ANALYSIS OF BIO-MOLECULES FOR SYSTEM LEVEL UNDERSTANDING OF ALZHEIMER'S DISEASE

By

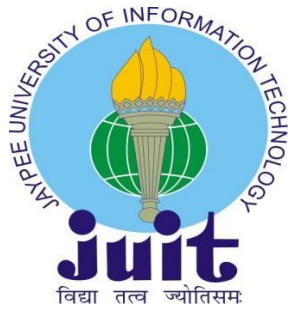
PRIYA PRADAYANI PANIGRAHI

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

BIOINFORMATICS



JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT

OCTOBER, 2014

DECLARATION

I certify that

a. The work contained in this thesis is original and has been done by me under the guidance of my supervisor.

b. The work has not been submitted to any other organisation for any degree or diploma.

c. Whenever, I have used materials (data, analysis, figures or text), I have given due credit by citing them in the text of the thesis.

Priya Panigrahi.

Priya Pradayani Panigrahi
(Email: bio4priya@gmail.com)

Date: 25/10/2014

CERTIFICATE

This is to certify that the thesis entitled, **"Analysis of bio-molecules for system level understanding of Alzheimer's disease"** which is being submitted by Priya Pradayani Panigrahi for the award of degree of Doctor of Philosophy in Bioinformatics by the Jaypee University of Information Technology at Waknaghat, is a record of the candidate's own work, carried out by her under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Tiratha Raj Singh
Assistant Professor (Senior Grade)
Department of BT and BI
Jaypee University of Information Technology
Waknaghat, Solan, H.P. India- 173234
Email: tiratharaj@gmail.com

Date: 25/10/2014

ACKNOWLEDGEMENTS

God gives blessings to us so we can give glory to him. I'm blessed and I thank God for everything in my life.

I deem it my privilege and honour to place and record my gratitude and indebtedness to the following without whose support, guidance, and concern I would not have been able to complete my Ph. D thesis.

I feel privileged to express my deep sense of reverence and gratitude to my revered mentor, ***Dr. Tiratha Raj Singh****, for his support, immaculate guidance, constructive criticism, constant encouragement and providing requisite facilities to carry on my research which otherwise would have remained incomplete. His nurturing and caring concern has been a stimulus which I will always cherish. I also appreciate his untiring efforts during the entire tenure of my research work and patience during writing of this thesis. I have no words to express my gratitude for everything he has contributed in my Ph. D and without his blessings it was surely impossible for me to finish my work. I thank him from bottom of my heart.*

I emphatically express my loyal and venerable thanks to ***Prof. Shiban Kishen Kak*** *(Vice Chancellor, JUIT),* ***Brig. (Retd.) Balbir Singh*** *(Director, JUIT),* ***Prof. T. S.Lamba*** *(Dean, JUIT) and* ***Dr. Y. Medury*** *(Ex-COO, Jaypee Education System) for providing opportunity to pursue a Doctorate Degree, teaching assistantship and advanced lab infrastructure to accomplish this scientific venture of my life.*

I gratefully acknowledge the help rendered by ***Prof. R. S. Chauhan*** *(Dean, BT and HOD, Dept. of BT& BI) for his encouragement, timely help and cooperation throughout my research work. It gives me immense pleasure to express my gratitude to* ***Chauhan Sir*** *for his ever smiling disposition coming to my rescue in solving my problems and suggestions which helped me in maintaining my confidence.*

I also feel indebted to my all time advisor, ***Mr. Sudipta Ku. Das*** *(Lect. in dept. of Zoology, P.S.College, gaisilat, Odissa),* ***Dr. Ragini Raj Singh*** *(Assist.Prof. Dept. of Physics and Material Science, JUIT), and* ***Dr. Anupriya Kaur*** *(Assist. Prof., Dept. of Humanities and Social Science, JUIT) for their valuable advice, unconditional support and guidance.*

This document would have remained an infant had it not received its necessary 'diet' in the form of comments, suggestions etc. by ***Dr. C. Rout, Dr. J. Ramana, Dr. Udayabanu M., Dr. Pradeep Kumar Naik, and Dr.Hemant Sood****. I am short of words in expressing my thanks to them, for their innovative ideas that shaped this document.*

I wish to convey my sincere thanks to all the faculty members of Department of

Biotechnology and Bioinformatics, for their help and guidance at the various stages of this study. I am also thankful to all the members of technical and non-technical staff of the department for their assistance and valuable contributions.

I am fortunate to have friends who have always stood beside me. I extend my heartfelt thanks to ***Jatin Sir, Varun Sir, Krishna Sir, Bharti Ma'am, Archit, Nikhil, Manika, Ashwani, Tarun, Ankita, Tamanna, Pawan, Dikshita and Deepika*** *for their sustained support and ever needed cooperation. I also thank* ***Asha, Pallavi, Nidhi, Dibyajyoti,*** *and* ***Yamini*** *for their inspiration throughout my life.*

It is my pleasure to express my gratitude to all research scholars of the Biotechnology & Bioinformatics Department for keeping me blessed with best wishes. It is worthwhile expressing my gratitude to ***Dr. Ahmed Moussa, Dr. MoacyrComar Junior,*** *and* ***Mr. Ramit Singla*** *who provided me assistance and valuable contributions.*

The successful culmination of this thesis could not have been possible without ***Dr. Kamal Raj Pardasani*** *(Professor, Department of Mathematics, Maulana Azad National Institute of Technology, Bhopal, India) and* ***Dr. Brigitte Vannier*** *(Assistant professor of cell biology, University of Poitiers, Poitiers, France). Their critical review and constructive and insightful comments were very helpful in strengthening this Thesis.*

The purpose of this acknowledgement will be incomplete if I fail to appreciate the moral support and encouragement rendered by my ***family*** *and my loving husband,* ***Preetam Biswal*** *(he is also one of my greatest advisor). I would like to express my love to my sweet nephew* ***"Aryaveer Nayak"*** *whose innocence refreshed me during my difficult times. The successful culmination of this thesis could not have been possible without the blessings of* ***my loving parents and Inlaws****. Without their help I would have been greatly incapacitated.*

All may not be mentioned, but no one is forgotten.

Priya Pradayani Panigrahi

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

Biology has currently become a 'big-data-science' principally supported by the advances in high-throughput experimental technologies. Newest advances in whole genome sequencing have changed the way biologists tackle problems in areas as diverse as biomedical research, plant biology, environmental microbiology etc [Callebaut, 2012]. Furthermore, molecular genetics has become integrated across the entire spectrum of model organisms because many cutting-edge techniques can be applied across the entire board of model organisms.Data-intensive science consists of three basic activities: capture, curation, and analysis. All three of these phases of managing huge data elevate many new research challenges to pursue in systems biology. Curation and analysis become imperative after capturing data from different experiments. Curation consists of storage, reclamation, filtering and integrating the data, spreading around the world.Scientists of the 21st century are rising up to the challenge of deciphering the workings of multifaceted processes that involve the interaction of numerous bio-molecules (genes, proteins etc.). The success of future investigators in this area will depend in huge part on broad, yet rigorous, training that accentuates the intertwine nature of biological systems, whether it is the amino acids and polypeptides in a protein complex, the gene products that make up a developmental pathway or signaling pathway. Jim Gray anticipated the fourth data paradigm and farming of the 'data deluge' explicitly, the capacity to measure, store, analyze and visualize data is the new actuality to which science must acclimatize. The center of the fourth paradigm is data and it sits alongside empiricism (1st paradigm), theory (2nd paradigm), and simulation (3rd paradigm), which collectively form the continuum we think of as the current scientific process [Bell *et al.*, 2009].

Systems biology is the study of systems comprising of biological components, which could be molecules, cells, organisms or whole species. Living systems are dynamic and multifaceted and their activities may possibly tough to envisage from the properties of individual components. We use quantitative measurements of the activities of groups of interacting components, systematic measurement technologies for example bioinformatics, genomics, proteomics, computational and mathematical models to depict and envisage dynamical behavior [Naylor and Chen, 2010].The purpose of systems biology is the system level understanding of a cell, organism, or disease, which can be recapitulated in the context of molecular networks as an understanding of the structure of all the components of a cell or

organism up to molecular level, the capability to predict the future condition of the cell or organism or disease under a normal environment, the capacity to predict the output responses for a given input stimulus, and the aptitude to estimate the changes in system behavior upon perturbation of the components or the environment. In an organism or cell the primary-level components, for instance, the molecules, are of numerous types and numbers, therefore system-level understanding is still a very complex task. Conversely to accomplish the theoretical target of systems biology, explicitly to understand life scientifically, many other useful applications will be invented. Useful applications include improvement of new generation medical tests, drugs, sensors, fuel, foods, materials etc. Presently Systems biology faces the challenges of analyzing huge amounts of molecular biological data and enormous biological networks. At present in systems biology some of the popularly-used data types are: Gene Expression, Sequences, Protein-Protein Interaction (PPI), Molecular Structure, Binding Sites and Domains, Metabolic Pathways etc.Complex diseases for instance cancer, Alzheimer's disease (AD), heart diseases, and mental disorder are very complicated and caused by manifold molecular irregularities.The drug discovery procedure of these complex diseases needs to target intact molecular pathways of diverse cellular omics networks rather than a single molecule. Recently biological networks, for instance PPI networks and gene expression networks are extensively used to discover drug targets [Lee *et al.*, 2012].

The progression in molecular biological experiments is generating huge piles of data interrelated to genome and RNA sequence, metabolite abundance and protein, gene expression, PPI, and so on. It is imperative to handle these huge amounts of data proficiently and scientifically to understand the system and to develop novel approaches in bioinformatics and biomedical fields. This, in turn, necessitates the handling of high speed computers and integrating understanding from other branches of science, such as statistics, physics, chemistry mathematics etc.The data we need to handle is of old formats, however the current challenge is that it has grown very big and needs the integration of diverse data types. This can be done by developing competent scaling techniques for the present software tools and statistical and mathematical models for data handling. The significance of network theory and algorithms can assist analyzing and integrating huge data [Altaf-Ul-Amin *et al.*, 2014].This research work has been planned by keeping in view of current scenario of AD research. All objectives have been designed according to the state-of-the-art requirements of AD study.

Following sections will describe about various aspects of dementias and AD along with associated features followed by major objectives and thesis plan.

1.2 DEMENTIA

Dementia is not a particular disease; it is a general term that describes an ample range of symptoms. In other words, it is a multi-faceted cognitive demolition that is typically progressive, and constantly involves functional impairments [Ripich and Horner, 2004]. With the ageing of the population and the estimated augment of dementia in the upcoming years, it is crucial that we understand the needs of people with dementia in order to furnish proper care [Cadieux *et al.,* 2013]. Persons affected by dementia are incapable to engage in daily activities with the similar level of independence as they had enjoyed previously in their life. While symptoms of dementia can vary significantly, at least two of the following core mental functions must be extensively impaired to be considered dementia:

- Memory
- Communication and language
- Ability to focus and pay attention
- Reasoning and judgment
- Visual perception

Dementia is often incorrectly referred to as ‘senility’ or ‘senile dementia’, which reflects the previously prevalent but inaccurate conviction that severe mental decline is a part of normal ageing (AG). People with dementia may have troubles with many general every day activities such as short-term memory, paying bills, keeping track of a purse or wallet, planning and preparing meals, remembering appointments etc. [MD Guidelines, 2009].

1.3DIAGNOSIS OF DEMENTIA

There has been a foremostswing in how scientists, neurologists, neuroradiologists, neuropsychologists, and other health professionals think about the dementias due to major scientific advances during 1990 to 1999 which is well known as the ‘Decade of the Brain’. These advances have affected diagnostic terms, pharmacologic options, and behavioral interventions. There are two categories of dementia that are; ‘presenile dementia’ (before age

65 years) and 'senile dementia' (after age 65 years) [Nakamura, 1990]. Neurologists are the medical specialists with proficiency in performing and interpreting clinical tests that rule in or rule out diverse probable causes. In recent years, several different working groups comprising panels of international experts have developed consensus statements regarding differential diagnosis. By testing patients at periodic intervals, the diagnosis is determined by clinical examination to be 'possible', 'probable' or 'definite'. Sadly, the gold standard for making a 'definite' diagnosis is only autopsy, which is really hazardous for brain [Lin *et al.*, 2013].

1.4 CAUSES OF DEMENTIA

Dementia is caused by damage to brain cells. This injure interferes with the aptitude of brain cells to communicate with each other. When brain cells cannot communicate normally, thinking, behavior, and feelings can be affected [Collin *et al.*, 2009]. The brain has many discrete regions, each of which is responsible for diverse functions (for example, memory, judgment and movement). When cells in a particular region are damaged, that region cannot carry out its functions normally. Different types of dementia are associated with particular types of brain cell damage in particular regions of the brain. For example, in AD, high levels of certain proteins inside and outside brain cells make it hard for brain cells to stay healthy and to communicate with each other. The brain region called the hippocampus is the center of learning and memory in the brain, and the brain cells in this region are often the first to be damaged. That's why memory loss is often one of the earliest symptoms of AD [Liang *et al.*, 2008].

While most changes in the brain that cause dementia are permanent and worsen over time, thinking and memory problems caused by the following conditions may improve when the condition is treated or addressed:

- Depression
- Medication side effects
- Excess use of alcohol
- Thyroid problems
- Vitamin deficiencies

1.5 TYPES OF DEMENTIA AND THEIR CHARACTERISTICS

There are many types of dementia; some are reversible, but most are irreversible. The neurodegenerative dementias are progressive and irreversible due to deterioration of brain cells and their interconnections. Depending on which part of the brain is suspected as the cause of dementia, the dementia is alienated into two broad categories that are cortical and sub-cortical dementias [Huber*et al.*, 1986].

1.5.1 Cortical dementias

Cortical dementia come up from a disorder affecting the cerebral cortex, the outer layers of the brain that play a critical role in thinking abilities like memory and language. AD and Creutzfeldt-Jakob disease are two forms of cortical dementia. People with cortical dementia typically show severe memory loss and aphasia: the inability to recall words and understand language [Whitehouse, 1986].

1.5.2 Sub-cortical dementias

Sub-cortical dementia occurs from dysfunction in the parts of the brain that are beneath the cortex. Usually, the forgetfulness and language difficulties that are characteristic of cortical dementias are not present. Rather, people with sub-cortical dementias, such as Parkinson's disease, Huntington's disease, and AIDS (Aquired Immuno Deficiency Syndrome) dementia complex, tend to show changes in their speed of thinking and ability to initiate activities [Cummings and Benson, 1984].There are cases of dementia where both parts of the brain have a tendency to be affected, for instance multi-infarct dementia or vascular dementia [Battistin and Cagnin, 2010].

1.6 HISTORY OF ALZHEIMER'S DISEASE

Progressive mental corrosion in aged people has been familiar anddemarcatedthroughout history. Conversely, in the year 1906 a German physician, Dr. Alois Alzheimer, distinctively identified an assortment of brain cell abnormalities as a disease. One of his patients died after years of rigorous memory problems, and confusion etc. The doctor observed eminent opaque deposits of neuritic plaques surrounding the nerve cells after her death, while performing a brain autopsy. He also observed twisted bands of fibers called NFTs (neurofibrillary tangles)

inside the nerve cells. This degenerative brain disorder bears his name, andthese plaques and tangles indicate a distinct diagnosis of AD [Wilkins and Brody, 1969].In the 1960s, scientists revealed an association among cognitive decline and the amount of plaques and tangles in the brain. Since thenAD is formally recognized as a disease by the medical community and not a normal part of AG [Berchtold and Cotman, 1998]. In the 1970s, researchers made enormous strides in understanding the human body as in one piece, and AD emerged as a substantial area of research attention.This amplified interest led to imperative discoveries as well as an improved perceptive of intricate nerve cells in the AD brains in the 1990s. Advance research was done on AD propensity genes; also a number of drugs were permitted to treat the symptoms of the disease [Derouesne, 2008].Over the last decade, scientists have significantly progressed in understanding prospective environmental, genetic, food and other risk factors for AD. Scientists have also revealed the processes most important to construction of plaques and tangles in the brain regions that are pretentious. Significant genes associated with both types of AD (early-onset (EOAD) and late-onset (LOAD) forms of AD), have been identified. However genetic risk factors single-handedly do not utterly clarify its causes. Therefore researchers are dynamically exploring environment and lifestyle to discover what function they might participate in the progress of this disease. New efficient treatment options have been approved by the Food and Drug Administration (FDA), but AD is still untreatable. The drugs presently in use can only treat the symptoms, not the cause of the disorder. Hence they can only deliberate the progression of cognitive decline [Osborn and Saunders, 2010].

1.7 ALZHEIMER'S DISEASE

The most familiar category of dementia amongst older people is AD, which primarily involves the parts of the brain that control thought, memory and language.AD continues to be one of the most complicated human diseases to treat. AD is a genetically intricate and heterogeneous disorder [Bertram *et al.*, 2007].It is an irreversible, progressive brain disease that gradually destroys memory and thinking skills. It is still not apparent whether AD is one disease with a particular foundation or manifold syndromes with familiar symptoms and/or a common pathology. Even though the possibility of developing AD increases with age; in the majority of people with AD, symptoms initially materialize after age of 60. AD is not a part of normal AG; it is a deadly disease that affects the brain [Scodellaro and Pin, 2013].Once

considered an exceptional disorder, it is now seen as a foremost public health trouble that is critically affecting millions of olders and their families. Younger's also get affected by AD but in very few percentages as compare to the older people. An expected 5.2 million Americans of all ages have AD in 2014. This comprises an expected 5 million people of age 65 and older and around 0.2 million individuals under age 65 who have EOAD (**Figure 1.1**) [Hebert *et al.*, 2013].

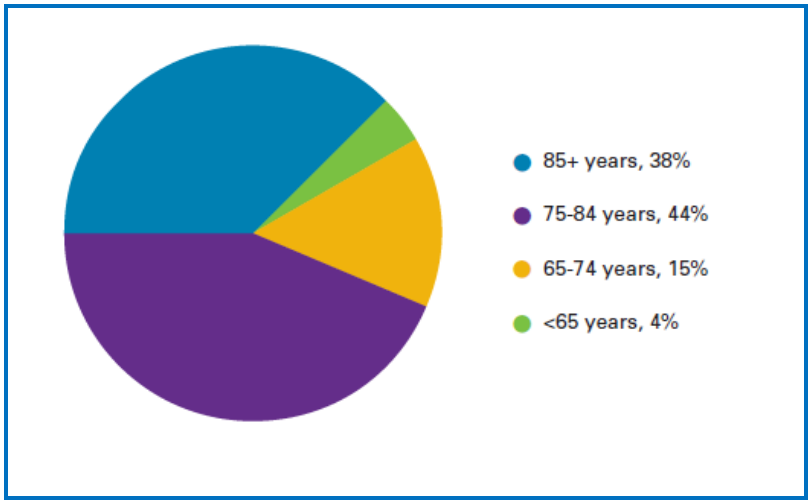


Figure 1.1.Proportion of people with AD (by age) in the United States.Percentages may not total 100 because of rounding, *Source: AD Facts and Figures [2014]*

The brain has billions of neurons, each with an axon and loads of dendrites. To reside healthy, neurons ought to communicate with each other, carry out metabolism, and refurbish themselves. AD disrupts all of these indispensable functions. No one knows what causes AD to begin, but we do know a lot about what happens in the brain once AD takes hold. The most fundamental features of AD incorporate the construction of extracellular protein deposits in the brain that consist of aggregates of beta amyloid protein (Aβ protein) (senile plaques) and NFTs (hyper-phosphorylated tau protein) in the intracellular compartments,turbulence in calcium homeostasis, and disintegration of synapses and neurons. That leads to nerve cell deathand tissue loss throughout the brain [Venugopal *et al.*, 2008]. Over time, the brain shrinks considerably, disturbing nearly all its functions (**Figure 1.2**).

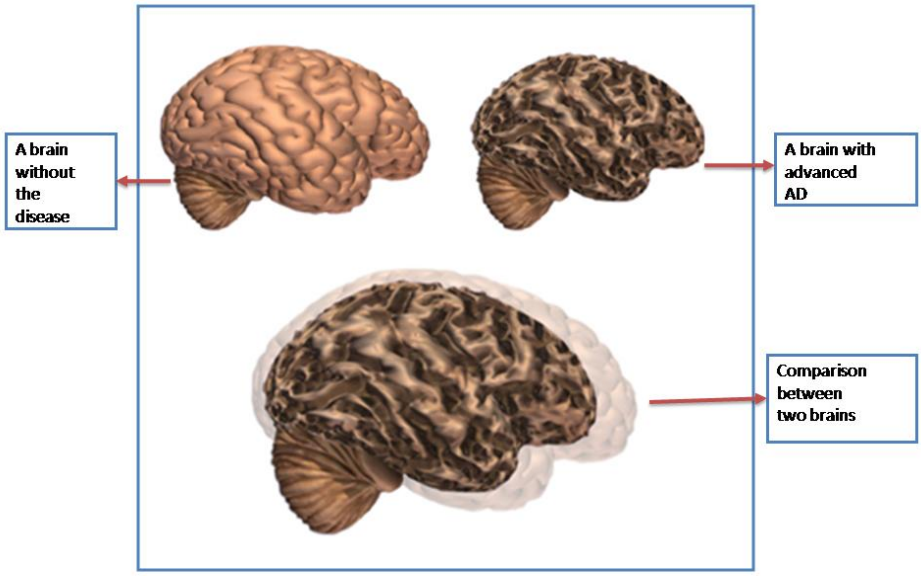


Figure 1.2.Difference between Normal Brain and AD Affected Brain: In case of AD, there is an overall shrinkage of brain tissue. The grooves or furrows in the brain called sulci, are noticeably widened and there is shrinkage of the gyri, the well-developed folds of the brain's outer layer (*With permission from:©2014 Alzheimer's Association. www.alz.org. All rights reserved. Illustrations by Stacy Jannis*)

Another viewpoint of how substantial cell loss changes the whole brain in advanced AD (**Figure 1.3**). This figure shows a diagonal 'segment' throughout the center of the brain flankedby the ears. In the AD brain, the cortex shrivels up, damaging areas involved in thinking, planning and remembering. Contraction is particularly rigorous in the hippocampus, a part of the cortex that plays a significant role in the construction of new-fangled memories. Ventricles (fluid-filled places contained by the brain) grow bigger.

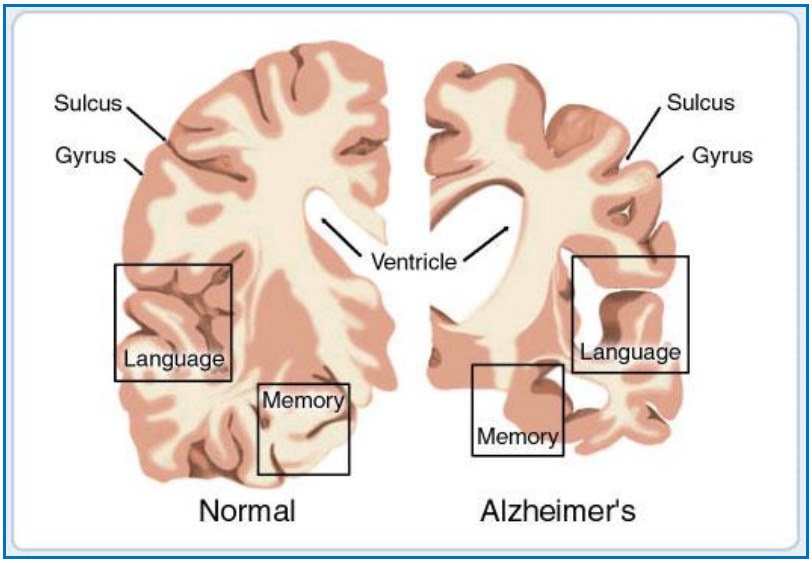


Figure 1.3.A cross-section of the brain as seen from the front. The cross-section on the left represents a normal brain and the one on the right represents a brain with AD (*With permission from: The medical illustration is provided courtesy of AD Research, a program of BrightFocus Foundation http://www.brightfocus.org/alzheimers*)

As we discussed above, one of the hallmarks of AD is the amassing of amyloid plaques among nerve cells in the brain. Amyloid is a common expression for protein fragments that the body produces generally. Aβ is a piece of a protein snipped from one more protein named amyloid precursor protein (APP). These protein pieceswould collapse and eliminated in case of a healthy brain. However in case of AD, the splinters mount up to form solid and impenetrable plaques inside the brain. NFTs are insoluble twisted fibers originate in the interior part of the brain's nerve cells. Succession of NFTs pathology is directly associated with both increased neurodegeneration and cognitive decline in AD and other tauopathies, for instance fronto-temporal dementia [Kubis and Janusz, 2008].

