵ models

I-TASSER generated five best protein models of the given sequence with the confidence (C) scores: -1.76, -1.53, -2.26, -2.44, and -2.30. In general, the C-score values of different protein models represent the quality of their structures; the greater the C-score, the better the model. Similarly, these models are also given ranks according to their quality which usually coincides with the values of C-score. However, in this case the best predicted models according to C-score and ranking did not meet the criteria of structural requirements of the proteins belonging to the same family, according to the literature review. Structural analysis of the predicted protein models was carried out using InterPro database which contains all the information about structural features of known proteins. The stretches of amino acid residues which constitute specific domains inside the PKC ϵ protein were identified and marked with different colors, in addition to the identification of the presence of the essential structural elements in the predicted models. According to the literature review and all the information obtained from InterPro (Blum et al., 2021). The predicted model 5 (fig.7) of PKC ϵ with C-score of -2.30 appears to be the best model for the protein. Model 2 (fig.6) which has the highest C-score (-1.53) among all the predicted protein models of PKC ϵ does not comply with the structural requirements of the protein. A large portion of this protein model does not form any secondary structure and can be seen in the form of abstract loop structure in most of the protein domains.

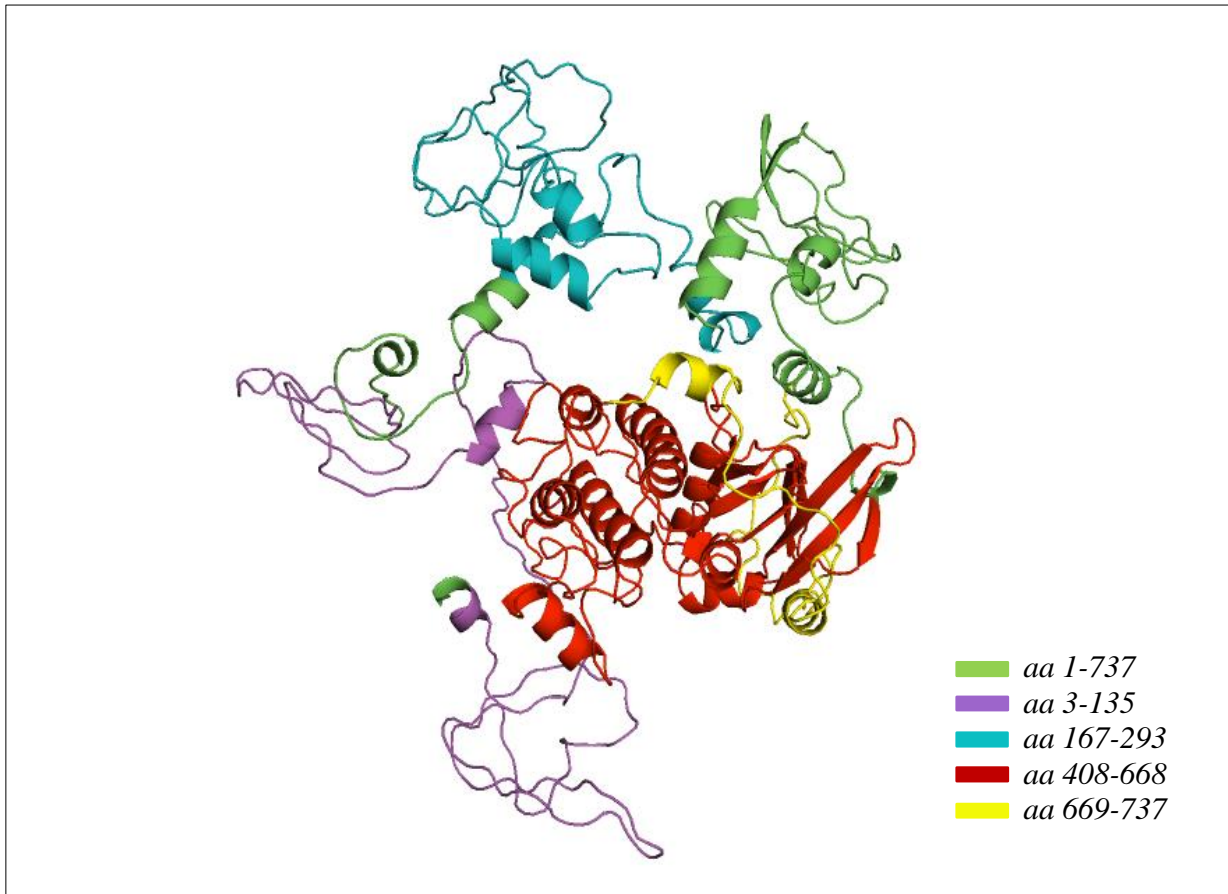


Figure 6. Predicted model 2 of PKC epsilon. The PKC epsilon structure predicted by I-TASSER with the highest C-score of -1.53. The whole protein structure is colored in green while different domains are shown in varying colors. Magenta (3-135) represents C2 domain in PKC epsilon, cyan (167-293) shows DAG/PE binding (C1) domain, the protein kinase domain (408-668) is highlighted in red while yellow color illustrates the presence of AGC kinase, superfamily, C-terminal (669-737).

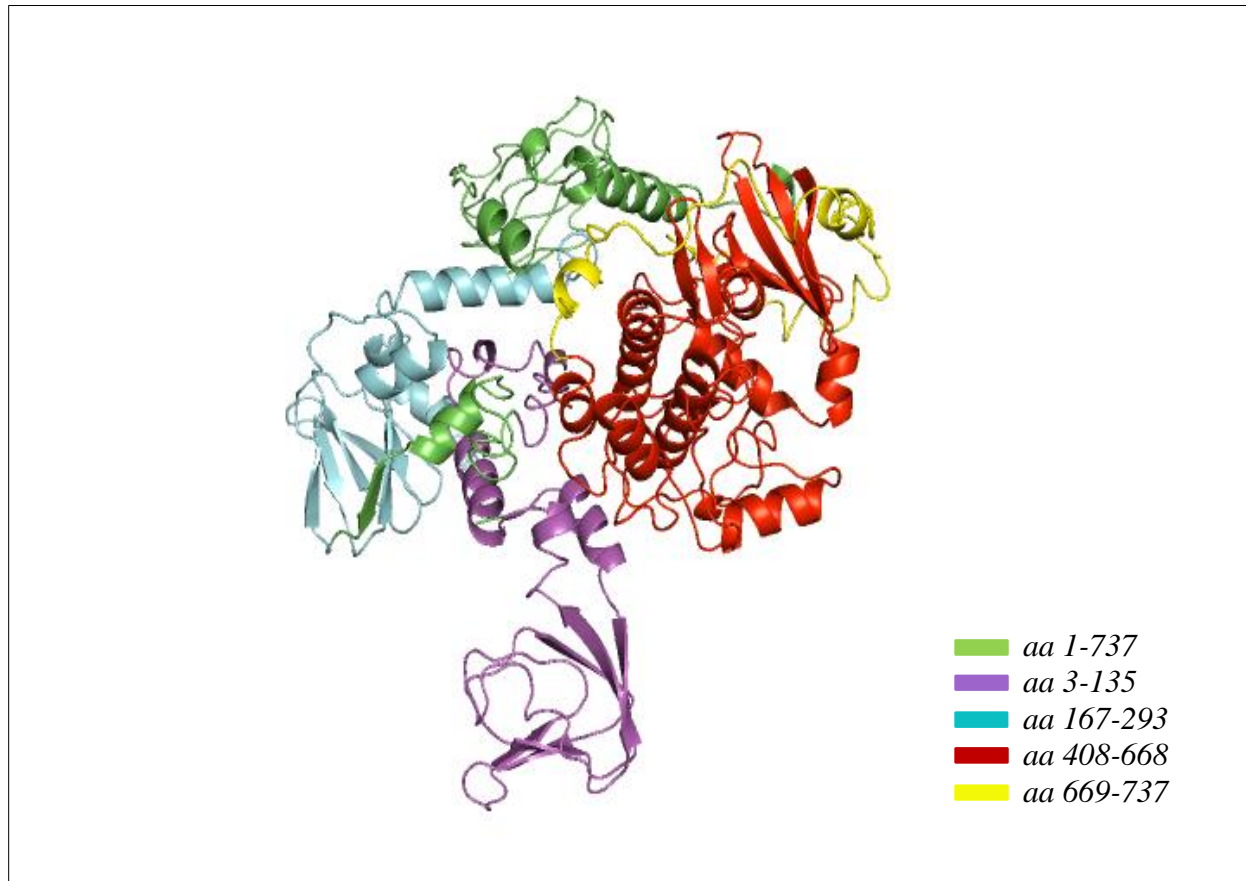


Figure 7. Predicted model 5 of PKC epsilon. The PKC epsilon structure predicted by I-TASSER with the C-score of -2.30. The whole protein structure is colored in green while different domains are shown in varying colors. Magenta (3-135) represents C2 domain in PKC epsilon, cyan (167-293) shows DAG/PE binding (C1) domain, the protein kinase domain (408-668) is highlighted in red while yellow color illustrates the presence of AGC kinase, superfamily, C-terminal (669-737).

After the induction of *in silico* mutagenesis the wild type structure (fig.7), both the mutagenesis induced structure and the wild protein structure were superimposed for comparison. The mutation is located inside the triangle while its effects can be seen with the change of color in other parts i.e., C2, C1 and hinge region of the protein.

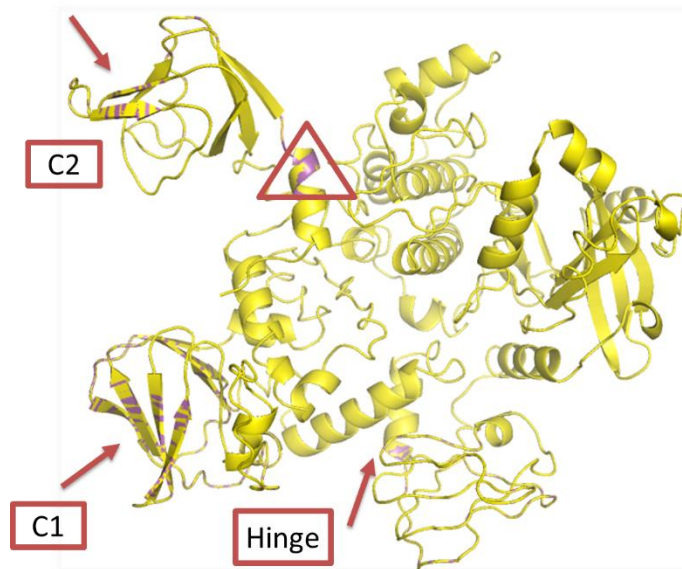


Figure 8. Superimposed structures of mutated and wild type PKCε.

2. Phylogenetic Tree and Subcellular Localization

The evolutionary relationship of the PKCs is depicted through the phylogenetic tree given in figure 9. It shows that all the members of this family of proteins originated from a common ancestor protein. The branches originate from common branch points which join the closely linked members of the protein family. The scores on each branch represents the substitution per site, showing the extent of evolution of each protein from its other family members.

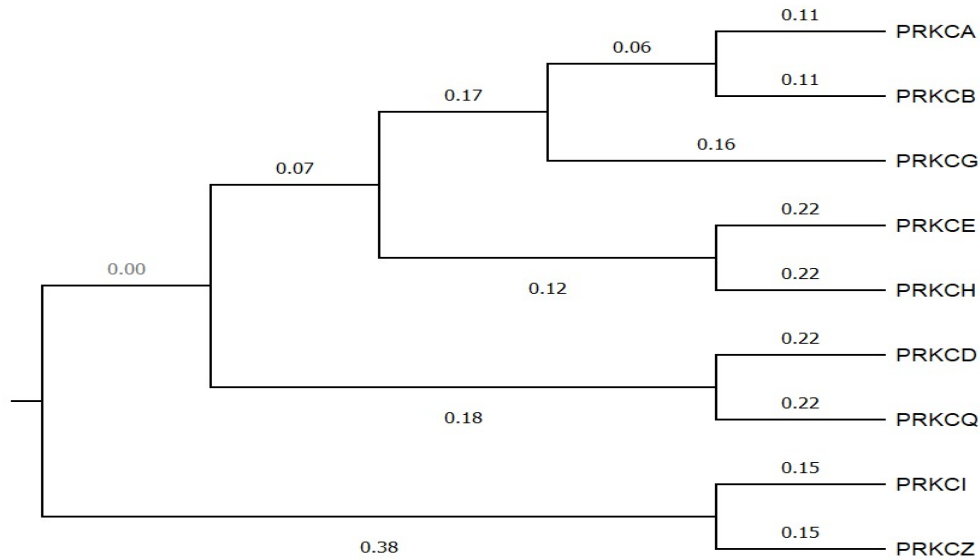


Figure 9. Phylogenetic tree of PKC protein family.

