COMPUTATIONAL INVESTIGATION ON INTERLEUKIN FAMILY FOR ITS POTENTIAL ROLE IN ALZHEIMER'S DISEASE

Dissertation submitted in partial fulfilment of the requirement for the degree of

Master of Science

in

Biotechnology

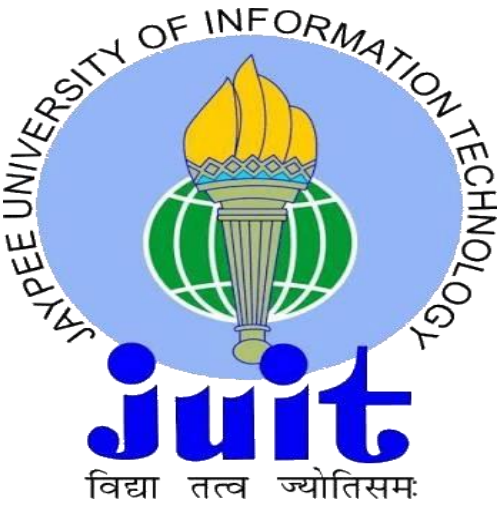
Submitted by

KHUSBOO PANDEY

Enrollment No.: 225111014

Under The Supervision of:

Prof. Dr. TIRATHA RAJ SINGH



Department of Biotechnology & Bioinformatics

JAYPEE UNIVERSITY OF INFORMATION AND TECHNOLOGY,

WAKNAGHAT, SOLAN-173234, HIMACHAL PRADESH

May, 2024

TABLE OF CONTENTS

DECLARATION

I hereby declare that the work presented in this report entitled **"Computational investigation on Interleukin Family on its potential role in Alzheimer's Disease"** for the award of the degree of Master of Science in Biotechnology submitted in the Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from August 2023 to May 2024 under the supervision of Dr. Tiratha Raj Singh (Professor, Department of Biotechnology and Bioinformatics). I also authenticate that we have carried out the above-mentioned project work under the proficiency stream. The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Khusboo Pandey
(225111014)

Department of Biotechnology and Bioinformatics, Jaypee University of Information and Technology, Waknaghat, Solan-173234, Himachal Pradesh, India

Date: 20/05/2024

SUPERVISORS' CERTIFICATE

This is to certify that the work reported in the M.sc dissertation entitled **"Computational investigation on Interleukin Family on its potential role in Alzheimer's Disease"**, submitted by Khusboo Pandey (225111014) at Jaypee University of Information and Technology, Waknaghat, India, is a bonafide record of her original work carried out under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Signature of Supervisor

Dr. Tiratha Raj Singh
(Professor)

Department of BT &BI

JUIT, Waknaghat, HP-173234

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to all those who have contributed to the successful completion of this dissertation report. Without their support, guidance, and encouragement, this report would not have been possible. First and foremost, I would like to thank my supervisor Dr. Tiratha Raj Singh, for their invaluable guidance, expertise, and continuous support throughout the entire process. Their insightful suggestions, constructive feedback, and unwavering encouragement have been instrumental in shaping this report. I would also like to extend my gratitude to the faculty members and staff of JUIT, especially the Department of Biotechnology and Bioinformatics, for providing me with a conducive environment for learning and conducting research. Their dedication to academic excellence and commitment to fostering a stimulating intellectual atmosphere has been instrumental in enhancing my knowledge and understanding of the subject matter. I would also like to express my appreciation to my peers who have provided support and encouragement throughout this journey. Their insightful discussions, constructive criticism, and motivation have been immensely valuable in shaping my ideas and strengthening my research. Lastly, I would like to thank my family for their unconditional love, unwavering support, and understanding. Their constant encouragement and belief in my abilities have been my driving force. In conclusion, I am deeply grateful to all the individuals and institutions who have contributed to the completion of this dissertation report. Their support, guidance, and encouragement have been invaluable, and I am truly honored to have had the opportunity to work with them.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Alzheimer's disease known as neurodegenerative disorder that led to abnormalities in entorhinal cortex, hippocampus and cerebral cortex. Usually in the beginning it affects the brain tissue like frontal lobe, temporal lobe and then extend towards other parts of neocortex. This disease cause, is linked with the abnormal protein build-up in and around of brain cell, includes amyloid plaque around neuron cell and tau protein which form tangles within brain cells. It has seen that Alzheimer's disease associated chronic neuroinflammation directly effects synaptic/neuronal loss and cognitive decline. In neuroinflammation, activation of glial cells happens and result in subsequent production of inflammatory factors such as cytokines and chemokines around deposited plaque at brain. These cytokine family includes major class of Interleukins which were observed in AD patients.

Therefore, by understanding the important role of different Interleukins participating in Alzheimer's disease we could contribute towards management of Alzheimer's disease, its functional and evolutionary aspects. With the aid of computational investigations on interleukin family our target is to identify the major class of interleukin which play potential role in Alzheimer's disease for the disease management. The computational investigation involves codon-based analysis, protein and sequenced based studies, structured based studies, conservation studies, functional and evolutionary studies.

CHAPTER 1

INTRODUCTION

Major elderly old population are affected with Alzheimer's Disease as trend going on everywhere currently it is hot issue need to be resolve for which scientists from all over the world focusing the severity of this condition to understand the cause with aim to resolve it with good solution. Various studies were happening in different aspects to target the cause, under that neuroinflammation taking lot of attention of scientist as in AD patient this is a major activity found involved cellular level and effecting metabolism. Considering the neuroinflammation in focus, within the classes of proteins involved in neuroinflammation of AD interleukin is interesting sub-class majorly contributing to neuroinflammations. Sequence and Structure bases computational analysis help in identification of the potential interleukins contributing in Alzheimer's disease for its better management.

Initially starting from the identification of interleukin members which are potentially involved and marked related to AD in different studies. Afterwords conducted different analysis toward these interleukins to embark the major hit potent interleukin could be the major factor aiding the Alzheimer's Disease. Later on, conducted the codon analysis for all potential interleukin and identified top expressing interleukins that are IL-1β, IL-1α, IL-10, IL-17β & IL-27, these are involved in cellular functioning and gene expression in major. Checked for variations and similarities of interleukins by using multiple sequences alignment of nucleotide and protein sequences. Additionally did functional and evolutionary based computational investigation for help in better understanding of relationship between interleukins and cause of Alzheimer's disease which have potential to open up new aspects regarding the disease and neuroinflammation. Conservation analysis illustrate that the interleukins sequences are less conserved in nature. SNPs analysis of interleukins help in understanding whether the single mutation in interleukin sequence can affect the structure and function of protein and to suggest its damaging or disease-causing properties. Initiate SNPs analysis with interleukin-1β (protein structure identifier 1HIB), filtered nsSNPs using sequenced and structure based computational tools and run molecular dynamic simulation with aim to understand its biological activity and importance. Evolutionary study vision the molecule conserving nature and adaptability of interleukin towards the environment.

Lastly, for any conclusion it is needed to check activity and importance of each potential interleukin member for better understanding and effect of its over neuroinflammation in parallel to Alzheimer's disease.

1.1. Aim:

Ultimate purpose of the computational investigation of interleukin family is to identify its potential role in Alzheimer's patient for better management of this disease.

"If Interleukins has potential involvement in Alzheimer's Disease management, functional and evolutionary area as its fundamental expression in CNS then with its aid we can control and recognise the cause and responses in AD patient."

1.2. Objectives:

- Cytokine class contribute to major aspects of neuroinflammation, including pro- and anti-inflammatory processes in Alzheimer's Disease.
- Under which Interleukin Family are observed in major scales. Interleukins study can play valuable role in understanding the evolutionary aspects and for the management of Alzheimer's Disease.

CHAPTER

REVIEW OF LITERATURE

Alzheimer's disease (AD) is a type of brain disorder also known as Senile dementia, which affects memory and task performing abilities. Brain cells degenerate and die, eventually resulting in poor mental functions. It is most common among older adults, as a cause of dementia. This disease was named on the behalf of Dr. Alois Alzheimer. It about the year 1906 when, Dr. Alzheimer found abnormal differences in the brain tissue of a woman, whose reason of death is due to an undefined mental illness. Her complaints include memory loss, language expressing problem, and unpredictable behaviour or reactions [1].

In a diseased person, symptoms get worse over time, however the rate of progression differs. According to different studies an average diseased person lives 4-8 years after diagnosis but sometime can live more this all depend on other factors including Alzheimer's stage.
It is a neurodegenerative disorder which lead to abnormalities in entorhinal cortex, hippocampus and cerebral cortex. Usually in the beginning it affects the brain tissue like frontal lobe, temporal lobe and then extend towards other parts of neocortex.

This disease is linked to the accretion caused at extracellular space of brain and in blood vessels, due to the deposition of plagues and tau protein. Plaques consist of amyloid-β (Aβ), which is insoluble in nature. However, tau is a microtubule protein which causes neurofibrillarytangles in nerve cells [2].

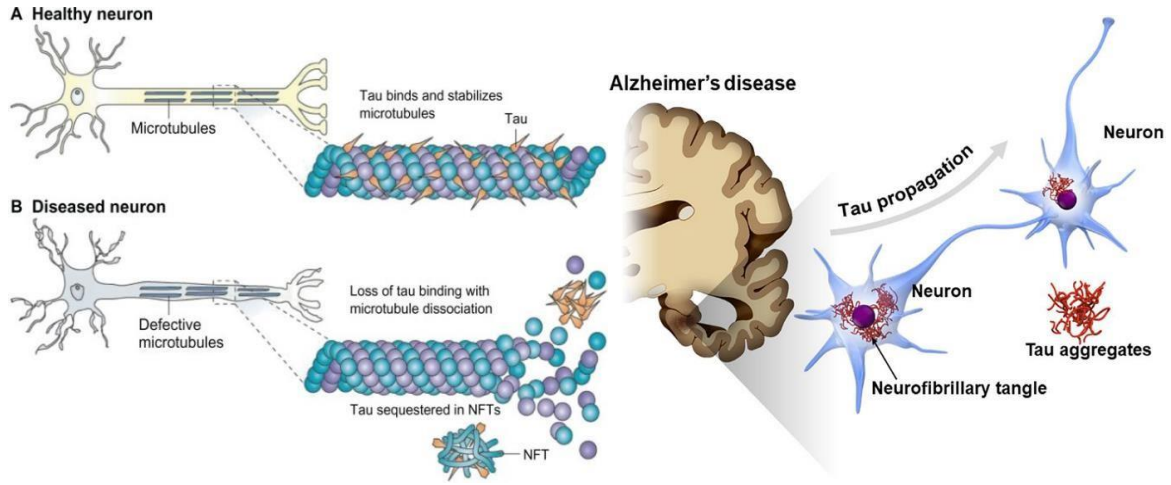


Fig1: (a) and (b) Shows hyperphosphorylation of Tau protein, forms insoluble aggregates that can fill the entire intracellular space of a neuron. Adapted from K. R. Brunden et al. (2009) and S. Takeda (2019).[4][5]

Categories of Alzheimer's Disease

- Early onset Alzheimer's disease (Familial): Age <65 year

  Less than 5% cases of early onset. Genes involved in this case are APP, PSEN1 and PSEN 2 which work intrinsic manner and result in increased Aβ level inflict overloading of late endosomes and lysosomes.

- Late onset Alzheimer's disease (Sporadic): Age >65 year

  More than 95% cases of late onset. Genes involved in this case are APOE(eg.APOe4), BIN1, CD2AP etc. Which work extrinsically and result in endolysosomal flux disturbances alter APP trafficking and processing.

This amyloid in plaque is a result of proteolytic cleavage of amyloid precursor proteins, belong to the complex enzymes family members (γ and β-secretases), concluding presenilin 1 (PS1) and presenilin 2 are encoded by PSEN1 and PSEN2. Overproduction of amyloid-β (Aβ) is a cause of pathogenetic mutation in these encoding genes are quit seen at early age of 40s in (<1%) of diseased patient having autosomal nature of dominant inherited Alzheimer's Disease (DIAD), these cases are generally unusual. According to studies, proteolytic functions in APP metabolism involves amyloid-β (Aβ) production which is normal. However, it may provide plasticity to the nerve cells but the exact role of APP is not known in neuron case.

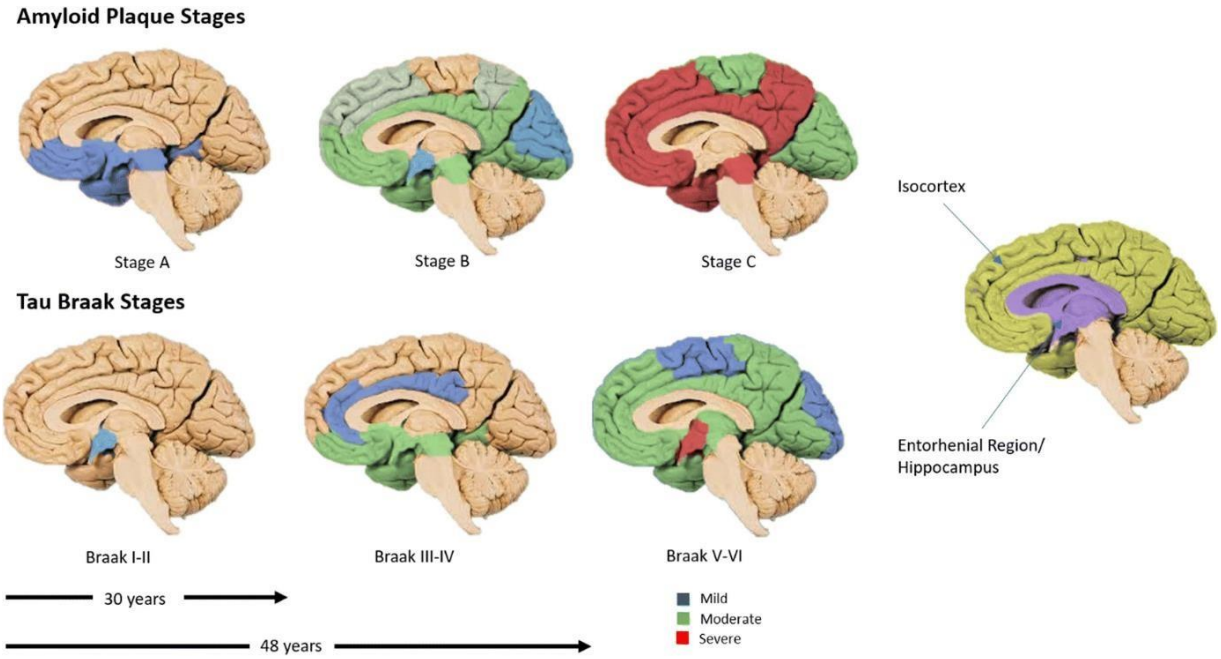


Fig.2: Different stages of plaque and tau protein deposition in the different brain parts of patient suffering from Alzheimer's disease. Adapted from S. Swarbrick.[6]

Environmental risk factor of sporadic Alzheimer's disease was not evident, frequently in the nature. However, one attention seeking risk factor come across via analyzing various studies

are linked with genetic relations. Among them the most reliable cause is related to polymorphism happened at the gene encoding stage of apolipoprotein E (APOE), present on chromosome19 which participate in the determination of Alzheimer's disease on the basis of age factors.

In addition to genetic factors bit alarming risks are found in different genomic studies, which involves PICALM, ATP binding cassette transporters (eg. ABCA7), clusterin, CD33 and TREM2 (both are molecules expressed by myeloid cells). In addition to APOE, many of these genes might play a part in the clearance pathways of Aβ. With help of SILK Technique (stable isotope labelling kinetics) we can measure the construction and clearance of Aβ with in the central nervous system. [3] This labelled protein help in analysing, positive or negative effect of different drugs that target Aβ generation to demonstrate any decrease in the production of amyloid. Somehow, the working mechanism participating in the amyloid removal is still unknown but could related to different factors such as APOE or misfunctioning of microglial cells or macrophages for the depletion of amyloid clogs [2].

APOE protein is constitute of 299 amino-acid with 3 types of isoforms which are common in humans they show variation of 1 to 2 amino acids only. In context to different studies APOE could be taken as effecting reason for the build of Alzheimer's disease, by working in a way like chaperone for Aβ protein, which affects the construction and clearance of Amyloid plaque.

Diagnostic method for Alzheimer's Disease (AD):

- Amyloid β PET imaging
- CSF profile of AD patient

Table 1: Alzheimer's diagnostic and clinical test.

Examination	Biomarkers	Pathological process	Pathology	Objective
Method	CSF profile, where rate of biochemical are measured.	Genetic testing (risk or protecting factors).	MRI structuring, Aβ and PET imaging.	Psychometric, neurological and CDR-SB examinations.
Result	CSF profile, where decreased level of Aβ42 and increased level of T-tau and P-tau are measured.	Mutation occurred at PSEN1, PSEN2, APP or APOE alleles (2/3/4).	Aβ and tau aggregation, cortical and hippocampal atrophy.	Dementia stage, attentiveness, memory, functional and cognitive functioning.