Scientists are not completely convinced what causes cell death and tissue loss in the AD brain, but plaques and tangles are key suspects. Scientists can see the dreadful effects of AD when they visualize at brain tissue under the microscope (**Figure 1.4**).

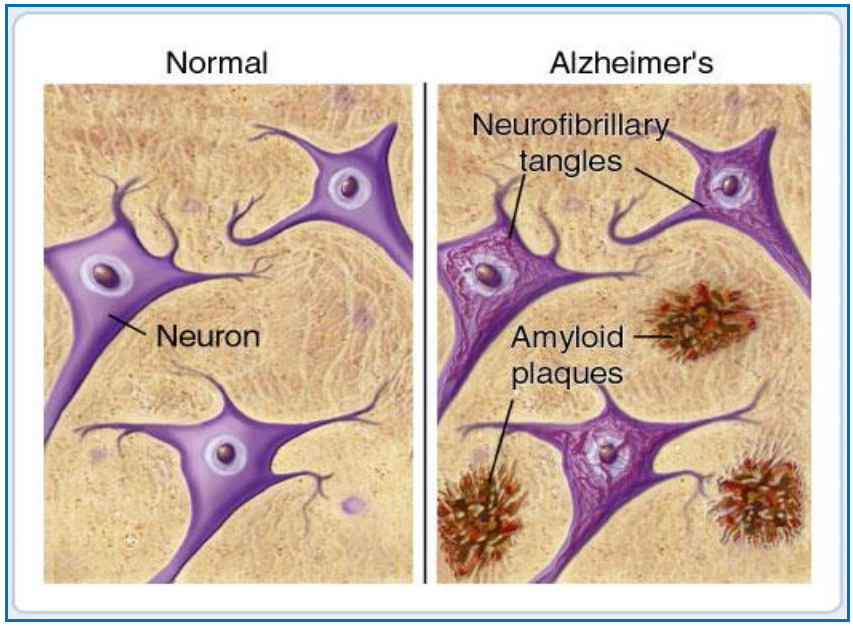


Figure 1.4.The construction of A*β* plaques and NFTs are thought to contribute to the degradation of the neurons in the brain and the consequent symptoms of AD (*With permission from: The medical illustration is provided courtesy of AD Research, a program of BrightFocus Foundation*http://www.brightfocus.org/alzheimers)**

1.7.1 Beta-amyloid Plaques

APP is the precursor to A*β* plaque. It sticks through the neuron membrane. Enzymes like *β*-secretase (beta-amyloid cleaving enzyme (BACE or *β*)) followed by *γ*-secretase (*γ*) divide the APP into fragments of protein, principally to form A*β*-42 type aggregations (**Figure 1.5**) [Oddo *et al.*, 2003; Hong *et al.*, 2004]. The construction of extracellular A*β* plaques is described by the amyloid cascade theory of plaque causing pathology (**Figure 1.6**) [Van Dam and De Deyn, 2006]. A*β* fragments come together in clumps to form plaques (**Figure 1.7**) in vulnerable brain regions, and disrupt the function of neurons. A*β* is chemically 'sticky' and progressively builds up into plaques. The most detrimental form of A*β* may be groups of a small number of pieces rather than the plaques themselves. The miniature clumps may obstruct cell-to-cell signaling at synapses. They may also activate immune system cells that generate inflammation and devour disabled cells.

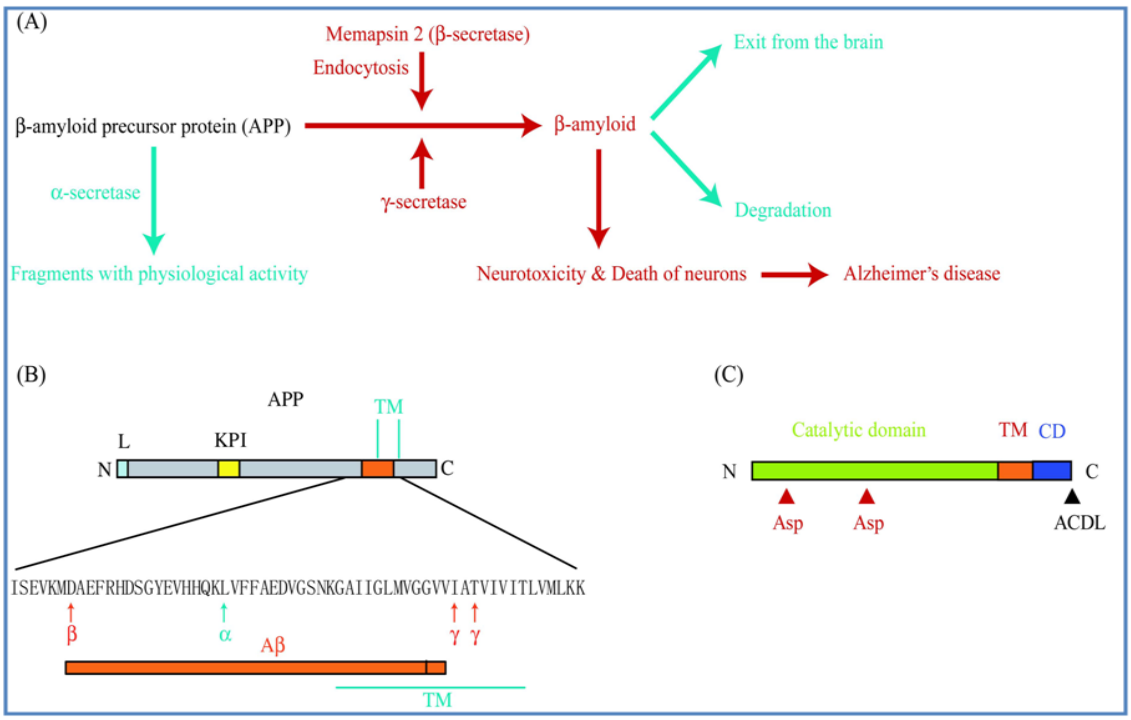


Figure 1.5.(A) Graphic presentation of the A*β* formation and role of memapsin 2 (*β*) in ADprogression.Pathway to AD in red and competing pathways that reduce A*β*are in blue. (B) Processing sites of APP by *α*, *β*, and *γ*-secretase. (C) Structural domains in *β*. The catalytic transmembrane (TM) and cytosolic domains (CD) are shown. Redtriangles stainthe positions of active site aspartic acids and the black one stain the position of ACDL motif. KPI, Kunitz protease inhibitor domain; L, leader sequence,

Adapted from Hong et al., [2004]

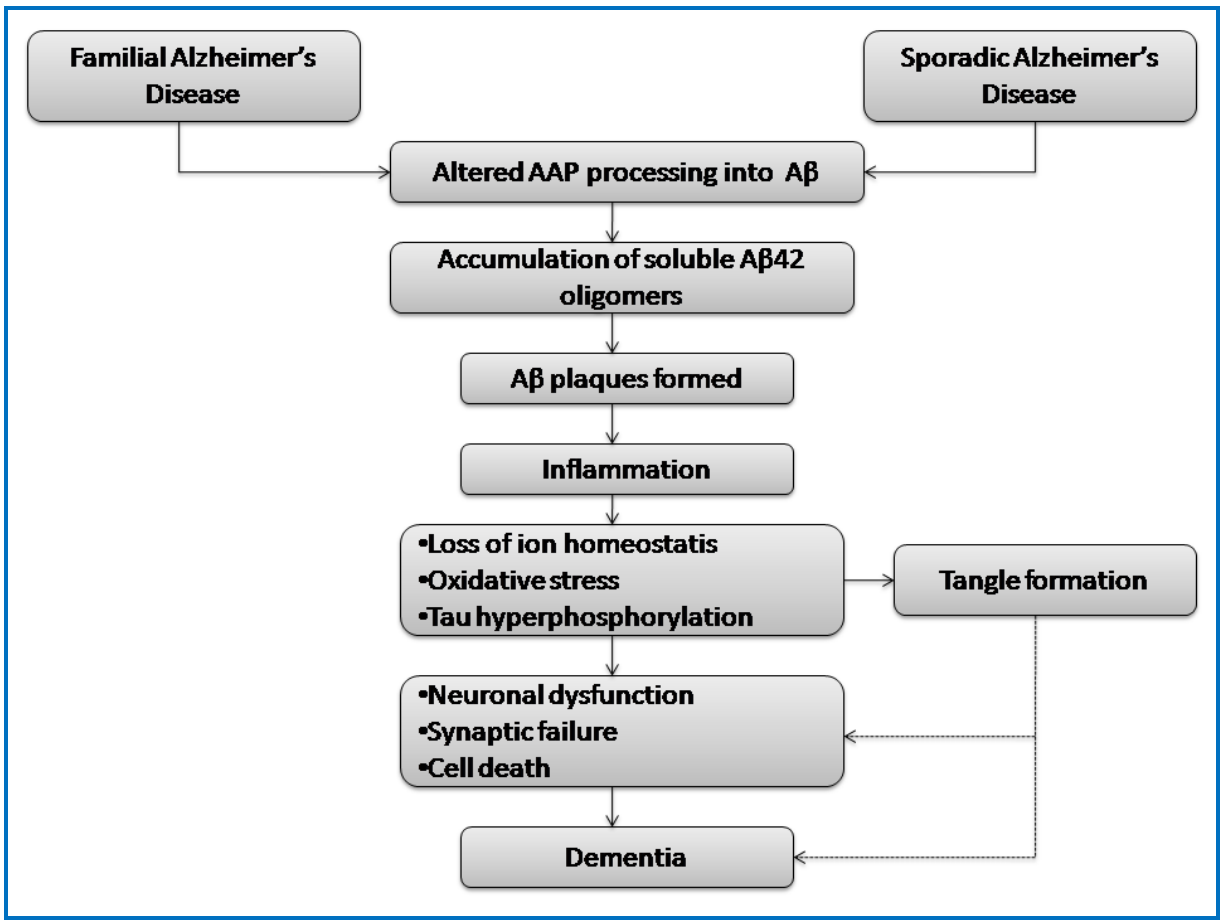


Figure 1.6.The Amyloid cascade hypothesis proposes that there is a primary inequality among A*β* production and its subsequent clearance, with amplified A*β*production in familial disease and decreased A*β*clearance in sporadic disease. A*β*oligomers may reduce hippocampal function and damage the synaptic function, as well as leading to inflammation and oxidative stress caused by the aggregation and deposition of A*β*. Both the processes unite to damage neuronal and synaptic function through ensuing neurotransmitter deficits and cognitive symptoms, *Adapted from Van Dam and De Deyn [2006]*

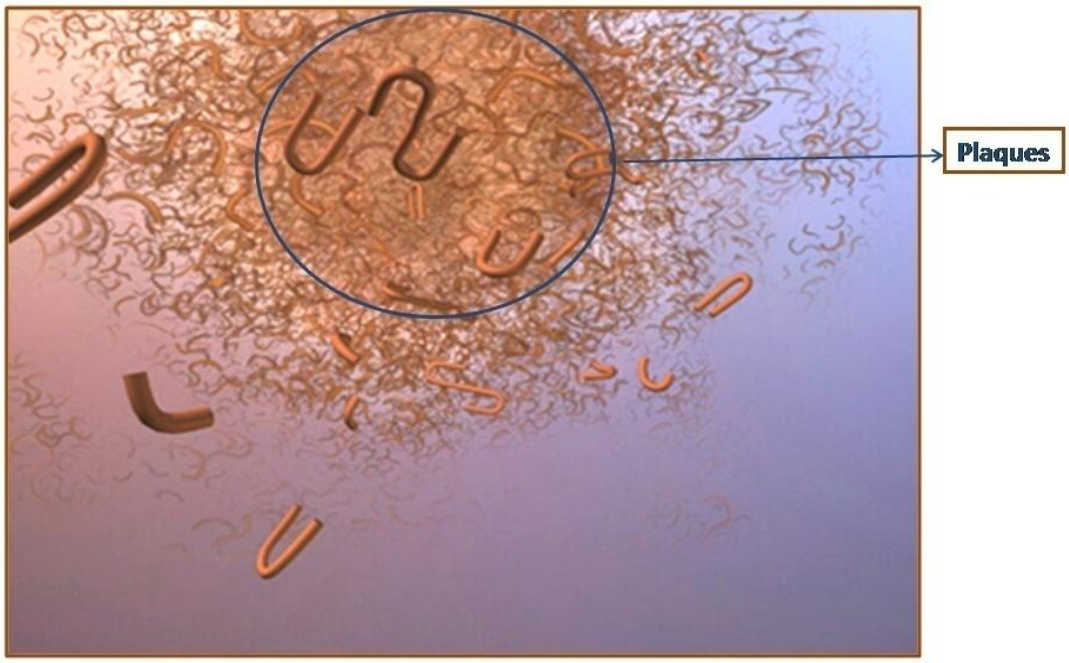


Figure 1.7.A*β* plaque, that visualize in vulnerable brain regions under the microscope (*With permission from: ©2014 Alzheimer's Association. www.alz.org. All rights reserved. Illustrations by Stacy Jannis)*

1.7.2 Neurofibrillary Tangles

The subsequent major hallmark of AD associated changes in the brain is intracellular formations named NFTs. NFTs are principally composed of paired helical filaments (PHF). The essential constituent of the NFTs is the tau protein, a microtubule associated protein (MAP), which binds with microtubulin to supply structural stability to a cell. In AD, the tau protein is abnormal and the microtubule structures collapse or dissociation of the tau protein from the microtublin leads to unbounded tau protein aggregation. The cause for the aggregation is elucidated by the tau hypothesis [Su *et al.*, 1996]. Tau, which is a soluble protein, undergoes phosphorylation and dephosphorylation, hence forming insoluble aggregates, under normal conditions. An inequality in this dynamic function, results in increased levels of unusually hyperphoshorylated tau (P-tau 181, P-tau 199, P-tau 231, P-tau 396, and P-tau 404), which in turn sequesters normal tau and other MAPs (MAP1 and MAP2) [Blennow *et al.*, 2007]. Hyperphosphorylated tau associates into PHF, and tangle construction. Corresponding to the progression of tangle construction is the disassembly of

microtubules. The collective results of both tangle formation and disassembly of microtubles is that they disrupt normal neuronal and synaptic function [Blennow *et al.*, 2006].

According to the amyloid cascade hypothesis, it is the raise in concentration levels of Aβ that stimulate the changes in tau, therefore leading to the formation of NFTs. Alim *et al.* (2004), revealed that the protein α-synuclein (aberrant forms of which are core components of lewy body based pathologies, recognized as synucleinopathies), having the same function like tau, which is engaged in microtubule assembly, serving as a fastening for the tublin 25 protein. However this fastening capability is vanished when α-synuclein becomes mutated, ensuing in tubulin aggregation. The microtubule helps transport nutrients and other crucial substances from one part of the nerve cell to another.Since microtubule function is essential for regular neuronal and synaptic function, dysfunction of the microtubules may be critical in neurodegeneration [Alim *et al.*, 2004].AD tissue has smaller quantity of nerve cells and synapses than a healthy brain. More specific comparison can be given as:

In healthy brain

- The transport system is organized in orderly parallel strands somewhat like railroad tracks. Food molecules, cell parts and other crucial materials travel along the 'tracks'.
- Tau protein helps the tracks reside straight.

In brain areas where tangles are forming

- Tau collapses into twisted strands named tangles.
- The tracks can no longer reside straight. They collapse and disintegrate.
- Nutrients and other necessary supplies can no longer travel through the cells, which ultimately die.

1.8 Aβ AND ITS ASSOCIATION WITH TAU

The important question at this point is how Aβ and tau are interrelated during the disease. Even though the association between these two proteins remains indistinct, data has lent support to the hypothesis that phosphorylation of tau protein could be the key linking

mechanism. Few years ago it was found that Aβ fibrils hasten the construction of abnormally phosphorylated NFTs in a tau transgenic mouse [Gotz *et al*., 2001]. Recently, it has been revealed that Aβ oligomers cause abnormal tau phosphorylation and morphology changes of spines by missorting of endogenous tau into dendrites. Aβ self-aggregates into oligomers of various sizes and forms disperse and neuritic plaques in the blood vessels and parenchyma. Aβ oligomers and plaques are potent synaptotoxins, block proteasome function, stimulate inflammatory processes, inhibit mitochondrial activity and alter intracellular Ca^{2+} levels. Aβ interacts with the signalling pathways that control the phosphorylation of the tau. Hyperphosphorylation of tau disrupts its regular function in regulating axonal transport and directs the accumulation of NFTs and toxic species of soluble tau [Zempel *et al*., 2010].Moreover, degradation of hyperphosphorylated tau by the proteasome is inhibited by the actions of Aβ. These two proteins and their connected signalling pathways consequently signify imperative therapeutic targets for AD.Some evidence has shown that Aβ might not be the single constituent involved in the pathogenesis of AD. There is one other cleavage fragments of APP, called amyloid C-terminal fragment (βCTF), which might involved in the pathophysiology of AD. The βCTF shows higher neurotoxicity than Aβ, including endosome dysfunction, neurodegeneration, and synaptic deficits [Choi *et al.*, 2001; Chang and Suh, 2005; Lee *et al.*, 2006; Chen *et al*., 2014].

1.9 SYMPTOMS AND STAGES OF ALZHEIMER'S DISEASE

In AD patient's brain, the ventricles, which are chambers within the brain that restrain cerebrospinal fluid, are noticeably enlarged. In the early stages of AD, short-term memory begins to decline when the cells in the hippocampus regiondisintegrate. As AD extends throughout the cerebral cortex, judgment get worse, emotional outbreaks may takes place and language is damaged. In the final stages, people may lose the capability to feed themselves, speak, recognize people and control bodily functions. Memory worsens and may become approximately non-existent. Constant care is usually compulsory. On standard, individuals with AD live for 8 to 10 years after diagnosis, but this incurable disease can last for as long as 20 years.The three stages listed below characterize the universal succession of the disease. Even though these symptoms will likely differ in severity, chronology, overlap, and swing the overall progress of the disease is reasonably conventional; however AD doesn't affect every person in a similar way [Forstl and Kurz, 1999].

1.9.1 Stage 1 (Mild)

Early in the sickness, persons with AD have a tendency to be less lively and spontaneous.This stage can last from 2 to 4 years.They may become introvert, circumvent peopleand are lethargic to learn and respond. They also have difficulty performing everyday duties, and facedifficulty in communicating and understanding written material. A number of specific examples of behaviors that people exhibit in this mild phase comprise [Alzheimer's Association, 2010; Mayo Clinic Medical Information and Tools for Healthy Living, 2010]:

- Getting lost
- Trouble managing money and paying bills
- Repetitive questions and conversations
- Deprived judgment
- Losing things or misplacing them in unusual places
- Noticeable changes in individuality or temper

1.9.2 Stage 2 (Moderate)

This is normally the longest phase and can last from 2 to 10 years. Individuals with AD are noticeably becoming disabled, in this period. Persons can still complete simple tasks autonomously, however may require support with other convoluted activities. They fail to remember current events and their individual history, and happen to more perplexed and disjointed from reality. Verbal communication problems occur, reading and writing are more complicated, and the person may instigate terminology. They may no longer be secure single-handedly and can wander. Since AD patients become attentive to this loss of control, they may become disheartened, short-tempered and impatient or apathetic and introvert. Individuals may experience sleep turbulence and have more problems such as; ingestion, grooming and dressing [Alzheimer's Association, 2010; Mayo Clinic Medical Information and Tools for Healthy Living, 2010].

1.9.3 Stage 3 (Severe)

During this phase patients may last 1 to 3 years. Individuals in this period may mislay the capability to feed themselves and manage bodily functions, for example bowel and bladder

control. They will sleep frequently and grunting or moaning can be ordinary. During this destabilized physical phase, individuals may become susceptible to additional illnesses, skin infections, and respiratory troubles, predominantly while they are incapable to move around [Alzheimer's Association, 2010; Mayo Clinic Medical Information and Tools for Healthy Living, 2010].

1.10 TYPES OF ALZHEIMER'S DISEASE

There are primarily three types of AD, and they are as follows [(Alzheimer's Disease Education and Referral (ADEAR) Center), *http://www.nia.nih.gov/alzheimers/about-adear-center*]:

1.10.1 Early Onset Alzheimer's Disease (EOAD)

EOAD is a rare form of AD that affects individual beneath 65 years of age. The symptoms start to emerge in the 40-50 age groups. This accounts for below 10% of overall AD patients. Individual with Down's syndrome, who experience premature AG seems more prone to develop EOAD. The majority EOAD is sporadic, although about 5% of patients with EOAD have an extremely penetrant genetic mutations in amyloid pathway genes including APP on chromosome 21, presenilin 1 (PSEN1) on chromosome 14, and presenilin 2 (PSEN2) on chromosome 1. These mutations lead to the accumulation of Aβ plaques [Bertram, 2009].

1.10.2 Late Onset Alzheimer's Disease (LOAD)

LOAD is a frequent form of AD that appears in individual having 65 years of age and over. LOAD accounts roughly 90% of overall AD cases. It strikes approximately half of all individual over 85 years of age. Various low-penetrant genetic risk factors conferring a modest increase in possibility of disease have been recognized for LOAD, the most studied one is the apolipoprotein *ε4* allele (APOE *ε4*). The overall population occurrence of APOE-ε4 is 22%, although roughly 60% of LOAD cases carry at least one allele. Large, multi-center genome-wide association studies (GWAS) approximate the population attributable threat for APOE variants is 19-35% [Ertekin-Taner, 2010]. There are extra polymorphisms connected with LOAD risk including genes, recognized by GWAS those are ABCA7, BIN1, CD2AP, CD33, CLU, CR1, EPHA1, MS4A, and PICALM. Further APOE *ε4* dose amendment reveals

50% of the population attributable threat for LOAD is accounted for by identified single nucleotide polymorphisms (SNPs). Although these variants are significant for both risk judgment and discovery of novel mechanisms of pathogenesis, still they are neither essential nor satisfactory for the progress of LOAD [Naj *et al.*, 2011].

1.10.3 Familial Alzheimer's Disease (FAD)

This category makes up less than 1% of cases, and is noticeably evidenced by multiple patients over 3 or more generations being diagnosed with the disease. FAD seems to show up in the patient's of age 40's. All FAD known has an EOAD. FAD is also known as EOFAD i.e. early onset familial AD. Some patients with EOFAD, however, may lack a known family history or may have deficient penetrance. For these causes, it is central to consider evaluating EOAD patients with an extremely early age of onset or anomalous neurological findings for genetic factors [Llado *et al.*, 2010].

1.11 RISK FACTORS FOR ALZHEIMER'S DISEASE

There are two types of risk factors

- Non-Genetic Risk Factors
- Genetic Risk Factors

1.11.1 Non-Genetic Risk Factors

1.11.1.1 Age

Right now, age is the distinct utmost risk factor for developing AD, along with family history. More women have AD than men, although this is estimated because women normally live longer than men. Most cases of AD are seen in olders i.e. ages 65 years or beyond. Between the ages of 65 and 74, roughly 5 to 10 percent of people have AD. For individuals over 85, the threat raises to 50% [Alzheimer's Association, 2010].

1.11.1.2 Education

There might be a correlation among educational level and the possibility of developing AD. Individuals having fewer years of education appear to be at a superior threat. It is theorized that a greater education level directs to the construction of more synaptic connections in the brain, however the accurate reason for this correspondence is unidentified. This generates a "synaptic preserve" in the brain, that enabling patients to reimburse for the loss of neurons as the disease progresses [Alzheimer's Disease Facts and Figures, 2010; Mayo Clinic Medical Information and Tools for Healthy Living, 2010].