The route of PKC ϵ localization is given in figure 10. The red line represents the path taken by the protein to localize into its compartment inside the cell. The score shows the probability/likelihood of the event. So, it is suggested based on the score that PKC ϵ localizes in the cytoplasm.

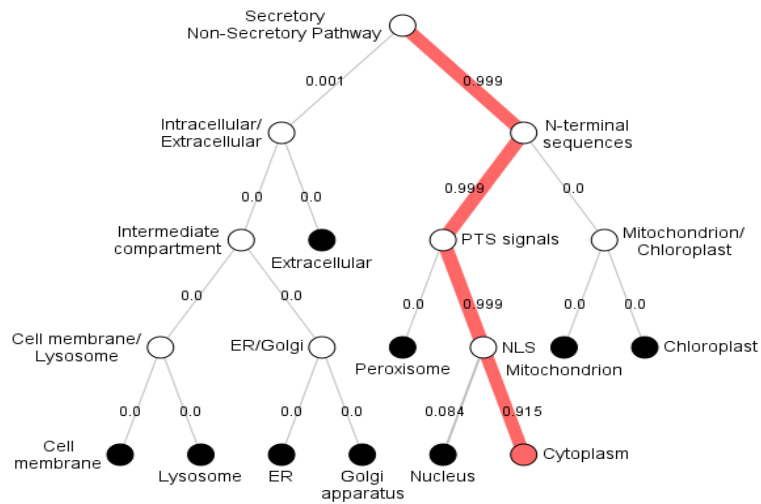


Figure 10. Subcellular localization of PKC ϵ .

3. Identified missense SNPs of PKC ϵ

The data from obtained from ensemble had all the information about the missense SNPs in PKC ϵ including the variant IDs, chromosomal location, allele changes and the resulting changes in respective amino acids encoded by them. A number of different *in silico* tools, SIFT, PolyPhen, CADD, MetaLR, Mutation Assessor and REVEL, were used to filter this data and to identify the SNPs with the potential to be most injurious structural and functional integrity of the protein. All the tools used generated results in accordance with their individual scoring systems which render the SNPs to different levels of damaging effects. The results from these tools are represented in the form of scatter plots (chart a,b,c,d,e,f) below which show the scores along with the classification of SNPs by each of the tools.

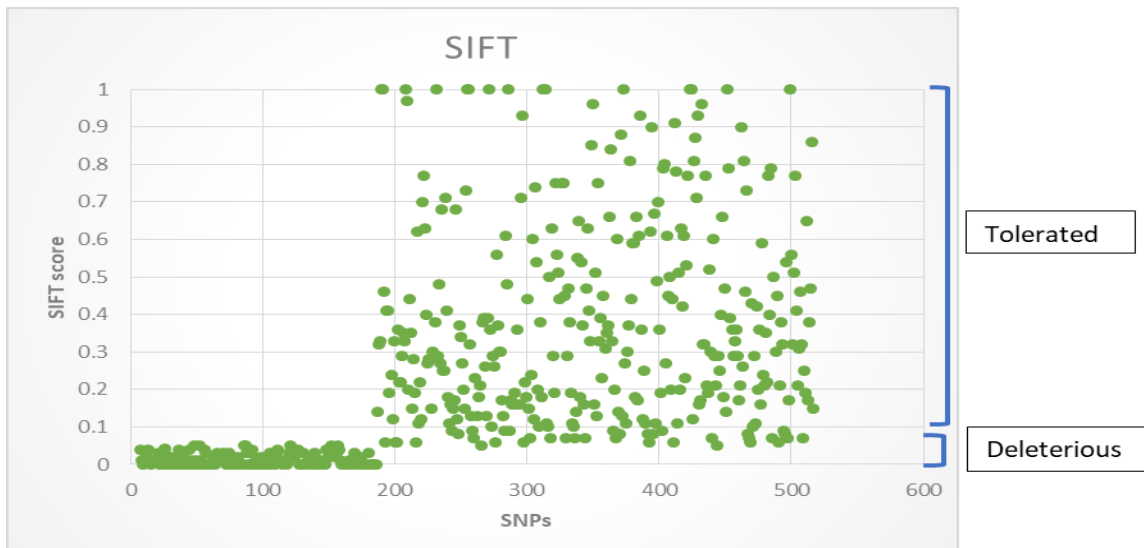


Chart a. Scatter plot of the tolerated and deleterious SNPs as classified by SIFT according to their score.

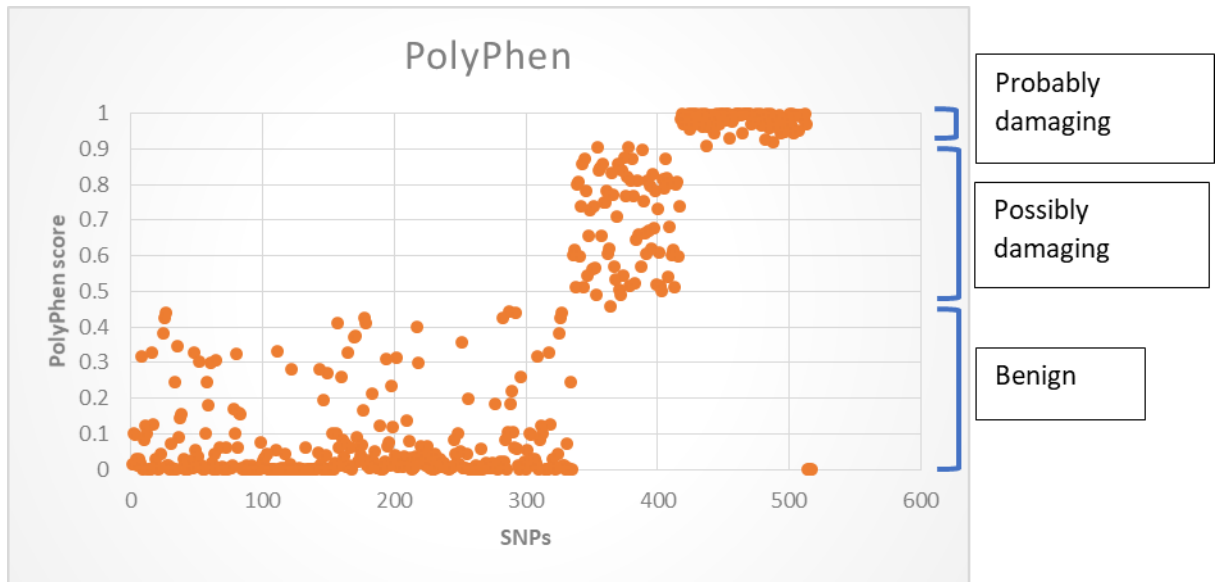


Chart b. Scatter plot of the probably damaging, possibly damaging and benign SNPs as classified by PolyPhen according to their scores.

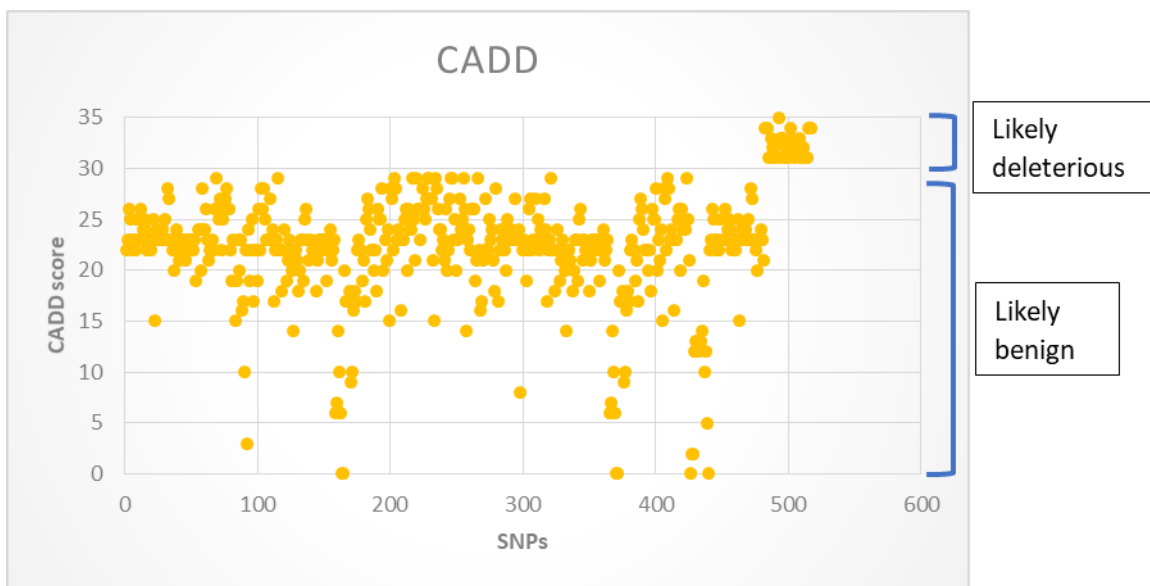


Chart c. Scatter plot of likely deleterious and likely benign SNPs as classified by CADD according to their scores.

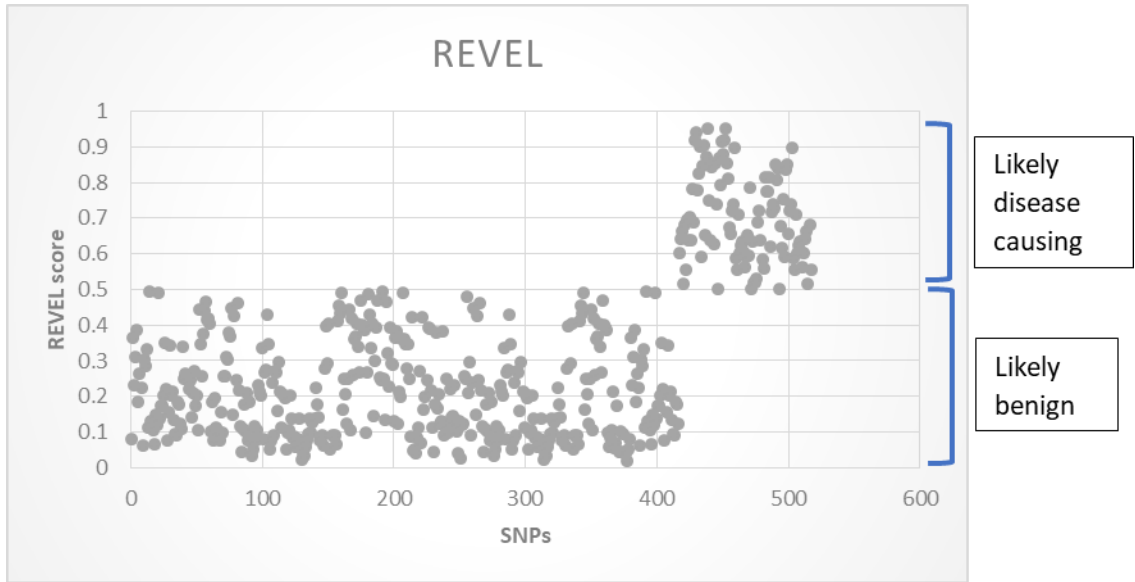


Chart d. Scatter plot of the likely disease causing, and likely benign SNPs as classified by REVEL according to their scores.