- **Neuroinflammation in Alzheimer's Patient:**

Activated of microglia or astrocytes produces inflammatory factors such as cytokines and chemokines which causes neuroinflammation in AD patient. Reported inflammatory secretions get deposited around the plaque and nerve cells of brain. Neuroinflammation is found to be key feature of Alzheimer's Disease.

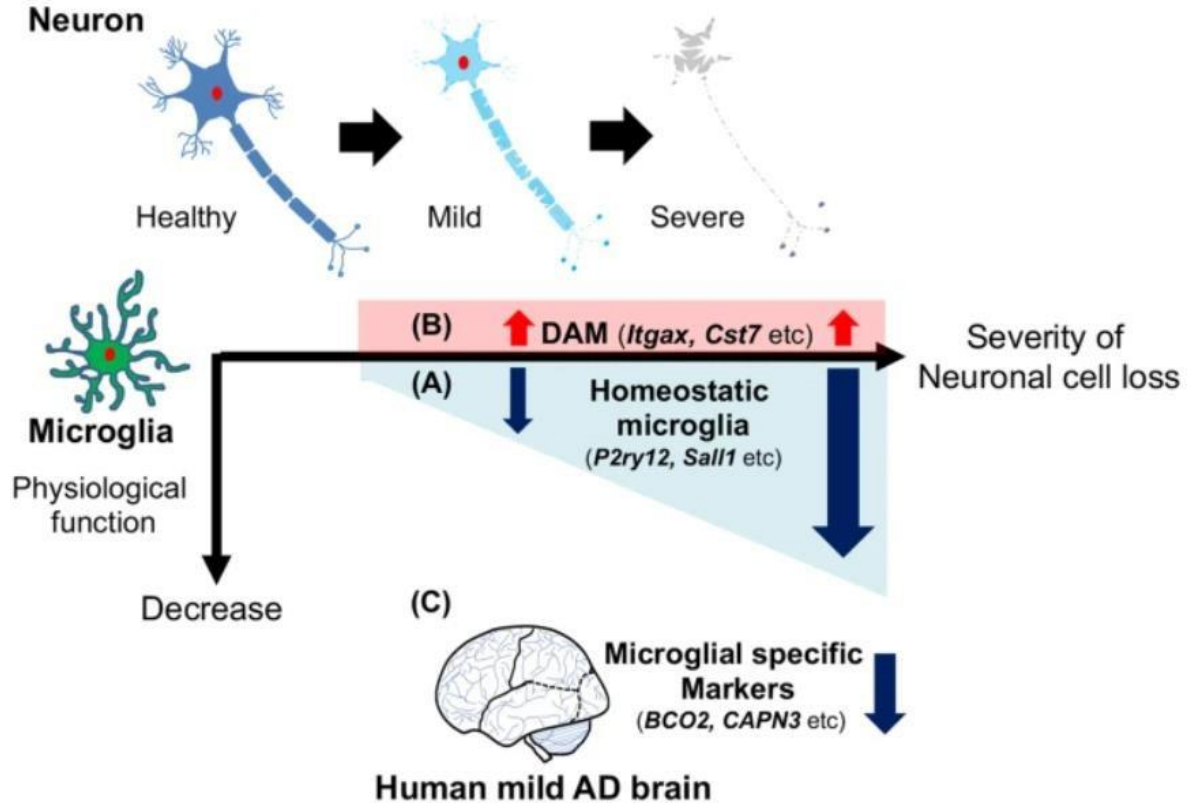


Fig.3: Neuron and microglia function in Alzheimer's patient. Adapted from A. Sobue, et al. (2023).[11]

Microglia are the cells which regulate neuronal network play important role in development, maintenance and injury repair of brain. Glia cells act as macrophages works but are bit different from the macrophages found at other body tissues, except brain. They have their extraordinary homeostatic along strong regulation over the nervous system. They play crucial role in removing dead cells, protein aggregates, redundant synapses, and other toxic soluble antigens from brain. In addition, microglia involve in production of proinflammatory cytokines, help in better and diverged cellular neural responses in brain with better synaptic transmissions and message delivery. Around $1/10^{th}$ of cells in found in central nervous system are microglia cells (mostly mononuclear phagocytes). Changes in microglia functionality result in neurodegeneration [10,[11].

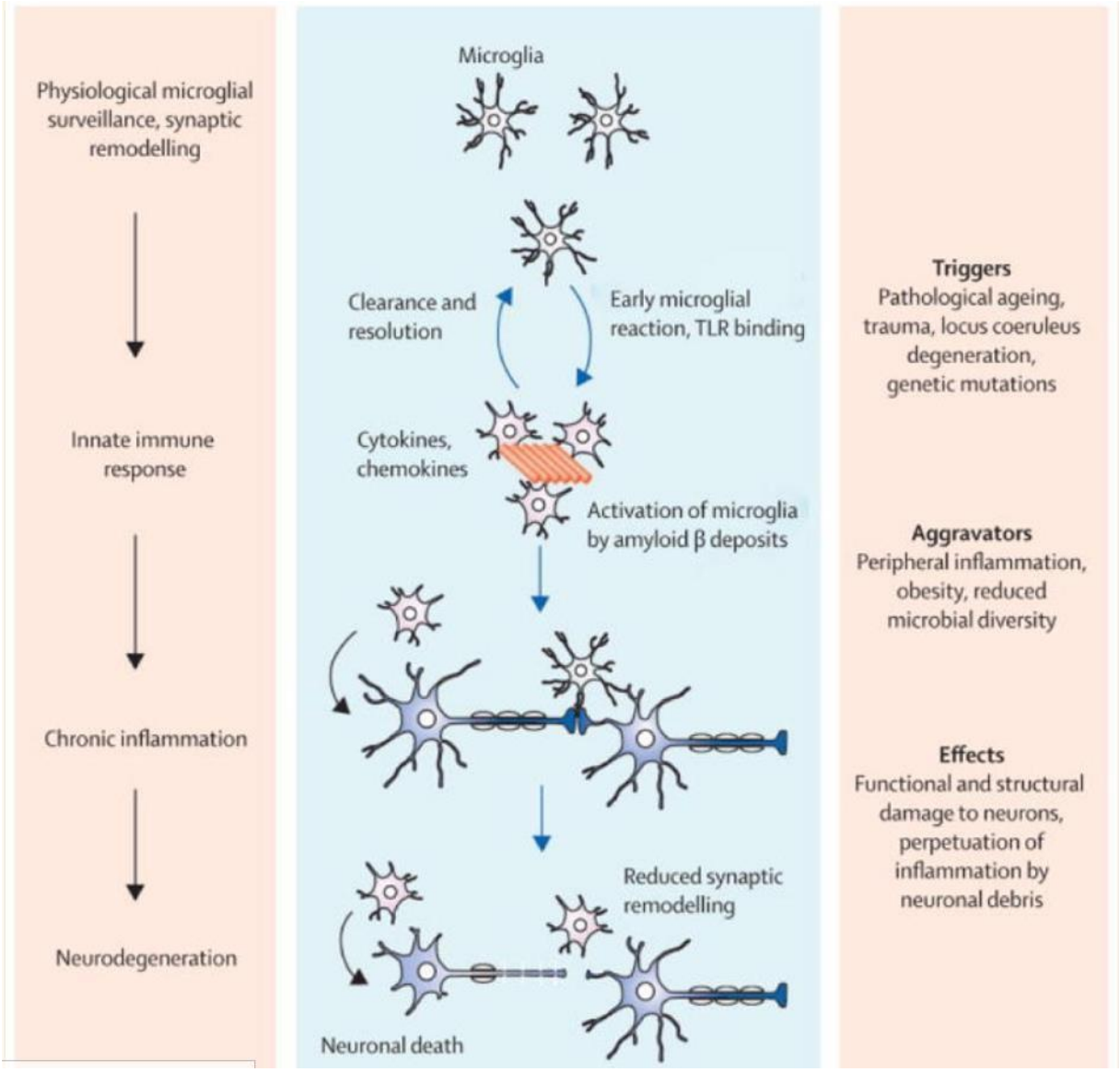


Fig.4: Microglia physiological role in neuronal death and neuroinflammation. Adapted from M. T. Heneka et al. (2013).[12]

Theory behind the pathological triggers (neural death or protein agitation shown that receptors present to pathogen molecule is greatly related to its molecular structure (DAMPs/PAMPs). It has been studies that microglia are capable to bind with amyloid oligomers and fibrils with the receptor molecules covering scavengers (A Class) A1, CD14/36/47, α6β1 integrin, and toll like receptors (TLR2/4/6/9), can be potential reason of the neuroinflammatory reaction in Alzheimer's patient. It is known that amyloid plaque peptide belongs to the large precursor protein. Larger precursor protein split due to action of two membrane bounded proteases, these proteases known as BACE1 (β site APP cleaving enzyme-1 and γ-secretase complex found at transmembrane area of amyloid precursor protein resulting in variable truncated C-termini which diverse from amino acid 37 to 42 chains. However, amyloid formed out of amino-acid with long chain show strong tendency to form the foregoing oligomers which are soluble in

nature and fibrillation. Binding of Aβ at the receptors which includes CD36/TLR4/TLR6, will lead to activation of glia cells of CNS which results in developing pro-inflammatory chemokines and cytokines. If any genetic disturbance occurs at CD36, TLR4/6 receptors by in-vitro means will, effect in less production of amyloid induced cytokines and help in preventing the intracellular amyloid accumulation and inflammasome activation [7],[8],[9].

Table 2: Represents the pro-inflammatory IL-family members.

Pro-inflammatory Interleukins	Its role in Neuroinflammation
IL-1β	Participate in neuroinflammation, induce neural death via downstream signal pathway.
IL-6	Participate in neuroinflammation, activates JAK-STAT pathway and maintain in blood brain barriers.
IL-17	Participate in neuroinflammation, activates pro-inflammatory pathway and promote tissue damage via neutrophils aid.
IL-23	Participate in neuroinflammation, activates pro-inflammatory pathway and promote tissue damage via Th17 cells differentiation and pro-inflammatory cytokines production.
IL-36	Promotes neuroinflammation along with elevation in production of inflammatory factors (such as cytokines, chemokines).

Table 3: Represents the anti-inflammatory IL-family members.

Anti-inflammatory Interleukins	Its role in Neuroinflammation
IL-2	Help in neuro-protection via promoting differentiation mechanism of nerve stem cells.
IL-4	Help in neuro-protection via inhibition of pro-inflammatory cytokines production in addition, it participates in inhibition of neuroinflammation and enhance in production of anti-inflammatory cytokines.
IL-10	Promotes neuroprotection via inhibition of pro-inflammatory cytokines production along with it participates in inhibition of neuroinflammation and enhance in production of anti-inflammatory cytokines.
IL-27	Promotes neuroprotection via inhibition of pro-inflammatory cytokines production and participate in inhibition of neuroinflammation and enhance the production of type 1 regulatory T cells.
IL-35	Promotes neuro-protection via inhibition of pro-inflammatory cytokines production and participate in inhibition of neuroinflammation and enhance the production of regulatory T cells.

Table 4: Represents the other potential IL-family members. [13-19]

Other Interleukins	Its role in Neuroinflammation
IL-1α	Induces neuroinflammation, promotes neural death via activation of inflammasome and downstream signaling.
IL-9	Maintain human inflammatory macrophages within the CNS (could be relevant to other CNS disorders).
IL-12	Studies suggests that it could be the part of neuroinflammation mechanisms in cognitive imbalance in AD.
IL-15	Its increase in CNS is less evident in studies with relation to cognitive impairment.
IL-1(ra)	Promotes neuroinflammation, neuronal cell damage and death via activation of the inflammasome pathway and other pathway like downstream signaling.

Computational investigations on potentially strong Interleukin Family members, which are specifically related to neuroinflammation for better functional and evolutionary understanding for the Alzheimer's Disease management.

CHAPTER 3

MATERIALS AND METHODS

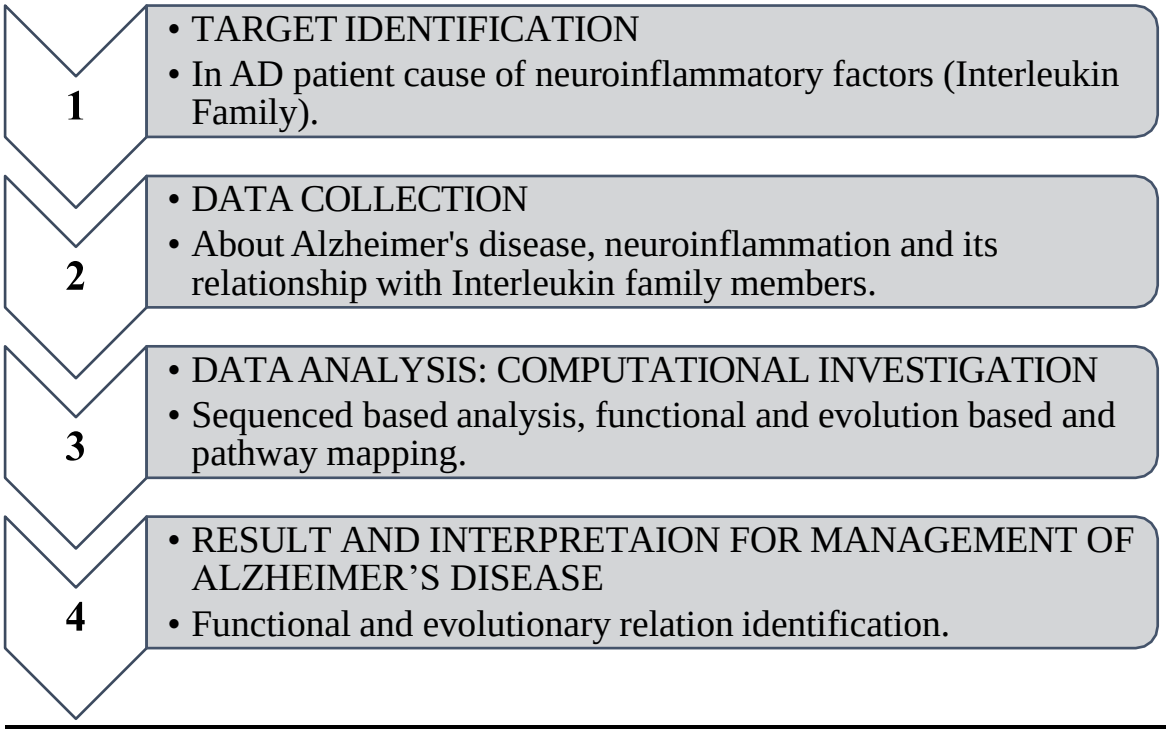


Fig 5: Methodology working.

- **Target Identification**

  Potential targeted Interleukin family members were identified which shows relation with Alzheimer's Disease or other dementia causing cognitive disorder, reported via literature studies, articles, etc.

  Interleukin members with which we are working are IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-9, IL-10, IL-12, IL-15, IL-16, IL-17, IL-18, IL-23, IL-27, IL-33, IL-35, IL-36 & IL-37.

- **Data Collection**

  After preparing list of Interleukins participate in this disease we need to collect data about all reported members of interleukin family.

  - Searched for cds nucleotide sequences by using database sever; NCBI and downloaded it in fasta format.
  - Similarly, searched for protein sequences of different interleukins and created a fasta file. Database used for protein sequence; UniProt.

- Checked for different type of transcripts variants of interleukin to identify the best fit coding sequence of interleukin members. Database used for this purpose, Ensembl.

- **Data Analysis**

  Sequence based analysis conducted under which we performed codon-based analysis, multiple sequence alignment, conservation analysis.

  - Codon-based analysis is performed with the aid of CAIcal software for the evaluation E-CAI value of different members of interleukin family, for better understanding of its potential target in Alzheimer's Disease.
  - Multiple sequence alignment conducted for the analysis of similarity among the members of interleukin family on the basis of their nucleotide sequences. Computational analysis conducted by using software MUSCLE, MAFFT & Mega11, along this looked for phylogenetic tree to check relativity among the members of Interleukin Family.
  - Multiple sequence alignment conducted for the analysis of similarity among the members of interleukin family on the basis of their protein sequences. Computational analysis conducted by using software MUSCLE, MAFFT & Mega11 and obtained phylogenetic tree for interpretation of relativity.
  - Conservation analysis conducted with interest to look for conservation index of amino acid sequences of different interleukins reported for Alzheimer's Disease. Conducted the test using AACon, Jalview server.
  - Protein-protein interactions between potential interleukins playing role in Alzheimer's Disease. Interleukins interaction predicts the relationship between them and with other protein, which are actively found in Alzheimer's patient. Conducted using Strings database.
  - Molecular Dynamics (MD) Simulation has been conducted for different mutants of interleukins involved in neuroinflammation causing AD. For MD simulation analysis initially filtered for non-synonymous SNPs, further conducted sequence-based and structure-based analysis and identified major effect causing deleterious mutants for simulation analysis. MD simulation completed with the help of Gromacs.
  - Evolutionary study of interleukin has been conducted to understand the rate of evolutionary change nucleic acid in each position of sequence. ConSurf is used to estimate the evolutionary parameters.