1.11.1.3 Concomitant Health Problems

Earlier history of head trauma is also frequently agreed upon as plausible risk factors for AD. There is a sturdy connection among cardiovascular health and brain health. Having heart disease, high blood pressure or high cholesterol can amplify the risk of developing AD. This is caused by damage to blood vessels in the brain, resulting in less blood flow and possible brain tissue death. Type 2 diabetes may also increase the risk for AD. Inefficiency of insulin to convert blood sugar to energy may cause higher levels of sugar in the brain, causing harm.The use of certain groups of drugs, including non-steroidal anti-inflammatory drugs (NSAIDs) and cholesterol-lowering drugs called statins, may also impact AD risk, according to a number of studies [Shobab *et al.*, 2005]. Compelling new evidence is now indicating that other "lifestyle aspects", such as one's dietary habits may impact one's risk for developing AD. The rationales why these factors enhance AD risk are unclear. Hypercholesterolaemia, hypertension, coronary heart disease, obesity, atherosclerosis, diabetes and smoking are all allied to AD, since these are all considered to be vascular risk factors, distressing the effectual supply of blood. Evidence implies that countering risks that incline a person to dementia consist of benefits resulting through diet modification i.e. escalating an intake of homocysteine-related vitamins (vitamin B12 and folate); antioxidants, like vitamin C and E; unsaturated fatty acids etc [Blennow *et al.*, 2006]. In a large populace based twin study, heritability for sporadic AD was high i.e. 79% by the similar genetic factors being as significant, irrespective of sex and other non-genetic risk factors [Gatz *et al.*, 2006].

1.11.2 Genetic Risk Factors

1.11.2.1 Familial Alzheimer's disease

All FAD recognized so far has an early onset and as many as 50% of the cases are now known to be caused by defects in three genes situated on three different chromosomes as discussed above. Even though one of these mutations is present in only one of the two copies of a gene inherited from a person's parents, the person will inevitably develop that form of EOAD. However, the overall known number of these cases is very few (between 100 and 200 worldwide), and there is as yet no evidence that any of these mutations play a major role in the more common, sporadic form of LOAD. Scientists are working to divulge the common function of APP and presenilins (PSEN1 and PSEN2) and to find out how mutations of these genes cause the onset of FAD [Mudher and Lovestone, 2002].

1.11.2.2 Sporadic Alzheimer's disease

Genetics appear to play a significant role in the development of the more frequent form of AD; LOAD. Scientists have found an increased threat for LOAD in populace who inherit one or two copies of a particular variation of gene, APOE *ε4*. The variations in the APOE gene that directs the manufacture of APOE, a protein that helps carry blood cholesterol all through the body, among other functions. It is found in neurons and further supportive brain cells (known as glia) of healthy brains; however it is also connected in surfeit amounts with the plaques found in the brains of AD people. Researchers are mainly interested in three frequent alleles of the APOE gene that are; *ε2*, *ε3* and *ε4*. The finding that increased threat is linked with inheritance of the APOE *ε4* allele, the more APOE *ε4* alleles one inherits, the lower the age of disease onset [Mattson, 2004]. The comparatively uncommon APOE *ε2* allele may protect some populace against the disease. APOE *ε3* is the most frequent version found in the general population and may perform as a neutral role in AD threat. An individual can have one or two APOE *ε4* alleles and still not get the disease, and an individual who develops the disease may not have any APOE *ε4* alleles. That means it increases the threat of developing AD, however it does not cause the disease. The mechanism how, APOE *ε4* increases the probability of developing AD are not known with conviction, but one possible mechanism is that it facilitates A*β* buildup in plaques. Supplementary theories engross interactions with

cholesterol levels and effects on nerve cell death that are independent of its effects on plaque buildup [Corder *et al.*, 1993].

1.12 DIAGNOSIS OF ALZHEIMER'S DISEASE

The only technique to definitively diagnosing AD is through a brain autopsy. On the other hand, mental and behavioral tests and physical evaluations allow physicians to make a correct diagnosis of AD in 90% of cases. The decisive factor for detecting mental disorders can be found in the Diagnostic and Statistical Manual of Mental Disorders (DSM-III), published by the 'American Psychiatric Association'. In this manual, AD falls into the group of principal degenerative dementia. The diagnostic criterion includes dementia (diagnosis of dementia includes loss of intellectual capabilities severe enough to obstruct with social or occupational functioning, memory impairment etc.), insidious onset with progressive deterioration and exclusion of all other types of dementia by history and physical inspection [American Health Assistance Foundation (AHAF)]. The initial step in finding a diagnosis is obtaining the patient history. Through this time, the physician will verify symptoms; those are present at the beginning stage, and how they have progressed over time. The family history of sickness is also significant. To rule out additional potential reasons of dementia, such as hormone imbalance, urinary tract infections and vitamin deficiency (vitamin B12), physician should perform a physical examination (blood tests and urinalysis etc).

Brain scans may also be performed to exclude other possible causes of dementia, including brain tumors, stroke, blood accumulation on the brain surface etc. These scans are also helpful in identifying the characteristic tangles and plaques seen in AD.Structural imaging scan includes; magnetic resonance imaging (MRI) and computed tomography (CT) which will provide information about the shape and volume of the brain. Functional imaging allows the physician to find out how efficiently the brain cells are functioning. A functional MRI or positron emission tomography (PET) scan can be used [Emilien *et al.*, 2004]. Physicians may direct an electroencephalogram (EEG) to determine the electrical activity in the brain. Rarely, spinal fluid may be tested throughout a lumbar puncture.

Neuropsychological tests also identify cognitive symptoms associated with brain injury or abnormal brain function. Physicians usually start with a concise screening tool, for instance the Mini-Mental Status Examination (MMSE), to verify that the patient is

experiencing tribulations with intellectual functions. In MMSE, the physician begins by asking a series of questions premeditated to check the patient's aptitude to recall and name a list of objects, perform simple arithmetic and follow instructions. The patient is then assigned a score out of 30 probable points, with a score of less than 12 indicating severe dementia. AD patient's usually scores 2 to 4 points. Further neuropsychological testing beyond the MMSE is typically not needed, if a patient has severe dementia. Conversely, for patients with mild intellectual deficits, further tests may be necessary to decide whether the patient is just showing signs of advanced age or is developing AD. The physician may also use the Alzheimer's disease Assessment Scale (ADAS) to determine the severity of the disease. The ADAS evaluates the patient's orientation, memory, language and reasoning on a scale of '0 to 70'. A higher score represents a higher level of cognitive impairment. ADAS is sensitive to a broad array of symptoms and evaluates many cognitive skills, including spoken language ability, ability to find correct words, recall of instructions, following commands and orientation to surroundings and time [Emilien *et al.*, 2004]. Besides mental tests, the doctor may perform a neurological test to evaluate the function of the patient's brain and nervous system. This will test reflexes, coordination and balance, muscle strength, speech, sensation, and eye function. Scientists are also looking at changes occurring in the blood and cerebrospinal fluid that may specify the progression of AD. Additionally, they are developing sophisticated brain imaging methods that helps in measuring the slightest changes in brain function or structure to detect AD prior to any noticeable symptoms occur.

1.13 IMPORTANT FACTS OF ALZHEIMER'S DISEASE

- AD is a distressing neurodegenerative disorder with a relentless progression.
- One person in the U.S is diagnosed with AD approximately every 69 seconds. According to data from the CDC, in 2010, over 82,000 deaths were traced as being caused by AD.
- AD is the 6^{th} leading cause of death in the U.S. It is anticipated that there were 35.6 million people living with dementia worldwide in 2010, and will rise to 65.7 million by 2030 and 115.4 million by 2050.

- It is expected that more than one in three Americans 85 year and older have AD. The lifetime threat of AD among those who reached the age of 65 is just about 1 in 5 for women and 1 in 10 among men.

- Around 5.1 million Americans are age 85 years or older, and this age group is one of the fastest growing segments of the population. It is also the group with the highest threat of AD. It is anticipated that at least 19 million people in America will be age 85 and older by the year 2050.

- Patients with AD live for approximately 8 to 10 years after diagnosis, but this fatal disease can last as long as 20 years, or as little as 3 to 4 years if the patient is over 80 years old when diagnosed.

- Just about 70% of AD patients receive care at home.In terms of health care expenses and lost wages of both patients and their caregivers, the cost of AD nationwide is estimated at $100 billion per year.For an individual with AD, the annual cost of home care is estimated at $76,000, including medical expenses and indirect costs such as a caregiver's time and lost wages.

- 58% of populace with dementia worldwide lives in low or middle income countries.One third of those whose lives have been touched by AD supply support to their loved ones.

- One third fears about getting AD, among those who do not personally have AD. Those who have a parent or parent in law with the disease are even more concerned.

- Approximately 50% of all caregivers are between the ages of 18 and 49, with the average age of the typical caregiver being 48.

1.14 ETIOLOGICAL HYPOTHESIS FOR ALZHEIMER'S DISEASE

1.14.1 Oxidative Stress

Oxidative stress is recognized to bring damage of diverse biological macromolecules in an unrestrained mode. It is also considered to be a characteristic of neurodegenerative diseases.

One deliberation is that the plaques and tangles rather than being critical in the beginning or pathology of the disease are potentially acting as an antioxidant defence, to facilitate a protective action. Consequently the subsequent appearance of Aβ deposits and tau hyperphosphorylation is a result of this defence [Smith *et al.*, 2002]. It has shown in animal model that, oxidative damage precedes the pathological modifications connected with AD [Nunomura *et al.*, 2001].

1.14.2 Inflammation

It is known that, brain regions which are affected by AD are containing augmented neuroinflammatory mediators (cytokines and microglia) through increased inflammatory cascades. Whether this is an innate response to control inflammation or an out of control immune procedure is anonymous. Cyclooxygenase (COX) a prime mediator of the inflammatory cascade is targeted by NSAIDs, affecting COX levels. The use of NSAIDs does not diminish the threat or setback the onset of AD [Van Gool *et al.*, 2001]. Microglia activation is thought to be an untimely event in the AD pathogenesis and may be important in synaptic disorder and hence early memory impairment. The AD Anti-inflammatory Prevention Trial (ADAPT) looked into the function of NSAIDs in people vulnerable to dementia, using COX-1 and COX-2 drugs. The testing was cancelled because of cardiovascular risks (ADAPT, 2006). COX-1 targeting NSAIDs are thought to be a better preference than COX-2 inhibitors [McGeer and McGeer, 2007].

1.14.3 Cholinergic hypothesis

Cholinergic hypothesis of AD proposes that destruction of the cholinergic pathway in the basal forebrain consequences in diminish of cholinergic neurons, which discharge the neurotransmitter acetylcholine (ACh). These neurons project to the hippocampus and neocortex, which are concerned in both memory interruption and cognitive symptoms [Bartus, 2000]. ACh is disintegrated by the enzyme acetylcholinesterase (AChE) (**Figure 1.8**). Concentration of this enzyme is reduced in moderate and severe AD patients. Cholinesterase inhibition progresses neurotransmitter function and provides relief to AD symptoms [Terry and Buccafusco, 2003].

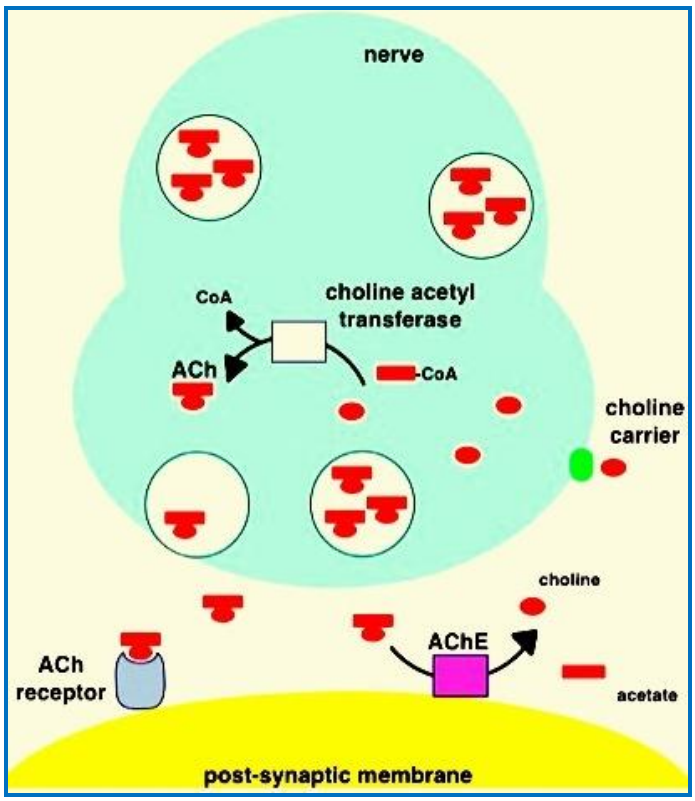


Figure1.8.Scheme summarizing the role of ACh and AChE in cholinergic activities, *Adapted from Katzung, [2001]*

1.14.4 Cholesterol metabolism

The function of lipid/cholesterol metabolism and AD pathogenesis is gaining adequacy. Cholesterol is known to influence the activity of enzymes implicated in the metabolism of APP in the construction of Aβ. As we have discussed above, cholesterol-lowering drugs statins associated with a lower threat of developing dementing illness; however a recent trial has revealed that the use of statins does not impact on enhancement in recorded cognitive impairment [Jones *et al.*, 2008]. APOE is involved in the transporting of cholesterol and APOE $\varepsilon 4$ allele is a universally recognized marker which enhances AD risk [Corder *et al.*, 1993]. A high cholesterol level throughout a person's mid-life is considered a threat factor for AD.

1.15 TREATMENT OF ALZHEIMER'S DISEASE

There are varied hypotheses proposed for the pathogenesis of AD. These comprise glutamate excitotoxicity as a result of obstruct of glutamate uptake into the astrocytes by Aβ aggregates, oxidative stress and membrane lipid peroxidation induced by Aβ aggregates, membrane lipid peroxidation due to C-terminal fragment of APP, microglial activation by Aβ aggregates and molecular pathways activated by Aβ induced stimulation of various kinases including MAP kinases and JNK (Jun amino-terminal kinase) [Suh and Checler, 2002]. A discrepancy of ACh in an AD brain is well identified. At the same instant, level of dementia show a relationship with the extent of neuronal death caused by surplus of glutamate, the most widespread excitatory neurotransmitter in the brain [Wenk *et al*., 1996] (**Figure 1.9**).

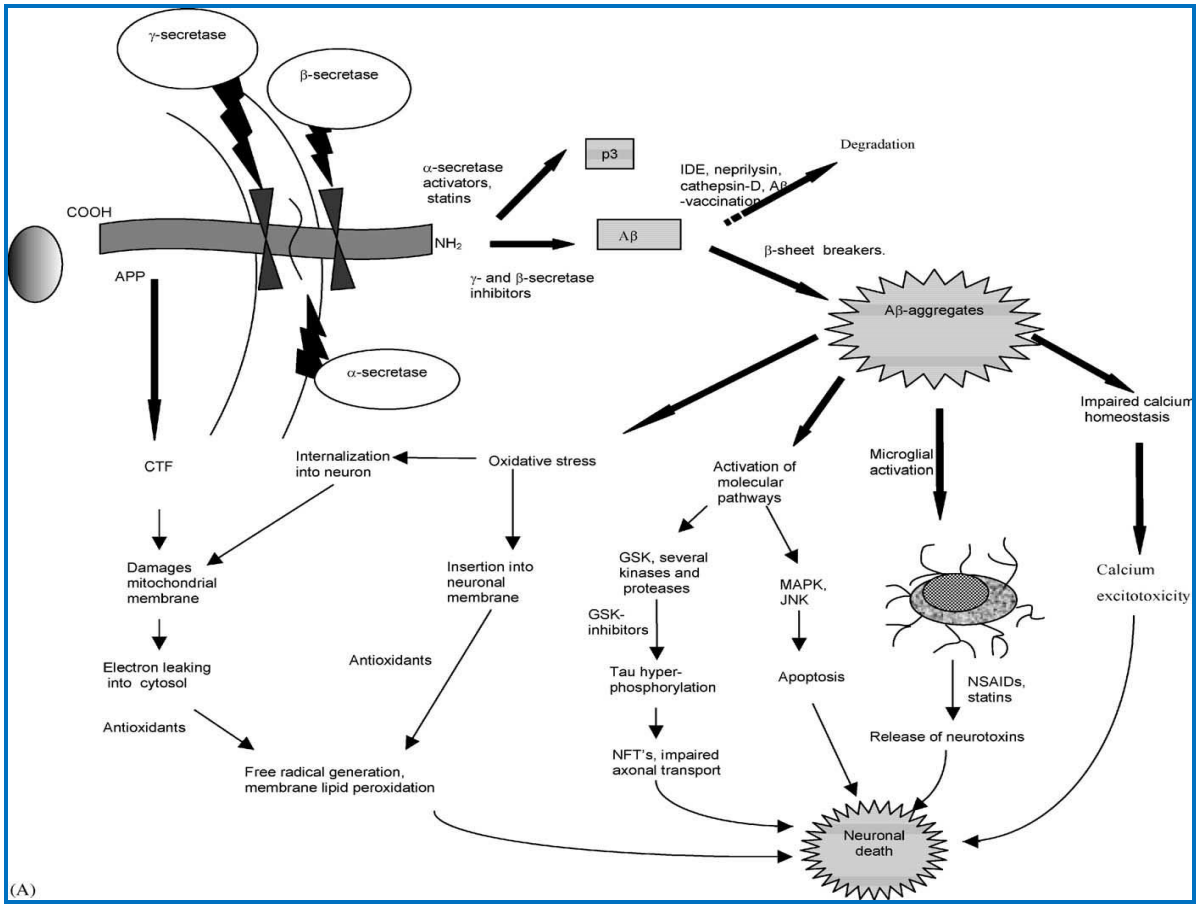


Figure 1.9.Steps involved in neurodegeneration of AD and targets there in for novel treatment strategies. GSK: glycogen synthase kinase; CTF: C-terminal fragment; MAPK: mitogen activated protein kinase (A: calcium channel), *Adapted from Sonkusare et al., [2005]*

Since the last decade, AD patients are being treated with substitute of neurotransmitters that are deficient in AD brain, based on 'cholinergic hypotheses'. Presently standard drugs for AD treatment are cholinesterase inhibitors that comprise tacrine (Cognex®), rivastigmine (Exelon®), donepezil (Aricept®) and galantamine (Reminyl®) (**Tables 1.1 and 1.2**). The methodology is to inactivate AChE, enzyme that cleaves synaptic ACh and terminates neuronal signaling. These drugs have inadequate success because they only improve memory in mild dementia however they cannot prevent the process of neurodegeneration. The magnitude of rejoinder to AChE inhibitors depends on integrity of pre-synaptic neurons. Evidently, efficacy of these agents will reduce with enhance in severity of AD. Furthermore, their use is connected with gastro-intestinal side effects [Hake, 2001].

Table 1.1 Proportional pharmacology of presently accepted drugs

Drug	**Mechanism**	**Adverse effects**	**Special comments**
Tacrine (Cognex®)	Non-selective cholinesterase inhibitor, more affinity towards synaptic G4 form of AChE than G1 form	Hepatotoxicity and gastro-intestinal symptoms such as nausea, anorexia, diarrhoea	Monitoring of liver transaminase levels is required; also inhibits K+-channels weakly
Donepezil (Aricept®)	Selective AChE inhibitor affinity for G1 and G4 forms varies from region to region	Diarrhoea, nausea, anorexia, vomiting, muscle cramps, fatigue in some cases	Reversible inhibitor, high oral bioavailability and no hepatotoxicity
Rivastigmine (Exelon®)	Non-selective inhibitor of cholinesterases, has higher affinity for cytoplasmic G1 Form of AChE	Nausea, vomiting and diarrhoea	Pseudo-irreversible inhibitor inactivated by enzymatic cleavage at the active site of enzyme and not by metabolism
Galantamine (Reminyl®)	Specifically inhibits AChE	Nausea, vomiting, anorexia, weight loss acutely at higher doses	Also binds to presynaptic nicotinic receptors, stimulating ACh release

Memantine (Ebixa®, Axura®, Akatinol®, Namenda®)	Moderate affinity, non-competitive inhibitor of NMDA receptors	Mild and not common, constipation, confusion, headache, dizziness, tiredness	Inhibits pathological but not physiological functions of NMDA receptors, also has antioxidant action and increases BDNF production

*** BDNF: Brain Derived Neurotrophic Factor; G1 and G4 forms: Isoforms of AChE; G1 form is a cytoplasmic whereas G4 form is synaptic.**

Table 1.2 Proportional clinical pharmacology of presently available drugs

Characteristic	Tacrine	Rivastigmine	Donepezil	Galantamine	Memantine
Introduced	1993	2000	1996	2001	2002[1]
Chemical class	Aminoacridine	Phenylcarbamate	Piperidine	Phenanthrene Alkaloid	Amino Adamantine
Maximum dose (mg per day)	160	12	10	24	20
Times per day	4	2	1	2	1
Initial dose (mg)	40	3	5	8	10
***t*max (h)**	1-2	1	3-4	1	3-8
Plasma protein binding (%)	55	40	96	18	45
Elimination *t*1/2 (h)	2-3	1.5	70	7	60-80
Metabolism by CYP450 system	Yes (CYP1A2)	No	Yes (CYP3A4)	Yes	(CYP3A4) No
Interaction	Not	To be taken	Not	To be taken	Not

with food	observed	with meals[2]	observed	with meals[2]	observed
Drug–drug interactions	Yes	None known	Yes	Yes, some	No
Metabolites	Active	Inactive	Active	Active	Inactive
Bioavailability (%)	10-30	40	99	90	100

[1] The drug is being used over 10 years in Germany for the treatment of dementia but was approved for this purpose by the European Medicines Evaluation Agency in February 2002.

[2] Food delays absorption lowers *C*max and reduces adverse effects.

First generation anticholinesterases comprise physostigmine, tacrine and NIK-247. These agents were non-selective and inhibited butyryl cholinesterase (plasma cholinesterase). Tacrine and NIK-247 also block potassium channels. Tacrine is well-known to cause hepatotoxicity and sedation. Selective inhibition of AChE noticeably diminishes tangential adverse effects [Cocabelos *et al*., 1994]. Second generation of cholinesterase inhibitors were developed using this principle and comprise donepezil (Pfizer and Eisai) [Bryson and Benfield, 1997], eptastigmine (Mediolanum) [Imbimbo *et al*., 1999] and galantamine (Janssen) [Scott and Goa, 2000]. One more approach is to develop prodrugs, which discharge the active constituent slowly in the blood and giving enduring inhibition of cholinesterase. Metrifonate (Bayer) is one such medicine that releases dimethyl-2, 2-dichlorvinylphosphate. Huperazine A., a plant alkaloid having both, AChE with antioxidant activity, is under research as a prospective medicinal option for cure of AD. Dual inhibitors of AChE-SERT (serotonin transporters) would be an enhanced restorative option, since depression is usually seen in AD patients. Inhibition of SERT may also diminish dose-related side effects of AChE inhibitors. Such twofold inhibitors were designed by hybridization of rivastigmine and fluoxetine [Toda *et al*., 2003].

1.16 NEW REMEDIAL ADVANCES FOR TREATMENT OF ALZHEIMER'S DISEASE

Since the cells of central and peripheral nervous system cannot revitalize, newer approaches are intended at preserving the surviving neurons by reducing their disintegration. Some of these potentially disease modifying treatments include use of β-sheet breakers, NMDA receptor blockage, Aβ peptide vaccination, secretase inhibitors, antioxidant strategies, cholesterol-lowering drugs, anti-inflammatory agents, metal chelators [Helmuth, 2002], activators of phosphatases, neuroregeneration by neurotrophic factors and immunophylline ligands, inhibitors of tau phosphorylation, supply of neuronal cells by gene therapy and human embryonic stem cells [Scarpini *et al.*, 2003]. A concise depiction of these newer therapeutic approaches has been given below:

- A potential disease modifying medicine for the treatment of AD targets energy production in neurons. Particularly, the medicine targets bio-energetic pathways upstream from the Aβ peptide production found in AD. In preclinical studies, the medicine verified significant enhanced cognition.

- Aβ fragments are cleared through intra- and extra-cellular proteolysis and phagocytosis. Neprilysin and insulin degrading enzyme degrade Aβ fragments extracellularly. Activation of these enzymes or administration of exogenous enzymes can be helpful. Deliverance of genes that express neprilysin into neurons of frontal cortex and hippocampus decreases Aβ levels and neuronal death in transgenic mouse model expressing human form of Aβ [Brown, 2003].