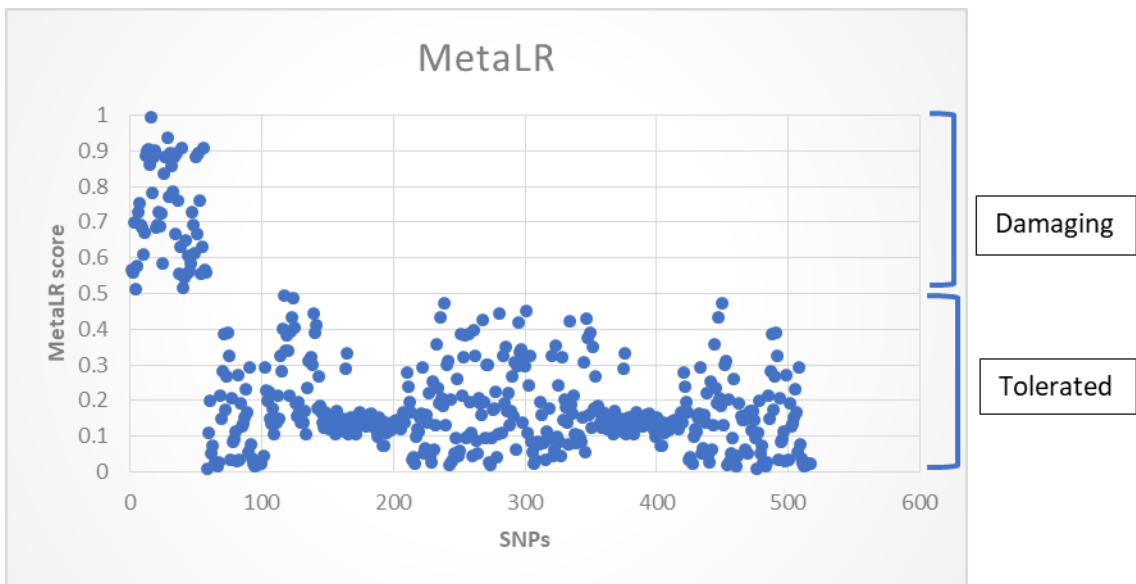


Chart e. Scatter plot of the damaging and tolerated SNPs as classified by MetaLR according to their scores.

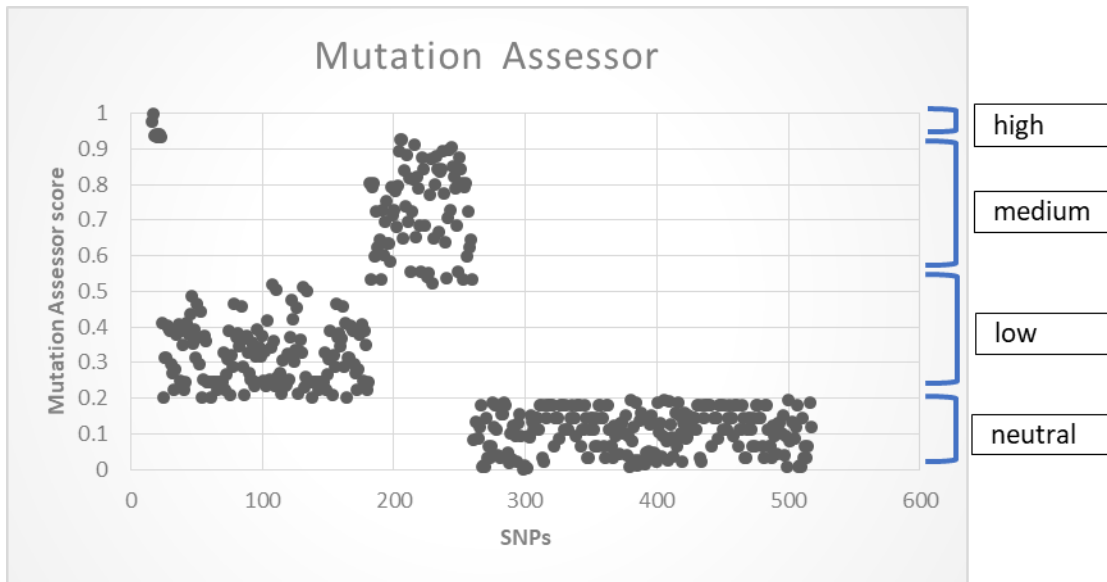


Chart f. Scatter plot of the high, medium, low and neutral SNPs as classified.

In addition to the scatter plots, bar graphs were also generated to show the relative amount of SNPs present in each of the different classes represented in these tools.

4. Data filtration of PKC ϵ SNPs

There are numerous types of variants in PKC ϵ gene. The different types of variants of PKC ϵ are enlisted below in table 3.

Table 1. Types of PKC ϵ variants and their numbers.

Variant type	Number of variants
Missense variants	357
Coding sequence variants	308
Synonymous variants	286
Splice region variants	91
Others	19

All the data about missense SNPs of PKC ϵ was filtered through the above-mentioned tools. In the first phase of filtration the threshold of scores of these tools was kept

at: SIFT; deleterious SNPs (score: 0-0.05), PolyPhen; probably damaging and possibly damaging SNPs (score: 0.4-1), Revel; likely disease-causing SNPs (score: 0.5-0.9), MetaLR; damaging SNPs (0.5-0.9), and Mutation Assessor; medium and high scoring SNPs (score: 0.5-0.9). The resulting SNPs are enlisted below along with their scores from each of these tools.

Table 2. Potentially harmful SNPs filtered through SIFT, PolyPhen, REVEL, MetaLR and Mutation Assessor.

Sr.	Variant ID	SIFT	PolyPhen	REVEL	MetaLR	Mutation Assessor
1.	rs1324772831	0.05	0.511	0.643	0.567	0.598
2.	rs765170809	0.02	0.856	0.697	0.698	0.755
3.	rs1470650959	0	0.988	0.783	0.728	0.792
4.	rs747860089	0	0.97	0.688	0.755	0.715
5.	rs950818717	0	0.873	0.918	0.691	0.729
6.	rs983848526	0	0.781	0.94	0.687	0.783
7.	rs1225886287	0	0.728	0.902	0.886	0.893
8.	rs966862357	0	0.999	0.846	0.902	0.925
9.	rs1472857456	0	0.956	0.904	0.906	0.928
10.	rs1276454009	0	0.999	0.872	0.863	0.649
11.	rs1345446290	0	0.977	0.953	0.996	0.975
12.	rs957679933	0	0.998	0.751	0.782	0.839
13.	rs570929334	0	0.971	0.639	0.882	0.738
14.	rs1338273606	0	0.999	0.843	0.903	0.882
15.	rs1320927580	0.01	0.739	0.739	0.729	0.694
16.	rs1433761543	0.01	0.566	0.87	0.689	0.819
17.	rs1388726033	0.02	0.492	0.793	0.724	0.555
18.	rs760292238	0.01	0.903	0.917	0.837	0.725
19.	rs1329893766	0	0.993	0.879	0.885	0.813

20.	rs1210254842	0	0.996	0.919	0.938	0.997
21.	rs924773684	0	0.995	0.953	0.772	0.936
22.	rs759149125	0	0.998	0.853	0.893	0.911
23.	rs1184220025	0.01	0.961	0.812	0.858	0.652
24.	rs904855630	0.01	0.846	0.656	0.883	0.821
25.	rs950073404	0.02	1	0.899	0.91	0.939
26.	rs1304555744	0	0.985	0.501	0.517	0.54
27.	rs748562237	0	0.998	0.813	0.546	0.871
28.	rs1446404436	0	0.975	0.851	0.649	0.879
29.	rs1446404436	0.02	0.812	0.807	0.555	0.844
30.	rs771680739	0	1	0.84	0.606	0.668
31.	rs891028429	0.01	0.981	0.677	0.564	0.892
32.	rs1294256395	0	0.999	0.618	0.584	0.896
33.	rs1399624263	0.01	0.798	0.753	0.728	0.729
34.	rs745492152	0.03	0.621	0.836	0.694	0.905
35.	rs769499415	0	0.985	0.852	0.615	0.85
36.	rs1324772831	0.01	0.511	0.643	0.567	0.598
37.	rs775012697	0.05	0.969	0.683	0.558	0.646
38.	rs1553369874	0	0.985	0.366	0.108	0.803
39.	rs1345511001	0	0.998	0.604	0.213	0.803

In the second phase of data filtration, another tool, CADD was also applied on the above given SNPs. The CADD score threshold was kept at 31. Following 10 SNPs came out to be the most potentially harmful ones after application of the CADD scores.

Table 3. Screened SNPs through CADD and their classification in all the applied tools.

Sr.	Variant ID	SIFT	PolyPhen	REVEL	MetaLR	Mutation Assessor	CADD
-----	------------	------	----------	-------	--------	-------------------	------

1.	rs950818717	deleterious	possibly damaging	likely disease- causing	damaging	medium	likely deleterious
2.	rs1345446290	deleterious	probably damaging	likely disease- causing	damaging	high	likely deleterious
3.	rs1320927580	deleterious	possibly damaging	likely disease- causing	damaging	medium	likely deleterious
4.	rs924773684	deleterious	probably damaging	likely disease- causing	damaging	high	likely deleterious
5.	rs1184220025	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious
6.	rs950073404	deleterious	probably damaging	likely disease- causing	damaging	high	likely deleterious
7.	rs748562237	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious
8.	rs891028429	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious
9.	rs1399624263	deleterious	possibly damaging	likely disease- causing	damaging	medium	likely deleterious
10.	rs769499415	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious

11.	rs1553369874	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious
12.	rs1345511001	deleterious	probably damaging	likely disease- causing	damaging	medium	likely deleterious

After all the screening first SNP with variant ID rs1553369874 was selected for further *in silico* analysis and wet lab validation process.

5. Structural and Functional Analysis

Structural and functional analysis of the SNP rs1553369874 was done using online tools: I-Mutant, MutPred and HOPE. I-Mutant generated DDG value of the SNP which showed its effect on the stability of overall protein structure. The RI of the given SNP information is 8 and its DDG value is -0.49 Kcal/mol which shows decreased stability of the protein structure with the induction of this mutation.

MutPred gave 0.77 general probability score for this SNP. According to the analysis results Mutpred generated, the induction of this mutation can cause gain of MoRF (molecular recognition features) binding ($P = 0.0042$), gain of methylation at E14 ($P = 0.0173$), and loss of ubiquitination at K11 ($P = 0.0432$).

In addition to the analysis of these tools, HOPE report showed that the mutated residue is much bigger than the wild-type residue as can be seen in fig.11. which might lead to bumps in the protein structure. The wild-type residue charge was also negative, while the mutant residue charge is positive. This can cause repulsion with other residues in the protein or ligands. Most importantly, the mutation is located

in the C2 domain of the protein near a highly conserved position which can be potentially damaging for the protein.

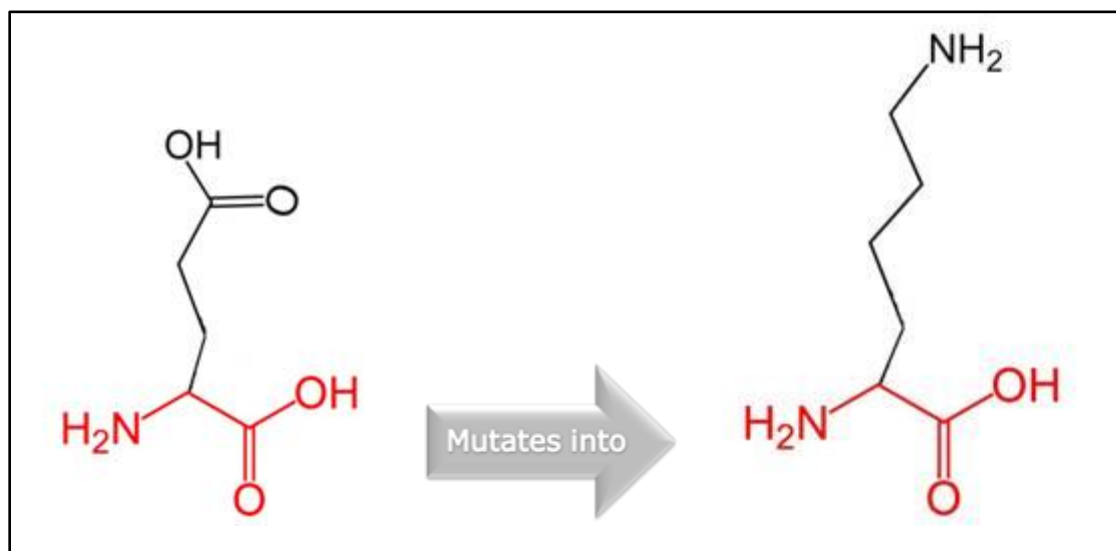


Figure 11. Change in the amino acid residue from glutamic acid to lysine as a result of the mutation rs1553369874 in PKC ϵ .