3.1. Codon Based Analysis:

Codons are present in DNA & RNA, each of codon is constitute of three nucleotides which codes for specific amino acid. Codon usage bias (CUB) varies with species, even among the organism genes. CUB help in determination of cellular functioning and gene expressions. Evolutionary effects cause change to codon bias through natural selection, genetic drift and mutations. Codon usage depicts the ornamentation of genome which describes phylogenetic relationship among organisms.

Genome composition, recombination rate, GC content, codon position, t-RNA, m-RNA interactions gene length and expression are some elements which influence the codon bias. Degree of gene expression regulated via transcriptional and post translation processes, according to which high GC content was observed at 3rd position. Variation in GC content is majorly affected via 3rd position nucleotide in a coding sequence reflects the increased stability, adaptation and affects the codon usage bias. Greater the GC (guanine, cytosine) content at 3rd position suggest variation in codon usage is due to mutation at DNA level rather than natural selection at cellular level. Codon bias analysis helps in understanding gene expression and modification of its expression level along with that we can use it for designing transgene [20].

Sequences of interleukin members were downloaded along with that codon frequency for homo sapiens using Kazusa codon usage database, in order to set a parameter for codon-based analysis. After all used CAI calculator for codon usage analysis.

```
UUU 17.6(714298)  UCU 15.2(618711)  UAU 12.2(495699)  UGU 10.6(430311)
UUC 20.3(824692)  UCC 17.7(718892)  UAC 15.3(622407)  UGC 12.6(513028)
UUA  7.7(311881)  UCA 12.2(496448)  UAA  1.0( 40285)  UGA  1.6( 63237)
UUG 12.9(525688)  UCG  4.4(179419)  UAG  0.8( 32109)  UGG 13.2(535595)

CUU 13.2(536515)  CCU 17.5(713233)  CAU 10.9(441711)  CGU  4.5(184609)
CUC 19.6(796638)  CCC 19.8(804620)  CAC 15.1(613713)  CGC 10.4(423516)
CUA  7.2(290751)  CCA 16.9(688038)  CAA 12.3(501911)  CGA  6.2(250760)
CUG 39.6(1611801)  CCG  6.9(281570)  CAG 34.2(1391973)  CGG 11.4(464485)

AUU 16.0(650473)  ACU 13.1(533609)  AAU 17.0(689701)  AGU 12.1(493429)
AUC 20.8(846466)  ACC 18.9(768147)  AAC 19.1(776603)  AGC 19.5(791383)
AUA  7.5(304565)  ACA 15.1(614523)  AAA 24.4(993621)  AGA 12.2(494682)
AUG 22.0(896005)  ACG  6.1(246105)  AAG 31.9(1295568)  AGG 12.0(486463)

GUU 11.0(448607)  GCU 18.4(750096)  GAU 21.8(885429)  GGU 10.8(437126)
GUC 14.5(588138)  GCC 27.7(1127679)  GAC 25.1(1020595)  GGC 22.2(903565)
GUA  7.1(287712)  GCA 15.8(643471)  GAA 29.0(1177632)  GGA 16.5(669873)
GUG 28.1(1143534)  GCG  7.4(299495)  GAG 39.6(1609975)  GGG 16.5(669768)

Coding GC 52.27% 1st letter GC 55.72% 2nd letter GC 42.54% 3rd letter GC 58.55%
```

Fig 6: Codon frequency parameter retrieve from Kazusa database.

3.1.1 CAI (Codon Adaptation Index):

Codon adaptation index is used to calculate, the degree of codon usage bias for which all codon frequencies are determined and compared. CAI value lies between 0 to 1.0. Greater the CAI

more will be expression level with strong CUB. Software used for calculation of CAI value are CAIcal, COUSIN, etc. [21]. Database used for codon usage analysis is Kazusa codon database is a tabular information accessed through GeneBank &NCBI Ref seq.

Effective number of codons (ENC) determines the degree of deviation in uniform codon usage however CAI measures degree of bias toward specific set of codons. Higher CUB is determined via low ENC value and high CAI values [20], [22].

3.1.2 RSCU (Relative synonymous codon usage):

It represents the ratio between the observed codon frequency to the expected frequency. RSCU values help in better understanding of codon usage. If value in 1, then it means no codon bias. However, value less than 1 shows relative scarcity and more than 1 stands for more codon expectant.[20]

3.2. Multiple Sequence Alignment of Interleukins:

MSA help in predicting the relationship between the sequences based on their evolutionary relatability or patterns. MSA could performed for protein or nucleotide (DNA/RNA) sequence in order to check the alignment between the sequences. There are many tools available for sequence alignments, e.g. MAFFT/MUSCLE/CLUSTAL Omega and Mega11.

Firstly, collected data of all interleukin sequences potentially affecting AD patient from NCBI database. Both RNA (coding) sequence and protein sequence are used. Preformed MSA using online tools and analysed results. Checked for phylogenetic relations also.

3.2.1 Multiple Sequence Alignment using nucleotide sequences of different Interleukins

All interleukins which are potentially occurs in Alzheimer's Disease are identified and m-RNA sequences are retrieved to evaluate sequence alignment. After this performed MSA by using tool MAFFT/CLUSTAL omega/ Mega and visualized the results.

Further on checked for phylogenetic relations among the interleukin members.

3.2.2 Multiple Sequence Alignment using protein sequences of different Interleukins

All interleukins which are potentially occurs in Alzheimer's Disease are identified and protein sequences are retrieved to evaluate sequence alignment. After this performed MSA by using tool MAFFT/CLUSTAL omega/ Mega and visualized the results.

Further on checked for phylogenetic relations among the interleukin members.

Sequence and Structure bases computational analysis help in identification of the potential interleukins contributing in Alzheimer's disease for its better management.

3.3. Conservation Analysis of Interleukins

Conservation analysis of potent interleukins was predicted via a help of AACon calculation have been performed by using Jalview servers. Conservation calculating tool is used to check for prediction of conserved nature of different interleukins sated in human.

Conservation of interleukins was measured on the basis of conserving index, representing physio chemical properties in sequences alignment.

3.4. Protein-Protein Interaction

Identification of interaction networks between the potential interleukin members playing role in neuroinflammation in AD. Also, their interaction with other important proteins found active in Alzheimer's Disease.

For this we used STRING software. It is based on association of biological properties and relationship between the proteins. This protein-protein interactions were aggregated via primary data available [23]. String database help in identification of known interaction, predicted interaction, co-expression and other parameters like co-occurrence and protein homology.

3.4.1. Protein-Protein Interaction among interleukins and other protein actively present in AD patient

Interleukins interactions were created along with other proteins which are associated with Alzheimer's disease, using STRING software. For which we input disease detail (Alzheimer's Disease) and actively present proteins in AD patient metabolism and follow up by more wide interaction with other favourable proteins.

3.4.2. Protein-Protein Interaction among potential interleukin members

In order to predict protein-protein interactions among the interleukin members (IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-9, IL-10, IL-12, IL-15, IL-16, IL-17, IL-18, IL-23, IL-27, IL-33, IL-35, IL-36 & IL-37) which we found relatable to Alzheimer's Disease**.** To retrieve string results, we follow up by inputting the interested interleukins and follow up by more wide interaction with other favourable proteins.

3.5. Single Nucleotide Polymorphisms (SNP) Analysis:

After identifying the potential Interleukin family members, which has shown participation in neuroinflammation caused by Alzheimer's Disease. We started with those Interleukin which scored high e-CAI value in previously conducted codon analysis.

SNPs Analysis help in identification of non-synonymous SNPs which affect the protein functioning and cause damage in the metabolism.

We initiated our SNP analysis with **Interleukin-1β** as it is noted most prominently active according to different researches.

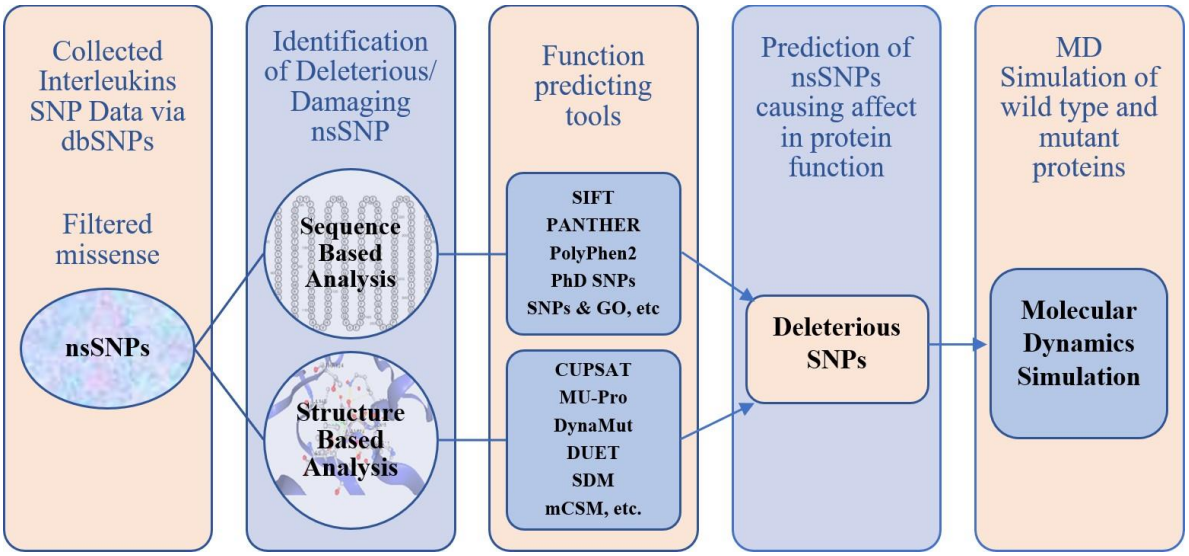


Fig 7: Methodology for SNPs Analysis.

Identification of non-synonymous (missense) mutation occur in Interleukin-1β with the help of dbSNP data (NCBI's database). Protein sequence was extracted in FASTA format from Ensembl, UniProt (UniProt ID: P01584), and protein structure of IL-1β (interleukin-1 beta, homo sapiens identifier) was downloaded from Protein Data Bank (PDB ID: 1HIB), in pdb format.

All non-synonymous SNPs were analyzed with SIFT tool for primary filtering of deleterious nsSNPs. Sorting of intolerant from tolerant (SIFT) was performed primarily in order to predict which amino acid substitution causing affect in protein function. A point mutation or single change in amino acid sequence can cause problematic effects to its 3D structure and protein functioning. Sequence and structure-based tools help in prediction of deleterious mutation by predicting its effect in functioning of protein. In sequence-based tools deleterious, diseased, probably and possibly damaging nsSNPs are amino acid substitution causing change in protein

function whereas benign are with low confidence score due to nearly no damage. For molecular dynamics simulation we use nsSNPs mutants which are destabilizing in nature. Structure validation prediction is done via help of MD simulations for better understanding of functional metabolism and structural ability of original protein (IL-1B) and mutants [24-29].

3.5.1. Sequence Based Analysis of nsSNPs of Interleukin-1β:

Tools incorporated for sequenced-based analysis are SIFT, PANTHER, PolyPhen2, PhD SNPs, SNPs & GO and Predict SNP.

For the purpose of sequence analysis, we require protein sequences which we retrieved from Ensembl of all protein IDs. Sort all deleterious nsSNPs resulted from SIFT analysis. Check for each mutation in protein from other tool to validate its effect on protein functioning.

3.5.2. Structure Based Analysis of nsSNPs of Interleukin-1β:

Tools incorporated for structure-based analysis are Structure based analysis: CUPSAT, MU-Pro, DynaMut, DUET, SDM, mCSM, SNPs & GO 3D and i- MUTANT.

Some of these tools predict the protein stability on the basic of their ΔΔG value (delta delta G, Gibbs free energy) among wild type and mutant amino acid of protein (interleukin-1β) to determine whether it cause damaging or deleterious effect on protein functioning [27]. nsSNP with higher negative ΔΔG value are more favourable reaction. Therefore, more damaging amino acid substitution to protein function whereas stabilizing mutants shows neutral effect to protein function.

3.5.3. Preparation of mutants and wild-type protein structures:

For wild type and mutant protein preparation download PDB protein file of interleukin-1β by using identifier 1HIB (pdb id). After downloading pdb file we open it in PyMol for introducing mutation to the original protein. After preparing mutant proteins individually saved it as pdb file.

3.5.4. Molecular Dynamics Simulations:

After preparation of wild-type and mutant proteins. Molecular Dynamic (MD) simulations were conducted to predict active biomolecular process, it evaluates microscopic molecular movements of atoms to understand properties of protein system. For MD simulation purpose we use Gromacs software runed in Linux operating system [29-31].

We check interleukin-1β protein activity in water environment. In simulation system we prepare a box containing our interested protein along with water and ions, to create a system which mimic the environment exist originally in biological system for the protein.

MD simulation run individually for wild type protein and for other 6 mutant type proteins. After analyzing simulations result, we compare them.

3.6. Evolutionary Conservation of Interleukins (protein):

Evolutionary study of interleukin-1β needed to conducted for better understanding whether the rate of evolutionary change nucleic acid in each position of sequence. For this prediction we use ConSurf to estimate the evolutionary conservation parameters.

Used the interleukin-1β structure identifier (PDB ID: 1HIB) for identification of conserved region of the sequence, this evolutionary conservation is a result of amino acid mutation occurred in sequence which cause effect in protein functioning.

Input all details in regard to interleukin-1β to the ConSurf database and submitted the query and analysed the resulted data.

CHAPTER 4

RESULTS & DISCUSSION

4.1. Codon Based Analysis:

Estimation of expected codon adaptive index (E-CAI) of different member of interleukins involved in Alzheimer's Disease using its nucleotide sequences.

Analysis is performed by the aid of E-CAI calculator tool, present at CAI calculator for codon usage analysis.

- Results:

Given below tables show all the interleukins its e-CAI value. which have to found potentially active in ad patient. According to the resultant parameters and comparison of e-CAI value among different members of Interleukin Family we can conclude that the most potent targets were those having e-CAI value near or greater to 0.800, as this value is more significant. Under the influence of e-CAI value we can conclude that IL-1β, IL-1α, IL-10, IL-17β & IL-27 are determination to be involved in cellular functioning and gene expressions. It can be determined that these are majorly affecting interleukins in Alzheimer's Diseased person and must be involve in neuroinflammation.

Graph present in *Fig-9* shows the GC occurrence in the 3rd potion of codon majorly affects in a coding sequence reflects the increased stability. According to which there is high probability of GC occurrence at #rd position in IL-17β & IL-27.

Table 5.1: Expected Codon Adaptive Index of different Interleukins.

INTERLEUKIN FAMILY	e-CAI (p<0.05)	G+C Content (%)			
		%GC	%GC1s	%GC2s	%GC3s
IL-1 alpha	0.800 (Average = 0.745)	48.9	56.5	32.2	60.0
IL-1 beta	0.798 (Average = 0.742)	48.9	56.9	32.2	60.0
IL-2	0.663 (Average = 0.606)	38.1	47.8	37.0	30.4
IL-4	0.811 (Average = 0.754)	52.4	47.3	40.0	70.7
IL-6	0.773 (Average = 0.715)	50.0	55.4	39.2	56.9
IL-9	0.780 (Average = 0.723)	48.7	45.3	45.3	58.4
IL-10	0.826 (Average = 0.766)	51.4	53.5	33.5	70.0
IL-12	0.751 (Average = 0.695)	49.5	54.2	43.3	53.8
IL-15	0.711 (Average = 0.659)	35.8	44.6	32.3	30.0
IL-16	0.750 (Average = 0.691)	55.0	57.9	52.4	56.3

Table 5.2: Expected Codon Adaptive Index of different Interleukins.

INTERLEUKIN FAMILY	e-CAI (p<0.05)	G+C Content (%)			
		%GC	%GC1s	%GC2s	%GC3s
IL-17 (beta)	0.832 (Average = 0.768)	66.4	63.3	49.1	82.2
IL-18	0.761 (Average = 0.709)	35.4	44.9	30.8	30.8
IL-23	0.735 (Average = 0.687)	38.4	43.8	35.7	34.3
IL-27 (receptor)	0.815 (Average = 0.754)	66.0	74.6	43.1	82.3
IL-33	0.763 (Average = 0.708)	40.1	45.0	35.8	39.6
IL-35	0.757 (Average = 0.701)	45.9	50.5	38.0	51.9
IL-36 (gamma)	0.750 (Average = 0.695)	48.0	54.6	42.3	47.9
IL-37	0.784 (Average = 0.729)	49.3	54.1	40.7	54.1
IL-1(ra)	0.788 (Average = 0.731)	57.7	62.8	42.6	69.6

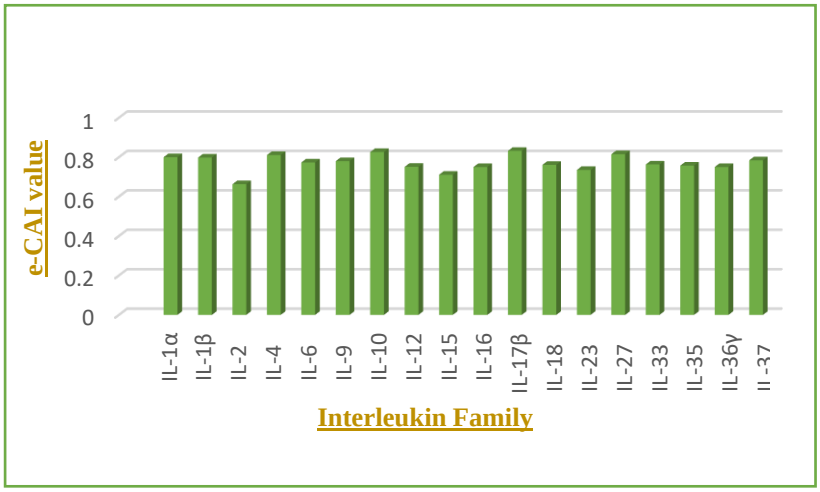


Fig 8: Following bar graph represents the e-CAI value for different interleukin members. Greatest value lies around 0.8 however least one 0.6, among the interleukins.

