

Computational Data Mining and Structural Analysis of Deleterious Non-Synonymous SNPs in Human CDK1



MS Project Thesis By

Kiran Najaf

66-FOC/MSBI/S23

Supervised by

Dr Mehrosh Khalid

Lecturer

Department of Bioinformatics

Faculty of Computing and Information

International Islamic University Islamabad

Department of Bioinformatics
International Islamic University Islamabad

FINAL APPROVAL

This is to certify that we have read the thesis submitted by **Kiran Najaf**. It is our judgment that this thesis is of sufficient standard to warrant its acceptance by International Islamic University, Islamabad for the degree of MS Bioinformatics.

Committee:

External Examiner:

Internal Examiner:

Supervisor:

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Table of Contents

DEDICATION	IX
DECLARATION.....	XI
ACKNOLWDGEMENTS.....	XII
LIST OF ABERVATIONS	XIII
LIST OF FIGURES	VI
LIST OF TABLES.....	VIII
ABSTRACT.....	XV
CHAPTER 1	1
INTRODUCTION	1
1.1 Variation	7
1.2 Types of Variations	7
1.2.1 Structural Variation.....	7
1.2.2 DNA Chain Variation	7
1.2.3 Substitution Variation	7
1.3 Role of Variation in Human and Diseases	8
1.4 SNP Analysis	9
1.5 Types of SNPs.....	12
CHAPTER 2	16
LITERATURE REVIEW.....	16
2.1 Critical Analysis.....	25
CHAPTER 3	30
MATERIALS & METHODS	30
3.1 SNPs Databases	31
3.2 Processing of Deleterious nsSNPs.....	31
3.3 Impact of coding nsSNPs Assessed Through Structure-Based Homology Analysis	34
3.4 Characterization of Functional nsSNPs.....	34
3.5 Prediction of the Impact of nsSNPs on the Molecular Phenotype of Proteins.....	35
3.6 Impact of nsSNPs on protein stability	36
3.7 Finding Functional SNPs in the Conserved Area.....	36
3.8 Predicting Structures	37

3.9	Refinement of Protein Model	37
3.10	3D Structure Confirmation.....	37
3.11	Site-directed mutagenesis and in-silico RMSD Computation	38
3.12	Forecasting the Impact of Mutations on Structure.....	38
3.13	Pathway and Network Analysis Tools and Software for Bioinformatics.....	39
3.13.1	CDK1 PPI network retrieval using the STRING database	39
3.13.2	KEGG	39
CHAPTER 4		40
RESULTS & DISCUSSION		40
4.1	Retrieval of Single Nucleotide Polymorphisms (SNPs) Dataset	41
4.2	Analysis of Non synonymous (ns) SNPs.....	42
4.3	Nature of Selected nsSNPs	47
4.4	Conservation and stability Profiles of selected Non synonymous SNPs	51
4.5	Prediction and Assessment of CDK1 Tertiary Structure	55
4.6	Verification of model	56
4.7	Effect of Mutation on Structure of CDK1 and their Functional Consequences.....	58
4.8	Structural Mapping of Deleterious Mutations in CDK1.....	58
4.9	In-silico site directed mutagenesis and RMSD Calculation.....	60
4.10	Network and Pathway Analysis	69
4.10.1	Retrieval of CDK1 Protein Protein network.....	69
4.10.2	KEGG	72
CHAPTER 5		77
CONCLUSION.....		78
REFERENCES		79

TABLE OF FIGURES

Figure 1.1: Summary of cellular pathways and molecular targets regulated by CDK1	4
Figure 1.2: Functional versatility and pharmacological relevance of CDK1	6
Figure 1.3 Different types of DNA chain variation [33]	10
Figure 1.4: Single-nucleotide polymorphism in a sequence [34]	10
Figure 1.5: Types of SNPs.....	14
Figure 3.1: Complete workflow of this study is represented.....	33
Figure 4.1: Distribution of Single Nucleotide Polymorphisms across the Human CDK1 Protein Coding Region	42
Figure 4.2: Comparison of aggregation propensity (Δ TANGO) between wild-type and L66F mutant CDK1 protein	48
Figure 4.3: Predicted variation in chaperone-binding affinity of CDK1 between wild-type and E40K mutant using LIMBO analysis	50
Figure 4.4: Comparative WALTZ-predicted amyloido propensity of wild-type and L66F mutant CDK1.	50
Figure 4.5: Color-coded ConSurf scores depicting the evolutionary conservation of amino acid residues in the CDK1 protein	52
Figure 4.6: Native monomer structure of CDK1 with one domain.....	56
Figure 4.7: Predicted structural stability profile of the CDK1 protein.....	57
Figure 4.8: Visualization of deleterious nsSNPs in the CDK1 tertiary structure, with six mutations is highlighted in red color on the tertiary structure of CDK1 protein	59
Figure 4.9: The superimposed structures of the native and mutant (L66F)	61
Figure 4.10: The superimposed structures of the native and mutant (I10M) forms of the CDK1 protein are shown.....	61
Figure 4.11: The superimposed structure of the native and mutant (I35F) forms of the CDK1 protein.	62
Figure 4.12: The superimposed structure of the native and mutant (R123C) forms of the CDK1 protein.	62
Figure 4.13: The superimposed structure of the native and mutant (E40K) forms of the CDK1 protein.....	63
Figure 4.14: The superimposed structure of the native and mutant (V185A) forms of the CDK1 protein	63
Figure 4.15: Close up look of mutation in a protein. (a)L66F (b) I10M (c) I35F (d) R123C (e) E40K (f) V185A.....	68
Figure 4.16: Shows CDK1 PPI network that involved in signal transduction	70
Figure 4.17: CDK1 PPI network differentiated into 3 k-means clusters.....	71
Figure 4.23 KEGG Pathway database map showing MAPK signaling pathway and ubiquitin mediated	

proteolysis	73
-------------------	----

LIST OF TABLES

Table 1.1: Differences between SNPs and SNVs [35].	11
Table 2.1: Shows the Summary of Literature Review on CDK1	26
Table 4.1: The most deleterious nsSNPs were identified, with steric symbols marking the six selected positions in the CDK1 protein structure	43
Table 4.2: Assessment of Deleterious Mutation Probability	46
Table 4.3: Predicted functional impact scores of nsSNPs identified in the CDK1 protein.	48
Table 4.4: ConSurf predicted evolutionary conservation of CDK1 amino acid residues, with scores ranging from 1 (variable) to 9 (highly conserved), highlighting regions potentially critical for structural or functional integrity.....	52
Table 4.5: Stability Analysis of Selected nsSNPs Using I-MUTANT and MUPRO.....	54
Table 4.6: RMSD of both native protein and mutant models	64
Table 4.7: 3 K-mean clusters of CDK1	71

DEDICATION

We dedicated our project to **ALLAH ALMIGHTY**.

Who created us and made us able to use our abilities.

In addition, it gave me the strength to complete this task as my final year project.

&

I dedicated my work and success to my parents my dear husband Syed Kamran
Abbas Naqvi and my mentor Dr, Ali Nawaz Khan.

**A thesis submitted to Department of Bioinformatics,
International Islamic University, Islamabad as a partial fulfillment
of requirement for the award of the degree of M.S in Bioinformatics.**

DECLARATION

I hereby declare that this Thesis **Computational Data Mining and Structural Analysis of Deleterious Non- Synonymous SNPs in Human CDK1** neither as a whole nor as a part has been copied out from any source. It is further declared that i have done this research with the accompanied report entirely on the basis of my supervisor **Dr, Mehrosh Khalid**. If any part of the system is proved to be copied out from any source or found to be reproduction of any project from any of the training institute or educational institutions, I will stand by the consequences.

Date _____

Kiran Najaf

(66-FOC/MSBI/S-23)

ACKNOWLEDGEMENTS

As in the Holy Quran, “**The seven heavens and the earth and all that is therein, glorify HIM and there is not a thing but glorifies HIS praise**”.

I, therefore, start my acknowledgment as a word of thanks to almighty **ALLAH**, the **Exalted**, the **Gracious** and the **Merciful**, and to whom i owe my very existence. HE bestowed upon me the strength, wisdom and perseverance to complete my research work successfully. I am nothing without my **ALLAH** but i can achieve everything with HIS assistance. I present my heartfelt gratitude from the bottom core of my heart to my **Holy Prophet Hazrat Muhammad (SAW)** whose life is an eternal source of guidance for me and for the entire world. His (SAW) teachings always show us the right path in times of difficulty.

I express my sincere and respectful appreciation to my worthy supervisor **Dr, Mehrosh Khalid** for her scholarly guidance, mentorship and vast knowledge. I am thankful for the patient guidance, enthusiastic encouragement and useful critiques for this thesis. Her assistance and advice kept my progress on schedule.

I would like to pay my sincere thanks to all my teachers who were very kind during my studies at International Islamic University, Islamabad.

I wish to express my special thanks to my family for their wonderful support and courage.

LIST OF ABBREVIATION

Acronym	Abbreviation
SNP	Single Nucleotide Polymorphism
nsSNP	Non-synonymous Single Nucleotide Polymorphism
SNV	Single Nucleotide Variation
RNA	Ribonucleic Acid
mRNA	Messenger Ribonucleic Acid
DNA	Deoxyribonucleic acid
HSCs	Hepatic stellate cells
ECM	Extracellular matrix
AD	Alzheimer's disease
GWAS	Genome-wide association study
EC	Endometrial Cancer
KEGG	Kyoto Encyclopedia of Genes and Genomes
CI	Complex I
ATP	Adenosine Tri-Phosphate
CPC	Chromosomal passenger complex complex/cyclosome
APC/C	Anaphase-promoting
AI	Artificial Intelligence
PCA	Principal Component Analysis
CNN	Convolutional Neural Networks
RF	Random Forest
TDP1	Tyrosyl-DNA phosphodiesterase 1
Top1cc	Top1-DNA complexes
WB	Western blotting

Acronym**Abbreviation**

NCBI

National Center for Biotechnology Information

PMUT

Piezoelectric Micromachined Ultrasonic
Transducers

PolyPhen-2

Polymorphism Phenotyping v2

STRING

Search Tool for the Retrieval of Interacting
Genes/Proteins

FASTA

Fast all

RBD

Receptor binding domain

AAS

Amino Acid Substitution

ProSA

Protein Structure Analysis

FACS

Fluorescence-activated cell sorting

PM

Polygonum multiflorum

AR

Androgen Receptor

RI

Reliability Index

SIFT

Sorting Intolerant from Tolerant

PROVEAN

Protein Variation Effect Analyzer

SNAP2

Screening for non-acceptable Polymorphism

PhD-SNP

Predictor of Human Deleterious Single
Nucleotide Polymorphism

PANTHER

Protein Analysis through Evolutionary
Relationships

dbSNP

Single Nucleotide Polymorphism Database

SVM

Support Vector Machine

DPP4

Dipeptidyl peptidase

RefSeq

Reference Sequence

NHGRI

National Human Genome Research Institute

CETSA

Cellular Thermal Shift Assay

RT-qPCR

Reverse transcription quantitative
polymerase chain reaction

ABSTRACT

CDK1 (Cyclin-Dependent Kinase 1) plays a vital role in G2/M process that's leads to mitotic progression and can start mitosis. The main role of CDK1 is to regulate the cell cycle process in the human body.

Amino acid substitutions resulting from non-synonymous single nucleotide polymorphisms (nsSNPs) can significantly impact protein structure and function, potentially leading to impaired molecular interactions and disease development. These structural alterations may disrupt normal cellular pathways, underscoring the importance of identifying functionally relevant nsSNPs in genes associated with human disorders. This remains a critical and ongoing challenge in genetic and molecular biology research. This research utilized computational and bioinformatics methodologies, incorporating data mining techniques, to systematically analyze non-synonymous single nucleotide polymorphisms (nsSNPs) with the aim of identifying genetic variants that may impact the function and expression of the CDK1 protein. In this study thirty-two nsSNPs in the CDK1 protein were predicted to be deleterious using the SIFT algorithm, while PolyPhen-2 identified twenty-seven potentially damaging variants. Cross validation using more than six computational tools led to the prioritization of six high-confidence deleterious nsSNPs. Structural and functional impacts of these variants were evaluated through I-Mutant and Project HOPE, which indicated significant alterations in protein stability and conformational dynamics. Functional domain analysis through InterPro and the Conserved Domain Database (CDD) confirmed the localization of these mutations within critical functional regions of CDK1. The six prioritized SNPs—(L66F) rs1460089100, (I10M) rs1179034229, (I35F) rs2080301489, (R123C) rs1322570597, (E40K) rs746720738, and (V185A) rs1234513195 were found to affect the structural integrity, functional properties, and stability of the CDK1 protein. These computational findings provide a valuable basis for future experimental validation and large-scale population-based association studies focused on the role of CDK1 variants in disease pathogenesis.

Keywords: CDK1; nsSNPs; most deleterious mutations; in-silico analysis; structure analysis

CHAPTER 1

INTRODUCTION

Cyclin-dependent kinase 1 (CDK1) is a serine/threonine kinase and a key member of the cyclin-dependent kinase (CDK) family, responsible for phosphorylating target proteins on serine or threonine residues to regulate cell cycle progression. Encoded by the cell division cycle 2 (CDC2) gene, CDK1 has a molecular mass of approximately 34 kDa. It plays a pivotal role in initiating and regulating mitosis, particularly in facilitating the transition from the G2 phase to the M phase. Uniquely, CDK1 is the only CDK essential for cell division in mammalian systems, acting as a central regulator of mitotic entry and progression in the human cell cycle. [1, 5].

The cell cycle is an essential and tightly controlled process that governs cell growth, replication, and division. It proceeds through a series of distinct phases G1, S, G2, and M phase ensuring precise DNA replication and equal distribution of genetic material, ultimately producing two daughter cells with identical genomes [94].

The cell cycle is regulated by highly conserved molecular machinery composed of cyclin-dependent kinases (CDKs), a subclass of serine/threonine protein kinases. CDK activity is tightly modulated through phosphorylation events and specific protein-protein interactions. Full activation of CDKs necessitates their association with regulatory cyclin subunits. Conversely, cyclin degradation via the ubiquitin-proteasome pathway plays a critical role in facilitating timely transitions between cell cycle phases by rapidly inactivating CDKs. In human cells, the prevailing model suggests that CDK1 drives the progression of cells from the G2 phase into mitosis by forming active complexes with mitotic cyclins, primarily cyclins A and B [3].

Cyclin-Dependent Kinase 1 (CDK1) has been implicated in the pathogenesis of several human diseases, including neurodegenerative disorders such as Alzheimer's disease and Parkinson's disease, as well as stroke, HIV infection, and various forms of cancer, due to its essential role in cell cycle regulation and cellular proliferation [2]. Humans have twenty CDKs (1–20). Originally CDKs (1–6, 11, 14–18) were identified as cell cycle regulatory proteins, i.e., those in the CDKs family of proteins. The CDKs have been defined based on their role in cell cycle control as interphase CDKs (CDK2, CDK4 and CDK5), mitotic CDK1 and transcriptional CDKs (CDK7 and CDK9) [1]. Cell cycle is a life supporting process which maintains the temporal and spatial ordering of DNA replication and segregation into daughter cells. Besides regulating cellular development, proliferation and differentiation and guaranteeing genetic continuity through meiosis and mitosis, this process also accounts for environmental signals in order to maintain genomic stability [6].

Research indicates that CDK1, in association with its cyclin partners Cyclin A and Cyclin B plays a pivotal role in regulating the progression of the cell cycle, particularly during the G1/S transition and the G2/M phase shift. The CDK1–Cyclin A complex is believed to initiate mitotic entry at the G2/M boundary, while the CDK1 Cyclin B complex is essential for both the initiation and completion of mitosis. Aberrant expression or activation of CDK1 has been implicated in various malignancies, including colorectal, lung, and breast cancers. Beyond its central role in cell cycle regulation, CDK1 has also been associated with promoting cellular survival mechanisms [5].

Cyclin-dependent kinases (CDKs) are essential for cellular proliferation and survival, with their loss often triggering apoptotic pathways underscoring their significance as potential therapeutic targets in oncology. Accordingly, CDK4/6 inhibitors such as abemaciclib, ribociclib, and palbociclib have received FDA approval for the treatment of breast cancer. Recent research suggest that CDK1 possesses the capacity to compensate for the loss of other interphase CDKs, implying functional redundancy within this family; however, CDK1 itself remains uniquely indispensable for mitotic entry and progression. Its non-redundant role establishes it as the sole CDK essential for proper cell cycle regulation [1]. Entry into mitosis is orchestrated by the stepwise activation of CDK1, primarily via its association with Cyclin B1. This complex becomes fully active following the dephosphorylation of CDK1 at threonine 14 and tyrosine 15, mediated by the Cdc25 phosphatase family. Among its three isoforms, Cdc25B initiates CDK1 activation at the centrosome, while Cdc25A and Cdc25C contribute to the amplification of the mitotic signal in the cytoplasm and nucleus, respectively [4].

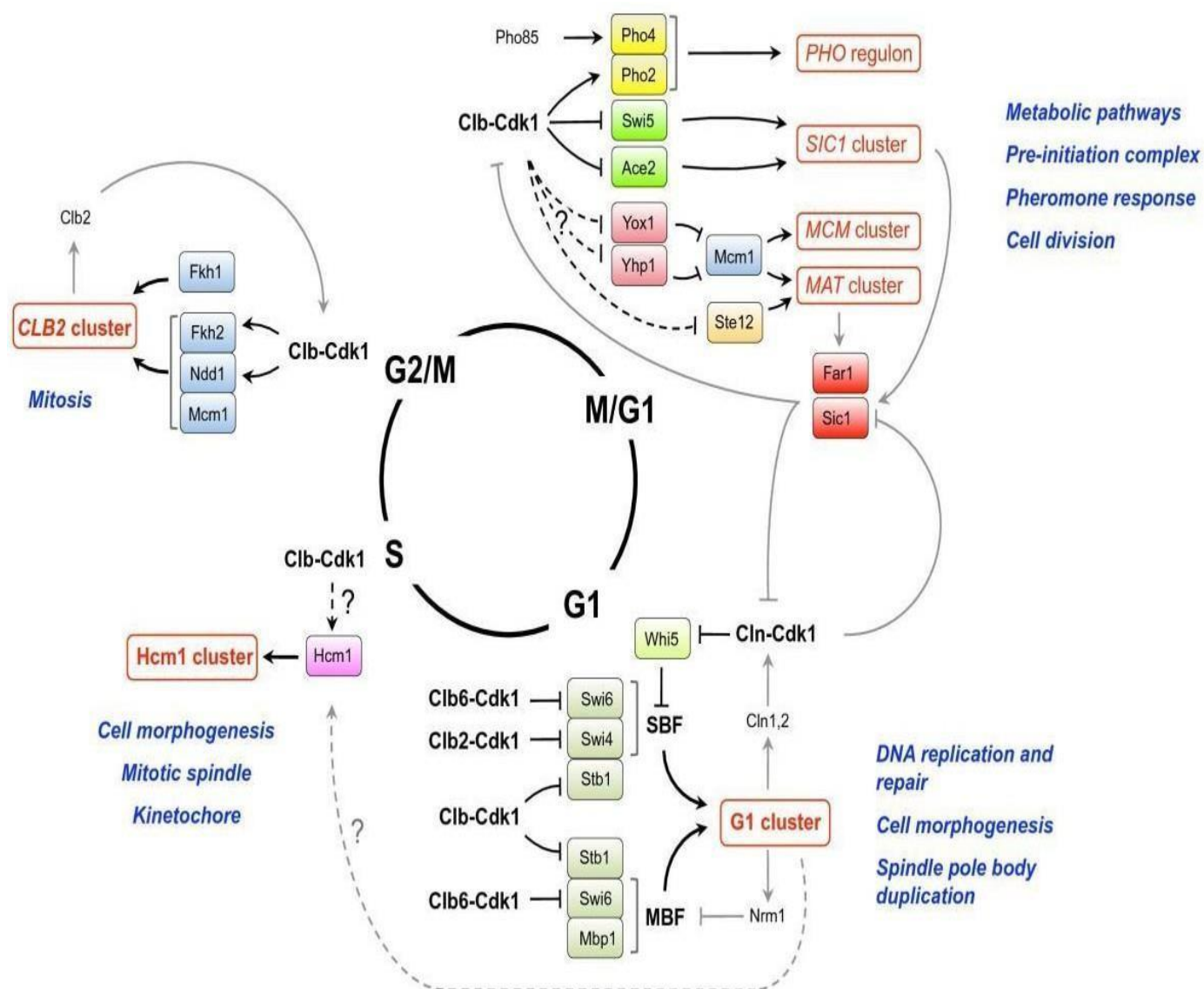


Figure 1.1: Summary of cellular pathways and molecular targets regulated by CDK1.

CDK1 regulates cell cycle progression by modulating transcriptional networks through integrated positive and negative feedback loops [39]

For diseases requiring unchecked cell proliferation, such as cancer, some CDKs particularly those that stimulate cell division are thought to be potentially intriguing therapeutic targets. CDK10 was not given much attention until it was shown to be a significant predictor of resistance to endocrine treatment for breast cancer, despite the fact that it was discovered during the pre-genomic period, when research on CDKs produced significant advances in our knowledge of cell cycle control [7].

There are twenty-one CDKs known to exist. Cell cycle checkpoint regulation is mostly regulated by CDK1, CDK2, CDK4, and CDK6, whereas transcriptional regulation is regulated by CDK8, CDK9, and CDK19. Although it is conceivable to expect that additional CDKs may contribute to cell cycle events in response to certain cytotoxic stressors, their molecular mechanisms are yet unclear [8].