6. Pharmacophore analysis

The ligand selected for docking is Betulin (CID: 72326) and its structure obtained from PubChem is given in the fig.12. The properties of Betulin obtained from admetSAR are given in the tables 4 and 5. AdmetSAR results show that this molecule fulfills the criteria

of a good drug candidate according to of Lipinski's rule of 5. In addition, its other properties gave significant insight about the molecule's properties and its behavior in human body.

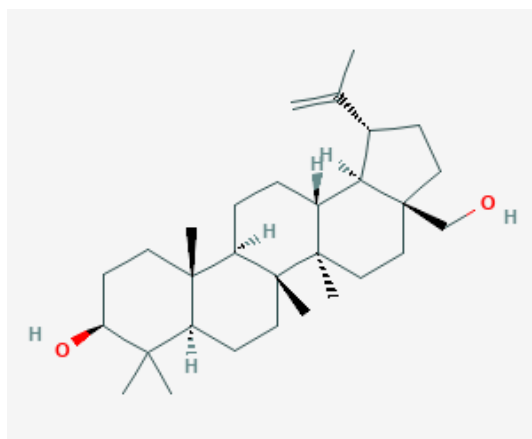


Figure 12. Chemical structure of Betulin obtained from PubChem.

Table 4. Pharmacokinetic properties of Betulin obtained from admetSAR.

Bioactivity	Value	Probability
Absorption		
Human Intestinal Absorption	+	0.9884
Blood Brain Barrier	-	0.4533
Human oral bioavailability	-	0.5857
P-glycoprotein inhibitor	-	0.8836
P-glycoprotein substrate	-	0.7347
Distribution and Metabolism		
Subcellular localization	Lysosomes	0.4831
CYP3A4 substrate	+	0.6751
CYP2C9 substrate	-	0.6284
CYP2D6 substrate	-	0.7222
CYP3A4 inhibition	-	0.8309
CYP2C9 inhibition	-	0.9071
CYP2C19 inhibition	-	0.9026
CYP2D6 inhibition	-	0.9297

CYP1A2 inhibition	-	0.9045
CYP inhibitory promiscuity	-	0.6441
Toxicity		
Acute Oral Toxicity (c)	III	0.7441
Hepatotoxicity	-	0.6250
Carcinogenicity	-	0.9857
Eye corrosion	-	0.9892
Eye irritation	-	0.9008
Ames mutagenesis	-	0.7500
Human ether-a-go-go inhibition	-	0.3617

Table 5. Physio-chemical properties of Betulin obtained from admetSAR.

Properties	Values
Lipinski's RO5 & Veber's rule:	
Molecular Weight	442.73 kDa
AlogP	7.00
H-Bond Acceptor	2
H-Bond Donor	2
Rotatable Bonds	2

Table 6. Regression properties of Betulin obtained from admetSAR.

Admet predicted profile --- Regressions:	
Water solubility	-4.319 (logS)
Plasma protein binding	0.891 (100%)
Acute Oral Toxicity	3.482 mol/kg
Tetrahymena pyriformis	1.282 (PIGC50 ug/L)

7. Molecular Docking Results

The structure of Betulin after analysis by admetSAR was docked with the PKC ϵ structure predicted by I-TASSER. Model 5 was selected for this purpose which is given in fig.7. The docking results generated by CB Dock (table.3) were in the form of vina scores and cavity sizes of the different interactions of protein with the ligand. The docked structure of the best interaction between the ligand and protein which has the score of -7.8 is given in the fig.13.

Table 7. Different binding modes along with vina scores and cavity sizes of model 5 of PKC ϵ and Betulin.

Vina score	Cavity size
-7.8	2243
-7.5	6772
-7.2	1181
-5.9	482
-5.2	329

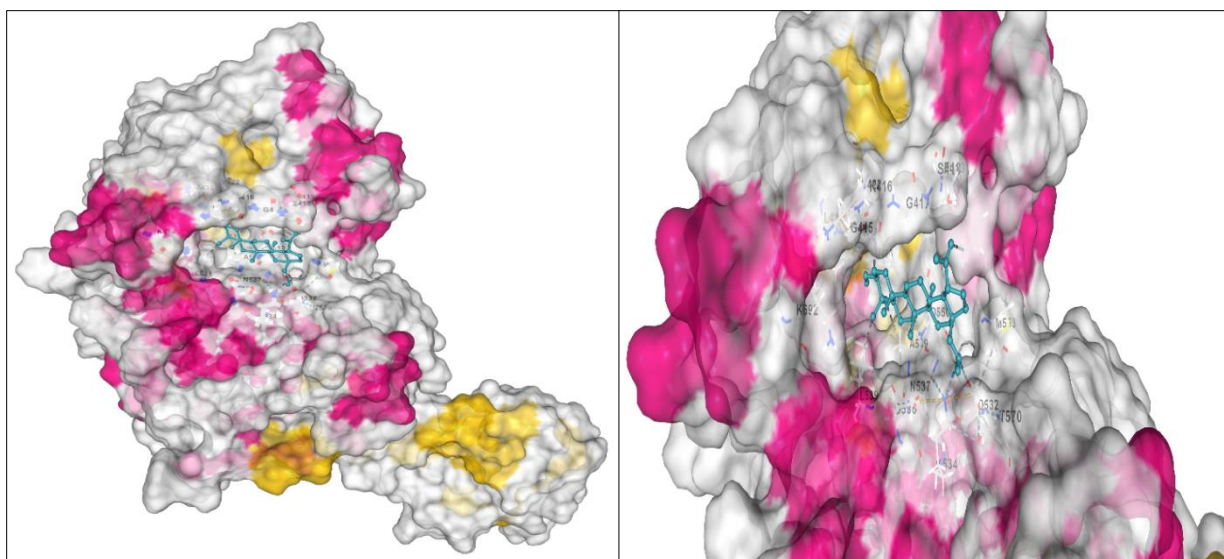


Figure 13. Surface view of PKCε and Betulin docked inside its binding cavity of size 2243 (left) Close up picture of docked betulin inside the binding cavity of PKCε (right).

The different kinds of molecular interactions of amino acids in the protein binding cavity with the ligand were visualized through LigPlot (fig.14). These interactions showed that the ligand was embedded deep in the hydrophobic pocket of the protein and formed strong hydrophobic interactions with it. Leucine 414 makes hydrophobic interaction with the first oxygen o while Threonine 570 makes hydrophobically interacts with the second oxygen of the ligand. Other residues which make hydrophobic interactions include Glycine 415 with C24, Leucine 539 with C27, Serine 418 with C26, C29 and C30, Phenylalanine 419 with C30, Aspartate 536 with C19, Aspartate 550 with C1 and C10, Asparagine 537 with C9 and C19, Aspartate 532 with C13 and O2, and Valine 422 with C23 and O1. Lysine 534 is

the only amino acid residue which interacts through Hydrogen bonds with the ligand having the bond length of 3.09 Angstrom.

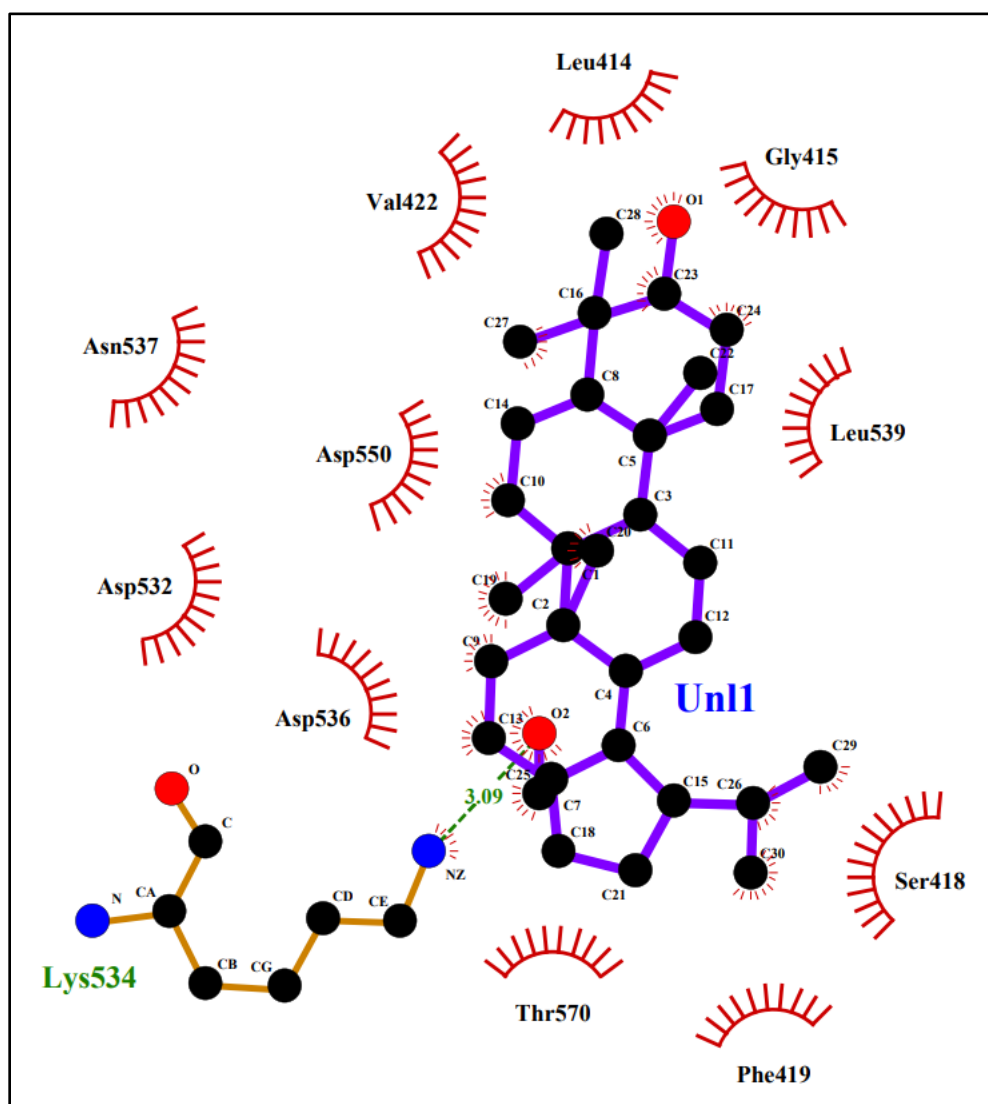


Figure 14. LigPlot visualization of docked Betulin in the binding pocket of PKC ϵ . The black and purple ball and stick structure is the ligand, Betulin, and the red spiked semi-circles surrounding it are the residues present in the binding pocket forming hydrophobic interactions with the ligand while the green dotted line gives the H-bond formed between Lysine 534 residue in the binding pocket with the ligand.