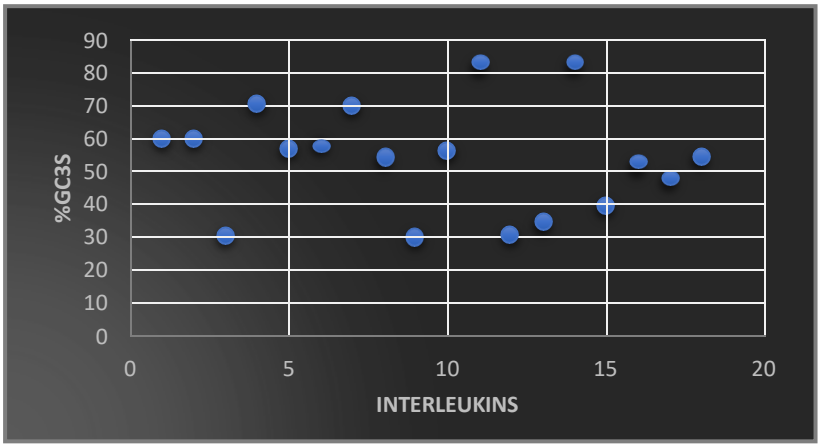


Fig 9: Following scattered graph shows %GC content occur at 3rd position of codon in sequence of all interleukin members.

Whereas blue dots in graph (***Fig-9)*** represents the different interleukin members and concentration of %GC at 3rd. Above bar graph in ***Fig-8*** represents e-CAI value in of all potential Interleukins.

4.2. Multiple Sequence Alignment using nucleotide sequences of different Interleukins

MSA is performed using online alignment tools like MAFFT/MUSCLE/CLUSTAL Omega and Mega11 for the visualization and identification of evolutionary relationship between the different interleukin family members for the analysis of any common patterns present among the sequences of Interleukins or not.

- Results:

The aligned results of multiple sequence alignment depicts that the nucleotide sequences of different interleukins involved in neuroinflammation of AD are less similar. Predicting their different metabolic functions and responsibilities.

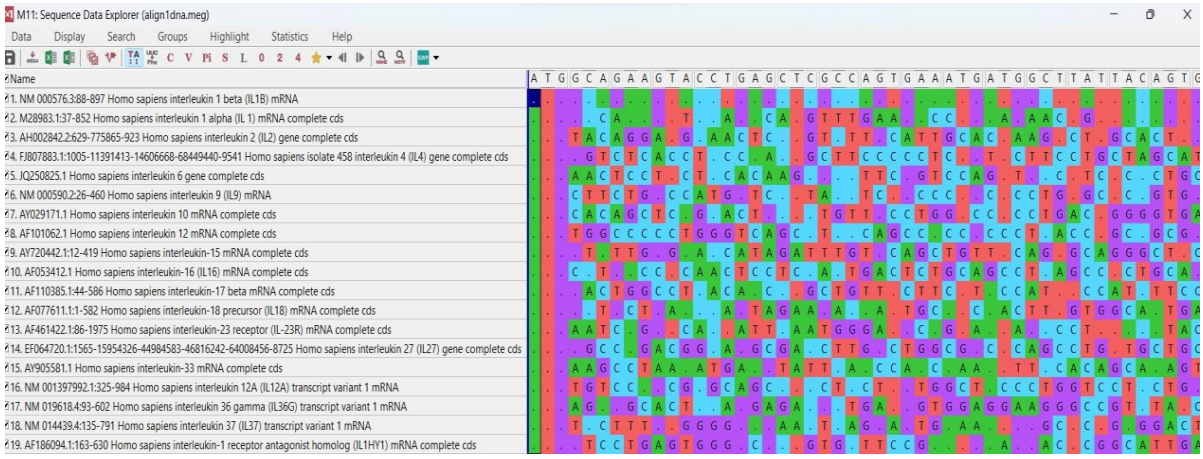


Fig 10.1: Mega11 nucleotide sequence of potential interleukin members in Alzheimer's Disease.

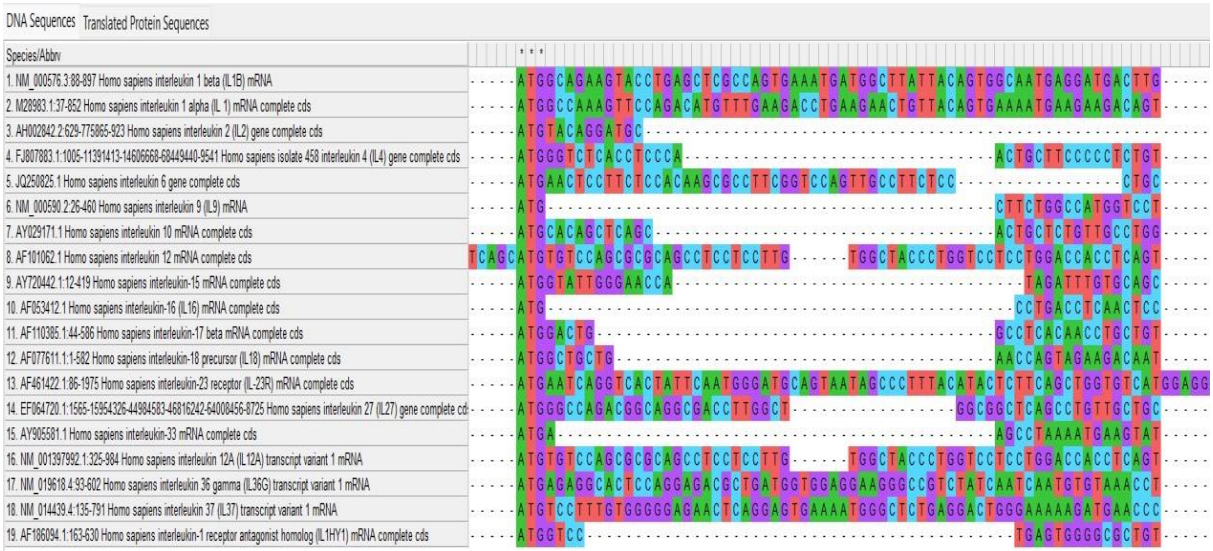


Fig 10.2: Mega11 nucleotide sequence of potential interleukin members in Alzheimer'sDisease alignment aligned result.

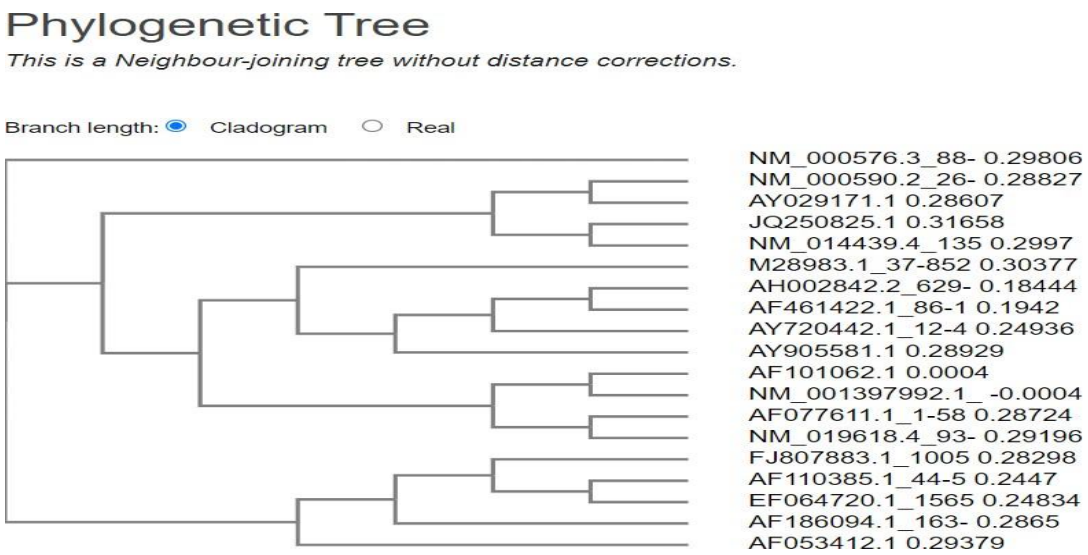


Fig 11: Phylogenetic tree representation of different Interleukins nucleotide sequences.

However, phylogenetic tree of nucleotide sequences of different interleukin members represents origin of some interleukin member remain original however other belong nearly related.

4.3. Multiple Sequence Alignment using protein sequences of different Interleukins

MSA is performed using online alignment tools like MAFFT/MUSCLE/CLUSTAL Omega and Mega11 for the visualization and identification of evolutionary relationship between the different interleukin family members for the analysis of any common patterns present among the protein sequences of Interleukins or not.

- Results

The aligned results of multiple sequence alignment depicts that the protein sequences of different interleukins involved in neuroinflammation of AD are less similar. Predicting their different metabolic functions and responsibilities.

However, phylogenetic tree of protein sequences of different interleukin members represents only one remain constant from origin of interleukin member remain original however majorly other belong nearly related.

Fig-12.2 represents the aligned MSA which was visualized from Mega 11. ***Fig-13*** show phylogenetic tree retrieved from Clustal omega.

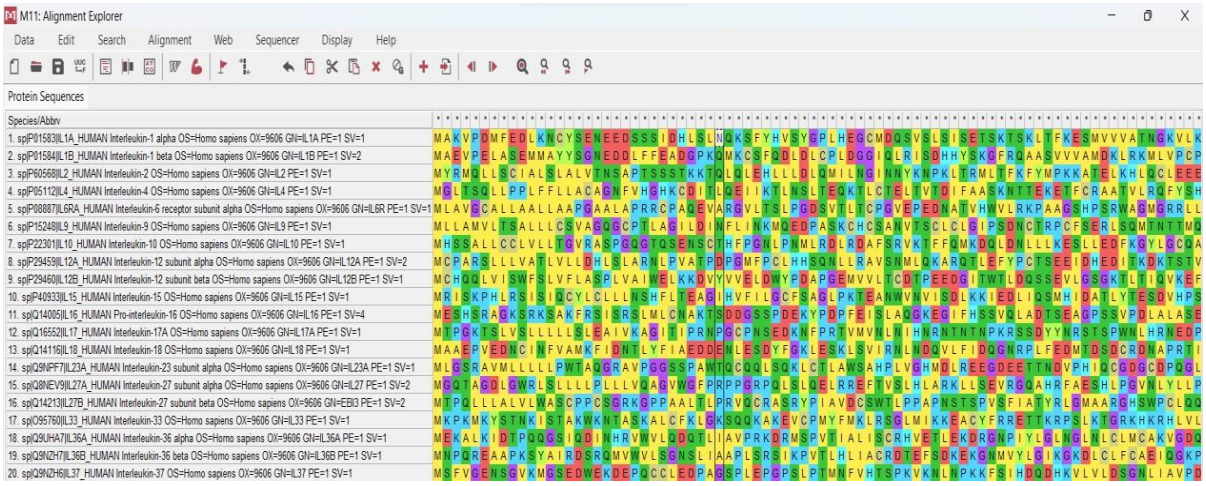


Fig 12.1: Mega11 protein sequence alignment of potential interleukin members in Alzheimer's Disease.

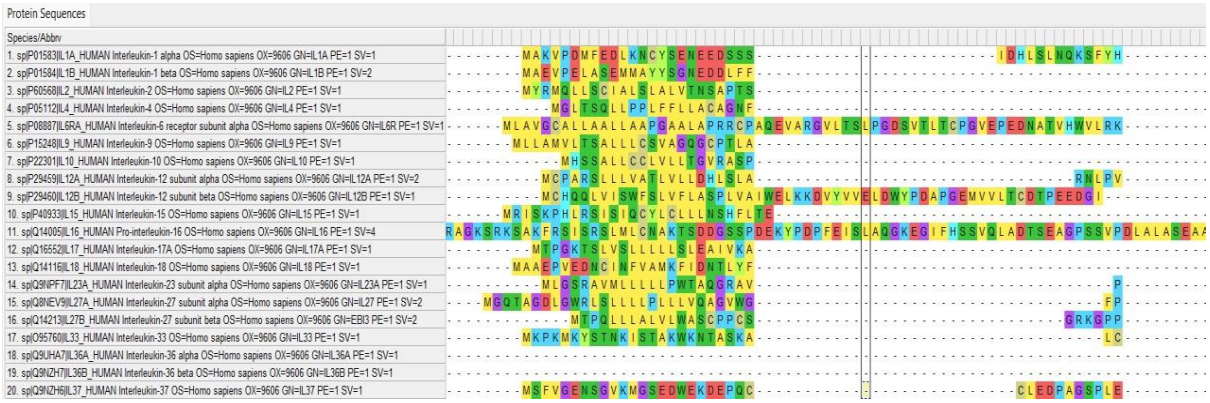


Fig 12.2: Mega11 protein sequence alignment of potential interleukin members in Alzheimer's Disease alignment aligned result.

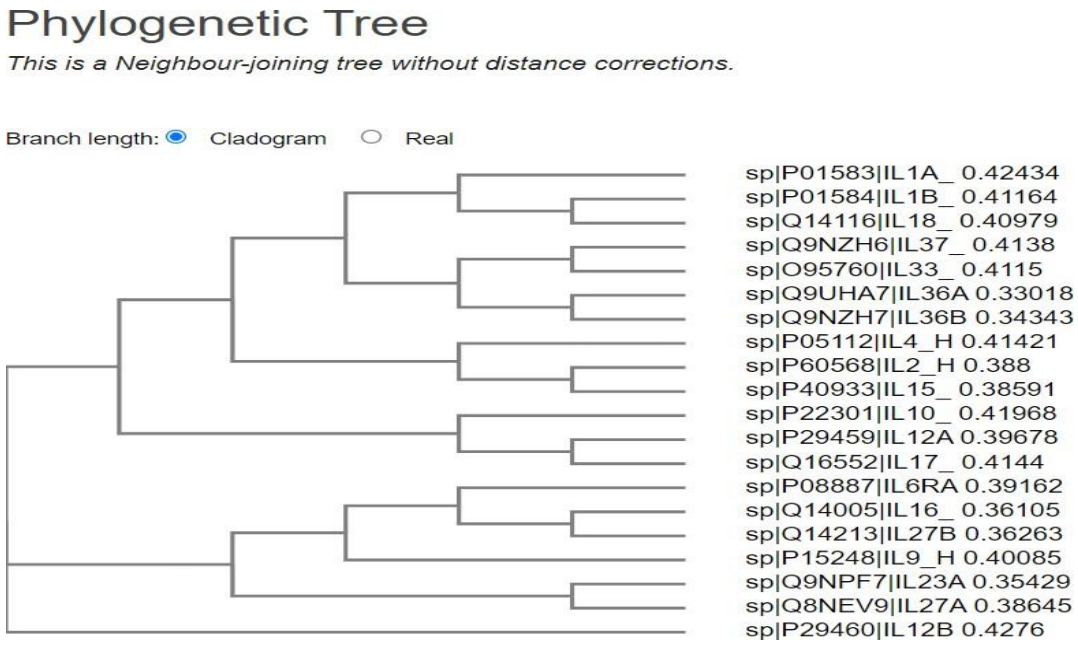


Fig 13: Phylogenetic tree representation of different Interleukins protein sequences.

4.4. Conservation Analysis of Interleukins

Conservation calculating tool is used to check for prediction of conserved nature of different interleukins sated in human. AACon calculation have been performed by using Jalview servers.

From the result we can conclude that there are few conserved regions, occupancy of sequence is high throughout the regions. Interleukins have less conserved nature.

This show that a lot of changes happened to interleukin sequences from initial time.