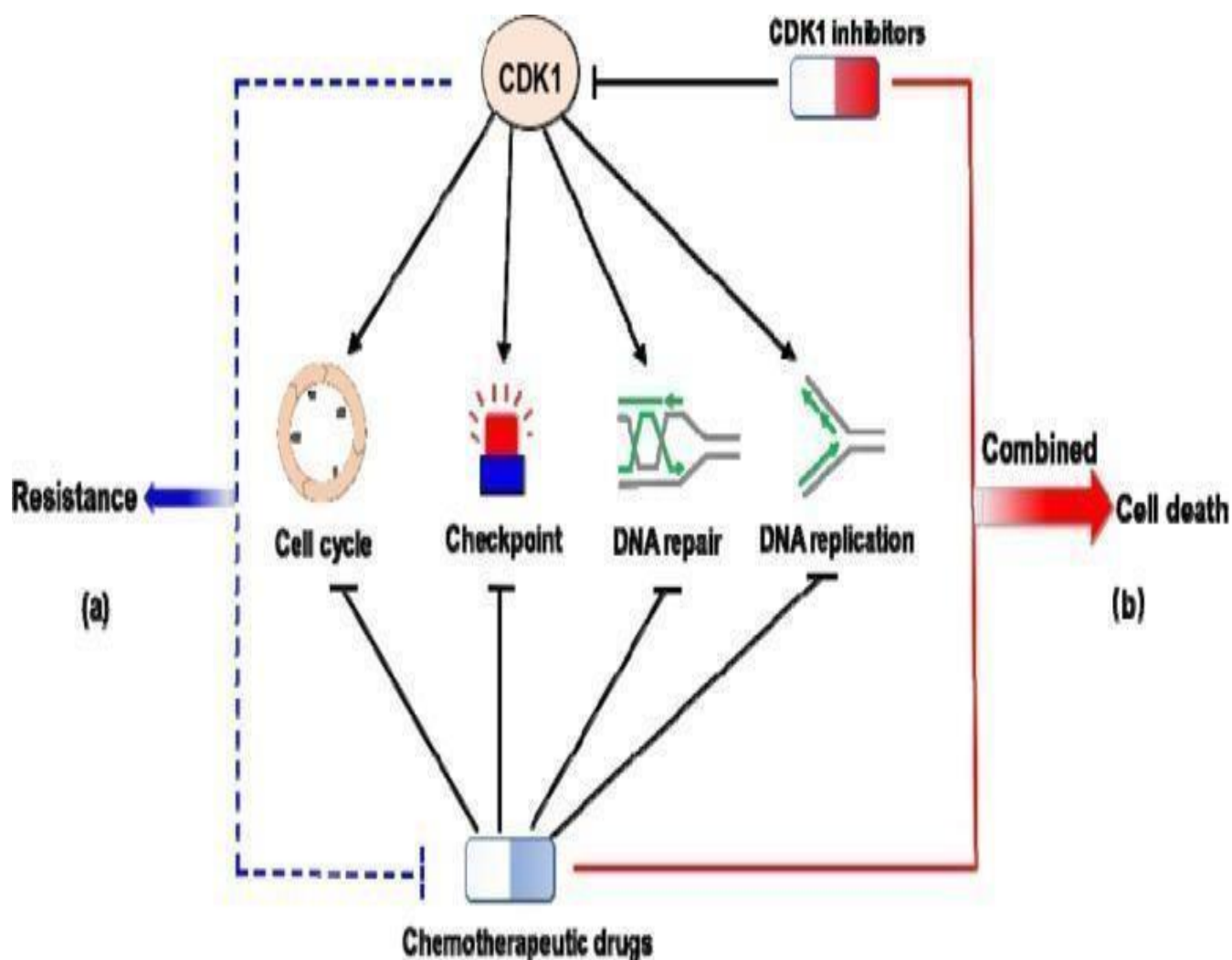


Figure 1.2: Functional versatility and pharmacological relevance of CDK1

(a) As we argue, CDK1 supports many biological processes that are essential for cell survival, such as DNA replication, checkpoint activation, G2/M transition, and DNA repair. Its involvement in these processes may lead to chemo resistance and jeopardize the effectiveness of chemotherapeutic medications. **(b)** Therefore, pharmacologically inhibiting CDK1 in conjunction with traditional anti-cancer medications may increase their toxicity to cancer cells and cause cell death [38].

1.1 Variation

Variation is the term used to describe the possible variances between nucleotide sequences in a gene. Any change between an organism's cells, groups of organisms, or creatures of any species is called variation. Genetic variations, known as genotypic variation, or the impact of environmental factors on gene expressions, known as phenotypic variation, may both cause it. Mutation, random mating, recombination, and gene immigration across homologous chromosomes during meiosis are the primary sources of variances. Protein variations may alter their structure or function, but they can also help identify illnesses and build resistance to them [31].

1.2 Types of Variations

Below is a discussion of the three main categories of variations:

1.2.1 Structural Variation

Structural variation refers to a segment of DNA that is 1 kb or larger. The term "copy number variations" also refers to genomic imbalances or balanced translocations that occur when a DNA region larger than 1 kb appears more than once in the genome. Techniques that use sequencing and ensemble algorithms may detect structural differences by obtaining information on non-synonymous SNPs' occurrence, associations with various disorders, and biological expressions [32].

1.2.2 DNA Chain Variation

DNA chain variation refers to genetic alterations occurring within a gene or its functional elements, potentially influencing gene expression, structure, or biological activity. Different DNA chain variations may include DNA translocation, DNA inversion, DNA deletion, and DNA indel (insertion-deletion). In DNA indel, one or more nucleotides are removed and replaced by other nucleotides, while in DNA deletion, a nucleotide is removed from the gene sequence. Insertion of DNA occurs when one or more additional nucleotides are added to a gene sequence. In a chromosome, DNA translocation occurs when a segment of a sequence is relocated to a new site; DNA inversion occurs when this translocation is reversed back to its original location [33]. Figure 1.3 lists many DNA chain variants as one example.

1.2.3 Substitution Variation

Genetic variation occurs both within and across populations, resulting in polymorphisms that may be linked to a phenotypic or a genetic characteristic when an environmental stimulus is present [9]. Substitution variation occurs when an amino acid in a protein sequence is changed to a different amino acid or when a base pair in a DNA sequence is changed to a different base pair.

Common names for this substitution include single-nucleotide substitution, single-nucleotide polymorphism (SNP), and single-nucleotide variation (SNV), as shown in Figure 1.4.

1.3 Role of Variation in Human and Diseases

One interbreeding species that is always changing is *Homo sapiens*. Each individual has unique genetic makeup. Approximately 85% of all human genetic diversity is found in human populations. SNPs are the most prevalent kind of genetic variation. They create alterations in a nucleotide sequence and are linked to a number of illnesses, including diabetes mellitus, breast cancer, Alzheimer's disease, and cystic fibrosis. They typically occur once per 1,000 nucleotides, meaning that there are 4–5 million SNPs in each individual's genome. 90% of inter-individual variance is thought to be caused by SNPs [25].

Scientists may discover genes connected to human illness by using the variation's function as biological markers. SNPs help forecast a person's likelihood of developing a certain illness, their susceptibility to environmental factors like pollution, and how they will react to particular therapies. SNPs can serve as genetic markers to trace the inheritance patterns of disease-associated genes within families [24].

The interdisciplinary domain of bioinformatics employs *in Silico* approaches to systematically manage, store, classify, and analyze biological data at the molecular level. Through the integration of computational algorithms, web-based servers, and curated biological databases, bioinformatics provides powerful tools for addressing complex biological questions [26]. A wide range of biological databases have been developed to facilitate the storage and retrieval of genomic, proteomic, and other molecular data, thereby supporting researchers in conducting efficient and accurate analyses. In parallel, specialized software applications have been introduced to assist in gene and protein function prediction, annotation, drug discovery, and molecular modeling [27].

The demand for bioinformatics has grown substantially, particularly following the completion of the Human Genome Project, which generated an unprecedented volume of genetic data, including information related to cancer biology [28]. In contemporary cancer research and therapeutics, bioinformatics plays an essential role. Numerous cancer-specific databases, continuously updated and based on *in silico* models, enable researchers to examine cancer cell behavior, phenotypic impacts, and molecular pathways in a quantifiable manner [29]. Furthermore, bioinformatics facilitates advanced applications such as rational drug design, prediction of protein secondary and tertiary structures, and the identification of gene loci involved in disease processes [30]

1.4 SNP Analysis

Point mutations known as SNPs are present in the population of functioning proteins. SNPs have the power to change the functional specificity of proteins that are enzymes engaged in important processes. Four primary factors protein-protein interaction, protein-ligand interaction, oligomerization, and the acquisition of divalent ions are used by proteins either alone or in combination to fulfill biological activities. All or some of the aforementioned variables may be impacted by the development of SNPs, which might result in a malfunctioning protein and ultimately cause any illness.

More than 9 million SNPs have been found in public databases, making them the most prevalent kind of variation. SNPs predominantly occur in noncoding regions of the genome, whereas coding regions tend to exhibit a lower frequency of these variations. Various experimental methods are employed for SNP genotyping, including polymerase chain reaction (PCR)-based amplification, primer extension assays, hybridization techniques, enzymatic cleavage, ligation reactions, and single-strand conformation polymorphism (SSCP) analysis [36].

Using various bioinformatics techniques, *in-silico* SNP analysis predicts if nsSNPs will have a negative or neutral impact on a certain gene or protein. It also predicts how those negative nsSNPs would influence the structure and activities of a protein. To forecast the impact of harmful SNPs on functions and structure, bioinformatics offers a variety of techniques for identifying neutral or harmful SNPs. Predicting harmful SNPs and their impact on protein function or structure may be done with the help of widely used in-silico tools like SIFT, PolyPhen, PredictSNP, SNAP2, PANTHER, PROVEAN, MUp, and I-Mutant. Both transition and transversion SNPs are types of single-nucleotide polymorphisms. A transition occurs when a pair of purine bases (A, G) is changed to another purine base pair or a pyrimidine base pair (C, T) is changed to another pyrimidine, while a transversion occurs when a purine base pair is changed to a pyrimidine or a pyrimidine to a purine [33]. Table 1.1 lists the distinctions between SNPs and SNVs, despite the fact that both are forms of genetic diversity.

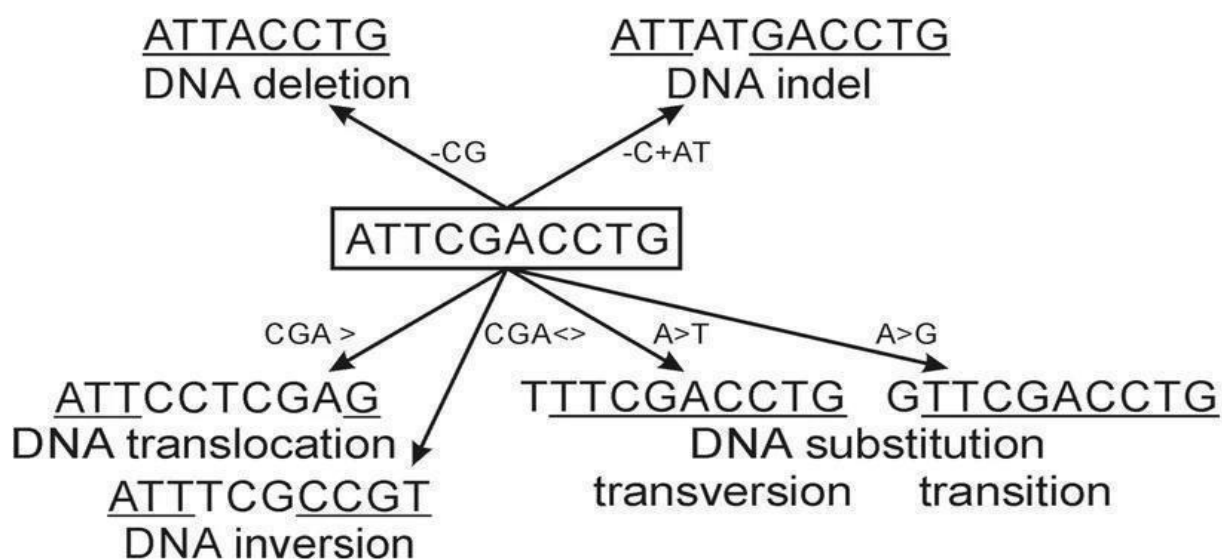


Figure 1.3 Different types of DNA chain variation [33].

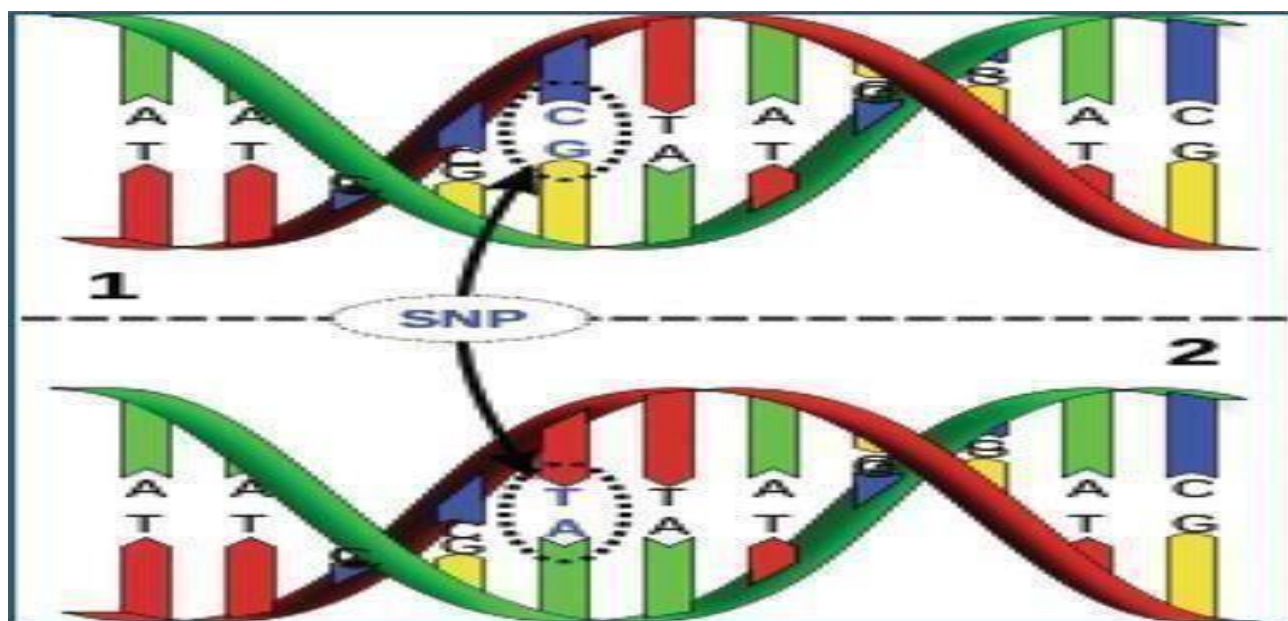


Figure 1.4: Single-nucleotide polymorphism in a sequence [34].

Table 1.1 Differences between SNPs and SNVs [35].

SNP	SNV
• SNP stands for single-nucleotide polymorphism. • SNP is considered to be a mutation if the variation is present in more than 1% of the population. • It is for germline mutations.	• SNV stands for single-nucleotide variation. • SNV is a mutation when there is a variation even in one nucleotide. • It is for somatic mutations

Variation in a single place in the DNA sequence in more than 1% of the population is known as a SNP. Depending on the location in which they are found, SNPs have an impact on the structure and function of genes. It has been found that single nucleotide variants are important in a number of disorders. Variants in gene regulatory regions are crucial to research because they have a direct impact on gene expression. Therefore, it is crucial to investigate SNPs found in non-coding areas such promoters and the 3'untranslated region (UTR). Variants in these regulatory regions may impact a gene's expression in a number of ways, including changing transcription factor binding sites and microRNA (miRNA) binding sites [11]. Exploration of non-coding DNA is crucial since it contains hundreds of known SNPs in human genomes that are linked to various diseases. Because of their impact on gene expression, posttranslational, and posttranscriptional activities, SNPs in the 3 and 5 untranslated regions (UTR) are of primary interest to researchers [12].

Point mutations known as SNPs proliferate across populations [13]. Although SNPs are scattered across the genome, the phenotypes of organisms are influenced by their location within the coding area. Over five hundred thousand SNPs may be discovered in the human genome's coding region [14]. It is simple to identify the new root and rising complex of genetic illness using the baseline information of SNPs. SNPs have a significant influence on both phenotypic traits and genetic diseases [14]. Certain SNPs alter both the structure and function of proteins and have an impact on how they interact. Therefore, SNPs that change a protein's function may contribute to the development of illness. [16].

1.5 Types of SNPs

Both coding and non-coding areas include SNPs. Figure 1.5 shows the distinction between synonymous and non-synonymous SNPs that are present in the coding area. Non-synonymous SNPs (nsSNPs) cause changes in protein sequence, which in turn affects changes in protein structure and function, while synonymous SNPs do not modify the DNA or amino acid sequence or the function of a gene or protein [15]. Additional SNPs are divided into tolerant and harmful nsSNPs. While harmless nsSNPs cause changes in the location of amino acids but do not impact on the structure or function of proteins, harmful nsSNPs cause significant disruptions in protein structure, function, and interactions [17 and 18].

In contrast, missense and nonsense SNPs are further subdivided from the SNPs classified as non-synonymous SNPs. While nonsense SNPs alter the stop codon in the DNA, which may have an impact on the protein output, missense SNPs may cause proteins to cease functioning by changing a codon to code for a different amino acid. Gene splicing, transcriptional binding, messenger RNA degradation, and non-coding RNA sequencing may all be directly impacted by SNPs in non-coding sections of genes [34].

The goal of pharmacogenomics is to identify harmful nsSNPs linked to illness, which is a very difficult task when working with large amounts of data. Sorting the SNPs according to their functions is one way to potentially solve this issue [17 and 18]. SNPs in the regulatory and coding regions may change the protein's structure and expression level, which might accelerate the course of the illness [19]. SNPs can contribute to the disruption in gene regulation by altering the DNA transcriptional binding factor [20].

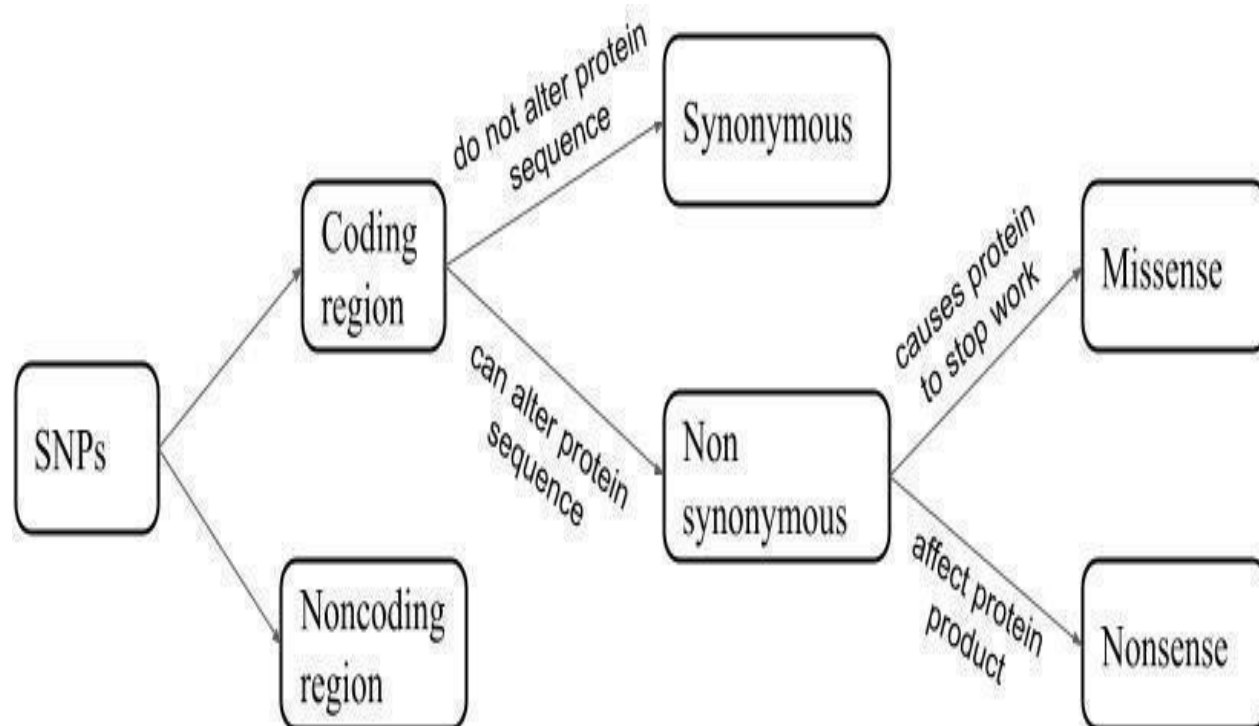


Figure 1.5: Types of SNPs

Finding the SNPs that directly contribute to the illness or raise the risk of the disease that is, an individual's likelihood of developing cancer is crucial. It's also critical to stratify the population's cancer-risk individuals. Certain dangerous alleles have negative effects that may contribute to a 5%–25% increase in risk [21 and 22]. Compared to those with fewer hazardous alleles, the population or people with many alleles are likewise at a higher risk of developing cancer [21, 23].

1.2.2 Role of SNP Analysis in Research & Diseases

Understanding human genetic variants and other prevalent disorders is made easier by advances in genomic variation, such as SNP analysis. By determining the likelihood of contracting a disease and detecting mutations linked to it, the predictions made by SNP analysis about the stability of protein structure and function are crucial for medication development [37].

1.3 Problem Statement

1. It is highly probable that previously unreported non-synonymous single nucleotide polymorphisms (nsSNPs) exist within the human genome that contribute to disease susceptibility by inducing alterations in protein structure and/or function.
2. To identify the deleterious nsSNPs through experimental method is time consuming and very costly

1.4 Research Questions

1. How do changes in the functional cascade of CDK1 due to SNPs impact various normal cellular pathways?
2. What are the potential impacts of SNPs in CDK1 on cellular function and how can these be predicted?
3. Can a detailed mapping of SNPs in CDK1 reveal insights into their role in disrupting normal cellular processes?

1.5 Proposed Solution

The current study design focuses on the understanding of genotypic factors and its potential impacts on phenotypic association in the form of human disease. We expect to achieve the following major and minor outcomes.

- Screening and mapping of the currently reported SNPs for CDK1 protein.
- Impacts of these SNPs on protein structure, in the form of stability.
- Prediction of altered function and its consequences in human diseases.

1.6 Aim and Objectives

- Insilco identification and screening of functional SNPs in CDK1 protein
- To forecast and chart the possible effects of nsSNPs on the composition and functionality of the CDK1 protein.
- This study aims to computationally screen and analyze deleterious non-synonymous SNPs in the human CDK1 protein.

CHAPTER 2

LITERATURE REVIEW

In this chapter, I have provided an in-depth review of the existing literature relevant to cyclin dependent kinase 1 (CDK1) in disease diagnostics and therapeutics. The review overviews the role of CDK1 in controlling cell cycle progression, its role in various diseases and CDK1 the emerging techniques for targeting of CDK1 in clinical settings.

Targeting CDK genes, multiplex PCR is used to diagnose Alzheimer's disease.

One of the most prevalent neurodegenerative disorders, Alzheimer's disease (AD), has no known biomarkers for early detection. To create a possible treatment intervention, a precise biomarker is required. Studies conducted on mammalian cells over ten years ago revealed a connection between the development of AD and the expression of cell cycle proteins, namely cyclin dependent kinases (CDKs). This research sought to ascertain if CDKs may serve as biomarkers for AD diagnosis. Consequently, a multiplex-PCR-based diagnostic technique was created for the quick and accurate identification of AD biomarkers. According to GWAS results, an AD mouse model showed increased CDK genes. Final optimization of multiplex- PCR based on the genomic analysis and specific primer sets and the conditions targeting CDK genes were extrapolated to RT-PCR on the AD mouse model and showed close concordance with the results. Moreover, an in silico meta-analysis showed that CDK genes have a significantly higher expression in moderate to severe AD patients. These findings indicate that the diagnostic method invented in this study would provide useful information for AD therapeutics and could also be used for early detection of AD [8].

CDK1 serves as a novel therapeutic target for endometrioid endometrial cancer

This study was to identify new therapeutics target for treating endometrial cancer (EC), a prevalent gynecologic malignancy. By performing bioinformatics analysis, CDK1 becomes a key gene in EC pathogenesis. CDK1 overexpression was confirmed by immunohistochemistry and cell proliferation and colony formation were increased by CDK1 in functional assays. Inhibition of CDK1 stimulated apoptosis, blocked tumor growth in xenograft models as well as stimulated G2/M phase cell enter in xenograft models. These results indicate a role of CDK1 in EC progression and provide a basis for development of CDK1 as a therapeutic target [1].

Mechanisms and Implications of Targeting CDK1 in Cancer

Cyclin dependent kinases (CDKs), especially CDK1 have become effective therapeutic target because they play an essential role in cell cycle regulation. While overexpression of CDKs in complex with cyclins has been observed in cancer tissue relative to normal, both evidence from the TCGA database and correlations for poor patient survival in multiple cancer types exists.

