DNA REPAIR MECHANISMS AND THEIR ROLE IN ALZHEIMER'S DISEASE

Dissertation submitted in partial fulfilment of the requirement for the degreeof

BACHELOR OF TECHNOLOGY

IN

BIOINFORMATICS

By

APURVA SISONDIA

141506

UNDER THE GUIDANCE OF

DR. TIRATHA RAJ SINGH

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

MAY 2018

TABLE OFCONTENTS

DECLARATION BY THE SCHOLAR

I hereby declare that the work reported in the B-Tech thesis entitled "**DNA repair mechanisms and their role in Alzheimer's disease**" submitted at **Jaypee University of Information Technology, Waknaghat, India,** is an authentic record of my work carried out under the supervision of **Dr.Tiratha Raj Singh.**

I have not submitted this work elsewhere for any other degree or diploma.

…….

Apurva Sisondia

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology, Waknaghat, India

Date:

CERTIFICATE

This is to certify that the project report entitled "**DNA repair mechanisms and their role in Alzheimer's disease**", submitted by **Apurva Sisondia (141506)** in partial fulfilment for the award of degree of Bachelor of Technology in Bioinformatics to Jaypee University of Information Technology, Waknaghat, Himachal Pradesh has been carried out under my supervision.

This work has not been submitted partially or fully to any other university or Institute for the award of this or any other degree or diploma.

Signature of Supervisor	……..
Name of Supervisor	Dr.Tiratha Raj Singh
Designation	Assistant Professor (Senior Grade)
Date:	……...

ACKNOWLEDGEMENT

The project in this report is an outcome of continual work over and intellectual support from various sources. I have taken many efforts in this project. However, it would not have been possible without the kind support and contribution of many individuals. I would like to express my sincere gratitude towards my project supervisor, **Dr.Tiratha Raj Singh**, Associate Professor, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, for his relentless support and guidance, and most of all for his unmatched patience. I am truly grateful to him.

I am highly indebted towards **Ms. Ankita Shukla** (PhD Scholar) for providing necessary information regarding the project, for her kind co-operation, valuable advice and encouragement which helped me in completion of this project.My special thanks also goes to HOD **Dr.Sudhir Sayal** for his invaluable suggestions, scientific attitude, continuous support, guidance, cooperation and encouragement.

I would also take this opportunity to thank **Mrs. Somlata Sharma** for her assistance during the project work.

…….

Apurva Sisondia

Date:

LIST OF FIGURES

LIST OF FLOWCHARTS

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
APP	Amyloid Precursor Protein
Aβ	Amyloid beta
MCI	Mild Cognitive Impairment
ROS	Reactive Oxygen Species
BER	Base Excision Repair
NER	Nucleotide Excision Repair
MMR	Mismatch Repair
NHEJ	Non-Homologous End Joining
HR	Homologous Recombination
AP	apurinic/apyrimidinicsite
SNP	Single Nucleotide Polymorphism

LIST OF TABLES

ABSTRACT

Alzheimer's is a type of neurodegenerative disorder that affects the functioning of neurons. Functional variants and polymorphisms of DNA repair genes have been the focus of several cancer association studies, but only in recent years some of them have been investigated as possible AD risk factors. The few studies performed so far suggest that some variants might play a role in AD pathogenesis and deserve further investigations. This is the basic line of thought that runs throughout the conceptualization of this project.

The idea was to find out the DNA Repair genes involved in the Alzheimer's disease mechanism. The DNA Repair genes that are involved in the disease are OGG1, XRCC1, BRCA1, and PARP1.These genes are extracted from various research papers. The SNPs for these genes are collected from dbSNP Database. Then we found out the damaging and non-damaging regions of those genes. In order to carry out this process we used five different Damage Prediction Tools named SIFT, PROVEAN, POLYPHEN, SNPs&GO, and SNAP. From analysing the results of these tools we concluded that OGG1 and BRCA1 are the two genes that have maximum number of damaging SNPs out of the total SNPs. The gene which is the most damaging is further taken to the structural level analysis by using Functional Site Predictions tools like UET(Universal Evolutionary Trace) and CONSURF-DB. After that, MD Simulations are performed with the help of GROMACS for the structural analysis. Simulations are performed for the wild type as well as for its mutants to compare the differences in the structure and function. There is a growing need to initiate studies on DNA damage and repair and unravel the molecular underpinnings entailed in the etiopathogenesis of the disease. The outcome of such studies substantiates the corner stone streamlined to employ therapeutic strategies.

CHAPTER 1

INTRODUCTION

1.1 Definition

Dementia is not technically a disease, but more of a way to describe a set of symptoms that are poor memory as well as difficulty in learning new information. Usually dementia is caused due to damage to the brain cells (neurons). In such cases, a condition arrives where neurons get damaged and start functioning abnormally. Alzheimer's disease (AD) is the most common type of dementia which is widely spread throughout the world affecting around 48 million people making it a global health crisis that needs to be addressed. In India, around 4.1 million people and in America 5 million people are suffering from AD. [1] It is an age-related progressive neurological brain disorder that slowly develops over a period of years and affects an individual's overall personality, way of talking, memory, and behaviour and finally the patient is left dependent on other caretakers for even carrying out daily activities. This condition is considered as a neurodegenerative disease, meaning it causes the degeneration or loss of neurons in the brain, particularly in the cortex region. The cortex shrinks up, damaging brain regions involved in planning, thinking and remembering. Hippocampus is an area of the cortex which plays a major role in the formation of new memories. Hippocampus is especially the region that suffers with severe shrinkage. Fluid filled spaces within the brain that are known as ventricles also grow larger in case of AD.

70% of all the dementia cases lead to Alzheimer's disease. Out of all the leading causes of deaths in the world, AD stands at 6^{th} position. The number of deaths caused by AD is even more than the number of deaths due to breast cancer and prostate cancer together. Every 66 seconds someone in the United States develops Alzheimer's. According to the statistics, it affects 10% of the population above 65 years of age and almost 50% of the population above 85 years of age. AD is frighteningly common in older people. Out of every 3 seniors, 1 dies due to AD or any other form of dementia. Since 2000, deaths due to heart disease have decreased by 14% while the deaths from AD have increased by 89%. According to the experts, the number of people suffering from AD will double and reach to 81.1 million by 2040, estimated rate of increase being almost 300%.[2]

1.2 History

On November 1906, Dr.Alois Alzheimer, a clinical psychiatrist and neuroanatomist, reported "A peculiar severe disease of the cerebral cortex" in the 37th meeting of the South-west German Psychiatrists in Tubingen. He, for the first time described the condition in his patient named Auguste D who experienced memory loss, paranoia, and other worsening psychological changes. Dr. Alzheimer saw dramatic shrinkage (atrophy) and abnormal deposits in and around nerve cells in her brain during the autopsy. The condition was later named as "Alzheimer's Disease" by Emil Kraepelin, a psychiatrist in 1910. Eventually, major factors such as amyloid protein and tau protein were found in 1984 and 1986 respectively that were declared as the pathological markers of AD. Tacrine, also known as Cognex was the first drug for AD which was approved by FDA.

1.3 Risk Factors of AD

In AD, there is a noticeable shrinkage in the brain that leads to change in structure and function of specific brain regions. The cause of this condition is still not known. However, scientists have found amyloid plaques, neurofibrillary tangles and imbalance of a neurotransmitter called acetylcholine in the brain of AD patients. "Amyloid Cascade Hypothesis" is the most widely researched and discussed hypothesis.

There are several other factors that are responsible to increase the risk for developing the condition:

- Old Age
- Family History
- Smoking
- Down's syndrome
- Severe head injuries
- Cardiovascular disease
- Lifestyle and heart health
- High blood pressure
- Mild Cognitive Impairment (MCI)
- Depression
- Gender

1.4 Signs and symptoms of AD

- Memory loss that intervenes daily activities
- Changes in mood& Social withdrawal
- Misplacing belongings
- Difficulty in completing familiar tasks
- Confusion of time and place
- Poor or decreased judgement
- Struggle in communicating
- Changes in vision and loss of coordination
- Confusion while speaking, writing or reading some words
- Confusing day from night
- Losing the ability to retrace the path
- Disorientation and forgetfulness

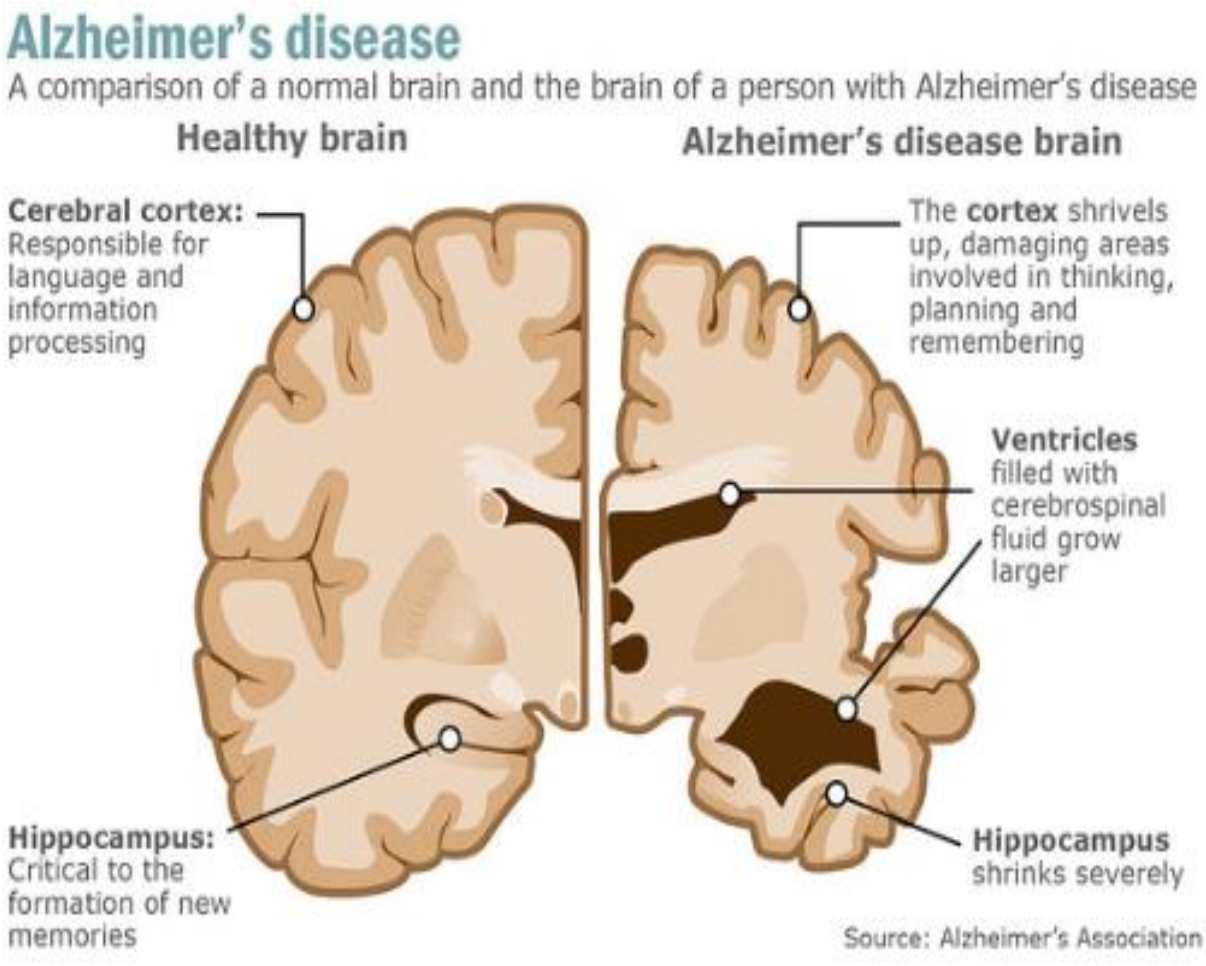


Figure 1.1: A comparison between a normal brain and the brain of a person with AD

1.5 Pathophysiology of AD

The major characteristics of a typical AD affected brain are progressive loss of brain cells, degeneration of synaptic communication, tau protein deposits inside the cell, plaques formation due to deposits of beta amyloid outside the cell. Talking about the cell membrane of the neuron, there is a molecule called Amyloid Precursor Protein (APP) which is embedded in the cell membrane in a way such that half of the APP molecule is inside the cell and half of it is outside the cell. The role of APP is to help the neurons grow and repair itself after some sort of injury or damage. APP is a protein, and just like every other protein, it gets used and over time gets broken down and recycled. Normally, APP protein gets chopped off by an enzyme called alpha-secretase and its partner, gamma-secretase. These chopped up fragments (peptides) that are obtained from the chopping off by the secretases are soluble in the cell environment and get dissolved. In case of AD, the scenario is different. Here, another enzyme beta-secretase teams up with gamma-secretase and this is where the problem begins. This leftover fragment chopped off by beta-secretase is not soluble and creates a monomer called amyloid-beta. These monomers tend to be sticky and bond together just outside the neurons and form clumps known as beta amyloid plaques. These plaque formations get between the neurons and disrupt neuron to neuron signalling. This leads to severe damage to brain cells and affects functions like memory. It is also considered that these plaques can start up an immune response and cause inflammation which might damage surrounding neurons.

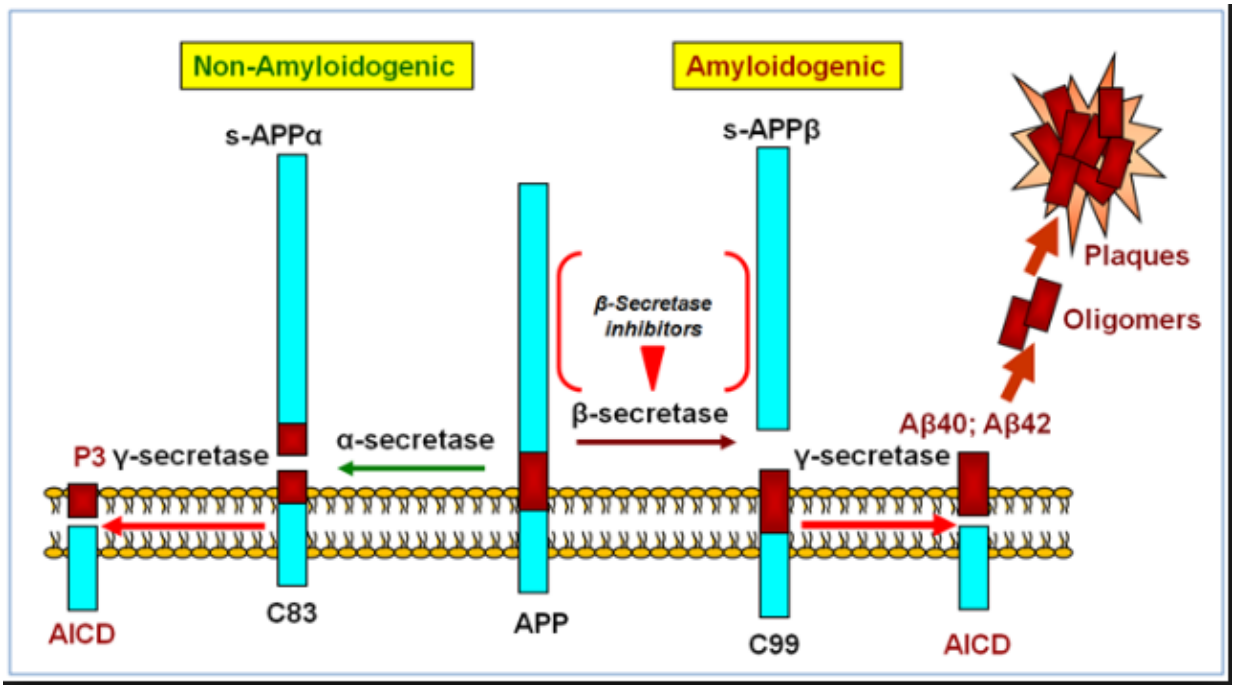


Figure 1.2: Mechanism of α, β and γ secretases in the formation of amyloid plaques

Another major characteristic of AD are neurofibrillary tangles, and these tangles are actually found inside the neurons. Neurons are held together by their cytoskeleton, which is made up of microtubules. The role of these microtubules is to ship the nutrients across the length of the cell. These microtubules contain a special protein called tau, which acts as a binding protein for this cytoskeleton such that it does not break apart. Although, the mechanism is still not completely understood by the scientists, it is assumed that the beta amyloid plaque formation initiates pathways inside the neurons which lead to activation of kinase, an enzyme that transfers phosphate group to the tau protein. At this stage, the structure of tau protein changes, now it no more holds the microtubules together and forms clumps outside the neurons along with the other tau proteins. These tau proteins together form tangles and are called neurofibrillary tangles. Due to the formation of these tangles, the cell is unable to perform its functions and therefore, lead to apoptosis or programmed cell death.

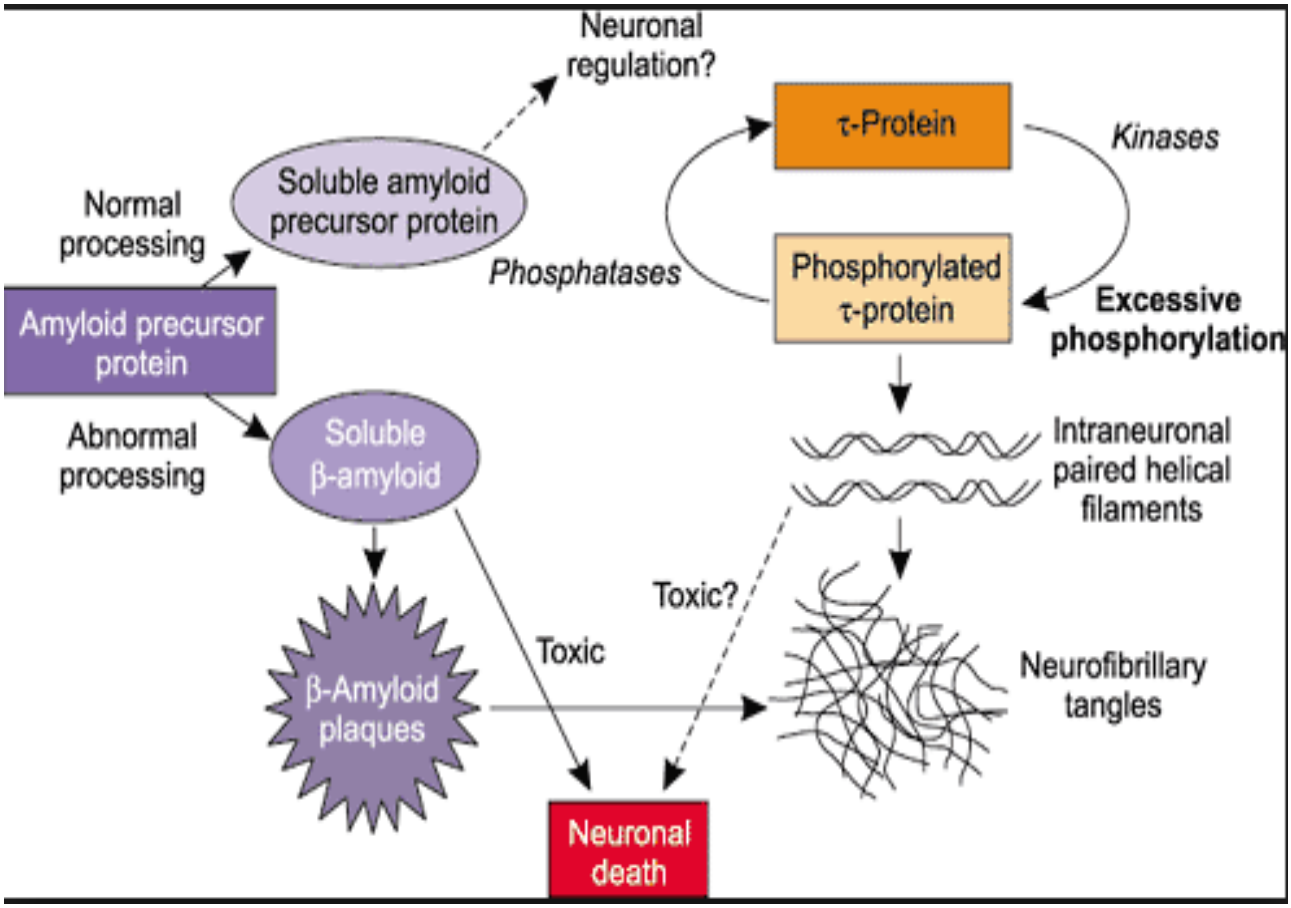


Figure 1.3: Pathophysiology of AD

1.6 Stages of AD

1.6.1 Mild Alzheimer's disease (early-stage)

Toward the starting, issues may occur with a few words or names. Patients experience difficulty recollecting names when acquainted with new individuals. They additionally confront challenges performing errands in social or work settings. Initially, individuals start losing or misplacing valuable items, general forgetfulness being the main characteristic symptom. Patients with early AD sometimes feel lost or disoriented and at times, they lose track of time. In the initial stages of AD, a man may act freely. He may at present drive, work and be a part of social exercises. Notwithstanding this, the individual may feel as though he is having memory lapses. Poor or diminished judgement may be noticed in many individuals.

