Thesis

Identification of Missense SNP of PKC ZETA Associated in HCV Induced Hepatocellular Carcinoma

A thesis submitted as final year project in partial fulfillment of the requirement for the degree of Master of Science in Healthcare Biotechnology

By

Ayesha Raza

Supervised By

Dr Yasmin Badshah

Co-supervised By

Dr. Maria Shabir

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Science and Technology (NUST) Islamabad, Pakistan

2021

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.

Ayesha Raza

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MBA/MS/MPhil Project/Thesis written by Mr. Ayesha Raza (00000319204), 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 **"**Identification of Missense SNP of PKC ZETA Associated in HCV Induced Hepatocellular Carcinoma" by Ayesha Raza 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: _____

Acknowledgements

I am very Grateful to Almighty Allah the most Beneficent and most Merciful Who gave me knowledge, wisdom, light, and courage in completion of my research work and thesis. I pay my deepest gratitude to Beloved Holy Prophet Hazrat Muhammad (S.A.W) for spreading Islam and honored us to be a part of Muslim nation.

My heartiest prayers and regards are for my supervisor Mrs. Yasmin Badshah. Her support, guidance and encouragement helped me a lot to complete my research work. She took keen interest throughout research and provided me all the facilities I needed.

I specially thank my honorable Principal ASAB, Dr Hassnain Janjua, and would like to thank all honorable teachers from the core of my heart.

Special thanks to my honorable and worthy co supervisor Dr Maria Shabir and my respected teacher Sir Naeem Mahmood, for their valuable guidance and encouragement during my course of study.

Finally, I would like to share the credit of my work with by beloved parents Mr&Mrs. Raza Muhammad (late) and my siblings Qasim Raza, Saqib Raza, Rashda Raza, Amna Raza and Naheed Raza for motivating me and supporting me at every step. I owe special thanks to my friends and lab-mates Areeba Rehman, Naila Khan, Talha Iqbal and Muhammad Irfan for encouraging and helping me in every step of my research project. I also want to thank my PhD seniors Hania Shah and Khushbakht Khan for helping me with softwares and guiding through whole research.

Lastly, I want to thank my friends, Khadija Irfan, Sundus Fahim, Sumayya Ijaz, and Maria Rehman for encouraging me throughout my master's program and believing in me.

In the end I wish good luck for everyone who has supported me in this effort.

Ayesha Raza

DEDICATED TO MY PARENTS, MR&MRS. RAZA MUHAMMAD (LATE) AND MY SIBLINGS

Table of Contents

List of Figures

List of Tables

LIST OF ABBREVIATIONS

ADMET	Absorption Distribution Metabolism Excretion and Toxicity
AFP	Alpha Fetoprotein
ALT	Alanine Transaminase
AMPK	AMP-activated Protein Kinase
aPKC	Atypical Protein Kinase C
ARMS	Amplification-refractory Mutation System
ATP	Adenosine Triphosphate
CADD	Combined Annotation Dependent Depletion
CI	Confidence Interval
cPKC	Conventional Protein Kinase C
CT	Cycle Threshold
DAG	Diacylglycerol
DNA	Deoxyribonucleic Acid
EDTA	Ethylenediaminetetraacetic Acid
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial–Mesenchymal Transition
ERK	Extracellular signal-regulated Kinase

FASTA	FAST-ALL
HCC	HEPATOCELLULAR CARCINOMA
HCV	HEPATITIS C VIRUS
HOPE	HAVE OUR PROTEIN EXPLAINED
ICCA	INTRAHEPATIC CHOLANGIOCARCINOMA
I-TASSER	ITERATIVE THREADING ASSEMBLY REFINEMENT
MAPK	MITOGEN-ACTIVATED PROTEIN KINASE
MD	MOLECULAR DYNAMICS
MM	MULTIPLE MYELOMA
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
PDK-1	PHOSPHOINOSITIDE-DEPENDENT KINASE-1
PE	PROBOLE ESTER
PKC	PROTEIN KINASE C

PLC	PHOSPHOLIPASE C
RI	RELIABILITY INDEX
RMSD	ROOT MEAN SQUARE DEVIATION
RMSF	ROOT MEAN SQUARE FLUCTUATION
RNA	RIBONUCLEIC ACID
RR	RELATIVE RISK
RTK	RECEPTOR TYROSINE KINASE
SDF	STANDARD DATA FORMAT
SDS	SODIUM DODECYL SULPHATE
SGPT	SERUM GLUTAMIC PYRUVIC TRANSAMINASE
SMILES	SIMPLIFIED MOLECULAR INPUT LINE ENTRY SYSTEM
SNP	SINGLE NUCLEOTIDE POLYMORPHISM
TAE	TRIS-ACETATE-EDTA
TE	TRIS-EDTA
TGF-B-1	TRANSFORMING GROWTH FACTOR BETA 1
UV	ULTRAVIOLET

Abstract

Cancer refers to uncontrolled division of abnormal cells and their ability to transfer to sites other than the origin. Hepatocellular carcinoma is second leading cause of death which arise from uncontrolled division of abnormal cells in liver that can spread to other parts of body. Viral induced carcinoma is widespread, and HCV is one of the potent risk factors in HCC. Single nucleotide polymorphism is widespread phenomenon resulting into mutation in single nucleotide which leads to genetic changes. There is well established fact that genetics are also involved in the occurrence of HCC so analysis of various mutations can lead to good diagnostic and prognostic marker. Protein kinase C is homologous family of protein involved in various cellular functions. Among 3 types of PKC being novel conventional and atypical PKCZ belongs to atypical PKC. This research focuses on the implication of PKCZ gene polymorphism by determining the association of PKCZ (G/A) mutation with liver cancer in Pakistan and conducting an expression analysis of the polymorphic gene. Samples were collected from liver cancer patients, followed by DNA extraction and PCR to analyze the allelic and genotypic frequency, followed by RT PCR for the expression analysis. This study also focuses on in-silico structural and functional analysis of mutant protein via HOPE an I-Mutant. The results of in vivo and in silico analysis were aligned with the fact that the heterozygous mutation in PKCZ is involved in inducing HCC. Whereas the homozygous mutation is more prone to HCV induced HCC. Molecular docking of antioncogenic molecule known as cangoronine was done which can be validated by further studies through wet lab.

Chapter 1

1. Introduction

Cancer is the second leading cause of death as reported by the World Health Organization (WHO). This disease is caused by the uncontrolled growth of cells beyond their boundaries and spreads all over the body, ultimately causing death. There are many factors contributing towards cancer including environmental, dietary and genetic changes. The genetic changes include mutations resulting in single nucleotide polymorphisms which is the variation of single nucleotide base pairs in a particular gene sequence (Gray, Campbell, and Spurr 2000). These are the most frequently occurring variations in the human genome causing various lethal diseases if present in exonic regions of the genome. Pathogenic SNPs cause damaging effects to the protein leading towards the structure dysregulation of protein or aberrant gene regulation and function. The SNPs are an important tool for the study of human genetics and genetic diseases (Riva &Kohane, 2002). There are some SNPs regarded as non-synonymous or missense SNPs which account for major deleterious effects on protein resulting in cancer (Kryukov, Pennacchio, &Sunyaev, 2007).

1.1. Liver cancer

Liver cancer is the fifth leading cause of death in men and eighth in women. It has various types depending on the tissue origin and some of them are named as, hepatocellular carcinoma, combined hepatocellular and cholangiocarcinoma, intrahepatic cholangiocarcinoma, bile duct cystadenocarcinoma, hepatoblastoma, and undifferentiated carcinoma.

The risk factors associated with all these types of cancers could be environmental exposures and hepatitis B and C viruses. Studies revealed a well-established connection between liver cancer and the infections caused by HBV and HCV (F. X. Bosch, J. Ribes, M. Díaz, & R. J. G. Cléries, 2004b). Several research studies determined that there is some association of mutated SNPs with hepatocellular carcinoma (Awan, Obaid, Ikram, & Janjua, 2017)

HCC is diagnosed using various imaging techniques such as CT scan and MRI in order to locate biomarkers of HCC. There are some common biomarkers of HCC including vascular endothelial growth factor (VEGF), alpha-1 fetoprotein, Golgi protein 73, hepatocyte growth factor, transforming growth factor-β1 (TGF-β1) and serum proteomics. In our study, we will aim to

detect a novel biomarker of pathological SNP of PRKCZ in liver cancer patients. This biomarker will serve as an important diagnostic and prognostic marker in hepatocellular carcinoma leading to targeted therapy of HCC patients.

1.2. Epidemiology

Liver cancer is the second leading cause of death, and it is estimated to have resulted in 564,000 cases per year in which rates of affecting men are twice higher than women. The incidence of liver cancer is related to the obesity pandemic and not only emerging in underdeveloped countries but developed countries like the USA. The increasing trend of liver cancer can be the aftermath of HBV and HCV virus which is likely to be decreased by emerging vaccines against HBV virus (Bosch et al., 2004b). HBV vaccines are being introduced into the infant vaccination program since 2008 as the HBV virus accounts for 60% of the total liver cancer incidence. Among all the types of liver cancers, hepatocellular carcinoma accounts for 80% of the total liver cancer worldwide. Whereas the other subtypes are rare but equally damaging (Jemal et al., 2011).

1.3. Molecular biomarkers of liver cancer

All types of liver cancers have different biological behaviors that make the diagnosis of the subtype of liver cancer difficult. Hence laboratory detection by blood biomarkers or histochemical biomarkers plays an important role in diagnosis, prognosis, and deciding the treatment plan for a specific type of liver cancer. Common serum biomarkers for liver cancer are AFP which is a glycoprotein whereas AFP-3 is heteroplast of AFP. Both markers are shown to be in higher concentration in the blood of patients with hepatocellular carcinoma. The sensitivity of AFP is not efficient, so AFP-3 is used as a supplementary test due to its increased efficacy and sensitivity towards liver cancer. Des-γ-carboxy prothrombin (DCP) is produced in HCC cells due to defects in the carboxylation of prothrombin precursors. Whereas the histochemical biomarkers include glypican 3 (GPC-3), hepatocyte parrafin 1 (Hep Par 1), heat shock protein (HSP70), glutamine synthetase (GS), arginase-1 (ARG1), cytokeratin (CK7 CK19). The emerging field of biotechnology is moving forward to find out more specific and sensitive biomarkers for the diagnosis of liver cancer so that the targeted therapy may be available readily (Gao et al., 2020).

1.4. Anti-Cancer Drug Cangoronine

This drug is a member of the triterpenoid family having a pentacyclic structure with formula C30H44O5, isolated from the bark of Tripterygium wilfordii. It has several roles in plant metabolism. Recent studies show that it is also an antiangiogenic agent(He et al., 2009). Angiogenesis is the process of reproducing new blood vessels from old blood vessels. Angiogenesis plays an important role in the progression of various diseases including tumor growth and its metastasis. The study was performed on zebrafish embryos which concluded that this drug which is a known Chinese herbal can be used against inflammatory diseases and as an inhibitor of angiogenesis(Kangli, Xiaokang, & Ying, 2000). In our research, we will study the binding of this drug with our protein PRKCZ and its potential role in HCC treatment along with its bioavailability and cytotoxicity.

1.5. Protein kinase C

Protein kinase C is a homologous group of structurally related proteins involved in signal transduction processes. They are responsible for modulating many biological functions of proteins in a reversible and rapid manner. Protein kinase C are enzymes that are involved in controlling the function of other proteins through phosphorylation of various groups present on proteins. The most common groups in these processes are hydroxyl groups of serine and threonine kinases. This protein kinase family is divided into three different subtypes based on their second messenger system. It includes classic (α, βl, βll, γ), novel (δ, ε, η, θ, and μ) and atypical (ζ and λ) isoforms (Azzi et al., 1992). In this study, we will focus on atypical isoform PKCZ and its effect after missense mutation on liver cancer.

1.5.1. Protein kinase C zeta

Protein kinase C zeta is a group of proteins belonging to serine-threonine kinases. There are 11 isoforms of protein kinase c which are classified into three main groups based on their structural and functional regulations. The three groups of isoforms formed by differential splicing are conventional, novel and atypical. These isoforms are differentiated based on the ligands which activate them for their regulation. One of these groups, the atypical protein kinases includes PKCZ. They are independent diacylglycerol, neutral lipids and calcium used for their catalytic

function. Instead, it gets activated by phosphatidylinositol 3,4,5-triphosphate (A. C. J. J. o. b. c. Newton, 1995)

Atypical PKCZ is involved in various cellular processes which include regulating proliferation, differentiation, secretion and determination of cellular polarity. PKCZ is also involved in the regulation of the cell cycle so its missense SNPs can increase the likelihood of cancer. PKCZ is seen to be at high levels in the cancerous liver as compared to the normal liver (Garg et al., 2014). The exact role of PKCZ contributing to the increased susceptibility with liver cancer is still unknown, therefore, this study will result in information regarding the role of non-synonymous or missense SNPs of PKCZ in liver cancer. There could also be a potential possibility to find therapeutic targets for liver cancer which would most likely target the main biomarker of liver cancer.

1.6. Single Nucleotide Polymorphism

Single nucleotide polymorphism is a phenomenon in which there is a difference between two DNA sequences of an individual nucleotide base. They are considered very significant as 90% of all the genetic variations in humans are occurring approximately one per 1000 bases. SNPs can either be transitions (C/T or G/A) or transversions (C/G, T/G, C/A), depending upon the type of nucleotide substitution. As nucleotide is the smallest unit of DNA, which is inherited, hence, it can be used as a molecular genetic marker for various genetic diseases. They are being used in various processes such as genetic mapping based positional cloning, linkage disequilibrium, and the role of pathological SNPs of PRKCZ gene Polymorphism in hepatocellular cancer patients. SNPs that arise in genes that are used to regulate DNA mismatch repair, cell cycle regulation, metabolism and immunity are often associated with a genetic susceptibility to cancer, so they can be used as diagnostic biomarkers in various cancer types. The study of such genetic variations can lead us to identify determinants in cancer, which could then go on to help us develop preventative and early intervention strategies. SNPs can be used to provide targeted therapy for cancer in every individual.

1.7. Objectives

- To study the association between PRKCZ polymorphism and HCV induced hepatocellular cancer in Pakistan.

- To investigate the PRKCZ gene polymorphism as a diagnostic and prognostic marker in hepatocellular cancer.
- To perform an expression analysis of PRKCZ in liver cancer patient samples.
- To predict the potential structure of protein PKCZ *in silico* and perform MD simulations.