Deregulation of the CDK1 is closely related to the tumorigenesis, and CDK1 activation impacts several pathways related to cancer via phosphorylation of critical substrates. It has also been further shown by Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis that CDK1 interacting proteins are involved in oncogenic pathways. Small molecules targeting CDK1 are already numerous and are efficacious in the preclinical stage taking some to clinic. Readings up on CDK1 in cancer progression and potential as a therapeutic target are reviewed in this article [40].

Cyclin B1/Cdk1 Coordinates Mitochondrial Respiration for Cell-Cycle G2/M Progression

While mitochondrial energy production is essential for proper cell cycle progression, the precise mechanisms that coordinate mitochondrial respiration with specific phases of the cell cycle particularly the G2/M transition remain incompletely understood. Recent investigations have demonstrated that a subset of cyclin B1/CDK1 complexes localize to the mitochondrial matrix, where they phosphorylate multiple mitochondrial proteins, including subunits of complex I (CI) within the electron transport chain. Phosphorylation by cyclin B1/CDK1 has been shown to enhance CI activity, whereas the absence of phosphorylation in key CI subunits results in a marked reduction of its function. This regulatory mechanism promotes elevated mitochondrial respiration, increased oxygen consumption, and enhanced ATP production to meet the heightened bioenergetic demands of the G2/M transition, thereby facilitating efficient cell cycle progression. The findings suggest that cyclin B1/CDK1-mediated phosphorylation of mitochondrial targets is a critical adaptive mechanism that enables cells to align metabolic output with mitotic entry [41].

Quantitative differences between cyclin-dependent kinases underlie the unique functions of CDK1 in human cells

Unlike higher eukaryotes that possess multiple cyclin-dependent kinases (CDKs) to regulate cell cycle transitions, *Saccharomyces cerevisiae* utilizes a single CDK to coordinate both the G1/S and G2/M phases. In human cells, studies employing conditional gene silencing systems have demonstrated that CDK1 can functionally compensate for CDK2 during the G1/S transition. However, although mitotic entry is compromised in cells lacking CDK1, its absence alone does not entirely prevent progression into mitosis. Instead, CDK2 complexes that lack cyclin B exhibit impaired phosphorylation of mitotic substrates, resulting in metaphase arrest and mitotic anomalies such as the uncoupling of anaphase A and B and defective cytokinesis. These mitotic irregularities are further intensified by CDK2 activity, indicating that the functional divergence between G1/S and G2/M CDKs in human cells is primarily governed by differences in their cellular concentrations rather than substrate specificity alone [2].

Survivin Mediates Mitotic Onset in HeLa Cells through Activation of the Cdk1-Cdc25B Axis

Survivin, a critical component of the chromosomal passenger complex (CPC), plays an essential role in ensuring accurate chromosome segregation by correcting microtubule–kinetochore attachment errors and facilitating proper cytokinesis, thereby preventing aneuploidy, polyploidy, and tumorigenesis. In cancer cells, such as HeLa cells, depletion of Survivin results in early prophase arrest characterized by an intact nuclear lamina, diminished CDK1 activity, and preserved spindle apparatus, suggesting a distinct role for Survivin in mitotic entry. Specifically, the absence of Survivin disrupts the Cdc25B–CDK1 signaling pathway, impairing CDK1 localization to the centrosome. This defect in mitotic initiation can be rescued by reintroducing recombinant Survivin, which restores activity to the Cdc25B–CDK1–Cyclin B1 complex in the cytoplasm of Survivin-deficient cells. Moreover, the expression of a gain-of-function mutant form of Cdc25B effectively circumvents the prophase arrest, reinforcing the notion that Survivin contributes to CDK1 activation through Cdc25B modulation. These findings position Survivin as a pivotal regulator of mitotic progression in cancer cells [4].

The Effect of Targeting Cyclin Dependent Kinase 1 Induces Apoptosis and Cell Cycle Arrest of Activated Hepatic Stellate Cells

Liver fibrosis, a defining feature of chronic hepatic diseases, is characterized by the activation of hepatic stellate cells (HSCs) and the excessive accumulation of extracellular matrix (ECM) components. Despite extensive research, the molecular mechanisms underlying the upregulation of cytokines, chemokines, and ECM proteins by proliferating HSCs during fibrogenesis remain incompletely understood. Cyclin-dependent kinase 1 (CDK1) is a pivotal regulator of eukaryotic cell cycle progression, particularly during the G1/S and G2/M transitions. Recent studies have demonstrated elevated expression of CDK1 and its regulatory partners Cyclin A2 and Cyclin B1 in activated HSCs, patient-derived fibrotic liver tissues, and murine models of hepatic fibrosis. Functional inhibition of CDK1 in cultured human HSCs, using either small-molecule inhibitors or shRNA-mediated gene silencing, results in apoptosis and cell cycle arrest, potentially mediated through modulation of Survivin expression. These findings underscore the critical role of CDK1 in promoting HSC proliferation and identify it as a promising therapeutic target for the treatment of liver fibrosis [5].

CDK1-PP2A-B55 interplay ensures cell cycle oscillation via Apc1-loop

Protein phosphatases such as PP1 and the PP2A-B55 complex play essential roles in reversing cyclin-dependent kinase 1 (CDK1)-mediated phosphorylation of key cell cycle regulators.

However, identifying the principal phospho-substrates responsible for driving cell cycle oscillations has remained a significant challenge. Recent evidence highlights the importance of phosphorylation dynamics in modulating the activity of the anaphase-promoting complex/cyclosome (APC/C), particularly through a regulatory region in Apc1 known as the loop300 domain (Apc1-300L). This domain is subject to phosphorylation by CDK1 and dephosphorylation by PP2A-B55. In the absence of Greatwall kinase, premature activation of PP2A-B55 during mitosis leads to dephosphorylation of Apc1-300L, resulting in APC/C inactivation and delayed degradation of Cyclin B. This delay can be mitigated either by genetic deletion of the Apc1-300L domain or through pharmacological inhibition of PP2A-B55 using phosphorylated Ensa (pEnsa). Additionally, APC/C association with its coactivator Cdc20 is modulated throughout the cell cycle, with Apc1-300L dephosphorylation impeding Cdc20 binding. These findings emphasize the role of the CDK1–PP2A regulatory axis in coordinating APC/C function and cell cycle progression through a central phospho-switch mechanism that underlies periodic oscillations in cell cycle activity [6].

Computational analysis of Cyclin D1 gene SNPs and association with breast cancer

The chromosomal translocation involving chromosome 11 leads to the generation of the *CCND1* gene product, cyclin D1, whose function may be altered by specific single nucleotide polymorphisms (SNPs). Of particular interest is the splice site variant c.1988+5G, which may affect cyclin D1 splicing and functionality. To explore the structural and functional consequences of SNPs in *CCND1*, computational approaches were employed to predict their potential impact on protein behavior. The integration of high-performance computing (HPC) and artificial intelligence (AI) has significantly enhanced the efficiency and depth of bioinformatics data mining, enabling the handling and interpretation of complex biological datasets.

Data Mining Techniques in Bioinformatics Enhanced by HPC and AI Convergence

A novel methodological framework has been proposed that combines dimensionality reduction and machine learning with deep learning techniques, including Principal Component Analysis (PCA), Random Forest (RF), and Convolutional Neural Networks (CNN). When benchmarked against six conventional data mining approaches, this integrated model demonstrated superior performance across critical evaluation metrics such as F1 score, accuracy, precision, and recall.

The application of PCA facilitates dimensionality reduction and feature extraction, RF contributes through ensemble learning and feature ranking, and CNNs effectively capture intricate patterns in biological sequences and imaging data.

Notably, the proposed hybrid model significantly reduces computation time, making it well-suited for processing large-scale genomic and proteomic data in real time. This efficiency supports rapid diagnostic evaluation, accelerates drug discovery pipelines, and yields novel insights into biological mechanisms. The approach exemplifies the transformative potential of combining AI and HPC in bioinformatics, enhancing both the accuracy and scalability of genomic analyses [23].

TDP1 phosphorylation by CDK1 in mitosis promotes MUS81-dependent repair of trapped Top1-DNA covalent complexes

During mitosis, cyclin-dependent kinase 1 (CDK1) plays a pivotal role in maintaining genome stability by phosphorylating tyrosyl-DNA phosphodiesterase 1 (TDP1), thereby promoting the resolution of topoisomerase I-DNA covalent complexes (Top1ccs) through a MUS81-dependent DNA repair pathway. Topoisomerase I (Top1) is essential for relieving torsional stress generated during DNA replication and transcription and for facilitating the transition from mitosis to the G1 phase. TDP1 is responsible for removing covalently trapped Top1-DNA intermediates. Recent studies have demonstrated that CDK1 phosphorylates TDP1 at the serine 61 (S61) residue specifically during mitotic progression. This phosphorylation event is critical for efficient DNA repair at the onset of anaphase.

Mutation of this residue to alanine (TDP1-S61A) results in the abnormal retention of the protein on mitotic chromosomes, which in turn causes defective DNA repair, chromosomal instability, and improper mitotic progression. Furthermore, MUS81 endonuclease activity is shown to be essential for resolving Top1ccs during mitosis. Under replication stress conditions—such as those induced by camptothecin or aphidicolin the presence of the TDP1-S61A variant leads to its accumulation at common fragile sites, triggering elevated levels of MUS81-mediated chromosomal aberrations including chromatid breaks, anaphase bridges, and micronucleus formation.

These defects persist into the G1 phase, as evidenced by the formation of 53BP1-positive nuclear bodies, indicating unresolved DNA damage. Collectively, these findings underscore a cell cycle-dependent mechanism in which CDK1-mediated TDP1 phosphorylation coordinates Top1cc resolution, ensuring genomic integrity during mitosis and mitigating replication stress-induced instability [43].

Cell Cycle Synchronization at the G2/M Phase Border by Reversible Inhibition of CDK1

Chemical synchronization of the cell cycle has been a critical approach for dissecting the molecular mechanisms governing cell cycle progression, with numerous agents successfully targeting the G1, S, and M phases.

However, achieving effective synchronization at the G2 phase has remained a major challenge due to the lack of selective inhibitors specific to this stage. A significant advancement in this area emerged with the development of RO-3306, a potent and selective cyclin-dependent kinase 1 (CDK1) inhibitor that enables reversible arrest of proliferating human cells precisely at the G2/M boundary.

RO-3306 allows for highly efficient synchronization, enabling more than 95% of cycling cancer cells to be arrested in the G2 phase using a simple one-step protocol. Upon removal of the inhibitor, cells proceed synchronously into mitosis, eliminating the need for microtubule-disrupting agents typically used in mitotic arrest protocols. This reversible inhibition strategy facilitates the isolation of mitotic cells with minimal cytotoxic effects and provides a robust model for investigating the functional roles of CDK1 during the G2/M transition.

The specificity and reversibility of RO-3306 make it an invaluable molecular tool in cell cycle research, particularly for elucidating the regulatory mechanisms of mitotic entry in human cells [44].

CDK1 inhibition sensitizes normal cells to DNA damage in a cell cycle dependent manner

Cyclin-dependent kinase 1 (CDK1), a central regulator of the G2/M transition, is frequently hyperactivated in various malignancies, positioning it as a promising tumor-specific therapeutic target. However, emerging evidence suggests that CDK1 inhibition also exerts cytotoxic effects on non-malignant cells in a manner that is dependent on the cell cycle phase. Specifically, the inhibition of CDK1 has been found to sensitize normal cells to DNA-damaging agents such as ionizing radiation, raising concerns about its selectivity and safety in therapeutic applications.

A study investigating the radio sensitizing effects of the selective CDK1 inhibitor RO-3306 utilized clonogenic survival assays across three cancer cell lines (HeLa, T24, and SQ20B) and three non-transformed human cell lines (HFL1, MRC-5, and RPE). Results demonstrated that although CDK1 signaling was efficiently suppressed in both cancerous and non-cancerous cells, normal fibroblasts and epithelial cells exhibited increased sensitivity to radiation when CDK1 was inhibited. This finding contradicts earlier assumptions that CDK1 inhibition would selectively enhance radio sensitivity in tumor cells while sparing healthy tissue. Further mechanistic insights revealed that the relative resistance of normal cells to CDK1 inhibition under standard culture conditions may be due to their extended residence in the G1 phase, thereby evading vulnerability during the CDK1-dependent G2/M

transition. However, when these non-cancerous cells were synchronized in the G2/M phase, CDK1 inhibition regardless of radiation exposure resulted in a marked reduction in clonogenic survival, highlighting a cell cycle-specific vulnerability.

Collectively, these findings indicate that CDK1 inhibition compromises the viability of normal proliferating cells, particularly during the G2/M phase, thereby posing potential risks in clinical applications. This underscores the importance of carefully considering cell cycle dynamics and tissue-specific proliferation rates when evaluating CDK1 inhibitors as anti-cancer agents [45].

Cell adhesion is regulated by CDK1 during the cell cycle

Cyclin-dependent kinase 1 (CDK1) has been identified as a key regulator of cell adhesion throughout the cell cycle. In most tissues, anchorage-dependent cell proliferation and progression through the cell cycle require stable interactions between cells and the extracellular matrix (ECM), mediated via integrin-based adhesion complexes. These complexes facilitate signaling and structural support essential for cell growth. However, during the transition to mitosis, cells typically lose substrate attachment, accompanied by disassembly of adhesion complexes and cellular rounding, suggesting a mechanistic link between adhesion regulation and cell cycle progression.

Recent findings have demonstrated that CDK1, in conjunction with cyclin A2, orchestrates the spatial and temporal organization of adhesion complexes and actin cytoskeletal architecture during interphase. This coordination ensures a progressive expansion of adhesion complex area from the G1 phase to the S phase, facilitating efficient cell cycle advancement. Conversely, as cells enter G2, a reduction in adhesion complex area is observed, with disassembly initiating several hours prior to mitotic onset.

The initiation of this disassembly process is regulated by rising levels of cyclin B1 and the inhibitory phosphorylation of the CDK1–cyclin B1 complex, which collectively prime cells for mitotic entry. This cyto skeletal remodeling and adhesion complex turnover reflect a preparatory phase necessary for the morphological and mechanical changes associated with mitosis. Thus, CDK1 serves as a critical integrator of cell cycle progression and adhesion dynamics, linking cytoskeletal reorganization with timely mitotic entry [46].

An overview of Cdk1-controlled targets and processes

The budding yeast *Saccharomyces cerevisiae* has served as a valuable model organism for elucidating the fundamental roles of cyclin-dependent kinase 1 (CDK1), a key regulator of cell cycle progression. Over the past decade, substantial progress has been made in understanding the molecular mechanisms that underlie cell cycle control.

Approximately 75 substrates of CDK1 have been identified and characterized in *S. cerevisiae*, revealing their involvement in diverse and essential cellular processes, including transcriptional regulation, chromosomal segregation, DNA replication, and cell morphogenesis.

These findings underscore the central role of CDK1 not only in orchestrating cell cycle transitions but also in maintaining genomic stability. The collective evidence highlights CDK1 as a critical integrator of cellular events required for faithful cell cycle execution and genome maintenance in eukaryotic cells [39].

CDK1-Mediated Cell Cycle Arrest Induced by *Reynoutria multiflora* in Prostate Cancer Suppression *Reynoutria multiflora* (Thunb.)

Moldenke, historically known as *Polygonum multiflorum* and widely utilized in traditional Chinese medicine, has been employed for centuries in the management of diverse ailments. Recent pharmacological investigations have revealed its potential anti-cancer activity, particularly in androgen receptor (AR)-dependent prostate cancer. However, the precise molecular mechanisms underlying its therapeutic effects against advanced prostate malignancies remain insufficiently understood. This study aimed to elucidate the anti-tumor properties of *R. multiflora* and its bioactive constituents in the context of prostate cancer.

A comprehensive set of in vitro assays including cell viability, colony formation, flow cytometry (FACS), and wound healing was employed to evaluate its cytostatic and anti-migratory effects. In vivo efficacy and safety were assessed using a xenograft mouse model. Integrative network pharmacology, transcriptomic profiling (RNA sequencing), and bioinformatics analyses were conducted to explore the underlying molecular mechanisms. Additionally, molecular docking, cellular thermal shift assays (CETSA), CRISPR-Cas13 screening, RT-qPCR, and Western blotting were used to identify key molecular targets.

The findings demonstrated that *R. multiflora* significantly inhibited prostate cancer cell proliferation and enhanced the responsiveness of these cells to AR antagonists. Mechanistically, the extract induced G2/M phase cell cycle arrest via modulation of CDK1 phosphorylation.

One of its key phytochemicals, polygalacic acid, and its structural analogs, further suppressed tumor cell growth through inhibition of CDC25B, a critical phosphatase responsible for activating CDK1. Collectively, the data indicate that *R. multiflora* and its active components exert anti-proliferative effects in prostate cancer by modulating the CDC25B-CDK1 signaling axis, leading to cell cycle arrest, with demonstrable safety and efficacy in preclinical models [47].

2.1 Critical Analysis:

Based on our research strategy our work's literature review explains how CDK1 functions as a therapeutic target and a diagnostic biomarker in a number of disorders. Although CDK1 inhibition is preferable to other approaches due to its significant promise in cancer treatment, especially in endometrioid endometrial cancer, more research is required to fully understand its resistance mechanisms and off-target consequences. Given that CDK1 is implicated in many neurodegenerative diseases, this specificity begs the issue of how CDK1 targeted multiplex PCR diagnostics for Alzheimer's disease might compare to another biomarker. According to the earlier research, CDK1 participates in DNA repair, cell cycle control, and mitochondrial respiration, indicating that it has many common functions in cell biology. However, it is still difficult to find CDK1-targeting therapeutic treatments that do not impair vital cellular functions, especially in non-tumorous tissue. Furthermore, the most cutting-edge bioinformatics research servers, such as AI and HPC, may be integrated to find pathways associated to CDK1, but regulatory obstacles and clinical validation are necessary to turn this into a workable medication. Our task involves identifying and characterizing harmful synonymous SNPs in the human CDK1 protein using structural analysis and data mining. Although CDK1 targeted medicines show promise, more study and clinical studies are required to broadly validate their safety and effectiveness in CDK1 dependent applications.

Table 2.1: Shows the Summary of Literature Review on CDK1

Title	Aim	Datasets	Methodology	Findings
Multiplex PCR advances that target CDK genes for Alzheimer's disease diagnosis.	Create a multiplex PCR technique that targets CDK genes and increases the effectiveness of biomarker identification for early Alzheimer's diagnosis.	Patient Samples, Control Samples NCBI Ensembl and ClinicalData	Primer Design & Optimization Multiplex PCR Setup, for Validation Performed using gel electrophoresis, qPCR, and Sanger sequencing and Data Analysis through expression profiling and statistical evaluation	The multiplex PCR assay accurately identified CDK5 overexpression in Alzheimer's patients with high sensitivity and specificity, offering a cost-effective and rapid diagnostic tool for early detection.
Endometrioid endometrial carcinoma has a new therapeutic target in CDK1.	Evaluate CDK1 inhibition as a therapeutic strategy in endometrioid endometrial cancer, focusing on tumor progression, cell proliferation, survival, and molecular mechanisms.	Utilized tumor tissues, EEC cell lines, RNA-seq data, and xenograft mouse models to evaluate CDK1 inhibition effects in endometrioid endometrial cancer.	The methodology involved qPCR, Western blotting, gene silencing, CDK1 inhibitor testing, xenograft studies, and bioinformatics analysis to assess CDK1's role in endometrioid endometrial cancer.	CDK1 overexpression in EEC tissues was linked to tumor progression, and its inhibition reduced cell proliferation and apoptosis, highlighting CDK1 as a potential therapeutic target for EEC.
Targeting CDK1 in cancer: mechanisms and implications	To explore the mechanisms by which CDK1 (Cyclin-Dependent Kinase 1) regulates cancer progression and evaluate its potential as a therapeutic target.	genomic and transcriptomic data from cancer cell lines, patient tumor samples, and public datasets (e.g., TCGA)	The methodology involved RNA sequencing, qPCR, functional assays with CDK1 inhibition, cell proliferation and apoptosis assays, in vivo xenograft models, and bioinformatics analysis to investigate CDK1's role in cancer.	CDK1 is overexpressed in various cancers and is linked to poor prognosis, with its inhibition reducing cell proliferation, inducing apoptosis, and slowing tumor growth, making it a promising therapeutic target.

Mitochondrial respiration is coordinated by Cyclin B1/Cdk1 throughout the progression of the cell-cycle G2/M.	To look into how Cyclin B1 and CDK1 control mitochondrial respiration throughout the cell cycle's G2/M transition and how it affects the course of the cell cycle.	Data on gene expression from publically accessible datasets (such as TCGA) and mitochondrial and transcriptome information from cell lines synchronized during various cell cycle stages)	The approach included mitochondrial respiration experiments, protein expression analysis, cell cycle synchronization, functional inhibition with CDK1 inhibitors/RNAi, flow cytometry for cell cycle analysis, and measurement of mitochondrial membrane potential.	During the G2/M transition, cyclin B1/CDK1 controls mitochondrial respiration; its suppression impairs mitochondrial function, slows the course of the cell cycle, and may have therapeutic implications for addressing the metabolism of cancer cells..
The distinct roles that CDK1 plays in human cells are based on quantitative variations among cyclin-dependent kinases.	To explore the mechanisms by which CDK1 (Cyclin-Dependent Kinase 1) regulates cancer progression and evaluate its potential as a therapeutic target.	Genomic and transcriptomic data from cancer cell lines and patient tumor samples. Publicly available cancer datasets (e.g., TCGA)	The methodology involved gene expression analysis (RNA- seq, qPCR), functional assays (CDK1 inhibition, knockdown, overexpression), cell proliferation and apoptosis assays, in vivo xenograft studies, and bioinformatics pathway enrichment analysis.	CDK1 is uniquely overexpressed in specific cell types, with its distinct role in cell cycle regulation and its inhibition affecting DNA repair, cell cycle progression, and apoptosis, offering potential for targeted therapies.
CDK1-PP2A-B55 interplay ensures cell cycle oscillation via Apc1-loop	To investigate the molecular interplay between CDK1 and PP2A-B55 in regulating the cell cycle, particularly through the Apc1-loop, and its role in ensuring proper cell cycle oscillation.	The study uses human cell lines, gene expression analysis (qPCR, RNA sequencing), proteomic data (Western blotting, mass spectrometry), and public databases (e.g., TCGA) to explore CDK1- PP2A-B55-Apc1 interactions in cancer.	The methodology includes gene expression analysis (qPCR, RNA sequencing), protein interaction studies (Co-IP, mass spectrometry), phosphorylation assays (Western blotting), cell cycle analysis (flow cytometry, live-cell imaging), and functional inhibition (RNAi, small molecule inhibitors).	The study finds that the CDK1- PP2A-B55 feedback loop, involving Apc1, regulates cell cycle oscillations and disruptions in this axis impair cell cycle progression, offering potential therapeutic targets for cancer.