1.6.2 Moderate Alzheimer's disease (middle-stage)

Direct Alzheimer's is the longest stage and can keep going for a long time. As the ailment advances, the individual with Alzheimer's requires a more prominent level of care. Major changes in behaviour as well as in one’s personality can be seen at this stage. One may see the individual with Alzheimer's getting mistaken for words, getting baffled or furious or behaving in most unexpected ways. Harm to neurons in the mind makes it hard to express considerations and perform routine undertakings. At times, patients can't recall their own address or phone number or the secondary school or college from which they graduated. Individuals face difficulties with language and communication and also require assistance with daily activities.

1.6.3 Severe Alzheimer's disease (late-stage)

In the last phase of this condition, people lose the capacity to react to their surroundings, have aggravated trouble in communicating and eventually, to control development. As memory and psychological abilities keep on degrading, huge identity changes occur and people require continuous help with every day exercises. Patients lose attention to late encounters and in addition of their environment, in this manner they require round-the-clock help with day by day exercises and individual care. People additionally encounter changes in physical capacities, including the capacity to walk, sit and, in the end, swallow (dysphagia). Reaching at this stage, the patient is almost bedridden, is unable to recognise familiar people and objects, walk of the individual becomes unstable and loses control over urination (incontinence).

1.7 Treatment of AD

Nowadays, two drugs are widely being used to treat the symptoms of AD. First group of drugs are Cholinesterase inhibitors: These drugs provide neurotransmitter named acetylcholine to the brain cells which keeps on depleting in cases of AD brains to boost up the cell to cell communication. Secondgroup of drugs in usage areMemantine (Namenda): This drug reduces the pace of symptoms in AD brains by working in communication network of the brain.AD is an irreversible condition; the medicines can only improve symptoms or slow down progression. Though, various preventive measures can be taken to slow down the progression, which are exercising regularly, taking a healthy diet, having adequate sleep, and taking part in an effective communication. These measures might help the patients to cope up with the condition.

1.8 DNA Damage

Constant exposure to various exogenous as well as endogenous agents poses huge threat of damage and alteration to our cells. Such exposures cause damage to our DNA, generate deleterious genetic variations, change the structure of DNA and alter basic biological processes like transcription and translation. DNA damage is mainly of four types: base and sugar modifications, single and double-strand breaks, DNA-protein cross-links, and base free sites. Defects in DNA repair genes can lead to various disorders such a neurodegenerative disease, reduced functioning of the immune system, premature ageing as well as increased risk for developing cancer. DNA damage occurs at a rate of 10,000 to 1,000,000 molecular lesions per cell per daydue to various environmental factors and normal metabolic processes being carried out inside the cell. Out of 6 billion bases (3 billion base pairs)in the human genome, this damage contributes only 0.000165%. These unrepaired damages in important genes (such as tumour suppressor genes) can impair the ability of the cell to perform its function and increase the chances of tumour formation. These damages to DNA change its double helical structure (spatial configuration) by introducing modifications to the nucleotide bases. DNA does not have a tertiary structure just like other proteins and RNA, therefore damage does not occur at that level but DNA and the histones around which it is coiled, both are vulnerable to the damage. There are various causes of DNA damage including cellular metabolism, UV light exposure, ionizing radiation and chemical exposure, replication errors, reactive oxygen species (ROS), X-rays and other alkylating agents.

1.9 DNA Repair

Cells have a specialized DNA repair mechanism developed within them in order to correct the damages introduced to DNA. DNA repair mechanism is a collection of various processes that identifies the damage caused to the DNA and corrects those lesions through various repair pathways. Correction of DNA is very important because in case, bases are not corrected, these mutations may be deleterious for the cell and might end up killing the cell. Due to various environmental and other factors, there are almost 1 million molecular lesions per cell per day and this is the reason why DNA repair mechanism is constantly active in the cell and keeps correcting the damages occurring. Based on the type of DNA lesion, DNA repair mechanisms can be subdivided into several distinct mechanisms.

1.10 DNA Repair Pathways

All six DNA repair pathways are listed below:

- Base Excision Repair (BER)
- Nucleotide Excision Repair (NER)
- Mismatch Repair (MMR)
- DNA strand cross-link repair
- Non-homologous end joining (NHEJ)
- Homologous Recombination (HR)

BER pathway is a mechanism that identifies the damaged bases in the DNA and repairs those bases through the entire cell cycle. BER is responsible for removing small and non-helix-distorting base lesions from the genome. On the other hand, NER is responsible to remove bulky helix distorting lesions. In mammals, this is the main pathway which is used to repair the damaged DNA. Role of mismatch repair pathway is identifying and repairing deleterious insertions, deletions and substitutions caused during the process of DNA replication and recombination. In crosslinking of DNA strand, formation of a covalent bond between two nucleotides of DNA takes place. These mutations hinder the cell's ability to function normally and can even cause cell death. NHEJ repair mechanism repairs the double strand breaks that occur due to various damage causing agents. This process is termed as non-homologous because there is no need to find a homologous strand for joining. HR is a process in which two strands of the DNA exchange few nucleotides (short fragments).

Here, in this study, the main aim is to study specifically about the BER Pathway.

1.10.1 Base Excision Repair (BER)

BER pathway is the most common pathway used to remove incorrect bases (e.g. uracil) and the damaged bases (e.g. 3-methyladenine). The role of BER is to repair the damages caused to DNA and its works throughout the entire cell cycle. It more specifically aims at removing small, non-helix distorting lesions (that do not distort the helical structure of DNA) from the genome. BER pathway is initiated by the enzyme called glycosylase. These glycosylases identify the damaged bases and removes the lesions from that site making that site an AP site (apurinic/ apyrimidinic site in DNA that does not contain any base). Enzyme called AP endonuclease then comes into action and cleaves these AP sites. At this step, the single strand break can be processed by two methods: either by short patch BER or by long patch BER. In short patch, single nucleotide is substituted whereas in case of long patch, a short fragment of 2-10 nucleotides are replaced.

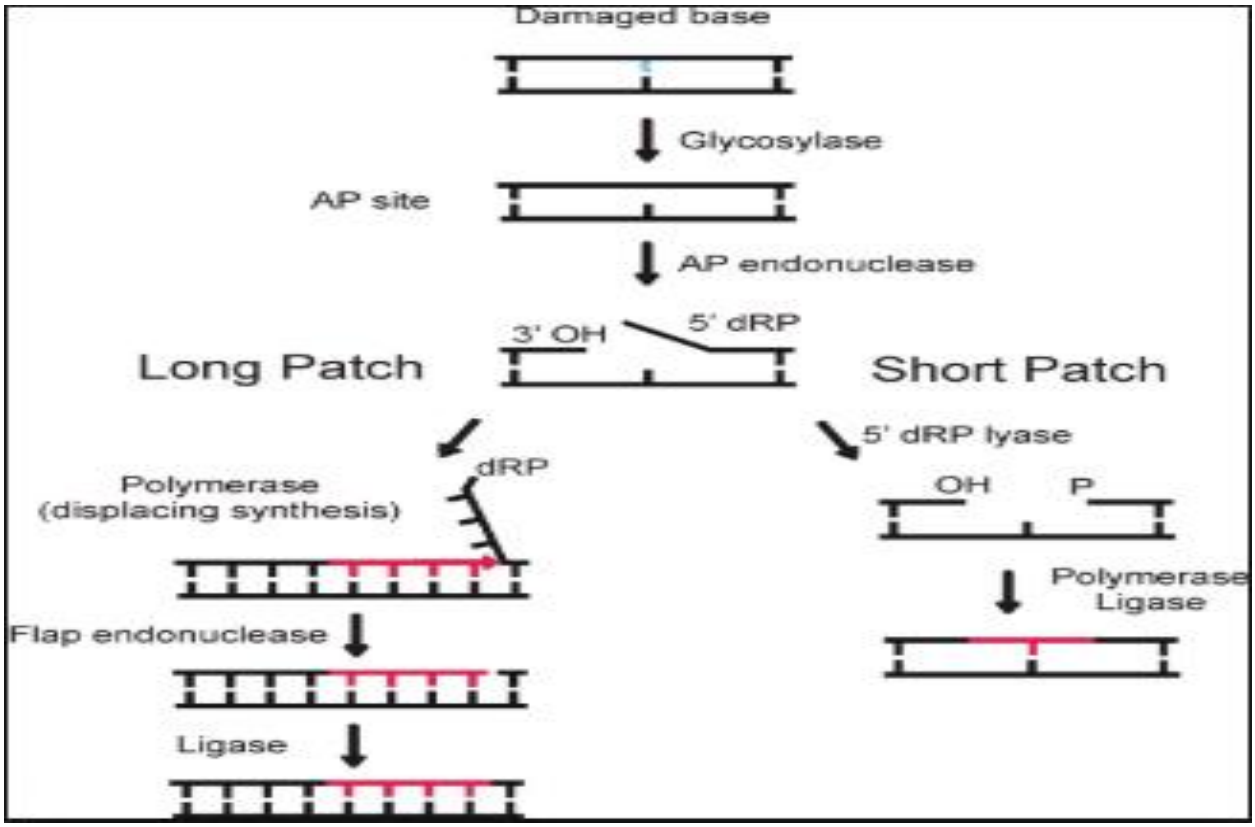
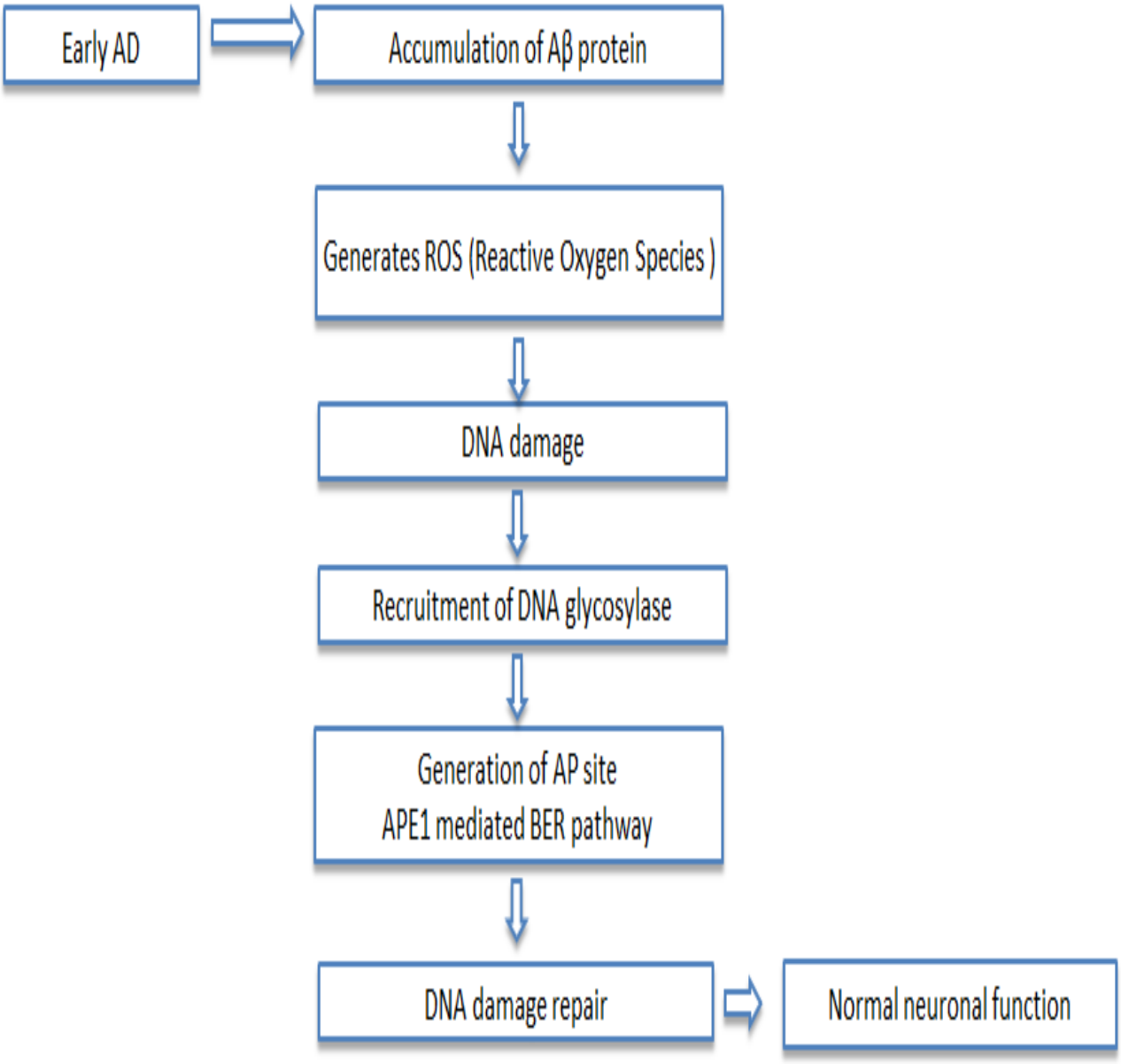


Figure 1.4: Mechanism of the BER Pathway (Long patch and short patch)

DNA Damage and Repair Mechanism:



Flowchart 1.1: DNA damage and repair mechanism

CHAPTER 2

REVIEW OF LITERATURE

Till date, many experiments have been performed on DNA repair mechanisms in mitotic cells like cancer cell lines but such kind of study linked with post-mitotic cells like neurons, still remains unexplored and therefore, it can be the most exciting topic for further studies. Beta amyloid plaques and neurofibrillary tangles, also called as markers of DNA are found in cerebral and peripheral regions of AD brains. Recent studies stated that oxidative DNA damage was one of the earliest detected damage that differentiated between a normal and an AD affected brain. In AD, amount of reactive oxygen species (ROS) increases and BER is unable to repair this oxidative damage. This process is observed by several researchers in which AD and MCI brains suffer from severe oxidative DNA damage followed by reduced BER activity. In this condition, DNA strands break and repair pathways are unable to perform their function. The exact cause of AD is not known, but according to the studies performed so far, all these are the possible factors that might lead to deaths of neurons and ultimately leading to memory loss. Studies carried out till date suggests that some DNA repair gene variants might play a role in AD mechanism.

Every cell in our body carries out two mechanisms at the molecular level – DNA damage and repair. Any imbalance caused in these two mechanisms can lead to DNA damage which can further have harmful effects (e.g. neuron-degeneration). DNA damage has remained a major cause of several diseases like ageing and disorders like AD since ages. Though DNA damage is involved in several processes, yet there are few studies that have been carried out in this field. Main cause of DNA damage is Reactive Oxygen Species in nucleus as well as in mitochondria. BER is the major pathway that helps in developing and maintaining the Central Nervous System (CNS) and by repairing the damages. DNA damage and repair related studies must be carried out to properly understand the underlying mechanism. There is an alarming need to carry out these studies, observing the pace at which AD is spreading. This study specifically focuses on BER and states that BER may worsen the neuronal loss in case of damaged DNA repair genes.

Recent studies have been performed to find out whether polymorphisms of BER related genes are responsible in causing AD or not. SNPs of BER genes might be a reason for reduced BER efficiency. Method used in the study was SNP genotyping. In this method, DNA was isolated from the peripheral regions of the mononuclear cells. The samples were collected from 120 patients suffering from AD and 110 healthy individuals (volunteers). Samples were genotyped on the basis of the presence of BER related SNPs.

A positive association was found between AD risk and presence of BER related SNPs in these three genes – XRCC1, MUTYH, and PARP1. The conclusion obtained from this study is that BER gene variants might be an important factor in causing AD.

The main hallmarks of AD are beta-amyloid plaques, neurofibrillary tangles and excessive neuronal loss. Recently, it has been observed that DNA repair mechanisms get damaged and are unable to function normally leading to more loss of neurons and the condition gets worse. This study shows that Arg194Trp polymorphism introduced in XRCC1 gene might play an important role in AD pathogenesis. A case study from turkey was carried out and found that SNP Arg194Trp in XRCC1 gene was linked with late onset AD (LOAD). This research suggests only a borderline association of XRCC1 with AD. This polymorphism requires further investigation in larger population. Another study was performed on XRCC1 gene in which they hypothesized that polymorphism Arg194Trp may be associated with increased risk of AD. To test this hypothesis, DNA samples were taken from 98 AD patients and 95 healthy individuals. The frequency of occurrence of Tryptophan at position 194 was more pronounced in AD cases rather than healthy subjects. This study also shows borderline association of XRCC1 gene polymorphism with AD since the results were not statistically significant to prove the point.

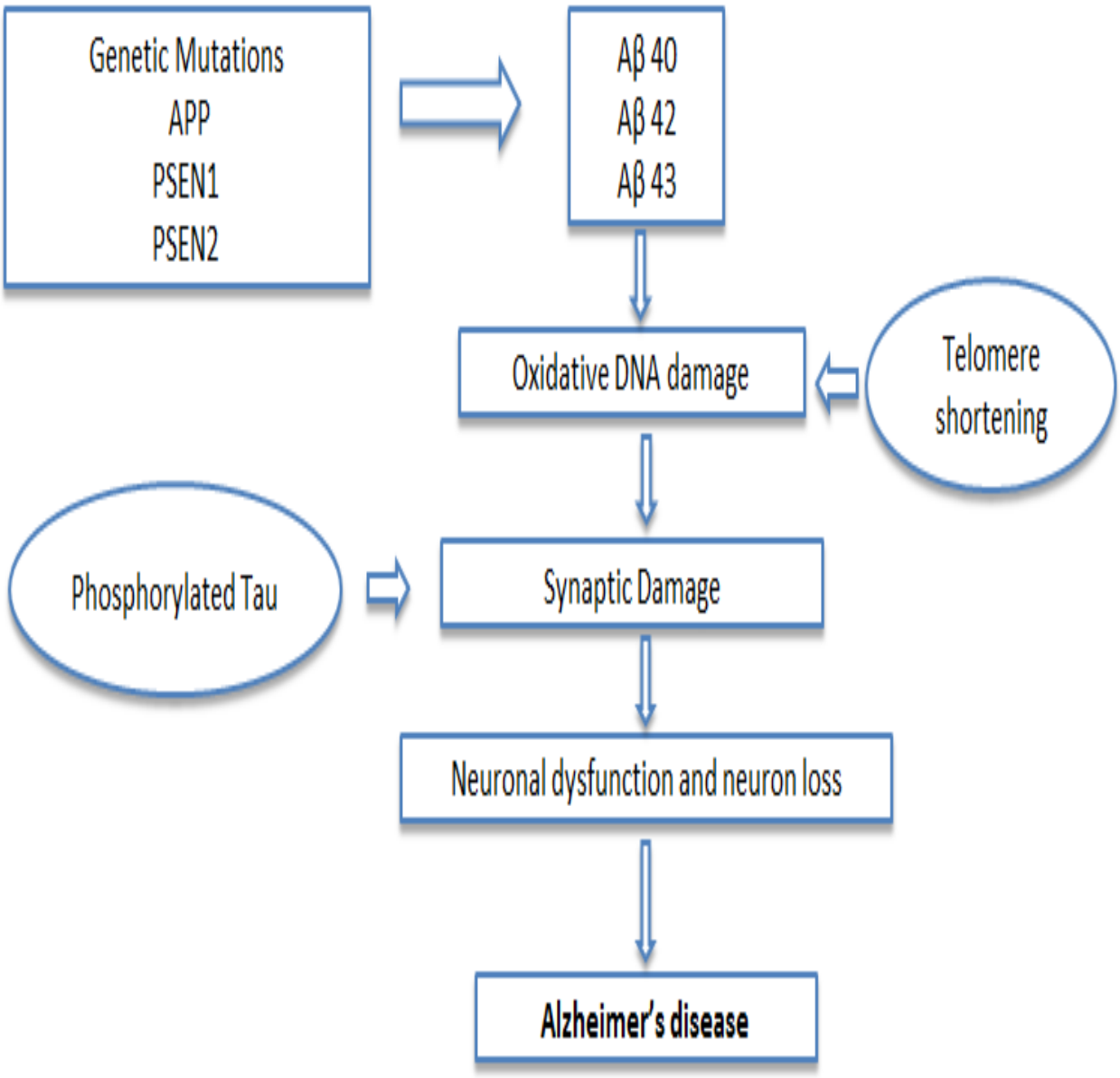
Another study shows that BRCA1 gene might play a role in increased risk of AD. Reduced amount of BRCA1 has been observed in the AD brains of humans as well as mice. In order to check whether there is any change in BRCA1 gene of humans with AD, post-mortem brain sections were immune stained. The brain sections were extracted from people who had no signs of cognitive deficiency and from patients with mild cognitive impairment (MCI) or AD within antibody against BRCA1.After performing this experiment, scientists found that level of BRCA1 gene in the inferior parietal cortex of AD and MCI brains was 50 to 70% lower than their healthy counterparts.