- Inhibition of β- and γ-secretase decreases Aβ levels in brain. Activation of α-secretase stimulates non-amyloidogenic pathway of APP proteolysis that forms undisruptive p3, thus diminishing Aβ levels. Inhibition of γ-secretase reduces cleavage of APP and Notch. Inhibition of notch cleavage alters embryogenesis, thymocyte, and haematopoiesis maturation. Partial inhibition of γ-secretase does not restrain its physiological functions [Doerfler *et al.*, 2001].

- Occurrence of oxidative stress is eminent in AD brain [Ames *et al.*, 1993]. Aβ aggregates can directly insert into cell membrane and cause membrane lipid

peroxidation. After being introduced into the cell, Aβ aggregates damage mitochondrial membrane leading to electron seepage from mitochondrial electron transport system. This outcome into formation of intracellular free radicals. C-terminal fragment of APP also results into intracellular free radical formation in a similar way. These free radicals cause membrane lipid peroxidation and perturb the integrity of neuronal membrane. This finally results in neuronal death. Free radical formation also couples with calcium excitotoxicity and together these aspects act synergistically to cause neuronal death. Incidentally antioxidants, predominantly those, which can cross cell membrane, can prove beneficial in AD patients [Suh and Checler, 2002].

- Purely the insoluble aggregates of Aβ are neurotoxic and β-folded oligomeric forms of Aβ are accountable for creation of insoluble Aβ-fibrils. Compounds that avert the construction of β-folded forms (β-sheet breakers) will be an apparent option. Various such agents, *e.g.* laminin derivatives, daunomycin, rifampicin, etc. are being worked upon [Bachurin, 2003].

- A scientific research has shown that long term treatment with NSAIDs reduces the threat of AD. Their efficacy is principally because of reduced production of Aβ 1-42 which is more amyloidogenic. Anti-inflammatory activity plays a diminutive role (by preventing microglial activation and discharge of neurotoxins), because all anti-inflammatory agents do not demonstrate advantage in AD. Disappointingly, results from clinical trials with NSAIDs are not promising [Breitner, 2003; McGeer and McGeer, 2007].

- High cholesterol levels favour APP processing by β-secretase pathway that is amyloidogenic, while low cholesterol favours processing by non-amyloidogenic β-secretase pathway. Therefore, High blood cholesterol is constantly a risk factor for AD. Cholesterol reduction inhibits Aβ-formation in brain hippocampus. Furthermore, cholesterol-lowering statins diminish Aβ-levels *in vitro* and *invivo* [Kwak *et al.*, 2000]. Statins as well block interferon-β-induced T-cell activation, therefore diminishes inflammation and has neuroprotective effect.

- Antibodies against Aβ may augment its excretion as observed in animal model of peripheral amyloidosis [Hrncic *et al.*, 2000]. Antibodies can traverse blood-brain-barrier (BBB) and activate microglia to phagocytose Aβ [Dominguez and De, 2002]. Such passive vaccination diminishes the level of Aβ plaques formation in transgenic mouse model of AD.

- Active immunisation with Aβ peptide was well tolerated in phase I trials; however lots of patients developed meningoencephalitis in phase II. These experiments confirmed the formation of antibodies to Aβ in humans, though the trials were stopped [McGeer and McGeer, 2003].

- A synthetic vaccine using an "affitope" a peptide designed to mimic Aβ antigens, stimulates antibody production against this protein without producing a systemic immune response (which would raise some safety concerns). Older vaccines carried the threat of triggering particular T-cells that could direct to meningoencephalitis.

- Aβ in brain exists as membrane connected soluble and aggregated forms. Later two fractions are aggravated in AD. Zn^{2+}precipitates soluble Aβ to form insoluble aggregate and Cu^{2+}as well as Fe^{3+}induces Aβ aggregation at acidic pH. All these ions are constitutively present in neocortex, which is the region most vulnerable to AD. Reduction of biometal causes Aβ deposits to dissolve. Clioquinol averts metal ions from binding to Aβ and stops Aβ deposition in transgenic mouse model for AD [Bush, 2003]. Metal ions catalyze H_2O_2 construction by Aβ, metal chelator act through preventing this process [Huang *et al.*, 1999].

- Neuronal Growth Factor (NGF) is a naturally occurring protein essential for neuron endurance. The gene treatment is injected into the brain area where the cells are damaged in AD patients. It is thought that the resulting constant expression of NGF in the neurons can reinstate their lost function, leading to memory and cognition enhancement. NGF can avert death due to excitotoxin and oxidative stress; however its delivery to brain is an issue as it cannot traverse BBB. Induction of neurotrophic factors is another option. NGF has a survival and growth promoting outcome on basal forebrain neurons [Kordower *et al.*, 1999].

- Co-localization of tau with microtubules is needed for transport of crucial nutrients and organelles and neurotransmitters along the microtubules. When tau protein is hyperphosphorylated, as in AD brain, co-localisation of tau and microtubules cannot take place and the transport is held back. GSK-3β and CDK5 (cyclin-dependent kinases) are implicated in tau-phosphorylation at sites that are phosphorylated in NFTs. These enzymes also accumulate in neurons among NFTs. Over-expression of GSK-3β outcome into tau-hyperphosphorylation. This specifies that inhibition of GSK-3 might be advantageous in AD [Mandelkow, 1999].

- Oestrogen protects against threat of AD in ageing women by enhancing the development of nerve processes and creation of synaptic connections. It averts neuronal death caused due to oxidative stress and excitotoxicity and upregulates cholinergic and monoamine neurotransmitter transport system. In restricted trials, estrogen replacement therapy was defensive against decline in verbal memory [Sherwin, 2002].

- Several disease modifying agents that are under phase II trials include phenserine of Axonyx (AChE I and Aβ formation inhibitor), Alzhemed of Neurochem (Aβ aggregation inhibitor) and clioquinol of Prana Biotechnology (Aβ aggregation inhibitor) [Witt *et al.*, 2004].

- Ultimately, if selective antagonists at NMDA receptors are used, death of cholinergic neurons in basal forebrain can be prevented [Wenk, 2003].

This study was undertaken with the following objectives:

1. **Introduction to System's biology and System level understanding for AD**
2. **To get a concrete knowledge about what is dementia it's mechanism and types and how it is associated with and affecting AD**
3. **Facts about the most common form of dementia among elderly people i.e. AD**
4. **Etiological Hypotheses for AD**

5. Latest curative Advancements for delaying in AD progression and its treatment.

Based upon overall discussion and current scenario of AD worldwide we have planned our study to analyze AD at molecular level. It has been planned to analyze molecular entities involved in AD by applying functional genomics, proteomics, and systems biology approaches. Here we aimed for deciphering the biomolecules such as TFs (Transcription factors), genes and proteins etc. and interactions among them at system level for better understanding of the bioprocesses involved in AD. We applied bottom up approach of systems biology where AD was a system and various biomolecules, biological processes and interactions were its components (**Figure 1.10, Figure 1.11**). Overall thesis has been divided into five chapters which include the current introduction chapter and other following four chapters:

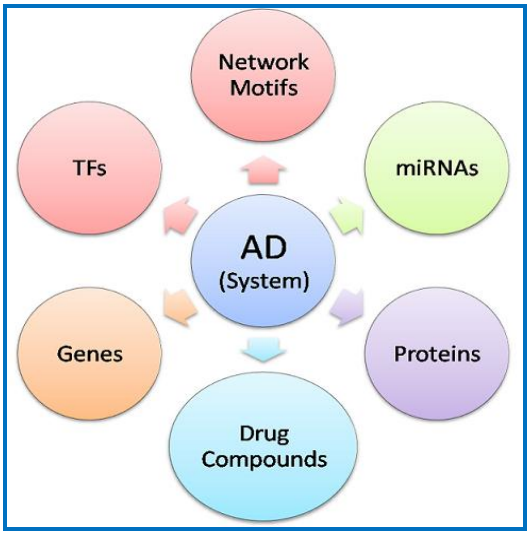


Figure 1.10.AD as a system and its various components

Chapter 2: *In silico* analysis for functional and evolutionary aspects of BACE1 (EC.3.4.23.46) and associated Alzheimer's related proteins

The functional and evolutionary aspects of*β*-secretase-1 (BACE1) and associated Alzheimer's related proteins have been analyzed.

Chapter 3: Enrichment analysis for Alzheimer's disease associated pathways and regulatory patterns with aging and other diseases using Microarray gene expression and network data

Using 3 different sets of Microarray data, (2 of AD and one of AG) genes were found that are common in both AD and AG. Also the association of novel genes and its variants for the interaction and association with other diseases were found that are either directly or indirectly implicated through AD and AG. TFBS (Transcription Factor Binding Sites) were identified as a mediocre biological process for AD. Structural and physico-chemical property analysis for the classes of TFBS revealed the association of biological processes involved with other severe human diseases.Novel information for network motifs such as BiFan, MIM (Multiple input module), and SIM (Single input module)and their close variants will help to improve research into AD.Unique miRNA targets such as LDB2, and DOPEY1 were identified as a regulatory process for Alzheimer's which could be potent targets for the process of gene regulation and inhibitory activities.

Chapter 4: A Genome wide association studiesfor all key genes identified in our earlier research having direct or indirect impact on the metabolic pathway of AD: devlopement of a web resource

A Genome wide association studiesfor all the key genes identified in my earlier research which has direct or indirect impact on the metabolic pathway of AD and developed a web-repository (*ADDGAP*) with SNP genotype data, haplotypes, phosphorylation states and other mutational analysis for all the prime genes.

Chapter 5: Study to test the binding of THC-Δ9-tetrahydrocannabinol & derivatives on Acetylcholine binding protein (AChBP): A Virtual Screening & Molecular Docking Study

Last objective i.e. "To test the binding of THC- Δ9-tetrahydrocannabinol & its derivatives on Acetylcholine binding protein (AChBP): A Virtual Screening & Molecular Docking Study"In this study 3000 molecules were initially selected and screened using the Lamarckian algorithm of AutoDock, from which top 20 lead molecules were chosen showing bond with TRP143 residue. Induced fit docking protocol and TOPKAT was applied to these top 20 lead molecules to determine their binding potential and toxicity. QIKPROP was utilized for ADME properties. Molecular dynamics study was done for the molecule which has shown

maximum affinity during induced fit docking (IFD). It is believed that the proposed set of putative molecules of THC and its derivatives along with most potential candidate molecular agent (S1) provide diagnostic tools to identify appropriate markers at earlier stages of the disease and will be helpful for AD drug development for clinical trials.

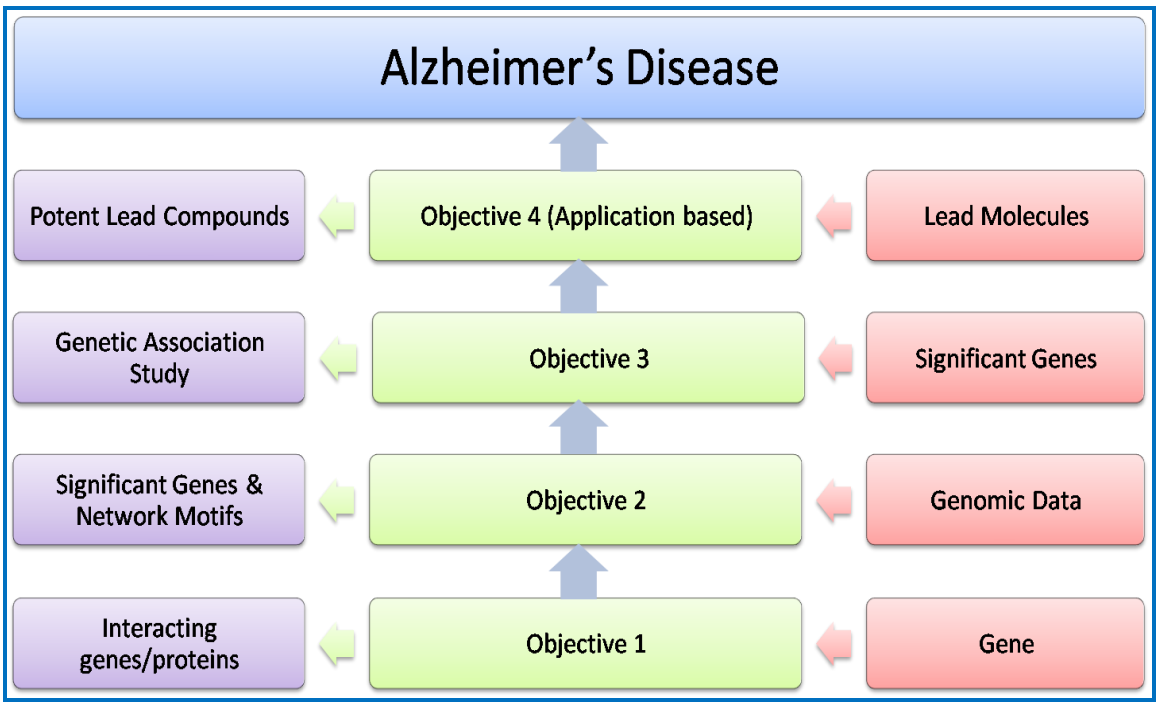


Figure 1.11.Proposed Bottom up approach being applied to deal with AD system

CHAPTER-2

IN SILICO ANALYSIS FOR FUNCTIONAL AND EVOLUTIONARY ASPECTS OF BACE1 AND ASSOCIATED ALZHEIMER'S RELATED PROTEINS

ABSTRACT

AD is a neurodegenerative disorder with unknown etiology leading to severe incapability and ultimately death. It is an irremediable, heterogeneous, progressive brain disease that slowly destroys memory and judgment skills. This fatal disease affects the brain meticulously and is seen as a major public health problem worldwide. As the primary cause and concern for AD is BACE, we studied the PPIs of BACE1 and select some important proteins for further annotation. Multiple sequence alignment (MSA) for homologs to BACE1 has been generated to identify functional motifs. To elucidate the structural, functional and evolutionary conservation of these motifs, we performed analysis through ConSeq and ConSurf servers. We found some specific information and clues of interest based on combinational phylogenetic and pattern finding analysis, which can be further, targeted for the design of new inhibitors and will help in the inhibition of β-secretase. Motifs include the active site of the molecule and are found functionally and evolutionarily conserved. This study will also help in elucidating regulation mechanism of proteins found from interaction studies and will help in the design of inhibitors for their activity leads to Alzheimer directly or indirectly. Design of such inhibitors will be a promising approach for the prevention and cure of AD.

2.1 INTRODUCTION

The most common form of dementia among older people is AD. It is an irremediable, heterogeneous and progressive neurodegenerative disorder, encompassing the corrosion of cognitive functions and behavioral changes [Bertram *et al.*, 2007; Kumar *et al*., 2011]. Although the risk of developing AD increases with age, but it is not a part of normal aging. It is a fatal disease that affects the brain meticulously [Berchtold and Cotman, 1998; Collin*et al.*, 2009; Scodellaro and Pin, 2013]. It is now seen as a major public health problem that is seriously affecting millions of older people and their families worldwide. Based on the disease progression patients are classified as EOAD and LOAD patients. The mean length of life following AD diagnosis is 8 and half years with a range of 1 to 25 years [Lynn *et al*., 2010]. The brain has billions of neurons and to stay healthy, neurons must communicate with each other, carry out metabolism, and repair themselves while AD disrupts all of these essential functions. No one knows what causes AD to begin, but we do know a lot about what happens in the brain once AD initializes or takes hold. The brains of people with AD have an abundance of two anomalous structures; plaques and tangles [Venugopal *et al*., 2008].

The Aβ peptide is resultant of proteolysis from the large precursor molecule called APP. It can undergo proteolytic processing by one of two pathways. Maximum is processed through the non-amyloidogenic pathway, which prevents Aβ formations incase of normal brain. The primary enzymatic cleavage is arbitrated by α-secretase, of which three putative candidates belonging to the family of disintegrin and metalloprotease (ADAM) have been identified: ADAM9, ADAM10 and ADAM17. Cleavage by α-secretase occurs within the Aβ domain, thus preventing the generation and release of the Aβ peptide. Two fragments are released the larger ectodomain (sAPPα) and the smaller carboxy-terminal fragment (C83). APP molecules that are not sliced by the non-amyloidogenic pathway become a substrate for BACE1, discharging an ectodomain (sAPPβ) and retaining the last 99 amino acids of APP (known as C99) within the membrane. The first amino acid of C99 is the first amino acid of Aβ C99 is subsequently cleaved 38-43 amino acids from the amino terminus to release Aβ, by the γ-secretase complex, which is made up of PSEN1 or 2, nicastrin, anterior pharynx defective and PSEN2. This cleavage largely produces Aβ 1-40, and the more amyloidogenic Aβ 1-42 at a ratio of 10:1 (**Figure 2.1**) [Frank *et al.*, 2007]. Aβ fragments come together in clumps to form plaques in vulnerable brain regions. In AD many of these clumps form,

disrupting the work of neurons. The neurons progressively die, which is reflected through memory loss and mental capacity in humans [Vassar *et al.*, 1999].

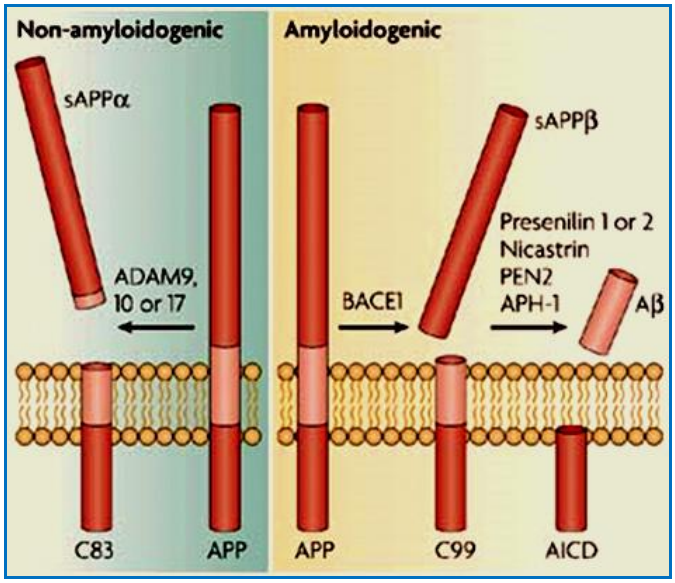


Figure 2.1.Graphic presentation of the A*β* formation.AICD: APP intracellular domain; APH-1: Anterior pharynx defective; PEN2: Presenilin enhancer2, *Adapted from Frank et al., [2007]*

BACE is an enzyme which is encoded by BACE1 gene. This enzyme is also termed as *β*-site APP cleaving enzyme 1 or memapsin-2 (membrane-associated aspartic protease 2) or aspartyl protease 2 (ASP2). A*β* peptide (plaques) is generated through the proteolytic cleavage of APP by *β*- and *γ*-secretases [Oddo *et al.*, 2003; Hong *et al.*, 2004; Hong *et al.*, 2005]. BACE1 is an aspartyl protease, and its practical function is specified cleavage of protein chains during maturation. It is generally found in the endoplasmic reticulum and Golgi apparatus, eradicating crucial proteins in the neural process. Examples of such proteins comprise neuregulin, a protein regulating construction of myelin sheaths near nerve axons and voltage-gated sodium channels (VGSC), which are essential for nerve signal communication [Hunt and Turner, 2009]. BACE1 restrains an active site cleft to clamp together protein chains and a pair of Asp amino acids to slash proteins. Moreover, its structure possesses an extended tail, which binds the enzyme to the membrane surface. The tail confines the enzyme to this surface therefore it does not ramble freely throughout the cell.

BACE1 is a compact globular protein with the molecular weight of 43722.54 Da, and isoelectric point (pI) is 5.04, formed by two domains specifically, residues 47-146 and residues 146-385 [Artimo *et al.*, 2012].Since, BACE1 is an aspartyl protease; it acquires two

lobes with an active site between the lobes. Each lobe yields one aspartate residue of the catalytic dyad. BACE1 is a type-I integral membrane glycoprotein, consists of a 21 residue cleavable signal sequence, a single transmembrane domain of 22 amino acids, a huge ectodomain of 434 amino acids, and a short cytoplasmic tail. There are two conserved aspartic acid residues namely, Asp32 and Asp228 present at the active site inside the conserved motifs of eukaryotes. The two aspartyl residues are in secure propinquity and are commonly found in eukaryotic aspartic proteinase [Artimo P. *et al.,* 2012]. The active site residues assemble in the middle of the cleft along with their surrounding hydrogen bond network [Tatsuno *et al.,* 2001]. BACE1 contains one subunit, which includes the VGSC used for signaling impulses in nerve cells. Troubles with the sodium channel will hinder the passage of neuronal signals throughout the cell, which can eventually lead to seizures in humans. Moreover, sodium channel metabolism is extensively distorted in the brains of AD patients contrasts to normal counterparts of equivalent age [Venugopal *et al*., 2008]. According to the crystal structure of BACE1, the catalytic region is located between the N- and C-terminal lobes, within the substrate binding site in the cleft [Vassar *et al.*, 1999].

By comparing the structures of other mammalian aspartyl proteases with BACE1, it seems that BACE1 has an extra loop, which means the addition of more sub-sites and enlarge the size of the target recognition site [Venugopal *et al.*, 2008]. Some of the significant key structural features of BACE1 are illustrated through its protein sequence (**Figure 2.2**). Cleaved signal peptide followed by the furin cleaved pro-domain is indicated on top of the image. The blue colored underlined fragments are the two active sites of BACE1, which are conserved in all aspartyl proteases. The limits of the transmembrane domain are shown as bold, italicised, and underlined sequences near to the C-terminal end and the core 17 residues are underlined and highlighted in grey. The cytoplasmic tail is shaded and the 'DDISLL' sequence present in that part is the acid cluster dileucine (ACDL) motif, which plays an important role in endocytosis of BACE1. In the cytoplasmic domain, the Cys residues are marked with an '*' to indicate that they are palmitoylated (The reaction of a membrane protein with a fatty acid, especially palmitic acid via the sulphur atom of a cysteine amino acid) [Venugopal *et al.,* 2008].

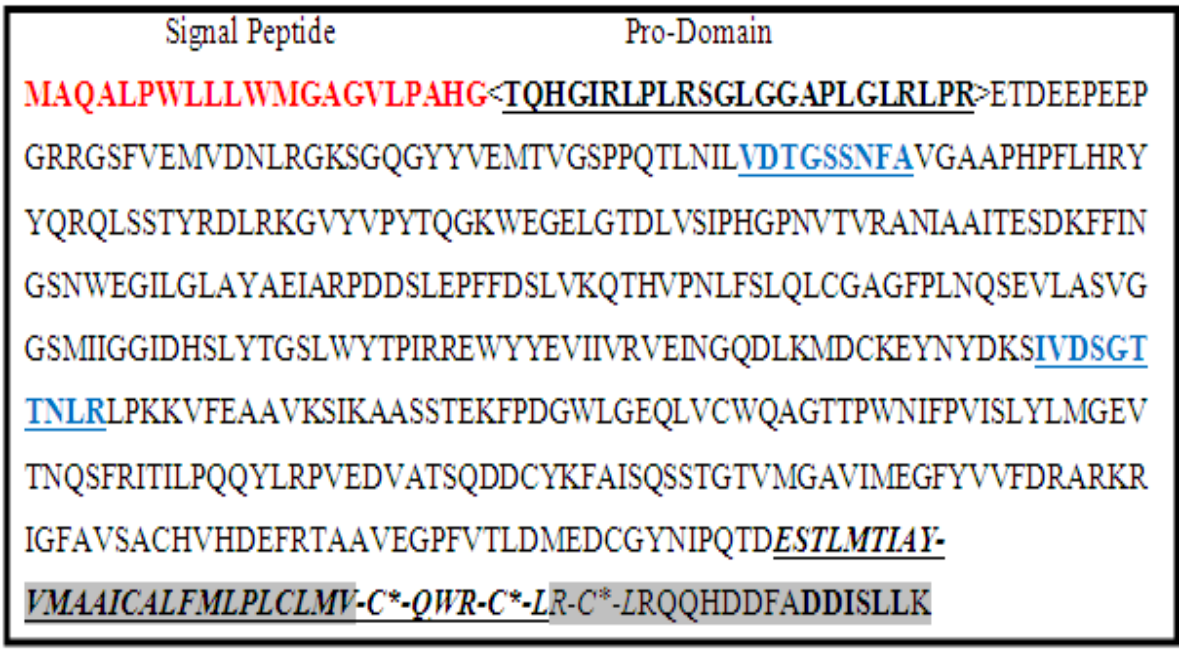


Figure 2.2.Diagrammatic representation of various motifs, active sites and key residues of BACE1, ***Adapted from Venugopal et al., [2008]***

The secondary structure of the human BACE1 monomer restrains two domains with a characteristic aspartic protease fold. The secondary structure also consists of α helices, β sheets, 3/10 helices and arbitrary coils. Distinctively, BACE1 is 14% helical (12 helices) and 40% β sheet (30 strands). Curbing the APP in the region between the BACE1 cleavage site and the membrane interrupts processing, suggesting the secondary structure persuades enzyme-substrate interaction. A lysine side-chain in the pro-segment constructs a salt bridge between two catalytic aspartates, in case of standard aspartyl proteases [Clarke *et al*., 2010]. There is no salt bridge and proline, in the tertiary structure of BACE1, as the corresponding residue does not intermingle with catalytic residues. This divergence in BACE1 divulges that it does not restrain activity, and it assists acceptable folding in the dynamic protease domain. Moreover, the tertiary structure consists of 3 disulfide bonds. The entire ectodomain cysteines are required for complete BACE1 activity and provide an efficient function in disulfide bonding. The connection of disulfide bonds assists enzyme folding. The three potted cysteines on the ectodomain are Cys-278, Cys-330 and Cys-380. Adjoining to Cys-216, Cys-330 and Cys-380 are added sequences with no homology to pepsin family members, which can be seen as loops inside the structure. These loops expand from and enclose the end of the active site cleft. This particular side of the active site cleft binds the N-terminal end of the

substrate, demonstrating the occurrence of an extensive substrate binding pocket [Tatsuno *et al.*, 2001].