Analysis of Cyclin D1 gene SNPs computationally and their correlation with breast cancer	To investigate the association between single nucleotide polymorphisms (SNPs) in the Cyclin D1 gene and breast cancer susceptibility.	Publicly available genetic datasets (e.g., TCGA dbSNP) for SNP analysis. Genomic data from breast cancer patients and healthy controls.	The methodology involves identifying SNPs in the Cyclin D1 gene, performing association analysis with breast cancer incidence, using bioinformatics tools for functional impact prediction, and validating significant SNPs in independent cohorts.	Several SNPs in the Cyclin D1 gene are linked to increased breast cancer risk, potentially affecting gene expression and functioning as biomarkers for early detection or prognosis.
Cell Cycle Synchronization by Reversible CDK1 Inhibition at the G2/M Phase Border	To explore the use of reversible CDK1 inhibitors to synchronize cells at the G2/M phase boundary, with potential applications in cancer therapy.	Gene expression and protein data	The methodology involves reversible CDK1 inhibition, flow cytometry for cell cycle analysis, gene/protein expression analysis via qPCR and Western blotting, and functional assays for cell proliferation, apoptosis, and DNA damage response.	Reversible CDK1 inhibition effectively synchronizes cells at the G2/M phase, enabling precise control over cell cycle progression without causing permanent arrest or widespread cell death.
Normal cells become more sensitive to DNA damage when CDK1 is inhibited, and this sensitivity is cell cycle dependent.	To investigate the molecular interplay between CDK1 and PP2A-B55 in regulating the cell cycle, particularly through the Apc1-loop, and its role in ensuring proper cell cycle oscillation.	Gene expression data from normal human cell lines with varying levels of CDK1 activity. Protein expression data for CDK1 and DNA damage markers in different phases of the cell cycle. Publicly available cancer datasets (e.g. TCGA)	Selective CDK1 inhibitors and RNAi were used to reduce CDK1 expression, followed by DNA damage induction, cell cycle analysis, DNA damage assays, and assessment of cell viability and apoptosis.	CDK1 inhibition sensitizes normal cells to DNA damage by impairing DNA repair and increasing apoptosis, with potential therapeutic implications for enhancing cancer treatments.
Throughout the cell cycle, CDK1	To investigate the role of CDK1 in regulating cell	Gene expression, protein expression (Western	CDK1 modulation (inhibition/overexpression), cell	CDK1 regulates cell adhesion molecule expression and localization,

controls cell adhesion.	cell adhesion during the cell cycle.	blotting), and publicly available datasets to examine CDK1 and cell adhesion molecule associations.	adhesion assays, immunofluorescence, flow cytometry, and functional assays to study CDK1's role in cell adhesion.	and its inhibition impairs adhesion, migration, and tissue integrity, linking cell cycle progression with adhesion regulation.
An outline of the targets and processes under Cdk1 regulation	To provide a comprehensive overview of the targets and processes regulated by CDK1 during cell cycle progression, particularly focusing on its role in cell division and mitotic events.	Gene expression data from cell lines and tissue samples with varying CDK1 activity, protein expression profiles from mass spectrometry and Western blotting, and TCGA databases for CDK1-related gene expression in cancers.	Gene expression analysis through RNA sequencing and qPCR, proteomic studies with mass spectrometry and Co-IP, functional assays using RNAi and inhibitors, flow cytometry for cell cycle analysis, and xenograft models for in vivo assessment.	CDK1 regulates key mitotic processes, including chromosome segregation and cytokinesis, with its dysregulation linked to cancer, suggesting therapeutic potential for CDK1 inhibitors.
Survivin Activates the Cdk1-Cdc25B Axis to Mediate Mitotic Onset in HeLa Cells	To investigate the role of Survivin in regulating the Cdk1-Cdc25B axis during mitotic onset in HeLa cells.	HeLa cell line (cancer cell model). Gene expression data on Survivin, Cdk1, and Cdc25B from qPCR and western blot analysis.	The methodology involved qPCR, Western blotting, RNA interference, immunofluorescent once, and flow cytometry to assess gene expression, cell cycle progression, and localization of Survivin, Cdk1, and Cdc25B during mitosis.	Survivin activates the Cdk1-Cdc25B axis to regulate mitotic onset in HeLa cells, and its knockdown disrupts cell cycle progression, highlighting Survivin as a potential therapeutic target for cancer.

CHAPTER 03
MATERIALS & METHODS

This chapter outlines the methodological framework adopted in this study to investigate the impact of non-synonymous single nucleotide polymorphisms (nsSNPs) on the structure and function of the CDK1 protein. A combination of computational and bioinformatics tools was utilized to analyze genetic variations, predict their deleterious effects, and assess structural and functional consequences. The study followed a systematic pipeline involving data retrieval, sequence analysis, pathogenicity prediction, structural modeling, and functional annotation. Each step of the analysis is described in detail to ensure reproducibility and transparency of the research process. Complete workflow of this study is represented in Figure 3.1.

3.1 SNPs Databases

We acquired the SNP data sets from the ENSEMBL, NCBI dbSNPs, UniProt, and Gene Card databases in order to filter and choose the functional nsSNPs in CDK1. Ensembl regularly shares and updates the dbSNP data. For data mining and further examination of recovered data, Ensembl is an important tool [48]. When the National Institute of Genomic and Advanced Biotechnology Information (NCBI) submit new entries, dbSNP is often updated [49]. The information that is shown in these two databases is essentially identical. Gene Card and UniProt primarily retrieve variant data, similar to dbSNP or Ensembl.

3.2 Processing of Deleterious nsSNPs

We employed the SIFT online tool to assess the functional consequences of non-synonymous SNPs (nsSNPs), categorizing them based on their likelihood to disrupt or preserve protein structure and activity.

SIFT

Sorting Intolerant from Tolerant is an algorithm that predicts how changes in amino acids or nucleotides may affect a protein or gene. SIFT is used in many investigations related to human genetics research and diseases. Given the SNPs' RefSeq IDs as input, it classifies a SIFT score below 0.005 as harmful. By comparing homologous protein sequences and physical characteristics, the SIFT method assesses how nsSNPs affect a protein's structure and function. To verify homologous sequences, the TrEMBL and SWISSPROT algorithm databases were used. SIFT analyzes conserved amino acid positions by aligning related protein sequences and assigning a Tolerance Index (TI) value between 0 and 1, with lower scores suggesting a greater likelihood of functional disruption due to substitution. When an in-tolerance location is identified, the TI score decreases, and vice versa [70]. In this tool SNPs are categorized as either tolerant or intolerant based on homologous alignment.

Any alteration to these important amino acids is seen as deleterious. An amino acid is expected to be tolerant if its predicted probability value is less than 0.05, or below the normalization probability threshold, and intolerant if its anticipated probability value is larger than 0.5. Higher value proteins will be less functionally affected by SNPs [50]. SIFT evaluates amino acid conservation by aligning evolutionarily related sequences and assigns a Tolerance Index (TI) score ranging from 0 to 1, where lower values indicate a higher probability of a deleterious effect. Highlighted are the in tools/databases and techniques used at every phase, ranging from data mining to structural and functional investigation of the mutant proteins. The techniques or resources for predicting harmful SNPs in the human CDK1 are covered in this chapter. Complete work flow of our methodology is shown below Figure 3.1.

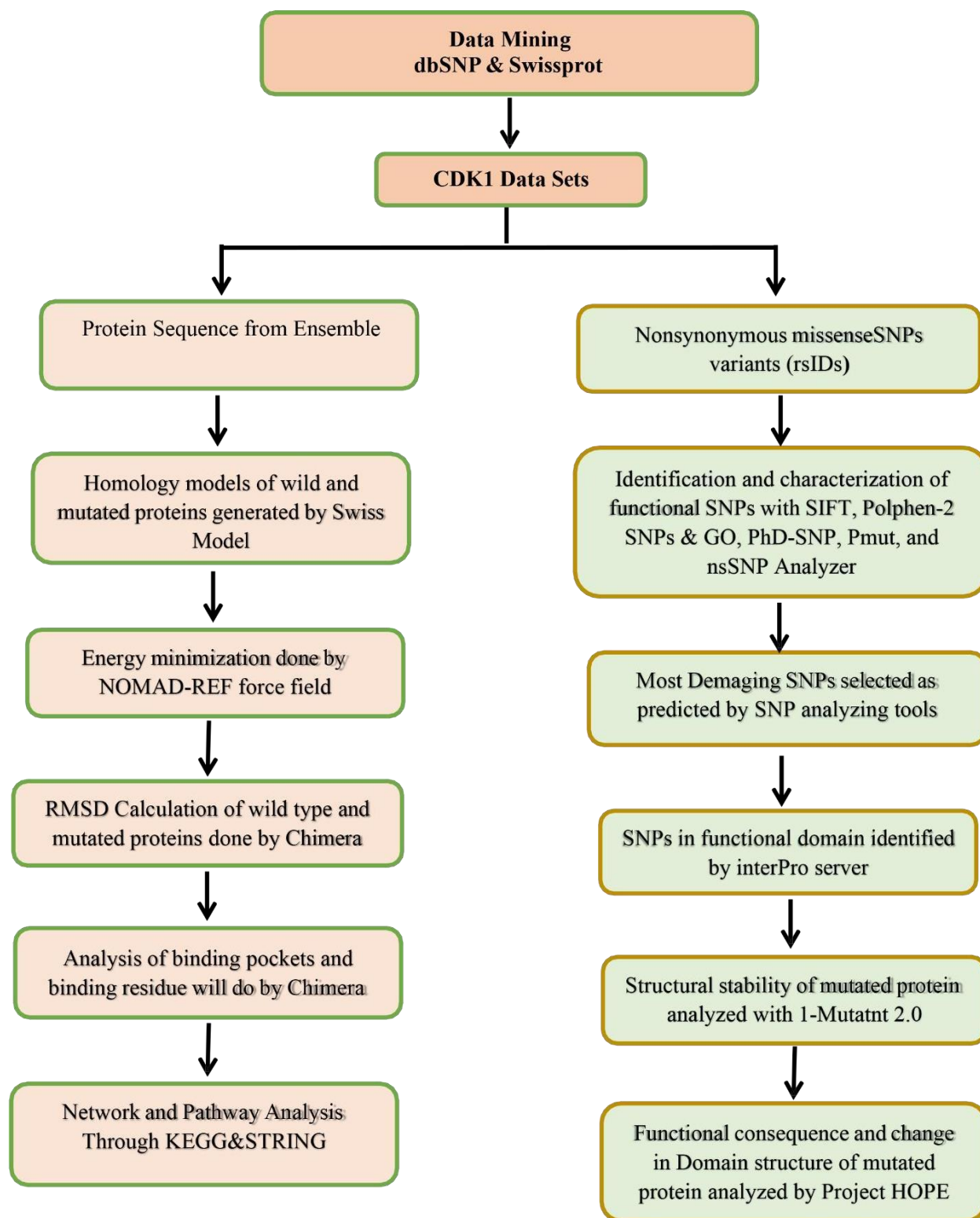


Figure 3.1: Complete workflow of this study is represented.

3.3 Impact of coding nsSNPs Assessed Through Structure-Based Homology

Analysis

Protein function is intrinsically linked to its three-dimensional structure, which is determined by its amino acid sequence. Alterations in this sequence can induce structural modifications that potentially disrupt functional integrity. In this study, PolyPhen-2 was employed to predict the deleterious effects of no synonymous SNPs (nsSNPs) on the structure and functional properties of the CDK1 protein.

PolyPhen2

Polyphen2 uses the native Bayesian classification method with a score of (0 1). According to a score, a mutation is either likely or potentially damaging; for instance, a value close to 1 would indicate that the protein structure is most likely severely damaged [51].

3.4 Characterization of Functional nsSNPs

Functional nsSNPs were found utilizing web-based programs such as phd-SNP, SNPS&GO, PANTHER, P-MUT, and PROVEAN.

PANTHER

PANTHER (Protein analysis through evolutionary relationships) is a sequence-based SNP analysis approach that uses evolutionary information and functional linkages in the protein sequence profile to predict SNPs. Considering predictions that are more than 0.5, 0.1, or 0.5 to be probable and probably harmful, respectively [93].

PhD-SNP

The Support Vector Machine is the foundation of the Predictor of Human Deleterious Single Nucleotide Polymorphism. This method may be used to use protein sequence data to ascertain if a new phenotypic nsSNP is linked to a human genetic disorder. Reliability index (RI) scores are used by PhD-SNP to forecast outcomes. Neutral and disease outcomes are kept apart. In the event that the probability scores are higher than 0.5 (<http://snps.biofold.org/phd-snp/phd-snp.html>) [5]. A correlation between the mutation and the disease is anticipated.

MutPred

MutPred2 is a machine learning-based method for estimating the likelihood of a harmful amino acid replacement. MutPred can accurately identify the function sites and structural fate of both wild type and mutant proteins. It determines generic probability ratings for every conceivable amino acid replacement. Variants with scores higher than 0.75 are classified as very hazardous, while those with scores higher than 0.5 are accused of having the potential to significantly alter protein function.

The P-value of the top five molecular characteristics is likewise provided by MutPred2. A p-value suggests that there's a chance an observed the discrepancy happened by accidently. A substitution is considered benign if the P-value is greater than 0.05 and harmful if it is less than 0.05 [88]. G is the MutPred mutation's G value, which shows how this particular mutation will affect the protein's structure and functionality. The effect is larger if the G value is included, and it should be between 0 and 1 or greater than 1 [52].

PROVEAN

A web-based program called Protein Variation Effect Analyzer (PROVEAN) can be used to analyze how changes in amino acids impact a protein's functionality. Prediction score values that are less than or equal to -2.5 are considered harmful. Provean uses the NCBI and the BLAST query sequence of mutations to classify mutations as either neutral or harmful using a cutoff value [53].

SNP&GO

The SVM algorithm is used by SNP&GO. Based on protein data, SNPs&GO determines the probability that a single SNP will be connected to a human illness. The likelihood of a correlation between the mutation and the disease is anticipated if the probability scores are more than 0.5.

P-MUT

In P-MUT, neural networks are used. Then, using the input sequence information that was acquired, it analyzes a portion of the sequence information. The mutants are categorized by P-MUT. Therefore, P-MUT was used to classify the mutations by using P-MUT on the query sequence. P-MUT is used to pass the query sequence in FASTA format, and the result for seed can be either diseased or neutral [54].

3.5 Prediction of the Impact of nsSNPs on the Molecular Phenotype of Proteins

The SNP effect online tool was employed to predict mutations induced by nsSNPs within the coding region of the CDK1 gene, potentially leading to alterations in protein structure and function. By examining the query protein's homeostasis landscape, this web-based program determined the conservation score [55]. The SNP effect employs a number of techniques for various tasks, including LIMBO to forecast chaperon binding, WALTZ to predict amyloid propensity, and TANGO to predict a mutant's propensity to assemble. To measure aggregation propensity, TANGO computes the scores of wild type and mutant residues using a unique TANGO score (dTANGO). The protein's amyloid-forming region was predicted using dWALTZ. The dLIMBO score served as the basis for estimating binding propensity.

Additionally, we obtained high resolution crystal structures from the Protein Data Bank (PDB) and applied the mutant structures of the SNP impact to FOLDX. Three categories are created based on the scores indicated above. The range of scores is -50 to 50. A score over 50 indicates a higher tendency. The propensity value should be zero, nevertheless, if the score is between - 5 and 50 and less than or equal to -50 or more than -5.

3.4 Impact of nsSNPs on protein stability

In order to solve the structure and function of proteins, protein stability is dynamic. I-MUTANT 3.0, a technique for predicting mutation changes in the stability of the CDK1 protein, was used. I-MUTANT was used to verify the findings after they were compared with MUPRO. The same technique, which relied on the SVM to determine whether stability increased or decreased with mutation, was used by both of these devices [10]. The SRide server was used to test the anticipated stabilization residue against both wild type and mutant proteins.

Mupro

This online server, which assesses the impact of single-site amino acid change on protein stability, was developed using support vector machines and neural networks. If a protein is less stable, its Mupro score will be smaller than 0 [96].

I-Mutant

An I-mutant may be used to assess the effect of non-synonymous SNPs on protein structural stability. This software determines the free energy change (DDG) caused by changes to the amino acid residues in a certain protein sequence. If $DDG > 0$ kcal/mol, both stabilities will be anticipated to rise; if $DDG < 0$ kcal/mol, both stabilities will be predicted to decline. We supplied the protein's position AAS, variant residue, and FASTA sequence as input [95]. Ultimately, CDK1's stability profile was identified. I-Mutant was used to compare the stability of the wild and mutant structures in order to achieve the aforementioned goal. First, the results of I-Mutant's production are derived from the proThern database, which provides results from thermodynamics experiments in the form of variations in the Gibbs free energy of protein stability [74]

3.5 Finding Functional SNPs in the Conserved Area

The ConSurf software measures the evolutionary conservation of amino acid locations in a protein by analyzing the evolutionary dynamics of amino acid changes between homologous sequences. The conservation score range of 1 to 9 is represented by the color scheme. Variable (or 1–3), intermediate (or 5–6), or highly preserved (or 7–9) are terms used to define such a place [56].

Both enzymatic and interaction sites are necessary for the protein to function. Amino acid change to these sites was more damaging than to other sites because they are strongly conserved across enzymatic and interface sites [57].

3.6 Predicting Structures

Understanding protein functions such as binding affinity, protein-protein interactions, and other significant characteristics both with and without mutation may be substantially aided by 3D visualization. Nuclear magnetic resonance spectroscopy (NMR) and X-ray crystallography are then used to experimentally identify a protein's three-dimensional model. The Protein Data Bank (PDB) has received the structures derived from these two approaches [58]. It takes a lot of money and effort to determine the 3D model via trials. Computational approaches provide an alternative to experimental methods for predicting 3D models of proteins. Computational prediction methods are rapid, inexpensive, and efficient in terms of time [59].

Our target protein's three-dimensional structure is listed in the Protein Data Bank (PDB). Swiss Model a web-based tool was used to predict the query protein as a 3D structure [60].

3.7 Refinement of Protein Model

As a result, tertiary structures produced using experimental or computational methods must be formed in local stereochemistry. Stereochemistry predicts certain values. The torsion angles that bind angles should fall within that exact range.

International conflicts should be avoided, and the distribution of polar and hydrophobic residues should be predicted, since the ring side chain must maintain planarity conflicts [61]. The CDK1 protein's predicted structure was submitted to locPREFMD for refinement. Thus, to improve the quality of the local and global structures of the predicted model, the protein predicted model's structure is improved [62].

3.8 3D Structure Confirmation

We have used many tools and services, such as ERRAT verify 3D and Ramachandran plot, to confirm our predicted 3D model of the CDK1 protein. Therefore, it is necessary to determine if 3D was used and whether it was done to detect misfolded areas or to confirm correct confirmation in the application that uses the molecular level and protein structure [63]. The angles phi and psi plot of the amino acids found in the protein of interest were shown by Ramachandran. Identification of the principal torsion seen in the Ramachandran plot [64] confirms the protein's stereochemistry, regardless of its use. Proteins were verified and their structural quality assessed using the ERRAT technique.

3.9 Site-directed mutagenesis and in-silico RMSD Computation

Based on the homology modeling approach, this was accomplished via the use of the program TRITON interfaced with MODELLER [66 and 67]. Each model's stereochemistry quality was carefully predicted using PROCHECK [68]. Meanwhile, the environmental profile of the mutant models was studied using ERRAT and CHIMERA [69] to lower their potential energies. Additionally, NOMAD-REF was employed.

CHIMERA

There are several possibilities for sequence and structural analysis in the UCSF Chimera molecular graphics tool. The core portion offers the fundamental services and visualization, whereas the extended or extensions portion consists of the extensions themselves that provide the majority of the higher-level content. In this research, we evaluate and compute utilizing the Matchmaker and Match Align of Chimera programs to determine the RMSD value of the wild type and mutant structures [90].

3.10 Forecasting the Impact of Mutations on Structure

Lastly, the project HOPE tool is used to assess the structural impact of mutations on the CDK1 protein in order to do further research.

HOPE

One of the upcoming web servers that serve as a single point mutation analyzer is Project HOPE (Have (y) Our Protein Explained). Additionally, this server was used to absorb the impact of mutations on the structure of the CDK1 protein and to perform molecular dynamic modeling.

In order to offer a mutation report with visuals that illustrate the impact of the mutation, HOPE gathers data from a variety of sources. The HOPE server receives input in the form of FASTA files, performs BLAST against the PDB, and then utilizes YASRA to predict the model. This website collects data from a variety of databases, including Uniprot sequence sequences and active site, motif, and domain data and 3D structure data that is accessible in the PDB [89].

3.9 Pathway and Network Analysis Tools and Software for Bioinformatics

Using a range of bioinformatics tools and resources, the CDK1 protein interaction network was retrieved from the STRING database.

3.9.1 CDK1 PPI network retrieval using the STRING database

The biological network was retrieved using the biological database STRING (String: Search Tool for Retrieving Interacting Genes/). The string then includes PPI interactions that are known and expected. The string is a publicly accessible, often updated source database that contains public, computational, and experimental text. Additionally, the site highlight's function enrichment in the user-provided protein list. Over 9.6 million proteins from over 2000 species are included in the most recent version [92]. These clusters were composed of the network's closely connected proteins. As shown in Figure 4.22, arrows in PPI network clusters indicate positive action effects, lines indicate negative action effects, and balls indicate nonspecific effects.

3.9.2 KEGG

A database resource for the depiction and study of biological systems is called KEGG. Sixteen manually curated databases for different data objects make up the component. These databases are divided into four categories: (i) molecular network system information, (ii) genomic information, (iii) chemical information, and (iv) health information gathered in the event of perturbed systems. Each data item is identified by the term child, and there are two different kinds of KEGG IDs. The fundamental form of all KEGG-original datasets is a five-digit number that comes after a dataset-dependent prefix. The combined form with the dataset and entry name separated by a colon is used for datasets introduced as part of entering from external sources.

The three manual reference objects Britte hierarchies, KEGG modules, and pathway maps are the basis for the expansion of PATHWAY, BRITE, and KEGG modules from molecular network objects to organism-specific molecular network objects in the systems information domain. For instance, the reference metabolic pathways (map01100) [91] are substituted with the human metabolic pathways (hsa01100) [91].

CHAPTER 04
RESULTS & DISCUSSION

4.1 Retrieval of Single Nucleotide Polymorphisms (SNPs) Dataset

In order to collect the SNPs data of the CDK1 protein, a number of distinct bioinformatics programs and databases were utilized. These included Ensemble genome browser, Gene card, Uniprot, and NCBI db-SNP. In order to get this knowledge, something was done in order to attain it. SNPs that may be found in the exons of the CDK1 protein are where the majority of our attention is being directed. Because the methodology of the study mandated that this was the only type of variation to be looked at, nsSNPs were the only ones to be chosen for further investigation into how they alter the structure and function of the CDK1 protein. This was done so in order to determine how these variations alter the structure and function of the CDK1 protein. The database was searched, and a total of 297 non-synonymous SNPs were found after passing through the screening process. During the process of extracting the data, it was discovered that the intron area of the protein, which was then followed by the 3-UTR region, had most single-nucleotide polymorphisms. This was followed by the 5'-untranslated region (SNPs) as shown in the figure below Figure 4.1.

4.2 Analysis of Non synonymous (ns) SNPs

To assess the potential structural impact of amino acid substitutions arising from mutations, 297 nsSNPs in the CDK1 protein were analyzed using the Sorting Intolerant From Tolerant (SIFT) algorithm. This tool estimates a Tolerance Index (TI) score between 0 and 1 for each variant, reflecting the evolutionary conservation of the altered residue across homologous protein sequences. A lower TI score indicates a higher likelihood that the substitution is intolerant and thus potentially deleterious to protein function, whereas a higher score suggests a more tolerated change.

Out of the 297 nsSNPs evaluated, 32 were identified as deleterious with TI scores ranging from 0.00 to 0.03. Among these, 28 mutations exhibited the minimum score of 0.00, indicating a strong likelihood of disrupting protein function. Four mutations had a TI score of 0.01, one had a score of 0.02, and five recorded a score of 0.03. Each of these 32 nsSNPs involved a single nucleotide substitution that resulted in a change in the corresponding amino acid, potentially altering the protein's structural integrity that show in table 4.1.