2.1 List of all the DNA repair genes and their association with AD obtained from the literature:

GENE/COMPLEX	ASSOCIATION WITH AD	VARIANT
poly-ADP-ribose polymerase-1 (PARP1)	inc. in AD brains	-
oxidized purine nucleoside triphosphatase (MTH1)	dec. in AD hippocampus and inc. in AD cortex	-
Mre11 DNA repair complex	dec. in the neurons of AD cortex	-
DNA protein kinase (DNA-PK) complex	dec. in AD brain extracts	-
apurinic/apyrimidinic endonuclease 1 (APE1)	inc. in AD brains	-
8-oxoguanine DNA glycosylase (OGG1)	dec. in several AD brain regions	-
DNA polymerase beta (Pol beta)	dec. in several AD brain regions	-
Uracil DNA glycosylase (UDG)	dec. in several AD brain regions	-
8-oxoguanine DNA glycosylase (OGG1)	Not associated with AD risk	Ser326Cys
8-oxoguanine DNA glycosylase (OGG1)	Found in AD samples, but not in controls	C796del
8-oxoguanine DNA glycosylase (OGG1)	Found in AD samples, but not in controls	Ala53Thr
8-oxoguanine DNA glycosylase (OGG1)	Found in AD samples, but not in controls	Ala288Val
X-ray epair crosscomplementing protein 1 (XRCC1)	Borderline association with AD risk	Arg194Trp
Xerodermapigmentosum group D (XPD)	Not associated with AD risk	XPD C156A
Xeroderma pigmentosum group D (XPD)	Not associated with AD risk	XPD A751C
Xeroderma pigmentosum group F (XPF)	Not associated with AD risk	XPF T824C
Breast cancer Associated gene (BRCA1)	Borderline association with AD risk	-

Table 2.1: List of DNA repair genes associated with AD (only BER pathway specific)

2.2 DISEASE MECHANISM:

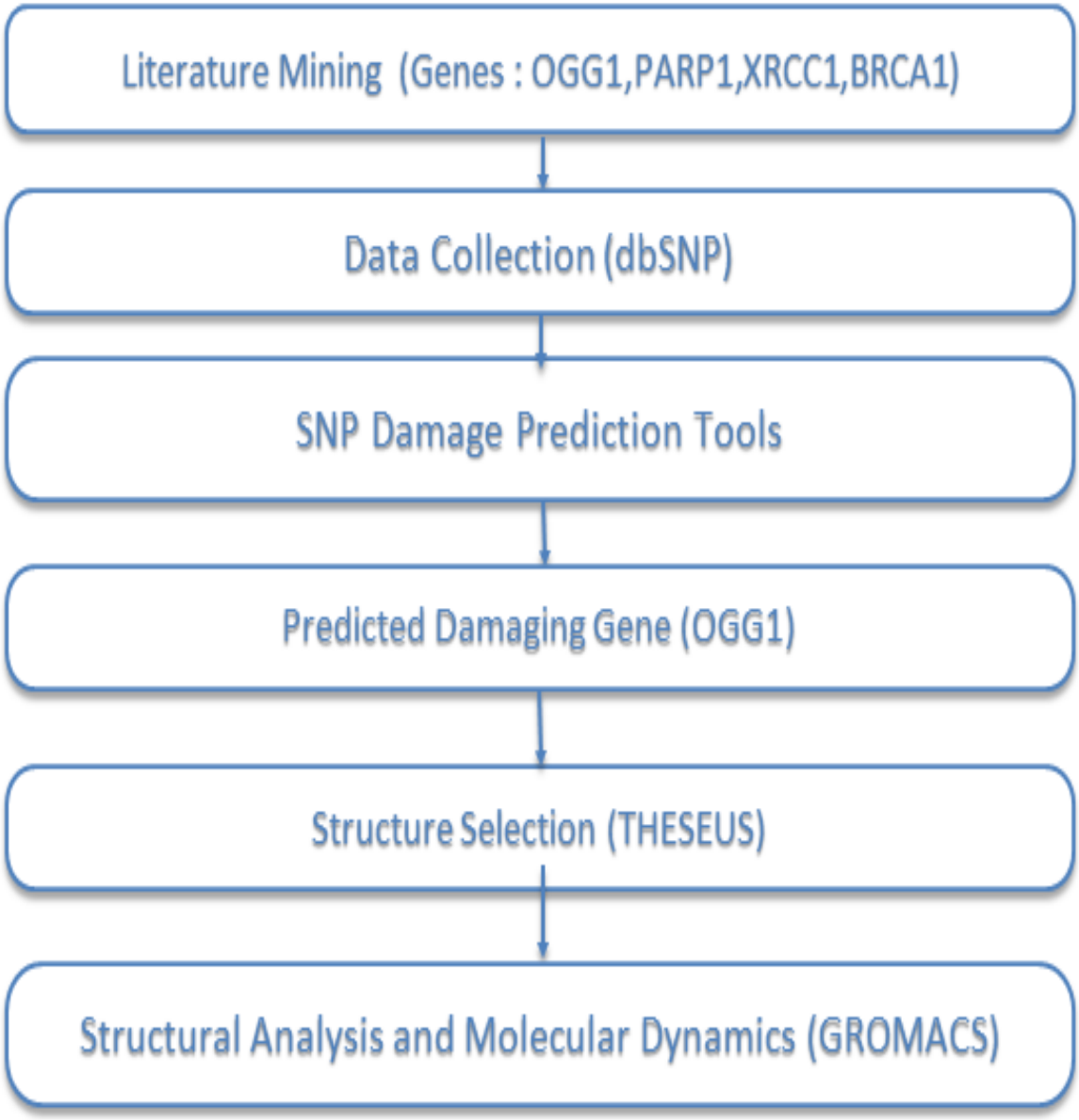


Flowchart 2.1: Alzheimer's disease mechanism

CHAPTER 3

MATERIALS AND METHODS

3.1 METHODOLOGY



Flowchart 3.1: Methodology followed in this study

3.2 LITERATURE MINING

Four main genes are extracted from mining the research papers. These four genes are involved in the BER pathway and are associated with DNA Repair mechanism in AD. A decrease in BER activity has been noticed in the past few years in the brains of AD patients thus compelling scientists to reach a hypothesis that the AD patients suffer from DNA damage as well as decreased activity of DNA repair pathways also occurs. There are many genes associated with DNA repair but these are particularly involved in the BER pathway.

These four DNA repair genes associated with AD are-

1. OGG1
2. BRCA1
3. XRCC1
4. PARP1

3.3 DATA COLLECTION

Next step is to extract SNPs for these genes. Aim is to look into mutations in the genes due to which there is decrease in the BER activity and these genes are not able to repair the DNA in case the DNA is damaged. SNPs are extracted from dbSNP database.

dbSNP:

dbSNP database is a catalogue of variations at a genetic level. It contains all the variations whether it is in form of single nucleotide polymorphisms, small insertion/deletion polymorphisms, invariant regions of sequence, microsatellite repeats, named variants, and uncharacterized heterozygous assays. Single nucleotide polymorphisms (SNPs) are the most common form of genetic variations. Only mis-sense mutations are considered and synonymous mutations are ignored since many of them are not harmful.

From dbSNP database the following data is obtained:

1. OGG1 - 10 SNPs
2. BRCA1 - 38 SNPs
3. XRCC1 - 24 SNPs
4. PARP1 - 18 SNPs

	OGG1	BRCA1	PARP1	XRCC1
1.	rs11548133	rs80357107	rs78381515	rs25487
2.	rs113572618	rs80356959	rs3219145	rs2307177
3.	rs104893751	rs80356959	rs61731502	rs2682557
4.	rs17050550	rs80356942	rs1059040	rs2307166
5.	rs56053615	rs80357268	rs11541664	rs2307167
6.	rs1805373	rs80357393	rs1136420	rs41561817
7.	rs55667729	rs80357332	rs1136410	rs25474
8.	rs3219012	rs80357212	rs1059011	rs72554204
9.	rs113561019	rs80357286	rs79529505	rs2307184
10.	rs115609368	rs80357477	rs3219062	rs2271980
11.	-	rs1800751	rs2230484	rs25491
12.	-	rs80357040	rs1805415	rs25490
13.	-	rs80357451	rs3219057	rs2307188
14.	-	rs80357216	rs61750985	rs25489
15.	-	rs28897698	rs111635488	rs111857343
16.	-	rs80357358	rs1805409	rs1799782
17.	-	rs80357241	rs113258217	rs56357789
18.	-	rs80356920	rs3738708	rs2307191
19.	-	rs80357149	-	rs2307180
20.	-	rs80357078	-	rs2228487
21.	-	rs80357069	-	rs25496
22.	-	rs80357069	-	rs2307171
23.	-	rs80357065	-	rs2307186
24.	-	rs55808233	-	rs11553659
25.	-	rs80357012	-	-
26.	-	rs80357474	-	-
27.	-	rs80357041	-	-
28.	-	rs80357112	-	-
29.	-	rs1800757	-	-
30.	-	rs41293463	-	-

31.	-	rs80357428	-	-
32.	-	rs80357324	-	-
33.	-	rs80357025	-	-
34.	-	rs80357025	-	-
35.	-	rs80357463	-	-
36.	-	rs80357281	-	-
37.	-	rs80357007	-	-
38.	-	rs80356905	-	-

Table 3.1: List of all the SNPs obtained from dbSNP for all the four genes

3.4 SNP DAMAGE PRECTION TOOLS

After obtaining the SNPs for all the four genes, Damage Prediction process is carried out to find out as to how much damaging these SNPs actually are. Various Damage Prediction tools are available online that predict whether an amino acid substitution affects protein function. If it affects the structure or function of the protein, then it is deleterious or non-deleterious. In case the SNP is deleterious then it is taken into consideration for further investigation at the structural level. For getting the maximum accuracy and maximum majority, **10 SNP Damage Prediction Tools** are used to perform the Damage Prediction process. These 10 tools are:

1. SIFT
2. PROVEAN
3. POLYPHEN
4. SNAP
5. SNPs&GO
6. PredictSNP
7. PANTHER
8. MutPred
9. MAPP
10. Phd-SNP

Further it is explained as to how each of these tools works, which algorithm each one of them follows and what output they provide:

- **SIFT** predicts whether an amino acid substitution has an effect on protein function. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignment derived from closely related sequences, obtained from PSI-BLAST. SIFT can be used for both naturally occurring non-synonymous polymorphisms and laboratory-induced missense mutations.

- **PROVEAN** (**Pro**tein **V**ariation **E**ffect **An**alyzer) is a tool which predicts whether an amino acid substitution or indel affects the protein function. It is used in filtering the sequence variants to identify non-synonymous or indel variants which are predicted as functionally important.

- **PolyPhen-2** (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.

- **SNAP** is a neural-network based method that uses in silico derived protein information (e.g. secondary structure, conservation, solvent accessibility, etc.) in order to make predictions regarding functionality of mutated proteins.

- **SNPs&GO** is a server for the prediction of single point protein mutations likely to be involved in the insurgence of diseases in humans.

- **Consensus** is a classifier for prediction of disease-related mutations. PredictSNP1 offers its users a consensus score based on the output of six different amino acid-based predictors.

- **PANTHER** (**P**rotein **AN**alysis **TH**rough **E**volutionary **R**elationships) Classification System was designed to classify proteins (and their genes) in order to facilitate high-throughput analysis. Proteins have been classified.

- **MutPred2 (**Predict the pathogenicity of amino acid substitutions and their molecular mechanisms) is a standalone and web application developed to classify amino acid substitutions as pathogenic or benign in human.

- **MAPP** (Multivariate Analysis of Protein Polymorphism) MAPP predictions are based on assessment of the physicochemical variation in each column of a sequence alignment.

- **PhD-SNP** is based a SVM-based classifier. It is a predictor of human Deleterious Single Nucleotide Polymorphisms. In the new version we developed a predictor based on a single SVM trained and tested on protein sequence and profile information.

These damage prediction tools predict the most damaging gene i.e. OGG1.

3.5 SEQUENCE ANALYSIS

FUNCTIONAL SITE PREDICTION TOOLS: In order to verify these damaging SNPs, functional site prediction tools are run for the sequence as well as the structural analysis of the protein. These tools predict the sites that are evolutionary more conserved and what all sites are functionally more important. After we have the most damaging genes with us, then structure analysis is performed by running Functional Site Prediction Tools **(CONSURF-db and UET).**The scores of these tools justify the predicted damaging and non-damaging regions.

3.6 STRUCTURE SELECTION

Structure selection is done using a program called THESEUS. THESEUS is a program that simultaneously superimposes multiple macromolecular structures. Instead of using the conventional least-squares criteria, Theseus finds the optimal solution to the superposition problem using the method of maximum likelihood (ML). The ML method down weights variable regions of the superposition and corrects for correlations among atoms, producing much more accurate results. After the study is narrowed down to one most damaging gene, structures of that protein are extracted from PDB. Since there are many structures that are obtained from PDB, we need to select just one structure for further structural analysis.

Therefore, this tool is required to superimpose all the structures obtained from PDB and find out the most accurate structure which is closest to the average of all the structures. As an output it gives the value of a median structure and this structure is best fit for further structure analysis and MD simulations.

3.7 MUTAGENESIS

Mutations are artificially induced in the wild type structure of the protein according to the SNPs found and predicted damaging by the Damage Prediction Tools. The steps of introducing a site-specific mutation are:

1. PyMol has a Mutagenesis Wizard to easily make mutations.
2. Load a PDB file in PyMol.
3. In the Wizard menu, select Mutagenesis option.
4. In the PyMol viewer window, select a residue.
5. Select No Mutation and select resultant residue.
6. Also choose the rotamer that best fits our structure.

3.8 STRUCTURAL ANALYSIS

Structural analysis is done using various online tools. Tools used for analysis are:

- CONSURF
- STRIDE
    - Visualization
    - Ramachandran Plot
- PDBSum

CONSURF server is a bioinformatics tool for estimating the evolutionary conservation of amino/nucleic acid positions in a protein molecule based on the phylogenetic relations between homologous sequences. The degree, to which an amino (or nucleic) acid position is evolutionarily conserved, is strongly dependent on its structural and functional importance.

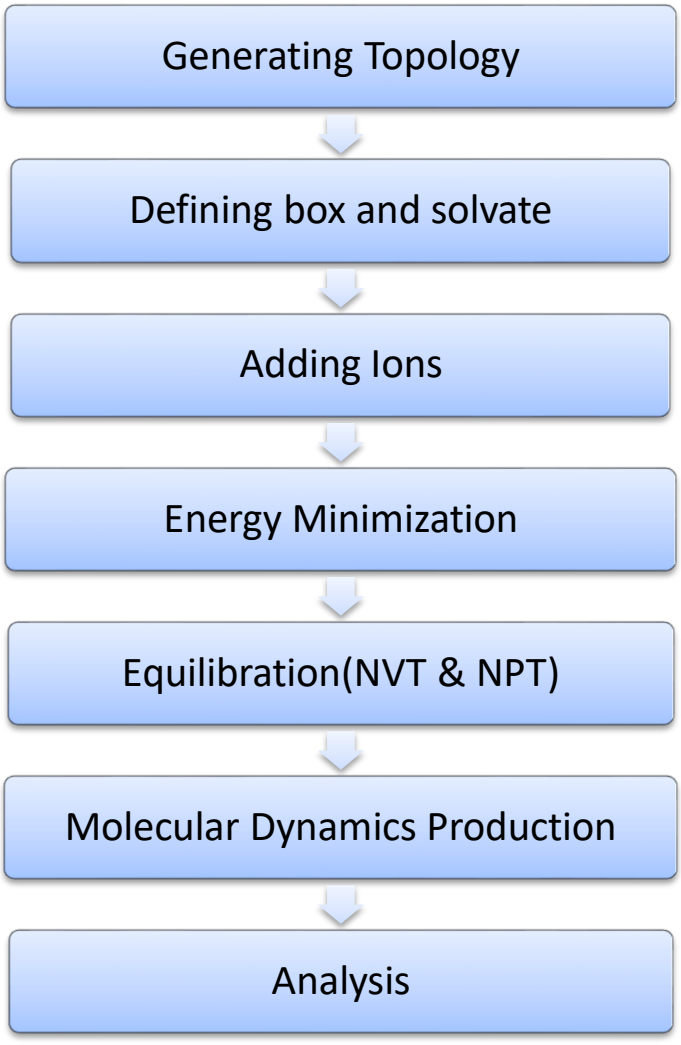
STRIDE is a software tool for secondary structure assignment from atomic resolution protein structures. It implements a knowledge-based algorithm that makes combined use of hydrogen bond energy and statistically derived backbone torsional angle information and is optimized to return resulting assignments in maximal agreement with crystallographers' designations.

This tool allows visualization of the secondary structure, as well as contact and Ramachandran maps for any file uploaded by the user with atomic coordinates in the Protein Data Bank (PDB) format. It also gives the conservation score for each amino acid at every position. A Ramachandran plot is a way to visualize energetically allowed regions for backbone dihedral angles ψ against φ of amino acid residues in protein structure.

PDBSumis a pictorial database that provides an at-a-glance overview of the contents of each 3D structure deposited in the Protein Data Bank (PDB). It shows the molecule(s) that make up the structure (i.e. protein chains, DNA, ligands and metal ions) and schematic diagrams of the interactions between them.

3.9 MD SIMULATIONS (GROMACS)

3.9.1 GROMACS STEPS



Flowchart 3.2: Steps in GROMACS to run a MD Simulation

Step 1: Reconstructing the structure with the missing residues by using Swiss PDB Viewer.

As the structure is opened in the software, it asks for reconstructing the entire structure with missing residues, atoms and chains. Without this step GROMACS does not accept the Structure and throws an error right in the first step of pdb2gmx. PDB file should only contain protein atoms, all other data must be removed from the file.

Step 2: Topology Generation

Convert pdb file to gromacs file format by executing pdb2gmx

Command – gmx pdb2gmx –f 1ko9.pdb –o 1ko9_processed.gro – water spce

Parameters -

-o: output

-water spce: solvent type

Immediately after this command runs, a force field option needs to be selected out of 15 force fields. OPLS-AA/L all-atom force field (option no. 15) is chosen. The next molecule type is the solvent, in this case SPC/E water option is chosen. Other options for water include SPC, TIP3P, and TIP4P. "-water spce" is chosen and passed to pdb2gmx.

The pdb2gmx command generates three files:

1. Topology for the molecule.
2. Position restraint file.
3. Post-processed structure file.

The command generated three new files: 1ko9_processed.gro, topol.top, and posre.itp. 1ko9_processed.gro is a GROMACS-formatted structure file that contains all atoms defined within the force field. The topol.top file is the topology of the system. The topology contains all the information necessary to define the molecule within a simulation. This information includes non-bonded parameters (atom types and charges) as well as bonded parameters (bonds, angles, and dihedrals). The posre.itp file contains information used to restrain the positions of heavy atoms. This file keeps atoms in place while equilibration.

Step 3: Defining Box and solvate

In this step, we provided the molecule with a simple aqueous system.There are two steps to define the box and fill it with solvent:

1. Define the box dimensions using the editconf module.
2. Fill the box with water using the solvate module

Command – gmxeditconf –f 1ko9_processed.gro –o 1ko9_newbox.gro –c –d 1.0 –bt cubic

Parameters –

-f: file input

-c: Places the box in the center

-d: Distance between the protein and the box edge

-bt: Box type

Option -bt defines the box shape: triclinic represents triclinic box, cubic represents rectangular box with all four sides equal, dodecahedron represents a rhombic dodecahedron and octahedron represents a truncated octahedron. After this step, the generated newbox.gro is converted to pdb file format and is visualized in PyMOL. The molecule must properly fit in the box and should be nearly in the centre of the box. In case, the molecule is slightly coming out of the box and is not completely fitting in the dimensions, then the shape of the box needs to be changed.

Command for conversion of .gro file to .pdb format –

trjconv –s solvated.gro –f solvated.gro –o solvated.pdb

After defining the box, next requirement is to fill the box with solvent(water). This process of solvation is completed by the command solvate.

Command – gmx solvate -cp 1ko9_newbox.gro -cs spc216.gro -o 1ko9_solv.gro -p topol.top

Parameters –

-cp: Configuration of protein

-cs: Configuration of solvent

Step 4: Adding ions

Total charge on the system = -1

Now the solvated system is ready but this system contains a net charge of -1. Since there is no system in this universe that exists with a net charge, therefore there is a need to add ions to balance the system. 1 Na (Sodium) ion is added to neutralize the negative charge.

Command – gmxgrompp -f ions.mdp -c 1ko9_solv.gro -p topol.top -o ions.tpr

ions.mdp is a sample file which is downloaded from the website GROMACS Tutorial. gromppcommand is used to assemble the structure, topology, and simulation parameters into a binary input file (.tpr).