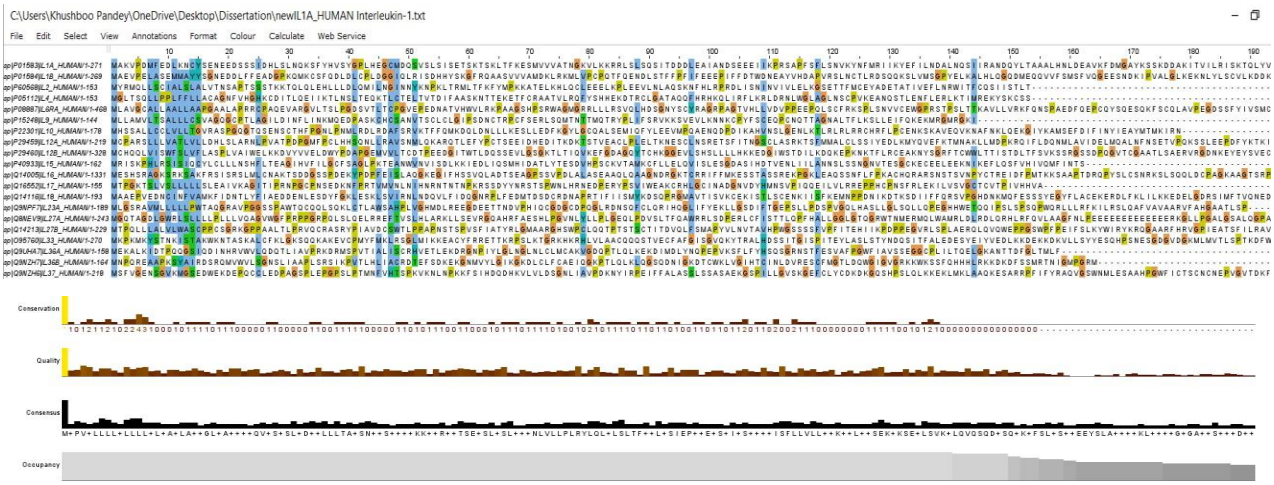


Fig 14.1: Protein sequences of different Interleukin members in Jalview server for amino acid prediction.

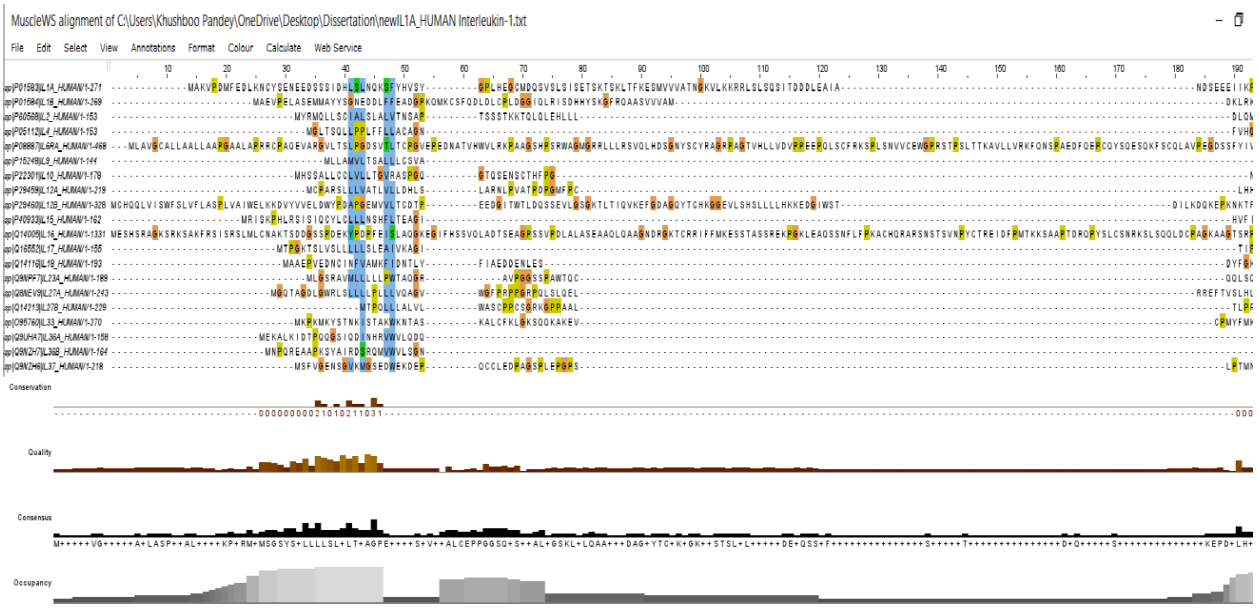


Fig 14.2: Protein sequences of different Interleukin members in Jalview server for amino acid prediction, aligned sequences.

In ***Fig 14.2*** we can visualize the quality, occupancy and conserved region in all over aligned interleukin protein sequences.

4.5. Protein-Protein Interaction:

Protein-protein network string colours symbolize different relations among the proteins.Yellow string connects actively present protein and interleukins in Alzheimer's Disease with repeatedly found from text mining. Black string shows the co-expression of proteins. Blue represents the gene co-occurrence among proteins. Their many known interactions are also formed which are from experimental and curated database. Predicted interactions help in determination of relation among the proteins and interleukins activity in Alzheimer's patient.

4.5.1. Interaction between interleukins and other proteins related to Alzheimer's Disease:

Interleukins interactions were created along with other proteins which are associated with Alzheimer's disease, using STRING software. String interactions shows co-occurrence of interleukins with other proteins among different experimental database.

Potentially participating interleukins with other protein in AD resulted from curated data of string analysis are IL-1β, IL-2, IL-4, IL-6, IL-10, IL-15, IL-16, 17α, IL-18, IL-22 & IL33.

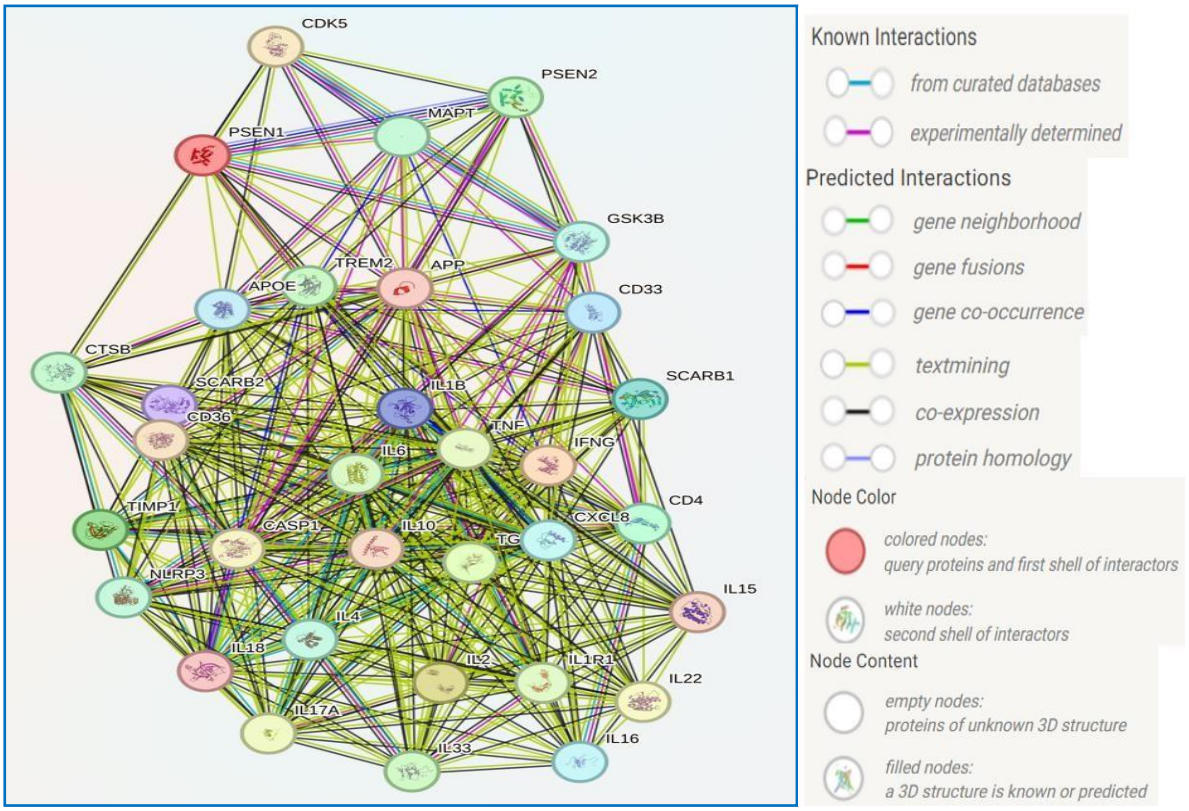


Fig 15.1: String connection explains the protein-protein interaction between interleukins with other proteins related to Alzheimer's Disease.

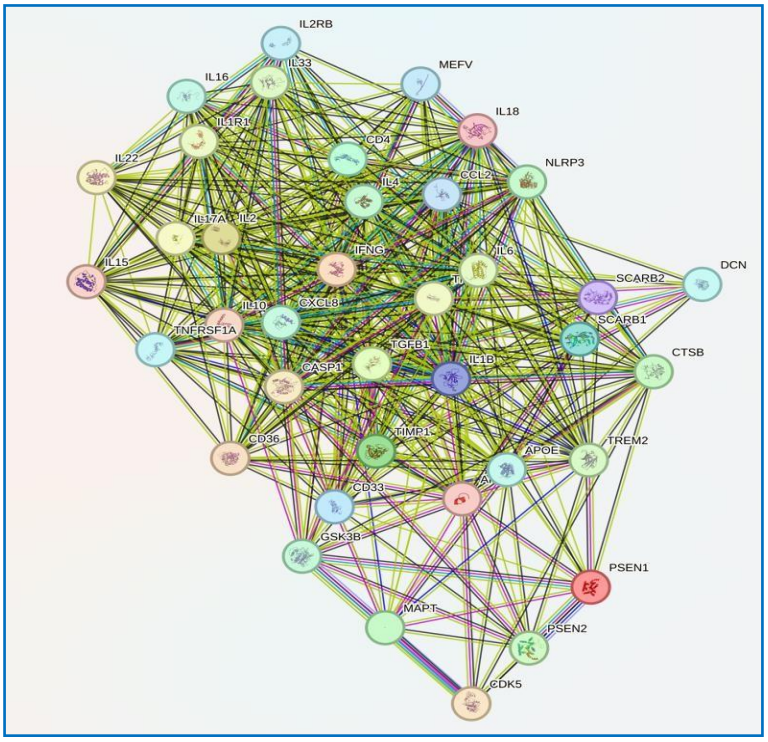


Fig 15.2: More dense protein-protein interaction between interleukins with other more proteins related to Alzheimer's Disease.

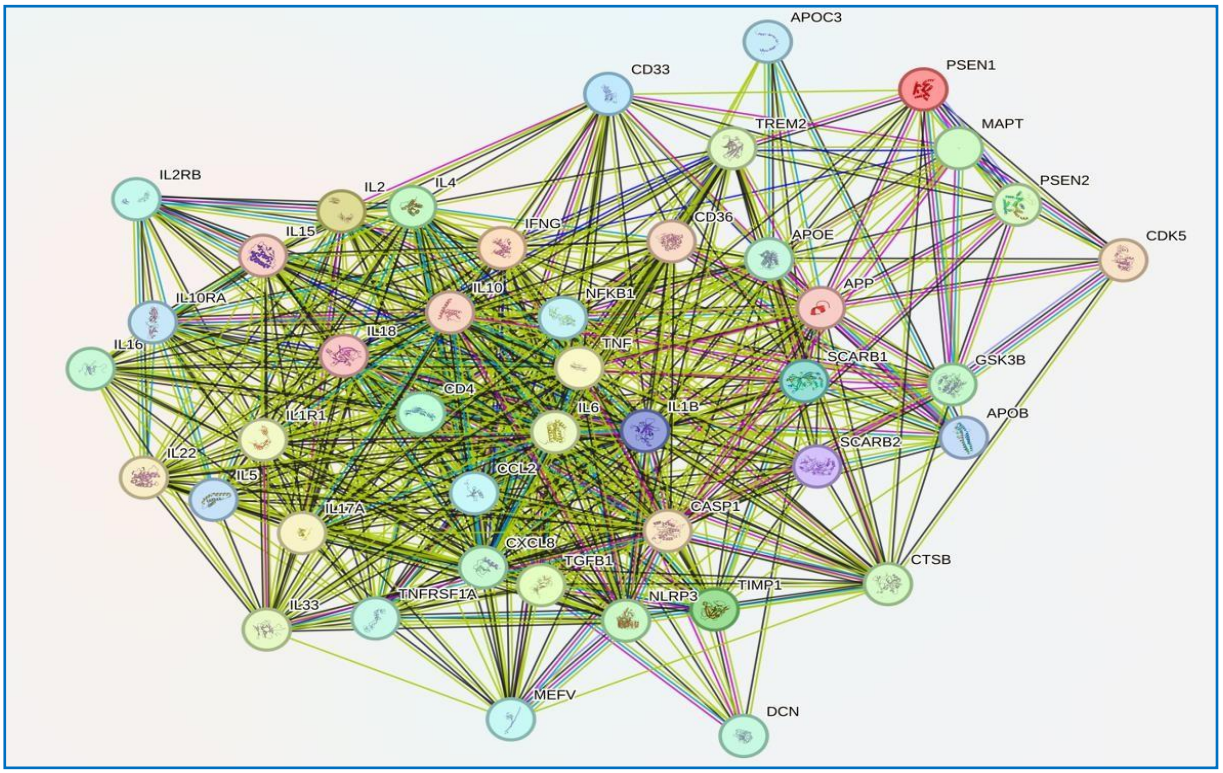


Fig 15.3: Further dense protein-protein interaction between interleukins with other more proteins related to Alzheimer's Disease.

More widely represented the other protein which effect closely and distantly participate in metabolic activity of Alzheimer's Disease, that are traced in different texts.

4.5.2. Interaction between interleukins related to Alzheimer's Disease

Interaction among interleukin family members related to Alzheimer's Disease, shows more co-expression less experimental data and predicted from large volume of text data base.Yellow string connects actively present among the interleukins potentially present in Alzheimer's Disease with repeatedly found from text mining, black string shows the co-expression behaviour of interleukins, association of biological properties and relationship between the proteins.

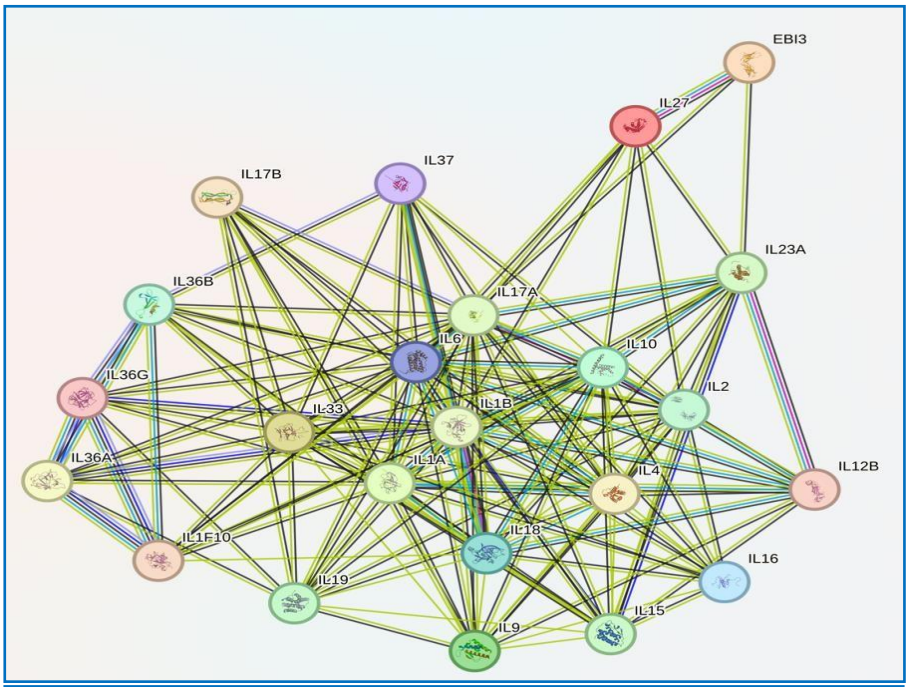


Fig 16.1: Interaction between interleukin family members which are related to Dementia.

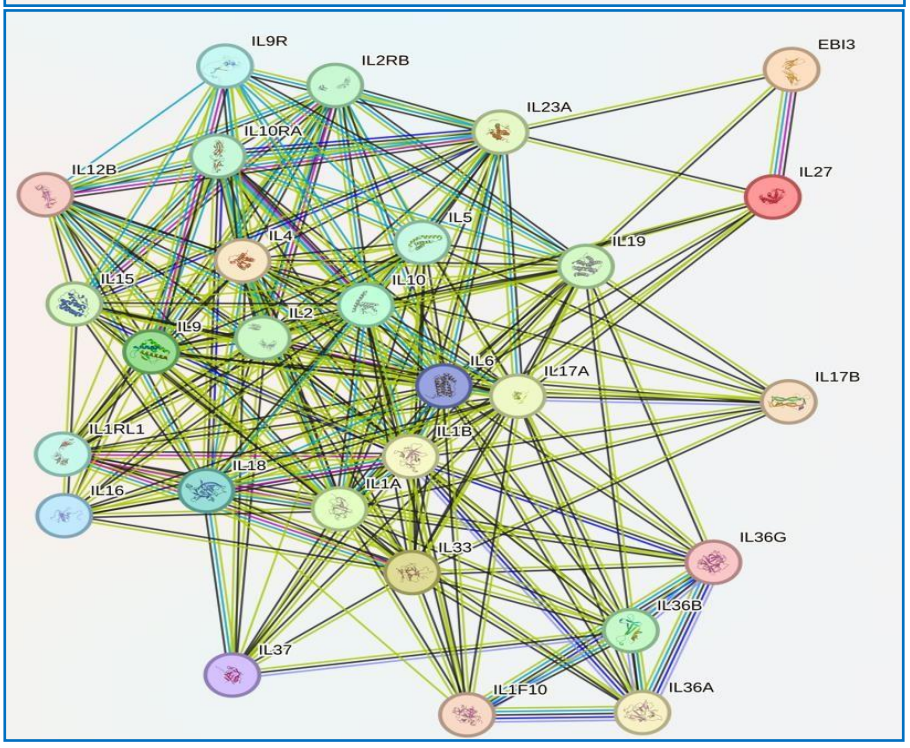


Fig 16.2: More interaction among interleukin family members which are related to Dementia.