There is a need to understand the whole mechanism related to the cleavage of APP by BACE which is the main initiation cause and may leads to severe AD. Also, we need to investigate pre- and post-modifications amongst processes involved regarding this putative event. Main objective of this study is to elucidate functional and evolutionary aspects of BACE1 and its association with related proteins involved in the regulation process of APP cleavage. This study is based on sequential and structural parameters of BACE1 and associated proteins involved to infer new insight for the regulatory processes involved in AD initiation and progression.

2.2 MATERIALS AND METHODS

Several online and offline tools, servers, and databases have been utilized to investigate the structural, functional and evolutionary aspects of BACE. The molecular evolutionary genetics analysis (MEGA 4.0 and sequence alignment) offline tool was used to generate multiple sequence alignment (MSA), to reconstruct phylogenetic tree, estimating rates of molecular evolution, inferring ancestral sequences, and testing evolutionary hypotheses [Nei and Kumar, 2000; Tamura *et al*., 2011]. This version of MEGA is accessible free of charge at '*http://www.megasoftware.net'*. MEGA 4.0 version is a native 32-bit Windows application with multi-threading and multi-user supports; it is also accessible to run in a Linux desktop environment and on Intel based Macintosh computers beneath the parallels program. MSA results were verified using different servers like MSA based on fast Fourier transform (MAFFT) and Multiple sequence comparison by log-expectation(MUSCLE) [Katoh *et al.*, 2002; Edgar, 2004]. We used conserved sequence (CONSEQ), a web server for the identification of structurally and functionally important residues in protein sequences and also used conserved surface (CONSURF, *http://www.consurf.tau.ac.il*), for protein 3D structure conservation analysis andfor high-throughput characterization of functional sites in the protein BACE1 at sequence and structure levels respectively [Ashkenazy *et al.*, 2010].

The search tool for the retrieval of interacting genes/proteins (STRING, *http://string-db.org*/) database is being used to generate PPIs among the sequences. STRING is a database of recognized and predicted protein interactions founded on the sources derived from the

genomic context, co-expression, high-throughput experiments, and previous knowledge. The interactions include direct (physical) and indirect (functional) associations [Szklarczyk *et al.*, 2011]. STRING specializes in three ways: (i) it offers distinctively all-inclusive coverage, with greater than 1000 organisms, 5 million proteins and more than 200 million interactions stored. (ii) It is one of very few sites to hold experimental, predicted and transferred interactions, simultaneously with interactions acquired during text mining. (iii) It incorporates an affluence of accomplice information, for instance protein domains and protein structures, improving its day-to-day assessment for users.STRING shows each functional pathway or term that can be connected to at least one protein in the network. The terms are arranged by their enrichment *P*-value, which is calculated using Hypergeometric test, as enlightened in Rivals *et al.*, [2007]. Different other online servers like molecular interaction database (MINT) and biomolecular object network databank (BOND) were also used for the cross verification and further accuracy of our results [Licata *et al.*, 2012; Lu *et al.*, 2007]. PAL2NAL has also been used for codon wise alignment of nucleotide and their respective protein sequences to confirm exact association among gene and protein sequences [Suyama *et al.*, 2006].

2.3 RESULTS AND DISCUSSIONS

An integrated approach applied in this study is given in **Figure 2.3**, which describes the whole methodology.Functional and evolutionary analysis of proteins is associated with the elucidation of sequential to structural conservation of biological sequences among close as well as distant lineages of various taxonomic groups. To evaluate such conservation first we have applied the traditional approach of phylogenetics. Phylogenetic tree has been reconstructed based on MSA of available complete BACE1 protein sequences from higher organisms, specifically from class mammalian, while some samples from other classes have also been included in the study to understand the evolutionary mechanisms for the BACE1 (**Figure 2.4.A and B**).

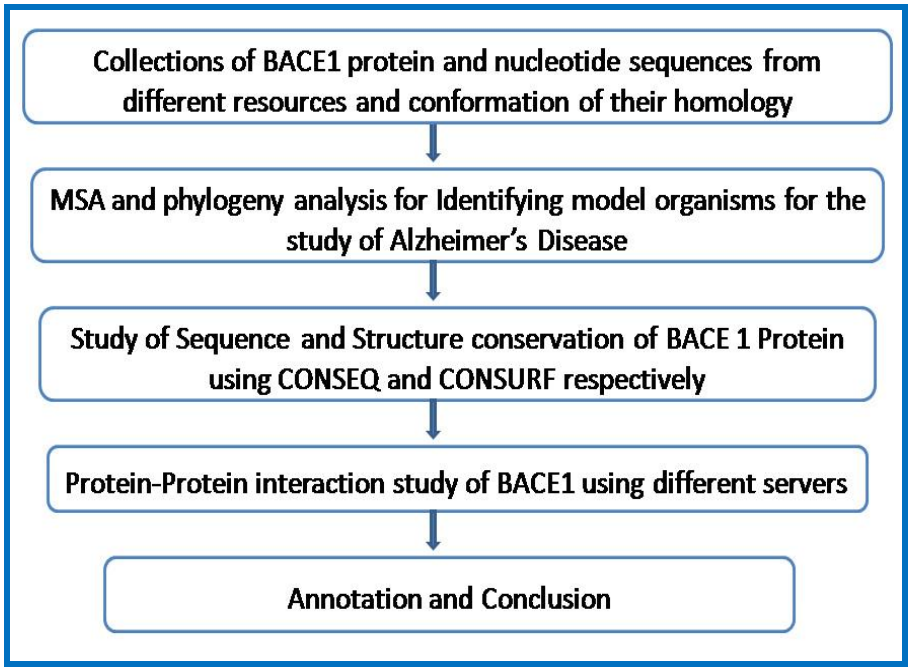
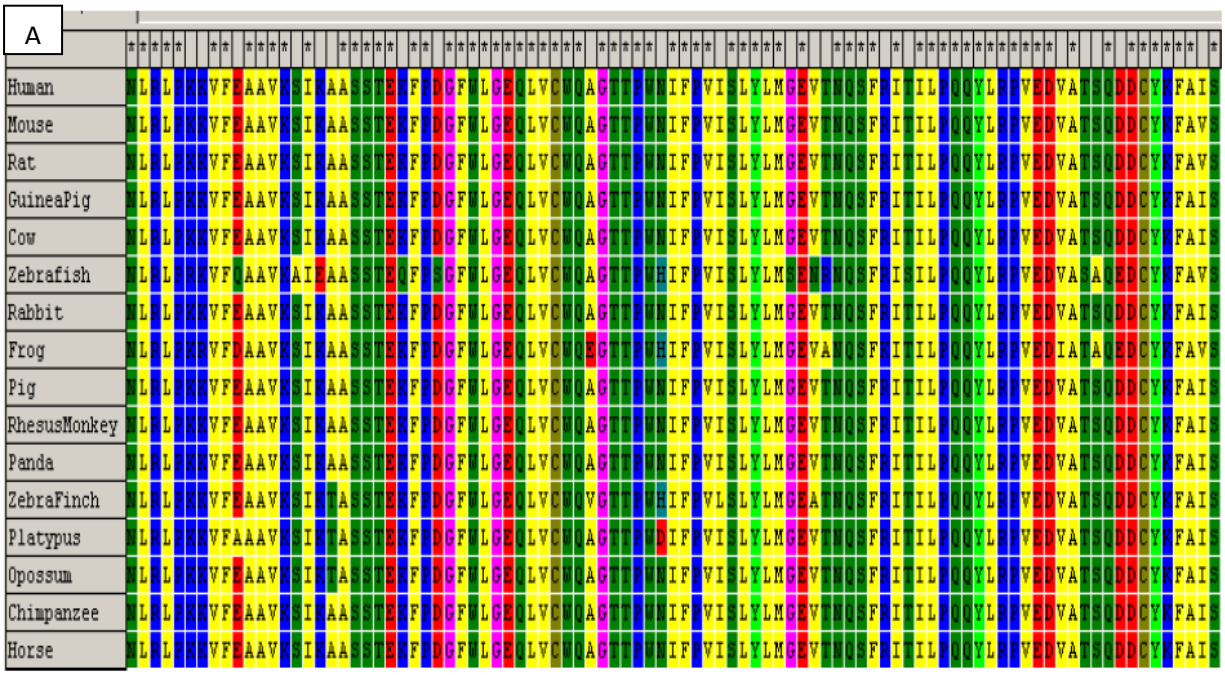


Figure 2.3.Pictographic representation of an integrated approach applied in this study



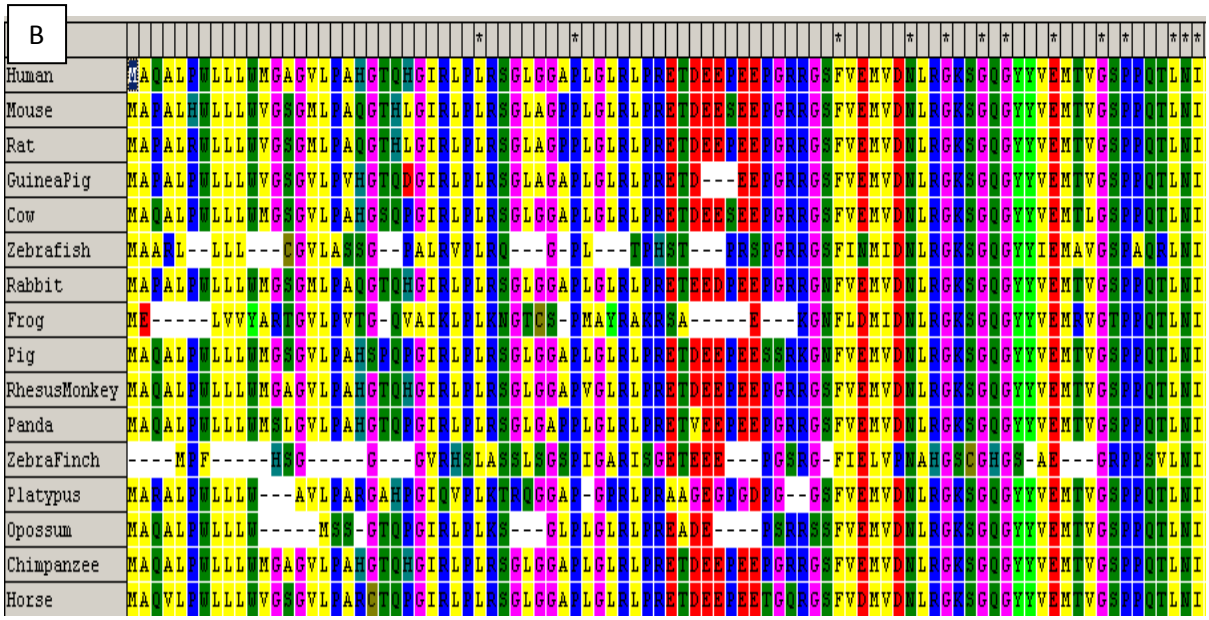


Figure 2.4.Parts of MSA of AA sequences (A. Conserved Sites, B. Mutation Sites) of BACE1 for various taxonomic classes and orders, *AA: Amino Acids

$$\textbf{Feature similarity Score (FSS)} = \frac{\textbf{Constant Sites} + \textbf{Singleton Sites}}{\textbf{Total no. of sequences}}$$

$$\textit{In this case,} \quad \textbf{FSS} = \frac{307 + 127}{501} = 86.6\%$$

According to the FSS score the sequences shows strong similarity hence we choose maximum parsimony (MP) method for the tree generation using MEGA software. We also try different methods like Neighbor joining, Maximum likelihood etc. and got the similar kind of tree. Phylogenetic tree is being reconstructed for BACE1 protein sequences for following species: human, mouse, rat, guinea pig, cow, rabbit, pig, rhesus monkey, chimpanzee, horse, panda, platypus, opossum, zebra fish, zebra finch and frog (**Figure 2.4**). Phylogenetic analysis revealed flexible conservation of BACE1 among various lineages where all mammalian, rodents and other class support traditional phylogeny. Interesting outcome is at order level where all available sequences belong to 11 different orders. This phylogeny is based on BACE1 protein sequences and supports the outlier positions of various

orders like Anura and Cypriniformes in close association with Passeriformes. Another interesting finding is the separate clade of rodentia with plausible leave out position of guinea pig. Reason behind this might be the little variation among rodentia as one of the active site of BACE1 is absent in rat and mouse, which take apart them in to separate clade from guinea pig and other mammals (**Figure 2.5**). Based on this analysis rat and mouse are hampered against guinea pig as model organism to perform experiments which can indirectly benefit human participation associated with AD control process. It is proposed that evolutionarily BACE1 sequence is more conserved in human and guinea pig than human and mouse and rat. Presence of two active sites in human and guinea pig BACE1 sequences make these sequences more robust and prone to experimental specifications.

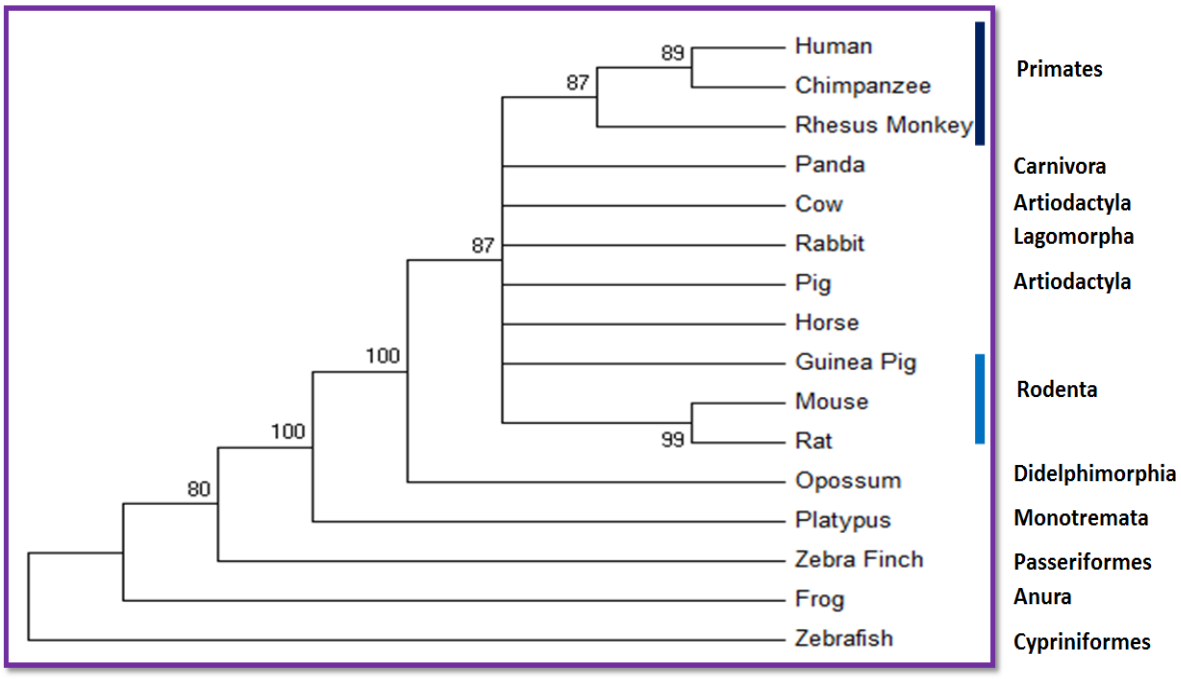


Figure 2.5.Phylogenetic tree using Maximum Parsimony [Note: Orders are mentioned against respective species or group of species.]

CONSEQ and CONSURF provide prediction of sequence and structure conservation with respect to their functional and structural residues on the basis of comparison among homologous sequences. It also provide information regarding the region of the protein residues in their respective structures weather residues belong to exposed or buried part of the structure. It helps in identifying the conservation of active sites and other important areas. In

CONSEQ and CONSURF results (**Figures 2.6.A, 2.6.B and 2.7**), the residues which are structurally, functionally and evolutionarily conserved are represented by dark maroon color. Same kind of conservation has been observed among sequence and structure residues. The conserved residues in 3D structure are indicated by red circles. Conserved residues of the protein sequence and structure correspond to the combination of exposed and buried residues with their respective structural and functional role and scattered all over the 3D structure in CONSURF and CONSEQ results respectively (**Figures 2.6.A, B and 2.7**). There have been several studies on the conservation of interacting residues among interacting protein partners [Mintseris and Weng 2005; Cui*et al.*, 2008; Von Eichborn *et al.*, 2010]. These studies are based on sequence and structure conservation, while structure conservation among interacting partners is having some advantage over sequential interactions. These conservations can be related to the functional and evolutionary aspects of interacting proteins. We performed similar analysis to reveal such kind of conservation based upon homologous partners and found that our results are in agreement with recent studies for the conservation and co-evolution of amino acid residues among interacting protein partners (**Figures 2.6.A, B and 2.7**). Different rate of evolution have been observed among different kind of complexes depending up on the selection pressure working on their residues. However conservation score was found maximum among core residues [Mintseris and Weng 2005], and our structural data with functional and evolutionary conservation are in agreement with this hypothesis (**Figure 2.7**).

Figure 2.6.A. CONSEQ result for Human BACE1

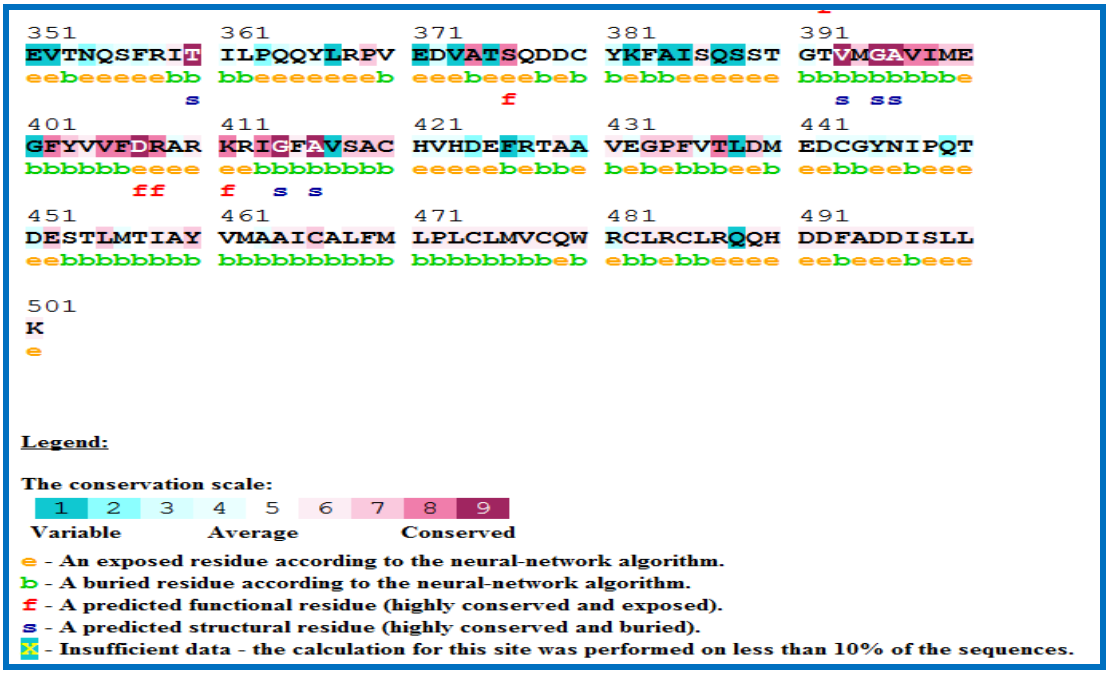


Figure 2.6.B.CONSEQ result for Human BACE1 with legend as conservation scoring color system. Same conservation system is applied on CONSURF results also. Active sites and surrounding residues have high to low conservation respectively

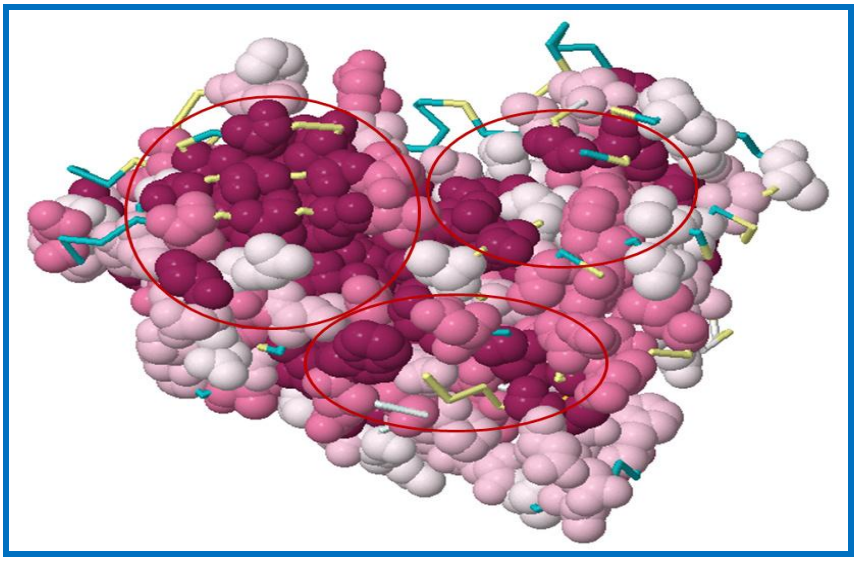


Figure 2.7.CONSURF result for Human BACE1. Conservation scale and score is as given in CONSEQ results. Results are in agreement with CONSEQ results for the structural andfunctional conservation of amino acid residues and group of residues for structural and functional motifs (highlighted circles/ellipses)

Tabulated variants were taken from NCBI (**Table 2.1**), first 6 variants are of human BACE1 as denoted by (H) and the other two variants are of *Mus musculus* as denoted by (M). Both nucleotide and their corresponding protein sequences were collected and then codon based multiple alignment was done using PAL2NAL server [Suyama *et al.*, 2006]. First we took the six variants of human transcripts and data analysis was done using the above mentioned server. We found that the first 4 variants are more conserved than the last two variants. The most significant finding of our study is that the first active site of the BACE1 found to be from position 90 to 101 (‘*VDTGSSNFA*’) is totally absent in the last two variants. For more detailed study we took the first 4 variants and analyzed them separately by the same server. No mutation was found in the variant 1. The two variants of Mouse and the six variants of human when analyzed using the PAL2NAL server, the two variants of Mouse were more similar with the first four variants of human than the last two variants of human. When the first four sequences of human and the two sequences of Mouse are altogether analyzed, we found that the 2nd active site of the BACE1 found to be from 286 to 296 (‘*IVDSGTTNLR*') is absent in the 2nd variant of Mouse due to mutation and this variation is

making significant difference in the topology of two type of variants. This difference might be used effectively in genomic studies.