Command – gmxgenion -s ions.tpr -o 1ko9_solv_ions.gro -p topol.top -pname NA -nname CL -np 1

Parameters –

-s: structure/state file

-p: Topology-

-pname: Positive ion name

-nname: Negative ion name

-np: number of positive ions needed to neutralize the system

-nn: number of negative ions needed to neutralize the system

Choose an option –

13: “SOL” for embedding ions

Step 5: Energy Minimization

Now the system which is obtained till this step is solvated as well as electroneutral. Before beginning with the Molecular Dynamics, it must be ensured that the system is relaxed and contains no steric clashes or any inappropriate geometry. For this Energy Minimization step is done.

Command – gmxgrompp –f minim.mdp –c 1ko9_solv_ions.gro –p topol.top –o em.tpr

Command – gmxmdrun –v –deffnmem

Parameters –

-v: Makes mdrun verbose such that it displays its progress on the screen at every step

-deffnm: Defines the file names of the input and output

Two factors to evaluate to determine if EM was successful-

1. E_{pot} should be negative (Potential Energy)
2. F_{max}< 1000 KJ/mol/nm

Command – gmx energy –f em.edr –o potential.xvg

Analysis:

Potential Energy

Command - gmx energy -f em.edr -o potential.xvg

Tool – xmgrace (2D plotting tool)

Now when the system is at an energy minimum, real dynamics can be carried out.

Step 6: Equilibration

After having the appropriate geometry and solvent orientation, next step is to equilibrate the solvent and ions around the protein.If at this point unrestrained dynamics are applied, the system may collapse. The main reason is that the solvent is optimized within itself, and not necessarily with the solute. It needs to be brought to the temperature we want to simulate and establish the proper orientation about the solute (the protein). After arriving at the correct temperature (based on kinetic energies), apply pressure to the system until it reaches the proper density. Equilibration is carried out in two phases – NVT and NPT. The first phase is carried out under an NVT ensemble (constant Number of particles, Volume, and Temperature) also known as "isothermal-isochoric" or "canonical." A 100 ps NVT equilibration is conducted in this process.

Command - gmxgrompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr

Command - gmxmdrun -deffnmnvt

Analysis of temperature progression:

Command - gmx energy -f nvt.edr -o temperature.xvg

After stabilizing the temperature, next step is to stabilize the pressure before fully preparing the system for simulations. The second phase is carried out under an NPTensemble,(constant Number of particles, Pressure, and Temperature).The ensemble is also referred to as "isothermal-isobaric" ensemble. A 100 ps NPT equilibration is conducted in this process.

Command – gmx grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr

Command – gmx mdrun –deffnm npt

Parameters –

-c: Coordinate file

-t: To include the checkpoint file

Analysis of pressure progression:

Command - gmx energy -f npt.edr -o pressure.xvg

Choose option: 16

Analysis of density progression:

Command – gmx energy -f npt.edr -o density.xvg

Choose option: 22

Step 7: Production MD

On the completion of the two equilibration phases, the system is now fully-equilibrated at the required temperature and pressure. We are now ready to release the position restraints and run production MD for data collection. We run a 20-ns MD simulation. Script named md.mdp is available in the documentation of GROMACS. This script is picked up from here and is modified according to our own requirements. The simulation is run on 20-ns (in script number of steps are updated to 10,000,000 and time is modified to 20-ns i.e. 20,000 –ps)

Command - gmxgrompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md_20.tpr

Command - gmxmdrun -deffnm md_20

Analysis Commands –

1. To concatenate the simulation files, command trjcat is used. This command produces a full trajectory file for complete 20 ns as output.

Command – gmxtrjcat –f md_0_10.trr/md_11_16.trr/md_0_1.trr –o 1ko9_full_20ns.trr

2. Next command is for concatenating multiple trajectory files. “nojump” checks if atoms jump across the box and then puts them back. This has the effect that all molecules will remain whole (provided they were whole in their initial conformation).

Command – gmxtrjconv –f 1ko9_full_20ns.trr –pbcnojump –o 1ko9_full_nojump.trr

3. Next command is the skip command to skip between the nanoseconds and also converts the trajectory file to PDB format for further analysis.

Command – gmxtrjconv –f 1ko9_full_nojump.trr –s md_20.tpr skip 10 –o 1ko9_full_nojump.pdb

Global Analysis –

1. **RMSD:**
   Command – gmxrms –s md_0_1.tpr –f md_0_1.trr –o fullmut1_rmsd.xvg
2. **SASA:**
   Command – gmxsasa –s md_0_1.tpr –f md_0_1.trr –o fullmut1_sasa.xvg
3. **Gyrate:**
   Command – gmx gyrate –s md_0_1.tpr –f md_0_1.trr –o fullmut1_gyrate.xvg
4. **RMSF:**
   Command – gmxrmsf –s md_0_1.tpr –f md_0_1.trr –o fullmut1_rmsf.xvg

3.9.2 GNUPLOT COMMANDS FOR VISUALIZING THE PLOTS –

PLOTS (GNUPLOT)

Steps for plotting a graph in gnu plot are –

1. gnuplot
2. gnuplot> set datafile commentschars “#@%$”
3. gnuplot> set terminal postscript color
4. gnuplot> set output :1ko9_full_20ns_rmsd.ps”
5. gnuplot> plot “1ko9_full_20ns_rmsd.xvg” u 1:2 w lines

   $ps2pdf 1ko9_full_20ns_rmsd.ps
6. quit

Steps for plotting more than one xvg file in a single graph are –

1. gnuplot> plot “1ko9_full_20ns_rmsd.xvg” u 1:2 w lines
2. gnuplot> replot “1ko9_full_20ns_sasa.xvg” u 1:2 w lines
3. gnuplot> replot “1ko9_full_20ns_gyrate.xvg” u 1:2 w lines
4. gnuplot> replot “1ko9_full_20ns_rmsf.xvg” u 1:2 w lines
5. quit

CHAPTER 4

RESULTS

4.1 RESULTS OBTAINED FROM DAMAGE PREDICTION TOOLS

OGG1

SNP	AA CHANGE	POLYPHEN	PROVEAN	SIFT	SNPs&GO	SNAP	PredictSNP	PANTHER	MutPred	MAPP	Phd-SNP	FINAL OUTCOME(>=7)
rs11548133	P27T	Possibly Damaging	DAMAGING	DAMAGING	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs113572618	P29R	Possibly Damaging	DAMAGING	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	Benign	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs104893751	R46Q	Damaging	DAMAGING	DAMAGING	NEUTRAL	EFFECT	DAMAGING	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	DAMAGING
rs17050550	A85S	Benign	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	Benign	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs56053615	R154H	Damaging	DAMAGING	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs1805373	R229Q	Benign	DAMAGING	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	DAMAGING	NON DAMAGING
rs55667729	E230Q	Benign	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	Benign	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs3219012	A288V	Benign	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs113561019	G308E	Damaging	DAMAGING	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs115609368	R324S	Damaging	DAMAGING	DAMAGING	DISEASE	EFFECT	NEUTRAL	Benign	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING

Table 4.1: Damage prediction for the entire SNPs of OGG1 gene

BRCA1

SNP	AA CHANGE	POLYPHEN	PROVEAN	SIFT	SNPs&GO	SNAP	PredictSNP	PANTHER	MutPred	MAPP	Phd-SNP	FINAL OUTCOME(>=7)
rs80357107	V1838E	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs80356959	W1837G	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs80356959	W1837R	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs80356942	E1836K	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	NEUTRAL	DAMAGING	DAMAGING	NEUTRAL	DAMAGING
rs80357268	V1833M	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	NEUTRAL	DAMAGING
rs80357393	A1830T	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NON DAMAGING
rs80357332	Q1826H	Benign	Neutral	DAMAGING	DISEASE	NEUTRAL	DAMAGING	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs80357212	A1823T	Possibly Damaging	Neutral	TOLERATED	DISEASE	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs80357286	N1819S	Benign	Neutral	TOLERATED	DISEASE	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs80357477	D1818G	Benign	Neutral	DAMAGING	DISEASE	EFEECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs1800751	P1812A	Benign	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs80357040	Q1811R	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING
rs80357451	V1810G	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	DAMAGING	DAMAGING
rs80357216	V1809A	Benign	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs28897698	V1809F	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs80357358	V1808A	Benign	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs80357241	P1806A	Benign	Neutral	TOLERATED	DISEASE	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs80356920	V1804D	Benign	Neutral	DAMAGING	DISEASE	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs80357149	G1803A	Benign	Neutral	DAMAGING	DISEASE	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs80357078	A1789S	Possibly Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING
rs80357069	G1788D	Damaging	Neutral	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING

Table 4.2: Damage prediction for the entire SNPs of BRCA1 gene

PARP1

SNP	AA CHANGE	POLYPHEN	PROVEAN	SIFT	SNPs&GO	SNAP	PredictSNP	PANTHER	MutPred	MAPP	Phd-SNP	FINAL OUTCOME(DAMAGE>=7)
rs78381515	P956S	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs3219145	K940R	Possibly Damaging	Deleterious	DAMAGING	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	DAMAGING	NEUTRAL	NEUTRAL	NON DAMAGING
rs61731502	G917A	Damaging	Deleterious	DAMAGING	NEUTRAL	EFFECT	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING	DAMAGING
rs1059040	C908Y	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs11541664	P882L	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs1136420	N827S	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NON DAMAGING
rs1136410	V762A	Possibly Damaging	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	DAMAGING	NEUTRAL	NON DAMAGING
rs1059011	H613Q	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs79529505	V408G	Benign	Deleterious	DAMAGING	NEUTRAL	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs3219062	S383Y	Possibly Damaging	Neutral	DAMAGING	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2230484	P377S	Benign	Neutral	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs1805415	K352N	Benign	Neutral	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs3219057	V334I	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs61750985	K262R	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs111635488	I248T	Benign	Deleterious	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs1805409	A188T	Benign	Neutral	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs113258217	L181F	Benign	Deleterious	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs3738708	F54L	Benign	Deleterious	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NON DAMAGING

Table 4.3: Damage prediction for the entire SNPs of PARP1 gene

XRCC1

SNP	AA CHANG	POLYPHEN	PROVEAN	SIFT	SNPs&GO	SNAP	PredictSNP	PANTHER	MutPred	MAPP	Phd-SNP	FINAL OUTCOME (DAMAGE>=7)
rs25487	Q399R	Damaging	Neutral	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2307177	Y576S	Benign	Deleterious	DAMAGING	NEUTRAL	EFFECT	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NON DAMAGING
rs2682557	Y576N	Benign	Deleterious	TOLERATED	DISEASE	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING	NON DAMAGING
rs2307166	R560W	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING	DAMAGING
rs2307167	R559Q	Damaging	Neutral	TOLERATED	NEUTRAL	EFFECT	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NON DAMAGING
rs41561817	H528Y	Benign	Neutral	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs25474	P514L	Damaging	Deleterious	DAMAGING	NEUTRAL	EFFECT	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs72554204	E491K	Damaging	Neutral	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2307184	S485Y	Possibly Damaging	Deleterious	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	NON DAMAGING
rs2271980	V381M	Damaging	Deleterious	DAMAGING	DISEASE	EFFECT	DAMAGING	DAMAGING	DAMAGING	DAMAGING	NEUTRAL	DAMAGING
rs25491	P309S	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs25490	T304A	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2307188	K298N	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs25489	R280H	Benign	Neutral	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs111857343	A214V	Benign	Neutral	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs1799782	R194W	Damaging	Deleterious	DAMAGING	NEUTRAL	EFFECT	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NON DAMAGING
rs56357789	N183S	Benign	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2307191	P161L	Benign	Deleterious	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	DAMAGING	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2307180	E157K	Possibly Damaging	Neutral	TOLERATED	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NEUTRAL	NON DAMAGING
rs2228487	R107H	Damaging	Deleterious	DAMAGING	NEUTRAL	EFFECT	DAMAGING	DAMAGING	NEUTRAL	DAMAGING	NEUTRAL	DAMAGING
rs25496	V72A	Damaging	Deleterious	TOLERATED	NEUTRAL	EFFECT	NEUTRAL	DAMAGING	DAMAGING	DAMAGING	NEUTRAL	NON DAMAGING

Table 4.4: Damage prediction for the entire SNPs of XRCC1 gene

After running these 10 tools, conclusion is that there are two DNA Repair Genes that consist of the most damaging SNPs. Looking at the majority in the result, those SNPs are chosen that are predicted damaging by at least 7 out of 10 Damage prediction tools. Those two damaging genes are –

- OGG1
- BRCA1

OGG1 gene has 3 Damaging SNPs out of 10 SNPs and BRCA1 gene has 22 Damaging SNPs out of 38 SNPs according to the prediction tools. Since complete structure of BRCA1 is not available in PDB, therefore OGG1 gene is chosen for further investigation at the structural level. There is one more reason not to choose BRCA1 gene at the structural level for AD. The reason is that BRCA1 is majorly involved in Breast Cancer and it just has a borderline association with Alzheimer's.

Final Conclusion: OGG1 selected as the most damaging gene

4.2 RESULTS OBTAINED FROM FUNCTIONAL SITE PREDICTION TOOLS

OGG1 UET RESULTS-

Residue	AA	type	coverage	variability	Score (rvET)
27	P	0.82319	19	NSPRKHC.TVEQIFLAGDY	79.00
46	**R**	**0.24058**	**16**	**KHSRVTILNGEA.MQC**	**28.21**
154	**R**	**0.11014**	**13**	**R.KQHMSVTNFGA**	**15.63**
308	**G**	**0.15072**	**12**	**GIE.PSATQKRY**	**22.35**
324	R	0.46667	18	KVDRAMS.QPLINGEHT	49.73

UET (Universal Evolutionary Trace) uses evolutionary distances to estimate functional distances and correlates genotype variations with those in the fitness phenotype. Thus, UET ranks are worse for sequence positions that vary among evolutionarily closer homologs but better for positions that vary mostly among distant homologs. This means that the residues with scores of lower values are more conserved evolutionarily as compared to higher scores.

OGG1 CONSURF-DB RESULTS-

POS	SEQ	SCORE	RESIDUE VARIETY
27	P	0.569	N,I,R,S,G,E,D,L,K,T,P,A,H
29	P	0.249	E,D,K,L,I,S,G,Q,A,P**,T,C**
46	**R**	**-0.922**	**N,R,G,S,D,K,L,A**
154	**R**	**-1.184**	**K,F,R**
308	**G**	**-0.903**	**T,A,K,Q,G,S**
324	**R**	**-0.319**	**N,Q,R,S,G,E,K,D,P,T,A,H**

ConSurf-DB provides evolutionary conservation profiles for proteins of known structure in the PDB. Visual inspection of the conservation patterns on the 3-dimensional structure often enables the identification of key residues that comprise the functionally-important regions of the protein. ConSurf-DB provides pre-calculated conservation profiles. The scores with the negative sign signify the conserved residues whereas the scores with the positive values signify non-conserved regions of a protein.

On analysing the results of these tools we found out the scores that justified the damaging and the non-damaging regions. All those SNPs that were predicted deleterious by Damage Prediction Tools are verified and get justified.

4.3 STRUCTURE SELECTION (THESEUS)

On running this software, 23rd structure is obtained as the median structure. This means that out of all the 23 structures given as input, 23rd structure is the closest to the average structure of all the 23 structures. Theseus calculates the root mean square distance for every structure from the average structure, which is obtained from the superimposition of all the structures. Here in our study, the 23rd structure is 1KO9. 1KO9 is extracted from PDB which is a complete structure of OGG1 gene. The result of Theseus is shown ahead as well as the structure of 1KO9 is displayed.

```
←[1;31m< BEGIN THESEUS 3.3.0 >←[0m |===-===-===-===-===-===-===-===-===-==
superpositioning        ||=-===-===-===-===-===-===-===-===-===-===-===-===-:
Successfully read 23 models and/or structures    Selecting coordinates for superposition
 23 models superimposed in 30.0 ms
 * Classical LS pairwise <RMSD>                   1.58642
* Least-squares <sigma>                           0.62832
* Maximum Likelihood <sigma>                      0.28391
~ Marginal Log Likelihood                        -6610.72
~ AIC                                            -7442.20
~ BIC                                            -10231.31
+ Omnibus chi^2                                   0.56 (P:0.00e+00)
+ Hierarchical var (1.21e-01, 1.50e+00) chi^2     3.88524 (P:0.00e+00)
+ Rotational, translational, covar chi^2          2.71 (P:0.00e+00)
+ Hierarchical minimum var (sigma)                4.32e-03 (6.57e-02)
< skewness                                        0.06 (P:1.39e-02)
< skewness Z-value                                2.46
< kurtosis                                        0.34 (P:1.58e-13)
< kurtosis Z-value                                7.38
* Data pts = 15424,
Free params = 950,
D/P = 2.0
* Median structure = #23
* N(total) = 7176,
N(atoms) = 312,
N(structures) = 23 |
Total rounds = 201
```

Figure 4.1: THESEUS result for Structure Selection

1KO9 STRUCTURE

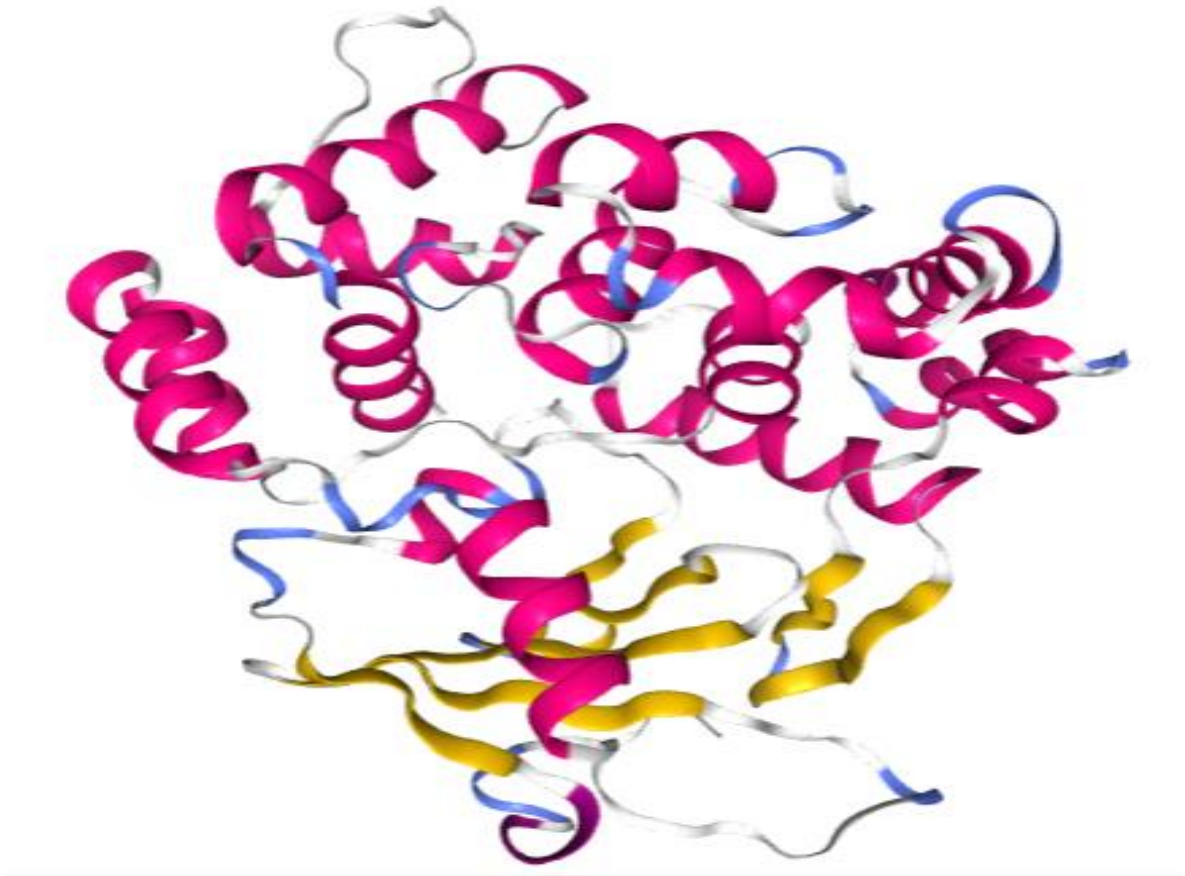


Figure 4.2: OGG1 gene complete structure from PDB (1KO9: PDB ID)