Chapter 2

2. Literature Review

Cancer is a major health concern around the globe. It is a disease that occurs due to abnormal division of cells in any organ of the body which results in cell proliferation, organ dysfunction, ultimately leading to death. Cancerous cells divide in an uncontrollable manner and spread from one organ to another which is termed metastasis. Normal cells grow and divide according to the body's requirements, they also die as they age or get damaged. However, in the case of cancer, the checkpoints in the cell cycle fail to stop the division of cells and result in an unwanted mass of cells called tumor (Ames, Gold, & Willett, 1995). Tumors can be classified into two categories: benign and malignant. A benign tumor is a mass of cancerous cells at one place in the body which has not spread into other organs. Whereas malignant tumors are more lethal as they spread from their site of origin and invade other organs of the body (H. Y. Kim et al., 2020). Cancer is further divided into multiple types depending upon the origin or location of a specific type of cancer.

First of all, the solid tumors which arise in skin or tissues that cover the internal surface of the organs are classified as carcinomas. Common examples of carcinomas are prostate cancer, lung cancer, liver cancer, and breast cancer. Carcinomas are the most common type of cancer. Next, the type of cancer that occurs in cells or tissues which connect or support the tissues of the body is known as sarcomas. Connective tissues of the body include all those supporting tissues including bone, cartilage, joints, muscles and fat that are affected in sarcomas. Blood cancers are known as leukemias. These cancers originate in blood cells which grow in an uncontrollable manner resulting in the conversion of healthy cells into cancerous cells. This type of cancer metastasizes more rapidly than other types, as it is already present in the bloodstream. Common examples include acute lymphocytic leukemia, chronic lymphocytic leukemia and all other types of leukemias. Cancer that originates in the lymphatic system of the body is known as lymphoma. In this type of cancer, the body's immune system is highly compromised because the cells in the lymphatic system which are responsible to fight infections in the body are affected. There are two main types of lymphomas known as Hodgkin's lymphoma and non-Hodgkin's lymphoma. Lastly, multiple myeloma arises in the plasma cells which function as immune cells. This type of cancer converts healthy plasma cells into abnormal cells known as myeloma cells. Myeloma

cells accumulate in the bone marrow resulting in the spread of tumors throughout the bones of the body.

2.1. Liver cancer

According to estimates, liver cancer is the fifth most common in men and the eighth-most common malignancy in women around the world. The incidence and mortality rate of liver cancer is constantly increasing in recent years (F. X. Bosch, J. Ribes, M. Díaz, & R. Cléries, 2004a). Liver cancer includes various subtypes which are hepatocellular carcinoma, hepatoblastoma, angiosarcoma and cholangiocarcinoma. Among all of these subtypes, hepatocellular carcinoma accounts for 80% of liver cancers (McGlynn, Tsao, Hsing, Devesa, & Fraumeni Jr, 2001). There is a well-established connection between liver cancer and the chronic infections caused by hepatitis C virus and hepatitis B virus. Other recognizable risk factors can be foodstuff with aflatoxins B1, alcohol, tobacco, and oral contraceptives (Bosch et al., 2004a). HBV infection can induce genetic instability due to DNA integration which leads to liver cancer. All tumor-inducing factors which result in alteration of signaling pathways lead to activation of oncogenic signaling pathways and suppression of the immune system (Fallot, Neuveut, &Buendia, 2012). Consequently, leading to tumor formation and cancer.

2.2. Protein Kinase C Zeta

Protein kinase c is a group of proteins that belong to serine threonine kinases which play a crucial role in cell survival and death (Reyland, 2009). There are 11 isoforms of protein kinase c which are classified into three main groups based on their structural and functional regulations. The three groups of isoforms formed by differential splicing are conventional, novel and atypical. These isoforms are differentiated on the basis of the ligands which activate them for their regulation. One of these groups, the atypical protein kinases, includes PKCZ (A. C. Newton, 1995). The main difference between other isoforms and atypical PKCs is that their catalytic activity is not dependent on diacylglycerol, neutral lipids and calcium. Atypical PKCZ has been widely implicated in the regulation of cellular processes including lipid synthesis in the liver, NFkB activation and in regulating other metabolic processes (Hirai & Chida, 2003).

2.2.1. Structure

All the isoforms of protein kinase C are made of a single polypeptide chain that harbors catalytic and regulatory domains. Both domains are separated by a region known as the hinge region - the site for proteolysis. All the isoforms have the same catalytic domain with different regulatory domains. Atypical PKCZ includes only 3 catalytic domains C1, C3and C4 attached with a pseudo substrate motif along with the regulatory domain. Studies showed that the C1 domain present in atypical isoforms does not bind to or gets activated by phorbol esters and diacylglycerols (A. C. Newton, 1995). Ligand binding and activation of PKCZ do not depend on calcium, unlike the other isoforms. Instead, it gets activated by phosphatidylinositol 3,4,5-triphosphate (Deacon, Pongracz, Griffiths, & Lord, 1997).

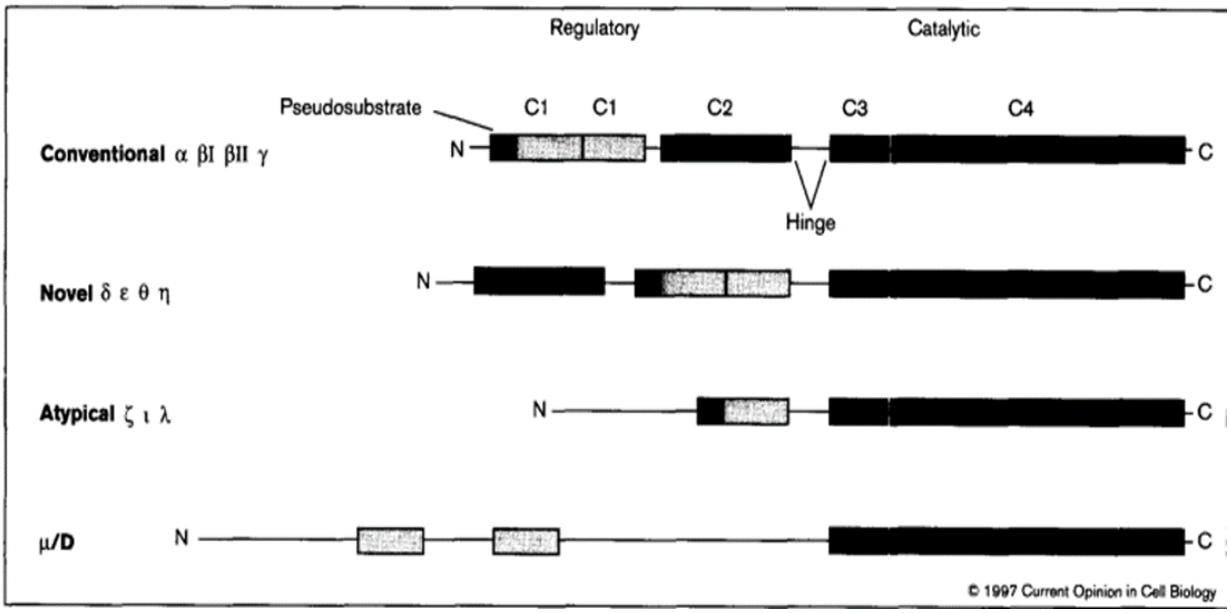


Figure 1: Schematic representation of catalytic and regulatory domains of various isoforms of Protein Kinase C

2.2.2. Role in the Cell Cycle

Atypical PKCZ is involved in various cellular processes which include regulating proliferation, differentiation, secretion and determination of cellular polarity. Primarily of all isoforms of PKCs, PKCZ is involved in the proliferation and survival of cells whereas some of the other isoforms are involved in the process of apoptosis (A. C. Newton, 1995). Activation and

inhibition of apoptosis is closely related to the expression of PKCZ. Over-expression of PKCZ results in apoptosis of leukemia cells whereas loss of PKCZ suppresses apoptosis. PKCZ is also very critical for cell survival as it is involved in the downstream processing of PI-3 kinase. It is involved in mediating the transcription of an important cell cycle regulator cyclin D1 which is involved in the progression of the cell cycle through the G1 phase to the S phase (Black & Black, 2013). Altogether, the data shows that PKCZ can be referred to as switches between cell survival and cell death (Black & Black, 2013; Reyland, 2009).

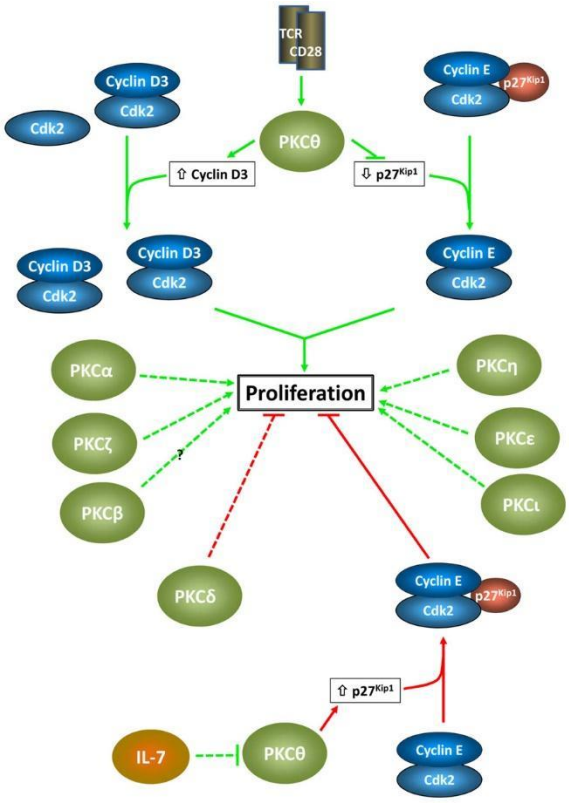


Figure 2: Effect of PKC Isozymes on Cell Cycle

2.2.3. Dysregulation in cancer

Understanding the role of the PKC family in cell survival and death allows us to perceive their involvement in mutations and various types of cancers. All isoforms interact with each other to regulate cell processes as oncogenes or tumor suppressors (Reyland, 2009). Upregulation and downregulation of PKCZ is known to be involved in the progression of various types of cancers

Ovarian cancer is one of the leading causes of death in women due to late diagnosis and prognosis. PKCZ is involved in cell survival, proliferation and cell motility which are crucial for the survival of cancer cells. Clinical studies and research have shown that the chances of survival of ovarian cancer cells increase by over-expression of PKCZ. It is involved in various cellular signaling pathways that lead to tumorigenesis (Seto&Andrulis, 2015).

PKCZ is involved in the progression and survival of breast cancer cells. It is involved in the phosphorylation of IKK-a/B which is an important component of the NFkB pathway. The continuous activation and phosphorylation lead to the survival of breast cancer cells (Hill, Win, & Acevedo-Duncan, 2008). Moreover, studies show that it is also involved in the chemical signaling pathways known as chemotaxis, in breast cancer. It is established that epidermal growth factors in cells can activate PKCZ by phosphoinositide 3- kinase. PI-3 kinase is a cofactor that acts as an important activator of PKCZ initiating and regulating the signaling pathways (Sun et al., 2005).

Pancreatic cancer is one of the most lethal cancers and is regarded as the tenth most diagnosed cancer. PKCZ is shown to be elevated in pancreatic cancer and subsequent loss of PKCZ from pancreatic cancer cells alters the phenotype of cancer cells. Knockout of PKCZ from pancreatic cancer cells has shown to decrease the mobility and anchorage of cancer cells restricting their independent growth (Butler et al., 2013).

There are multiple factors that contribute to lung cancer and its lethal outcome. Atypical PKCs are involved in the promotion of lung cancer when they are phosphorylated by a protein known as YAP1 (yes associated protein1). YAP1 is a protein that acts as a transcriptional regulator of the genes involved in cell survival and progression along with suppression of apoptosis. In lung cancer, YAP1 phosphorylates atypical PKCs leading to tumorigenesis (K.-H. Kim et al., 2019).

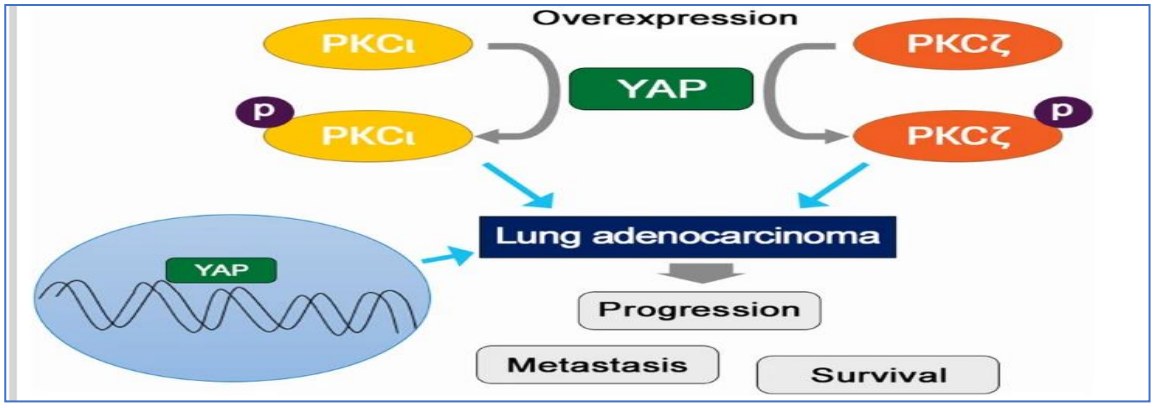


Figure 3: An illustration explaining the relationship between YAP1 and atypical PKCs in lung cancer development

In liver cancer, the membrane-bound and cytosolic fractions of isoforms of protein kinase C are shown to be altered. PKCZ is higher in cancerous cells as compared to normal liver cells. HIV B infection plays a crucial role in the initiation of liver cancer. As it releases a surface protein that activates PKC activity leading to lethal hepatocellular carcinoma (Tsai et al., 2000). Phosphorylation of heat shock protein-27 is correlated with invasion and metastasis of liver cancer (Kang, 2014).

Melanoma is a type of cancer that arises in melanocytes which are pigment-producing cells in the skin. Melanin is the pigment present in the skin for the maintenance of skin color and its protection from UV rays. PKCZ is over-expressed in melanomas leading to enhanced metastasis in cancerous cells and their survival.