Table2.1 Details of *Homo sapiens*or Human and *Mus musculus*or Mouse BACE1 Sequences

Nucleotide	mRNA Length	ORF Length	Protein Length
Transcript variant a (H)	5864	1503	501
Transcript variant b (H)	5789	1428	476
Transcript variant c (H)	5732	1371	457
Transcript variant d (H)	5657	1296	432
Transcript variant e (H)	5226	1203	401
Transcript variant f (H)	5151	1128`	376
Transcript variant a (M)	4194	1503	501
Transcript variant b (M)	4092	1401	467

While analysing and aligning the human BACE1 and its various isoforms manually, we have identified that major sites of deletions among BACE1 isoforms B, C, and D are between the positions 146 to 214 with reference to BACE1 precursor where as no deletion in isoform A has been found. Another interesting finding was the mutation at –21 sites (C terminal) in all the isoforms where Cys has been substituted by Arg amino acid. This is neighbor to another consecutive Cys residue present next to this substituted position and might play a crucial role for the activity of the enzyme. Conservation of intracellular motif *DDISLL*among all isoforms and in other species was found. *DDISLL* sequence also termed as ACDL motif, it is recognized that two leucines (residues) inside this motif which appear to be indispensable for BACE1 endocytosis. The substitute of these residues with alanines resulted in the accretion of the mutant protease on cell surface. Dileucine is a well recognized endocytotic signal, and it appears to function autonomous of the GGA/ACDL interaction [He

et al., 2003; He *et al.*, 2005]. It is believed that the proper use of this conserved motif could be helpful in regulating transport of ions and their recycling which can facilitate the communication better amongst neurons.

There are several proteins available which are directly or indirectly associated with the AD regulation process. To identify crucial proteins which are directly associated with BACE1 and also have some kind of PPIs, we have generated PPI network. STRING database has been used for this purpose while interaction was also confirmed using other available resources for identified proteins. Some important proteins interacting with BACE1 have been selected for further annotation studies (**Table 2.2**). There are different color representations depending on their interactions category with BACE1. Like magenta colored interactions (for instance, between BACE1 and APP in **Figure 2.7**) is denoted for experimentally verified connections, sky-blue color is denoted for interactions those are assembled from different databases, light green colored interactions stands for text mining and so on.Identified proteins having important association with AD regulation processes are APP,PSEN1, Furin Precursor (FURIN), Nicastrin Precursor (NCSTN), Low density Lipo-protein Receptor- related Protein 1 Precursor (LRP1), and Phospholipid Scramblase (PLSCR). Golgi associated, gamma adaptin ear (GGA1), reticulon 3 (RTN3), sortilin-related receptor (SORL1), and amyloid beta (A4) precursor protein-binding (APBB1) have also been found connected with BACE1 and its associated proteins but due to lack of experimental information they have been ranked in second group for further studies (**Table 2.2**). These proteins may also be promising as they are not too far in functional and evolutionary facet in comparison to other proteins in the generated interaction network (**Figure 2.7**).

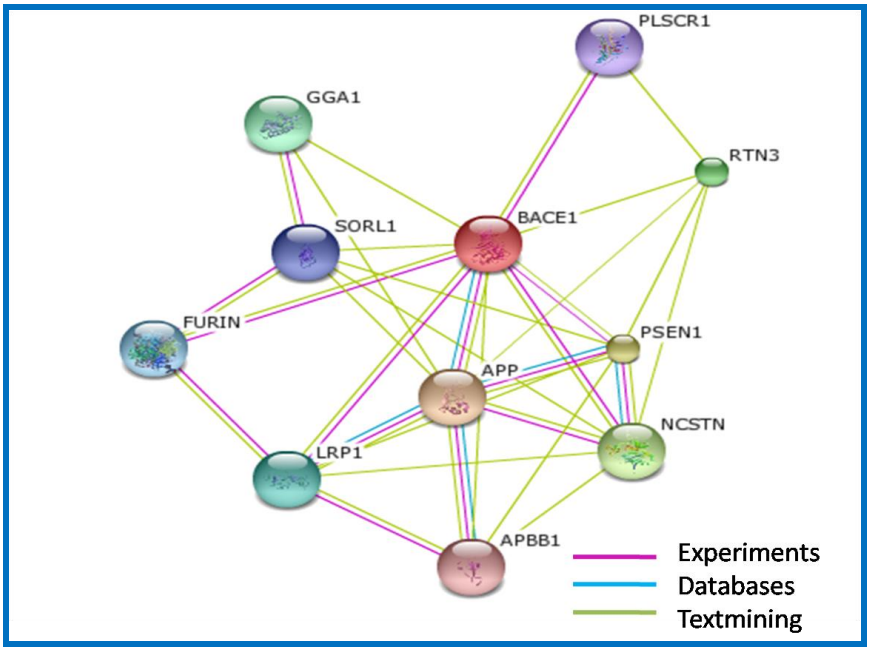


Figure 2.7.PPI network of BACE1, 10 close interacting partners were selected

Only proteins having some experimental information for their involvement with BACE1 other than database and text-based connectivity have been selected. Further referential annotation has been done for these proteins to elaborate their specific properties related to AD or BACE1 (**Table 2.2**).These proteins on further exploration of their interaction mechanism will be helpful in providing further clues for the APP cleavage mechanism, identification of catalytic sites, and discovering drug like molecules and inhibitors for targeting BACE1 regulation mechanism. Functional annotation will help the researchers to further enhance the activities for the betterment of existing molecular and biomedical technologies and will also help to plan for the new ones.

Table 2.2 BACE1 Interacting proteins identified through protein-protein interaction network. Further functional information compiled from literature and other annotations

Protein	Description	Information	Association with BACE1
APP	Amyloid Precursor Protein	N-APP binds TNFRSF21 triggering caspase activation and degeneration of both neuronal cell bodies (via caspase-3) and axons (via	Cleaved by BACE1 to form Aβ plaques, which is the main cause for AD [Johnston *et al.*, 2006; Zhang*et al.*, 2011].

		caspase-6).	
PSEN1	Presenilin-1(PS-1)	Play a role in intracellular signaling and gene expression or in linking chromatin to the nuclear membrane. Stimulates cell-cell adhesion though its association with the E-cadherin/catenin complex.	Aβ is derived from the APP after cleavage by the recently identified BACE1 and the putative γ-secretase complex containing PS1. In an attempt to develop a functional secretase enzymatic assay they demonstrate a direct binding between BACE1 and PS1 [Hebert *et al.*, 2003].
NCSTN	Nicastrin Precursor	Probably represents a stabilizing cofactor required for the assembly of the γ-secretase complex.	BACE1 binds to nicastrin, a component of γ-secretase complexes *in vitro*, and that NCSTN activates β-secretase activity [Hattori *et al.*, 2002].
LRP1	Low density Lipo-protein Receptor-related Protein 1 Precursor	Low density lipoprotein receptor-related protein associated protein 1; Interacts with LRP1/alpha-2-macroglobulin receptor and glycoprotein 330.	Interaction between LPR1 and BACE1 may enhance the beta secretase cleavage of APP and increases Aβ generation [Von Arnim *et al.*, 2005].
FURIN	Furin Precursor	Furin is likely to represent the ubiquitous endoprotease activity. Within constitutive secretory pathways and capable of cleavage at the RX (K/R) R consensus motif.	Converts the pro-BACE to mature BACE by cleaves out the propeptide present in the pro-BACE [Haniu *et al.*, 2000].
PLSCR1	Phospholipid Scramblase(PL Scramblase 1)	Mediate accelerated ATP-independent bidirectional trans-bilayer migration of phospholipids upon binding calcium ions that result in a loss of phospholipid asymmetry in the plasma membrane.	Both proteins share a common trafficking pathway in cells [Kametaka *et al.*, 2003].
GGA1	Golgi associated, gamma adaptin ear	Containing, ADP-ribosylation factor (ARF) binding protein 1; Plays a role in protein sorting and trafficking between the trans-Golgi	Small interfering RNA (siRNA) knockdown studies of GGA1, GGA2, and GGA3 signify an astounding degree of specificity headed for GGA1, suggesting

		network (TGN) and endosomes. Arbitrates the ARF- dependent recruitment of clathrin to the TGN and binds ubiquitinated proteins and membrane cargo molecules with a cytosolic acidic cluster-dileucine (AC-LL) motif.	that it is a candidate regulator of LR11 traffic. GGA1 is necessary for both LR11 and BACE1 inflection of APP processing to Aβ. Mutagenesis of BACE1 serine 498 to alanines enhances BACE1 targeting to LR11-positive compartments and invalidates LR11-mediated diminution of Aβ. GGA1 assists LR11 endocytic traffic and that LR11 modulates Aβ levels by helping APP traffic to the endocytic recycling section [Herskowitz *et al.*, 2012].
RTN3	reticulon 3	May be involved in membrane trafficking in the early on secretory pathway. Inhibits BACE1 activity and APP processing. May persuade caspase-8 cascade and apoptosis. May help in BCL2 translocation to the mitochondria upon endoplasmic reticulum stress. In case of enteroviruses infection, RTN3 may be implicated in the viral replication or pathogenesis.	Reticulon (RTN) proteins for instance RTN3 and RTN4-B/C interact with BACE1 and inhibit its β-cleavage activity. RTN3 and RTN4-B/C generally disseminated in nonraft domains where they appear to control BACE1 [He *et al.* 2004; Murayama *et al.* 2005].
SORL1	sortilin-related receptor	Probable to be a multifunctional endocytic receptor, that may be concerned in the uptake of lipoproteins and of proteases. Binds LDL, the major cholesterol-carrying lipoprotein of plasma, and transports it into cells by endocytosis. Binds the receptor-associated protein (RAP). Possibly will play a role in cell-cell interaction.	Two haplotypes in the sortilin-related receptor (*SORL1*) gene associated with LOAD were identified [Meng *et al.*, 2007; Reitz *et al.*, 2011]. *SORL1* is involved in trafficking of APP from the cell surface to the golgi-endoplasmic reticulum complex and γ-secretase processing of APP [Lee *et al.*, 2008], also in line with the ACH.
APBB1	amyloid beta	Adapter protein that forms a	Genetic depletion of

	(A4) precursor protein-binding	transcriptionally active complex with the γ-secretase-derived APP intracellular domain. Plays an essential role in the response to DNA damage by translocating to the nucleus and inducing apoptosis. May act by distinctively distinguishing and binding histone H2AX phosphorylated on 'Tyr-142' (H2AXY142ph) at double-strand breaks (DSBs), recruiting other pro-apoptosis factors for example MAPK8/JNK1. Essential for histone H4 acetylation at double-strand breaks (DSBs).	APBB1was shown to impair function in learning and memory tasks, with a prominent discrepancy in reversal spatial learning [Wang *et al.*, 2004]. APBB1 and APBB2 interact with the cytoplasmic tail of APP and were verified to potently stimulate transcription [Csiszar *et al.*, 2013]. Genetic diminution of both APBB1 and APBB2 outcomes in cortical dysplasia [Guenette *et al.*, 2006].

2.4 CONCLUSION

Discovery of BACE1 as integral membrane aspartyl protease envisaged regulation mechanism for the AD. There is an urgent need to further explore the mechanisms underlying various processes and pathwaysassociated with the AD. Here, in this chapter we performed functional and evolutionary analysis of BACE1 and identified its associated proteins related to AD. Phylogenetic results gave new direction to the position of lineages based on BACE1 evolutionary tree. Based upon combination of phylogenetic and functional motif analysis, we suggest that "guinea pig" has an advantage over traditional rat and mouse as model organism, because it is more close to mammalian class, even though all three belongs to class rodentia. Identification of Cys-Arg mutation at C-terminal will explore plausible secondary cause of variations among human BACE1 isoforms. Our functional and evolutionary conservation studies on sequential and structural data are in agreement to some recent hypothesis on evolution of residues among interacting protein partners and revealed new aspects and directions for evolutionary conservation of BACE1 and associated protein partners. Absolute knowledge of all direct and indirect interactions among proteins in a given cell would represent a significant milestone towards an all-inclusive depiction of cellular mechanisms and functions. From PPI studies, we have found five additional proteins closely associated with BACE1 that can be helpful in elucidation of regulation mechanism of BACE1. These identified BACE1 associated proteins (also known as interacting partners), will also be helpful in designing new drug targets and inhibitors for prevention and cure of AD.

CHAPTER-3

Enrichment analysis for alzheimer's disease associated pathways and regulatory patterns with aging and other diseases using microarray gene expression and network data

ABSTRACT

AD is a far-reaching condition, one that rips through not only the lives of those who have individually suffered through the diagnosis, but the lives of family members, friends and caretakers who brush up against the illness as well. It is one of the most common age-associated neurodegenerative disorders, affects millions of people worldwide. Due to its polygenic nature, AD is believed to be caused not by defects in single genes, but by variations in a large number of genes and their complex interactions, which ultimately contribute to the broad spectrum of disease phenotypes. Extraction of insights and knowledge from microarray and network data will lead to a better understanding of complex diseases. The present study aimed to identify genes with differential topology and their further association with other biological processes that regulate causative factors for AD, AG and other diseases. Our analysis revealed a common sharing of important biological processes and putative candidate genes among AD and AG. Some significant novel genes and other variants for various biological processes have been reported as being associated with AD, AG, and other diseases, and these could be implicated in biochemical events leading to AD from AG through pathways, interactions, and associations. Novel information for network motifs such as BiFan, MIM, and SIM and their close variants has also been discovered and this implicit information will help to improve research into AD. Ten major classes for TFs have been identified in our data, where hundreds of TFBS patterns are being found associated with AD. Structural and physico-chemical properties analysis for these TFBS classes revealed association of biological processes involved with other severe human diseases. Nucleosomes and linkers positional information could provide insights into key cellular processes. Unique miRNA (micro RNA) targets were identified as another regulatory process for AD. The association of novel genes and variants of existing genes have also been explored for their interaction and association with other diseases that are either directly or indirectly implicated through AD.

3.1 INTRODUCTION

AD affects millions of people worldwide and is one of the most common age associated neurodegenerative disorders [Chou, 2004]. AD is characterized by a progressive decline in memory associated with other cognitive deficits: judgment, abstraction, language, attention and visuoconstructive abilities [Hommet *et al.*, 2011]. It is polygenic in nature and involves large number of variations in genes and their critical interactions that lead to this disease [Ray *et al.*, 2008]. Extracellular Aβ plaques, intracellular NFTs, cerebrovascular amyloid, dystrophic neuritis and loss of synaptic connections are standard markers of neurodegeneration in AD [Tarawneh and Holtzman, 2010; Panigrahi and Singh, 2012]. It is found that the aberrant toxic Aβ aggregation causes synaptic dysfunction, oxidative stress, ionic dyshomeostasis, tau aggregation, and apoptosis [Hardy and Selkoe, 2002]. The cytotoxic Aβ fibril is one of the coherent contenders for causing the starting damage to neurons in AD [Carter and Chou, 1998]. Another noticeable fact is that AD does not affect entire brain at once as middle temporal gyrus (MTG) shows early AD pathology compared to the other regions of brain like entorhinal cortex (EC), hippocampus (HIP), and posterior cingulate cortex (PCC) [Ray and Zhang, 2010]. Various techniques have been used for the analysis of gene expression data associated with neurodegenerative disorders like AD. DNA microarrays constitute a contemporary tool for hypothesis generation [Newman and Weiner, 2005]. Using this technique large amount of gene expression data have been easily accumulated. Now the challenging task is to extract valuable biological information from this immense amount of data [Kong *et al.*, 2011]. It is done either by identifying 'critical genes' which might singlehandedly produce a biological effect or by finding patterns in the list that point to an underlying biological process. Then annotating each gene on the list and looking for groups of genes that share a particular characteristic [Stekel, 2003]. This shared or interacting nature of genes is crucial for the analysis of complex polygenic disease like AD.

The development of microarray technology provides researchers, a tool that measures the expression levels of thousands of genes at once, offering possible molecular clues regarding mechanisms underlying the disease pathophysiology [Huang *et al.*, 2009]. The Gene Ontology (GO) consortium has brought systematic order to the field of gene annotation by pre-categorizing genes by biological process, molecular function, and cellular component [Ashburner *et al.*, 2000].The focus of bioinformatics development has now shifted from

understanding networks encoded by model species to understand the networks underlying human disease, by the increase of the various kinds of human interaction data [Kann, 2007]. Combining these network-based disease studies with the original analyses of network properties in model organisms may override the conclusion that genes associated with a particular phenotype or function, including the progression of disease, are not randomly positioned in the network. Rather, they tend to exhibit specific patterns such as high connectivity, cluster together, and occur in central locations of the network as hub genes. There are evidences based on network property values where it has been concluded that overall degree or average distance to one another tends to lie between essential and nonessential genes, and provide patterns for the inclusion of all available interacting partners for a specific biological network [Said *et al.*, 2004; Shachar *et al.*, 2008]. Network motifs play a central role in the identification and analysis of such specific patterns in biological networks and yield significant new insights into understanding complex biological processes involved in the intricate human disease such as AD.

Recently some studies have been proposed which have commonalities in methods, while objectives are discrete [Miller *et al.*, 2008; Ray and Zhang, 2010]. Additionally approaches as well as findings are novel in all of these studies. Also there are scientific proofs and suggestions for common features between AD and other diseases to establish a link and common pathogenic mechanisms for the treatment strategies [Gotz *et al.*, 2009]. Additionally several efforts have been made recently to investigate AD by using myriad computational approaches [Chou, 2004; Chou, 2005; Wei *et al.*, 2005; Gu *et al.*, 2009]. This overall coordination of studies designates the functional commonalities for the complex mechanisms involved in AD and its links to other diseases and suggest common prediction practices and treatment strategies. Objective of this study was to find out the relationship between one of the most threatening disease, AD with the normal AG or in other words the impact of AG factor on this disease. This study also identifies genes with differential topology and their further association with other biological processes regulating causing factors for AD, AG and other diseases. This analysis has been performed by applying integrative approach on various aspects of molecular data, markers, and networks to study a complex disease AD. This analysis has implications and applications for early AD detection and novel marker identification for AD.

3.2. MATERIALS AND METHODS

3.2.1 Data

Three separate microarray data sets were used in this study: one consists of microarrays assessing gene expression from the CA1 region of the hippocampus from 31 individuals, comprising nine controls, seven with incipient AD, eight with moderate AD, and seven with severe AD [Blalock *et al.*, 2004]. Second data set is of 30 microarrays representing a study of the effects of aging on frontal lobe gene expression of individuals who died of natural causes between the ages of 26 and 106 [Lu *et al.*, 2004]. The AD study used Affymetrix HG-U133A chips containing 22,283 probe sets, and the aging study used HG-U95A chips with 12,625 probe sets. In addition to these data sets, one more dataset used for comparative analysis consists of 14 normal controls and 14 AD affected samples obtained from Gene Expression Omnibus (GEO Accession Number: GDS2601) [Maes *et al.*, 2007]. Additionally to incorporate network profile and network motif studies, network/pathways data associated with AD has been utilized from Kyoto Encyclopedia of Genes and Genomes(KEGG, *http://www.genome.jp/kegg*) [Kanehisa and Goto, 2000] and other popular interaction resources.

3.2.2 Data pre-processing

Data filtering or normalization can reduce the dataset by removing poor or questionable data, data deemed uninteresting or irrelevant to the analysis. In this study normalization of the datasets was done using one of the tool of TM4 called Microarray Data Analysis System (MIDAS) and normalization modules used were locally weighted linear regression [Cleveland and Devlin, 1988] and total intensity normalization [Yang *et al.*, 2002]. The factors considered in the filtration of dataset include low-intensity cutoff, intensitydependent *Z*-score cutoffs and replicate consistency trimming, creating a highly customizable method for preparing expression data for subsequent comparison and analysis. MIDAS provides scatter plots that illustrate the effects of each algorithm on the data [Saeed *et al.*, 2003]. Preprocessed data was subjected to differential gene expression followed by manual scrutinization.

3.2.3 Differential gene expression

Differential expression of probe sets for each dataset was performed using significance analysis of microarrays (SAM) [Tusher *et al.*, 2001]. This supervised learning software for genomic expression data mining determines differentially expressed genes in a two class experiment based on the statistical analysis of modified gene specific *t*-test [Dziuda, 2010]. The expressed genes of control and various AD stages (incipient, moderate, and severe) have been considered and compared for the identification of final set of differentially expressedgenes in AG and AD. After the process of normalization inclusive of all available methods and then scrutiny of genes based on the differential gene expression, and comparative analysis for both AD and AG individually, amongst all three datasets we select 62 and 658 numbers of informative or important genes for AD and AG, respectively.

3.2.4 Clustering of co-expressed genes

After selection of differentially expressed genes from all the datasets in this study, the genes were clustered based on the expression level to find the co-expressed gene clusters. Multi Experiment Viewer (MEV) package from TM4 was used for clustering of microarray data, using Euclidean distance metrics [Deza and Deza, 2009] and average linkage clustering algorithms [Johnson, 1967]. WebGestalt (WEB-based GEne SeT AnaLysis Toolkit), was used to analyze functional, genomic, proteomic, and large-scale genetic studies from which large number of gene lists (e.g. differentially expressed gene sets, co-expressed gene sets etc.) were generated [Zhang *et al.*, 2005;Duncan *et al.*, 2010]. WebGestalt incorporates information from different public resources and provides an easy way for biologists to make sense out of gene lists. Self-Organizing Maps (SOM) or Kohonen network [Kohonen, 1982a, 1982b] was used to generate distinct clusters based upon the functional parameter and expression profile of each cluster independently. SOM is a visualization tool based on the unsupervised artificial neural network (ANN) as a learning algorithm. SOM is mostly used to cluster either genes or biological samples. In case of gene expression analysis, this method is used to group genes into clusters of similar expression profiles. We applied SOM with the input layer of N inputs representing the original N variables (here, AD and AG common genes), and the output layer being the grid of neurons corresponding to cluster prototypes. Principal Component Analysis (PCA) [Hotelling, 1933] was also performed to determine the

relationship between each modules of AD gene set and their phenotypic assessments for all the three data types.

3.2.5 Hypergeometric distribution and association of ranked genes

Ranked list of genes is also informative while dealing with multiple (polygenic) processes which are required in AD case as stated earlier. To implement this aspect GOrilla was used that identifies enriched GO terms in ranked lists of genes [Eden *et al.*, 2009]. GOrilla computes an exact *p*-value for the observed enrichment, taking threshold multiple testing into account without the need for simulations. BothWebGestalt and GOrilla uses the same statistical approach that is hyper-geometric distribution (HGD) for the enrichment analysis or statistical significance testing, while WebGestalt additionally uses the Fisher's exact test for two independent gene sets [Sealfon *et al.*, 2006; Eden *et al.*, 2007]. In a group of *N* number of genes there are *K* genes which are associated with a particular GO term. If we take sample *n* genes out of *N*, then we found *k* associated genes and the probability of obtaining *k* or more GO term associated genes in a sample of n could be calculated via HGD.

$$p-value = 1-\sum_{i=0}^{k-1} f_{HG}(i;N,K,n) = 1-\sum_{i=0}^{k-1}\frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

GOrilla employs a flexible threshold statistical approaches to discover GO terms that are significantly enriched at the top of a ranked gene list. It applies a variant of standard HGD based on a complete theoretical characterization of the underlying distribution called mHG (minimum hypergeometric) [Eden *et al.*, 2007]. In most cases a fixed threshold (*n*) is not known but rather a ranking of all the elements, to find n which minimizes the HGD. Formally, if a ranked gene list: $g_1,\ldots,g_N$ is provided in place of a target set, we define a label vector $\lambda = \lambda_1,\ldots,\lambda_N \varepsilon \{0, 1\}N$ according to the association of the ranked genes to the given GO term, λ_i=1 if g_i is associated with the term [Eden *et al.*, 2007]. The mHG score is then defined as:

$$mHG(\lambda) = \min_{1 \le n \le N} (HGT(N, B, n, b_n(\lambda)))$$

$$Where, b_n(\lambda) = \sum_{i=1}^{n} \lambda_i$$

3.2.6 Topological overlap between co-expressed networks and other associated factors

For analyzing topological overlap (TO) between co-expressed networks, method developed by Ray and Zhang (2010) was implemented. We identified genes with a topological difference (i.e. low TO) between co-expressed networks, where the actual amount of similarity between two neighborhoods (brain regions such as EC, HIP, MTG, and PCC, and their co-expressed networks) is only 5%. Random additions or deletion of links to the original network while keeping the degree of the genes equal to the original network using *t* statisticswere made to perform comparisons against 1000 random networks to assess the significance of the zero TO genes. The significance values (*p*-values) were calculated (with 999 degrees of freedom) using the formula and method given in Ray and Zhang (2010).