4.4 STRUCTURAL ANALYSIS

4.4.1 CONSURF OUTPUTS

R46Q

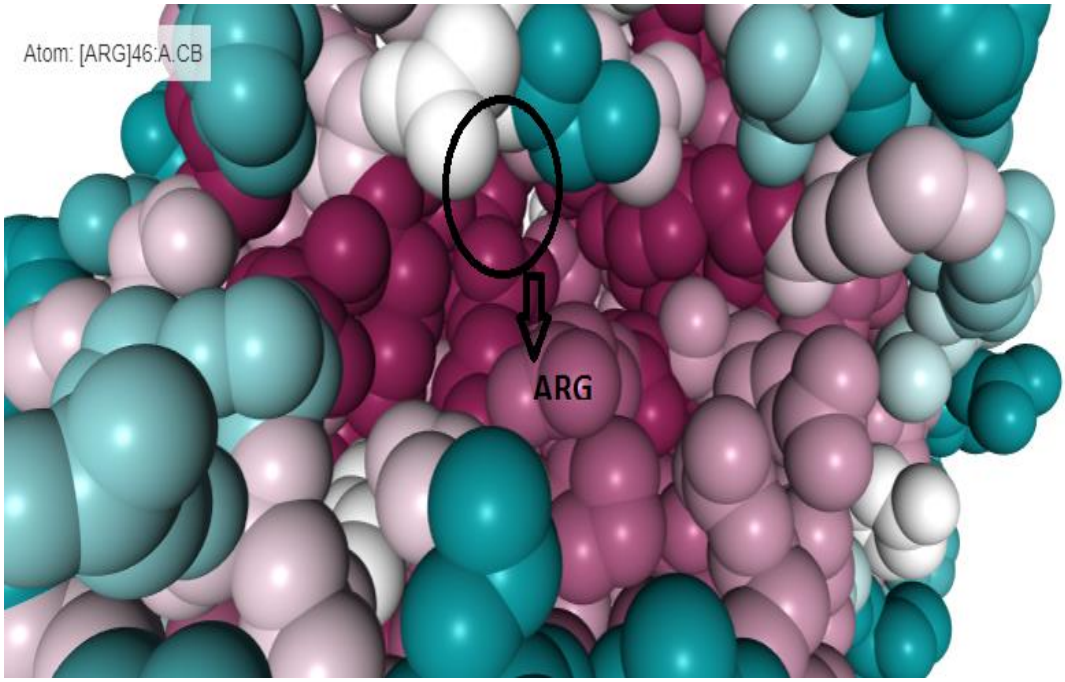


Figure 4.3: Consurf output showing Arginine (R) at position 46

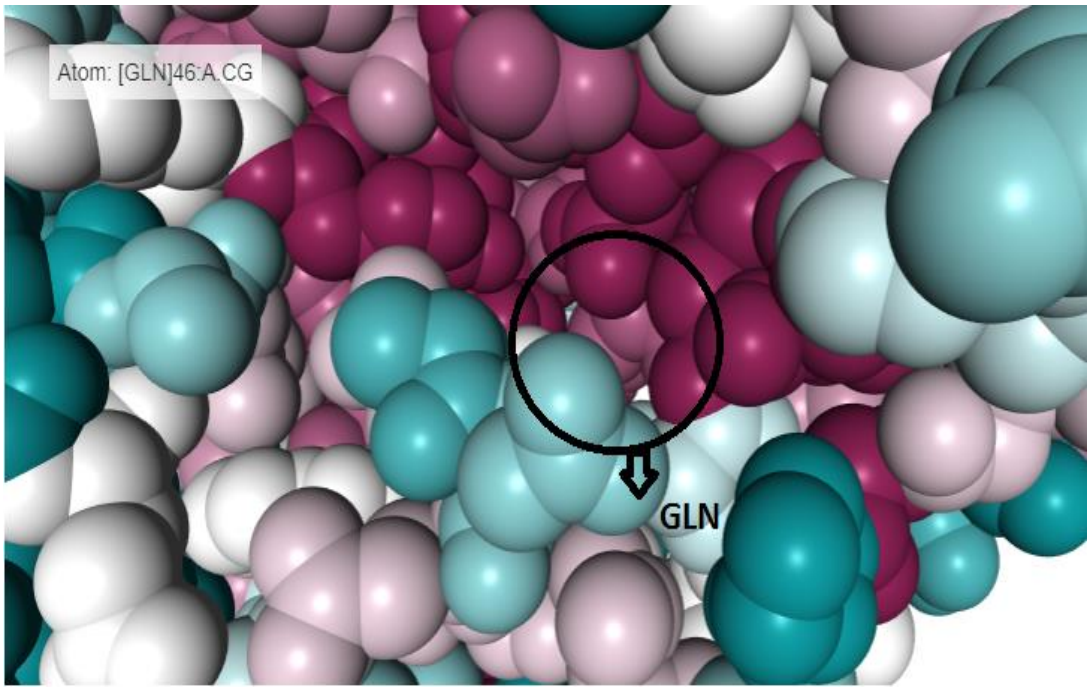


Figure 4.4: Consurf output showing Glutamine (Q) at position 46

R154H

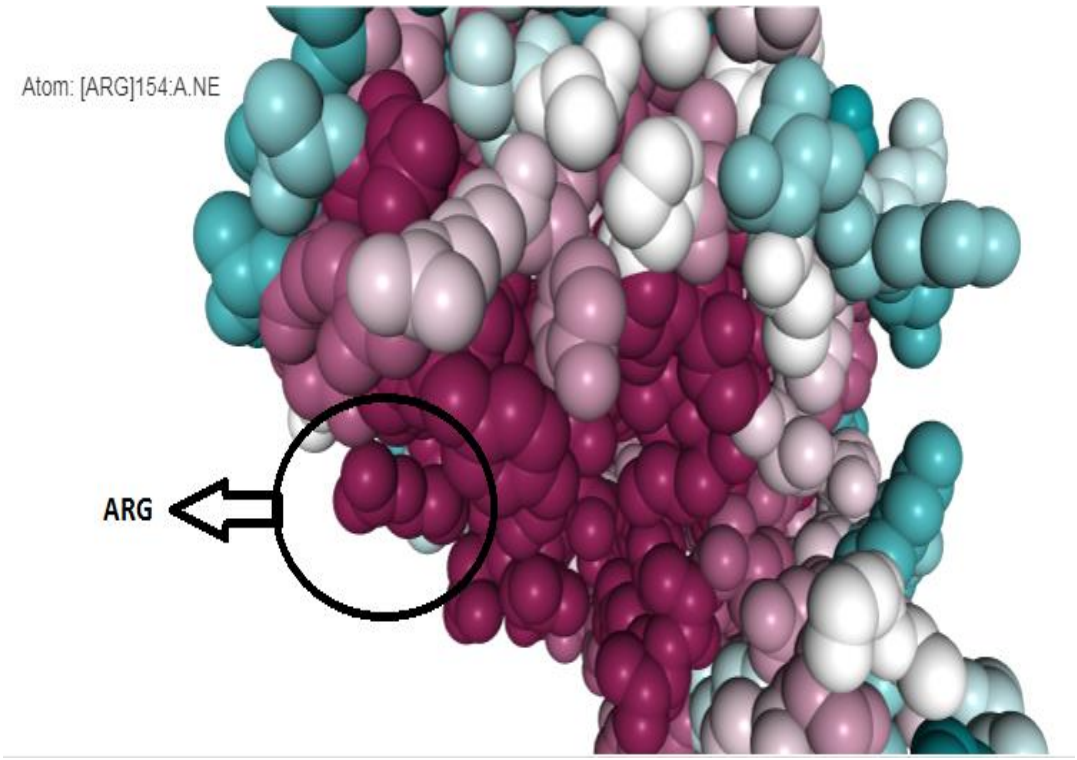


Figure 4.5: Consurf output showing Arginine (R) at position 154

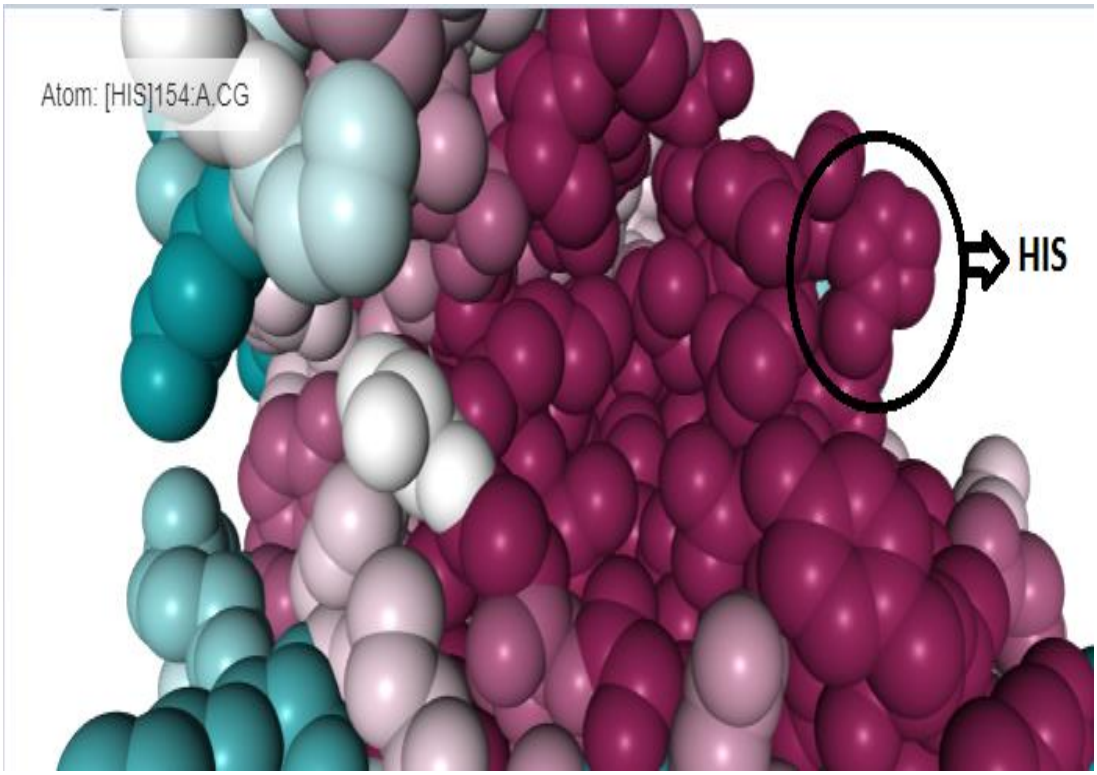


Figure 4.6: Consurf output showing Histidine (H) at position 154

G308E

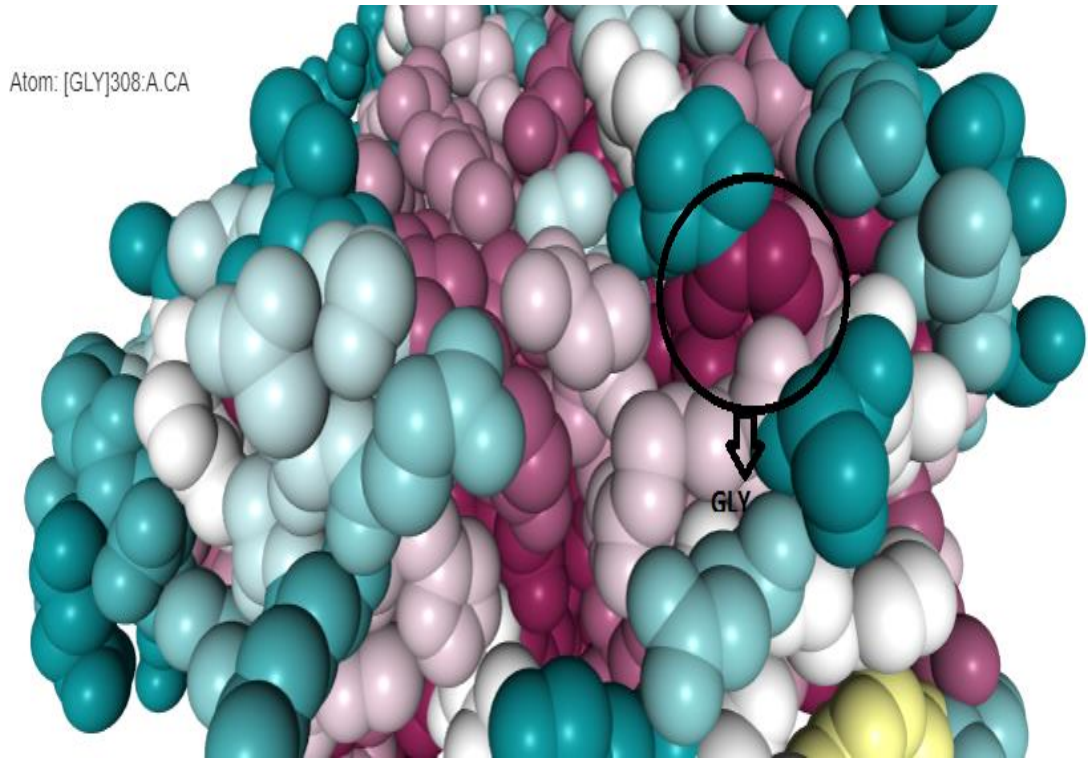


Figure 4.7: Consurf output showing Glycine (G) at position 308

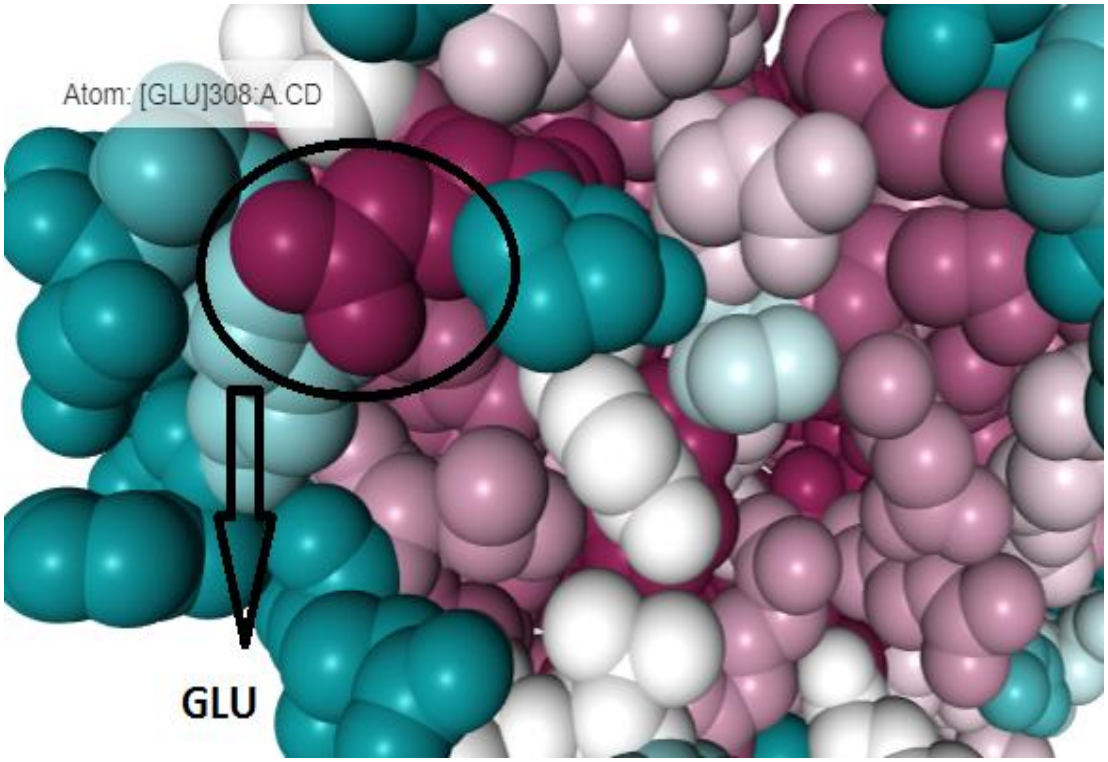


Figure 4.8: Consurf output showing Glutamic Acid (E) at position 308

4.4.2 STRIDE OUTPUT

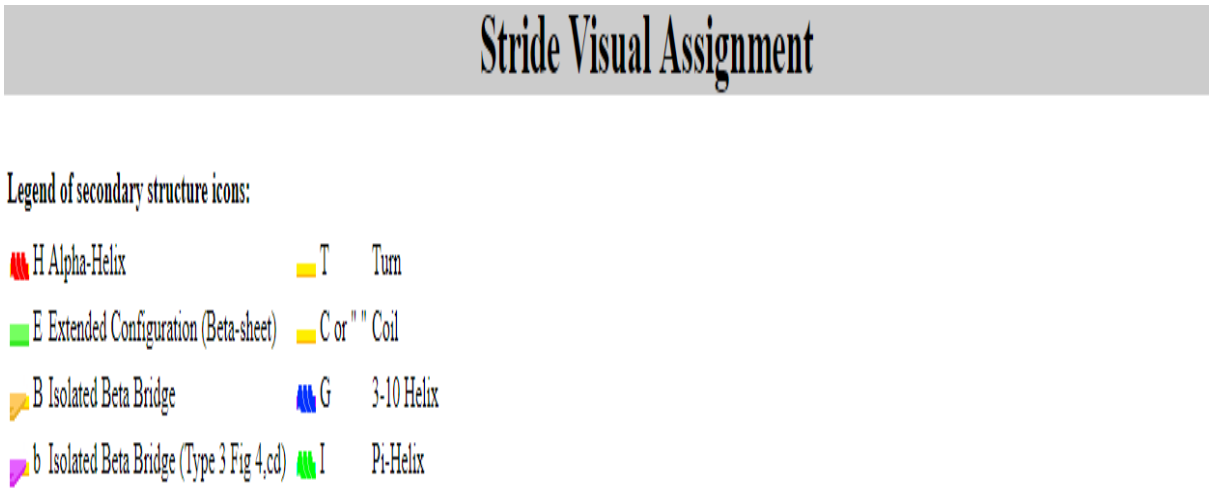


Figure 4.9: STRIDE visualization colour scheme to represent secondary structure

This image represents the colour scheme for the representation of secondary structure in a protein molecule. STRIDE has two options – one is the visual assignment and the other one is the Ramachandran Plot. Here, with the help of visual assignment, it is shown whether the mutation affects the secondary structure of the molecule or not.

In case of Ramachandran Plot, it is shown whether the torsion angles change after mutation of a certain amino acid and after that mutation, the amino acid lies in which region of the plot (allowed or the disallowed regions).

MUTATION R46Q –

R (Arginine) at position 46

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
ARG	A	46	35	C	Coil	-96.14	9.07	26.9

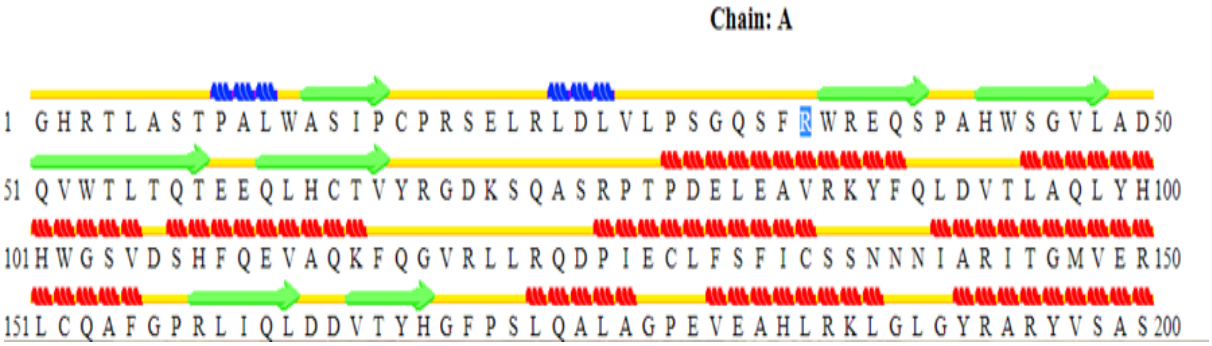


Figure 4.10:STRIDE visual assignment output for Arginine at position 46

Q (Glutamine) at position 46

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
GLN	A	46	35	C	Coil	100.28	9.31	25.0