These 32 deleterious nsSNPs were subsequently analyzed using a panel of additional predictive tools, including PolyPhen-2, SNP&GO, PANTHER, PhD-SNP, PROVEAN, and P-Mut. Based on consensus predictions across all six tools, six nsSNPs (rs1460089100, rs1179034229, rs2080301489, rs1322570597, rs746720738, and rs1234513195) were consistently classified as deleterious. These variants were identified as likely to contribute to disease or significantly disrupt protein structure. Therefore, they were selected for further in-depth analysis that shows in Table 4.1.

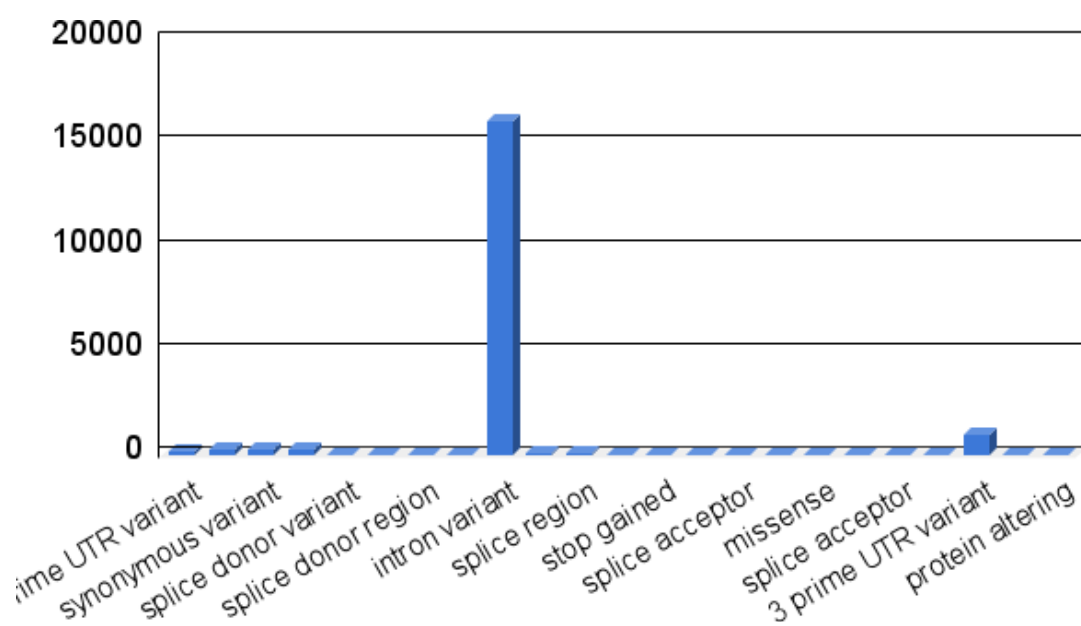


Figure 4.1: Distribution of Single Nucleotide Polymorphisms across the Human CDK1 Protein Coding Region.

Table 4.1: The most deleterious nsSNPs were identified, with steric symbols marking the six selected positions in the CDK1 protein structure.

Variant ID	Allel change	A.A change	SIFT	Plephen	SNP& GO	Panther	PHD SNPs	Provean	P-Mut
rs771969255	G/T	V30L	0.01	0.958	Disease	Neutral	Neutral	Deleterious -2.677	Disease
rs768015814	A/T	T47S	0	0.986	Neutral	Disease	Neutral	Deleterious -3.867	Disease
rs1230964680	G/C	E57Q	0.01	0.977	Neutral	Disease	Neutral	Deleterious -2.958	Neutral
rs753490763	A/G	I193V	0.03	0.998	Neutral	Disease	Neutral	Neutral -0.949	Neutral
rs373718394	T/S	L234S	0	0.984	Disease	Neutral	Neutral	Deleterious -5.377	Disease
rs1260427336	G/T	D291Y	0	0.949	Neutral	Neutral	Disease	Deleterious -4.912	Disease
rs768015814	A/T	T47S	0	0.986	Neutral	Disease	Neutral	Deleterious -3.867	Neutral
rs1230964680	G/C	E57Q	0.01	0.977	Neutral	Disease	Neutral	Deleterious -2.958	Neutral
rs753490763	A/G	I136V	0.03	0.998	Neutral	Neutral	Neutral	Neutral- 0.529	Neutral
rs1260427336	G/T	D291Y	0	0.949	Disease	Neutral	Neutral	Deleterious -4.912	Disease
*rs1460089100	C/T	L66F	0	1	Disease	Disease	Disease	Deleterious -3.911	Disease
*rs1179034229	T/G	I10M	0	0.673	Disease	Disease	Disease	Deleterious -2.680	Disease
*rs2080301489	A/T	I35F	0	0.457	Disease	Disease	Disease	Deleterious -3.889	Disease

rs780626115	C/T	L70F	0	0.866	Neutral	Disease	Neutral	Deleterious -2.897	Neutral
*rs1322570597	C/T	R123C	0	0.069	Disease	Disease	Disease	Deleterious -7.380	Disease
rs2080365738	A/T	D257V	0.03	0.672	Neutral	Neutral	Disease	Deleterious -7.001	Disease
rs2080365841	T/C	L261S	0	0.153	Disease	Neutral	Neutral	Deleterious -3.905	Neutral
rs1382290424	C/T	P285L	0.01	0.547	Neutral	Neutral	Disease	Deleterious -6.715	Neutral
rs1468681761	T/C	I7T	0.03	0.018	Neutral	Neutral	Neutral	Deleterious -3.944	Neutral
*rs746720738	G/A	E40K	0.03	0.7	Disease	Disease	Disease	Deleterious -3.784	Disease
rs2080362913	G/A	G177R	0	0.993	Disease	Disease	Disease	Deleterious -7.583	Neutral
rs1322570597	C/T	R180C	0	0.069	Neutral	Disease	Disease	Deleterious -6.161	Neutral
*rs1234513195	T/C	V185A	0.02	0.695	Disease	Disease	Disease	Deleterious -3.666	Disease
rs1271439313	C/G	Q235E	0	0.021	Neutral	Neutral	Neutral	Neutral -0.497	Neutral
rs1198047745	G/T	K238N	0	0.145	Neutral	Neutral	Neutral	Deleterious -4.336	Neutral
rs771969255	G/T	V30L	0	0.799	Neutral	Neutral	Disease	Deleterious -2.677	Neutral

We made use of a server that is known as Mutpred, which is a standalone application that was developed to evaluate whether changes in amino acid levels in people are safe or detrimental. Because of this, we were able to develop a prediction on the role that the six harmful nsSNPs that were found play in human diseases and their link to proteins. More specifically, the focus of our prediction was on the manner in which these nsSNPs are associated with proteins. In addition to information on post-translational modifications and enzyme active sites, it makes use of characteristics such as molecular recognition features (MoRFs), solvent accessibility (SA), and changes in the stability of the protein. The results that are produced by MutPred each have a p score as well as a general score connected with them, and these scores may be compared to one another. The confidence level score that was indicated by a number of different features inspired the development of a number of different theories. Three of these hypotheses qualify as actionable hypotheses, meeting the criteria of having a confident hypothesis that falls within the range of ($p < 0.01$ and ($g > 0.75$). These hypotheses represent the hypothesis with the highest level of confidence according to MutPred ($p < 0.01$, $g > 0.75$). The mutation I316T exhibited the lowest p value ($p = 0.302$) for the characteristic of gain of ubiquitination, while having the mutation I35F display the maximum g value in the range of 0.807 for the characteristic of gain of ubiquitination. These results are based on the study that we conducted. For the stability loss that is typical of nsSNPs, the mutation I346T displayed the highest g value, which was 0.78, and the lowest p value, which was 0.0358 as shown in the Table 4.2.

Table 4.2: Assessment of Deleterious Mutation Probability.

Mutation	Probability Mutation	deleterious	Top 5 Features
L66F	0.688		Gain of Strand (Pr = 0.28 P = 8.2e-03)
I10M	0.344		-
I35F	0.807		Altered Ordered interface (Pr = 0.28 P = 0.04); Gain of Relative solvent accessibility (Pr = 0.25 P = 0.03); Loss of Methylation at K33 (Pr = 0.22 P = 4.4e-03); Loss of Acetylation at K33 (Pr = 0.21 P = 0.04); Loss of SUMOylation at K33 (Pr = 0.20 P = 0.03); Altered Metal binding (Pr = 0.15 P = 0.03)
R123C	0.693		Altered Metal binding (Pr = 0.35 P = 0.01); Gain of Allosteric site at H126 (Pr = 0.30 P = 1.9e-03); Gain of Catalytic site at H120 (Pr = 0.15 P = 0.02)
E40K	0.708		Gain of Acetylation at E40 (Pr = 0.42 P = 9.4e-04); Gain of SUMOylation at E40 (Pr = 0.36 P = 1.9e-03); Gain of Relative solvent accessibility (Pr = 0.28 P = 0.02); Loss of B-factor (Pr = 0.26 P = 0.03); Loss of Catalytic site at E42 (Pr = 0.26 P = 3.9e-03); Gain of Ubiquitylation at E40 (Pr = 0.18 P = 0.02); Altered Metal binding (Pr = 0.16 P = 0.04); Gain of Methylation at E40 (Pr = 0.16 P = 0.01)
V185A	0.622		Altered Metal binding (Pr = 0.28 P = 3.9e-03); Altered Ordered interface (Pr = 0.28 P = 0.04); Loss of Strand (Pr = 0.27 P = 0.02); Loss of Allosteric site at Y181 (Pr = 0.25 P = 0.01); Loss of Relative solvent accessibility (Pr = 0.24 P = 0.04); Gain of Catalytic site at W188 (Pr = 0.22 P = 9.0e-03); Altered Transmembrane protein (Pr = 0.18 P = 9.0e-03)

4.3 Nature of Selected nsSNPs

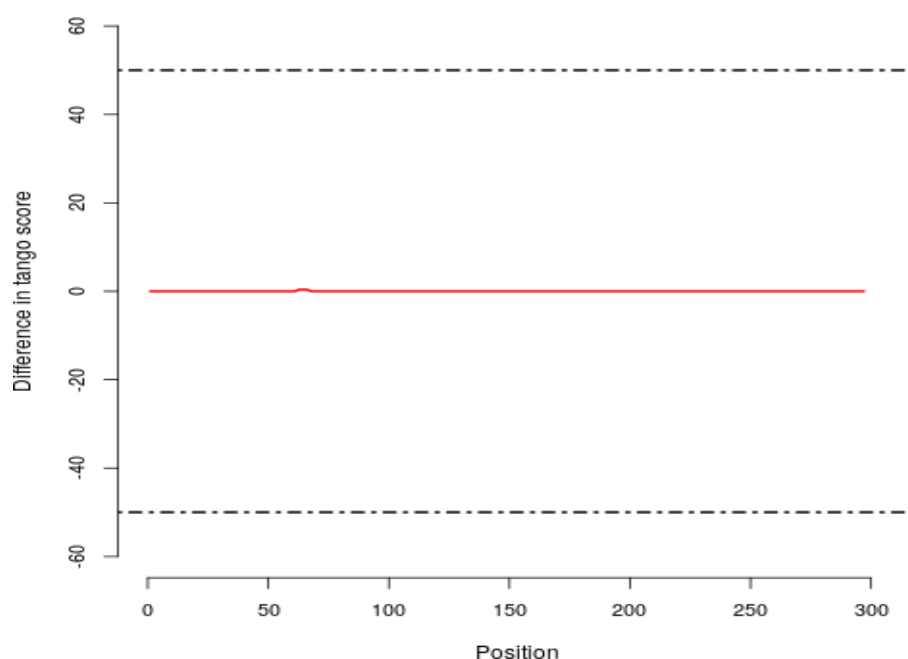
To evaluate the functional consequences of the selected non-synonymous single nucleotide polymorphisms (nsSNPs), the SNPeffect database was utilized. This tool accepts primary protein sequences as input and predicts the impact of SNPs on various structural and functional properties. Specifically, SNPeffect categorizes the effects into four main domains: chaperone-binding prediction, aggregation propensity, amyloid formation, and protein stability analysis. These predictions are generated using integrated computational tools, including TANGO (for aggregation), WALTZ (for amyloid formation), LIMBO (for chaperone binding), and FoldX (for stability prediction).

Analysis of the six selected nsSNPs revealed that only two variants significantly reduced the aggregation tendency of the CDK1 protein. No considerable effects were observed in relation to chaperone-binding capacity. Regarding amyloid propensity, two variants exhibited opposing trends one increased and the other decreased amyloid formation potential, as summarized in Table 4.3.

We also compared the aggregation score for wild type to the estimated delta Tango since the reported aggregate score for both of the mutants, I35F and E40K was higher than what was predicted. This was done as a result of the fact that both of these mutants got greater scores than was anticipated. Because the aggregate score that was achieved for both mutants was greater than that which was anticipated, this action was performed because of this finding. According to the results of the study, both mutant forms of delta TANGO exhibited an overall difference that ranged from -35 to -45. This suggests that a specific location in CDK1 may be to blame for producing a tendency toward aggregation. At this specific site, both of the several mutant versions of delta TANGO may be discovered (Figure 4.2). It has been proposed that either a rise in the quantity or a decrease in the amount is something that ought to take place in this case as a potential course of action. Proteins are extremely sensitive to a diverse range of alterations, some of which have the potential to create issues in living organisms. Both amyloidosis and protein aggregation are ailments that can be caused by irreversible beta-strand abnormalities. Both amyloidosis and protein aggregation impose a significant burden on the bodies of the animals who are afflicted by these conditions. The substance known as amyloid is a kind of protein. Because of these properties, the protein is no longer able to execute the function for which it was designed, which causes the cell to be subjected to a considerable load as a result. In addition to this, the protein exerts a load on the entire body and causes it to work harder than it normally would. For these activities to take place, a significant quantity of energy is necessary. This energy is required for the expression of genes as well as the breakdown of damaged proteins.

Table 4.3: Predicted functional impact scores of nsSNPs identified in the CDK1 protein.

Variant ID	A.A change	TANGO score	Aggregation tendency	WALT Z	Amyloid propensity	LIMBO	Chaperone binding tendency
rs1460089100	L66F	1.97	No effect	88.31	Increase	0	No effect
rs1179034229	I10M	0.22	No effect	-0.22	No effect	0	No effect
rs2080301489	I35F	0.00	No effect	0.00	No effect	-2.97	No effect
rs1322570597	R123C	0.22	No effect	-0.01	No effect	0	No effect
rs746720738	E40K	-1.53	No effect	0.18	No effect	-5.69	No effect
rs1234513195	V185 A	-0.14	No effect	0.23	No effect	0	No effect

**Figure 4.2: Comparison of aggregation propensity (Δ TANGO) between wild-type and L66F mutant CDK1 protein.**

This graph illustrates the residue-wise TANGO aggregation score, representing the difference in aggregation propensity between the wild-type (WT) CDK1 protein and its mutant variant L66F. The x-axis displays the position of amino acid residues from the N-terminus to the C-terminus, while the y-axis represents the change in TANGO aggregation scores. A flat line across the plot suggests that the L66F mutation does not affect the aggregation profile of the protein. In contrast, positive peaks indicate regions where the mutation increases the aggregation tendency, whereas negative peaks reflect a decrease in aggregation propensity due to the mutation. In the case of L66F, the presence of negative peaks highlights its role in reducing the aggregation potential of specific regions within the CDK1 protein.

The LIMBO chaperone-binding score was analyzed on a per-residue basis to compare the wild-type (WT) CDK1 protein with its E40K mutant. The graph displays the differences in LIMBO scores from the N-terminus to the C-terminus, highlighting changes in chaperone-binding propensity induced by the mutation. A flat line in the plot suggests that the E40K variant does not affect any chaperone-binding regions within the CDK1 protein, as shown in Figure 4.3. In contrast, positive peaks indicate an increased likelihood of chaperone interaction at specific residues due to the mutation, while negative peaks reflect a reduction in chaperone-binding propensity.

The WALITZ aggregation score graph illustrates a comparison between the wild type protein and the mutant variant of the protein, plotting the findings per residue. Moving from left to right across the graph displays all of the LIMBO scores as well as the disparities between those scores and those obtained from the N-C terminal. In this instance, the fact that the line does not have a slope implies that the variance does not have any effect on the aggregate profile. Because of this mutation, the plot displays positive peaks that indicate an increase in chaperone binding, and the plot displays negative peaks that show a decrease in amyloid propensity. Both of these outcomes are the consequence of the mutation.

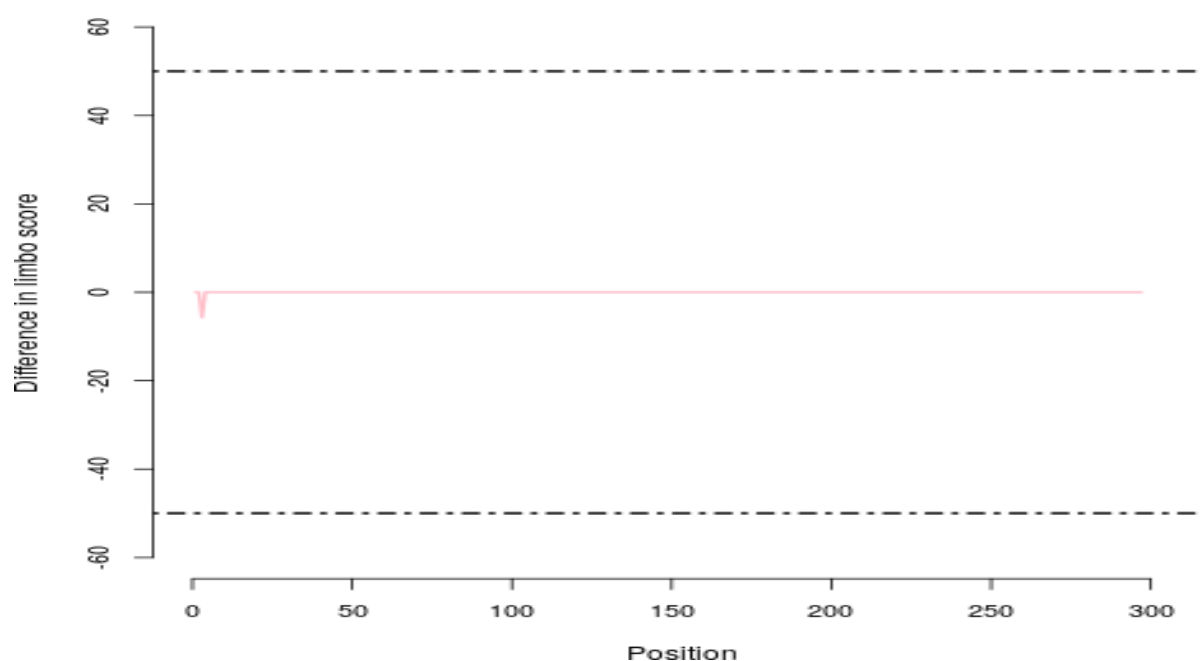


Figure 4.3: Predicted variation in chaperone-binding affinity of CDK1 between wild-type and E40K mutant using LIMBO analysis.

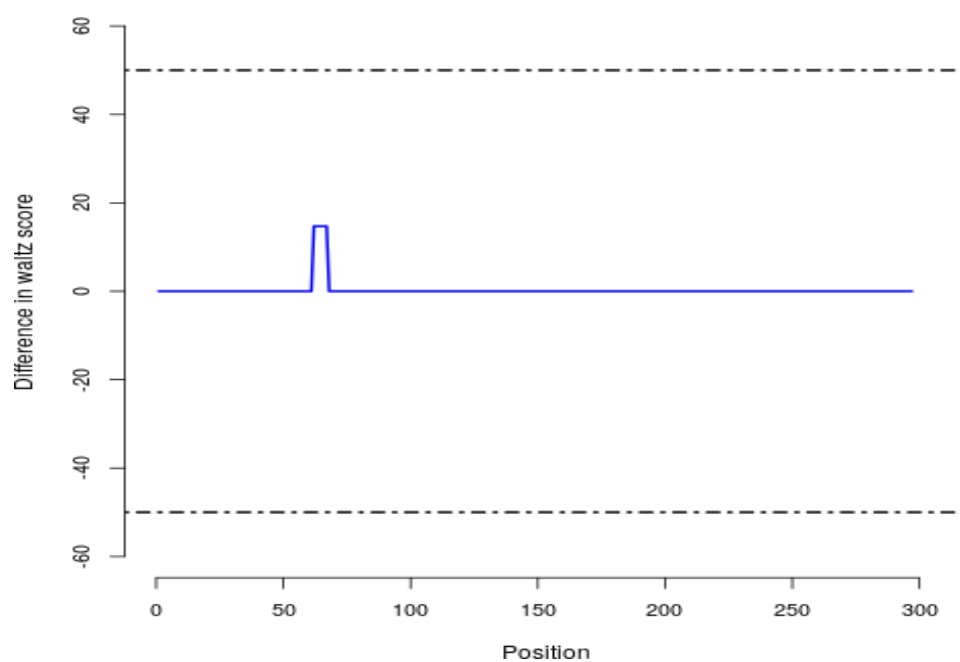


Figure 4.4: Comparative WALTZ-predicted amyloid propensity of wild-type and L66F mutant CDK1.

4.4 Conservation and stability Profiles of selected Non synonymous SNPs

To identify functionally and structurally important regions within the CDK1 protein, conservation analysis was performed using the ConSurf server. This tool assesses the evolutionary conservation of amino acid residues by aligning homologous sequences and mapping conservation scores onto the protein structure.

Although the CDK1 protein exhibits some degree of sequence variability across its family, the six selected non-synonymous SNP (nsSNP) positions displayed distinct conservation patterns. Of these, five residues were classified as buried, while V185A was found to be exposed on the protein surface. Based on ConSurf's conservation scale, all six residues were highly conserved, indicating potential functional or structural importance.

Each of the six mutation sites was located and visually represented with arrows in figure 4.5. According to previous studies, residues that play a crucial role in protein structure or function or both tend to exhibit high conservation throughout evolution, which supports the significance of these selected positions.

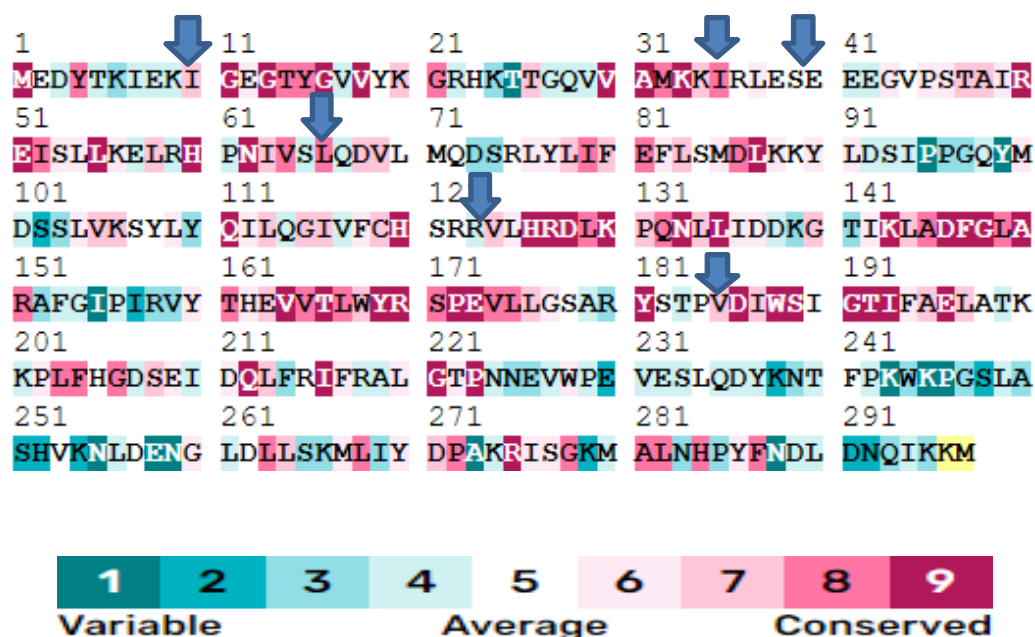


Figure 4.5: Color-coded ConSurf scores depicting the evolutionary conservation of amino acid residues in the CDK1 protein.