3.2.7 Prioritization of gene candidates with molecular triangulation

Presence of large molecular networks that encompass multiple genes harboring disease-related genetic variation, and are available in a computer-accessible form, motivated to prioritize candidate genes. Distance-dependent decay function was implicated. Let each seed node project its evidence value to its immediate neighbor nodes, such that the secondary-evidence value decays with the distance from the seed node. Then the secondary-evidence value is calculated in the following way:

$$E(u) = \sum_{v \varepsilon B} E_p(v) f(d_{uv})$$

Where,*E(u)* is the secondary evidence for node *u*, *Ep(v)* is the primary evidence for seed node *v*, *B* is the set of all seed nodes, d_{uv} is thedistance between nodes *u* and *v*, and finally, f is a distancedependent decay function [Krauthammer *et al.*, 2004].

STRING [Szklarczyk *et al.*, 2011], oPOSSUM [Ho *et al.*, 2007], and JASPAR [Sandelin *et al.*, 2004] web applications were used for interaction studies (network mapping

for association), and enrichment analysis of microarray gene expression data. Each cluster of gene set was analyzed for enrichment of TFBS using the oPOSSUM. The conserved non-coding regions of the promoters were searched for matches to all TFBS profiles in the JASPAR database. The positioning of nucleosomes has played important roles in key cellular processes such as mRNA splicing, DNA replication, and DNA repair [Yasuda *et al.*, 2005; Tilgner *et al.*, 2009; Berbenetz *et al.*, 2010; Sehgal and Singh, 2012], and to evaluate this phenomenon, a sequence-based predictor named iNuc-PhysChem [Chen *et al.*, 2012] was used for identifying nucleosomes of the genes (investigated in this study), by their physico-chemical properties. All the above mentioned tools/web applications have been applied to infer knowledge from the processed data about various aspects of gene expression, their interactions, pathways involved, and role of other biological processes. Identified processes involved in our expression data after comprehensive analysis are ranging from PPI, transcription targets (TTGS) which includes TFBS, miRNA targets, and KEGG file results (KEGG).

3.2.8 Network motif analysis

Network motifs are statistically over represented sub-structures (sub-graphs) in a network, and have been recognized as 'the simple building blocks of complex networks' [Milo *et al.*, 2002; Alon, 2006]. Network motifs are important to understand the modularity and the large-scalestructure of biological networks. In this study AD associated biological networks were used for network motifs identificationand analysis. Available pathways and networks were taken from KEGG, REACTOME, BioGRID and other sources [Stark *et al.*, 2006; Croft*et al.*, 2010; Kandasamy *et al.*, 2010]. MFinder [Kashtan *et al.*, 2004], FANMOD [Wernicke and Rasche, 2006], and MAVisto [Schreiber and Schwobbermeyer, 2005] tools were used to identify and analyze network motifs. Motifs in the range of 3-8 nodes were selected for the study. Statistically significant motifs (depending upon the criteria of tools utilized) were used for further annotations and analysis.

3.3. RESULTS AND DISCUSSION

3.3.1 Differential gene expression, clustering of co-expressed genes

It is relatively difficult to find lists of differentially expressed genes with significant overlap between microarray studies [Kuo *et al.*, 2002]. To overcome from this limitation, multiple

data sets were analyzed in this study to reveal novel patterns amid interacting and associating entities for AG, AD, and other diseases by applying myriad statistical and computational techniques. An integrated multifaceted approach is applied to analyze this polygenic disease (**Figure 3.1**).

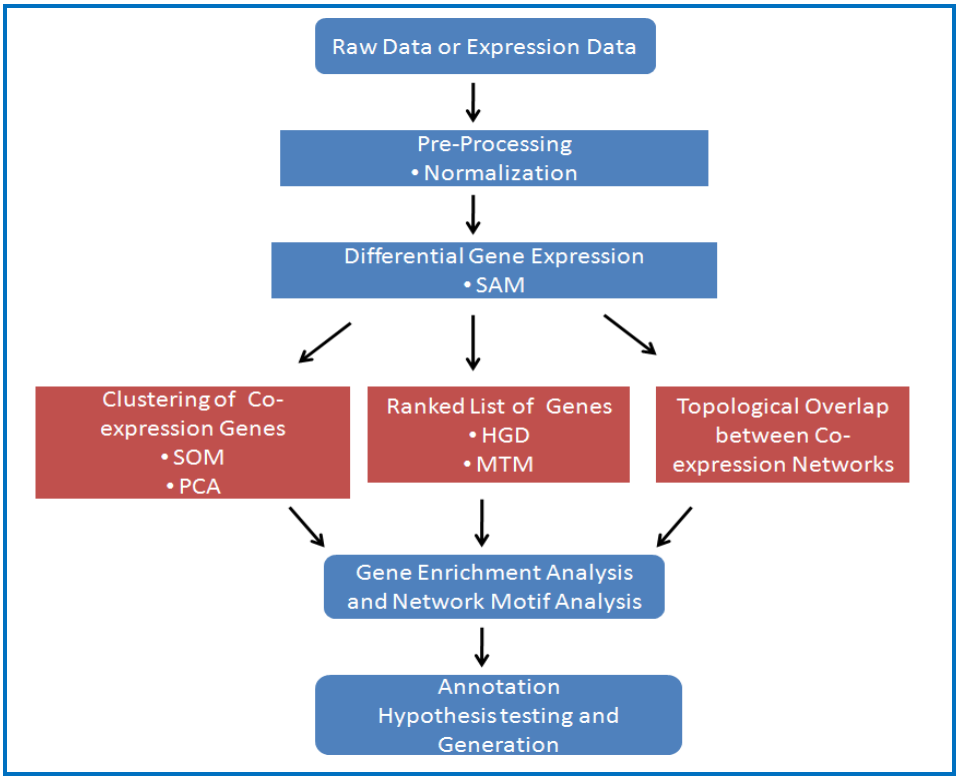


Figure 3.1.Pictographic representation of an integrated multifaceted approach applied in this study which describes the whole methodology

Final set of important genes as robust candidates have been selected through preliminary preprocessing and manual scrutiny based upon the comparative analysis. The results obtained through SAM revealed total 720 genes from all three datasets that are differentially expressed at a false discovery rate of 0.1%. Dynamic feature of SAM through *t*-statistics and ANOVA *F*-statistic was applied in MeV for SAM plot generation (**Figure 3.2**). Our study is based on multiple datasets and for the *t* test; the overall *p*-value of the *F* test has to be adjusted for multiple comparisons. After adjustments, we obtained final *p*-value of 0.55, which covers all plausible genes, which were subjected to further annotation and analysis. When manually compared, 36 genes out of these 720 genes were common for both AD and

AG based on relevant factors involved in both processes. After the GO enrichment analysis it was found that out of the 36 genes that are common in both AD and AG, 26 genes are found to have essential role in AD, while 622 are AG specific (**Table 3.1**).

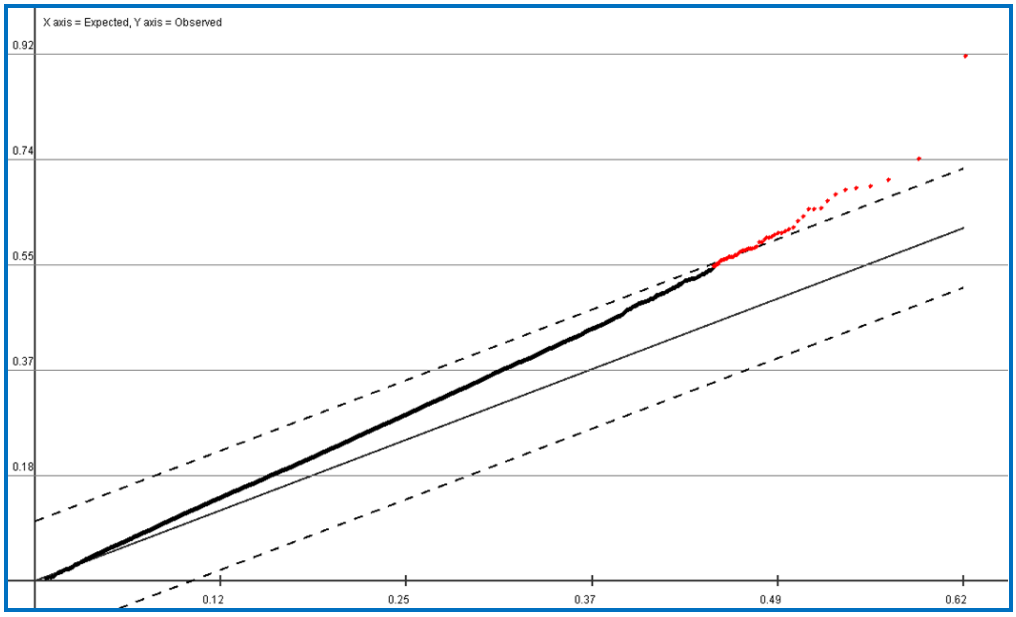


Figure 3.2.A scattered plot of the observed relative difference d (Y-axis) versus the expected relative difference dE (X-axis) indicates the result of SAM on 720 genes where the dash lines are at a threshold distance Δ from the d¼dE identity line. In the above plot the red colored points, that are outside the threshold lines represents common genes for AG and AD at the threshold level Δ. Plot was generated for multiple comparisons for three data sets by applying *t*-statistic and ANOVA *F*-statistic with adjustable *p*-value

Volcano plot was generated between the controls and diseased stage genes of AD. Here we were able to clearly measure the difference in the gene expressions between two groups. Grouping was done based on the available expression data. Over and under expressed genes are shown in scattered form on right and left side of volcano, respectively (**Figure 3.3**).

Table 3.1 Significant number of genes identified through multifaceted integrative approach. 7 different types of analyses applied on all data types and number of genes identified for all categories indicated against each one respectively

Analysis Types	**No. of Genes Involved in Interactions and Associations**
Differential Gene Expression (After multiple comparisons through *t* and ANOVA *F*-statistics)	62 (AD), 658 (AG)
Clustering of Co-expressed Genes	36
Ranked List of Genes	43
Gene Enrichment	43
Network Motif	34
Expression of Genes in Brain Regions	7
Over-represented transcription factor binding sites	54

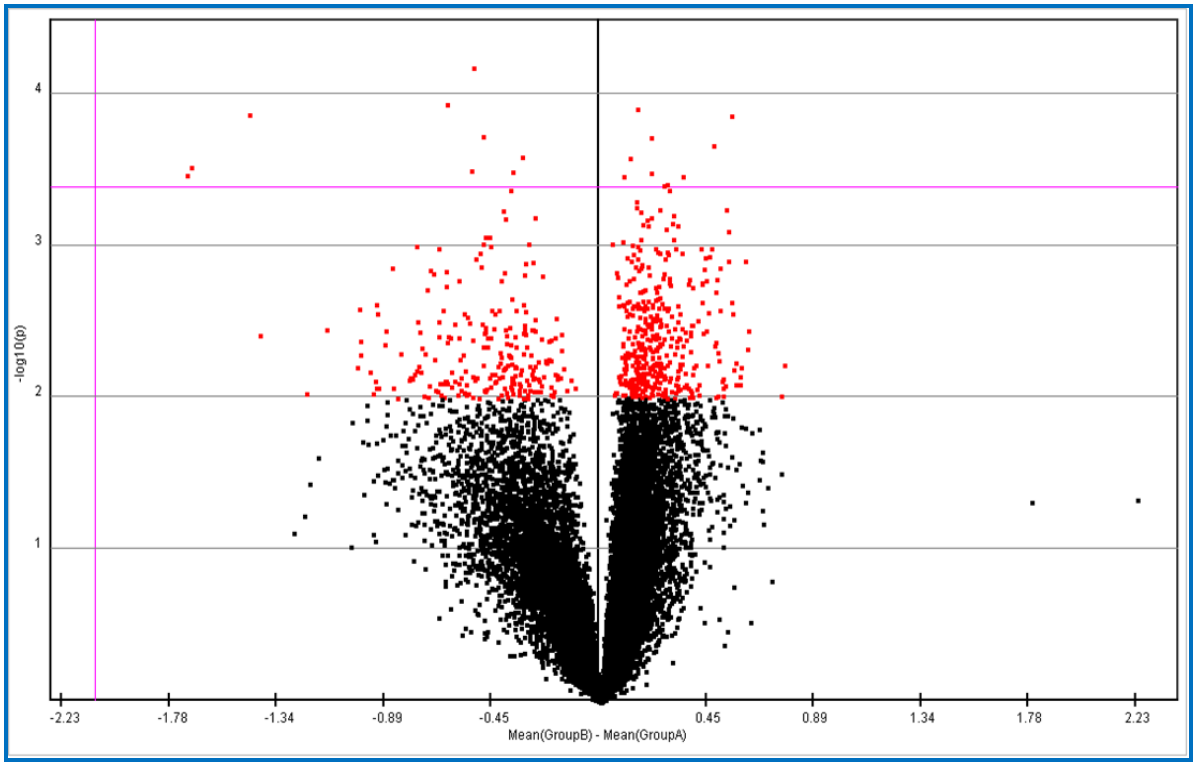


Figure 3.3.Volcano plot for differentially expressed genes. The black longitudinal line between the group A and B is the mean of both groups. I.e. over and under expressed genes

SOM image was generated to group AD genes (62 genes) into clusters of similar expression profiles (**Figure 3.4**). Scale of the clusters in image ranges from black to white through few gray shades indicating maximal (black), moderate (gray), and minimal (white) co-expressed genes fall under respective clusters. Out of 9, 5 clusters were identified with significant measures (cutoff for black and dark gray color only). While looking for genes common in AD and AG, we were able to map all 36 genes which were found in these 5 clusters where 20 were found in 2 clusters (cluster 1, and 2:**Figure 3.4**) and remaining 16 were distributed among 3 other clusters (clusters 3-5:**Figure 3.4**). Similar information could also be generated through SOM (distance based scale). We verified these two scaled conditions for SOM and found same set of genes in both the conditions.

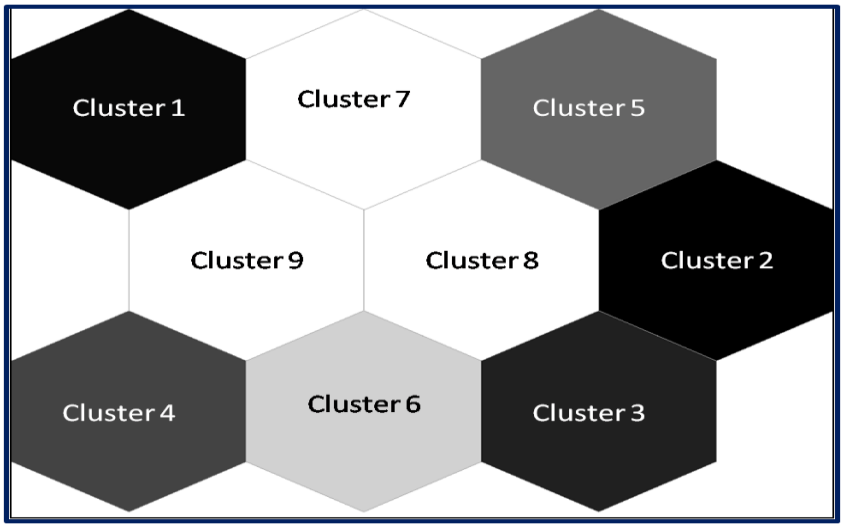


Figure 3.4.SOM result, showing grouped genes into 9 hexagonal clusters of similar expression profiles. Black for more conserved and white for less conserved expression profiles. 36 common genes were found distributed amongst 5 clusters (clusters 1–5)

PCA revealed the relationship between each module of AG and AD gene set (total 89 samples from three experiments) and their phenotypic assessments. PCA projections have been observed for 3 different conditions for 3 component combinations. Analysis wasperformed for data 1, 2, and 3 for all three data sets, respectively, in this study. It has been observed that component 1 (data set 1) interacting with components 2 (data set 2) (**Figure 3.5.A**) and 3 (**Figure 3.5.B**) respectively have shown similarpatterns while

interaction of components 2 and 3 (data set 3) (**Figure 3.5.C**) indicates different pattern than others and support the differential expression levels at component level. Similar kind of modular patterns have been observed in other studies [Ray and Zhang, 2010] and support the modularity of expression levels with reference to regions in the brain. One possible region for this differentiation would be association of diverse brain regions in these studies. PCA mainly applicable for continuous data while our data setscould be categorized into different groups i.e. control, and various disease states. To evaluate this parameter and to further extend our PCA analysis, we applied correspondence analysis projections for all three data sets and were able to trace co-expression patternsfor genes in all three combinations.

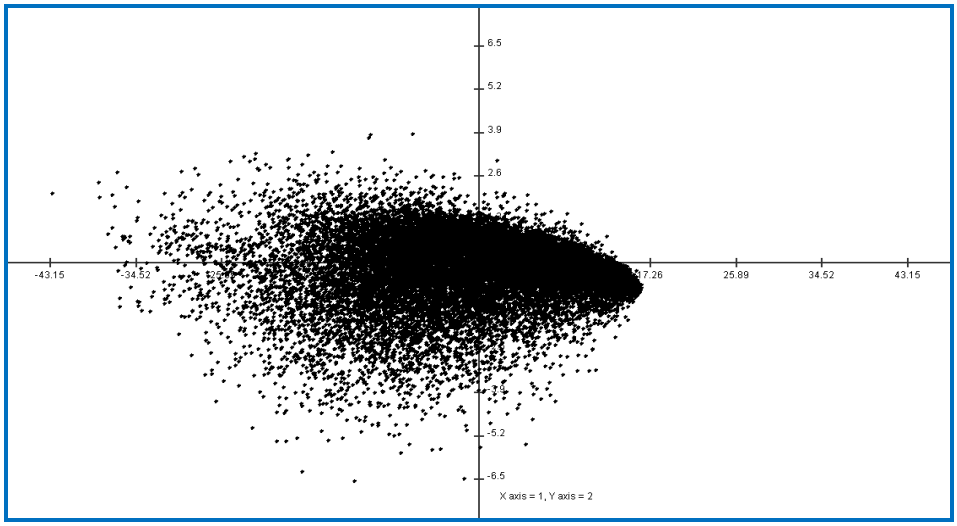


Figure 3.5.A. PCA projections for the components 1 and 2

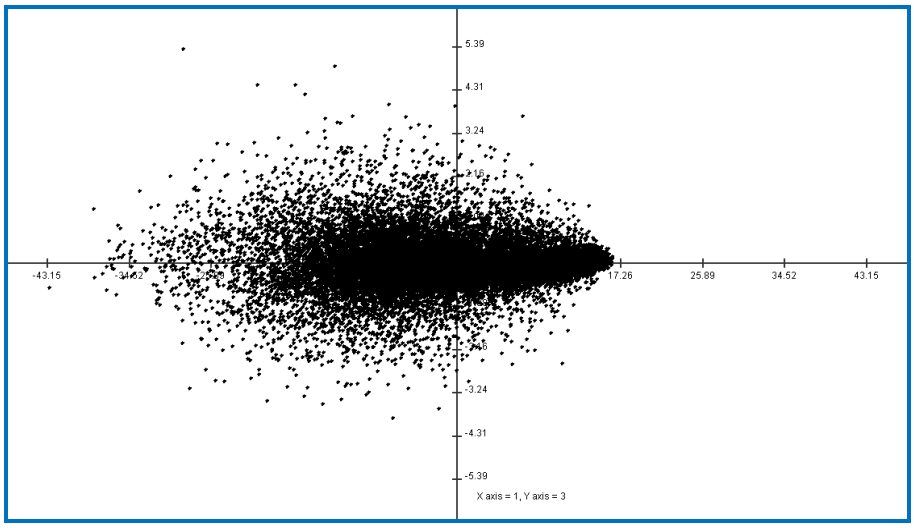


Figure 3.5.B. PCA projections for the components 1 and 3

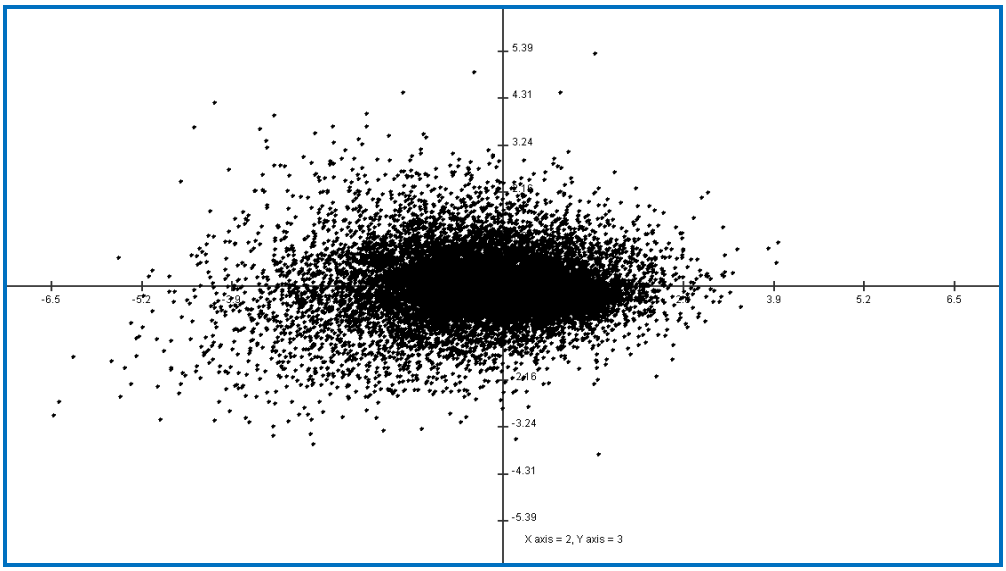


Figure 3.5.C. PCA projections for the components 2 and 3

3.3.2 Enrichment analysis through co-expressed networks and ranked list of genes

GO enrichment analysis shows the significantly enriched GO categories for biological process, molecular function and cellular components (**Figures 3.6.A, B and C**). Based upon the functional enrichment analysis, finally 43 genes were found to be involved in PPI, TTGS, miRNAs, and KEGG (**Table 3.2**). While manually analyzed amongst 658 processed genes of AG and 62 processed genes of AD, we found 36 genes common (**Figure 3.4 and 3.11**). This finding prompted us to perform co-expression analysis of networks. Co-expression network method developed by Ruan and Zhang [Ruan and Zhang, 2006; Ruan *et al.*, 2010] was applied to these genes to measure the pairwise expression similarity between genes and to construct gene co-expressed networks. Besides primary hub genes, some experimentally verified entries were identified and pairwise expression similarity was applied between genes (co-expressed).