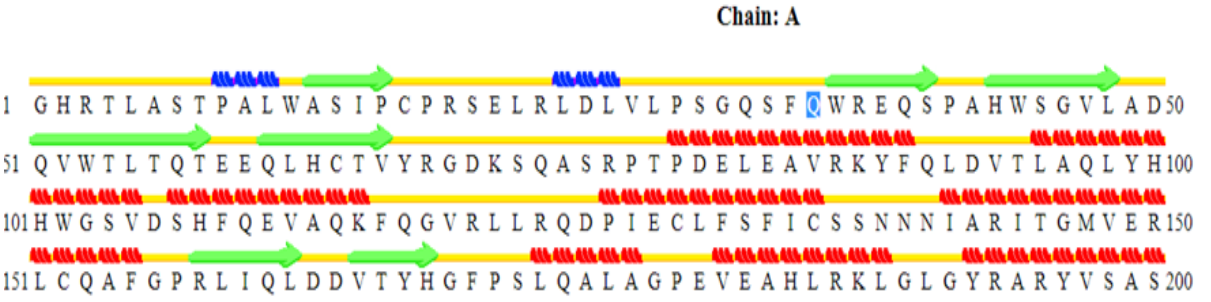


Figure 4.11:STRIDE visual assignment output for Glutamine at position 46

RAMACHANDRAN PLOTS FOR MUTATION R46Q -

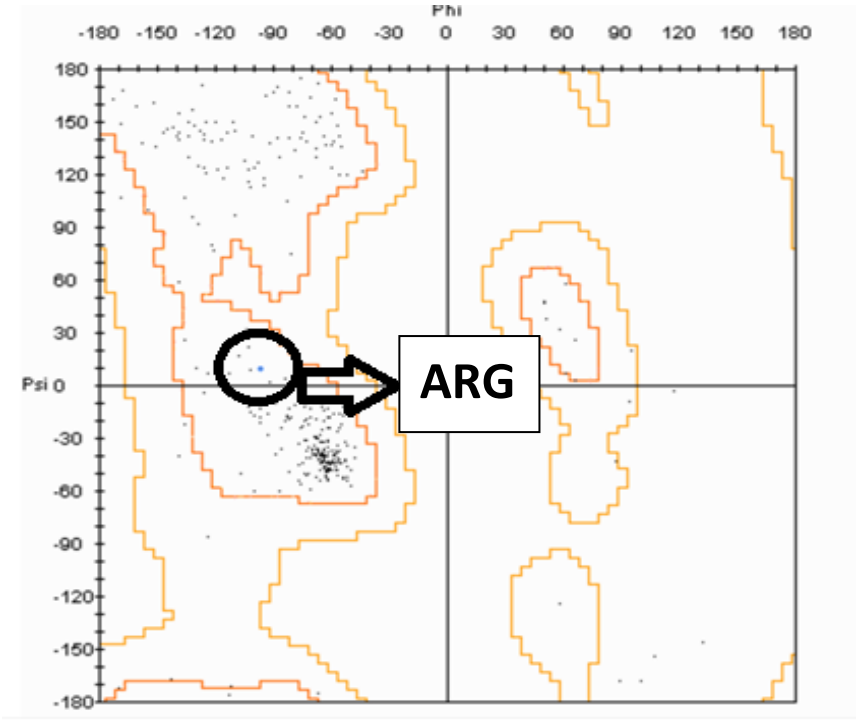


Figure 4.12: Ramachandran Plot showing the region where Arginine at position 46 lies

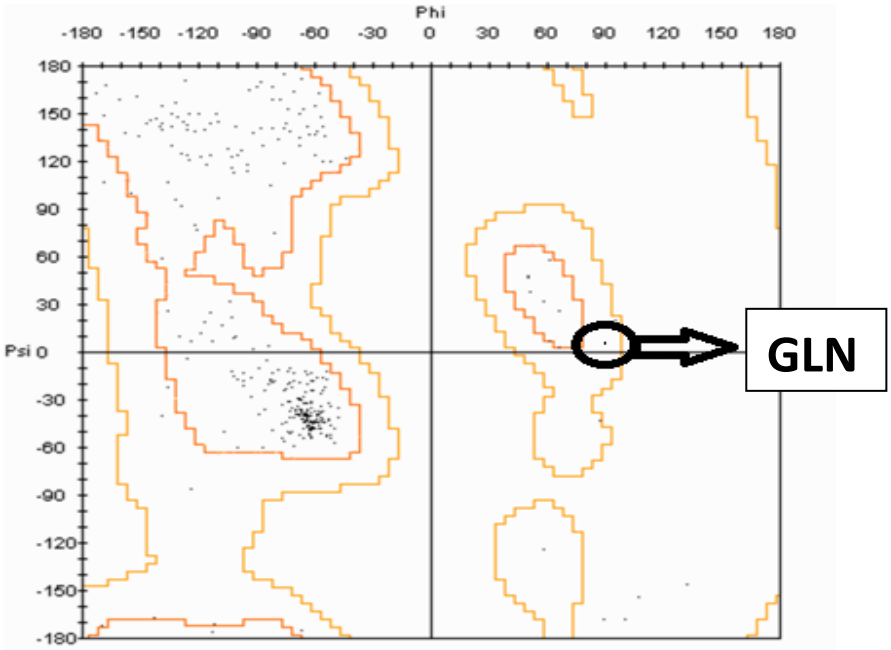


Figure 4.13: Ramachandran Plot showing the region where Glutamine at position 46 lies

MUTATION R154H -

R (Arginine) at position 154

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
ARG	A	154	143	H	AlphaHelix	-70.55	-59.15	27.6

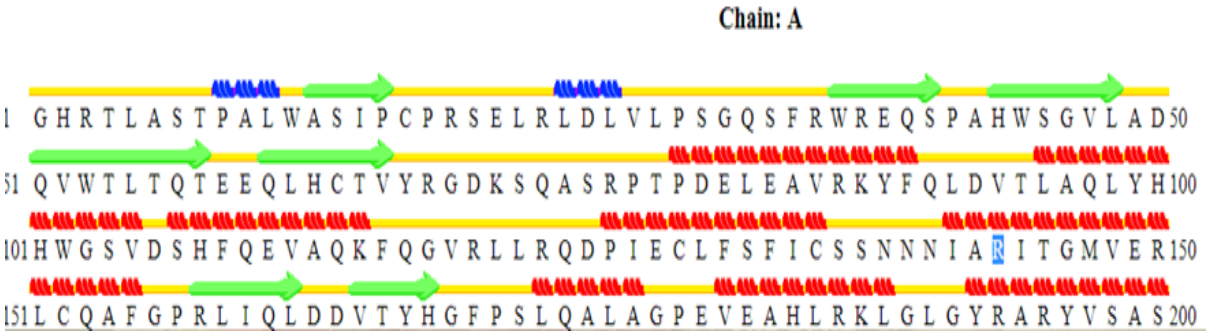


Figure 4.14:STRIDE visual assignment output for Arginine at position 154

H (Histidine) at position 154

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
HIS	A	154	143	H	AlphaHelix	-72.69	-58.03	07.6

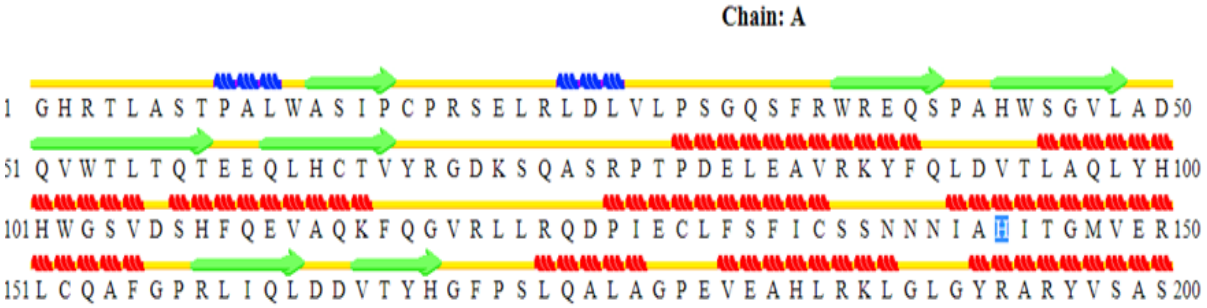


Figure 4.15: STRIDE visual assignment output for Histidine at position 154

RAMACHANDRAN PLOTS FOR MUTATION R154H -

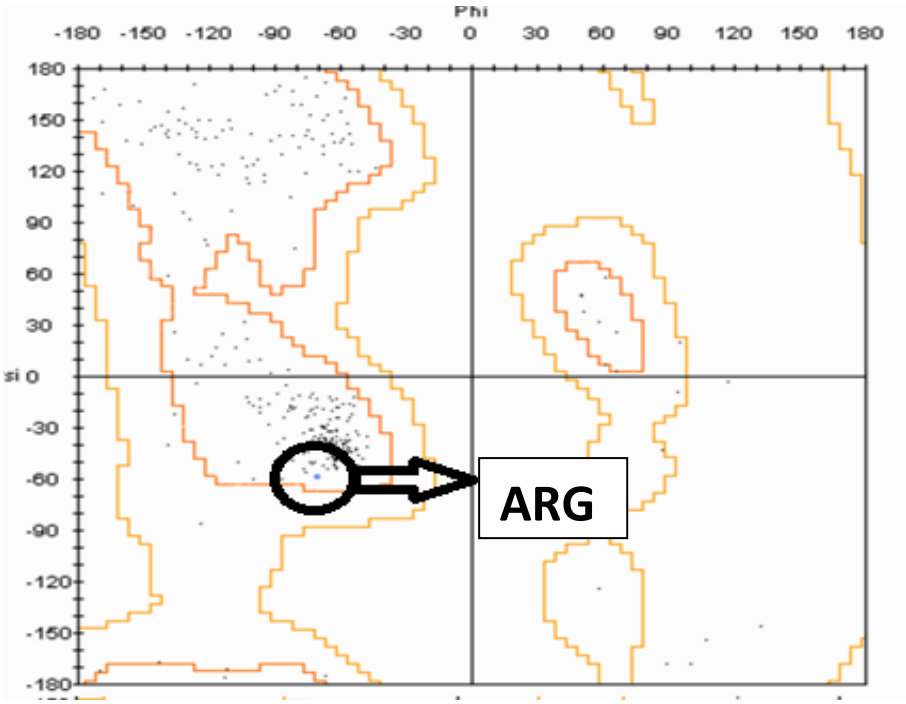


Figure 4.16: Ramachandran Plot showing the region where Arginine at position 154 lies

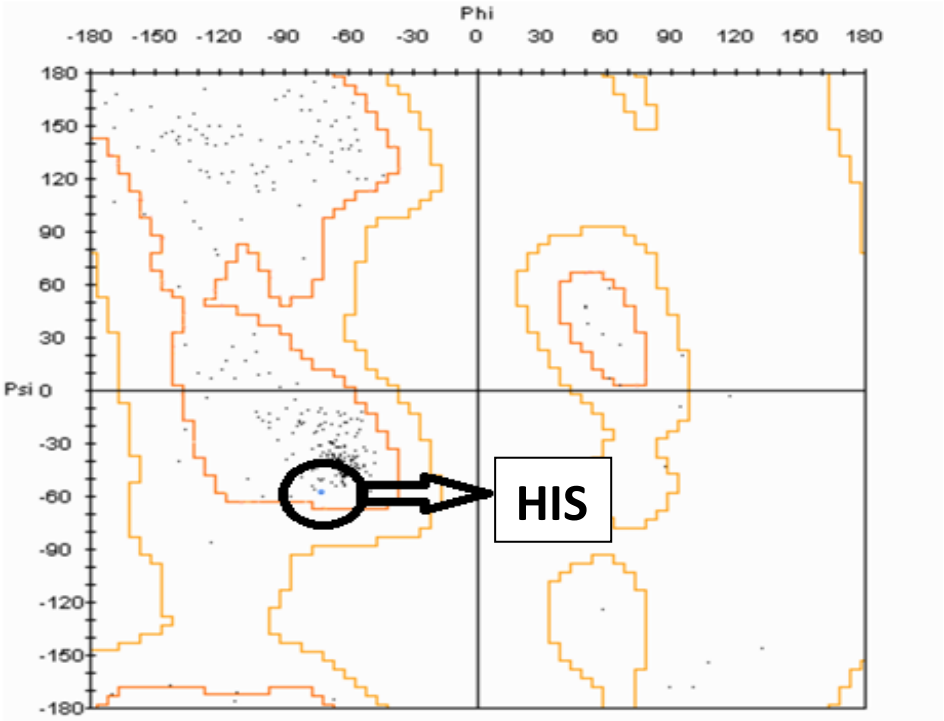


Figure 4.17: Ramachandran Plot showing the region where Histidine at position 154 lies

MUTATION G308E -

G (Glycine) at position 308

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
GLY	A	308	297	T	Turn	100.55	168.97	17.6

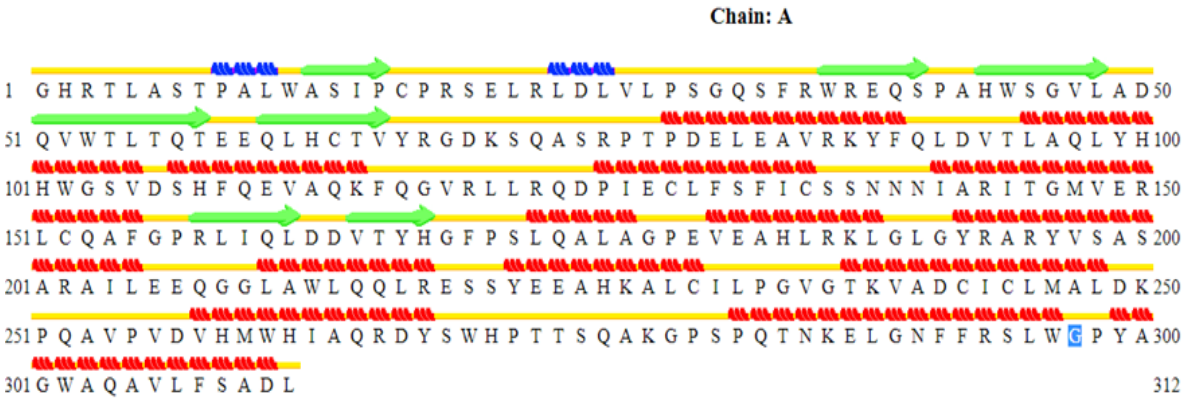


Figure 4.18: STRIDE visual assignment output for Glycine at position 308

E (Glutamic acid) at position 308

Residue Information: (point on the structure of a residue to get the detailed assignment information)

Residue name	Chain	Pdb residue #	Ordinal residue #	One letter code	Structure	Phi angle	Psi angle	Solvent accessible area
GLU	A	308	297	T	Turn	103.31	169.26	87.7

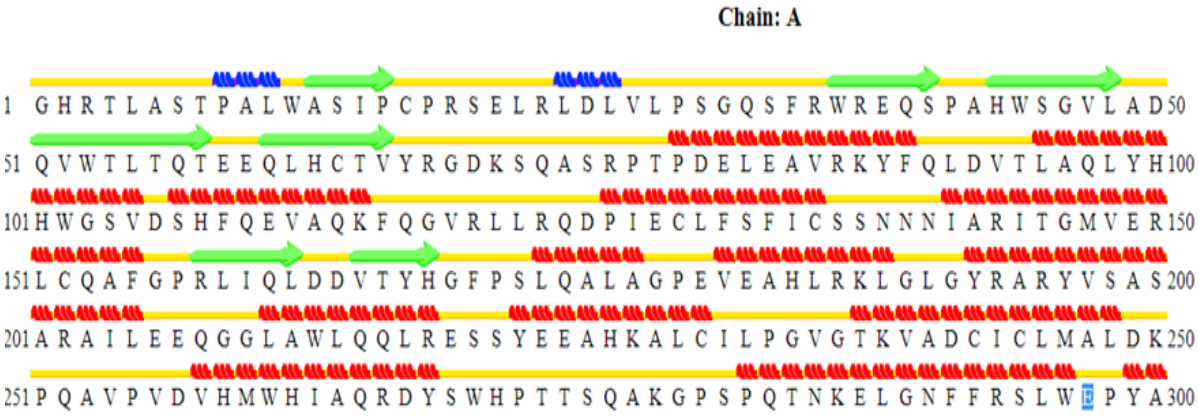


Figure 4.19: STRIDE visual assignment output for Glutamic acid at position 308

RAMACHANDRAN PLOTS FOR MUTATION G308E -

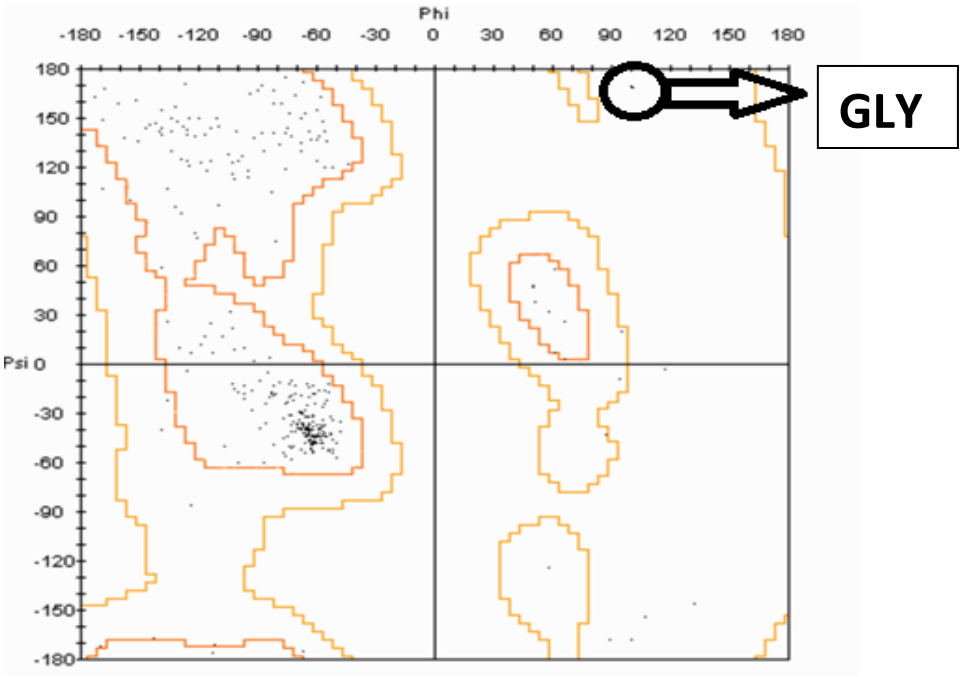


Figure 4.20: Ramachandran Plot showing the region where Glycine at position 308 lies

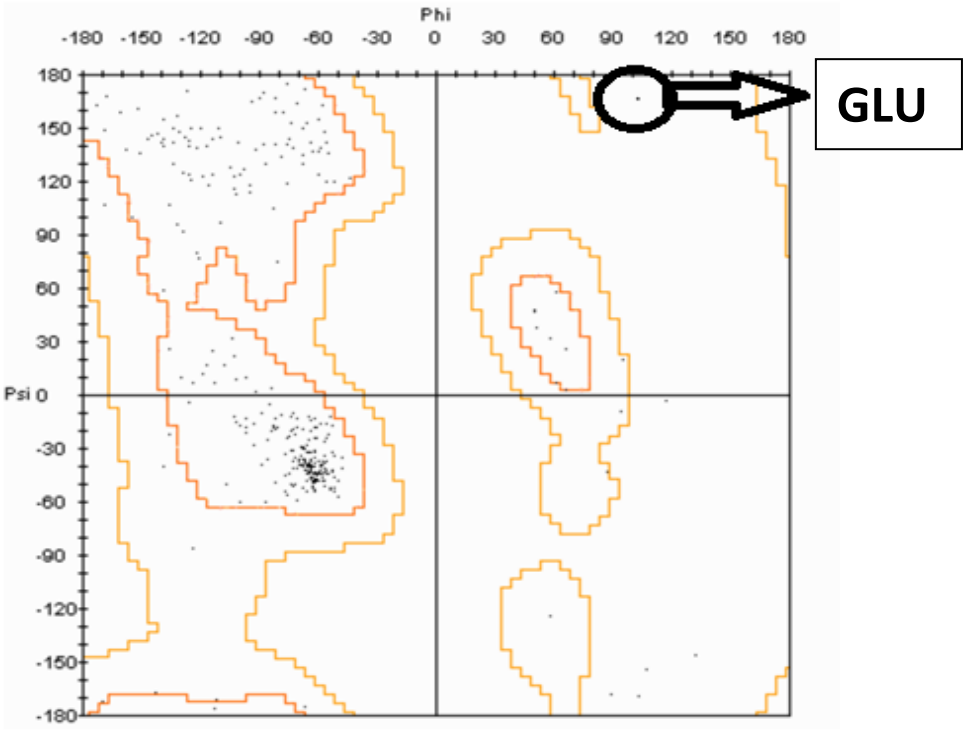


Figure 4.21: Ramachandran Plot showing the region where Glutamic Acid at position 308 lies

4.5 MD SIMULATIONS ANALYSIS

BOX VISUALIZATION IN PyMOL -

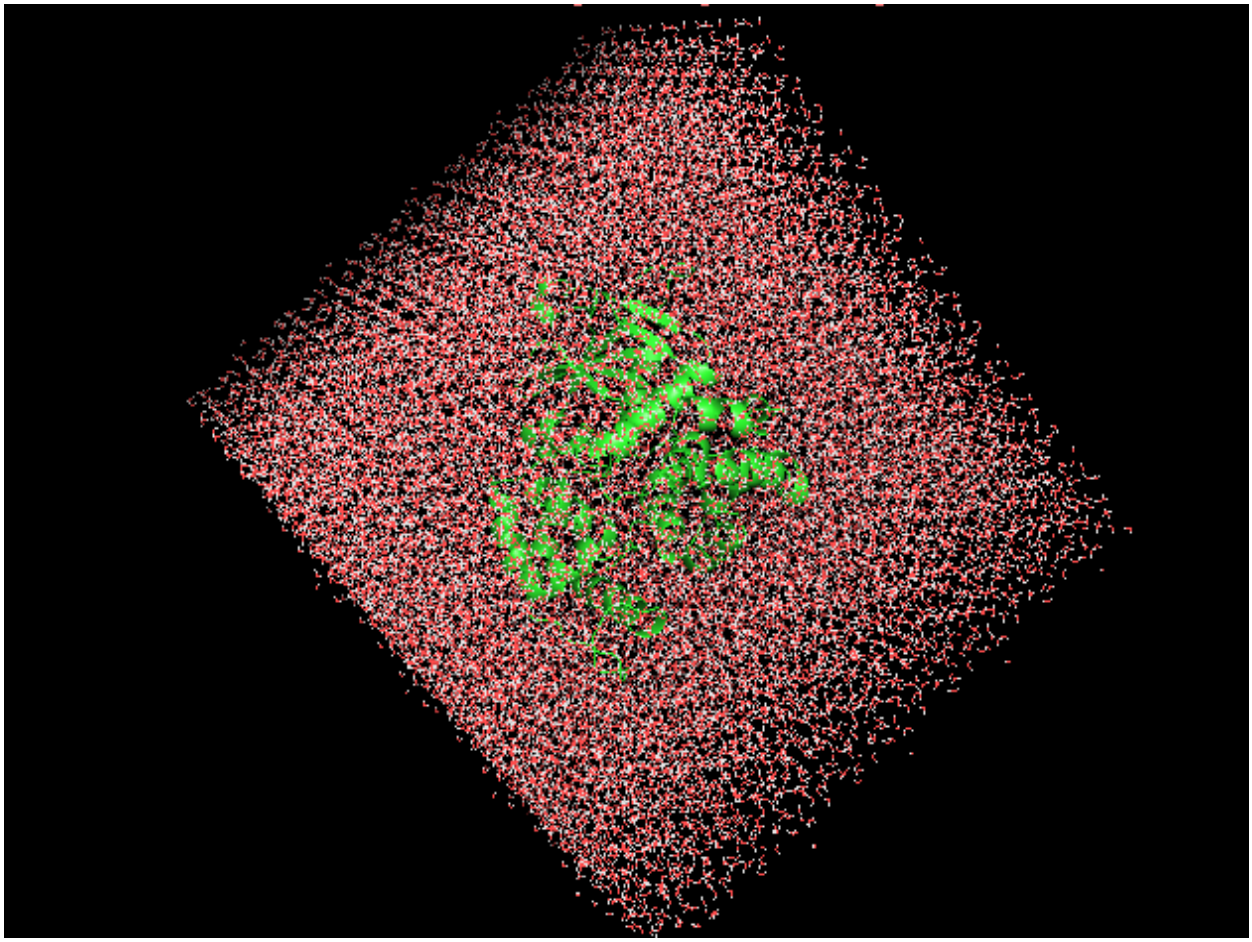


Figure 4.22: Visualization of 1ko9 structure inside the box using Pymol

While performing the newbox.gro command in GROMACS, we need to look the new structure (which is generated after successfully running the command to define a box around the molecule) in any visualization tool like PyMol or VMD. The protein molecule must be in the centre of the box and neither of its atoms should be coming out of the box. In case, atoms come out, then the box's shape must be redefined and the dimensions of the box must be changed. The molecule must lie within the boundaries of the defined box and should be almost in the centre. Here, in our study, this image shows that our molecule (1ko9), is completely within the bounds of the box and is also almost in the centre. Therefore, we can carry out further GROMACS steps.