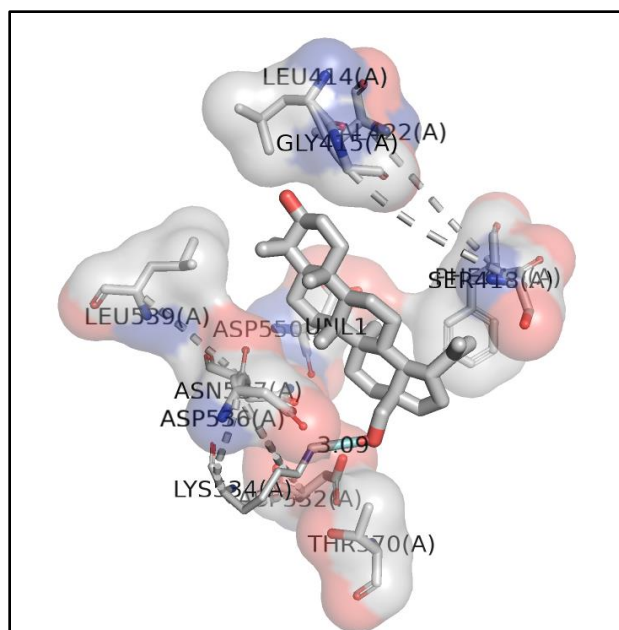
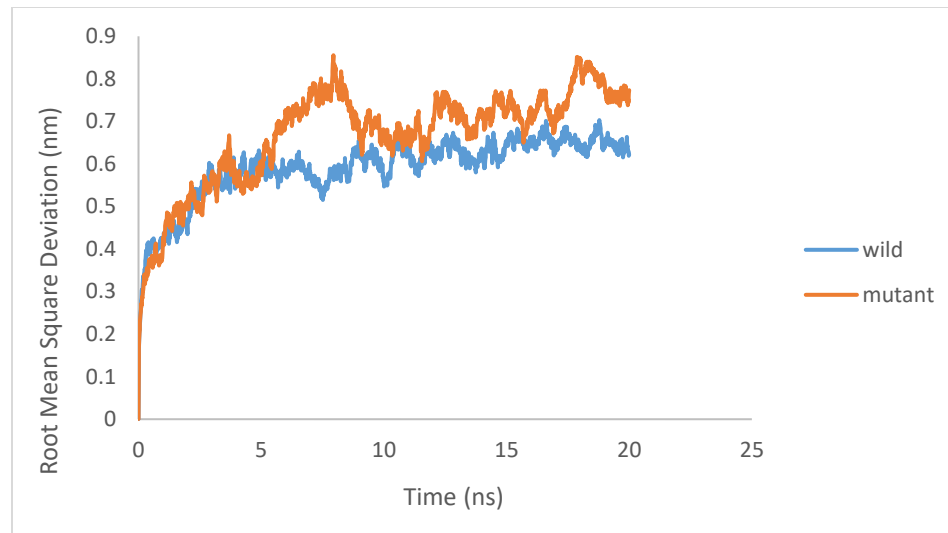


Figure 15. 3D visualization of docked PKC ϵ and Betulin on PyMOL.

8. Molecular Dynamic Simulation Results

Molecular dynamic simulations which were carried out using Gromacs software package generated files of ceratin important parameters related to both wild and mutated protein structures. The data in these files was converted to graphs(1,2,3,4). The root mean square deviation (RMSD) of the protein is plotted against time and it shows deviations in its structure from a reference point at the start of simulation. The RMSD of both wild and mutant PKC ϵ protein can be seen from the graph (graph1.) to remain closer to their initial confirmation, stable and almost similar till 5ns of the simulations resulting in the deviation of upto 0.6nm. After 5ns the RMSD of wild structure still ramins closer to 0.6nm and gradually reaches the maximum value of 0.7nm in the 20ns of simulations. But the mutant structure of PKC ϵ after 5ns suddenly deviates upto o.87nm and then comes back to 0.6nm at 10ns of the simulation. Through the rest of the time of simulation the RMSD of mutant remains higher than that of the wild structure. The abonormal peaks in the graph for both

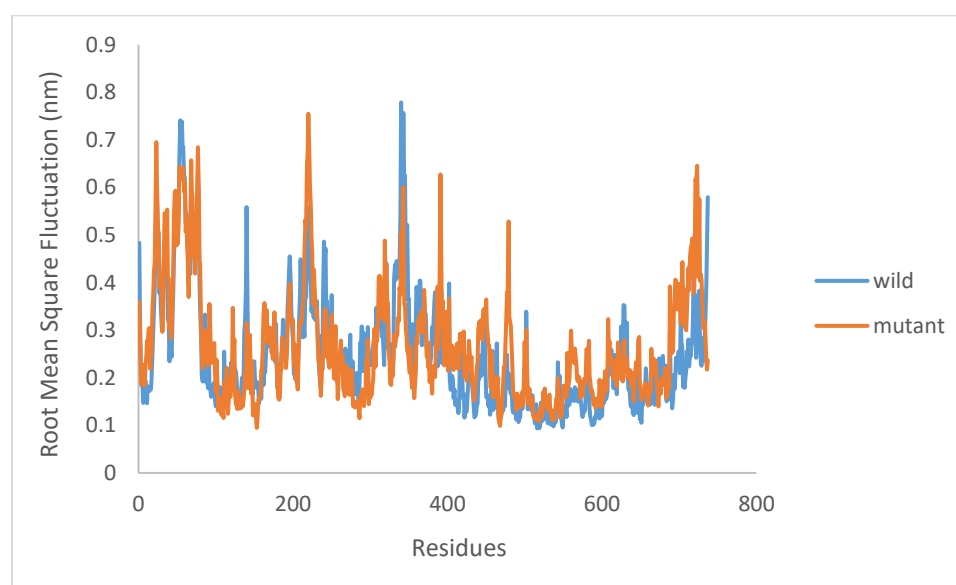
wild and mutant RMSDs indicate the presence of undefined loop structure in the protein which deviates a lot more than the defined structures such as α -helices and β -sheets. The mutated structure shows slightly larger deviation from the reference point at the start of simulations than the wild structure. However, this deviation is not significant enough to consider the mutant structure as unstable.



Graph 1. Root mean square deviation of PKC ϵ wild and mutant structures.

Root mean square fluctuation (RMSF) is the fluctuation in nm of individual residues of the protein from their average positions in a dynamic setting. RMSFs of both wild and mutant PKC ϵ structures were calculated to see the difference in the dynamic behavior of residues due to the mutation. The peaks show maximum kinetic energies of the residues. The number of peaks in the RMSF graph (graph2.) differ between the wild and mutant PKC ϵ structure. The fluctuation in residues of both wild and mutant PKC ϵ structures range from 0.1nm to 0.8nm. The initial upto 100 residues of the mutant structure which lie in the C2 regulatory domain of the protein and where the mutation is located, show more fluctuation peaks than the wild type residues in that region. Next the residues of mutant from about 250 to 310 which make up the end of C1 domain and start of hinge region show even lesser

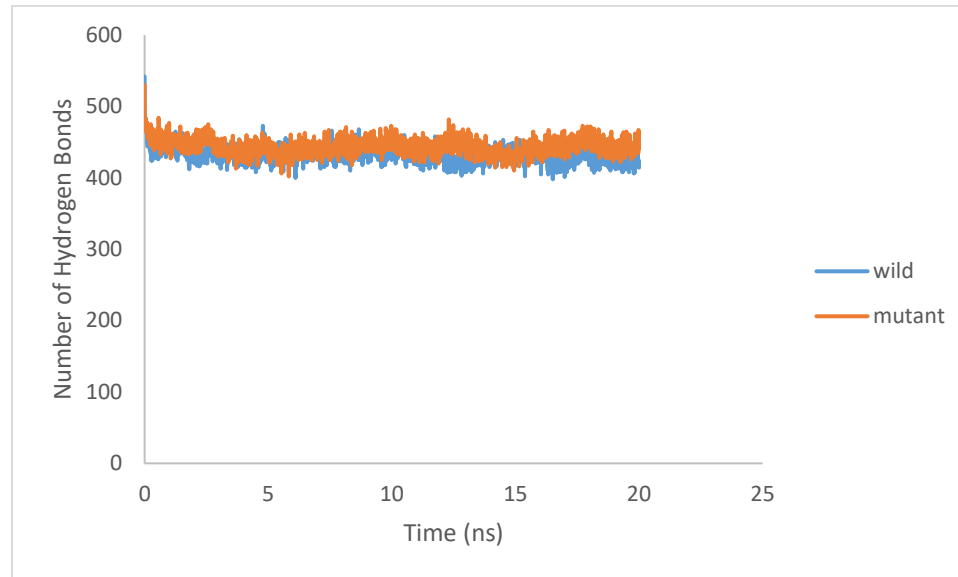
fluctuation than the wild protein structure. The mutant residue fluctuation slightly increases than the wild protein at the end of hinge region and start of the kinase domain from residue 390 and reaches the maximum of 0.64nm gradually decreasing and coming at the level of wild structure to ~0.3nm at after 490 residues. The fluctuation of mutant residues in the C-terminal domain (from about 680 residues onwards) again slightly increases than the wild and reaches 0.65nm while that of wild type residues remain upto 0.59nm. The overall fluctuation difference between the residues of mutant and wild structure is not big although from some of the fluctuation peaks of the mutant structure, its residues appear to have more kinetic energy than the wild type.



Graph 2. Root mean square fluctuation of PKC ϵ wild and mutant structures.

Hydrogen bonds form between the polar side chains of the amino acid residues. The number of hydrogen bonds formation in both the structures of PKC ϵ over the time period of 20ns of the simulations is shown in graph 3. The total number of H-bonds forming in both structures remain between 400 to 500 and this number remains almost constant throughout the simulations. There is no significant difference in the number of H-bonds of

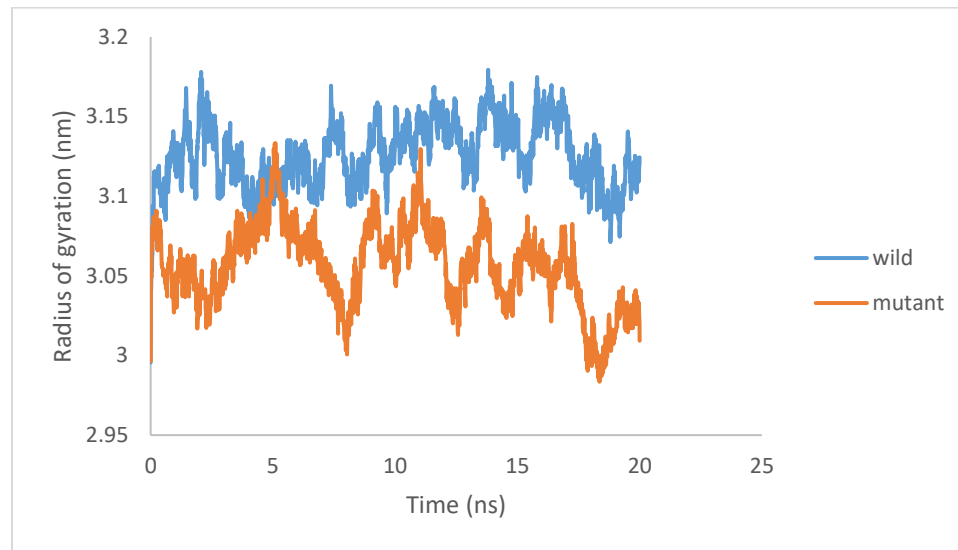
both PKC ϵ structures because the change is of only one amino acid residue. The slight difference in H-bonding of both structures might be due to the variation in bond length of the residue-residue contacts around the mutation.



Graph 3. Number of hydrogen bonds formation over the time of 20ns in PKC ϵ wild and mutant structures.

Radius of gyration (Rg) is the measure of distance of all the atoms of a protein from their collective center or axis of rotation. Rg is the radius of a protein structure in a dynamic setting which shows compactness of the protein and change in its folding over the time. The change in Rg of wild and mutant PKC ϵ structures can be seen in graph 3. Rg of the wild structure immediately increases upto 3.17nm from 3.08nm in the first 2.5ns of the simulations after which it comes back to 3.08nm at the same rate till 5ns. At 5ns we can again see a gradual increase in Rg where it keeps on resonating between 3.09 to 3.17nm for nearly 17 ns after which we drop to 3.07nm and again start increasing gradually. In contrast to the wild structure, Rg of mutant PKC ϵ decreases from 3.09 to 3.02 in the first 2.5ns, after which it increases upto 3.14nm at 5ns and again decreases upto 3nm at 7ns of the simulations. At 8ns we again see a peak of 3.11nm which comes down to 3.04nm at

10ns and again climbs upto 3.14nm at 11ns. At 13ns the mutant R_g again comes down to 3.01nm and then resonates between 3.1nm and 3.03nm until 17ns before dropping to 2.99nm and again climbing upto 3.04nm till the end of simulations. This graph shows that the total radius of gyration of the mutant decreases markedly with the passage of time than the wild structure. This means the overall compactness of mutant structure increases which implies that the mutated structure became more stable than the wild type PKC ϵ structure.



Graph 4. Radius of gyration of PKC ϵ wild and mutant structures.