4.6. SNPs Analysis:

From dbSNP database (NCBI's database) searched for **Interleukin-1β (IL1B)** resulted **2193 SNPs**, out of which non-synonymous SNPs were filtered. Listed **167 nsSNPs** out of 2193 SNPs. Identification of non-synonymous (missense) mutation occur in Interleukin-1β with the help of dbSNP data.

Sorting of intolerant from tolerant (SIFT) was performed primarily in order to predict which amino acid substitution causing affect in protein function.

4.6.1. Sequence Based Analysis of nsSNPs of Interleukin-1β:

Sorted the deleterious nsSNPs from the tolerant nsSNPs with the help of SIFT tool with the help of nsSNPs rsIDs. Deleterious nsSNPs are problem causing nsSNPs whereas tolerant are those which resist the change in protein function or are non-problematic in nature.

Table 6: Listed are deleterious nsSNPs resulted from SIFT analysis.

S.No.	SNP	ORGANISM/ BUILD	COORDINATE	REF ALLELE	ALT ALLELE	AMINO ACID CHANGE	GENE ID	TRANSCRIPT ID	PROTEIN ID	SIFT SCORE	SIFT MEDIAN	NO OF SEQS AT POSITION	SIFT PREDICTION
1	rs375479974	Homo_sapiens/ GRCh37.74	113588981	A	G	F162S	ENSG00000125538	ENST00000263341	ENSP00000263341	0.013	2.67	48	DELETERIOUS
2	rs373127037	Homo_sapiens/ GRCh37.74	113591132	G	T	D40E	ENSG00000125538	ENST00000418817	ENSP00000407219	0.01	3.12	27	DELETERIOUS
3	rs373127037	Homo_sapiens/ GRCh37.74	113591132	G	T	D40E	ENSG00000125538	ENST00000416750	ENSP00000400854	0.011	3.31	27	DELETERIOUS
4	rs373103547	Homo_sapiens/ GRCh37.74	113590391	A	G	F105S	ENSG00000125538	ENST00000416750	ENSP00000400854	0.037	3.29	27	DELETERIOUS
5	rs371339015	Homo_sapiens/ GRCh37.74	113590987	C	T	D89N	ENSG00000125538	ENST00000418817	ENSP00000407219	0.03	3.13	29	DELETERIOUS
6	rs371339015	Homo_sapiens/ GRCh37.74	113590987	C	T	D89N	ENSG00000125538	ENST00000416750	ENSP00000400854	0.033	3.32	29	DELETERIOUS
7	rs370988408	Homo_sapiens/ GRCh37.74	113593794	G	T	P5T	ENSG00000125538	ENST00000263341	ENSP00000263341	0	2.85	32	DELETERIOUS
8	rs370988408	Homo_sapiens/ GRCh37.74	113593794	G	T	P5T	ENSG00000125538	ENST00000416750	ENSP00000400854	0	3.33	24	DELETERIOUS
9	rs370988408	Homo_sapiens/ GRCh37.74	113593794	G	T	P5T	ENSG00000125538	ENST00000418817	ENSP00000407219	0	3.21	24	DELETERIOUS
10	rs369312177	Homo_sapiens/ GRCh37.74	113593775	A	G	M11T	ENSG00000125538	ENST00000418817	ENSP00000407219	0.022	3.21	24	DELETERIOUS
11	rs369312177	Homo_sapiens/ GRCh37.74	113593775	A	G	M11T	ENSG00000125538	ENST00000416750	ENSP00000400854	0.025	3.33	24	DELETERIOUS
12	rs201772036	Homo_sapiens/ GRCh37.74	113590374	C	T	E111K	ENSG00000125538	ENST00000418817	ENSP00000407219	0.029	3.13	29	DELETERIOUS
13	rs201772036	Homo_sapiens/ GRCh37.74	113590374	C	T	E111K	ENSG00000125538	ENST00000416750	ENSP00000400854	0.039	3.4	17	DELETERIOUS
14	rs200401035	Homo_sapiens/ GRCh37.74	113590977	G	A	T92I	ENSG00000125538	ENST00000418817	ENSP00000407219	0.011	3.13	30	DELETERIOUS
15	rs200401035	Homo_sapiens/ GRCh37.74	113590977	G	A	T92I	ENSG00000125538	ENST00000416750	ENSP00000400854	0.012	3.3	30	DELETERIOUS
16	rs200223008	Homo_sapiens/ GRCh37.74	113593150	T	A	Q31L	ENSG00000125538	ENST00000418817	ENSP00000407219	0.028	3.12	26	DELETERIOUS
17	rs200223008	Homo_sapiens/ GRCh37.74	113593150	T	A	Q31L	ENSG00000125538	ENST00000416750	ENSP00000400854	0.04	3.3	26	DELETERIOUS
18	rs141525769	Homo_sapiens/ GRCh37.74	113590375	G	T	N110K	ENSG00000125538	ENST00000416750	ENSP00000400854	0.022	3.36	21	DELETERIOUS
19	rs376289593	Homo_sapiens/ GRCh37.74	113593779	C	T	E10K	ENSG00000125538	ENST00000416750	ENSP00000400854	0.009	3.33	24	DELETERIOUS
20	rs376289593	Homo_sapiens/ GRCh37.74	113593779	C	T	E10K	ENSG00000125538	ENST00000418817	ENSP00000407219	0.01	3.21	24	DELETERIOUS
21	rs376341819	Homo_sapiens/ GRCh37.74	113590998	A	C	F85C	ENSG00000125538	ENST00000416750	ENSP00000400854	0.002	3.3	30	DELETERIOUS
22	rs376341819	Homo_sapiens/ GRCh37.74	113590998	A	C	F85C	ENSG00000125538	ENST00000418817	ENSP00000407219	0.002	3.13	30	DELETERIOUS
23	rs376341819	Homo_sapiens/ GRCh37.74	113590998	A	C	F85C	ENSG00000125538	ENST00000263341	ENSP00000263341	0.004	2.65	45	DELETERIOUS

From SIFT table (***Table 6***) we get **23** deleterious nsSNPs as primarily data of interleukin-1β (*Homo sapiens,* found at chromosome 2) along with protein IDs, mutation position, SIFT scores and prediction. Retrieved the protein sequence using Protein ID from Ensembl databasefor further sequenced based tool analysis for nsSNPs.

Run sequenced based analysis through different tools for validation of most deleterious nsSNPs. Tools incorporated for sequenced-based analysis are SIFT, PANTHER, PolyPhen2, PhD SNPs, SNPs & GO and Predict SNP. Given below are table representing the nature of nsSNPs their possible effect on protein functioning. Using the sorted table result for identification most problematic mutation. IL-1B mutant (nsSNPs) found deleterious are **F162S, D40E, F105S, D89N, P5T, M11T, T92I, Q31L, N110K, E10K** and **F85C (11 mutants).**

Table 7.1: Listed are deleterious nsSNPs resulted from different sequenced-based tools (PolyPen-1, PolyPhen-2, SNPs&GO).

IL-1β		PolyPhen 2		PolyPhen 1	SNPs&GO	
nsSNPs	AA substitution	Score Range	Possible Effect	Possible Effect	Reliability Index	Possible Effect
rs375479974	F162S	1	Probably Damaging	Deleterious	9	Disease
rs373127037	D40E	0.908	Possibly Damaging	Neutral	3	Disease
rs373103547	F105S	0.002	Benign	Deleterious	7	Disease
rs371339015	D89N	0.981	Probably Damaging	Neutral	9	Disease
rs370988408	P5T	1	Probably Damaging	Deleterious	2	Disease
rs369312177	M11T	0.324	Benign	Neutral	3	Disease
rs201772036	E111K	0.001	Benign	Neutral	7	Disease
rs200401035	T92I	0.761	Probably Damaging	Neutral	4	Disease
rs200223008	Q31L	0.624	Possibly Damaging	Deleterious	5	Disease
rs141525769	N110K	0.089	Benign	Neutral	4	Disease
rs376289593	E10K	0.957	Probably Damaging	Deleterious	4	Disease
rs376341819	F85C	0.989	Probably Damaging	Deleterious	8	Disease

Table 7.2: Listed are deleterious nsSNPs resulted from different sequenced-based tools (PredictSNP, MAPP, PhD-SNP, SNAP, PANTHER).

IL-1β		PredictSNP	MAPP	PhD-SNP	SNAP	PANTHER
nsSNPs	AA substitution	Possible Effect	Possible Effect	Possible Effect	Possible Effect	Possible Effect
rs375479974	F162S	Deleterious	Deleterious	Deleterious	Deleterious	Probably Damaging
rs373127037	D40E	Deleterious	Deleterious	Neutral	Neutral	Probably Damaging
rs373103547	F105S	Deleterious	Deleterious	Deleterious	Deleterious	Probably Damaging
rs371339015	D89N	Deleterious	Neutral	Deleterious	Deleterious	Probably Damaging
rs370988408	P5T	Deleterious	Deleterious	Deleterious	Deleterious	Probably Damaging
rs369312177	M11T	Neutral	Deleterious	Neutral	Deleterious	Probably Damaging
rs201772036	E111K	Neutral	Deleterious	Deleterious	Neutral	Probably Damaging
rs200401035	T92I	Neutral	Neutral	Neutral	Neutral	Probably Damaging
rs200223008	Q31L	Deleterious	Deleterious	Neutral	Deleterious	Probably Damaging
rs141525769	N110K	Deleterious	Neutral	Neutral	Neutral	Probably Damaging
rs376289593	E10K	Deleterious	Deleterious	Neutral	Deleterious	Probably Damaging
rs376341819	F85C	Deleterious	Deleterious	Deleterious	Deleterious	Probably Damaging

Deleterious, probably and possibly damaging nsSNPs are amino acid substitution causing change in protein function. Benign and neutral are with low confidence score due to nearly no damage. Amino Acid substitution with high score range have more probability of causing damage. Out of 11 mutants 5 mutants are most deleterious in nature these are **F162S, F105S, F85C, E10K, P5T** and **D89N.**

4.6.2. Structure Based Analysis of nsSNPs of Interleukin-1β:

For the purpose of structure-based analysis of interleukin-1β nsSNPs, initially determined the protein structure determiner, which was 1HIB then download the .pdb file of identifier (1HIB) from PDB database to understand the position of amino acid in protein chain. 1HIB protein contains only A-chain and there is particular number for different amino acid in .pdb file of 1HIB (interleukin-1β) protein e.g. F162S mutation according to .pdb structure file will be F42S, as Phenylalanine was represented as 42 position in A-chain.

Tools incorporated for structure-based analysis are Structure based analysis: CUPSAT, MU-Pro, DynaMut, DUET, SDM, mCSM, SNPs & GO 3D and i- MUTANT.

Table 8.1: Listed are deleterious nsSNPs resulted from different structure-based tools (DynaMut, DUET, SDM and SNP&GO-3D SNAP).

IL1B(mutant)		DynaMut		DUET		SDM		SNP & GO-3D	
1HIB	A-chain	ΔΔG (kcal/mol)	Stability	ΔΔG (kcal/mol)	Stability	ΔΔG (kcal/mol)	Stability	Reliability Index	Prediction
F162S	F42S	-3.371	Destabilizing	-3.567	Destabilizing	-3.58	Destabilizing	8	Disease
D40E	D12E	0.148	Stabilizing	-1.019	Destabilizing	-0.81	Destabilizing	7	Disease
F105S	F42S	-3.371	Destabilizing	-3.567	Destabilizing	-3.58	Destabilizing	8	Disease
D89N	D12N	-0.697	Destabilizing	-1.197	Destabilizing	0.34	Stabilizing	7	Disease
P5T	P78T	0.143	Stabilizing	-0.914	Destabilizing	1.51	Stabilizing	7	Disease
M11T	M36T	0.734	Stabilizing	0.668	Stabilizing	-0.43	Destabilizing	7	Disease
E111K	E37K	0.767	Stabilizing	0.374	Stabilizing	-0.52	Destabilizing	6	Neutral
T92I	T124I	1.079	Stabilizing	-0.322	Destabilizing	1.1	Stabilizing	6	Disease
Q31L	Q126L	0.233	Stabilizing	0.611	Stabilizing	0.9	Stabilizing	0	Neutral
N110K	N129K	0.414	Stabilizing	0.02	Stabilizing	-0.72	Destabilizing	3	Neutral
E10K	E37K	0.767	Stabilizing	0.374	Stabilizing	-0.52	Destabilizing	6	Neutral
F85C	F42C	-1.457	Destabilizing	-2.464	Destabilizing	-1.33	Destabilizing	8	Disease

Table 8.2: Listed are deleterious nsSNPs resulted from different structure-based tools (Mcsm. I-Mutant 2.0, CUPSAT and MUPro).

IL1B(mutant)		mCSM		I-Mutant2.0		CUPSAT		MUPro	
1HIB	A-chain	ΔΔG (kcal/mol)	Stability	ΔΔG (kcal/mol)	Stability	ΔΔG (kcal/mol)	Stability	ΔΔG (kcal/mol)	Stability
F162S	F42S	-3.454	Destabilizing	-2.83	Destabilizing	-3.5	Destabilizing	-1.83768	Destabilizing
D40E	D12E	-1.091	Destabilizing	-1.22	Destabilizing	-4.3	Destabilizing	-0.99411	Destabilizing
F105S	F42S	-3.454	Destabilizing	-2.83	Destabilizing	-0.99	Destabilizing	-1.77789	Destabilizing
D89N	D12N	-1.476	Destabilizing	-2.06	Destabilizing	-0.42	Destabilizing	-1.38672	Destabilizing
P5T	P78T	-1.552	Destabilizing	-1.45	Destabilizing	0.16	Stabilizing	-0.2314	Destabilizing
M11T	M36T	0.276	Stabilizing	-0.85	Destabilizing	-1.12	Destabilizing	-1.58459	Destabilizing
E111K	E37K	0.139	Stabilizing	-1.24	Destabilizing	2.88	Stabilizing	-1.28077	Destabilizing
T92I	T124I	-0.814	Destabilizing	-0.87	Destabilizing	-2.3	Destabilizing	-0.47142	Destabilizing
Q31L	Q126L	0.165	Stabilizing	1.29	Stabilizing	-0.16	Destabilizing	0.25689	Stabilizing
N110K	N129K	-0.166	Destabilizing	-0.99	Destabilizing	0.92	Stabilizing	-1.01414	Destabilizing
E10K	E37K	0.139	Stabilizing	-1.24	Destabilizing	-1.02	Destabilizing	-1.32262	Destabilizing
F85C	F42C	-2.413	Destabilizing	-2.94	Destabilizing	-2.72	Destabilizing	-0.867	Destabilizing

nsSNP with higher negative ΔΔG value are more favourable reaction. Therefore, more damaging amino acid substitution to protein function. Stabilizing mutants shows neutral effect to protein function. For molecular dynamics simulation we use nsSNPs mutants which are

destabilizing in nature. Out of 11 mutants 6 mutants are most deleterious in nature these are **F42S, F42C, E10K, P78T, D12E, D12N** and **M36T.**

Table 9: Listed are most deleterious and destabilizing nsSNPs resulted from sequence and structure-based analysis F42S, F42C, E10K, P78T, D12E, D12N and M36T.

IL1B			
S.No.	**1HIB (mutation)**	**Stability Index**	**ΔΔG (kcal/mol)**
1.	F42S	Destabilizing	-3.477
2.	D12E	Destabilizing	-1.003
3.	D12N	Destabilizing	-1.03
4.	P78T	Destabilizing	-1.17
5.	M36T	Destabilizing	-0.99
6.	E37K	Stabilizing	0.381
7.	T124I	Stabilizing	1.1
8.	Q126L	Stabilizing	0.233
9.	N129K	Stabilizing	0.414
10.	E37K	Stabilizing	0.374
11.	F42C	Destabilizing	-2.464

After both the analysis and listing the destabilizing mutants we prepared the mutant variants of protein.

4.6.3. Preparation of mutants and wild-type protein structures:

To prepare the mutant structures, PyMol software is used. Initially original protein structure 1HIB is downloaded in pdb format and mutation is induced at particular position. For example: mutagenesis is prepared at the phenylalanine (original/ wild-type amino acid), position is altered with serine (mutation/ new type).