PLOTS FOR 1ST MUTANT R46Q -

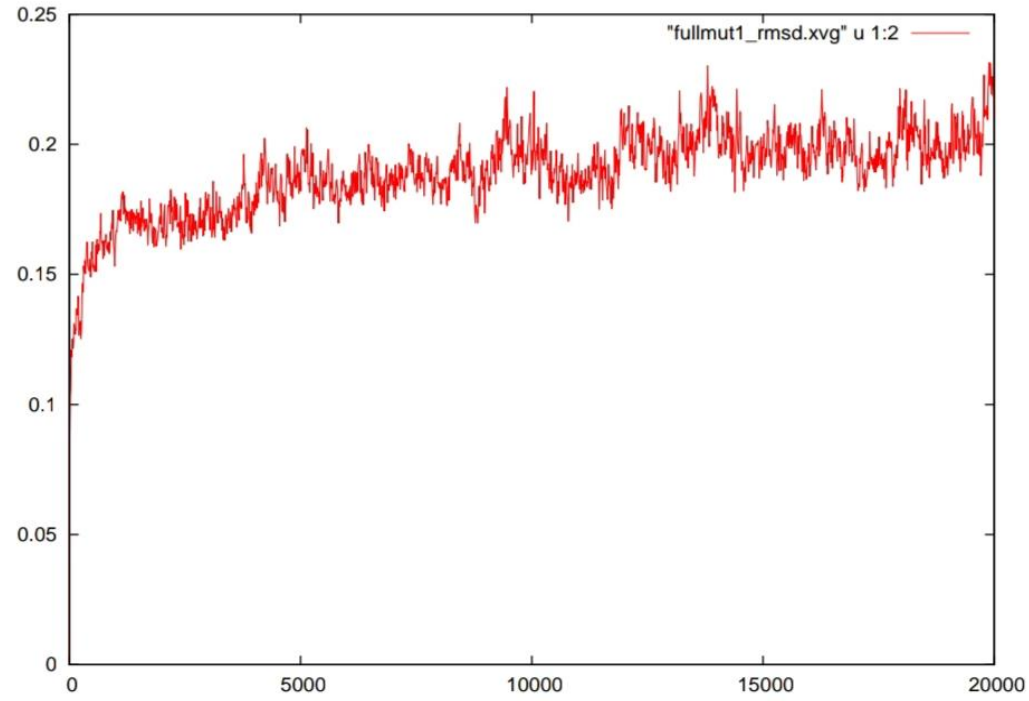


Figure 4.23: Root mean square deviation plot for mutant R46Q

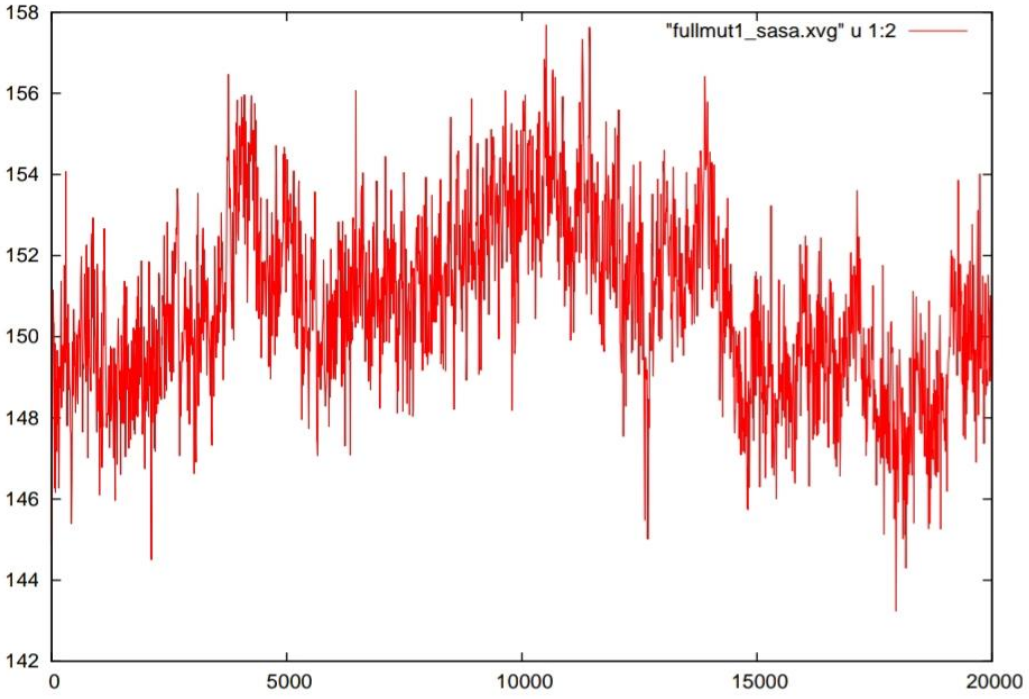


Figure 4.24: Solvent accessible surface area plot for mutant R46Q

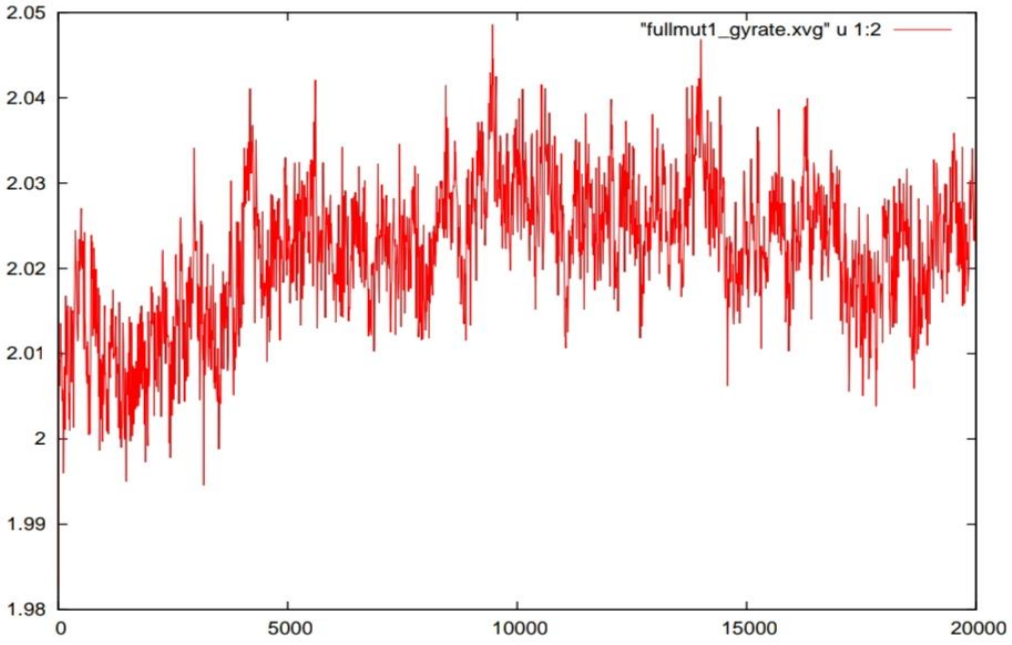


Figure 4.25: Radius of Gyration plot for mutant R46Q

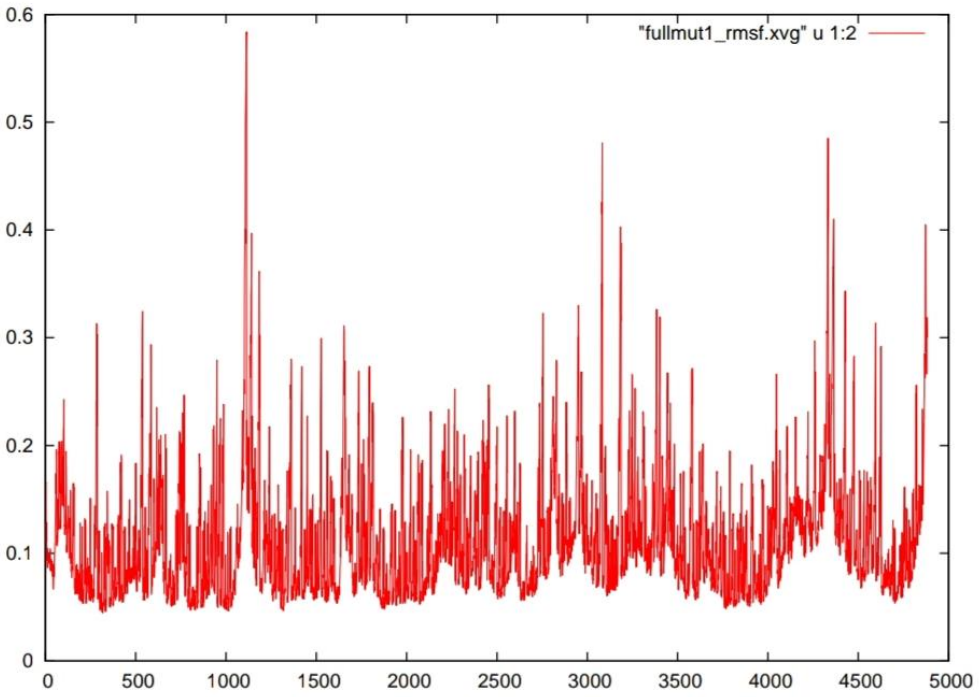


Figure 4.26: Root mean square fluctuation plot for mutant R46Q

PLOT FOR 2nd MUTANT R154H -

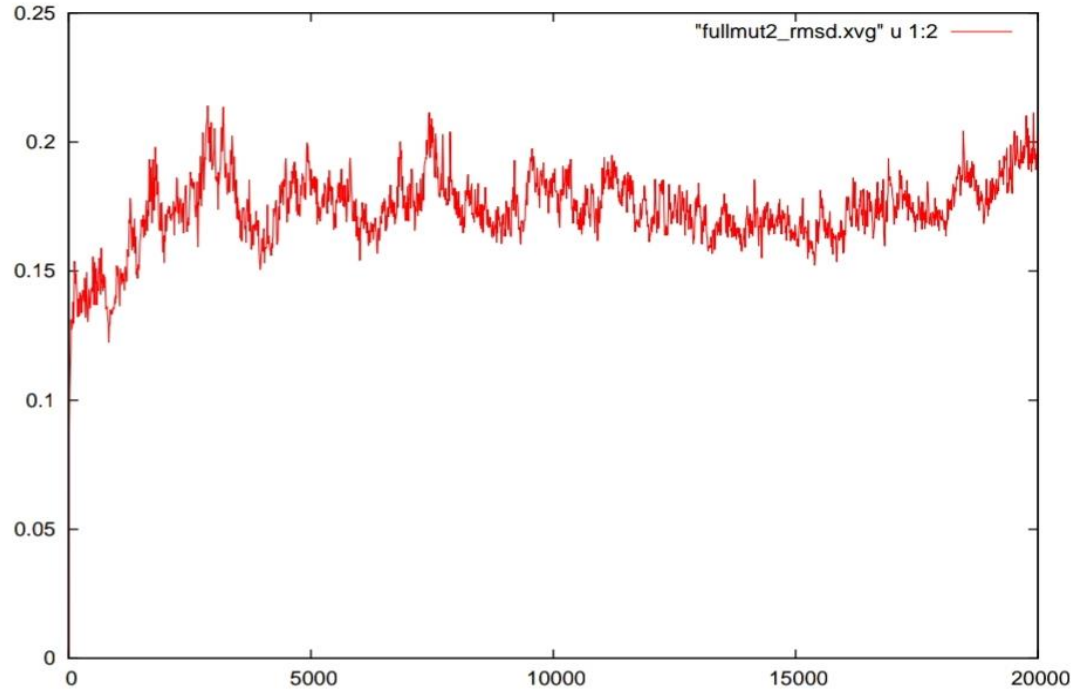


Figure 4.27: Root mean square deviation plot for mutant R154H

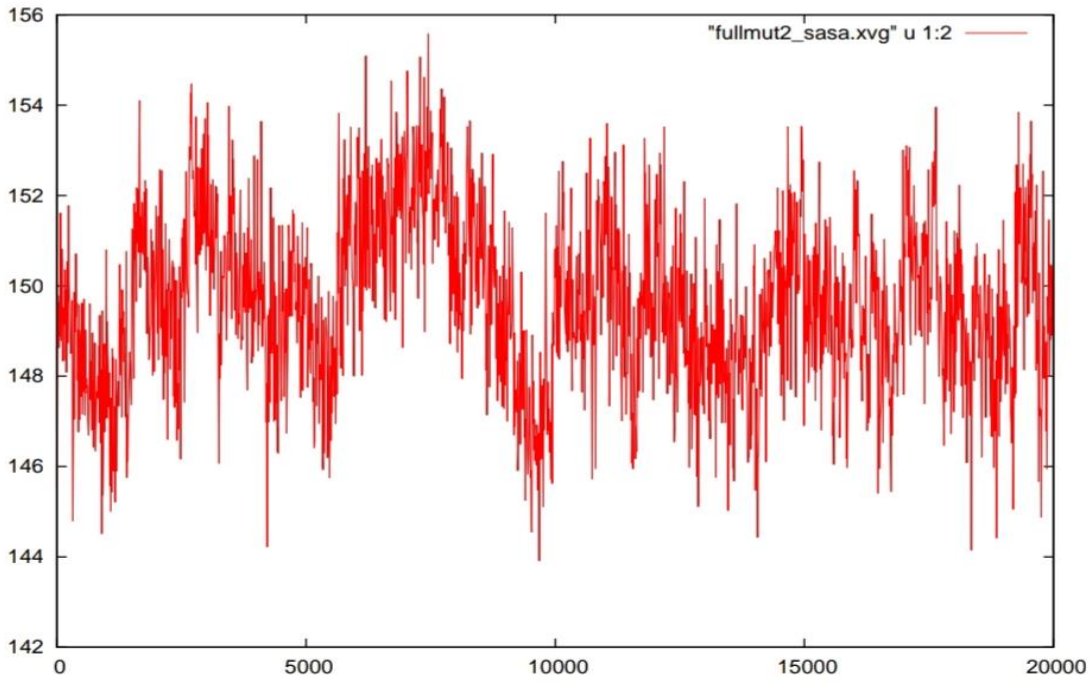


Figure 4.28: Solvent accessible surface area plot for mutant R154H

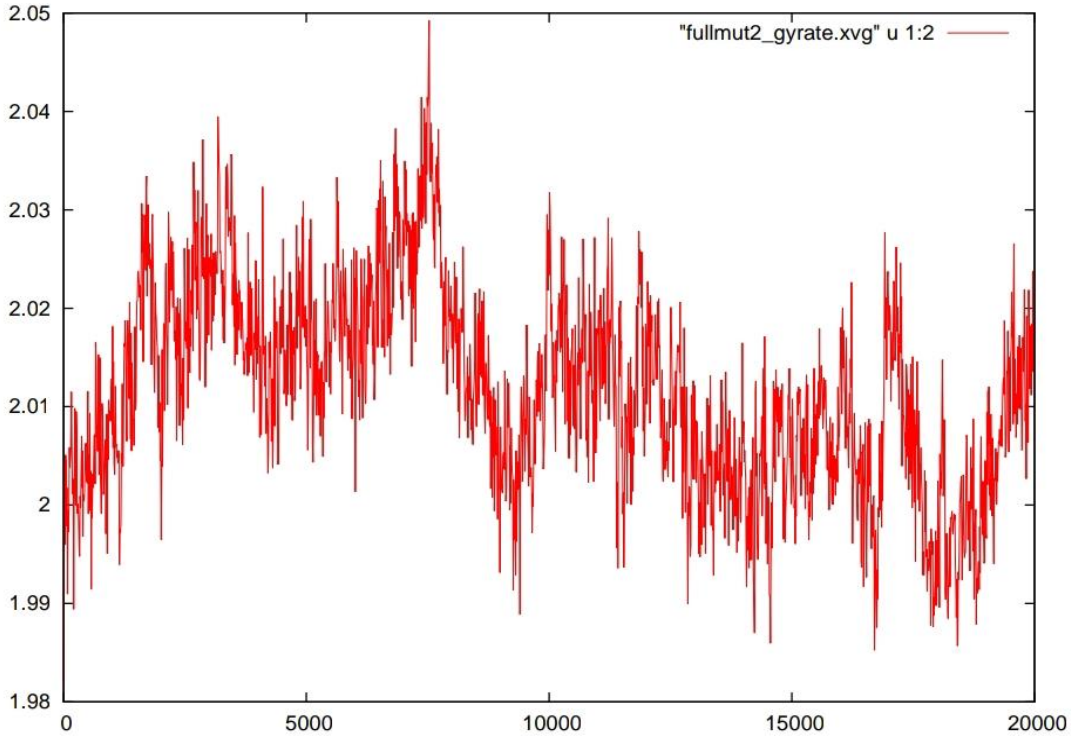


Figure 4.29: Radius of Gyration plot for mutant R154H

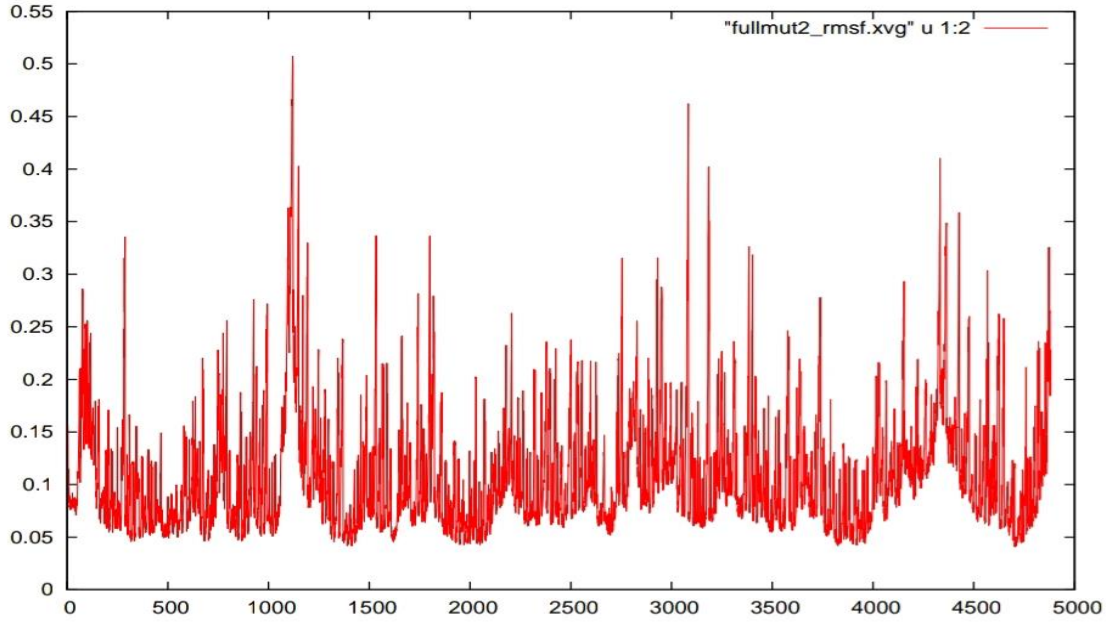


Figure 4.30: Root mean square fluctuation plot for mutant R154H

PLOT FOR 3rd MUTANT G308E –

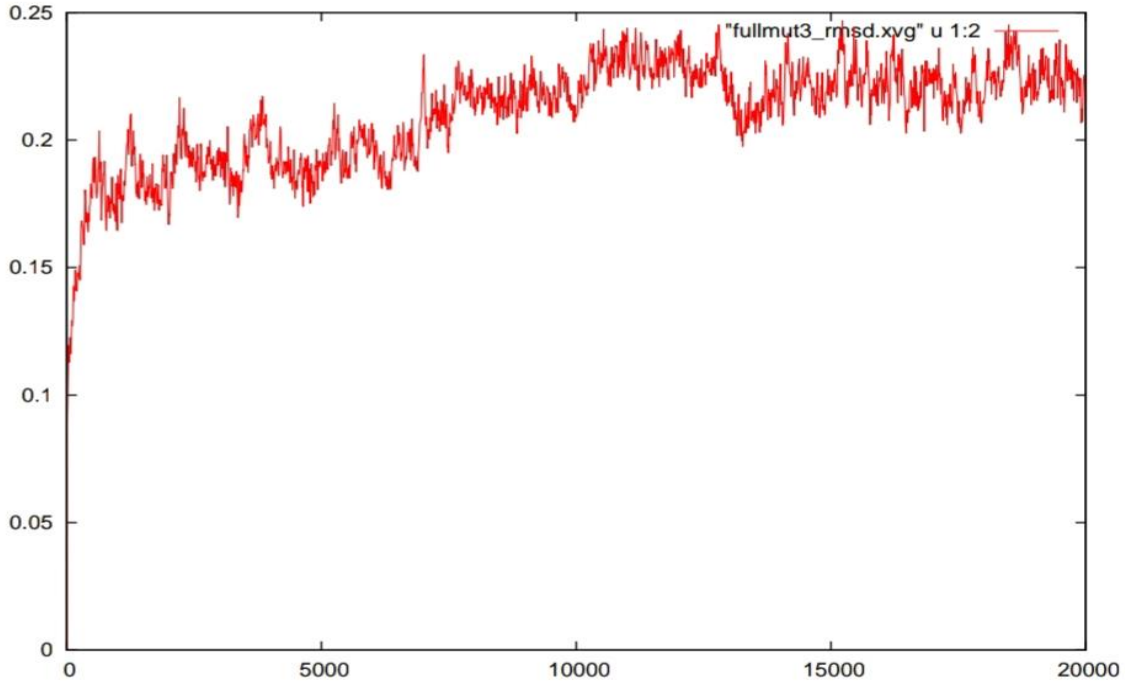


Figure 4.31: Root mean square deviation plot for mutant G308E

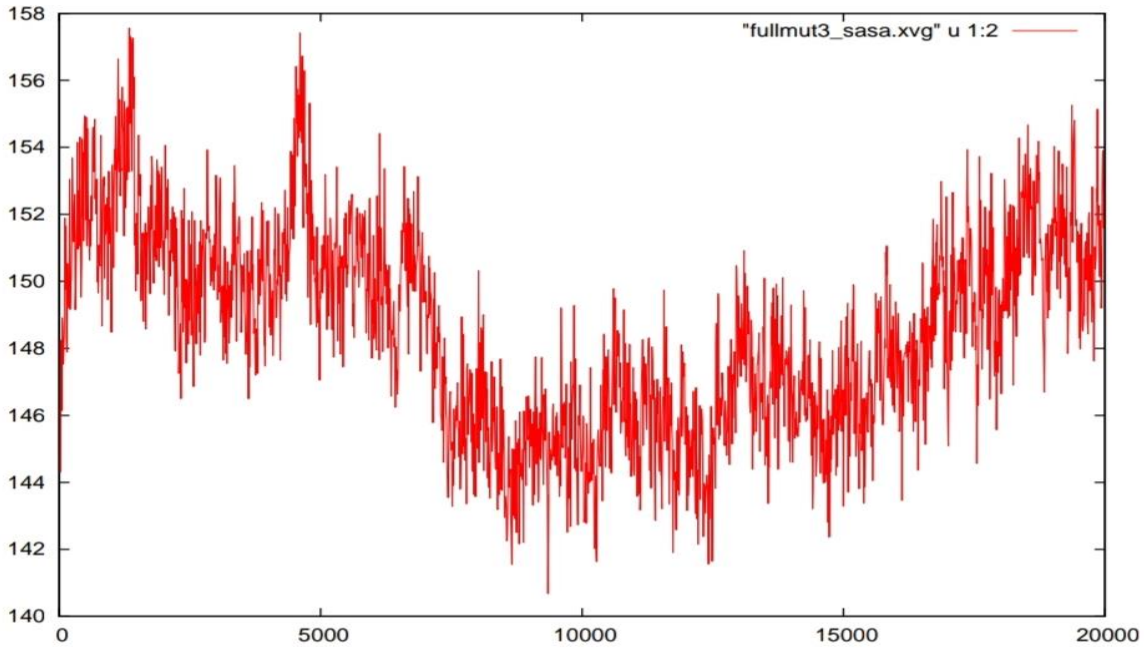


Figure 4.32: Solvent accessible surface area plot for mutant G308E

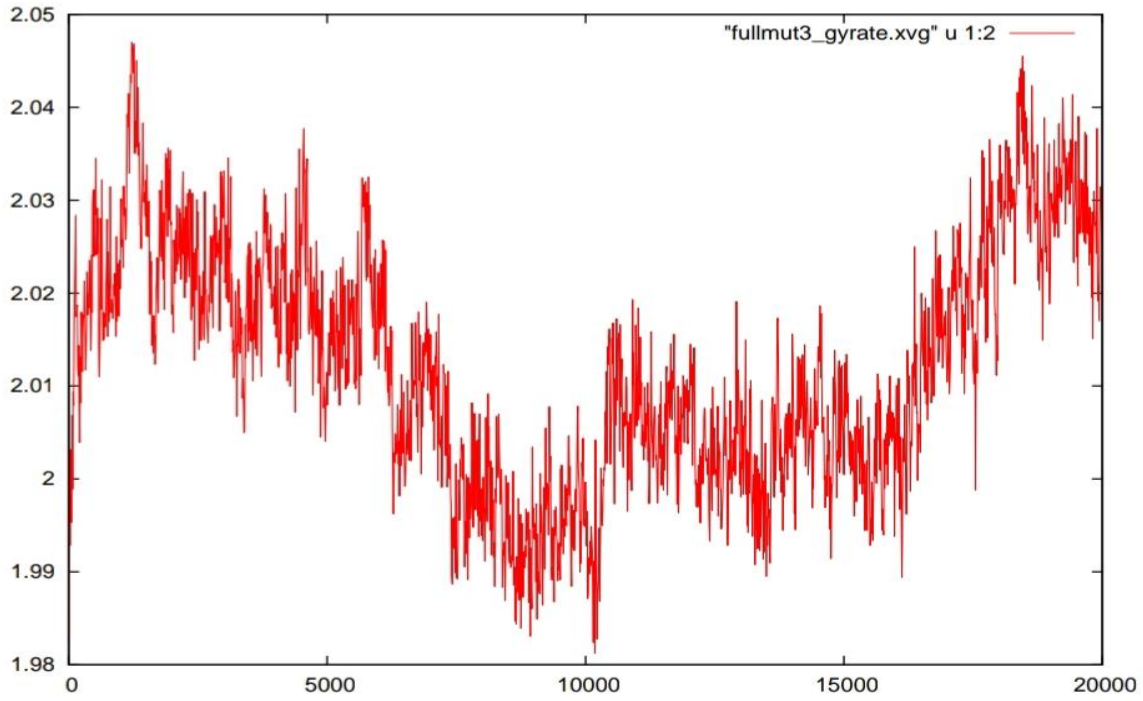


Figure 4.33: Radius of Gyration plot for mutant G308

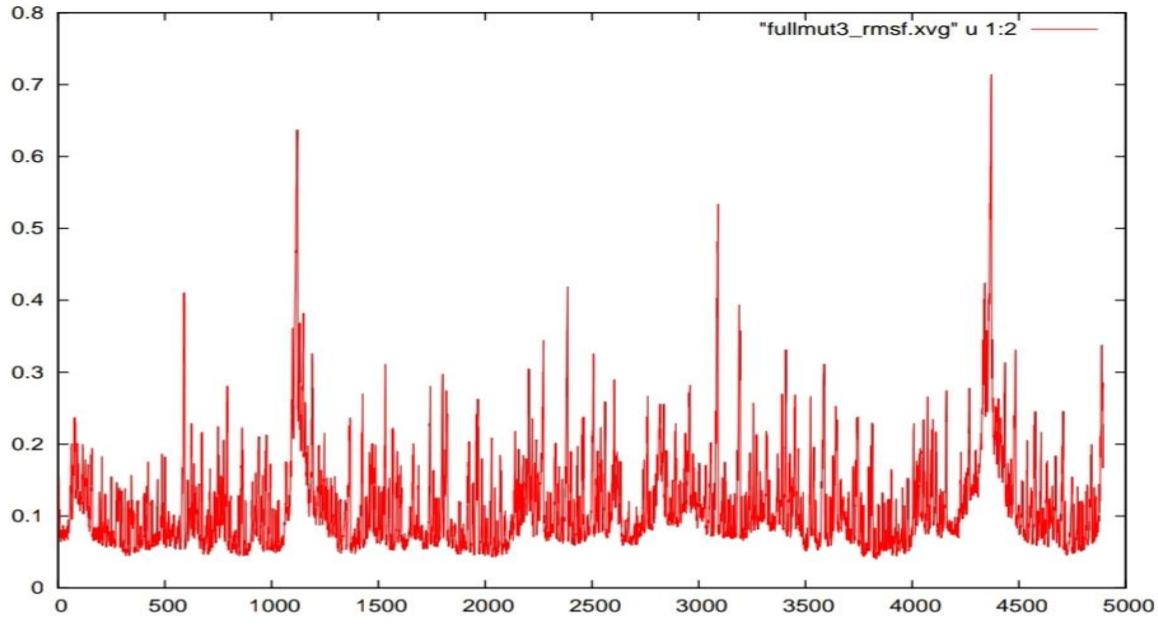


Figure 4.34: Root mean square fluctuation plot for mutant G308E

CHAPTER 5

CONCLUSION

Functional variants and polymorphisms of DNA repair genes have been the focus of several cancer association studies, but only in recent years some of them have been investigated as possible AD risk factors. The few studies performed so far suggest that some variants might play a role in AD pathogenesis and deserve further investigations. This is the basic line of thought that runs throughout the conceptualization of this project. The genes that we extracted from the research papers are – OGG1, PARP1, XRCC1 and BRCA1. OGG1 gene has 5 Damaging SNPs out of 10 SNPs and BRCA1 gene has 21 Damaging SNPs out of 38 SNPs according to the prediction tools.

The project participant wish to come up with meaningful results as the project reaches a conclusion as to how DNA Repair Mechanisms are involved in AD. There is a growing need to initiate studies on DNA damage and repair and unravel the molecular underpinnings entailed in the etiopathogenesis of the disease. The outcome of such studies substantiates the corner stone streamlined to employ therapeutic strategies.

CHAPTER 2

REFERENCES

[1] S. A. Hutchinson and H. S. Wilson, "The Theory of Unpleasant Symptoms and Alzheimer's disease.," *Sch. Inq. Nurs. Pract.*, vol. 12, no. 2, pp. 143-58; discussion 159-62, 1998.

[2] "Skilled carers' ways of understanding people with Alzheimer's disease. - Abstract - Europe PMC." [Online]. Available: http://europepmc.org/abstract/med/10189809. [Accessed: 02-Dec-2017].

[3] K.-H. Chen, S. J. Lwi, A. Y. Hua, C. M. Haase, B. L. Miller, and R. W. Levenson, "Increased subjective experience of non-target emotions in patients with frontotemporal dementia and Alzheimer's disease," *Curr. Opin. Behav. Sci.*, vol. 15, no. Supplement C, pp. 77–84, Jun. 2017.

[4] E. B. Gonzales and N. Sumien, "Acidity and Acid-Sensing Ion Channels in the Normal and Alzheimer's Disease Brain," *J. Alzheimers Dis.*, vol. 57, no. 4, pp. 1137–1144, Jan. 2017.

[5] "Structural variation in amyloid-β fibrils from Alzheimer's disease clinical subtypes | Nature." [Online]. Available: https://www.nature.com/articles/nature20814. [Accessed: 02-Dec-2017].

[6] "2017 Alzheimer's disease facts and figures," *Alzheimers Dement.*, vol. 13, no. 4, pp. 325–373, Apr. 2017.

[7] H. Fu *et al.*, "Tau Pathology Induces Excitatory Neuron Loss, Grid Cell Dysfunction, and Spatial Memory Deficits Reminiscent of Early Alzheimer's Disease," *Neuron*, vol. 93, no. 3, p. 533–541.e5, Feb. 2017.

[8] E. R. Lindemer, D. N. Greve, B. Fischl, J. C. Augustinack, and D. H. Salat, "Differential Regional Distribution of Juxtacortical White Matter Signal Abnormalities in Aging and Alzheimer's Disease," *J. Alzheimers Dis.*, vol. 57, no. 1, pp. 293–303, Jan. 2017.

[9] C. Haass and D. J. Selkoe, "Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide," *Nat. Rev. Mol. Cell Biol.*, vol. 8, no. 2, p. 101, Feb. 2007.

[10] "Review: Alzheimer's Amyloid β-Peptide-Associated Free Radical Oxidative Stress and Neurotoxicity - ScienceDirect." [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S1047847700942746. [Accessed: 02-Dec-2017].

[11] C. Carvalho *et al.*, "Role of mitochondrial-mediated signaling pathways in Alzheimer disease and hypoxia," *J. Bioenerg. Biomembr.*, vol. 41, no. 5, pp. 433–440, Oct. 2009.

[12] M. T. Lin and M. F. Beal, "Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases," *Nature*, 18-Oct-2006. [Online]. Available: https://www.nature.com/articles/nature05292. [Accessed: 02-Dec-2017].

[13] M. Hashimoto, E. Rockenstein, L. Crews, and E. Masliah, "Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases," *NeuroMolecular Med.*, vol. 4, no. 1–2, pp. 21–35, Oct. 2003.