9. PCR Results

The tetra ARMS PCR data for the selected SNP of PRKCE (rs1553369874) was collected and statistical tools were applied for exploring its association with HCC. The statistical analysis of this data is given in tables 7,8 and 9. The total analysis (table7.) of all the genotypes of selected locus shows that the data was significant with the P-values ranging smaller than 0.05. The homozygous wild genotype GG has the highest chance of being associated with HCC with its odd ratio 4.19 and relative risk 2.1 and P value <0.0001. The

other two genotypes have less than 1 odds ratio and relative risk which means they cannot be associated with HCC development and progression.

The analysis of this SNP data with respect to gender shows significance of the data for all genotypes in both genders except for AA in females. The data shows that there is slight difference in the odds ratio and relative risk of GG in males and females. The odds ratio and relative risk of GG association with HCC is higher in males than in females. In males the odds ratio of GG is 6.824 and relative risk is 2.737 whereas in females its odds ratio is 2.794 and relative risk is 1.711. The P-value of other two genotypes AG in both males and females, and AA in males show their protective significance because of their insignificant (less than 1) odds ratio and risk factor.

The analysis with respect to age, showed significance of data of all the genotypes in age groups 20-39 and 40-59, except for the data of AA genotype in 40-59 age group. These results also show the association of homozygous wild genotype GG with the HCV induced HCC in all age groups and the protective significance of AG genotype in all ages and AA genotype in age group of 20-39.

Table 8. Statistical analysis of PRKCE SNP data obtained from PCR.

Total Patient and Control Data							
Genotype	Frequency Distribution		Odds Ratio		Relative Risk		P value
	%		Value	95% CI	Value	95% CI	
	Patients	Control					
AA	16.16%	32.32%	0.4036	0.2100 to 0.8024	0.6024	0.3837 to 0.8877	0.0124
GG	72.00%	38.00%	4.195	2.264 to 7.468	2.104	1.527 to 2.981	<0.0001

AG	12.00%	30.00%	0.3182	0.1535 to 0.6541	0.5130	0.3031 to 0.8039	0.0029
-----------	--------	--------	--------	---------------------	--------	---------------------	--------

Table 9. Statistical analysis of PCR data of PRKCE SNP with respect to gender.

Patient and Control Data with respect to Gender							
Genotype	Frequency Distribution		Odds Ratio		Relative Risk		P value
	%		Value	95% CI	Value	95% CI	
	Patients	Control					
AA (F)	20.00%	30.19%	0.5781	0.2237 to 1.472	0.7404	0.4168 to 1.188	0.2635
GG (F)	64.00%	38.89%	2.794	1.262 to 6.054	1.711	1.130 to 2.675	0.0118
AG (F)	16.00%	31.48%	0.4146	0.1607 to 1.106	0.6019	0.3144 to 1.028	0.0716
AA (M)	12.24%	34.78%	0.2616	0.09103 to 0.7278	0.4630	0.2186 to 0.8535	0.0141
GG (M)	80.00%	36.96%	6.824	2.766 to 15.84	2.737	1.642 to 4.919	<0.0001
AG (M)	8.00%	28.26%	0.2207	0.07444 to 0.6796	0.4041	0.1615 to 0.8405	0.0146

Table 10. Statistical analysis of PCR data of PRKCE SNP with respect to gender.

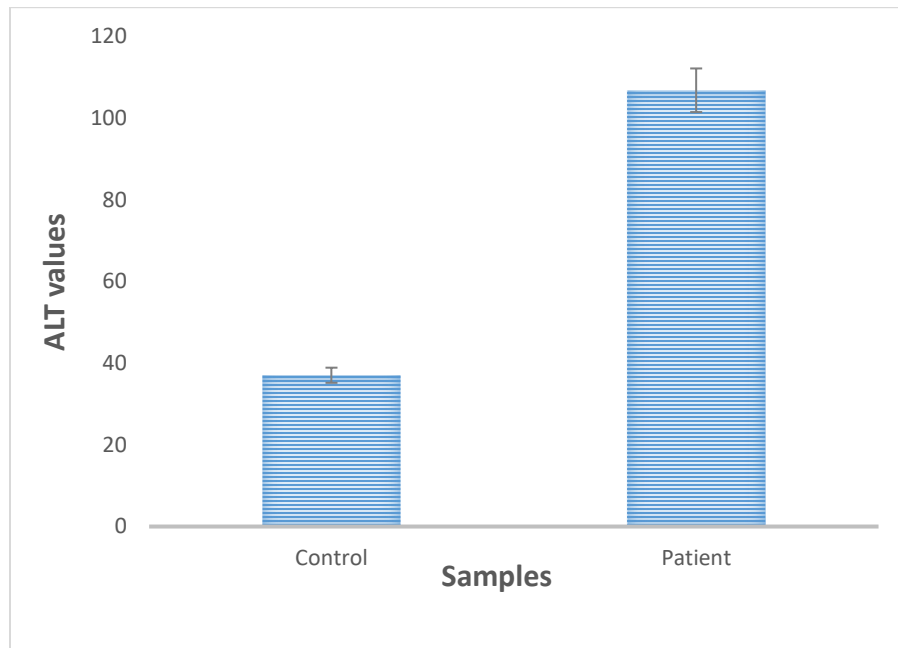
Patient and Control Data with respect to Age							
Genotype	Frequency Distribution		Odds Ratio		Relative Risk		P value
	%		Value	95% CI	Value	95% CI	
	Patients	Control					

1-19 AA	0.00%	0.00%	—	—	0.000	0.000 to 1.000	>0.9999
1-19 GG	100.00%	100.00%	—	—	Infinity	6.939e- 018 to Infinity	>0.9999
1-19 AG	0.00%	0.00%	—	—	0.000	0.000 to 1.000	>0.9999
20-39 AA	16.67%	34.09%	0.3867	0.1420 to 1.001	0.6000	0.3158 to 1.007	0.0904
20-39 GG	72.92%	36.36%	4.712	1.986 to 11.30	2.164	1.380 to 3.610	0.0007
20-39 AG	10.42%	29.55%	0.2773	0.1022 to 0.8984	0.4780	0.2110 to 0.9138	0.0337
40-59 AA	17.78%	27.66%	0.5655	0.2036 to 1.533	0.7310	0.3850 to 1.221	0.3236
40-59 GG	69.57%	39.58%	3.489	1.501 to 7.755	1.927	1.227 to 3.181	0.0041
40-59 AG	13.04%	33.33%	0.3000	0.1044 to 0.8337	0.4909	0.2310 to 0.9108	0.0279
60+ AA	0.00%	40.00%	0.000	0.000 to 5.724	0.000	0.000 to 2.774	>0.9999
60+ GG	100.00%	40.00%	Infinity	0.3717 to Infinity	Infinity	0.5871 to Infinity	0.4286
60+ AG	0.00%	33.33%	0.000	0.000 to 7.077	0.000	0.000 to 3.358	>0.9999

10. ALT Results

The mean levels of ALT in control versus HCV induced HCC patients is illustrated in graph 5. The ALT values of control group lie around 37.5 ± 10.2 whereas those of HCV

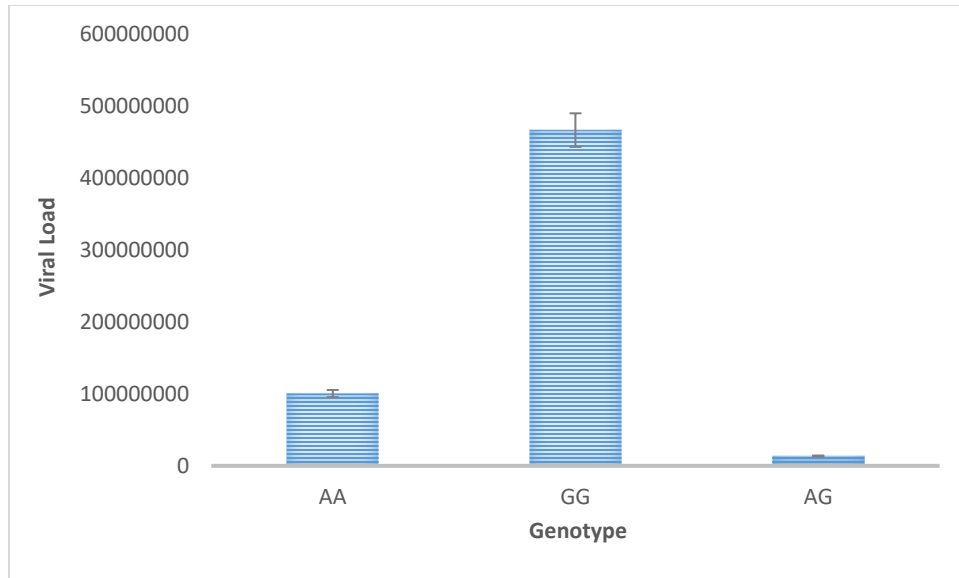
iduced HCC patients are close to 106.84 ± 52.5 . These values indicate significant difference between ALT levels of both the groups.



Graph 5. ALT levels in HCV induced HCC patients versus control samples.

11. Viral Load

The viral load measured by qRT-PCR was plotted in the graph with respect to the three genotypes of selected locus of PRKCE (graph 6.). This graph shows clear association of the homozygous wild GG genotype with the abundance of viral load in HCV patients which might be a leading cause for the development and progression of HCC.



Graph 6. Viral load of HCV induced HCC patient genotypes.

CHAPTER 5

Discussion

The structure of PKC ϵ predicted using its protein sequence. From the five structures predicted by I-TASSER one (fig.7) was selected based on its structural similarities with the already known proteins present in PDB database. Each of the domain of this structure was compared and cross checked using InterPro and other literature sources. The first C2 domain is highly conserved in the C2 domains of InterPro homologous superfamily. In rest of the proteins such as Synaptotagmin I it has been observed to fold into an 8-stranded β -sandwich which can adopt Type I or Type II structural arrangements. These arrangements are distinguished by the circular permutation involved in the N- and C- terminal β -strands (Sutton, Davletov, Berghuis, Sudhof, & Sprang, 1995).

There is a high degree of structural homology between PKC ϵ model 5 (fig.7) C2 domain and the ones found in literature. Apart from the top and bottom loop structure, PKC ϵ model 5 consists of 6-stranded core β -sandwich. The C2-domain sequences involved in the core β -sandwich are highly conserved between C2-domains whereas the sequences of the loops, particularly loop 1 and the three bottom loops, are not (Brose, Hofmann, Hata, & Südhof, 1995). The high degree of structural homology between the core β -sandwiches of C2-domains suggests that the β -sandwich represents a scaffold. This scaffold allows the emergence of variable loops at the top and bottom of the domain (Farah & Sossin, 2012). The next C1 domain which is a DAG/PE binding domain does not have much supportive evidence for its structure but in a broader view, cysteine-rich sequences of this domain resemble the DNA-binding zinc finger domain present in several proteins. The classical ZF is a compactly folded globular structure consisting of two cysteine and two histidine residues (Cys₂His₂) in each repeated unit of 28–30 amino acids (tandem repeat) bound to a single

zinc atom. This structure usually consists of a polypeptide backbone fold having two conserved antiparallel β -sheets forming a hairpin opposite to a conserved α -helix structure (Azzi, Boscoboinik, & Hensey, 1992; Ono et al., 1989). PKC ϵ model 5 has the C1 domain consisting of both secondary structure, α -helix and β -sheets.

The catalytic protein kinase domain is highly conserved among different protein kinases (Hunter, 1991). The crystal structure of the catalytic subunit of cyclic AMP-dependent protein kinase of AGC superfamily shows the smaller lobe which consists mostly of amino-terminal sequence, is associated with nucleotide binding, and its largely antiparallel beta sheet architecture constitutes an unusual nucleotide binding motif. The larger lobe is dominated by helical structure with a single beta sheet at the domain interface. This lobe is primarily involved in peptide binding and catalysis. (Hanks, Quinn, & Hunter, 1988). The structure of PKC ϵ also displays these features. There are two lobes in its catalytic domain, one dominated by β -sheets and the other mainly constituting α -helices. The C-terminal sequence has high degree of conservation among all proteins of the AGC kinase superfamily. Generally, it consists of two phosphorylation sites. The first site is in a turn motif while the second one is present at the end of the domain in a hydrophobic pocket. The secondary structure of this terminal, however, is not well defined in any of the proteins (P. Parker & Parkinson, 2001; Jing Yang et al., 2002). In PKC ϵ , we can see this region largely as an undefined loop structure.