Table 4.4: ConSurf predicted evolutionary conservation of CDK1 amino acid residues, with scores ranging from 1 (variable) to 9 (highly conserved), highlighting regions potentially critical for structural or functional integrity.

Variant ID	Residue Position	Residue	Color
rs1460089100	66	L	8
rs1179034229	10	I	7
rs2080301489	35	I	8
rs1322570597	123	R	4
rs746720738	40	E	5
rs1234513195	185	V	7

To assess the impact of selected mutations on the structural stability of the CDK1 protein, a stability analysis was performed using the I-Mutant 2.0 server. This tool predicts changes in protein stability upon mutation by comparing the wild-type and mutant forms. I-Mutant utilizes the ProTherm database, which is a comprehensive resource of thermodynamic experimental data. The predictions are based on variations in Gibbs free energy ($\Delta\Delta G$), providing insights into whether a mutation increases or decreases protein stability.

Our findings showed that L66F explained the increase in stability while I10M, I35F, R123C, E40K, and V185A decreased stability. The DDG values of the other mutations ranged from -3.13 kcal/mol to 0.42 kcal/mol, with the V185A mutation recording the lowest value at -3.13 kcal/mol. According to table 5, a DDG value of less than 0 indicates a decrease in structural stability, whereas a number larger than 0 indicates a gain.

All damaging mutations were uploaded to the MUPRO server in order to validate the I-MUTANT findings. With the exception of L66F, which indicated an increase via I-MUTANT, other mutants produced findings that were almost identical, and vice versa, on the MUPRO server. As seen in table 5, only L66F produces findings that are in opposition to those of both softwares.

Table 4.5: Stability Analysis of Selected nsSNPs Using I-MUTANT and MUPRO

Variant ID	Position	Wild Type	Mutant	DDG Value	I-Mutant	Mupro
rs1460089100	66	L	F	0.42	Increase	Decrease
rs1179034229	10	I	M	-0.52	Decrease	Decrease
rs2080301489	35	I	F	-0.49	Decrease	Decrease
rs1322570597	123	R	C	-1.46	Decrease	Decrease
rs746720738	40	E	K	-0.76	Decrease	Decrease
rs1234513195	185	V	A	-3.13	Decrease	Decrease

4.5 Prediction and Assessment of CDK1 Tertiary Structure

The structure of a protein is critical for its specific biological function, and this structure is inherently determined by its amino acid sequence. Therefore, accurate prediction of a protein's three-dimensional (3D) structure is essential for understanding its functional mechanisms. The 3D structure of the CDK1 protein was predicted using both MODELLER (version 9.19) and the SWISS-MODEL online server, as illustrated in Figure 4.6. The crystal structure PDB ID: 4YC3, exhibiting 100% sequence identity and maximum alignment coverage, was selected as the most appropriate template. Specifically, chain A of 4YC3 was used to generate the homology model of CDK1 protein.

4.6 Verification of Model

Using LocPREFMD, the 3D structure of the CDK1 protein model was further refined to enhance structural accuracy, as illustrated in Figure 4.7. The refinement yielded an average 3D1D score ≥ 0.1 for approximately 81.48% of the residues, indicating good model quality for both the wild-type CDK1 protein and its variants. The majority of residues were positioned within the core region of the protein, supporting the structural stability of the model. Furthermore, 94% of residues were found within the favored regions of the Ramachandran plot, 5% in the allowed regions, and only 1% in the outlier regions, reflecting the overall reliability of the predicted protein conformation.

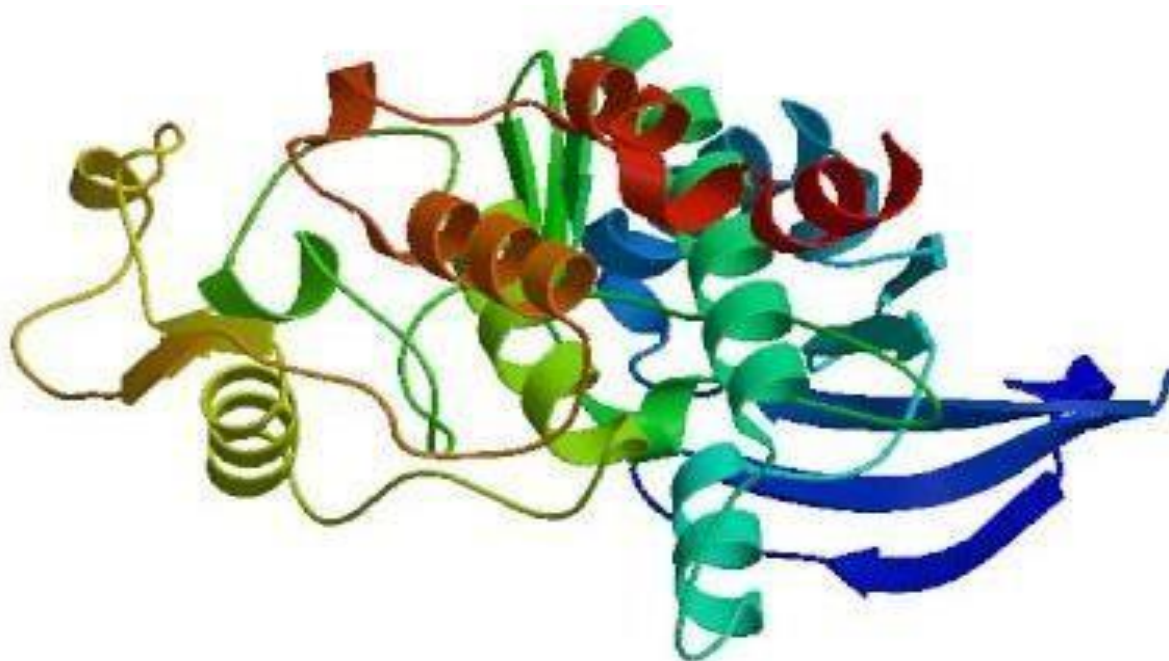


Figure 4.6: Native monomer structure of CDK1 with one domain.

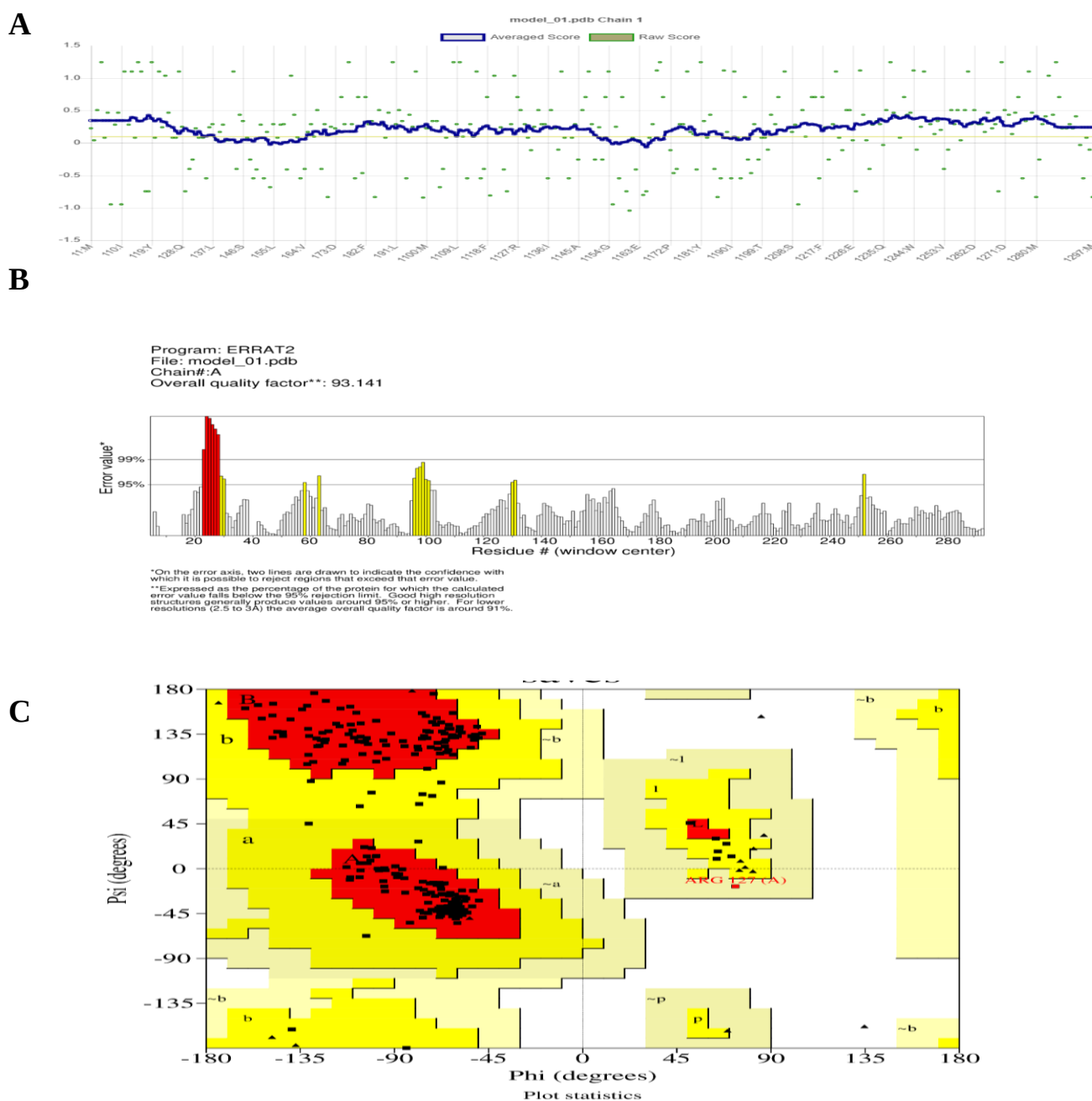


Figure 4.7: Predicted structural stability profile of the CDK1 protein.

Structural validation of the CDK1 3D model: (a) ERRAT plot indicating regions rejected at 95% (grey) and 99% (black) confidence levels.

(b) Verify3D profile showing residue-specific 3D-ID scores along the sequence, with positive values indicating favorable environments.

(c) Ramachandran plot depicting residue distribution within allowed conformational regions.

4.7 Effect of Mutation on Structure of CDK1 and their Functional Consequences

CDD (conserved domain database) helps to find conserved domains in protein sequence and CDD uses methods such as RPS-BLAST. CDD helps researchers to predict the function of unknown proteins by identifying similarities to well characterized domains. Conserved domains often represent functional or structural units within proteins, providing insights into their roles in biological processes. Therefore, we subjected our sequence of CDK1 protein to CDD to confirm the domain regions in our protein. In our results CDD confirmed only one Catalytic domain of the serine/threonine Kinase domain was detected. Where all the mutations are present in this domain as shown in Figure 4.8.

4.8 Structural Mapping of Deleterious Mutations in CDK1

All deleterious non-synonymous SNPs (nsSNPs) were localized on the three-dimensional (3D) structure of the CDK1 protein. Specifically, the mutations were mapped within the conserved STKc_CDK1_euk domain. Six predicted damaging mutations were highlighted in red on the tertiary structure, as illustrated in Figure 4.8.

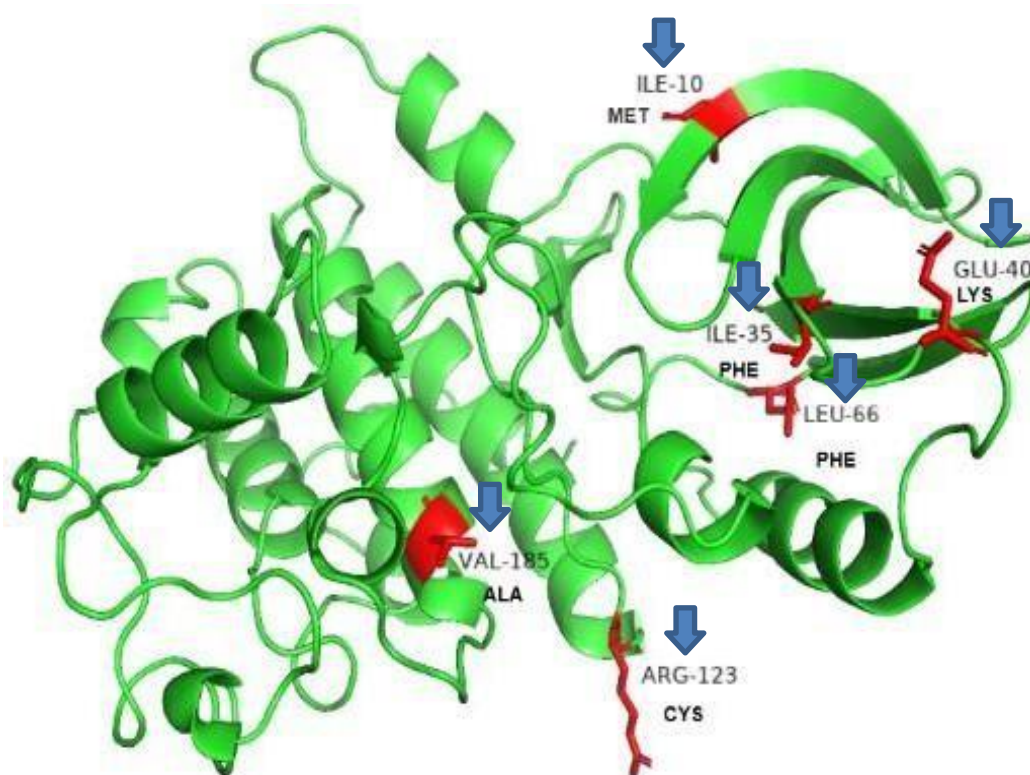


Figure 4.8: Visualization of deleterious nsSNPs in the CDK1 tertiary structure, with six mutations is highlighted in red color on the tertiary structure of CDK1 protein.

4.9 *In silico* site directed mutagenesis and RMSD Calculation

Using SNP analysis methods, the structure and function significance of several mutations in CDK1 were determined. Pymol were used to create the mutant models. The Swiss model produced the mutant models. Galaxy Web Server was used to develop the model, and all of the improved models were then submitted to Gromacs program for energy minimization. In order to minimize energy, NOMAD-REF was employed.

Since mutations alter a protein's structure, the RMSD value was used to assess the structural change caused by mutations. The structural variation and the RMSD value are always directly proportional. Variability will have a significant impact on protein structure. Table 6 displays the RMSD for each mutant structure in our results. The RMSD values were calculated using CHIMERA software. The RMSD values for the other mutants vary from 0.009 to 0.034, with the highest RMSD value for L66F being 0.656 and the lowest for R123C being 0.007.

CDK1 proteins are presented, with the native structure displayed in green and the mutant model superimposed for comparison, as illustrated in the figure. This structural alignment enables visualization of conformational differences introduced by the L66F mutation.

The native structure is represented in cyan, and the mutant model is displayed in medium blue. This structural overlay facilitates the comparison of conformational changes resulting from the I10M mutation.

The native structure depicted in cyan and the mutant structure highlighted in magenta, allowing visual comparison of structural alterations resulting from the I35F mutation.

The native structure is represented in cyan and the mutant structure is displayed in red, enabling visual assessment of structural variations induced by the R123C substitution.

The superimposed structure of the CDK1 protein displays the native form in cyan, while the mutant variant is highlighted in hot pink, allowing for a clear visual comparison of structural differences between the two forms.

The superimposed structure of the CDK1 protein shows the native form in cyan, while the mutant structure is visualized in orchid, facilitating a comparative analysis of their structural conformations.

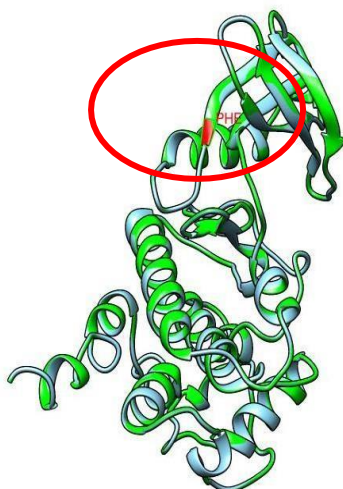


Figure 4.9: The superimposed structures of the native and mutant (L66F)

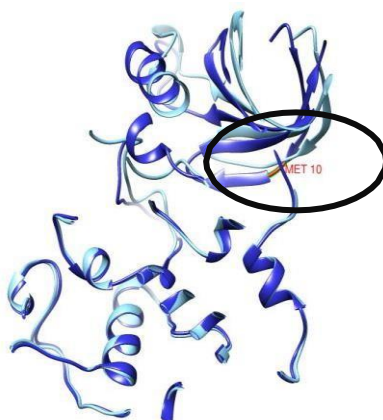


Figure 4.10: The superimposed structures of the native and mutant (I10M) forms of the CDK1 protein are shown.

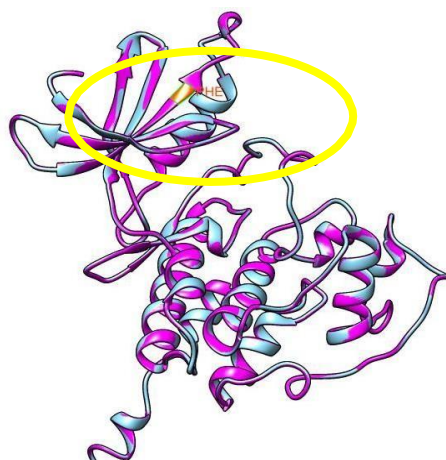


Figure 4.11: The superimposed structure of the native and mutant (I35F) forms of the CDK1 protein.

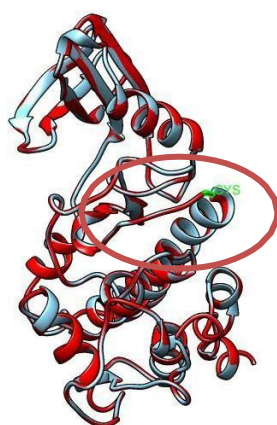


Figure 4.12: The superimposed structure of the native and mutant (R123C) forms of the CDK1 protein.

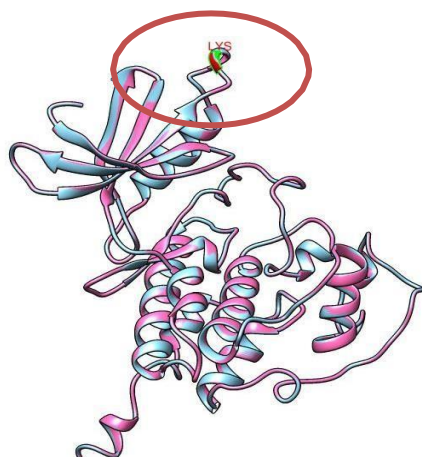


Figure 413: The superimposed structure of the native and mutant (E40K) forms of the CDK1 protein.

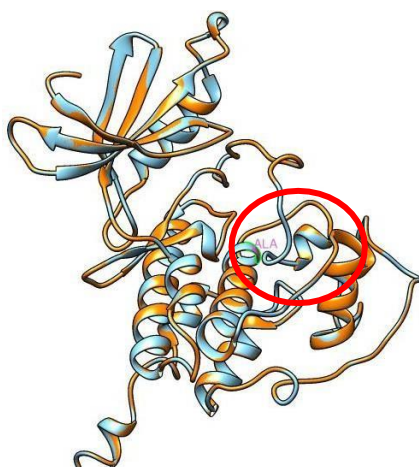


Figure 4.14: The superimposed structure of the native and mutant (V185A) forms of the CDK1 protein

Table 4.6: RMSD of both native protein and mutant models.

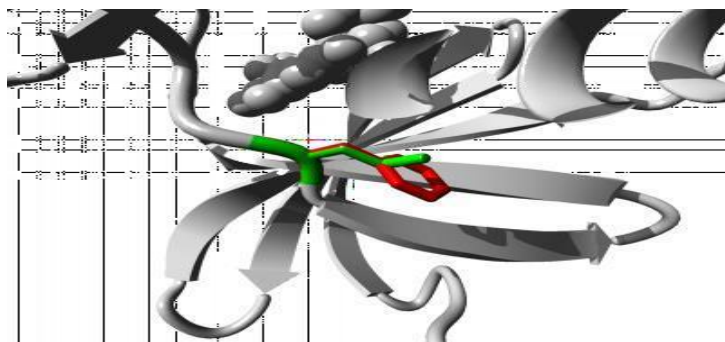
Mutations	RMSD value
L66F	0.034
I10M	0.656
I35F	0.017
R123C	0.007
E40K	0.013
V185A	0.009

HOPE

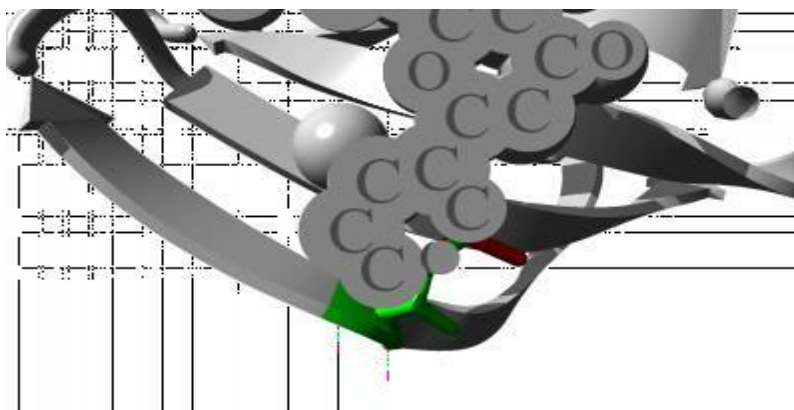
The Project HOPE web server was utilized to analyze the potential impact of the mutation on the protein domain and its interactions with surrounding molecules or substrates.

- (A)** In our results we find the amino acid Leucine changes into Phenylalanine at position 66. The wild-type residue is smaller than the mutant residue. Because of this, the mutant valine residue is more hydrophobic than the natural glycine residue. This mutation affects how a protein interacts with molecules or other proteins because it occurs on the protein's surface. A detailed structural analysis of the L66F mutation in the CDK1 protein was performed to observe local conformational changes.
- (B)** For the I10M mutation, the substitution may potentially disrupt external protein interactions, as the native residue (Isoleucine) is larger than the mutant residue (Methionine). Additionally, differences in hydrophobicity between the wild-type and mutant residues could further influence the structural integrity and interaction potential of the protein.
- (C)** In the I35F mutation, the wild-type residue (Isoleucine) is larger than the mutant residue (Phenylalanine). Due to this difference in side-chain size, the mutant residue may not occupy the same spatial position as the original, resulting in the loss of hydrogen bonding capability that was facilitated by the wild-type residue.
- (D)** At position 123, the larger residue Arginine (R) is substituted with the smaller residue Cysteine. This mutation may disrupt hydrophobic interactions on the protein surface, potentially leading to a loss of important intermolecular interactions and affecting protein stability or function.
- (E)** The substitution of Glutamic Acid with Lysine at position 40 may impair the protein's function, as the wild-type residue, Glycine, is known for its high flexibility and minimal steric hindrance. Glycine is neutral and smaller compared to the positively charged and bulkier mutant residue, Lysine. This replacement could lead to local conformational destabilization, potentially affecting the structural and functional integrity of the protein at this position.
- (F)** The substitution of Valine with Alanine at position 185 may result in potential protein folding issues. This is attributed to the introduction of a charged residue within a buried region of the protein. Since the mutation occurs in the protein's core, the altered residue may not fit properly due to differences in side-chain size and properties, potentially affecting the overall structural integrity.