PKCZ is involved in various types of cancers and its effect varies from induction to inhibition of cancers depending upon the cancer types. Studies have shown that PKCZ along with other isoforms of PKC acts as a prognostic and diagnostic marker in different cancers, which can potentially be used as therapeutic targets for cancer treatment. (Kang, 2014).

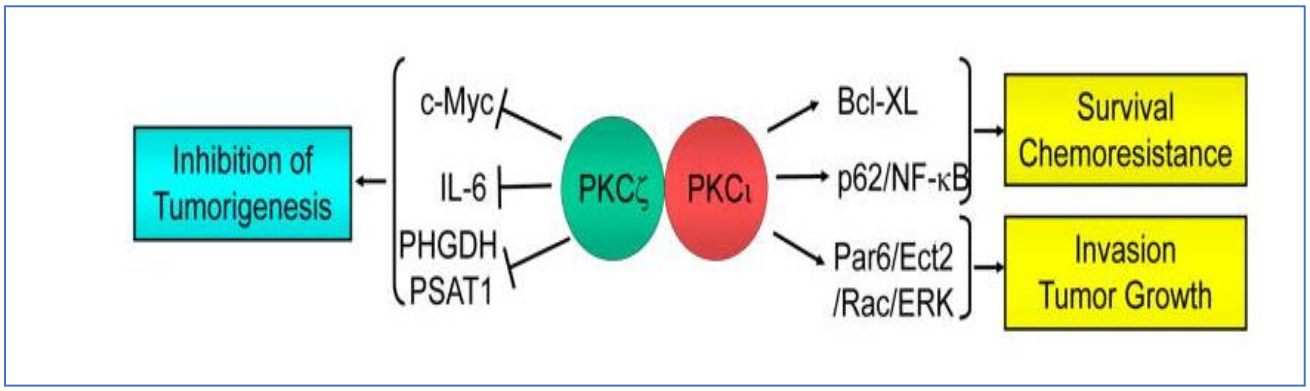


Figure 4: Atypical PKCs are involved in induction or inhibition of cancers

2.2.4. Molecular targets and Interaction of PKCZ

PKCZ is involved in various signaling pathways and interacts with molecules involved in cellular signaling. Studies have shown that PKCZ plays an important role in signaling which involves MAPK, transcription factor NFkB and ribosomal S6- protein kinase (Xu & Cong, 2005).

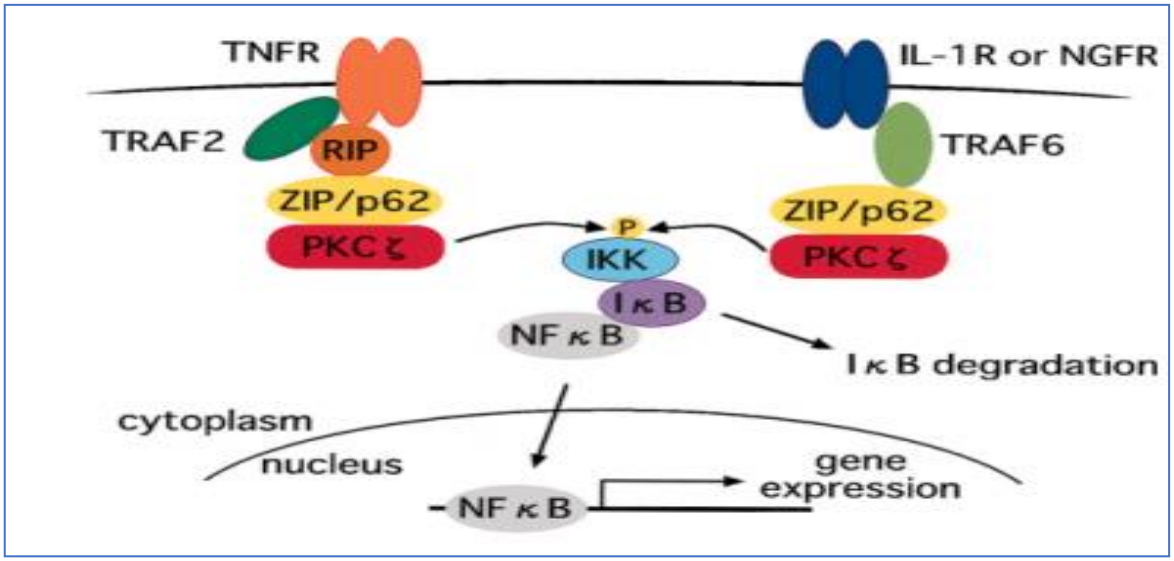


Figure 5: Schematic representation of molecular interactions of KCZ and its involvement in various signaling pathways of receptor complexes

PKCZ is also involved in the activation of epidermal growth factor (EFG) and its kinase, in both normal and cancerous cells (Cohen et al., 2006). It is involved in cell survival growth and proliferation by interacting with an adapter protein known as p62 or aPKC interaction domain. This domain is made up of short stretches of amino acid and is also present in MEPK. It serves as a potential target for PKCZ in cellular processes (Diaz-Meco&Moscat, 2001). Some other molecules involved in PKCZ activation include PIP3 and PDK1. Kinases phosphorylate the kinase domain present on PKCZ, resulting in subsequent activation of PKCZ (Xu & Cong, 2005)

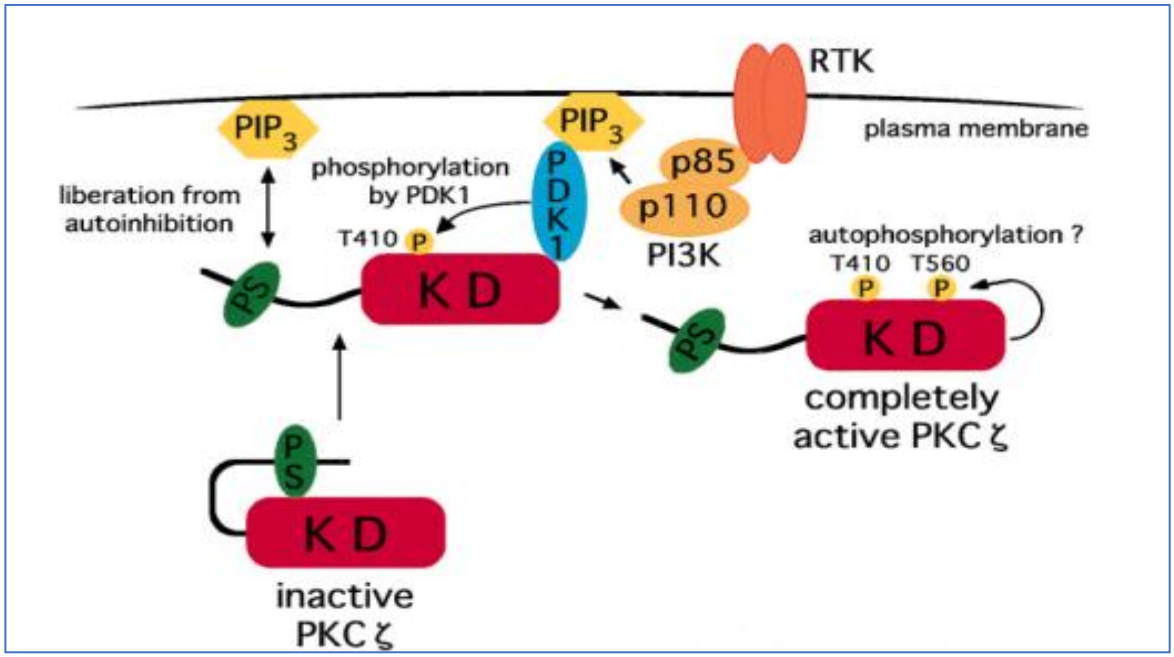


Figure 6: Molecular interactions resulting in activation of PKCZ

2.2.5. Inhibitors of PKCZ

Studies have identified various inhibitors of atypical PKCs but there are very few which bind to PKCZ. These inhibitors bind with PKCZ and interfere with the survival, growth and proliferation of cells. The two novel inhibitors 2-acetyl-1,3-cyclopentanedione (ACPD) and 3,4-diaminonaphthalene-2,7-disulfonic acid (DNDA) are shown to bind with PKCZ and block its survival activities. These inhibitors decrease the levels of phosphorylated PKCZ and induce apoptosis in melanoma cells. This approach can yield therapeutic results by inhibiting the PKCZ in various cancer cells (Ratnayake, Apostolatos, Ostrov, & Acevedo-Duncan, 2017).

2.2.6. Conclusion and Perspective

Protein kinase C is a serine threonine kinase with significant oncogenic as well as tumor-suppressing properties. Data has shown that isoforms of PKC are involved in cell survival, growth and proliferation but their loss of function, dysregulation or overexpression can result in various types of cancers. PKCZ is one of the isoforms involved in cell motility, progression and survival along with inhibition of apoptosis. This gene can serve as a potential diagnostic and prognostic marker in cancer studies.

Pathological SNPs related to PKCZ and their role in liver cancer are an emerging research interest. These gene targets can be used as therapeutic targets, as they can inhibit the growth of cancer cells by initiating apoptosis in them.

Thus, this study aims to identify pathological SNPs of PKCZ responsible for liver cancer and to specify the novel therapeutic targets for liver cancer treatments.

Chapter 3

3. Methodology

There were two phases of the study including computational analysis and in-vitro analysis. First of all, we aimed to predict the structure of PRKCZ protein along with localization of the domains present in this protein.

Following is the methodology of the *in-silico* analysis in detail:

3.1. Computational Analysis

3.1.1. Structure Prediction

The FASTA sequence of PKC-Zeta obtained from the Ensemble genome browser was submitted to an online tool I-TASSER (Yang et al., 2015) for structure prediction. I-tasser is a famous online tool used for structure predictions of target proteins which helps in further process of identification of domains and ligand binding sites in proteins. Within 24 hours of submission, I-TASSER predicted detailed 5 models for PKC-ZETA which were further analyzed in PyMol (Yuan, Chan, Filipek, & Vogel, 2016).

3.1.2. Uniqueness and Localization

To predict the localization of protein the FASTA sequence was submitted to web-based software known as DeepLoc. This software predicts the compartmentalization of protein in the cell. Whereas the uniqueness of PRKCZ was found by using another web-based software known as MegaX, that predicted the phylogenetic tree of the protein along with other family members and their uniqueness.

3.1.3. Prediction of Protein Family and Domains

After structure prediction, the FASTA format of target protein was submitted to another online tool known as InterPro (Mitchell et al., 2015). This tool is used for the prediction of protein family and the various domains that are present in the target protein

3.1.4. Identification of Pathogenic SNPs

The Ensembl genome browser predicted 358 nonsynonymous SNPs. To further analyze the pathogenicity of these SNPs 6 online tools were being used i.e., CADD (Combined Annotation Dependent Depletion) (Kircher et al., 2014), SIFT (Sorting Intolerant from Tolerant) (Ng &Henikoff, 2003), Mutation assessor (Reva, Antipin, & Sander, 2011), PolyPhen-2 (Polymorphism Phenotyping 2) (Adzhubei, Jordan, &Sunyaev, 2013), REVEL (Rare Exome Variant Ensemble Learner) (Ioannidis et al., 2016) and MetaLR (Hintzsche, Robinson, & Tan, 2016). Each of these tools categorizes the nsSNPs into various categories from highly deleterious to not deleterious or benign. These tools work according to specific algorithms which result in different groups according to their intensity of pathogenicity. The scoring algorithm of these tools is as follows:

SIFT	POLYPHEN	REVEL	metaLR	Mutation accessor	Provean
1- Deleterious	0.9 and above probably damaging	0-0.4 likely disease-causing	0-0.4 damaging	0.9 high	-2.5 and below deleterious
2- Tolerated	0.4 – 0.8 possibly damaging	0.5-0.9 likely benign	0.5 – 0.9 tolerated	0.5-0.8 medium	-2.5 above tolerated

3.1.5. SNPs analysis

The two online tools HOPE and I-Mutant were used for the analysis of pathogenic SNPs. The sequence was inserted specifying the mutation in it in both software's which send results to mail mentioned in the tool. I-mutant categorizes the wild type of SNP and mutant SNP according to reliability index (RI value) and energy values (DDG values). Whereas HOPE on the other hand specifies the structural changes occurring in the mutant protein and the possible location of the mutation in the structure with respect to various domains of the structure.

3.1.6. MD simulations

To determine the root, mean square value of deviation, function, gyration, and hydrogen bonding, molecular dynamics simulations are used. The GROMACS tool was used to perform Molecular Dynamic Studies on a 3D model saved in PDB format. GROMACS is an online program that assists with data input to supercomputers. Furthermore, Putty was also used to connect to the supercomputer and feed commands during the simulation, while WinSCP was used to move files between the supercomputer and the personal computer.

3.2. CB DOCKING

For molecular docking, in addition to the predicted structure of protein from I-tasser, the structure of cangoronine (CID 101616678) ligand is also required which was downloaded from PubChem in SDF format. Both ligand and receptor were docked using CB dock which is an online server that identifies the binding sites of given protein and calculates its centers and sizes (Liu et al., 2020). In this study CB dock generated 5 models of different interactions between the ligand and protein. Based on lowest vina score and highest cavity size we selected one model and prepared it for visualization on ligplot. Ligplot is a molecular visualization program which generates the automatic 2D schematic representation of the interaction between the ligand and the receptor protein. Ligplot enabled the visualization of the 2D structure as well as the molecular hydrogen and hydrophobic interactions of the docked structure (Wallace, Laskowski, & Thornton, 1995). The docked structure was further analyzed in 3D conformation using PyMOL.

3.3. Pharmacophore analysis

The interactions of the drug cangoronine with the target protein are studied using an online tool known as admetSAR. This tool showed results in accordance with Lipinski's rule of 5. Results are shown in tabular form below in Chapter 4 (Pollastri, 2010). According to these rules the drug can have poor absorption or poor permeability if it fulfills the four rules mentioned below:

- h-bond donors are more than 5
- h-bond acceptor is more than 10
- molecular weight of drug is more than 500
- The AlogP is more than 5

3.4. Wet lab experimentation

3.4.1. Study subjects and sample collection

50 diseased samples of liver cancer were collected from patients at the Oncology ward of Combined Military Hospital (CMH). All the patients had given their consent for sample collection. Medical history of each patient was also collected.