The change in protein expression and its structural and functional properties has often been associated to the presence of certain kinds of polymorphisms in the genetic sequence (Decock et al., 2007; Fu et al., 2010; Robert & Pelletier, 2018). The presence of missense SNPs has been observed to cause structural changes and enhance the function of oncogenes due to their occurrence in the functional sites of the genes (Stehr et al., 2011). Thus, missense SNPs of certain genes can

potentially increase the susceptibility to develop cancers (Mita et al., 2010; Thirumal Kumar & George Priya Doss, 2017). In this research, the Ensemble database for missense SNPs for PKC ϵ was accessed and all the missense variants were evaluated through widely used SNP screening tools which are SIFT, PolyPhen, CADD, REVEL, MetaLR and Mutation Assessor. The selected SNP with variant ID rs1553369874 when sifted through all these tools, showed promising results in terms of its association with pathogenic properties (Suybeng, Koeppel, Harlé, & Rouleau, 2020). The structural and functional properties of the selected SNP were also assessed using tools such as I-Mutant, MutPred and HOPE (Kumar, Swetha, Anbarasu, & Ramaiah, 2014). The results from these tools suggested that with the induction of the selected point mutation which is located in the C2 regulatory domain of PKC ϵ , the stability of it is likely to decrease (-0.49 Kcal/mol DDG value predicted by I-Mutant) along with the gain of MoRF binding ($P = 0.0042$), gain of methylation at E14 ($P = 0.0173$), and loss of ubiquitination at K11 ($P = 0.0432$). The wet lab results, however, suggest otherwise. According to the ARMS-PCR data generated after amplifying the SNP in nearly 200 samples (100 controls and 100 HCV induced HCC patients) and applying statistical tools using GraphPad Prism, it was observed that the homozygous wild genotype GG has significant association with HCV induced HCC with 2.1 relative risk and 4.19 odds ratio. The analysis of PCR data with respect to gender also indicated the association of the same genotype in both males and females, although its relative risk in males and odds ratio is slightly greater than in females. This difference could also be due to the relatively greater incidence rate of liver cancer, more specifically HCC in males than in females. The male to female ratio of incidence and mortality of HCC exceeds 2.5 for both (Hafeez, Mahmood, Khan, & Malkani, 2020; Singal, Lampertico, & Nahon, 2020). The association of homozygous GG genotype was also found in the PCR data

analysis with respect to age. From 20yrs to 59yrs of age, GG genotype was found more frequently in all HCV induced HCC patients.

MD simulations were carried out to virtually observe the behavior of both mutated and wild protein phenotypes in a dynamic environment. The results from MD simulation support our wet lab generated results. The slight fluctuation in the RMSD and RMSF and Rg of mutant structure from the wild PKC ϵ structure might be because of the change in the residue-residue interaction within the protein. In wild type glutamate is present at the 14th position in primary structure of the protein whereas in mutant it is replaced by lysine. Glutamate is acidic in nature and also smaller in size than lysine which is basic in nature. Due to their ionic properties, both glutamate and lysine can form salt bridges in the protein structure (Marqusee & Baldwin, 1987). Salt bridges are formed when the charged residues experience electrostatic attraction due to their close proximity leading to the formation of ionic interaction in addition to the H-bonds. Formation of salt bridges is highly dependent on the nature and distance between the charged residues. The change in nature as well as size of the residue in the mutant might have affected its number of contacts with the neighboring residues, electrostatic interactions and H-bonds which led to the difference in RMSD and RMSF values of both the structures. There is a slight difference between the number of H-bonds that form over time in mutant and wild structures. Moreover, the number of contacts also determines the folding of protein which is not necessarily correct with an increased number of contacts. Another topological feature of proteins that is the rotatable angle of individual residues can also affect the folding and compactness of protein. The number of rotatable angles is 5 for glutamate and 6 for lysine (Galzitskaya & Garbuzynskiy, 2006). These angles can also change the intra-residue interactions of the protein structure. All these properties are manifested in the MD simulation results especially in the Rg graphs of both wild and mutant PKC ϵ structure where the radius of the

mutant structure clearly decreases than the wild type due to possible changes in size, residue interactions, rotatable angles and protein folding.

In addition to the variant association with HCV induced HCC, the liver function and viral load was also checked for these samples. The ALT values of patients versus control clearly pointed out increased liver damage in HCC. Apart from that, elevated viral load with the GG genotype in HCV induced HCC patients was found. This could be an indication of the increased susceptibility of individuals with this genotype to hepatitis C viral pathogenecity. Previously SNPs have been associated with increased viral load and disease progression in individuals with chronic viral infections (Anokhin et al., 2016; Zhong et al., 2016).

The association of PKC ϵ variant with hepatocellular carcinoma open new direction for therapeutics. A potential drug, Betulin, against PKC ϵ was selected on the basis of ametsAR predictions. The predictions for Lipinski's rule of 5 indicated betulin as promising drug candidate. Lipinski's rule of 5 gives standard values of molecular mass less than 500 Da, H-bond donors no more than 5, H-bond acceptors no more than 10 and logP (octanal-water partition coefficient) upto 5. Betulin has all of these vlues within the given range except logP which is 7. Besides that admet profile of drugs also include prediction of drug behavior against 4 regression models which include logS (water solubility), plasma protein binding, acute oral toxicity and PIGC50 in Tetrahymena pyriformis model organism. Water solunbility of Betulin is -4.319 which is less than 1, thus indicating its hydrophobic properties. Besides that, 0.891 out of 100% molecules of Betulin can bind to plasma proteins and it can cause oral toxicity at the dose of 3.482 mol/kg. The concentration which can cause 50% growth inhibition of Tetrahymena pyriformis is 3.482 mol/kg. The pharmacokinetic properties of Betulin predicted by admetSAR idicate its subcellular localization in lysosomes and human intestinal absorption with 0.9884 probability. Apart from that, it is not

hepatotoxic, biodegradable, orally available and does not cross blood brain barrier. All these properties indicate that betulin is a promising drug candidate. Betulin has already been studied and tested for its antitumor, apoptosis inducing properties against HCC and other cancers (Wang et al., 2019; Yin et al., 2019). Therefore, molecular docking of betulin with PKC ϵ was performed to see its interaction with it. It showed good interaction with multiple hydrophobic interactions and a H-bond in the binding pocket of PKC ϵ with vina score of -7.8. Docking of this molecule in PKC ϵ was not specific for the mutated or wild type structure which keeps an open window for potential drug against this protein irrespective of its mutations.

Conclusion

In silico identification of disease associated variants of PKC ϵ was done and a potentially pathogenic SNP was selected. Tertiary structure of PKC ϵ was also predicted and mutagenesis of the selected SNP was induced in it. MD simulations of both predicted structures were carried out to get an insight about the effect of selected mutation on the protein structure and function. Association of the selected SNP with HCV induced HCC was investigated through tetra ARMS PCR. The results of PCR indicated the association of GG variant of selected SNP of PKC ϵ with HCV induced HCC specifically in the age group of 20-59yrs. Viral load of this GG genotype was also very high among all three genotypes of HCV induced HCC patients. Thus, the variant GG of rs1553369874 was identified as the risk factor for HCV induced HCC occurrence. For a potential therapy, physiochemical and pharmacokinetic properties of Betulin were inspected and its interaction in the binding pocket of PKC ϵ was analyzed. Betulin has potential to be therapeutic agent against PKC ϵ but *in vitro* and *in vivo* testing for its potential as a drug is needed.

References

- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, 76(1), 7.20. 21-27.20. 41.
- Akita, Y. (2002). Protein kinase C- ϵ (PKC- ϵ): its unique structure and function. *The Journal of Biochemistry*, 132(6), 847-852.
- Ali, A. S., Ali, S., El-Rayes, B. F., Philip, P. A., & Sarkar, F. H. (2009). Exploitation of protein kinase C: a useful target for cancer therapy. *Cancer treatment reviews*, 35(1), 1-8.
- Ann, J., Yoon, S., Baek, J., Kim, D. H., Lewin, N. E., Hill, C. S., . . . Lee, J. (2015). Design and synthesis of protein kinase C epsilon selective diacylglycerol lactones (DAG-lactones). *European journal of medicinal chemistry*, 90, 332-341.
- Anokhin, V. V., Bakhteeva, L. B., Khasanova, G. R., Khaiboullina, S. F., Martynova, E. V., Tillett, R. L., . . . Rizvanov, A. A. (2016). Previously unidentified single nucleotide polymorphisms in HIV/AIDS cases associate with clinical parameters and disease progression. *BioMed research international*, 2016.
- Aziz, M. H., Shen, H., & Maki, C. G. (2012). Glucocorticoid receptor activation inhibits p53-induced apoptosis of MCF10Amc cells via induction of protein kinase C ϵ . *Journal of biological chemistry*, 287(35), 29825-29836.
- Azzi, A., Boscoboinik, D., & Hensey, C. (1992). The protein kinase C family. *European Journal of Biochemistry*, 208(3), 547-557.
- Bae, K.-M., Wang, H., Jiang, G., Chen, M. G., Lu, L., & Xiao, L. (2007). Protein Kinase C ϵ Is Overexpressed 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 research*, 67(13), 6053-6063.
- Basta, P., Strickland, M. B., Holmes, W., Loomis, C. R., Ballas, L. M., & Burns, D. J. (1992). Sequence and expression of human protein kinase C- ϵ . *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1132(2), 154-160.
- Basu, A., & Sivaprasad, U. (2007). Protein kinase C ϵ makes the life and death decision. *Cellular signalling*, 19(8), 1633-1642.
- Black, A. R., & Black, J. D. (2013). Protein kinase C signaling and cell cycle regulation. *Frontiers in immunology*, 3, 423.
- Black, J. D. (2000). Protein kinase C-mediated regulation of the cell cycle. *Front Biosci*, 5(1), D406.
- Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., . . . Raj, S. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, 49(D1), D344-D354.
- Brandman, R., Disatnik, M.-H., Churchill, E., & Mochly-Rosen, D. (2007). Peptides derived from the C2 domain of protein kinase C epsilon (epsilon PKC) modulate epsilon PKC activity and identify potential protein-protein interaction surfaces. *The Journal of biological chemistry*, 282(6), 4113.
- Brose, N., Hofmann, K., Hata, Y., & Südhof, T. C. (1995). Mammalian Homologues of Caenorhabditis elegans unc-13 Gene Define Novel Family of C2-domain Proteins (*). *Journal of Biological Chemistry*, 270(42), 25273-25280.
- Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*, 33(suppl_2), W306-W310.
- Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., . . . Tang, Y. (2012). admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. In: ACS Publications.
- Collins, A., & Ke, X. (2012). Primer1: primer design web service for tetra-primer ARMS-PCR. *The Open Bioinformatics Journal*, 6(1).
- Coutinho, I., Pereira, C., Pereira, G., Gonçalves, J., Côte-Real, M., & Saraiva, L. (2011). Distinct regulation of p53-mediated apoptosis by protein kinase C α , δ , ϵ and ζ : Evidence in yeast for transcription-