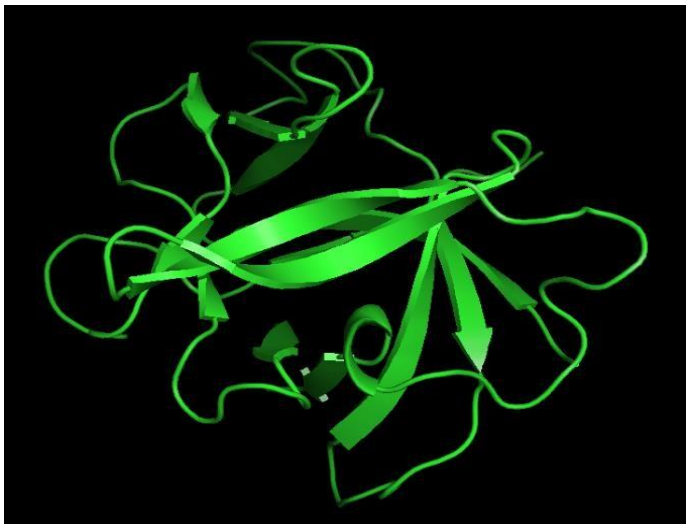


Fig 17: Wild type protein 1HIB (IL-1β) cartoon structure, visualized via PyMol.

Mutation caused at P78T position where proline is replaced by threonine.

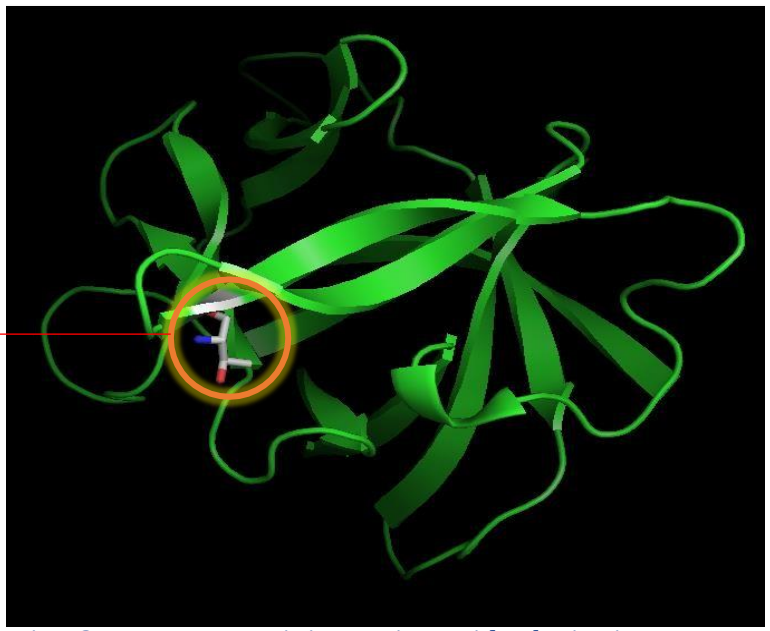


Fig 18: Mutant containing amino acid substitution, prepared using PyMol software.

Similarly all mutant type proteins (F42S, F42C, E10K, P78T, D12E, D12N and M36T) are prepared with new alteration so that these proteins can be used to do molecular dynamic simulation of protein in water to understant its activity and comparition between them. All the mutents and wild-tupe protein are saved in pdb format.

Table 10.1: List of mutant proteins (F42S, D12E and D12N).

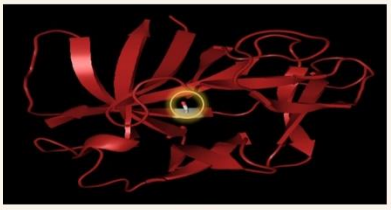
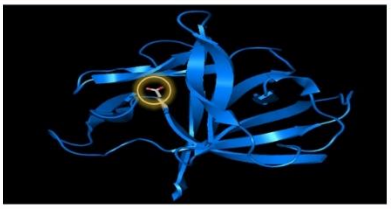
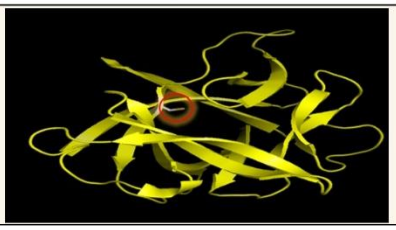
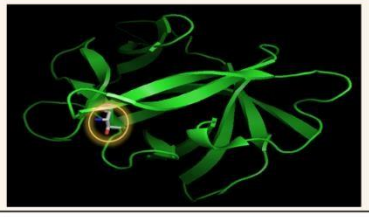
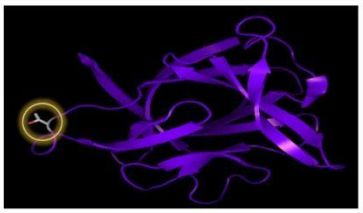
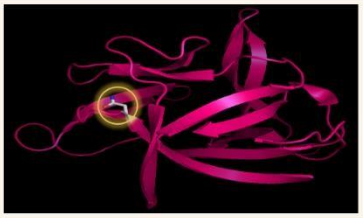
S. No.	Mutation Position	Original Amino Acid	New Amino Acid	Mutant Protein
1.	F42S	Phenylalanine (PHE)	Serine (SER)	
2.	D12E	Aspartic acid (ASP)	Glutamic acid (GLU)	
3.	D12N	Aspartic acid (ASP)	Asparagine (ASN)	

Table 10.2: List of mutant proteins (P78T, M36T and F42C).

S. No.	Mutation Position	Original Amino Acid	New Amino Acid	Mutant Protein
4.	P78T	Proline (PRO)	Threonine (THR)	
5.	M36T	Methionine (MET)	Threonine (THR)	
6.	F42C	Phenylalanine (PHE)	Cysteine (CYS)	

4.6.4. Molecular Dynamics Simulations:

Md simulation of wild-type and 6 mutant type proteins were conducted individually later compared and analyzed. Gromacs software is used to perform simulation and it is runed in Linux operating system. We check 1HIB protein activity in water along with ions to mimic the environment exist originally in biological system for the protein. With the aid of Gromacs software we generated protein topology, created cubic box with solvates and ions, proceed with energy minimization of structure and lastly produced the molecular dynamics of protein.

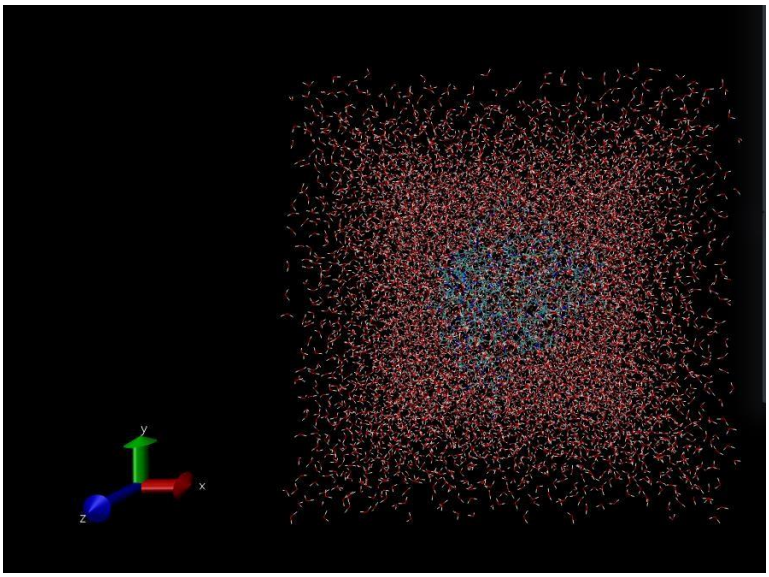


Fig 19: Wild type protein 1HIB (IL-1β) with in the cubic box, having solutes, water molecules and ions. Visualized using vmd computer program. Blue core region represents protein and red region are cubic box with different solutes and ions.

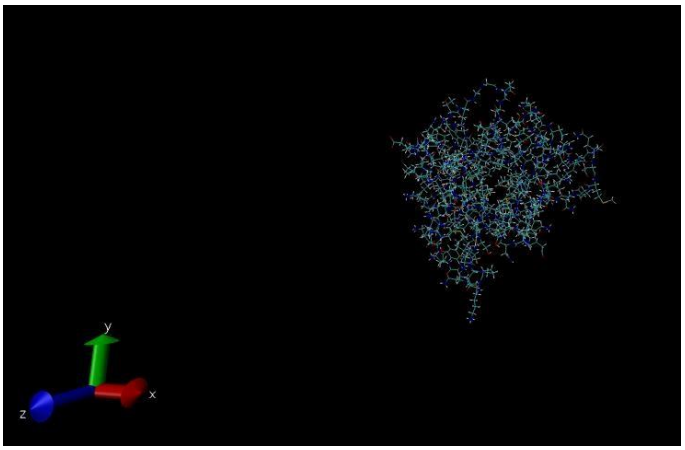


Fig 20.1: Blue lines represents protein, visualized using vmd computer program.

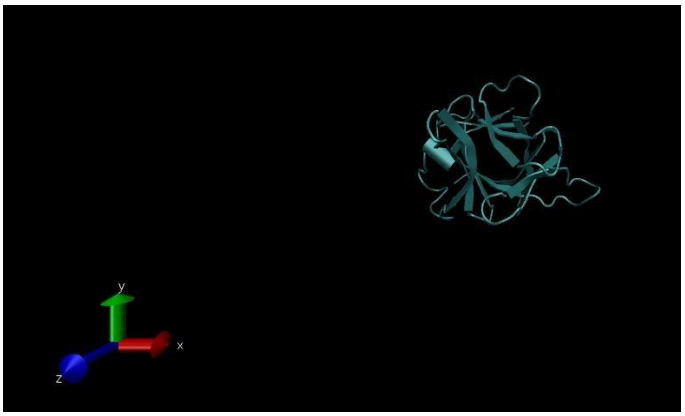


Fig 20.2: Blue ribbon like cartoon structure represents protein, visualized using vmd computer program.

Resulted graphs were comprised and visualized through QtGrace program to analyze data. Type of graphs obtained includes energy minimization graph, density graph, temperature graph, pressure graph, RMSD (root mean square deviation) and radius of gyration which measures compact structure of protein.

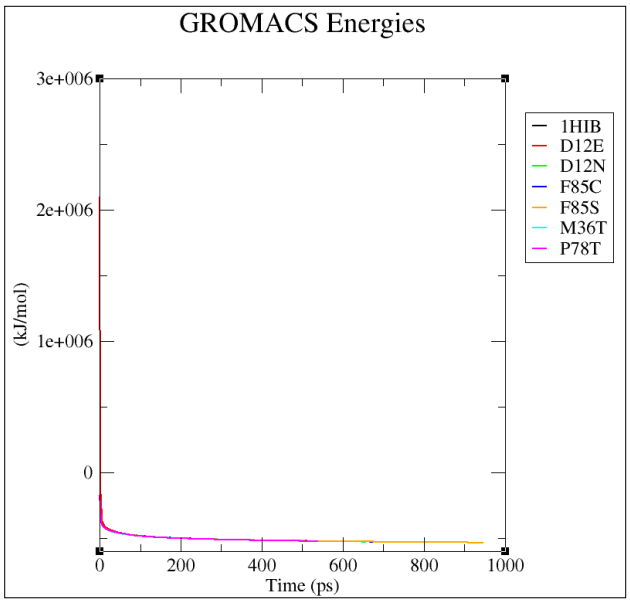


Fig 21: Potential Energy (kJ/mol) vs Time (ps) Energy minimization graph depicting wild-type and different mutants proteins following nearly same energies in the molecular system. Showing the stable system. At the right top there are different color idertifier for the protein molecules.

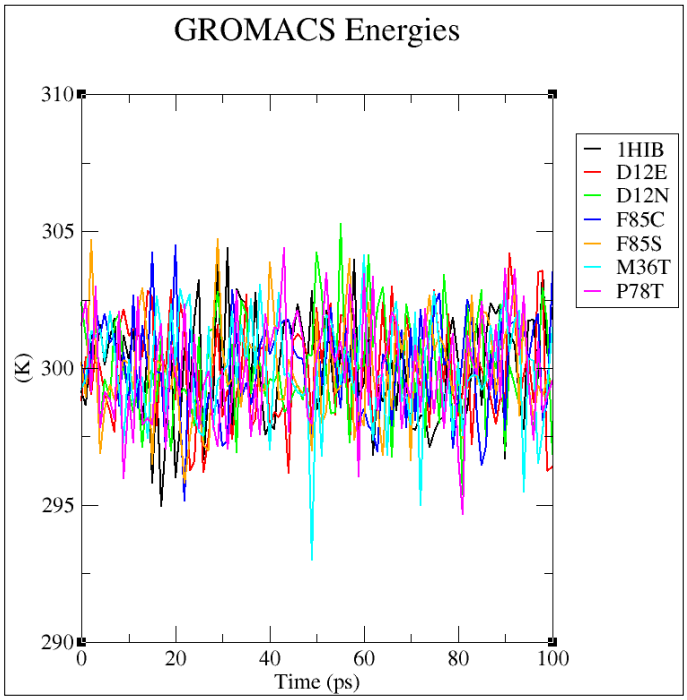


Fig 22: Temperature (K) vs Time (ps) graph depicting wild-type and different mutants proteins temperature showing their thermal behavior in the molecular system. Showing the stable system. At the right top there are different color idertifier for the protein molecules.

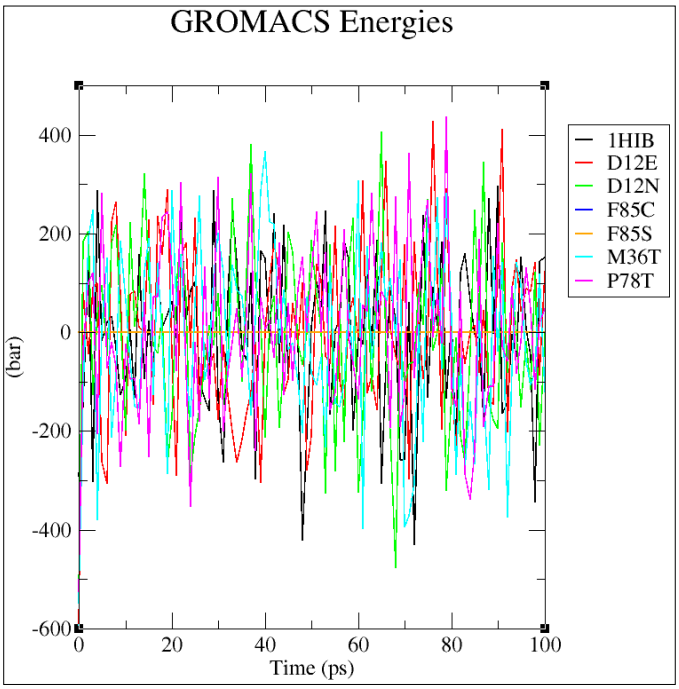


Fig 23: Pressure (bar) vs Time (ps) graph depicting wild-type and different mutants proteins showing for F85C and F85S pressure remain 0 (system is in equilibrium state) throughout the however for other proteins pressure remain changing. At the right top there are different color idertifier for the protein molecules.

All these Graph suggest the activity of original protein with its mutant variety protein and aid in understanding how it work or react in biological system. From the results we can illustrate the conclusion that these mutant proteins show its impact towards biological system and change its functioning and structural properties of protein and could be the reason of damage. Therefore, it can be suggested that IL-1B activity can affect or influence neuroinflammation in AD patient.

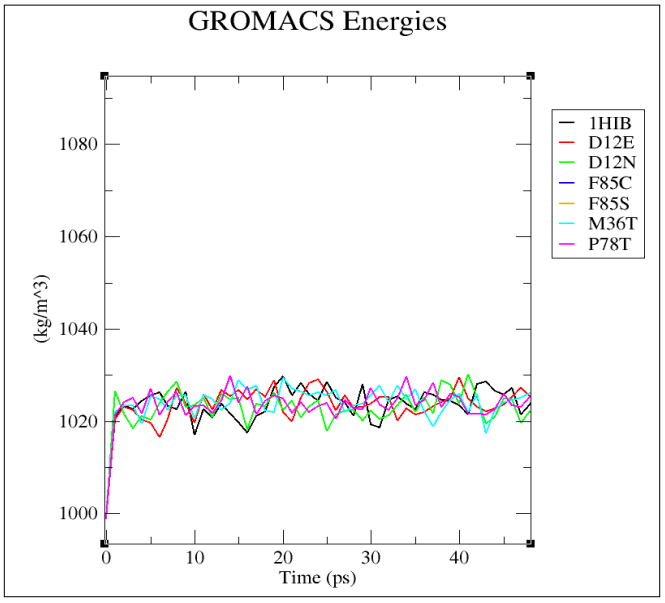


Fig 24: Density (kg/m^3) vs Time (ps) graph depicting wild-type and different mutants proteins showing their density keep changing through out thesimulation but remained below 1030 kg/m^3 and vary for each proteins. At the right top there are different color idertifier for the protein molecules.