5ml blood was collected from each patient and stored in EDTA tubes. EDTA (Ethidium-Diamine-Tetra-Acetic acid) is a known anticoagulant which prevents clogging of blood samples and facilitates long term storage of blood samples. Blood samples were stored at 4^0C.

3.4.2. Genomic DNA extraction

DNA extraction from samples taken by HCC patients was done following the Chloroform-Phenol method (Köchl, Niederstätter, & Parson, 2005). This method takes two days. The detailed procedure for DNA extraction is explained below:

Day 1:

1. Tke750 µL of blood sample in an Eppendorf tube.
2. Add the same volume i.e., 750 µL of Solution A to the Eppendorf tube as well. Invert the tube repeatedly four to six times to mix Solution A with the blood. Solution A is used to break the cell membrane to release the DNA from the cells.
3. Keep the mixture at room temperature for 10 minutes.
4. Centrifuge the mixture at 13000 RPM for 1 minute.
5. Discard the supernatant and resuspend the pellet in 400 µL solution A.
6. Centrifuge the mixture again at 13000 RPM for 1 minute.
7. After centrifugation discard the supernatant and dissolve the pellet formed due to centrifugation in 400 µL of solution B by applying mechanical force. Solution B is responsible for the separation of proteins and the precipitation of the DNA.
8. Centrifuge the mixture again at 13000 RPM for 1 minute.
9. After centrifugation, discard the supernatant again and resuspend the pellet in 400 µL of solution B, 12 µL of 20% SDS and 5 µL of proteinase K. In this step SDS is used to

remove the lipid membranes and proteinase K for the digestion of contaminating proteins.

10. Now leave the mixture overnight for incubation at 37℃.

Day 2:

1. The next day add equal volume (400 µL) of solution C and solution B in the tube which was incubated overnight. Solution C is a Phenol that facilitates the separation of aqueous phase and organic phase in the solution clearly. Solution D is freshly prepared mixture of isoamyl alcohol and chloroform, responsible for the stabilization of the coagulated proteins and prevention of foaming in the solution to obtain pure DNA
2. Centrifuge the mixture at 13000 RPM for 10 minutes.
3. After centrifugation, the mixture will be separated in two layers. Transfer the aqueous layer in another tube and discard the remaining layer.
4. Now submerge the aqueous layer in 55 µL of 3M sodium acetate which neutralizes the charges on the phosphate backbone of the DNA and 500 µL chilled isopropanol responsible for concentrating and desalting the nucleic acids.
5. Invert the tube several times so that the DNA precipitates. Centrifuge the solution at 13000 RPM for 10 minutes.
6. After centrifugation discard the supernatant and then resuspend the DNA pellet in 200 µL of chilled ethanol. Ethanol is used for the washing of the DNA.
7. Centrifuge the mixture at 13000 RPM for 8 minutes and after centrifugation let it air dry for one hour.
8. After air drying, submerge the DNA in 200 µL of PCR water or TE buffer to dissolve the DNA in the buffer.
9. Preserve the DNA at -20℃ to prevent degradation of any component of the DNA.

The composition of solution A, B, and D is given below:

Table 1: Composition of solutions used in DNA extraction

Chemical	**Composition**
Solution A	0.32 Sucrose 10mM Tris (pH 7.5) 5mM Magnesium Chloride

	(MgCl2) Triton X-100 1% (V/V). Add triton after autoclave.
Solution B	10mM Tris (pH 7.5) 400mM Sodium Chloride (NaCl) 2mM Ethylene Diamine Tetra Acetic Acid (EDTA)
Solution C (Phenol)	As per protocol
Solution D	Chloroform: Isoamyl alcohol (24:1)

The purpose of using different chemicals is described below:

Table 2: Purpose of solutions in DNA extraction

Solution	**Purpose**
Solution A	Facilitates the breaking of cells and membranes to allow for their components, including the DNA, to be released
Solution B	Allows precipitation of the DNA and separation of the proteins
Solution C	Allows clear separation of the aqueous and the organic phases
Solution D	Stabilizes the coagulated proteins and reduce foaming to obtain pure DNA
Sodium Dodecyl Sulphate (SDS)	Removes lipid membranes
Proteinase K	Digests contaminated proteins
Sodium acetate	Neutralizes the charges on phosphate backbone of the DNA

3.4.3. Polymerase chain reaction

Polymerase chain reaction is an analytical technique used for the amplification of DNA. The process occurs in three stages and forms millions of identical copies of a single DNA strand. The three stages of this process are discussed below.

1. Separation of the double stranded DNA molecule into two single stranded DNA molecules. This step occurred at 95^0C and is known as denaturation.
2. Second step includes the binding of forward and reverse primers to their respective single stranded template DNA molecules. Optimum temperature for this step was 64^0C. This step is known as annealing.
3. The third and the final step is known as extension. During this step the DNA is amplified using Taq polymerase. The temperature for this step was 72^0C.

In order to analyze the pathogenic SNPs of PRKCZ in the DNA obtained from the blood sample of hepatocellular carcinoma patients we used Tetra-Primer Amplification Refractory Mutation System-Polymerase Chain Reaction (Tetra ARMS PCR).

Tetra ARMS PCR uses four primers in two sets. In this technique we used gradient PCR technique in which we can vary the temperature of each block making optimization of Tm (melting temperature) easy. This is an easy and economical technique used for the amplification of the target DNA. Moreover, it allows us to study SNPs in a fast and reliable way (Medrano & de Oliveira, 2014). PCR is followed by gel electrophoresis to evaluate the amplified DNA under UV light in a transilluminator.

3.4.4. PCR Protocol

The materials and their quantity used for PCR are shown in table below:

Table 3: Composition of PCR reaction mixture

Material	**Quantity**
DNTPs	2μL
MgCl2	4.5μL

Taq buffer	2μL
Taq Polymerase	0.4μL
Water	10.6μL
Template DNA	1.5μL
Internal forward primer	1μL
Internal reverse primer	1μL
Outer forward primer	1μL
Outer reverse primer	1Ml

- A reaction mixture is prepared in a falcon tube by adding all components of PCR together in the quantities described in the above table.
- Calculated amounts of DNA samples are added into multiple PCR tubes. Afterwards, a calculated amount of reaction mixture is also added into each PCR tube as well.
- Tubes are centrifuged in a micro centrifuge for efficient mixing of template DNA with the reaction mixture.
- Tubes are placed in a PCR machine which is set to run a program of optimized temperatures for the three stages of PCR for specific durations.

3.4.5. Gel electrophoresis

Gel electrophoresis is performed to separate the extracted DNA molecules according to their size. The process also acts as a confirmation of successful DNA extraction. If the DNA extraction fails, no DNA bands are shown when the gel is viewed under UV (Voytas, 2000). The protocol of gel electrophoresis is given below:

- To prepare the Agarose gel, start by heating 1g of Agar in 100 mL of 1X TAE buffer. Add 3-5 μL of ethidium bromide to the solution. 2% (w/v) agarose gel was made for viewing the PCR products. 2g of agarose is mixed in 100ml of 1X TBE or TAE buffer

- Pour the mixture in a casting tray. The casting tray should have a comb in position. The comb is used to form inundations or wells in the gel, where DNA is later loaded. Make sure that there are no bubbles formed in the tray after pouring the gel. Leave the tray until the mixture solidifies into a gel.
- After the gel has set and solidified, remove the comb gently. Gently slide the gel out of the tray and submerge it in a tank filled with 1X TAE buffer.
- Load the DNA sample mixed with 2 µL of bromophenol dye in the wells. connect the electrodes and run the gel electrophoresis at 100 V and 500 mA for 30-45 minutes.
- For loading the PCR products, 10 – 12µl of the PCR product from each tube was combined with 1 – 2µl of loading dye and then loaded in each well.
- Electrophoresis was performed at 80 V for 15 minutes in a buffer tank and bands of genomic DNA and PCR products were visualized in the UV transilluminator.

3.4.6. ALT/SGPT Test

Alanine aminotransferase (ALT) is an enzyme used to analyze liver damage. if the levels of ALT are higher that indicates any abnormality in liver function. Normal range of ALT is 7 to 55units per liter

The ALT test was performed for both the patient and the control samples to find out the mean value of ALT and its relationship with HCV induced HCC.

1. A kit-based method was used for this purpose. Blood samples from both groups i.e., patient group and control group were collected in empty tubes without anticoagulant.
2. After a few minutes the blood clotted. The clotted blood was then centrifuged at 3000 RPM for 2 minutes to extract the serum which contains ALT enzymes.
3. The kit contains two reagents R1 (Alanine tris buffer) and R2 (Alpha keto glutamate). Both the reagents were added in 4:1. 400 µl of R1 and 100 µl of R2 were taken in a separate tube and 50 µl of sample was then added to it. The step required no incubation period.
4. The ALT concentration was checked using Microlab 300 semi-automated spectrophotometer at 340 nm. The formula for concentration is:
   concentration = ΔAbs/min * kit factor

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

3.4.7. Viral RNA Extraction

Viral present in infected patient's body expressed and number of viral particles in each millimeter of blood. To examine viral load, viral RNA was extracted from patients' blood using FAVOREGEN KIT. Following steps were followed in order to extract the viral RNA.

1. 200 µl of blood was taken in an Eppendorf tube and 500 µl of VNE buffer was added. The tube was then vortexed for 5-7 seconds.
2. After the vortex, 500 µl of 75% ethanol was added to the tube. The tube was again vortexed for 5-7 seconds.
3. In the next step, a spin column was used. The sample from the Eppendorf was transferred to the spin column and centrifuged at 8000 RPM for 1 minute. The collecting tube was discarded, and RNA was retained on filter of spin column and then RNA was transferred to another collecting tube
4. 500 µl of wash buffer 1 was added to the spin column and the tube was centrifuged at 8000 RPM for 1 minute.
5. The collecting tube was changed and 750 µl of wash buffer 2 was added and centrifuged at 14000 RPM for 1 minute. This step was repeated twice to dry the filter.
6. Filter tube was transferred to an Eppendorf tube and 50 µl of RNase free water was added to the tube. The sample mixture was centrifuged again at 8000 RPM for 1 minute.
7. The RNA from the sample was transferred from the filter to the Eppendorf tube and kept at -20°C.

Chapter 4

4. Results

4.1. Structure of PKC-ZETA

The online tool named as I-TASSER was used to predict the 3-dimensional structure of PKCZ. I-TASSER predicted 5 models for this protein based on the multiple threading approach. The final selected model was Model-2 based on its C score and protein folding, according to the literature which revealed that structure with higher C score has better folding structure. The 3Dstructure was further visualized in PyMol highlighting the domain 4 present in PRKCZ. Large portion of this protein structure was in the form of an abstract loop instead of secondary or tertiary structures. The structure has binding domains for the drug understudy i.e. cangoronine.

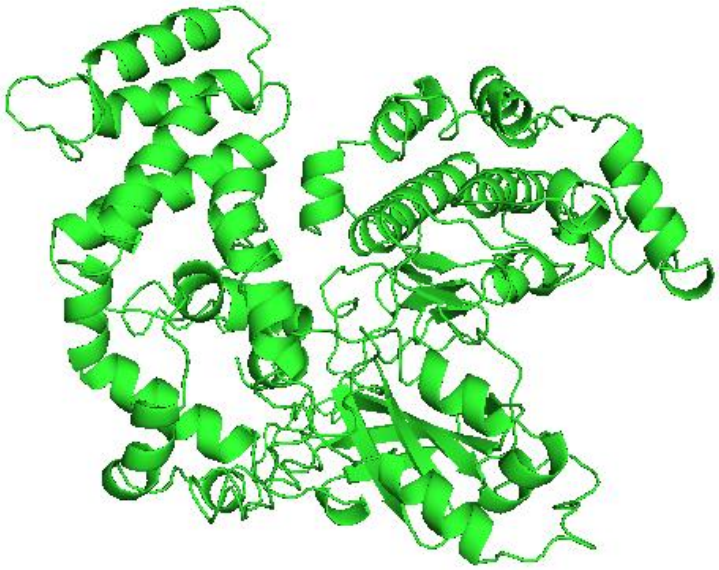


Figure 7: 3D model predicted by I-TASSER

4.2. Domains of PKC-Zeta

The 3-dimensional structure was further validated by an online tool, the InterPro. InterPro predicts the domains of proteins and categorizes them into various protein families. The domains of PKC-Zeta predicted by InterPro were PB1_ domain, PE/DAG domain, Prot_kinase_domain, and AGC-kinase domain.

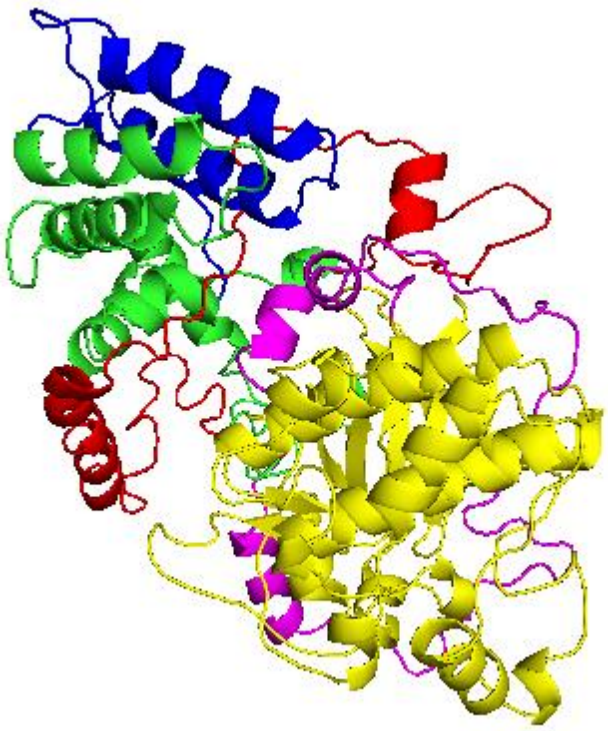


Figure 8: 3D model of PKCZ depicting 4 domains.

4.3. Phylogenetic tree and subcellular localization

The subcellular localization predicted by DeepLoc software shows that PRKCZ is mainly present in cytoplasm. The route of PKCZ localization is shown with red line in the figure 9. The score mentioned in the figure represents the likelihood or probability of the event. The phylogenetic tree predicted by MegaX shows that all family members belong to same ancestor protein. The score represents the substitution per site, mean that how much a particular protein evolute with the time from its other family members.