- dependent and-independent p53 apoptotic mechanisms. *Experimental cell research*, 317(8), 1147-1158.
- Decock, J., Long, J.-R., Laxton, R. C., Shu, X.-O., Hodgkinson, C., Hendrickx, W., . . . Paridaens, R. (2007). Association of matrix metalloproteinase-8 gene variation with breast cancer prognosis. *Cancer research*, 67(21), 10214-10221.
- Deng, N., Zhou, H., Fan, H., & Yuan, Y. (2017). Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget*, 8(66), 110635.
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., & Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human molecular genetics*, 24(8), 2125-2137.
- Dragani, T. A. (2010). Risk of HCC: genetic heterogeneity and complex genetics. *Journal of hepatology*, 52(2), 252-257.
- Duquesnes, N., Lezoualc'h, F., & Crozatier, B. (2011). PKC-delta and PKC-epsilon: foes of the same family or strangers? *Journal of molecular and cellular cardiology*, 51(5), 665-673.
- Farah, C. A., & Sossin, W. S. (2012). The role of C2 domains in PKC signaling. *Calcium Signaling*, 663-683.
- Fu, Y.-P., Edvardsen, H., Kaushiva, A., Arhancet, J. P., Howe, T. M., Kohaar, I., . . . Fosså, S. D. (2010). NOTCH2 in breast cancer: association of SNP rs11249433 with gene expression in ER-positive breast tumors without TP53 mutations. *Molecular cancer*, 9(1), 1-11.
- Galzitskaya, O. V., & Garbuzynskiy, S. O. (2006). Entropy capacity determines protein folding. *Proteins: Structure, Function, and Bioinformatics*, 63(1), 144-154.
- Garg, R., Benedetti, L. G., Abera, M. B., Wang, H., Abba, M., & Kazanietz, M. G. (2014). Protein kinase C and cancer: what we know and what we do not. *Oncogene*, 33(45), 5225-5237.
- Gorin, M. A., & Pan, Q. (2009). Protein kinase C ϵ : an oncogene and emerging tumor biomarker. *Molecular cancer*, 8(1), 1-8.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., . . . Stevens, C. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132), 153-158.
- Hafeez, S., Mahmood, A., Khan, R. U., & Malkani, N. (2020). Trends in Cancer Prevalence in Punjab, Pakistan: A Systematic Study from 2010 to 2016. *Journal of Bioresource Management*, 7(2), 8.
- Han, E. K.-H., Cacace, A. M., Sgambato, A., & Bernard Weinstein, I. (1995). Altered expression of cyclins and C-FOS in R6 cells that overproduce PKC ϵ . *Carcinogenesis*, 16(10), 2423-2428.
- Hanks, S. K., Quinn, A. M., & Hunter, T. (1988). The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science*, 241(4861), 42-52.
- Hunter, T. (1991). [1] Protein kinase classification. *Methods in enzymology*, 200, 3-37.
- Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Karyadi, D. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, 99(4), 877-885.
- Isakov, N. (2018). *Protein kinase C (PKC) isoforms in cancer, tumor promotion and tumor suppression*. Paper presented at the Seminars in cancer biology.
- Jain, K., & Basu, A. (2014). The multifunctional protein kinase C- ϵ in cancer development and progression. *Cancers*, 6(2), 860-878.
- Jerdeva, G. V., Yarber, F. A., Trousdale, M. D., Rhodes, C. J., Okamoto, C. T., Dartt, D. A., & Hamm-Alvarez, S. F. (2005). Dominant-negative PKC- ϵ impairs apical actin remodeling in parallel with inhibition of carbachol-stimulated secretion in rabbit lacrimal acini. *American Journal of Physiology-Cell Physiology*, 289(4), C1052-C1068.
- Kim, M. H., Jung, Y.-S., Moon, C.-H., Jeong, E.-M., Lee, S. H., Baik, E. J., & Moon, C.-K. (2003). Isoform-specific induction of PKC- ϵ by high glucose protects heart-derived H9c2 cells against hypoxic injury. *Biochemical and biophysical research communications*, 309(1), 1-6.

- Kumar, C. V., Swetha, R. G., Anbarasu, A., & Ramaiah, S. (2014). Computational analysis reveals the association of threonine 118 methionine mutation in PMP22 resulting in CMT-1A. *Advances in bioinformatics*, 2014.
- Lehel, C., Oláh, Z., Jakab, G., Szállási, Z., Petrovics, G., Harta, G., . . . Anderson, W. B. (1995). Protein Kinase C ϵ Subcellular Localization Domains and Proteolytic Degradation Sites A MODEL FOR PROTEIN KINASE C CONFORMATIONAL CHANGES. *Journal of biological chemistry*, 270(33), 19651-19658.
- Liu, Y., Grimm, M., Dai, W.-t., Hou, M.-c., Xiao, Z.-X., & Cao, Y. (2020). CB-Dock: a web server for cavity detection-guided protein–ligand blind docking. *Acta Pharmacologica Sinica*, 41(1), 138-144.
- Livneh, E., & Fishman, D. D. (1997). Linking protein kinase C to cell-cycle control. *European Journal of Biochemistry*, 248(1), 1-9.
- Maeno-Hikichi, Y., Chang, S., Matsumura, K., Lai, M., Lin, H., Nakagawa, N., . . . Zhang, J.-f. (2003). A PKC ϵ –ENH–channel complex specifically modulates N-type Ca²⁺ channels. *Nature neuroscience*, 6(5), 468-475.
- Marqusee, S., & Baldwin, R. L. (1987). Helix stabilization by Glu... Lys+ salt bridges in short peptides of de novo design. *Proceedings of the National Academy of Sciences*, 84(24), 8898-8902.
- Mehta, K. D., Radominska-Pandya, A., Kapoor, G. S., Dave, B., & Atkins, B. A. (2002). Critical role of diacylglycerol-and phospholipid-regulated protein kinase C ϵ in induction of low-density lipoprotein receptor transcription in response to depletion of cholesterol. *Molecular and cellular biology*, 22(11), 3783-3793.
- MELLOR, H., & PARKER, P. J. (1998). The extended protein kinase C superfamily. *Biochemical Journal*, 332(2), 281-292. doi:10.1042/bj3320281
- Mita, Y., Yasuda, Y., Sakai, A., Yamamoto, H., Toyooka, S., Gunduz, M., . . . Shimizu, K. (2010). Missense polymorphisms of PTPRJ and PTPN13 genes affect susceptibility to a variety of human cancers. *Journal of cancer research and clinical oncology*, 136(2), 249-259.
- Newton, A. C. (1995). Protein kinase C: structure, function, and regulation. *Journal of biological chemistry*, 270(48), 28495-28498.
- Newton, A. C. (2010). Protein Kinase C: Relaying signals from lipid hydrolysis to protein phosphorylation. In *Handbook of Cell Signaling* (pp. 1123-1129): Elsevier.
- Newton, A. C. (2018). Protein kinase C: perfectly balanced. *Critical reviews in biochemistry and molecular biology*, 53(2), 208-230.
- Ono, Y., Fuji, T., Ogita, K., Kikkawa, U., Igarashi, K., & Nishizuka, Y. (1987). Identification of three additional members of rat protein kinase C family: gd-, ϵ -and ξ -subspecies. *FEBS letters*, 226(1), 125-128.
- Ono, Y., Fujii, T., Igarashi, K., Kuno, T., Tanaka, C., Kikkawa, U., & Nishizuka, Y. (1989). Phorbol ester binding to protein kinase C requires a cysteine-rich zinc-finger-like sequence. *Proceedings of the National Academy of Sciences*, 86(13), 4868-4871.
- Parker, P., & Parkinson, S. (2001). AGC protein kinase phosphorylation and protein kinase C. *Biochemical society transactions*, 29(6), 860-863.
- Parker, P. J., Coussens, L., Totty, N., Rhee, L., Young, S., Chen, E., . . . Ullrich, A. (1986). The complete primary structure of protein kinase C--the major phorbol ester receptor. *Science*, 233(4766), 853-859.
- Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H.-J., . . . Iakoucheva, L. M. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature communications*, 11(1), 1-13.
- Perletti, G., Tessitore, L., Sesca, E., Pani, P., Dianzani, M. U., & Piccinini, F. (1996). ϵ PKC Acts Like a Marker of Progressive Malignancy in Rat Liver, but Fails to Enhance Tumorigenesis in Rat Hepatoma Cells in Culture. *Biochemical and biophysical research communications*, 221(3), 688-691.
- Pravdic, D., Sedlic, F., Mio, Y., Vladic, N., Bienengraeber, M., & Bosnjak, Z. J. (2009). Anesthetic-induced Preconditioning Delays Opening of Mitochondrial Permeability Transition Pore via Protein Kinase

- C- ϵ -mediated Pathway. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, *111*(2), 267-274.
- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research*, *47*(D1), D886-D894.
- Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*, *39*(17), e118-e118.
- Robert, F., & Pelletier, J. (2018). Exploring the impact of single-nucleotide polymorphisms on translation. *Frontiers in genetics*, *9*, 507.
- Schrödinger, L. (2010). The PyMOL Molecular Graphics System, Version 1.2 r3pre, Schrödinger, LLC.
- Shankar, E., Sivaprasad, U., & Basu, A. (2008). Protein kinase C ϵ confers resistance of MCF-7 cells to TRAIL by Akt-dependent activation of Hdm2 and downregulation of p53. *Oncogene*, *27*(28), 3957-3966.
- Shastri, B. S. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. *Journal of human genetics*, *52*(11), 871-880.
- Sia, D., Villanueva, A., Friedman, S. L., & Llovet, J. M. (2017). Liver cancer cell of origin, molecular class, and effects on patient prognosis. *Gastroenterology*, *152*(4), 745-761.
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research*, *40*(W1), W452-W457.
- Singal, A. G., Lampertico, P., & Nahon, P. (2020). Epidemiology and surveillance for hepatocellular carcinoma: New trends. *Journal of hepatology*, *72*(2), 250-261.
- Soh, J.-W., & Weinstein, I. B. (2003). Roles of specific isoforms of protein kinase C in the transcriptional control of cyclin D1 and related genes. *Journal of biological chemistry*, *278*(36), 34709-34716.
- Srivatanakul, P., Sriplung, H., & Deerasamee, S. (2004). Epidemiology of liver cancer: an overview. *Asian Pacific journal of cancer prevention*, *5*(2), 118-125.
- Stehr, H., Jang, S.-H. J., Duarte, J. M., Wierling, C., Lehrach, H., Lappe, M., & Lange, B. M. (2011). The structural impact of cancer-associated missense mutations in oncogenes and tumor suppressors. *Molecular cancer*, *10*(1), 1-10.
- Sutton, R. B., Davletov, B. A., Berghuis, A. M., Sudhof, T. C., & Sprang, S. R. (1995). Structure of the first C2 domain of synaptotagmin I: a novel Ca²⁺/phospholipid-binding fold. *Cell*, *80*(6), 929-938.
- Suybeng, V., Koepfel, F., Harlé, A., & Rouleau, E. (2020). Comparison of pathogenicity prediction tools on somatic variants. *The Journal of Molecular Diagnostics*, *22*(12), 1383-1392.
- Thirumal Kumar, D., & George Priya Doss, C. (2017). Role of E542 and E545 missense mutations of PIK3CA in breast cancer: a comparative computational approach. *Journal of Biomolecular Structure and Dynamics*, *35*(12), 2745-2757.
- Uchiyama, K., Saito, M., Sasaki, M., Obara, Y., Higashiyama, S., & Nakahata, N. (2009). Thromboxane A2 receptor-mediated epidermal growth factor receptor transactivation: Involvement of PKC- δ and PKC- ϵ in the shedding of epidermal growth factor receptor ligands. *European Journal of Pharmaceutical Sciences*, *38*(5), 504-511.
- Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L., & Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC bioinformatics*, *11*(1), 1-10.
- Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering, design and selection*, *8*(2), 127-134.
- Wang, W., Wang, Y., Liu, M., Zhang, Y., Yang, T., Li, D., . . . Shi, L. (2019). Betulinic acid induces apoptosis and suppresses metastasis in hepatocellular carcinoma cell lines in vitro and in vivo. *Journal of cellular and molecular medicine*, *23*(1), 586-595.

- Yang, J., Cron, P., Good, V. M., Thompson, V., Hemmings, B. A., & Barford, D. (2002). Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. *Nature structural biology*, *9*(12), 940-944.
- Yang, J., & Zhang, Y. (2015). I-TASSER server: new development for protein structure and function predictions. *Nucleic acids research*, *43*(W1), W174-W181.
- Yin, F., Feng, F., Wang, L., Wang, X., Li, Z., & Cao, Y. (2019). SREBP-1 inhibitor Betulin enhances the antitumor effect of Sorafenib on hepatocellular carcinoma via restricting cellular glycolytic activity. *Cell death & disease*, *10*(9), 1-12.
- Zhang, C., Freddolino, P. L., & Zhang, Y. (2017). COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic acids research*, *45*(W1), W291-W299.
- Zhong, Y.-W., Zhang, H.-F., Shi, Y.-M., Li, Y.-L., Chu, F., Xu, Z.-Q., . . . Gu, M.-L. (2016). IL28B SNP rs12979860 is the critical predictor for sustained viral response in Chinese children aged 1 to 6 years with chronic hepatitis C. *International journal of biological sciences*, *12*(11), 1357.