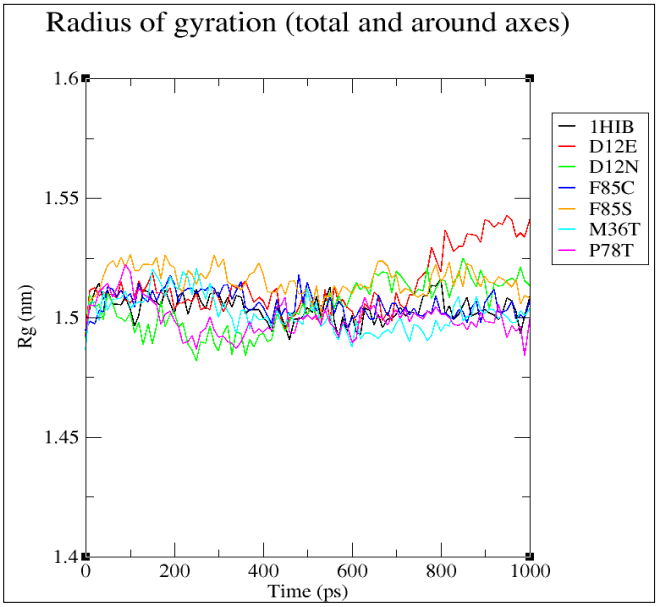


Fig 25: Radius Gyration (nm) vs Time (ps) graph depicting wild-type and different mutants proteins showing their compact molecular structure of proteins, dynamic changes can be observed from graph. Protein with mutation D12E show major diflection with time. At the right top there are different color idertifier for the protein molecules.

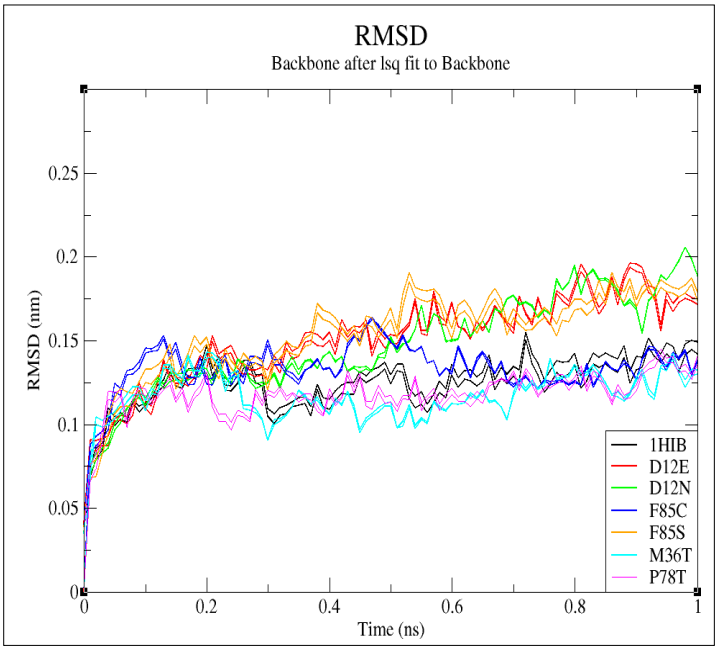


Fig 26: RMSD (nm) vs Time (ns) graph depicting wild-type and different mutants proteins structure deviations through out the simulations. It can be analyzed from graph that mutant protein (D12E, D12N & F85S) shows more deviation from original protein (1HIB). At the right bottom there are different color idertifier for the protein molecules.

4.7. Evolutionary Conservation of Interleukins (protein):

Evolutionary conservation was analysed to understand the nature of sequence and for identification of region which remain conserved throughout the evolutionary generations. Using ConSurf database obtained and visualize the interleukin 1β (1HIB).

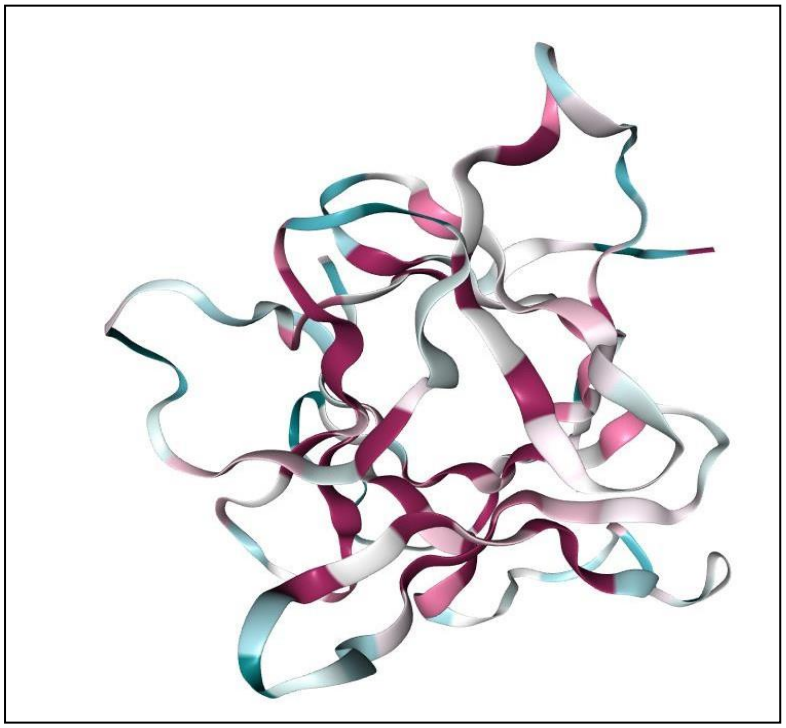
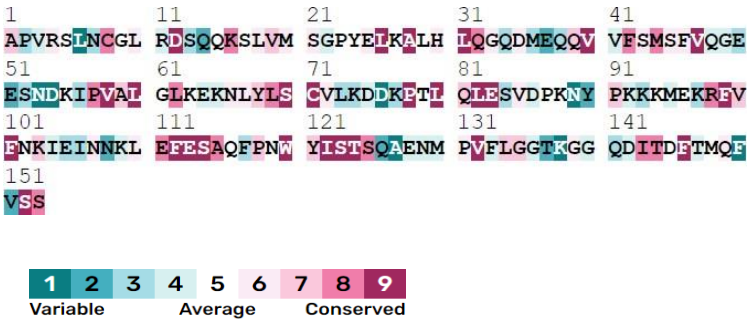


Fig 27: Structure of interleukin -1β (1HIB), where blue region shows the variable part and reddish region show the conserved part of the protein structure.

Below the figure there is protein sequence and colour scale for prediction of conserved region range probability.



Result illustrated that the major sequence part remains conserved throughout the evolution. However, there are some variable regions also lies in the protein sequence could be the happened due to change or mutation in nucleotide sequence.

CHAPTER 5

CONCLUSION

Potentially active interleukins participating in metabolism of Alzheimer's Disease and causing neuroinflammation, major symptom of cognitive disorder. Interleukin members with which we are working with IL-1α, IL-1β, IL-2, IL-4, IL-6, IL-9, IL-10, IL-12, IL-15, IL-16, IL-17, IL-18, IL-23, IL-27, IL-33, IL-35, IL-36 & IL-37.

The resultant parameters analyzed via computational investigation of different Interleukins member involved in neuroinflammation, linked with Alzheimer's Disease suggest according to codon analysis conclude that the most potent targets were those having e-CAI value near or greater to 0.800, as this value is more significant. Under the influence of e-CAI value we can conclude that IL-1β, IL-1α, IL-10, IL-17β & IL-27 are determination to be involved in cellular functioning and gene expressions.

However, different members belonging to same Interleukin Family shows huge diversity when compared. Multiple sequence alignment result and conservation analysis of amino acid depicted less similarity, nearly less similar and minimal conservation index. Phylogenetic tree of protein sequences of different interleukin members represents only one remain constant from origin of interleukin member remain original, majorly other belong nearly related but for nucleotide sequences few interleukins remain same throughout generations.

From the conservation analysis it can be conclude interleukin member have less conserved nature in human, which suggest that lot of changes have happened from generations. Protein-protein interactions results depict the involvement and reference studies showing the participation of interleukins in major activity of AD metabolism.

Further on the SNPs analysis help in understanding the cause of alteration in sequence amino acid leading to single mutation can affect the protein functioning, parallelly it be the reason of damage or disease factor element. In addition to this molecular dynamic simulation of interleukin-1β (1HIB) protein along with its top hit 6 mutation (F42S, F42C, E10K, P78T, D12E, D12N and M36T) validate the molecular effect in biological environment will affect the cellular functioning.

Apart from conducted analysis our aim to study potential role of major working interleukins in a wider prospect for which functional and evolutionary based analysis needed to conduct along with pathway designing. Parallelly, multiple analysis will be conducted among other species for identification of similarity of target molecule for better understanding.

It can be concluded from results that interleukin-1β could be one of the majorly active and affecting interleukin member toward neuroinflammation in Alzheimer's Disease. Before any conclusion there is a need to analyze every potent interleukin member in AD for better understanding and management of Alzheimer's Disease.

REFERENCES

[1] "What is Alzheimer's disease?," *National Institute on Aging*. [Online]. Available: https://www.nia.nih.gov/health/what-alzheimers-disease.

[2] "Alzheimer's disease," Nat. Rev. Dis. Primers, vol. 1, no. 1, pp. 1–2, 2015.

[3] R. W. Paterson *et al.*, "SILK studies — capturing the turnover of proteins linked to neurodegenerative diseases," *Nat. Rev. Neurol.*, vol. 15, no. 7, pp. 419–427, 2019.

[4] K. R. Brunden, J. Q. Trojanowski, and V. M.-Y. Lee, "Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies," *Nat. Rev. Drug Discov.*, vol. 8, no. 10, pp. 783–793, 2009

[5] S. Takeda, "Tau propagation as a diagnostic and therapeutic target for dementia: Potentials and unanswered questions," *Front. Neurosci.*, vol. 13, 2019.

[6] S. Swarbrick, N. Wragg, S. Ghosh, and A. Stolzing, "Systematic review of miRNA as biomarkers in Alzheimer's disease," Mol. Neurobiol., vol. 56, no. 9, pp. 6156–6167, 2019.

[7] A. Sobue, O. Komine, and K. Yamanaka, "Neuroinflammation in Alzheimer's disease: microglial signature and their relevance to disease," Inflamm. Regen., vol. 43, no. 1, 2023.

[8] S. Thakur, R. Dhapola, P. Sarma, B. Medhi, and D. H. Reddy, "Neuroinflammation in Alzheimer's disease: Current progress in molecular signaling andtherapeutics," *Inflammation*, vol. 46, no. 1, pp. 1–17, 2023.

[9] S.-B. Choi, S. Kwon, J.-H. Kim, N.-H. Ahn, J.-H. Lee, and S.-H. Yang, "The molecular mechanisms of neuroinflammation in Alzheimer's disease, the consequence of neural cell death," *Int. J. Mol. Sci.*, vol. 24, no. 14, p. 11757, 2023.

[10] Y. Liu *et al.*, "Targeting neuroinflammation in Alzheimer's disease: from mechanisms to clinical applications," *Neural Regen. Res.*, vol. 18, no. 4, p. 708, 2023.

[11] M. Colonna and O. Butovsky, "Microglia function in the central nervous system during health and neurodegeneration," *Annu. Rev. Immunol.*, vol. 35, no. 1, pp. 441–468, 2017.

[12] M. T. Heneka et al., "Neuroinflammation in Alzheimer's disease," Lancet Neurol., vol. 14, no. 4, pp. 388–405, 2015.

[13] A. W. Khan, M. Farooq, M.-J. Hwang, M. Haseeb, and S. Choi, "Autoimmune neuroinflammatory diseases: Role of interleukins," *Int. J. Mol. Sci.*, vol. 24, no. 9, p. 7960, 2023.

[14] G. Donninelli *et al.*, "Interleukin-9 regulates macrophage activation in the progressive multiple sclerosis brain," *J. Neuroinflammation*, vol. 17, no. 1, 2020.

[15] E. Lin, P.-H. Kuo, Y.-L. Liu, A. C. Yang, and S.-J. Tsai, "Association and interaction effects of interleukin-12 related genes and physical activity on cognitive aging in old adults in the Taiwanese population," *Front. Neurol.*, vol. 10, 2019.

[16] A. S. Mendiola and A. E. Cardona, "The IL-1β phenomena in neuroinflammatory diseases," *J. Neural Transm. (Vienna)*, vol. 125, no. 5, pp. 781–795, 2018.

[17] N. M. Lyra e Silva *et al.*, "Pro-inflammatory interleukin-6 signaling links cognitive impairments and peripheral metabolic alterations in Alzheimer's disease," *Transl. Psychiatry*, vol. 11, no. 1, pp. 1–15, 2021.

[18] R. J. Bishnoi, R. F. Palmer, and D. R. Royall, "Serum interleukin (IL)-15 as a biomarker of Alzheimer's disease," *PLoS One*, vol. 10, no. 2, p. e0117282, 2015.

[19] X. Rao *et al.*, "Dual roles of interleukin-33 in cognitive function by regulating central nervous system inflammation," *J. Transl. Med.*, vol. 20, no. 1, 2022.

[20] S. T. Parvathy, V. Udayasuriyan, and V. Bhadana, "Codon usage bias," *Mol. Biol. Rep.*, vol. 49, no. 1, pp. 539–565, 2022.

[21] S. Bahiri-Elitzur and T. Tuller, "Codon-based indices for modeling gene expression and transcript evolution," *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 2646–2663, 2021.

[22] P. Puigbò, I. G. Bravo, and S. Garcia-Vallvé, "E-CAI: a novel server to estimate an expected value of Codon Adaptation Index (eCAI)," *BMC Bioinformatics*, vol. 9, no. 1, p. 65, 2008.

[23] D. Szklarczyk et al., "STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimentaldatasets," Nucleic Acids Res., vol. 47, no. D1, pp. D607–D613, 2019.

[24] A. K. Yadav and T. R. Singh, "Novel structural and functional impact of damaging single nucleotide polymorphisms (SNPs) on human SMYD2 protein using computational approaches," Meta Gene, vol. 28, no. 100871, p. 100871, 2021.

[25] R. Kumar et al., "In silico screening of deleterious single nucleotide polymorphisms (SNPs) and molecular dynamics simulation of disease associated mutations in gene responsible for oculocutaneous albinism type 6 (OCA 6) disorder," J. Biomol. Struct. Dyn., vol. 37, no. 13, pp. 3513–3523, 2019.

[26] A. Shukla and T. R. Singh, "Network-based approach to understand dynamic behaviour of Wnt signaling pathway regulatory elements in colorectal cancer," Netw. Model. Anal. Health Inform. Bioinform., vol. 7, no. 1, 2018.

[27] A. Girish et al., "Comprehensive bioinformatics analysis of structural and functional consequences of deleterious missense mutations in the human QDPR gene," J. Biomol. Struct. Dyn., pp. 1–17, 2023.

[28] R. Shukla, A. Kumar, D. J. Kelvin, and T. R. Singh, "Disruption of DYRK1A-induced hyperphosphorylation of amyloid-beta and tau protein in Alzheimer's disease: An integrative molecular modeling approach," Front. Mol. Biosci., vol. 9, 2023.

[29] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free," J. Comput. Chem., vol. 26, no. 16, pp. 1701–1718, 2005.

[30] S. Pronk et al., "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit," Bioinformatics, vol. 29, no. 7, pp. 845–854, 2013.

[31] A. A. Al-Karmalawy et al., "Molecular docking and dynamics simulation revealed the potential inhibitory activity of ACEIs against SARS-CoV-2 targeting the hACE2 receptor," Front. Chem., vol. 9, 2021.

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 27/05/2024

Type of Document (Tick): PhD Thesis | M.Tech/M.Sc. Dissertation ✓ | B.Tech./B.Sc./BBA/Other

Name: KHUSBOO PANDEY Department: BT & BI Enrolment No 225111014

Contact No. 8808378343 E-mail. khushboopandeykp10@gmail.com

Name of the Supervisor: Prof. Dr. TIRATHA RAJ SINGH

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters):

COMPUTATIONAL INVESTIGATION ON INTERLEUKIN FAMILY FOR ITS POTENTIAL ROLE IN ALZHEIMER'S DISEASE

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

- Total No. of Pages = 54
- Total No. of Preliminary pages = 08
- Total No. of pages accommodate bibliography/references = 02

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at 03 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

27/05/2024

(Signature of Guide/Supervisor)

27/05/2024

Signature of HOD

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Abstract & Chapters Details	
27th/05/2024	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String	04%	Word Counts	7,192
			Character Counts	43,392
Report Generated on		Submission ID	Page counts	46
27th/05/2024		2389178249	File Size	11.77M

Checked by
Name & Signature

27.05.24

Librarian

LIBRARIAN
LEARNING RESOURCE CENTRE
Jaypee University of Information Technology
Waknaghat, Distt. Solan (Himachal Pradesh)
Pin Code - 173234

Please send your complete Thesis/Report in (PDF) & DOC (Word File) through your Supervisor/Guide at plagcheck.juit@gmail.com