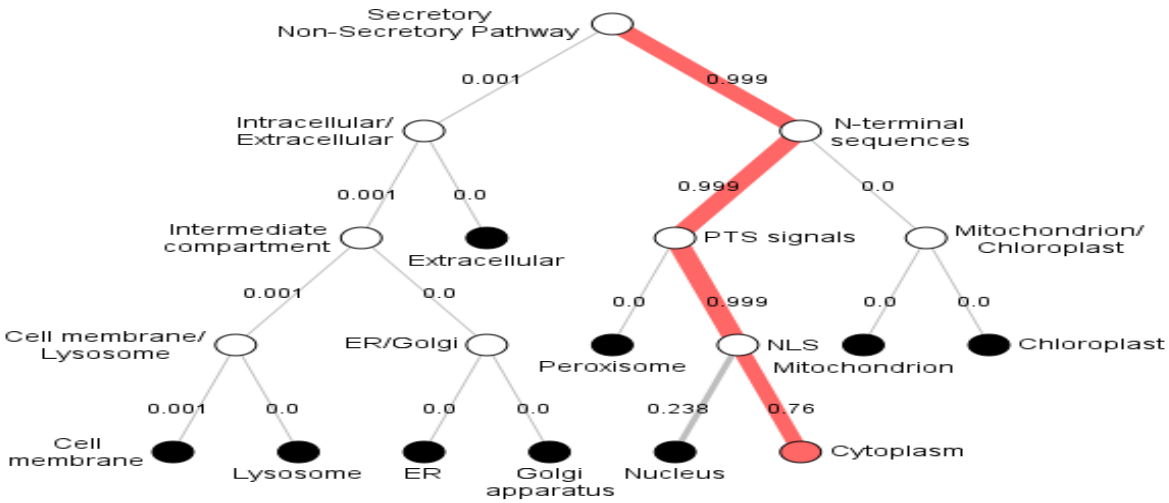


Figure 9 The red line lead to the localization point of PKCZ.

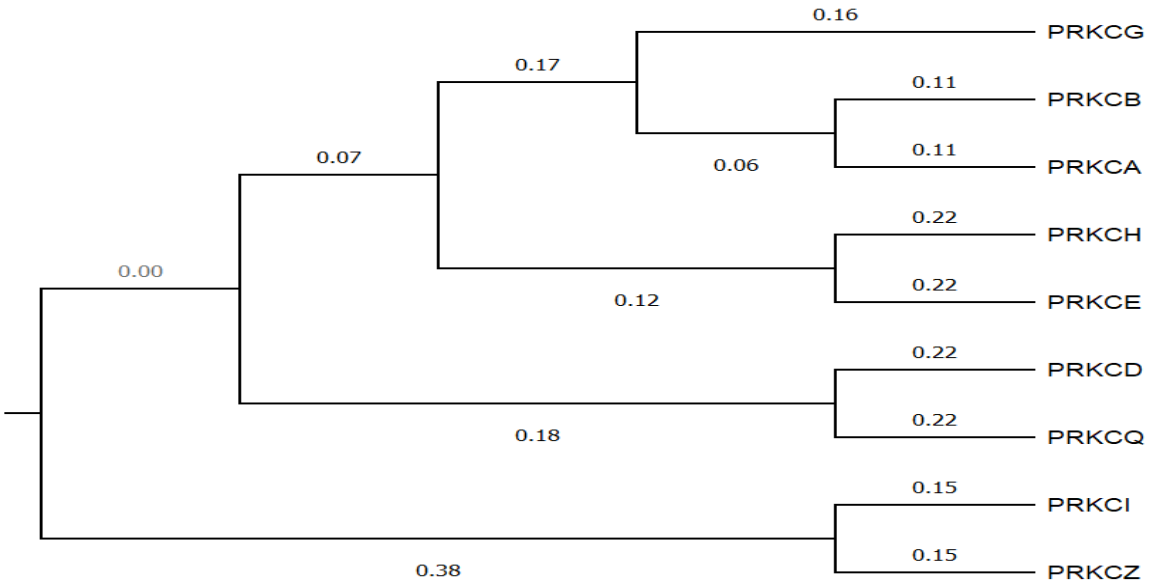


Figure 10 Represents the Phylogenetic Tree of the PKCs proteins. All the PKCs are originated from common root and then evolute into three different classes i.e., cPKCs, nPKCs and aPKCs.

4.4. Pathogenic SNPs

There was a total of 358 nonsynonymous SNPs predicted by Ensemble Genome browser that gives access to the genomes of the vertebrates at single point the nsSNPs were further analyzed by 6 online tools to classify into pathogenic and non-pathogenic SNPs. The online tools used for this purpose were SIFT, PolyPhen, CADD, REVEL, MetaLR and Mutation assessor. In the caseof PKC-Zeta only 1 SNP among 392 missense SNPs was selected for experimentation

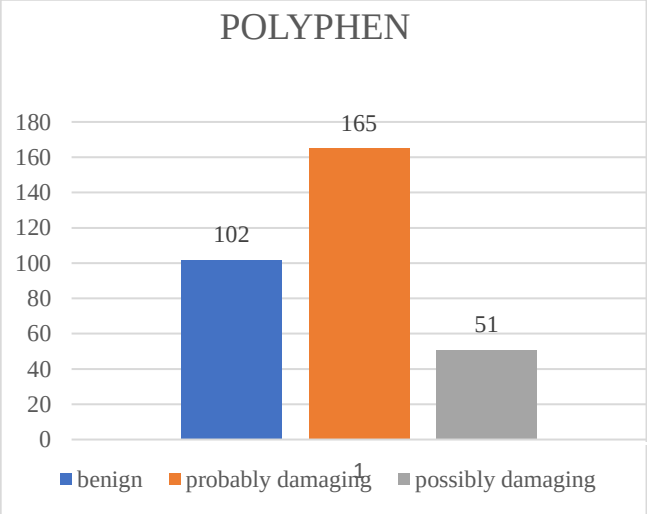


Figure 9a

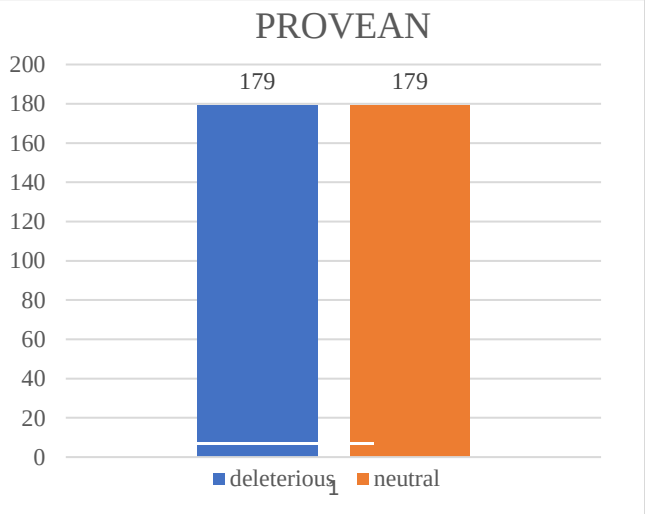


Figure 9b

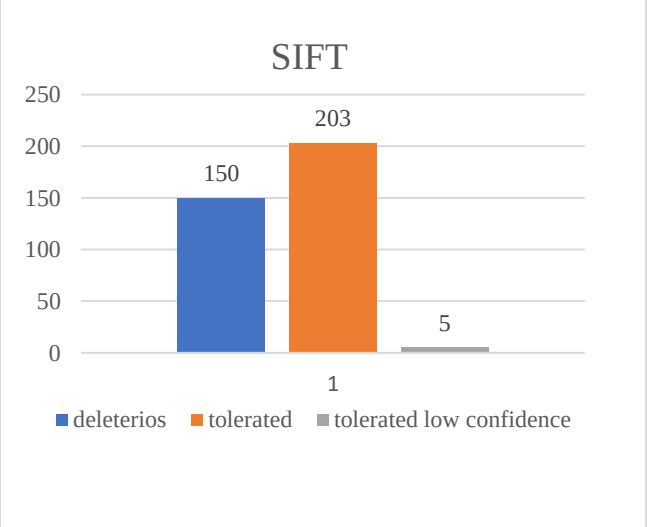


Figure 9c

Figure 11: Graphical representation of pathogenic SNPs using a: POLYPHEN, b: PROVEAN and c: SIFT

4.5. Filtration of pathogenic SNPs

According to the criteria mentioned in the methodology section 20 pathogenic SNPs were sorted on the basis SIFT and POLYPHEN score.

S.no	Variant ID	SIFT	SIFT Score	POLYPHEN Class	POLYPHEN Score
1.	rs367917640	Deleterious	0.01	Probably damaging	0.969
2.	rs781419399	Deleterious	0	Probably damaging	0.792
3.	rs1557685738	Deleterious	0	Probably damaging	1
4.	rs1557476592	Deleterious	0	Probably damaging	0.998
5.	rs1193199669	Deleterious	0.04	Benign	0.183
6.	rs750444575	Deleterious	0.01	Probably damaging	1
7.	rs1453902078	Deleterious	0.01	Probably damaging	0.995
8.	rs1242409778	Deleterious	0	Probably damaging	0.999
9.	rs1241344879	Deleterious	0.04	Probably damaging	1

10.	rs777106756	Deleterious	0.01	Probably damaging	0.99
11.	rs775001561	Deleterious	0.02	Probably damaging	0.703
12.	rs1366142734	Deleterious	0	Probably damaging	0.928
13.	rs371023344	Deleterious	0.04	Benign	0.01
14.	rs765030314	Deleterious	0.02	Probably damaging	0.935
15.	rs1457165982	Deleterious	0	Probably damaging	0.901
16.	rs1351375256	Deleterious	0	Possibly damaging	0.793
17.	rs1040957975	Deleterious	0.03	Benign	0.217
18.	rs759002098	Tolerated	0.49	Benign	0.031
19.	rs945617424	Tolerated	1	Benign	0
20.	rs972252981	Tolerated	0.1	Benign	0.107

These variants were further analyzed by another online tool known as Provean.

S. No	Variant ID	SIFT	SIFT Score	POLYPHEN Class	POLYPHEN Score	PROVEAN Score	PROVEAN class
1.	rs367917640	Deleterious	0.01	Probably damaging	0.969	-5.257	deleterious
2.	rs781419399	Deleterious	0	Probably damaging	0.792	-6.61	Deleterious
3.	rs1557685738	Deleterious	0	Probably damaging	1	-8.821	Deleterious
4.	rs1557476592	Deleterious	0	Probably damaging	0.998	-9.515	Deleterious
5.	rs1193199669	Deleterious	0.04	Benign	0.183	-3.365	Deleterious

After the 2nd phase of filtration by using multiple online tools' one variant: rs 367917640 of mutated PRKCZ, was selected for further *in silico* and wet lab experimentation.

4.6. Analysis by HOPE

Hope report highlighted the following points about the given mutation:

- Wild type protein glycine is mutated into aspartic acid.

- The mutant residue is bigger than the wild type which may lead to bumps.
- The mutant residue is more hydrophobic than the wild type.
- The mutation is located within the PB1 domain of the structure.
- Mutant residue is present near a highly conserved position which might disturb the protein.
- The flexibility of wild residue is abolished in mutant residue.
- The charge of mutant protein is changed to negative from neutral

4.7. Analysis by I-mutant

I-mutant is an online tool which predicts the change in stability of protein structures after mutation. In the given case, the RI value is 4 and its DDG value is -0.78, which clearly shows the decreased stability of the structure after induction of mutation.

4.8. MD simulations results

The MD simulations for wild type and mutant structure were effectively run for 20ns. The results are explained below:

4.8.1. RMSD

The term root means square deviation (RMSD) refers to the difference in structural changes between wild type and mutant structures over time. There isn't any significant change in the structure in this study, indicating that the mutant type does not destabilize itself after mutation, thus having no effect on the overall structure of the protein residues. Over a 20ns period, the average value of RMSD is 0-0.5nm. The RMSD value of the mutant structure is somewhat lower

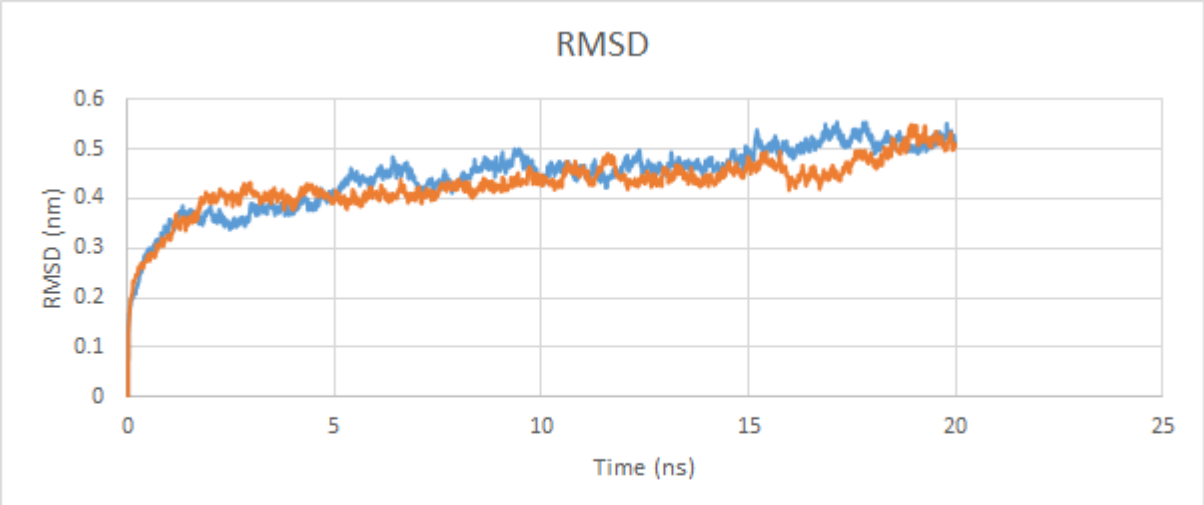


Figure 12: Graphical representation of RMSD value of wild and mutant protein

than that of the wild type, indicating that the protein is still stable, and no significant structural changes have occurred.