[14] P. H. Reddy, "Role of Mitochondria in Neurodegenerative Diseases: Mitochondria as a Therapeutic Target in Alzheimer's Disease," *CNS Spectr.*, vol. 14, no. S7, pp. 8–13, Aug. 2009.

[15] "Amyloid beta, mitochondrial dysfunction and synaptic damage: implications for cognitive decline in aging and Alzheimer's disease - ScienceDirect." [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1471491408000038. [Accessed: 02-Dec-2017].

[16] "Mitochondrial bioenergetic deficit precedes Alzheimer's pathology in female mouse model of Alzheimer's disease." [Online]. Available: http://www.pnas.org/content/106/34/14670.short. [Accessed: 02-Dec-2017].

[17] P. I. Moreira, "Alzheimer's Disease and Diabetes: An Integrative View of the Role of Mitochondria, Oxidative Stress, and Insulin," *J. Alzheimers Dis.*, vol. 30, no. s2, pp. S199–S215, Jan. 2012.

[18] C. Hureau and P. Faller, "Aβ-mediated ROS production by Cu ions: Structural insights, mechanisms and relevance to Alzheimer's disease," *Biochimie*, vol. 91, no. 10, pp. 1212–1217, Oct. 2009.

[19] W. R. Markesbery, "Oxidative Stress Hypothesis in Alzheimer's Disease," *Free Radic. Biol. Med.*, vol. 23, no. 1, pp. 134–147, Jan. 1997.

[20] T. Teaktong *et al.*, "Nicotinic acetylcholine receptor immunohistochemistry in Alzheimer's disease and dementia with Lewy bodies: differential neuronal and astroglial pathology," *J. Neurol. Sci.*, vol. 225, no. 1, pp. 39–49, Oct. 2004.

[21] M. Ihara *et al.*, "Quantification of myelin loss in frontal lobe white matter in vascular dementia, Alzheimer's disease, and dementia with Lewy bodies," *Acta Neuropathol. (Berl.)*, vol. 119, no. 5, pp. 579–589, May 2010.

[22] “Diagnosis of Alzheimer’s Disease | JAMA Neurology | The JAMA Network.” [Online]. Available: https://jamanetwork.com/journals/jamaneurology/article-abstract/584730. [Accessed: 02-Dec-2017].

[23] S. L. Risacher *et al.*, “Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline,” *Neurology*, vol. 89, no. 21, pp. 2176–2186, Nov. 2017.

[24] “Strategic roadmap for an early diagnosis of Alzheimer’s disease based on biomarkers - ScienceDirect.” [Online]. Available: https://www.sciencedirect.com/science/article/pii/S147444221730159X. [Accessed: 02-Dec-2017].

[25] E. Morris, A. Chalkidou, A. Hammers, J. Peacock, J. Summers, and S. Keevil, “Diagnostic accuracy of 18F amyloid PET tracers for the diagnosis of Alzheimer’s disease: a systematic review and meta-analysis,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 43, no. 2, pp. 374–385, Feb. 2016.

[26] U. Andreasson *et al.*, “CSF biomarkers for Alzheimer’s pathology and the effect size of APOE ε4,” *Mol. Psychiatry*, vol. 19, no. 2, pp. 148–149, Feb. 2014.

[27] H. P. Vranová *et al.*, “Tau protein, beta-amyloid1–42 and clusterin CSF levels in the differential diagnosis of Parkinsonian syndrome with dementia,” *J. Neurol. Sci.*, vol. 343, no. 1, pp. 120–124, Aug. 2014.

[28] S. Lisi *et al.*, “Toward sensitive immuno-based detection of tau protein by surface plasmon resonance coupled to carbon nanostructures as signal amplifiers,” *Biosens. Bioelectron.*, vol. 93, no. Supplement C, pp. 289–292, Jul. 2017.

[29] L. D. Baker *et al.*, “Aerobic exercise reduces phosphorylated tau protein in cerebrospinal fluid in older adults with mild cognitive impairment,” *Alzheimers Dement. J. Alzheimers Assoc.*, vol. 11, no. 7, p. P324, Jul. 2015.

[30] M. Ueno, Y. Chiba, K. Matsumoto, T. Nakagawa, and H. Miyanaka, “Clearance of Beta-Amyloid in the Brain,” *Curr. Med. Chem.*, vol. 21, no. 35, pp. 4085–4090, Nov. 2014.

[31] T. Schirinzi *et al.*, “Levels of amyloid-beta-42 and CSF pressure are directly related in patients with Alzheimer’s disease,” *J. Neural Transm.*, vol. 124, no. 12, pp. 1621–1625, Dec. 2017.

[32] N. Andreasen *et al.*, “First Administration of the Fc-Attenuated Anti-β Amyloid Antibody GSK933776 to Patients with Mild Alzheimer’s Disease: A Randomized, Placebo-Controlled Study,” *PLOS ONE*, vol. 10, no. 3, p. e0098153, Mar. 2015.

[33] “Molecular Structure of β-Amyloid Fibrils in Alzheimer’s Disease Brain Tissue - ScienceDirect.” [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S0092867413010295. [Accessed: 02-Dec-2017].

[34] R. M. E. Steketee *et al.*, "Early-stage differentiation between presenile Alzheimer's disease and frontotemporal dementia using arterial spin labeling MRI," *Eur. Radiol.*, vol. 26, no. 1, pp. 244–253, Jan. 2016.

[35] Y. Zhang *et al.*, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Front. Comput. Neurosci.*, vol. 9, Jun. 2015.

[36] N. Okamura *et al.*, "Non-invasive assessment of Alzheimer's disease neurofibrillary pathology using 18F-THK5105 PET," *Brain*, vol. 137, no. 6, pp. 1762–1771, Jun. 2014.

[37] "Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: Phase 3 study - ScienceDirect." [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1552526015000606. [Accessed: 02-Dec-2017].