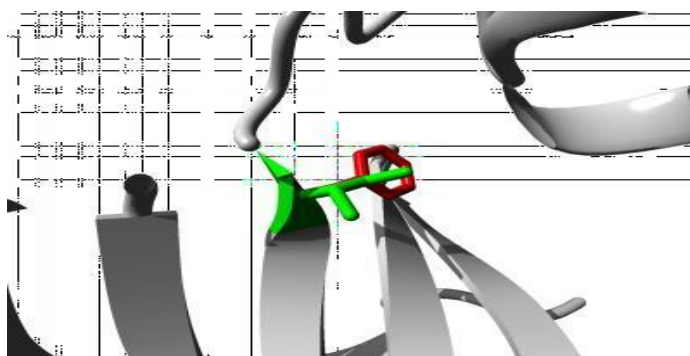
A.



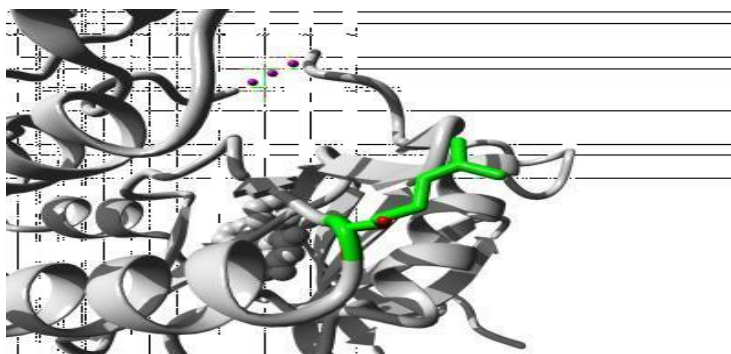
B



C



D



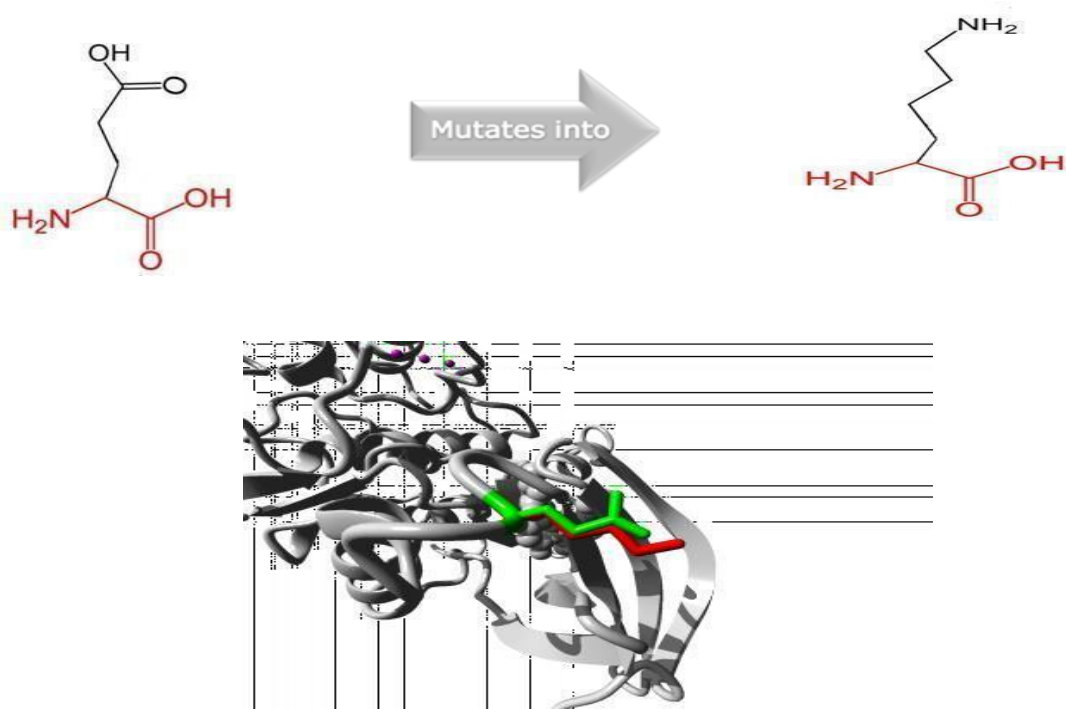
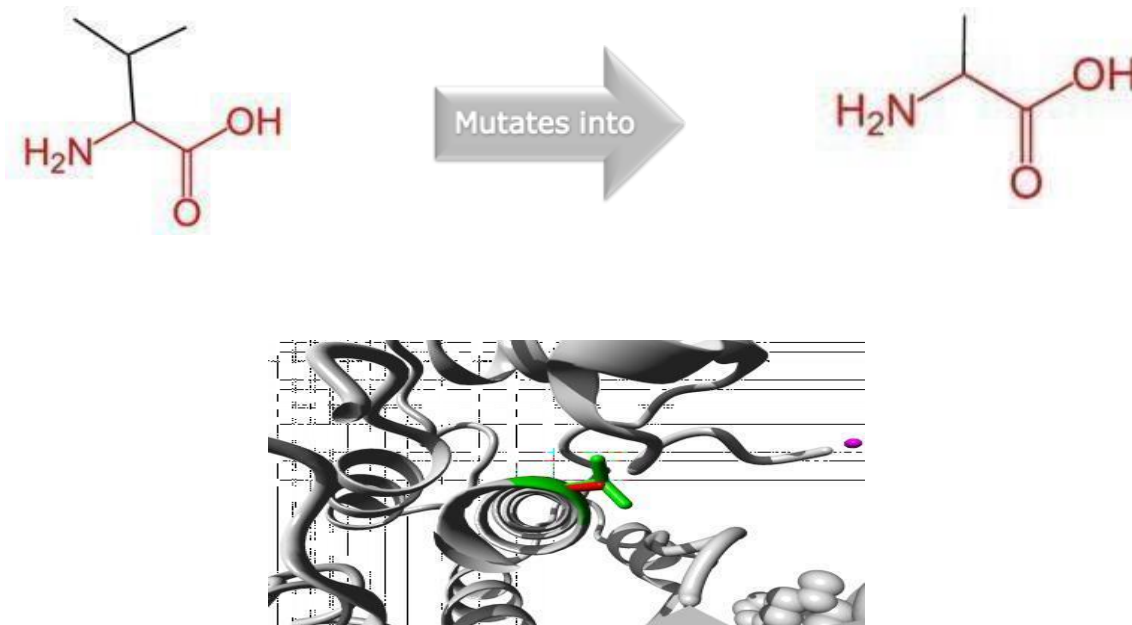
E**F**

Figure 4.15: Close up look of mutation in a protein. (a) L66F (b) I10M (c) I35F (d) R123C (e) E40K (f) V185A

The molecular and three-dimensional structures of both native and mutant residues in the protein are illustrated in the above Figure. In the 3D representation, the mutant residues are highlighted in red, while the native residues are depicted in green, allowing for a clear visual comparison of their structural differences.

4.10 Network and Pathway Analysis

4.10.1 Retrieval of CDK1 Protein Protein network

The CDK1 PPI network was retrieved using the String database. Using the CDK1 protein as a query, the CDK1 PPI interaction network was obtained from the String Database. It showed a dense network of 133 nodes and 1312 edges, with an average of 19.7 nodes, 0.762 average local clustering coefficients, 164 anticipated edges, and a PPI enrichment p-value $<10.e-16$. (Fig. 4.21)

In order to better understand the interactions, the interacting network was divided into three clusters using the string database's k-mean cluster option. Table 4.7 displays the number of genes and their descriptions.

The closely related proteins in the network made up these clusters. Positive action effects are represented by arrows in PPI network clusters, negative action effects by lines, and nonspecific effects by balls as shown in Figure 4.22.

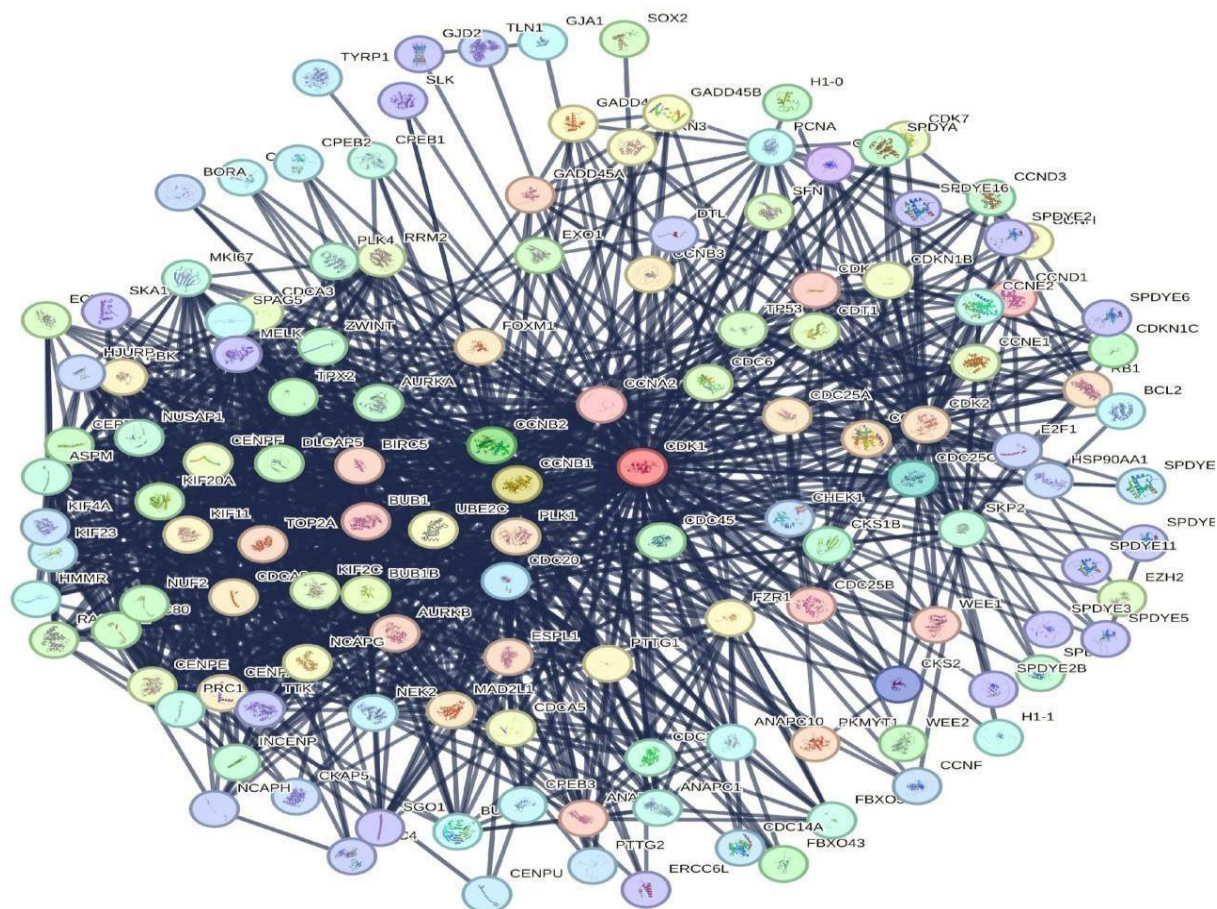


Figure 4.16: Shows CDK1 PPI network that involved in signal transduction.

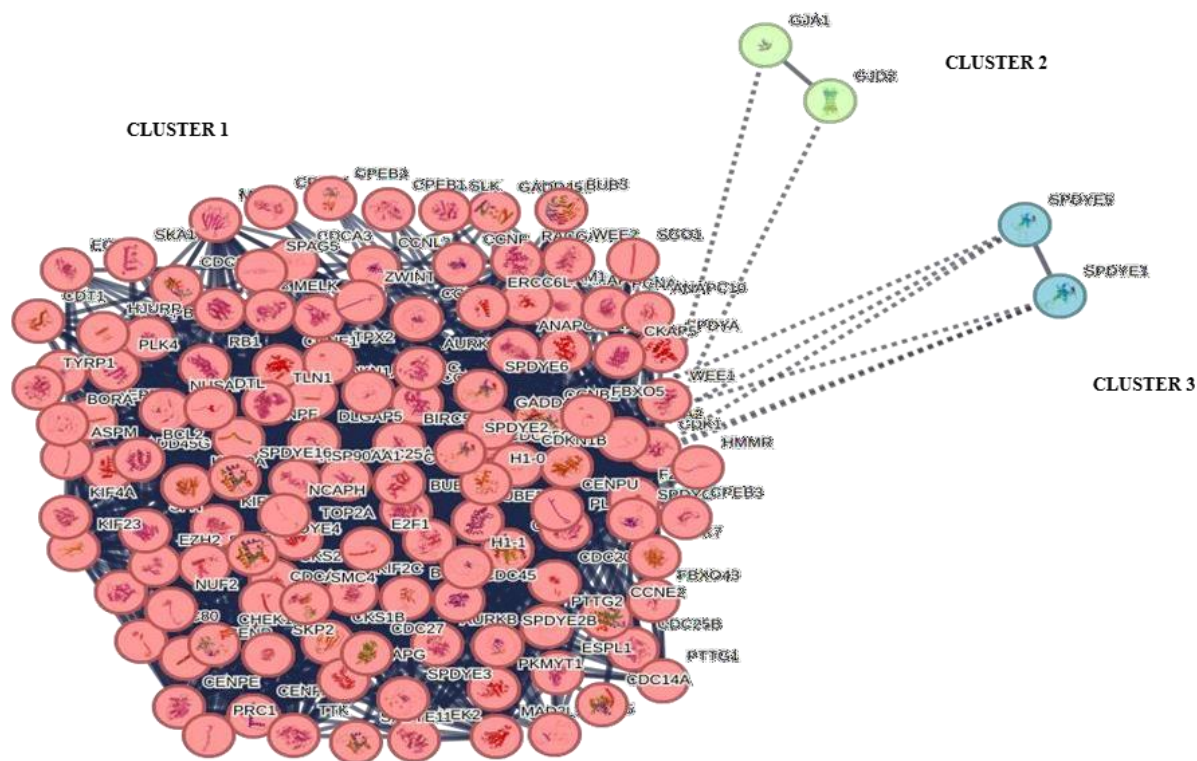


Figure 4.17: CDK1 PPI network differentiated into 3 k-means clusters.

Table 4.7: 3 K-mean clusters of CDK1

Cluster ID	Gene Count	Description
Cluster 1	129	Cell cycle Mixed, incl. Mitotic Spindle Checkpoint, and Mitotic sister chromatid segregation
Cluster 2	2	Gap junction Connexin homologues
Cluster 3	2	Cell cycle regulatory protein

4.10.2 KEGG

Based on KEGG pathway enrichment analysis, the proteins were found to be involved in multiple cellular signaling pathways. Specifically, eukaryotic cells respond to DNA damage by activating complex signaling cascades that promote DNA repair and induce cell cycle arrest. Upon DNA damage, the checkpoint kinase ATM phosphorylates and activates Chk2, which subsequently activates the tumor suppressor protein p53 through phosphorylation. Both the G1 and G2 cell cycle checkpoints are regulated by p53 and its downstream transcriptional targets. Additionally, intra-S-phase checkpoint activation occurs through the degradation of the Cdc25A phosphatase, mediated by the ATR-Chk1 signaling axis. The cell cycle is also negatively regulated by cyclin-dependent kinase inhibitors (CKIs), such as p16Ink4a, p15Ink4b, p27Kip1, and p21Cip1, which inhibit CDK activity and thereby control cell cycle progression.

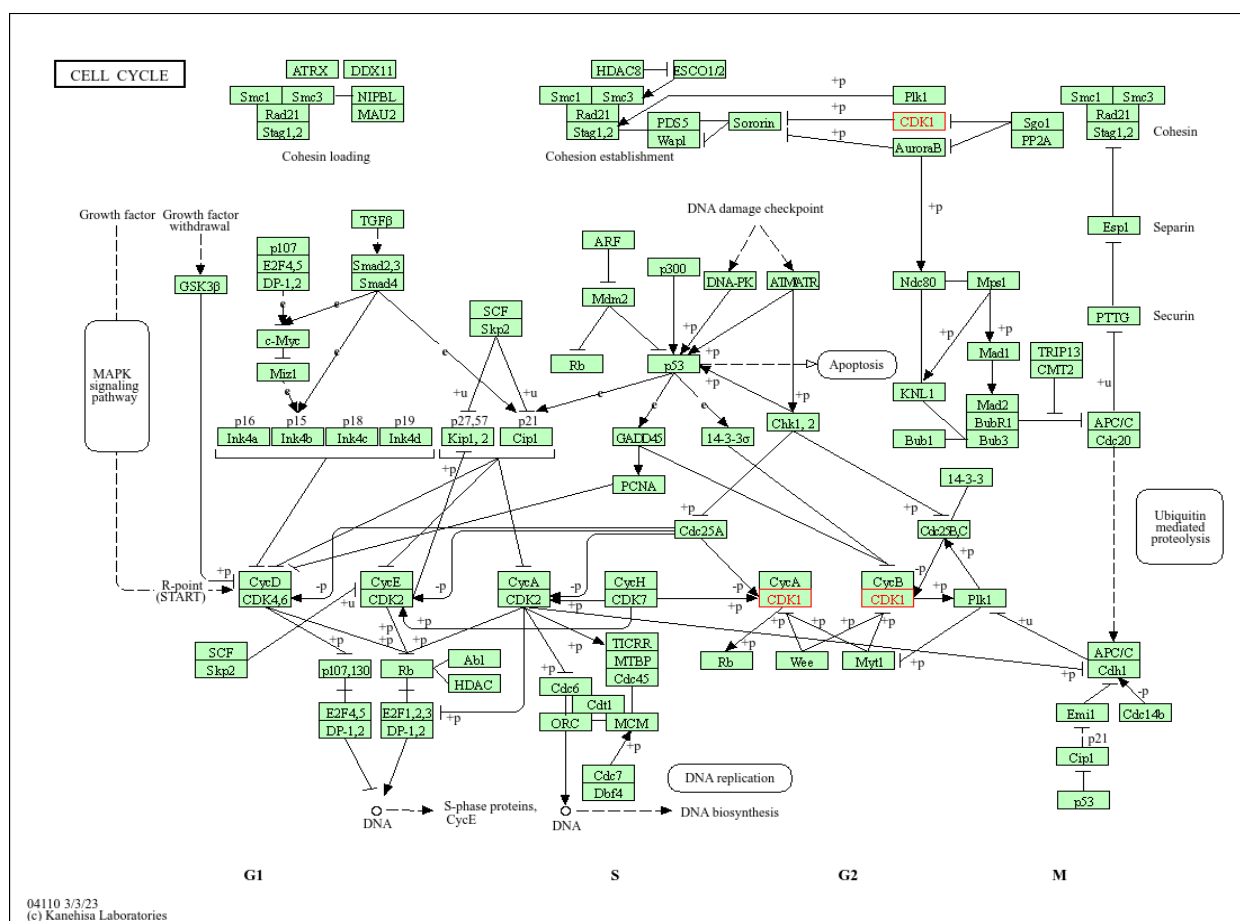


Figure 4.23: KEGG Pathway database map showing MAPK signaling pathway and ubiquitin mediated proteolysis.

4.11 DISCUSSION

Cyclin-dependent kinase 1 (CDK1), a 34 kDa serine/threonine protein kinase encoded by the *CDC2* gene, is a pivotal regulator of the eukaryotic cell cycle. As a central mediator of mitotic entry, CDK1 governs the G2/M phase transition, thereby playing an essential role in the orchestration of mitotic progression and overall cell cycle regulation in human cells [1]. The cell cycle is a fundamental biological process responsible for the precise replication and equitable segregation of genetic material into daughter cells. It plays a critical role in maintaining genomic integrity by coordinating cellular proliferation, growth, and differentiation, while dynamically integrating environmental cues to preserve genomic stability during both mitotic and meiotic divisions [6].

Non-synonymous single nucleotide polymorphisms (nsSNPs) are a primary source of amino acid substitutions within protein sequences, which can consequently alter both the structure and function of proteins. These alterations are often implicated in the development of various genetic disorders. Functionally important regions of proteins are generally subject to strong selective pressure, which resists sequence variation despite the degeneracy of the genetic code. Identifying specific amino acid changes that contribute to disease pathogenesis remains a complex, labor-intensive, and costly task. Furthermore, distinguishing between deleterious and neutral nsSNPs presents a significant challenge in genomic studies [80].

Computational data mining techniques play a vital role in distinguishing between neutral and deleterious SNPs by leveraging a range of algorithmic approaches and curated biological databases. These tools utilize available structural and phylogenetic information to assess the potential impact of amino acid substitutions resulting from mutations [64].

Amino acid substitutions resulting from non-synonymous SNPs (nsSNPs) in the coding regions of genes can lead to alterations in protein structure and function, potentially contributing to disease susceptibility. Screening for deleterious nsSNPs is a valuable strategy for assessing an individual's genetic predisposition to various disorders. However, not all nsSNPs exert harmful effects; some may be functionally tolerated without disrupting protein activity. Despite this, many nsSNPs possess molecular significance, as they are associated with disease phenotypes or other biologically relevant traits [81]. Non-synonymous single nucleotide polymorphisms (nsSNPs), which result in amino acid substitutions, may influence protein structural integrity, interfere with bimolecular interactions, or affect catalytic activity. The exploration and functional assessment of nsSNPs associated with disease phenotypes remain a major focus in molecular biology and population genetics research. [82].

A considerable number of SNPs associated with the CDK1 gene are available in public databases. These variants require comprehensive screening to identify those that are most deleterious i.e., capable of altering protein structure or function and potentially contributing to disease. In the present study, all reported SNPs for CDK1 were systematically analyzed using computational approaches to determine which variants may have the potential to disrupt the normal structural and functional properties of the protein.

To evaluate the functional impact of the selected nsSNPs, their corresponding rsIDs were submitted to widely used prediction tools such as PolyPhen-2 and SIFT. PolyPhen-2 classified 22 variants as probably or possibly damaging, while SIFT identified 28 nsSNPs as non-tolerated. For further validation, the same set of rsIDs was analyzed using additional tools including P-Mut, SNP&GO, PROVEAN, and PANTHER. Among these, six variants rs1460089100, rs1179034229, rs2080301489, rs1322570597, rs746720738, and rs1234513195 were consistently predicted as deleterious by all tools used.

According to a thorough review of existing literature, no prior studies have directly linked these damaging nsSNPs in CDK1 to any specific disease phenotypes. This highlights the need for experimental validation to confirm the clinical significance of these computational predictions. The findings from this study also emphasize the effectiveness of combining multiple data mining algorithms for the accurate identification and prioritization of functionally important nsSNPs.

Among the various computational tools assessed, SNAP and PolyPhen-2 demonstrated particularly strong performance in predicting the functional consequences of nsSNPs [83]. These tools offer valuable support in distinguishing between causal and non-causal variants, thereby aiding in the understanding of genotype–phenotype relationships.