4.8.2. RMSF

Similar to the RMSD, the root mean square of fluctuation is plotted against wild type and mutant type to demonstrate positional differences in structure over time. The fluctuation was investigated over a 20-second period, and it was discovered that the stability of the residues, following a mutation, has been altered very little. The first peak in the graph is at the mutated area, which affected the other PE/DAG domain and protein kinase domain residues. The peaks are still smaller than 1nm in size, indicating that the structure is still stable and there are no significant changes in residues structure

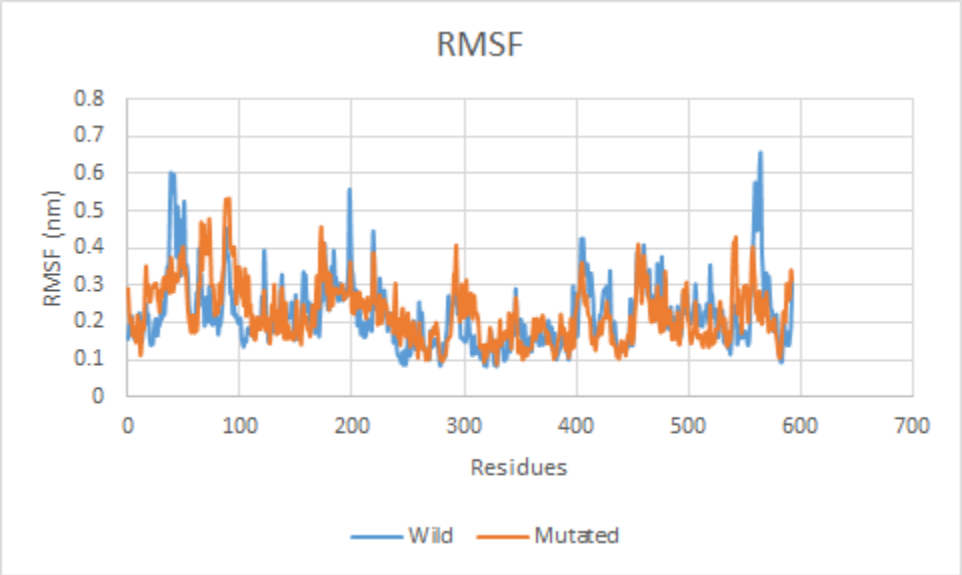


Figure 13: Graphical representation of RMSF value of wild and mutant protein

4.8.3. Hydrogen Bonding

The most important factor in maintaining the 3D and 2D structure is the hydrogen bond. It is critical that the structure maintains its hydrogen bonding following mutations so that other residues are not harmed. In this work, we're looking at a single mutation that shouldn't cause any

major changes to the protein's bonding or general structural and functional aspects Our graph

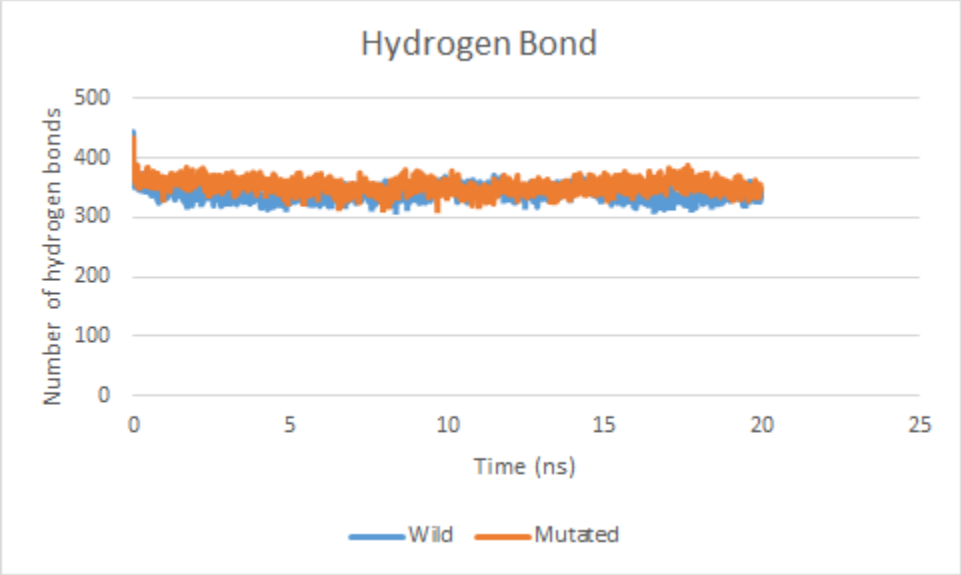


Figure 14: Graphical representation of hydrogen bonds in wild and mutant protein

shows the same results with no significant changes in hydrogen bonding.

.

4.8.4. Radius of Gyration

The radius of gyration is the distribution of atoms in a protein around its axis. It allows us to compare the compactness of the mutant and wild type structures. It is calculated by measuring the distance between the point of rotation and the maximum energy point. In this study, the radius of gyration was identical for both wild and mutant types until 5ns, after which it began to decrease, indicating that the structure became compact and more rigid. After 10 seconds, the radius begins to increase from in comparison to the wild type until 20 seconds. The change in this structure depicts that the kinetic energy of mutated structure decreases over the period of 5ns followed by increase in kinetic energy as the simulation progresses.

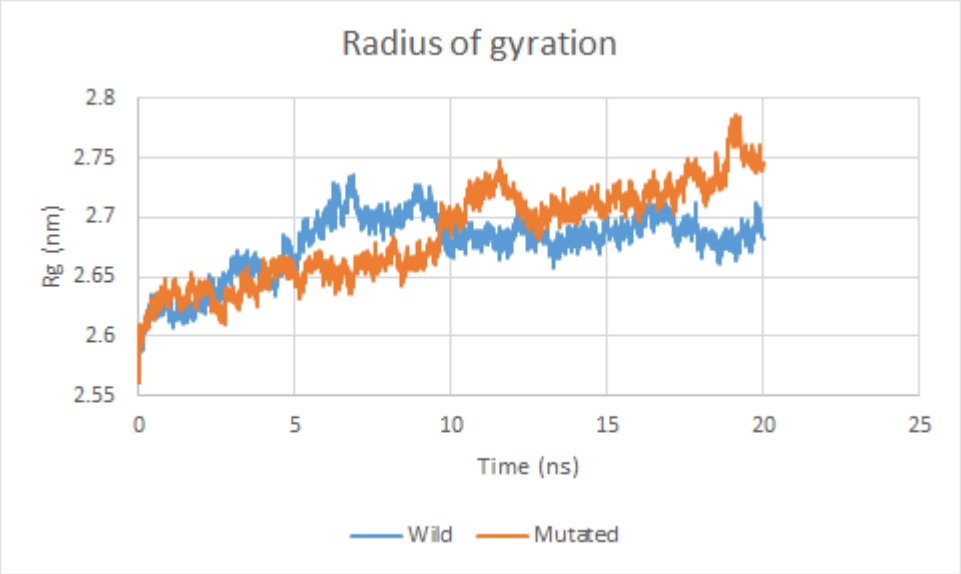


Figure 15: Graphical representation of radius of gyration for wild type and mutated protein

4.9. Drug Ligand Binding by CB DOCK

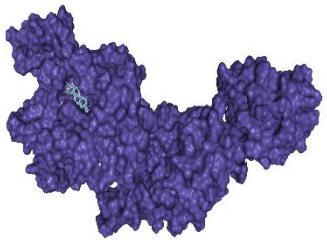
CB dock resulted in the combined structure of ligand cangoronine (CID 101616678) with its receptor protein known as PRKCZ. There were five predicted models based on the vina score and cavity sizes, from which we selected model number 1. Model number 1 had the lowest vina score and the most stabilized energy interactions.

The following table shows the cavity sizes with the respected vina scores:

Table 4: Statistical analysis by CB-DOCK

Vina score	Cavity size
-10	9045

-8.3	4085
-7.1	2007
-6.5	1000



-5.9	2485

Figure 16: 3D Structure of ligand binding drug

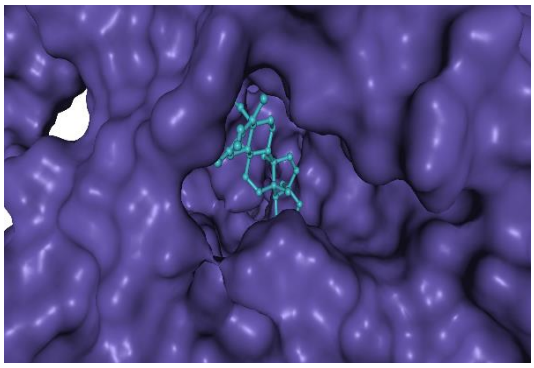


Figure 17: 3D Representation of ligand bonded with drug

Ligplot software resulted in figures specifying the molecular interactions of ligand and receptor. The model understudy has various hydrogen bonds and hydrophobic interactions. The spiked circle in the given figure shows the hydrophobic interactions of the ligand and receptor in the binding pocket ligand. The amino acids involved in these hydrophobic interactions are

glutamine, asparagine, isoleucine, phenylalanine, and alanine. Whereas there are 4 hydrogen bonds present in this complex, that are being formed between residues of arginine, tryptophan, and glutamine. The distance of these hydrogen bonds is measured in angstrom.

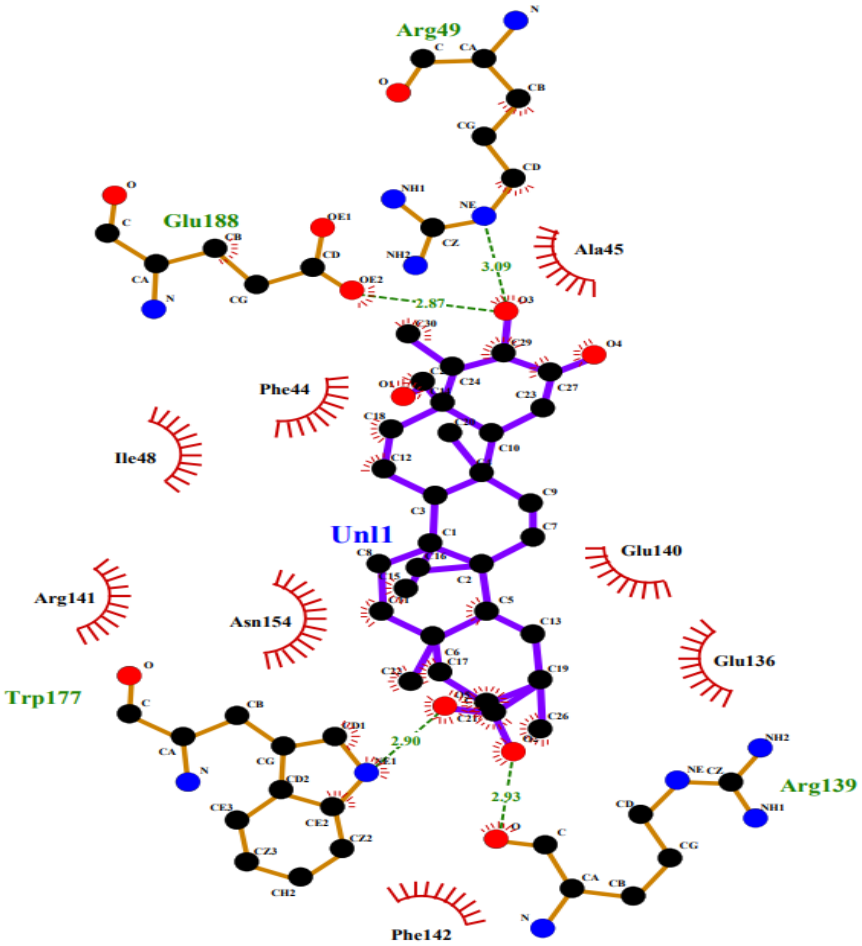


Figure 18: Protein-ligand bonding and molecular interaction. Amnio acid with hydrophobic interaction, Hydrogen bonding, covalent bonding ligand and non-ligand, distance of hydrogen bonding are shown in semi-spiked circle, green doted lines, orange, and purple lin

Further analysis of this 2D structure was carried out on PyMOL for better understanding of the 3D structure

.

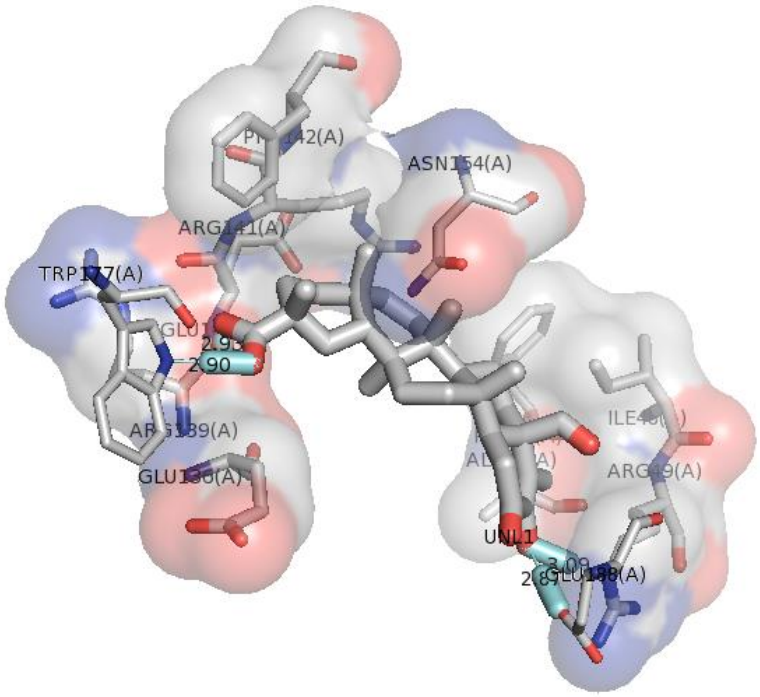


Figure 19: 3D Representation of protein ligand bonding via PyMOL

4.10. Pharmacophore

According to rules mentioned in the methodology section, the drug under study fulfills the criteria for being a molecule with good absorption and permeability. Moreover, the drug does not exhibit cancer activity or organ toxicity, making it safer for usage.

Table 5: Pharmacophore properties of cangoronine estimated by AdmetSAR

Lipinskis rule of five	
Property	Value
Molecular weight	484.68
AlogP	6.51
H-bond donor	4
H-bond acceptor	2
Rotatable bonds	2

Table 6: Physiochemical properties of cangoronine

Pharmacokinetic features	**Score**
Absorption	
Human intestinal absorption	0.9753 (+)
Caco-2	0.5698 (-)
Blood brain barrier	0.5500 (-)
Water solubility	-4.024 logS
Plasma protein binding	1.024 (100%)
Human oral bioavailability	0.5143 (-)
Metabolism	
CYP3A4 substrate	0.6382 (-)
CYP2C9 substrate	0.8038 (-)
CYP2D6 substrate	0.8038 (-)
CYP3A4 inhibition	0.7979 (-)
CYP2C9 inhibition	0.9563 (-)
CYP2C19 inhibition	0.9499 (-)
CYP2D6 inhibition	0.9443 (-)
Excretion	
UGT catalyzed	0.9000 (+)

Table 7: Toxic profile of drug

Toxicity	**Score/ probability**
Carcinogenicity	0.8663 (-)
Hepatotoxicity	0.7500 (-)
Acute oral toxicity	3.238 mol/kg
Eye irritation	0.9397 (-)
Eye corrosion	0.9948 (-)
Human either-a-go-go inhibition	0.6556 (+)

biodegradation	0.8500 (-)

4.11. Wet lab analysis of pathogenic SNP

The genotype data comparison between patient and control is shown in table 1. According to this data the p-value of genotype GG and AG is less than 0.05 which shows that both SNPs are significant. Whereas, according to the relative risk and odds ratio the values more than 1 are involved in increasing the risk factor of disease and the genotypes with p-values less than 1 contribute to the protective effect. In this study genotype AG is disease causing whereas GG is shown to be a protective genotype.

Table 8: Genotypic data of patient and control

Total Patient and Control Data							
Genotype	**Frequency Distribution %**		**Odds Ratio**		**Relative Risk**		**P value**
	Patients	**Control**	**Value**	**95% CI**	**Value**	**95% CI**	
AA	27.00%	32.00%	0.7860	0.4348 to 1.471	0.8839	0.6291 to 1.194	0.5353
GG	10.00%	40.00%	0.1667	0.08146 to 0.3491	0.3333	0.1849 to 0.5616	<0.0001
AG	64.00%	28.00%	4.571	2.455 to 8.197	2.087	1.562 to 2.840	<0.0001

Statistical analysis of data based on gender shows that the mutation with genotype AG is a contributing factor in increased risk of HCV induced HCC. Both males and females with genotype AG are more prone to cancer as compared to patients with GG and AG genotype,

which has a relative risk and odds ratio less than 1 and is shown to have a protective role. Table 8 shows the gender based statistical analysis.

Table 9: Statistical analysis of PCR data with respect to gender

Patient and Control Data with respect to Gender							
Genotype	**Frequency Distribution %**		**Odds Ratio**		**Relative Risk**		P value
	Patients	**Control**	**Value**	**95% CI**	**Value**	**95% CI**	
AA (M)	28.00%	34.78%	0.7292	0.3104 to 1.758	0.8556	0.5305 to 1.284	0.5144
GG (M)	16.00%	39.13%	0.2963	0.1126 to 0.8034	0.5128	0.2684 to 0.8752	0.0126
AG (M)	56.00%	26.09%	3.606	1.566 to 8.302	1.782	1.220 to 2.648	0.0038
AA (F)	26.00%	29.63%	0.8345	0.3650 to 1.984	0.9087	0.5500 to 1.390	0.8271
GG (F)	4.00%	40.74%	0.06061	0.01363 to 0.2387	0.1389	0.03824 to 0.4385	<0.0001
AG (F)	70.00%	29.63%	5.542	2.344 to 12.09	2.425	1.560 to 3.938	<0.0001

Lastly, the statistical analysis was applied on the data based on the age of the patients. The results of analysis showed that patients in the age group of 1-19yrs are not significantly susceptible to HCC regardless of the mutation in genotype. However, as age increases the risk factor according to P value, odds ratio and relative risk also increase in patients having AG

genotype mutations. Whereas the genotype GG has shown to have a protective role against disease. Table 9 shows the statistical analysis based on various age groups.

Table 10: Statistical analysis of PCR data with respect to age

Patient and Control Data with respect to Age							
Age group-Genotype	**Frequency Distribution %**		**Odds Ratio**		**Relative Risk**		**P value**
	Patients	**Control**	**Value**	**95% CI**	**Value**	**95% CI**	
1-19 AA	0.00%	100.00%	0.000	0.000 to 9.000	0.000	0.000 to 1.921	>0.9999
1-19 GG	100.00%	0.00%	Infinity	0.1111 to Infinity	Infinity	0.5206 to Infinity	>0.9999
1-19 AG	0.00%	0.00%	__	__	0.000	0.000 to 1.000	>0.9999
20-39 AA	25.00%	27.91%	0.8611	0.3391 to 2.184	0.9306	0.5610 to 1.404	0.8143
20-39 GG	16.67%	38.64%	0.3176	0.1184 to 0.7965	0.5360	0.2808 to 0.9114	0.0209
20-39 AG	58.33%	34.09%	2.707	1.202 to 6.142	1.595	1.077 to 2.421	0.0231
40-59 AA	30.61%	31.11%	0.9769	0.3921 to 2.299	0.9888	0.6257 to 1.464	>0.9999
40-59 GG	2.04%	39.58%	0.03180	0.002993	0.08021	0.01417	<0.0001

				to 0.2100		to 0.3835	
40-59 AG	67.35%	27.08%	5.553	2.211 to 13.04	2.287	1.501 to 3.636	<0.0001
60+ AA	0.00%	42.86%	0.000	0.000 to 4.348	0.000	0.000 to 2.600	0.5000
60+ GG	0.00%	57.14%	0.000	0.000 to 2.550	0.000	0.000 to 1.758	0.4444
60+ AG	100.00%	0.00%	Infinity	2.492 to Infinity	Infinity	2.000 to Infinity	0.0278

4.11.1. Comparison between average secretion of ALT among patients and control

The comparison of enzymatic secretion among patient and control group showed that ALT is secreted more in patients. The average secretion of ALT in our study on 100 control groups and patients is 37.5 106.84 respectively. There is a significant increase of ALT in patients with HCV induced HCC. The standard deviation value for control group is ±10.16567 and for patients it is ±52.44655

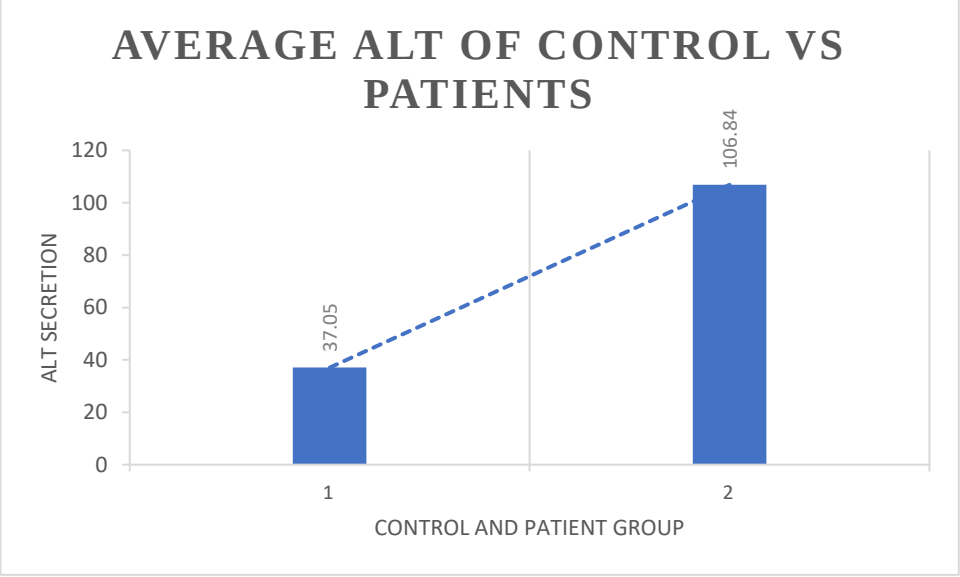


Figure 20: Graphical representation of average ALT among patient and control group

4.11.2. Viral load of 3 genotypes

Viral load is the amount of virus present in an infected patient. In this study, the viral load for HCV in HCC patients is highest for genotype GG. The mean value for AA, GG, and AG is 368507608.8, 582359752.5 and 353832156.6 respectively. Whereas the standard deviation of these genotypes is 1147890187, 1304868648 and 1126780682.

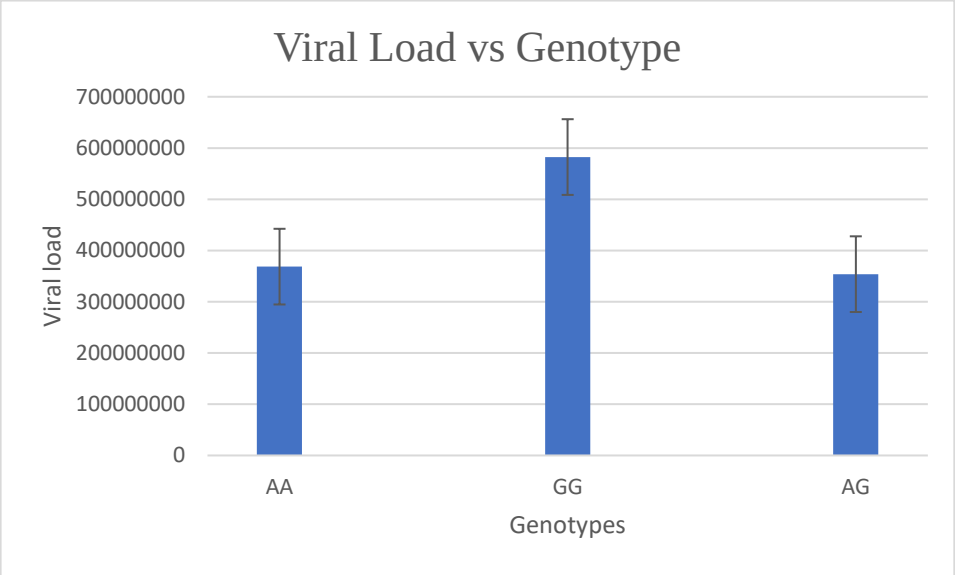


Figure 21: Graphical representation of viral load with respect to 3 genotypes

CHAPTER 5

5. Discussion

There is extensive literature about protein kinase C and its isoforms, including their role in various diseases, specifically cancer. HCC diagnosis and its prognosis is explored in detail, however the role of single nucleotide polymorphism in PRKCZ is not elucidated. We aimed to find out the association of PRKCZ gene polymorphism in liver cancer patients through wet lab analysis along with drug conjugation and structural analysis *in silico.*

The gene PRKCZ is located on chromosome number 1 and the resultant protein has 592 amino acids. The structure of PRKCZ and its domains were aligned with the previous research(Bignell et al., 2010). The atypical family of PKC has shown similar domains and structure which was confirmed through PyMOL and I-Tasser. I-Tasser is known to determine protein structures based on c-scores and other algorithms, whereas PyMOL is used for visualization of structure and its respective domain(Yang et al., 2015). All the family members contain calcium sensitive domain C2 but atypical PKCs have C1 domain instead, which binds with phosphatidylinositol 4,5,6 triphosphate. There were a total 358 nonsynonymous SNPs retrieved from the ensemble genome browser, and we studied one novel SNP rs367917640 for its pathogenic or protective role in liver cancer. The mutation discussed in this study at position 65 from arginine to guanine lies in the PB1 domain. Previously, it has been reported that PRKCZ is involved in various steps of cellular transformation, the activation of EFG induced mitogen activated protein kinase and stimulation for basic fibroblast growth factor. EFGR is present in squamous head and neck carcinoma and studies revealed that malignant tissues have significantly increased expression of PRKZ(Cohen et al., 2006). Moreover, it is involved in tumor progression and activation of MAPK which helps in the proliferation of tumor cells. It is also reported that expression of PRKCZ is increased by two folds in glioblastoma cell lines. Previous literature has shown that PRKCZ is hypermethylated and transcriptionally downregulated in HCC. PRKCZ is also believed to be a tumor suppressor as the knockdown of PRKCZ is shown to result in resistance against tumor treatments and inhibiting oncogenic signaling pathways(Deevi et al., 2018). There are no significant studies elucidating the relation of PRKCZ with HCC. However, in my study, I validate this finding by performing various tests and bioinformatic tools. As mentioned in the

results, there were three genotypes involved in HCC which contribute through their mutated version. The genotype AA was confirmed to be nonpathogenic according to its P-value, relative risk and odd ratio. Whereas the other two genotypes GG and AG could be involved in inducing HCC.

The SNP analysis of PKCZ was being done by an ensemble genome browser which categorizes various mutations and variants based on different online tools. The data collected was similar to online results of pathogenic SNPs present in PRKCZ. All of the isoforms of PKC have numerous variants having missense SNPs contributing to various genetic mutations. 20 missense SNP retrieved from ensemble genome browser out of 358 and 1 SNP was selected for further analysis.

The structural and functional analysis of mutant PRKCZ was performed using I-mutant and HOPE. Both of these tools are widely being used for structural modification in mutant proteins, as reported in previous studies that the p63 protein and human hemochromatosis protein were studied using these tools. Results of HOPE suggested that the size of the mutant protein is bigger than the wild type which may cause alteration in stability of protein structure and disruption in its function. Moreover, it is also stated that the mutation is in a highly conserved region which not only destabilizes the residue, but also the entire protein as well as diminishing the flexibility of the wild type protein structure for binding of various drugs. On the other hand, I-mutant predicts changes in mutated protein based on energy values and reliability values which are 4 and -0.78 respectively. These values clearly lead to the decreased stability of the protein structure after induction of mutation.

Further analysis of mutant PRKCZ was performed through wet lab experimentation which involved DNA extraction of patients with HCC followed by tetra ARMS-PCR. Two sets of primers, including 2 outer and 1 inner primer, were designed using Primer1 to study association between single nucleotide polymorphism in PRKCZ and hepatocellular carcinoma. The PCR results showed that the mutation AG has strong relation with HCC as compared to AA and GG. Whereas the genotype GG is more prone to viral induced HCC. The study was also conducted based on age and gender that revealed that both males and females are equally susceptible to HCC induced by this mutation and patients aged greater than 20 are at a higher risk of disease. However, the studies based on age remain doubtful due to smaller sample size. Previous studies

on HCC revealed that HCC is more common in females as compared to males of all ages(Lee et al., 1999).