In this context, the aforementioned computational tools can be utilized to differentiate causal relationships between non-synonymous SNPs (nsSNPs) and the associated phenotypic outcomes from those that are non-causal. In order to estimate the function of six damaging nsSNPs in human diseases and their relationship to proteins, we used the Mutpred server, a stand-alone program designed to classify amino acid changes in humans as either benign or pathogenic [71]. MutPred evaluates several structural and functional properties, including solvent accessibility; molecular recognition features (MoRFs), alterations in protein stability, post-translational modifications, and enzymatic active site involvement. The tool generates predictions based on two key metrics: the general (g) score and the probability (p) score. Actionable hypotheses are defined by thresholds of $p < 0.5$ and $g > 0.5$, while high-confidence predictions meet more stringent criteria of $p < 0.01$ and $g > 0.75$.

These scores collectively contribute to the classification of nsSNPs into different hypothesis categories based on their predicted functional impact [72].

The potential effects of selected SNPs on amyloid propensity, chaperone binding, and protein aggregation were evaluated using the SNPeff server. Results revealed that the variant rs1460089100 exhibited a marked increase in amyloid potential, with a d WALTZ score of 88.31. However, for the majority of the analyzed nsSNPs, SNPeff did not predict significant changes in chaperone-binding affinity, aggregation propensity, or amyloid formation in the CDK1 protein, indicating that these variants may not substantially alter the protein's biophysical characteristics through these specific mechanisms.

To identify functionally important regions within the CDK1 protein, evolutionary conservation analysis was conducted using the ConSurf tool [73]. This approach provided insight into residues under strong evolutionary constraint, which are more likely to be involved in maintaining structural integrity or executing critical biological functions.

The ConSurf server utilizes the multiple sequence alignment of the query protein as input to assess the evolutionary conservation of each residue. In this study, the alignment for the CDK1 protein was provided in FASTA format [14].

If these variants result in damaging alterations to the CDK1 protein, they are likely to influence its amyloid propensity, aggregation behavior, and chaperone-binding capacity, as predicted by the SNPeff tool. To further investigate the structural consequences of these mutations, the HOPE web server was employed. This tool provides insights into how specific amino acid substitutions may affect protein domains and alter the interactions of the protein with its surrounding environment [78].

Various research groups have investigated and documented the spatial distribution of non-synonymous single nucleotide polymorphisms (nsSNPs) within protein structures [76, 85]. The three-dimensional (3D) structure of the CDK1 protein is publicly available in the Protein Data Bank (PDB). In this study, energy minimization was applied to both the wild-type and mutant models to enhance structural accuracy. The Root Mean Square Deviation (RMSD) was calculated to quantify structural differences between the native and mutant forms. These deviations are attributed to mutations such as insertions and deletions, which may significantly impact the functional performance and structural stability of the protein [86, 87].

According to KEGG Pathway enrichment analysis, the proteins were implicated in several pathways [93].

CHAPTER 05
CONCLUSION

The CDK1 protein is located on chromosome 10 at the 10q21.2 region in the human genome. CDK1 encodes a highly conserved serine/threonine kinase that plays a critical role in cell cycle regulation, particularly in facilitating the transition from the G2 phase to mitosis (M phase). This kinase remains inactive in its monomeric form and requires association with Cyclin B to become functionally active. The formation of the CDK1–Cyclin B complex induces conformational changes that properly align the active site of CDK1, enabling its catalytic activity. Notably, CDK1 is unique among cyclin-dependent kinases (CDKs) in that it is indispensable for mitosis and can partially compensate for the absence of other CDKs in certain biological contexts.

Mutations or dysregulation in CDK1 have been implicated in various pathological conditions, including tumor genesis. Therefore, current research has expanded from merely understanding the genotypic features of CDK1 to investigating phenotypic associations of single nucleotide polymorphisms (SNPs) linked to this gene. In the present study, a comprehensive screening of CDK1-associated SNPs was conducted using a suite of bioinformatics tools, including SIFT, PolyPhen-2, SNP&GO, PHD, PROVEAN, P-Mut, MutPred, and SNPeffect. Through this analysis, six nsSNPs rs1460089100, rs1179034229, rs2080301489, rs1322570597, rs746720738, and rs1234513195 were consistently predicted to have damaging effects on protein structure and function. All of these variants were found to lie within functional domains, suggesting their potential involvement in disease pathogenesis.

Given their predicted impact, these unreported deleterious nsSNPs may significantly alter CDK1's structural integrity and biological activity, thereby increasing the likelihood of association with disease development. The findings of this study underscore the importance of prioritizing these high-risk nsSNPs for further experimental validation. Although this study is limited to *in silico* analysis, wet-laboratory experimentation is essential to confirm the biological consequences of these mutations. Additionally, the use of disease models carrying these specific CDK1 mutations could provide deeper insights into their functional roles in pathological conditions.

REFERENCES

1. Ying X, Che X, Wang J, Zou G, Yu Q, Zhang X. CDK1 serves as a novel therapeutic target for endometrioid endometrial cancer. *J Cancer*. 2021 Feb 22;12(8):2206-2215. doi: 10.7150/jca.51139. PMID: 33758599; PMCID: PMC7974891.
2. Zhang, M. et al. CDK inhibitors in cancer therapy, an overview of recent development. *Am. J. Cancer Res*. 11, 1913 (2021).
3. Lau, Ho Wai, et al. "Quantitative differences between cyclin-dependent kinases underlie the unique functions of CDK1 in human cells." *Cell Reports* 37.2 (2021).
4. Cánovas PM. Survivin Mediates Mitotic Onset in HeLa Cells Through Activation of the Cdk1-Cdc25B Axis. *Res Sq [Preprint]*. 2024 Feb 28:rs.3.rs-3949429. doi: 10.21203/rs.3.rs-3949429/v1. PMID: 38464014; PMCID: PMC10925418.
5. X. Kang, H. Chen, Z. Zhou, S. Tu, B. Cui, Y. Li, S. Dong, Q. Zhang, Y. Xu, Targeting Cyclin-Dependent Kinase 1 Induces Apoptosis and Cell Cycle Arrest of Activated Hepatic Stellate Cells. *Adv. Biology* 2024, 8, 2300403.
6. Chia, Kim Hou, et al. "CDK1-PP2A-B55 interplay ensures cell cycle oscillation via Apc1-loop300." *Cell Reports* 43.5 (2024).
7. Guen, Vincent J., et al. "The awakening of the CDK10/Cyclin M protein kinase." *Oncotarget* 8.30 (2017): 50174.
8. Lim, Key-Hwan, et al. "Advances in multiplex PCR for Alzheimer's disease diagnostics targeting CDK genes." *Neuroscience Letters* 749 (2021): 135715.
9. Vallejos-Vidal, Eva, et al. "Single-nucleotide polymorphisms (SNP) mining and their effect on the tridimensional protein structure prediction in a set of immunity-related expressed sequence tags (EST) in Atlantic salmon (*Salmo salar*)." *Frontiers in genetics* 10 (2020): 1406.
10. Izarzugaza, Jose MG, Martin Krallinger, and Alfonso Valencia. "Interpretation of the consequences of mutations in protein kinases: combined use of bioinformatics and text mining." *Frontiers in Physiology* 3 (2012): 323.
11. Essadssi, S., Krami, A. M., Elkhatabi, L., Elkarhat, Z., Amalou, G., Abdelghaffar, H., & Barakat, A. (2019). Computational analysis of nsSNPs of ADA gene in severe combined immunodeficiency using molecular modeling and dynamics simulation. *Journal of immunology research*, 2019(1), 5902391.
12. "In silico analysis of coding/noncoding SNPs of human RETN gene and characterization of their impact on resistin stability and structure." *Journal of Diabetes Research* 2019.1 (2019):

- 4951627.
13. Tawfik NS, Spruit MR. The SNPcurator: literature mining of enriched SNP-disease associations. Database (Oxford). 2018 Jan 1;2018:bay020. doi: 10.1093/database/bay020. Erratum in: Database (Oxford). 2021 Nov 25;2021(2021):baab070. doi: 10.1093/database/baab070. PMID: 29688369; PMCID: PMC5844215.
 14. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. Genome Res. 1998 Dec;8(12):1229-31. doi: 10.1101/gr.8.12.1229. Erratum in: Genome Res 1999 Feb;9(2):210. PMID: 9872978.
 15. Savas S, Schmidt S, Jarjanazi H, Ozcelik H. Functional nsSNPs from carcinogenesis-related genes expressed in breast tissue: potential breast cancer risk alleles and their distribution across human populations. Hum Genomics. 2006 Mar;2(5):287-96. doi: 10.1186/1479-7364-2-5-287. PMID: 16595073; PMCID: PMC3500178.
 16. Sava V, Reunova O, Velasquez A, Song S, Sanchez-Ramos J. Neuroanatomical mapping of DNA repair and antioxidative responses in mouse brain: Effects of a single dose of MPTP. Neurotoxicology. 2006 Dec;27(6):1080-93. doi: 10.1016/j.neuro.2006.05.023. Epub 2006 Jun 9. PMID: 16831462.
 17. Schork NJ, Fallin D, Lanchbury JS. Single nucleotide polymorphisms and the future of genetic epidemiology. Clin Genet. 2000 Oct;58(4):250-64. doi: 10.1034/j.1399-0004.2000.580402.x. PMID: 11076050.
 18. Emahazion T, Feuk L, Jobs M, Sawyer SL, Fredman D, St Clair D, Prince JA, Brookes AJ. SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. Trends Genet. 2001 Jul;17(7):407-13. doi: 10.1016/s0168-9525(01)02342-3. PMID: 11418222.
 19. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol. 2010 Jan 26;8(1):e1000294. doi: 10.1371/journal.pbio.1000294. PMID: 20126254; PMCID: PMC2811148.
 20. Barroso, I., et al. "Dominant negative mutations in human PPAR γ associated with severe insulin resistance, diabetes mellitus and hypertension." *Nature* 402.6764 (1999): 880-883
 21. Zhang, Ben, et al. "Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk." *Nature genetics* 46.6 (2014): 533-542.
 22. Al Olama, Ali Amin, et al. "A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer." *Nature genetics* 46.10 (2014): 1103-1109.

23. Michailidou, Kyriaki, et al. "Large-scale genotyping identifies 41 new loci associated with breast cancer risk." *Nature genetics* 45.4 (2013): 353-361.4.
24. Yu, Z., Coorens, T. H., Uddin, M. M., Ardlie, K. G., Lennon, N., & Natarajan, P. (2024). Genetic variation across and within individuals. *Nature Reviews Genetics*, 25(8), 548-562..
25. Yu, Z., Coorens, T. H., Uddin, M. M., Ardlie, K. G., Lennon, N., & Natarajan, P. (2024). Genetic variation across and within individuals. *Nature Reviews Genetics*, 25(8), 548-562.
26. "Structural bioinformatics-based design of selective, irreversible kinase inhibitors." *Science* 308.5726 (2005): 1318-1321.
27. "Fluorescent carbon nanoparticles: synthesis, characterization, and bioimaging application." *The Journal of Physical Chemistry C* 113.43 (2009): 18546-18551
28. Simon, Richard. "Bioinformatics in cancer therapeutics--hype or hope?." *Nature Clinical Practice Oncology* 2.5 (2005): 223-224. Cohen, Michael S., et al.
29. Landgren O, Linet MS, McMaster ML, Gridley G, Hemminki K, Goldin LR. Familial characteristics of autoimmune and hematologic disorders in 8,406 multiple myeloma patients: a population-based case-control study. *Int J Cancer*. 2006 Jun 15;118(12):3095-8. doi: 10.1002/ijc.21745. PMID: 16395700.
30. Pişkin E. Biodegradable polymers as biomaterials. *J Biomater Sci Polym Ed*. 1995;6(9):775-95. doi: 10.1163/156856295x00175. PMID: 7772566. Graves Jr, J. L. (2021).
31. Human biological variation and the “normal”. *American Journal of Human Biology*, e23658.
32. Ho, S. S., Urban, A. E., & Mills, R. E. (2020). Structural variation in the sequencing era. *Nature Reviews Genetics*, 21(3), 171-189.
33. Vihinen, M. (2018). Systematics for types and effects of DNA variations. *BMC genomics*, 19(1), 1-16.
34. Sukhumsirichart, W. (2018, October 17). Polymorphisms. IntechOpen. Retrieved September 16, 2021, from <https://www.intechopen.com/chapters/61204>.
35. Li, H. (2021, March 15). SNP vs SNV. Sitewide ATOM. Retrieved September 16, 2021, from <http://lh3.github.io/2021/03/15/snp-vs-snv>.
36. Kim, S., & Misra, A. (2007). SNP genotyping: technologies and biomedical applications. *Annu. Rev. Biomed. Eng.*, 9, 289-320.
37. Brown, D. K., & Bishop, Ö. T. (2017). Role of structural bioinformatics in drug discovery by computational SNP analysis: analyzing variation at the protein level. *Global heart*, 12(2), 151-161.

38. Liao H, Ji F, Ying S. CDK1: beyond cell cycle regulation. *Aging* (Albany NY). 2017 Dec 14;9(12):2465-2466. doi: 10.18632/aging.101348. PMID: 29242409; PMCID: PMC5764383.
39. Enserink, J. M., & Kolodner, R. D. (2010). An overview of Cdk1-controlled targets and processes. *Cell division*, 5, 1-41.
40. Wang, Q., Bode, A. M., & Zhang, T. (2023). Targeting CDK1 in cancer: mechanisms and implications. *NPJ precision oncology*, 7(1), 58.
41. Wang, Z., Fan, M., Candas, D., Zhang, T. Q., Qin, L., Eldridge, A., ... & Li, J. J. (2014). Cyclin B1/Cdk1 coordinates mitochondrial respiration for cell-cycle G2/M progression. *Developmental cell*, 29(2), 217-232.
42. Aftab, A., Khan, R., Shah, W., Azhar, M., Unar, A., Jafar Hussain, H. M., & Waqas, A. (2021). Computational analysis of Cyclin D1 gene SNPs and association with breast cancer. *Bioscience Reports*, 41(1), BSR20202269.
43. Paul Chowdhuri, S., & Das, B. B. (2024). TDP1 phosphorylation by CDK1 in mitosis promotes MUS81-dependent repair of trapped Top1-DNA covalent complexes. *The EMBO Journal*, 1-23.
44. Vassilev, L. T. (2006). Cell cycle synchronization at the G2/M phase border by reversible inhibition of CDK1. *Cell cycle*, 5(22), 2555-2556.
45. Prevo, R., Pirovano, G., Puliyadi, R., Herbert, K. J., Rodriguez-Berriguete, G., O'Docherty, A., ... & Higgins, G. S. (2018). CDK1 inhibition sensitizes normal cells to DNA damage in a cell cycle dependent manner. *Cell Cycle*, 17(12), 1513-1523.
46. Jones, M. C., Askari, J. A., Humphries, J. D., & Humphries, M. J. (2018). Cell adhesion is regulated by CDK1 during the cell cycle. *Journal of Cell Biology*, 217(9), 3203-3218.
47. Zhou, Q., Wu, F., Chen, Y., Fu, J., Zhou, L., Xu, Y., ... & Yuan, F. (2024). Reynoutria multiflora (Thunb.) Moldenke and its ingredient suppress lethal prostate cancer growth by inducing CDC25B-CDK1 mediated cell cycle arrest. *Bioorganic Chemistry*, 152, 107731
48. William McLaren, Bethan Pritchard, Daniel Rios, Yuan Chen, Paul Flicek, Fiona Cunningham, Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor, *Bioinformatics*, Volume 26, Issue 16, August 2010, Pages 2069–2070,
49. S. T. Sherry, M.-H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, K. Sirotkin, dbSNP: the NCBI database of genetic variation, *Nucleic Acids Research*, Volume 29, Issue 1, 1 January 2001, Pages 308–311,

50. Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L.A., Edwards, K.J., Day, I.N.M. and Gaunt, T.R. (2013), Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Human Mutation*, 34: 57-65.
51. I. Adzhubei, et al., Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2, current protocols in human genatics, vol. 76, pp.7.20, 1-7.20. 41, 2013.
52. B.Kamaraj, et al., In silico screening of cancer associated mutation on PLK1 protein and its structural consequence *Journal of Molecular Modeling*, Vol. 7, pp. 5587-5599, 2013.
53. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* 7(10): e46688. doi:10.1371/journal.pone.0046688
54. B.A. Hassel, et al., A dominant negative mutant of 2-5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon. *The EMBO Journal* vol. 12, pp.3297 - 3304
55. Powers, Evan T., et al. "Biological and chemical approaches to diseases of proteostasis deficiency." *Annual review of biochemistry* 78.1 (2009): 959-991.
56. Mayrose, Itay, et al. "Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior." *Molecular biology and evolution* 21.9 (2004): 1781-1791.
57. Miller, Mark P., and Sudhir Kumar. "Understanding human disease mutations through the use of interspecific genetic variation." *Human molecular genetics* 10.21 (2001): 2319-2328.
58. Westbrook, John, et al. "The protein data bank and structural genomics." *Nucleic acids research* 31.1 (2003): 489-491
59. Morad, Fatima A., et al. "In silico approach to investigate the structural and functional attributes of familial hypercholesterolemia variants reported in the saudi population." *Journal of Computational Biology* 25.2 (2018): 170-181.
60. Danishuddin, Mohd, Ravish Chatrath, and Rajender Singh. "Insights of interaction between small and large subunits of ADP-glucose pyrophosphorylase from bread wheat (*Triticum aestivum* L.)." *Bioinformation* 6.4 (2011): 144.
61. Feig, Michael. "Local protein structure refinement via molecular dynamics simulations with locPREFMD." *Journal of chemical information and modeling* 56.7 (2016): 1304-1312.
62. Heo, Lim, Hahnbeom Park, and Chaok Seok. "GalaxyRefine: Protein structure refinement driven by side-chain repacking." *Nucleic acids research* 41.W1 (2013): W384-W388.

63. Grotthuss, Marcin von, et al. "Application of 3D-Jury, GRDB, and Verify3D in fold recognition." *Proteins: Structure, Function, and Bioinformatics* 53.S6 (2003): 418-423.
64. Hooft, Rob WW, Chris Sander, and Gerrit Vriend. "Objectively judging the quality of a protein structure from a Ramachandran plot." *Bioinformatics* 13.4 (1997): 425-430.
65. Van Durme, Joost, et al. "Accurate prediction of DnaK-peptide binding via homology modelling and experimental data." *PLoS computational biology* 5.8 (2009): e1000475.
66. Overington, John P., et al. "Molecular recognition in protein families: a database of aligned three-dimensional structures of related proteins." (1993): 597-604.
67. Prokop, Martin, et al. "TRITON: a graphical tool for ligand-binding protein engineering." *Bioinformatics* 24.17 (2008): 1955-1956.
68. Laskowski, Roman A., et al. "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." *Journal of biomolecular NMR* 8 (1996): 477-486.
69. Rashid, Tamir, Toshihiro Kobayashi, and Hiromitsu Nakauchi. "Revisiting the flight of Icarus: making human organs from PSCs with large animal chimeras." *Cell stem cell* 15.4 (2014): 406-409.
70. Vaser, Robert, et al. "SIFT missense predictions for genomes." *Nature protocols* 11.1 (2016): 1-9.
71. Pejaver, Vikas, et al. "Inferring the molecular and phenotypic impact of amino acid variants with MutPred2." *Nature communications* 11.1 (2020): 5918.
72. Seehausen, Ole, et al. "Genomics and the origin of species." *Nature Reviews Genetics* 15.3 (2014): 176-192.
73. Ashkenazy, Haim, et al. "ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids." *Nucleic acids research* 38.suppl_2 (2010): W529-W533.
74. Capriotti, Emidio, Piero Fariselli, and Rita Casadio. "I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure." *Nucleic acids research* 33.suppl_2 (2005): W306-W310.
75. Bhargava, Kiran, et al. "Molecular Docking studies of D2 Dopamine receptor with Risperidone derivatives." *Bioinformation* 10.1 (2014): 8.
76. Eisenberg, David, Roland L  thy, and James U. Bowie. "[20] VERIFY3D: assessment of protein models with three-dimensional profiles." *Methods in enzymology*. Vol. 277. Academic Press, 1997. 396-404.

77. Ho, Bosco K., and Robert Brasseur. "The Ramachandran plots of glycine and pre-proline." *BMC structural biology* 5 (2005): 1-11.
78. Venselaar, H., et al. "Protein structure analysis of mutations causing inheritable diseases." *An e-Science approach with life scientist friendly interfaces.*(1471-2105 (Electronic)) (2003).
79. Wang, B., & Zhao, D. (2024). HOPE: Homomorphic Order-Preserving Encryption for Outsourced Databases--A Stateless Approach. *arXiv preprint arXiv:2411.17009*.
80. Yates, Christopher M., and Michael JE Sternberg. "Proteins and domains vary in their tolerance of non-synonymous single nucleotide polymorphisms (nsSNPs)." *Journal of molecular biology* 425.8 (2013): 1274-1286.
81. Johnson, Melissa M., John Houck, and Chu Chen. "Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response." *Cancer Epidemiology Biomarkers & Prevention* 14.5 (2005): 1326-1329.
82. George Priya Doss, C., et al. "Extrapolating the effect of deleterious nsSNPs in the binding adaptability of flavopiridol with CDK7 protein: a molecular dynamics approach." *Human genomics* 7 (2013): 1-15.
83. Thusberg, Janita, and Mauno Vihinen. "Pathogenic or not? And if so, then how? Studying the effects of missense mutations using bioinformatics methods." *Human mutation* 30.5 (2009): 703-714.
84. Rickert, Andreas M., et al. "First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome." *Plant biotechnology journal* 1.6 (2003): 399-410.
85. Yue, Peng, Eugene Melamud, and John Moul. "SNPs3D: candidate gene and SNP selection for association studies." *BMC bioinformatics* 7 (2006): 1-15.
86. Han, Jung-Hoon, et al. "Divergence of interdomain geometry in two-domain proteins." *Structure* 14.5 (2006): 935-945.
87. Varfolomeev, S. D., I. V. Uporov, and E. V. Fedorov. "Bioinformatics and molecular modeling in chemical enzymology. Active sites of hydrolases." *Biochemistry (Moscow)* 67 (2002): 1099-1108.
88. Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, 22(22), 2729-2734.
89. Piotrowska-Nowak, A., Elson, J. L., Sobczyk-Kopciol, A., Piwonska, A., Puch-Walczak, A., Drygas, W., & Tonska, K. (2019). New mtDNA association model, MutPred variant load,

- suggests individuals with multiple mildly deleterious mtDNA variants are more likely to suffer from atherosclerosis. *Frontiers in genetics*, 9, 702.
90. Saih, A., Baba, H., Bouqdayr, M., Ghazal, H., Hamdi, S., Kettani, A., & Wakrim, L. (2021). In Silico Analysis of High-Risk Missense Variants in Human ACE2 Gene and Susceptibility to SARS-CoV-2 Infection. *BioMed research international*, 2021.
 91. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., & Ferrin, T. E. (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC bioinformatics*, 7(1), 1-10.
 92. Damian Szklarczyk, Andrea Franceschini, Michael Kuhn, Milan Simonovic, Alexander Roth, Pablo Minguez, Tobias Doerks, Manuel Stark, Jean Muller, Peer Bork, Lars J. Jensen, Christian von Mering, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored, *Nucleic Acids Research*, Volume 39, Issue suppl_1, 1 January 2011, Pages D561–D568,
 93. Kaur, Sargeet, Jitendraa Vashistt, and Harish Changotra. "Computational Investigation of Regulatory Region SNPs of Autophagy Gene BECN1." *Defence Life Science Journal* 9.1 (2024): 27-34.Elkhatabi, Lamiae, et al