Wet lab analysis also included the role of hepatitis c virus in inducing HCC and its correlation with our respective gene mutations. For viral load analysis, RNA was extracted using reverse transcriptase method and then analyzed via real time PCR. The results of viral load suggested that the homozygous mutation GG is at a high risk of inducing HCV induced carcinoma as compared to wild type and AG mutation. Literature stated the fact that HCV plays a significant role in inducing HCC with alterations in human genome polymorphism, metabolic and growth signaling systems(Teo et al., 2021). Our study revealed that there is correlation of PRKCZ mutation and HCV induced HCC as both alter the signaling of growth and proliferation of cells. Moreover, another contributing factor increasing the risk of HCC is serum alanine aminotransferase levels in patients as compared to control groups. The risk of HCV induced HCC is seen to be increased by previous studies(Goossens& Hoshida, 2015).

The drug selected for this study was cangoronine, belonging to the triterpenoids family involved in plant metabolism. It is known to be used as an angiogenic agent in various diseases which also includes cancer(He et al., 2009). Binding of cangoronine with our subject protein using CB DOCK for docking analysis is shown to be stable and it can be used as potential treatment for disease. The stability validation of this drug was performed via an online tool known as admetSAR. It revealed that the drug is potation anticarcinogenic with high bioavailability and lower toxicity. However, there was not enough data elucidating the role of drugs previously.

Finally, MD simulations were performed on both wild type and mutant protein to predict molecular dynamics. It is widely used in molecular studies of various protein variants. Recently binding affinity of COV-19 with noscapines is being studied via MD simulations at different temperatures(Keretsu, Bhujbal, & Cho, 2020). In our study the values of RMSD and RMSF shows that there is slight change in the values between wild type and mutant type which shows there is less change in structure of mutated protein, also there is very little effect on other residues of protein by this mutation. Hydrogen bonding remains almost similar in both cases as one single mutation does not affect the bonding at significant levels. Lastly the radius of gyration was being studied in which the mutated protein is constantly changing its structure ranging from

compact and rigid to wide and more flexible than wild type. Both alterations make the protein less stable and less suitable for binding with drugs.

6. Conclusion

Several studies show that cancer arises from malignancy in one cell or its cell cycle which leads to uncontrolled division and then cancer. Single nucleotide polymorphism is widely being studied for mutations in various diseases. Therefore, scientists are working hard to achieve more success in finding new pathogenic SNPs to find cure through targeted therapies. In this study we have focused on the pathogenic SNP of PRKCZ which is contributing to HCV induced HCC. We selected rs367917640 from the pool of 392 missense SNPs for its correlation with HCC and HCV induced HCC. The structure prediction via I-tasser revealed the most appropriate structure of our respective protein based on the c-score which is -1.95 and has the best folding structure. CB dock was used for the binding of a relatively novel drug for cancer known as cangoronine with PRKCZ based on vina score; the appropriate ligand-drug molecule was selected for its further analysis. This drug binds with protein structure without altering its stability or inducing any cytotoxicity as studied using an online admetSAR revealing its various pharmacophore properties including 3.238 kg/mol acute oral toxicity. There were three genotypes under study including AA, GG, AG and after validating from PCR results and in silico analysis it was shown that AG is the most pathogenic SNP for inducing HCC. Moreover, the studies of HCC based on gender with respect to specific mutation under study have shown that both males and females are equally susceptible to the disease. Moreover, there was an interesting finding that the mutation with GG genotype has shown increased viral load which could possibly lead to HCV induced HCC.MD simulates were being run on both wild and mutated type which shows there is slightly significant changes in the structure after mutation with respect to the values of RMSD, RMSF hydrogen bonding and radius of gyration. However, there is much work needed towards targeted therapy of HCC induced by PKCZ mutation as it is also shown to be a tumor suppressor in some cases which can be very useful in therapies. Let’s hope that the miracles of science continue to grow and the dreadful diseases like cancer can be cured 100.

7. References

Adzhubei, I., Jordan, D. M., &Sunyaev, S. R. J. C. p. i. h. g. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *76*(1), 7.20. 21-27.20. 41.

Ames, B. N., Gold, L. S., & Willett, W. C. (1995). The causes and prevention of cancer. *Proceedings of the National Academy of Sciences, 92*(12), 5258-5265.

Awan, F. M., Obaid, A., Ikram, A., & Janjua, H. A. J. I. j. o. m. s. (2017). Mutation-structure-function relationship based integrated strategy reveals the potential impact of deleterious missense mutations in autophagy related proteins on hepatocellular carcinoma (HCC): A comprehensive informatics approach. *18*(1), 139.

Azzi, A., Boscoboinik, D., &Hensey, C. (1992). The protein kinase C family. *European Journal of Biochemistry, 208*(3), 547-557.

Black, A. R., & Black, J. D. (2013). Protein kinase C signaling and cell cycle regulation. *Frontiers in immunology, 3*, 423.

Bosch, F. X., Ribes, J., Díaz, M., &Cléries, R. (2004a). Primary liver cancer: worldwide incidence and trends. *Gastroenterology, 127*(5), S5-S16.

Bosch, F. X., Ribes, J., Díaz, M., &Cléries, R. J. G. (2004b). Primary liver cancer: worldwide incidence and trends. *127*(5), S5-S16.

Butler, A. M., Buzhardt, M. L. S., Li, S., Smith, K. E., Fields, A. P., & Murray, N. R. (2013). Protein kinase C zeta regulates human pancreatic cancer cell transformed growth and invasion through a STAT3-dependent mechanism. *PloS one, 8*(8), e72061.

Cohen, E. E. W., Lingen, M. W., Zhu, B., Zhu, H., Straza, M. W., Pierce, C., . . . Rosner, M. R. (2006). Protein kinase Cζ mediates epidermal growth factor–induced growth of head and neck tumor cells by regulating mitogen-activated protein kinase. *Cancer research, 66*(12), 6296-6303.

Deacon, E., Pongracz, J., Griffiths, G., & Lord, J. (1997). Isoenzymes of protein kinase C: differential involvement in apoptosis and pathogenesis. *Molecular Pathology, 50*(3), 124.

Diaz-Meco, M. a. T., &Moscat, J. (2001). MEK5, a new target of the atypical protein kinase C isoforms in mitogenic signaling. *Molecular and cellular biology, 21*(4), 1218-1227.

Fallot, G., Neuveut, C., &Buendia, M.-A. (2012). Diverse roles of hepatitis B virus in liver cancer. *Current opinion in virology, 2*(4), 467-473.

Gao, Y. X., Yang, T. W., Yin, J. M., Yang, P. X., Kou, B. X., Chai, M. Y., . . . Chen, D. X. (2020). Progress and prospects of biomarkers in primary liver cancer. *International journal of oncology, 57*(1), 54-66.

Garg, R., Benedetti, L. G., Abera, M. B., Wang, H., Abba, M., &Kazanietz, M. G. J. O. (2014). Protein kinase C and cancer: what we know and what we do not. *33*(45), 5225-5237.

Gray, I. C., Campbell, D. A., &Spurr, N. K. J. H. m. g. (2000). Single nucleotide polymorphisms as tools in human genetics. *9*(16), 2403-2408.

Hill, D., Win, H., & Acevedo-Duncan, M. (2008). PKC-zeta associates and phosphorylates IKK-α/β in breast cancer cells. In: AACR.

Hintzsche, J. D., Robinson, W. A., & Tan, A. C. J. I. j. o. g. (2016). A survey of computational tools to analyze and interpret whole exome sequencing data. *2016*.

Hirai, T., & Chida, K. (2003). Protein kinase Cζ (PKCζ): activation mechanisms and cellular functions. *Journal of biochemistry, 133*(1), 1-7.

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., . . . Karyadi, D. J. T. A. J. o. H. G. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *99*(4), 877-885.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: a cancer journal for clinicians, 61*(2), 69-90.

Kang, J.-H. (2014). Protein kinase C (PKC) isozymes and cancer. *New Journal of Science, 2014*.

Kim, H. Y., Jung, E. K., Lee, D. H., Yoon, T. M., Lee, J. K., & Lim, S. C. (2020). Clinical difference between benign and malignant tumors of the hard palate. *European Archives of Oto-Rhino-Laryngology, 277*(3), 903-907.

Kim, K.-H., Chung, C., Kim, J.-M., Lee, D., Cho, S. Y., Lee, T. H., . . . Yeo, M.-K. (2019). Clinical significance of atypical protein kinase C (PKCι and PKCζ) and its relationship with yes-associated protein in lung adenocarcinoma. *BMC cancer, 19*(1), 804.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., &Shendure, J. J. N. g. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *46*(3), 310-315.

Köchl, S., Niederstätter, H., & Parson, W. (2005). DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. In *Forensic DNA typing protocols* (pp. 13-29): Springer.

Kryukov, G. V., Pennacchio, L. A., &Sunyaev, S. R. J. T. A. J. o. H. G. (2007). Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *80*(4), 727-739.

McGlynn, K. A., Tsao, L., Hsing, A. W., Devesa, S. S., & Fraumeni Jr, J. F. (2001). International trends and patterns of primary liver cancer. *International Journal of Cancer, 94*(2), 290-296.

Medrano, R. F. V., & de Oliveira, C. A. (2014). Guidelines for the tetra-primer ARMS–PCR technique development. *Molecular biotechnology, 56*(7), 599-608.

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., . . . Pesseat, S. J. N. a. r. (2015). The InterPro protein family's database: the classification resource after 15 years. *43*(D1), D213-D221.

Newton, A. C. (1995). Protein kinase C: structure, function, and regulation. *Journal of biological chemistry, 270*(48), 28495-28498.

Newton, A. C. J. J. o. b. c. (1995). Protein kinase C: structure, function, and regulation. *270*, 28495-28498.

Ng, P. C., &Henikoff, S. J. N. a. r. (2003). SIFT: Predicting amino acid changes that affect protein function. *31*(13), 3812-3814.

Ratnayake, W. S., Apostolatos, A. H., Ostrov, D. A., & Acevedo-Duncan, M. (2017). Two novel atypical PKC inhibitors; ACPD and DNDA effectively mitigate cell proliferation and epithelial to mesenchymal transition of metastatic melanoma while inducing apoptosis. *International Journal of Oncology, 51*(5), 1370-1382.

Reva, B., Antipin, Y., & Sander, C. J. N. a. r. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *39*(17), e118-e118.

Reyland, M. E. (2009). Protein kinase C isoforms: multi-functional regulators of cell life and death. *Frontiers in bioscience (Landmark edition), 14*, 2386.

Riva, A., &Kohane, I. S. J. B. (2002). SNPper: retrieval and analysis of human SNPs. *18*(12), 1681-1685.

Seto, K. K., &Andrulis, I. L. (2015). Atypical protein kinase C zeta: potential player in cell survival and cell migration of ovarian cancer. *PloS one, 10*(4), e0123528.

Sun, R., Gao, P., Chen, L., Ma, D., Wang, J., Oppenheim, J. J., & Zhang, N. (2005). Protein kinase C ζ is required for epidermal growth factor–induced chemotaxis of human breast cancer cells. *Cancer research, 65*(4), 1433-1441.

Tsai, J.-H., Hsieh, Y.-S., Kuo, S.-J., Chen, S.-T., Yu, S.-Y., Huang, C.-Y., . . . Liu, J.-Y. (2000). Alteration in the expression of protein kinase C isoforms in human hepatocellular carcinoma. *Cancer letters, 161*(2), 171-175.

Voytas, D. (2000). Agarose gel electrophoresis. *Current protocols in molecular biology, 51*(1), 2.5 A. 1-2.5 A. 9.

Xu, S.-J., & Cong, B. (2005). Activation mechanisms of protein kinase Czeta and its cellular function. *Sheng li kexuejinzhan [Progress in physiology], 36*(4), 309.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. J. N. m. (2015). The I-TASSER Suite: protein structure and function prediction. *12*(1), 7-8.

Yuan, S., Chan, H. S., Filipek, S., & Vogel, H. J. S. (2016). PyMOL and Inkscape bridge the data and the data visualization. *24*(12), 2041-2042.

Bignell, G. R., Greenman, C. D., Davies, H., Butler, A. P., Edkins, S., Andrews, J. M., . . . Latimer, C. (2010). Signatures of mutation and selection in the cancer genome. *Nature, 463*(7283), 893-898.

Cohen, E. E. W., Lingen, M. W., Zhu, B., Zhu, H., Straza, M. W., Pierce, C., . . . Rosner, M. R. (2006). Protein kinase Cζ mediates epidermal growth factor–induced growth of head and neck tumor cells by regulating mitogen-activated protein kinase. *Cancer research, 66*(12), 6296-6303.

Deevi, R. K., Javadi, A., McClements, J., Vohhodina, J., Savage, K., Loughrey, M. B., . . . Campbell, F. C. (2018). Protein kinase C zeta suppresses low-or high-grade colorectal cancer (CRC) phenotypes by interphase centrosome anchoring. *The Journal of pathology, 244*(4), 445-459.

Goossens, N., & Hoshida, Y. (2015). Hepatitis C virus-induced hepatocellular carcinoma. *Clinical and molecular hepatology, 21*(2), 105.

He, M.-F., Liu, L., Ge, W., Shaw, P.-C., Jiang, R., Wu, L.-W., & But, P. P.-H. (2009). Antiangiogenic activity of Tripterygium wilfordii and its terpenoids. *Journal of ethnopharmacology, 121*(1), 61-68.

Kangli, M., Xiaokang, Z., & Ying, D. (2000). Studies on triterpenoids constituents of Tripterygium wilfordii Hook. F. *Natural Product Research and Development, 12*(4), 1-7.

Keretsu, S., Bhujbal, S. P., & Cho, S. J. (2020). Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Scientific reports, 10*(1), 1-14.

Lee, C. M., Lu, S. N., Changchien, C. S., Yeh, C. T., Hsu, T. T., Tang, J. H., . . . Chen, W. J. (1999). Age, gender, and local geographic variations of viral etiology of hepatocellular carcinoma in a hyperendemic area for hepatitis B virus infection. *Cancer: Interdisciplinary International Journal of the American Cancer Society, 86*(7), 1143-1150.

Teo, K., Abeysekera, K. W., Adams, L., Aigner, E., Anstee, Q. M., Banales, J. M., . . . Bhatnagar, P. (2021). rs641738C> T near MBOAT7 is associated with liver fat, ALT and fibrosis in NAFLD: A meta-analysis. *Journal of hepatology, 74*(1), 20-30.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods, 12*(1), 7-8.