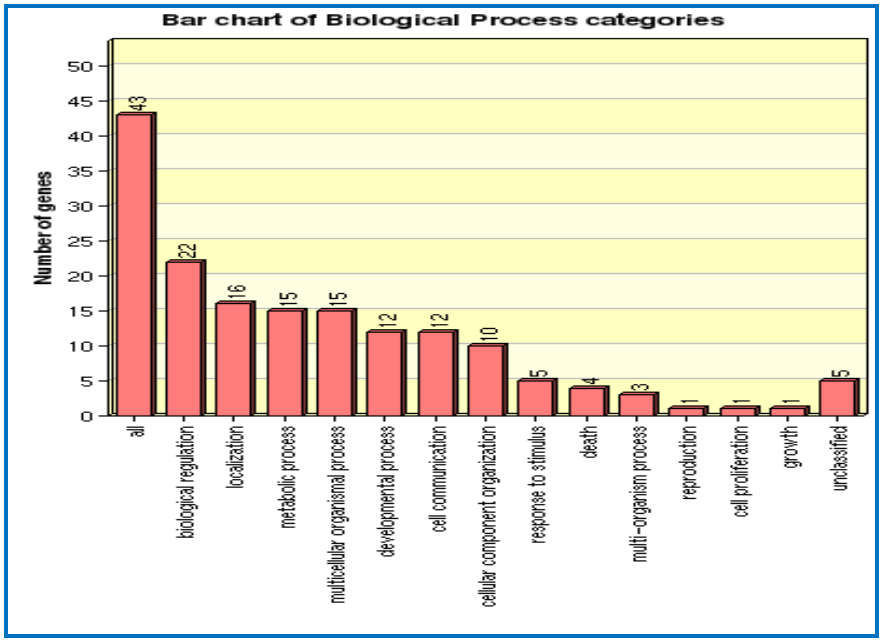


Figure 3.6.A. GO enrichment analysis shows the significantly enriched bar chart of biological process categories for differentially expressed genes. Major genes fallunder the categories biological regulation, localizations, and various metabolic, organism, and development process

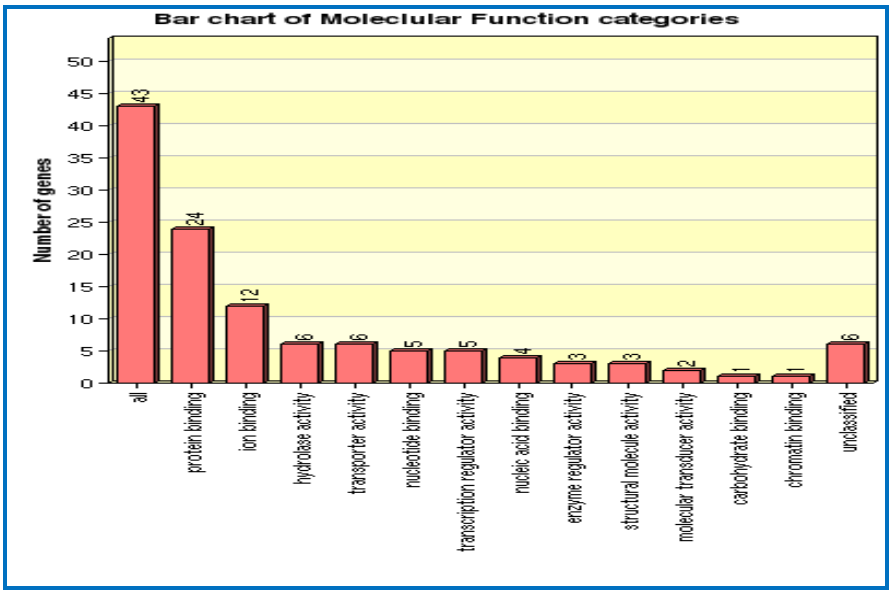


Figure 3.6.B. GO enrichment analysis shows the significantly enriched bar chart of molecular function categories for differentially expressed genes. Major genes fallunder

the categories protein, and ion binding, various activities which includes hydrolase, transporter, and transcription regulation etc.

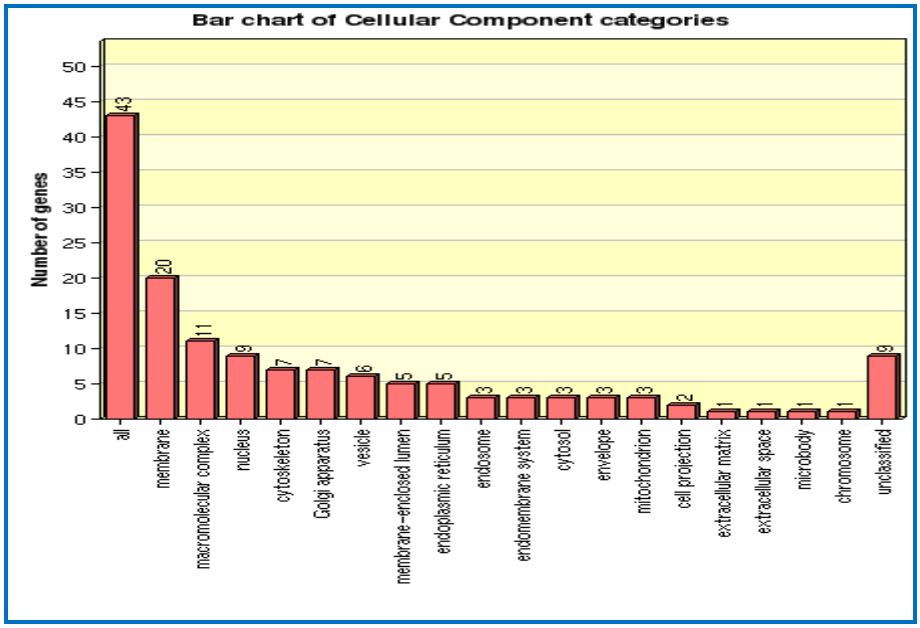


Figure 3.6.C. GO enrichment analysis shows the significantly enriched bar chart of cellular components categories for differentially expressed genes. Major genes fall under membrane, macromolecular complex, and nucleus. Others are distributed amongst Golgi apparatus, cytoskeleton, vesicle, and endoplasmic reticulum

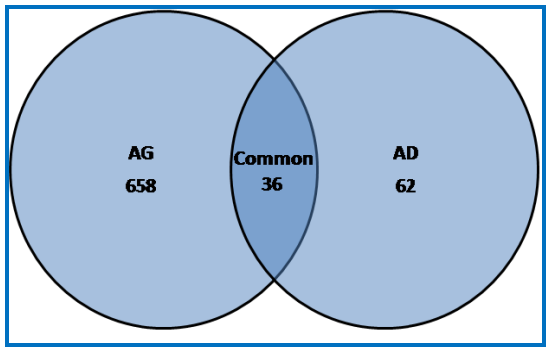


Figure 3.7. Diagram showing independent and common genes amongst AD and AG after multiple comparison analysis through various statistical measures

Table 3.2 Important putative genes identified through manual screening and computational analyses by applying myriad statistical techniques. Abbreviations used are: PPI Gene Set Enrichment Results (PPI), TTGS Enrichment Results (TTGS), MicroRNA Target Gene Set Enrichment Results (miRNA), Main KEGG File Results (KEGG). 'Y' is for presence and 'N' for the absence of that particular gene's association with the concerned biological process

Gene Official Symbol	Detailed Gene Names	Association with other disease	PPI	TTGS	mi RNA	KEGG	Common (AD&AG)	Involved in Pathways
ATXN1	Ataxin 1	Spinocerebellar ataxia type 1	Y	Y	Y	N	N	Not found
KCNAB2	Potassium voltage-gated channel, shaker-related subfamily, beta member 2	Not found	Y	Y	N	N	N	Not found
PSME3	Proteasome (prosome, macropain) activator subunit 3 (PA28 gamma; Ki)	Hepatitis C	Y	Y	N	Y	Y	Proteasome , Antigen processing and presentation, Hepatitis C
ATP5A1	ATP synthase alpha chain, mitochondrial	Parkinson's disease, Huntington's disease	N	Y	N	Y	Y	Oxidative phosphorylation, Metabolic pathways, Alzheimer's disease, Parkinson's disease, Huntington's disease
PSMD4	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 4	Epstein-Barr virus infection	Y	Y	N	Y	Y	Proteasoe, Epstein-Barr virus infection
TBL1X	Transducin (beta)-like 1X-linked	Sensorineural deafness	N	Y	Y	N	Y	Wnt signaling pathway

CDK5	Cyclin-dependent kinase 5	Cocaine addiction	N	Y	N	Y	Y	Alzheimer's disease, Cocaine addiction, Axon guidance
LDB2	LIM domain binding 2	haematopoiesis	N	Y	Y	N	Y	Not found
SLC35A1	Solute carrier family 35 (CMP-sialic acid transporter), member A1	congenital disorder	N	Y	Y	N	N	Not found
YWHAZ	14-3-3 Protein zeta/delta	Pathogenic Escherichia coli infection, Epstein-Barr virus infection, Viral carcinogenesis	N	Y	N	Y	Y	Cell cycle, Oocyte meiosis, Pathogenic Escherichia coli infection, Epstein-Barr virus infection, Viral carcinogenesis
EFNB2	Ephrin-B2	neuroblastoma	N	Y	Y	Y	Y	Axon guidance
DOPEY1	Dopey family member 1	mental retardation in down syndrome	N	Y	Y	N	Y	Not found
RAB6A	RAB6A, member RAS oncogene family	Not found	N	Y	N	N	Y	Not found
KIAA0528	KIAA0528	Not found	N	Y	Y	N	Y	Not found
TMSB10	Thymosin beta 10	ovarian cancer, thyroid carcinoma	N	Y	Y	N	N	Not found
TRIM36	Tripartite motif containing 36	Prostate tumorigenesis	N	Y	Y	N	N	Not found
PTN	Pleiotrophin	Tumor angiogenesis	N	Y	Y	N	N	Not found

USP25	Ubiquitin specific peptidase 25	Down syndrome	N	Y	Y	N	Y	Not found
FAR2	Fatty acyl CoA reductase 2	Not found	N	Y	N	Y	Y	Peroxisome
DNM1L	Dynamin 1-like	dominant optic atrophy	N	Y	N	Y	Y	Synaptic vesicle cycle, Endocrine and other factor-regulated calcium reabsorption, Endocytosis, Fc gamma R-mediated phagocytosis ,Bacterial invasion of epithelial cells
KCNA5	Potassium voltage-gated channel, shaker-related subfamily, member 5	Atrial fibrillation and sudden cardiac death	N	Y	N	N	N	Not found
BRI3	Brain protein I3	TNF-induced cell death	N	Y	N	N	Y	Not found
RND2	Rho family GTPase 2	Not found	N	Y	N	N	Y	Not found
LARP4	Laribonucleoprotein domain family member4	Not found	N	Y	N	N	Y	Not found
TUBG2	Tubulin, gamma 2	Not found	N	Y	N	N	Y	Not found
PBRM1	Polybromo 1	Clear cell renal cell carcinoma, lung cancer	N	Y	N	N	Y	Not found
LTF	Lactotransferrin	Not found	N	Y	N	Y	N	Arachidonic acid metabolism

CSPG5	Chondroitin sulfate proteoglycan 5 (neuroglycan C)	breast tumor	N	Y	N	N	N	Not found
SLC25A6	Solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	Not found	N	Y	N	Y	Y	Calcium signaling pathway, Parkinson's disease, Huntington's disease, Influenza A, HTLV-I infection
ATP6V1G2	Vascular ATP synthese subunit G2	Rheumatoid arthritis	N	Y	N	Y	Y	Phagosome, Oxidative phosphorylation, Metabolic pathways, Synaptic vesicle cycle, Collecting duct acid secretion, Vibrio cholerae infection, Epithelial cell signaling in Helicobacter pylori infection, Rheumatoid arthritis
NOTCH2NL	Notch 2 N-terminal like	Not found	N	Y	N	Y	Y	Notch signaling pathway
C1orf115	Chromosome 1 open reading frame 115	Not found	N	Y	N	N	N	Not found
F5	Coagulation factor V (proaccelerin,	Parahemophilia, venous thrombosis	N	Y	N	Y	N	Herpes simplex infection,

	labile factor)							RNA transport, Dopaminergic synapse, Pathways in cancer, Basal transcription factors etc.
NEFM	Neurofilament, medium polypeptide	Neuronal damage	N	Y	N	Y	N	Amyotrophic lateral sclerosis
LZTS1	Leucine zipper, putative tumor suppressor 1	lung cancer	N	Y	N	N	N	Not found
DPF3	D4, zinc and double PHD fingers, family 3	breast cancer	N	Y	N	N	N	Not found
CDH10	Cadherin 10, type 2 (T2-cadherin)	Autism	N	Y	N	N	Y	Not found
CHRDL1	Chordin-like 1	external canal atresia and neoplasms, myringosclerosis, middle ear anomalies, effusion, cholesteatomas, and neoplasms	N	Y	N	N	N	Not found
GPR22	G protein-coupled receptor 22	Not found	N	Y	N	N	Y	Not found
PRKCB1	Protein kinase C, beta 1	Autism	N	N	N	Y	Y	diverse cellular signaling pathways
HERC1	Hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain and RCC1 (CHC1)-	Not found	N	Y	N	Y	N	Ubiquitin mediated proteolysis

	like domain (RLD) 1							
EHD1	EH-domain containing 1	Bardet-Biedl syndrome 1	N	Y	N	Y	Y	Endocytosis
DICER 1	Dicer1, Dcr-1 homolog (Drosophila)	DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration	N	N	N	N	Y	
MS4A1	Membrane-spanning 4-domains, subfamily A, member 1	Not found	N	Y	N	Y	N	Hematopoietic cell lineage

This set of genes was also been mapped on a sub-network through co-expression analysis (**Table 3.3**). Some genes found to be involved in multiple processes such as proteasome activator subunit 3 (PSME3) and proteasome (prosome, macropain) 26S subunit, non-ATPase, 4 (PSMD4) genes are involved in PPI, TTGS, and KEGG pathways. These two genes are also found to be involved in normal ageing process. While cross checking the output of GO enrichment analysis (using WebGestalt and GOrilla) the above mentioned two genes were found in nucleoplasm with high level of significance (*p*-value 5.43E-04; GOrilla). It indicates how a particular method might not capture all the information latent in biological data and similar analysis with other tools, or methods could provide insightful annotations. Similarly genes such as transducin (beta)-like 1X-linked (TBL1X) and KIAA0528 are found to be involved in TTGS, miRNA targets, and also involved in both AD and AG related metabolic processes.

Table 3.3 Gene details mapped on network through co-expression analysis

Gene Official Symbol	Full Names	Disease Involved	Significance
ITM2B	Integral membrane protein 2B	Familial Danish Disease, Familial British Disease	Plays a role in APP processing regulating the physiological production of the beta amyloid peptide. Restricts docking of gamma-secretase to APP and access of alpha- and beta-secretase
A2MR	Alpha-2-macroglobulin receptor	Breast cancer	One of the neurotoxic mechanisms triggered by Apo*ε4* is to activate a cell type specific apoptogenic program involving A2 MR or LRP and the G (i) class of GTPases and that the apoE4 gene may play a direct role in the pathogenesis of AD and other forms of dementia
PDIA3	Protein disulfide isomerase family A, member 3	Chronic hepatitis or cirrhosis	Essential for formation of the final antigen conformation and export from the endoplasmic reticulum to the cell surface
ITM2C	Integral membrane protein 2C	Not found	Negative regulator of beta amyloid peptide production. May inhibit the processing of APP by blocking its access to alpha- and beta-secretase
PSAP	Prosaposin	Gaucher disease, Tay-Sachs disease, and metachromatic leukodystrophy	Isolating the lipid substrate from the membrane surroundings, thus making it more accessible to the soluble degradative enzymes
SELPLG	Selectin P ligand	Coronary heart disease	Mediates rapid rolling of leukocytes over vascular surfaces during the initial steps in inflammation
RANBP9	RAN binding protein 9	Prostate cancer, breast tumor	Act as an adapter protein to couple membrane receptors to intracellular signaling pathways
MMP2	Matrix metallopeptidase 2	Torg-Winchester syndrome, multicentric osteolysis and arthritis syndrome	Gelatinase A, 72kDa gelatinase, 72kDa type IV collagenase. In addition to gelatin and collagens, it cleaves KiSS1 at a Gly-|-Leu bond
KHSRP	KH-type splicing regulatoryprotei	Apoptosis	Binds to the dendritic targeting element and may play a role in mRNA trafficking

	n		
BRI3	Brain protein I3	Apoptosis	BRI3 may be localized to lysosome and function through lysosome. overexpression of BRI3 induced apoptosis in L929 cells
GGA2	Golgi associated, gamma adaptin ear containing, ARF binding protein 2	Not found	Plays a role in protein sorting and trafficking between the trans-Golgi network and endosomes.
ATP1B1	ATPase, Na+/K+ transporting, beta 1 polypeptide	Not found	catalyzes the hydrolysis of ATP coupled with the exchange of Na(+) and K(+) ions across the plasma membrane
CMPK1	Cytidine monophosphate (UMP-CMP) kinase 1	Haemophilus influenzae type B disease	Catalyzes specific phosphoryl transfer from ATP to UMP and CMP
ATP1B4	ATPase, (Na+)/K+ transporting, beta 4 polypeptide	Not found	Non-catalytic component of a yet unknown sodium or proton exchange ATPase
CSNK1D	Casein kinase 1, delta	Not found	Phosphorylate a large number of proteins. Participates in Wnt signaling. Central component of the circadian clock

3.3.3 Novel genes, their variants, transcription factors, and miRNA targets

Another interesting facet of this study was the identification of some novel gene entries or variants of genes which were not reported inthe previous studies. These novel entries are chromosome 1 open reading frame 115 (C1orf115); D4, zinc and double PHD fingers,family 3 (DPF3); PSMD4; ubiquitin specific peptidase 25 (USP25);potassium voltage-gated channel, shaker-related subfamily, member 5 (KCNA5); leucine zipper, putative tumor suppressor 1 (LZTS1);chondroitin sulfate proteoglycan 5 (neuroglycan C) (CSPG5); and solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6 (SLC25A6). These entire novel variant's information found associated with AD could be useful for targeting eitherdifferent brain regions (conditioned to their presence) or various biomolecular entities for designing treatment strategies for AD andother diseases.

Furthermore we identified over-representation of TFBS in the promoter regions of co-expressed genes. For this analysis, first, differentially expressed genes along with other significant genes involved in AD pathways were taken as input (all the genes from**Tables3.2 and 3.3**). Then, we grouped all TFs according to the TFBS classes and found 10 different groups ofclasses for over-represented TFBS for all the genes. One representative TF was selected from each class based on its frequency ofoccurrence (most frequent in a class is being selected as putative representative candidate for that class) in the input set of genes (**Table 3.4**). Another aspect that we considered while selecting representative TF for each class was availability of its binding sitein almost all the genes that were important to AD (**Table 3.4and Table 3.5**).Furthermore literature analysis of theserepresentative TFs revealed that AP1 (representative of Class 1) or activating protein-1 (IUPAC code: NNNSTCA) is a leucine-zipper TF, which is a heterodimer formed by c-Jun and c-Fos. AP1 acts synergistically with NFAT family proteins on composite regulatoryelements involved in the regulation of the immune system [Macian *et al.*, 2001]. The second class representative ZEB1, zinc finger/ homeodomain serve as DNA binding domain. ZEB1/zfh-1 transcriptional repressor regulates muscle differentiation andexpressed in central nervous system (CNS). Schmalhofer *et al.*, (2009) reveal in their study, the molecular interconnection of ZEB1with E-cadherin, β-catenin, and WNT signaling in cancerogenesis. FEV, the third class representative functions as a transcriptionalregulator essentially in the differentiation and the maintenance of the adult human brain central serotonergic neurons. Kriegebaum *et al.*, (2010) in their study assumed that any severe mutation in FEV would result in fetal death. Class 4 representative SRY (sexdetermining region Y) is a sex-determining gene on the Y chromosome in the therians (placental mammals and marsupials).It plays a major role in determining gender in humans [Ottolenghi *et al.*, 2007].

Table 3.4 Details of 10 representative TFs for 10 different TFBS classes that were identified for genes involved in this study. Representative TF for a class was selected based on its frequency of occurrence in that class. Most frequent was selected as putative representative for a class

Class No.	Class Name	Representative TFs	Most Frequently Occurring TFBS	Gene Hits	No. of TFs in one Class	Total No. of Target TFBS Hits
1	Zipper-Type	AP1	TGAGTCA	46	21	2216
2	Zinc-coordinating	ZEB1	CACCTG	49	37	4798
3	Winged Helix-Turn-Helix	FEV	CAGGAAGT	48	19	3792
4	Other Alpha-Helix	SRY	ATAAACAAT	45	8	1403
5	Other	Tcfcp2l1	CCAGTCTGAGCCAG, CCAGACTGAACCAG	26	2	151
6	Ig-fold	RUNX1	Not Found	41	8	908
7	Helix-Turn-Helix::Other	TLX1::NFIC	TGGCAGCATGCCAA	2	1	2
8	Helix-Turn-Helix	HOXA5	CATTAGTG, AATTTATG	48	18	5435
9	Beta-sheet	TBP	Not Found	30	1	158
10	Beta-Hairpin-Ribbon	T	CTAGGTGTGAA	6	1	9

Table 3.5 Details of over-represented conserved TFBS for all the genes involved in this study along with gene hits, TFBS hits and respective *z*-score for statistical predictions. JASPAR IDs, major structural class and family information is also being provided

Transcription Factors	JASPAR ID	Class	Family	Gene Hits	TFBS hits	Z-score
HOXA5	MA0158.1	Helix-Turn-Helix	Homeo	48	952	21.502
Nkx2-5	MA0063.1	Helix-Turn-Helix	Homeo	46	925	20.972
Pdx1	MA0132.1	Helix-Turn-Helix	Homeo	46	684	18.781
Foxd3	MA0041.1	Winged Helix-Turn-Helix	Forkhead	30	251	18.459
Prrx2	MA0075.1	Helix-Turn-Helix	Homeo	45	617	17.991
ARID3A	MA0151.1	Helix-Turn-Helix	Arid	46	762	15.936
SRY	MA0084.1	Other Alpha-Helix	High Mobility Group	45	471	15.86
Sox5	MA0087.1	Other Alpha-Helix	High Mobility Group	39	357	15.654
Nobox	MA0125.1	Helix-Turn-Helix	Homeo	41	357	13.008
TBP	MA0108.2	Beta-sheet	TATA-binding	30	158	12.392
Foxq1	MA0040.1	Winged Helix-Turn-Helix	Forkhead	25	113	12.134
Foxa2	MA0047.2	Winged Helix-Turn-Helix	Forkhead	34	202	12.128
FOXD1	MA0031.1	Winged Helix-Turn-Helix	Forkhead	37	291	11.722
Gfi	MA0038.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	39	296	11.682
CEBPA	MA0102.2	Zipper-Type	Leucine Zipper	38	253	10.801
NKX3-1	MA0124.1	Helix-Turn-Helix	Homeo	35	312	10.403
MEF2A	MA0052.1	Other Alpha-Helix	MADS	23	81	10.147
Gata1	MA0035.2	Zinc-coordinating	GATA	40	324	9.709

NR3C1	MA0113.1	Zinc-coordinating	Hormone-nuclear Receptor	15	19	9.373
FOXI1	MA0042.1	Winged Helix-Turn-Helix	Forkhead	36	214	9.003
FOXA1	MA0148.1	Winged Helix-Turn-Helix	Forkhead	36	270	8.968
Pou5f1	MA0142.1	Helix-Turn-Helix	Homeo	14	29	8.956
AP1	MA0099.2	Zipper-Type	Leucine Zipper	46	477	8.608
FOXO3	MA0157.1	Winged Helix-Turn-Helix	Forkhead	37	293	7.295
NFIL3	MA0025.1	Zipper-Type	Leucine Zipper	25	74	7.199
Sox17	MA0078.1	Other Alpha-Helix	High Mobility Group	40	276	7.13
PBX1	MA0070.1	Helix-Turn-Helix	Homeo	16	34	6.953
Lhx3	MA0135.1	Helix-Turn-Helix	Homeo	17	59	6.681
ESR1	MA0112.2	Zinc-coordinating	Hormone-nuclear Receptor	3	3	6.633
FOXF2	MA0030.1	Winged Helix-Turn-Helix	Forkhead	14	32	6.562
TAL1::TCF3	MA0091.1	Zipper-Type	Helix-Loop-Helix	25	67	6.463
TEAD1	MA0090.1	Helix-Turn-Helix	Homeo	23	41	6.423
IRF2	MA0051.1	Winged Helix-Turn-Helix	IRF	6	6	6.064
NFE2L2	MA0150.1	Zipper-Type	Leucine Zipper	20	43	6.02
SOX9	MA0077.1	Other Alpha-Helix	High Mobility Group	32	175	6.011
Nkx3-2	MA0122.1	Helix-Turn-Helix	Homeo	44	369	5.98
IRF1	MA0050.1	Winged Helix-Turn-Helix	IRF	19	53	5.953
CREB1	MA0018.2	Zipper-Type	Leucine Zipper	30	87	5.12
HLF	MA0043.1	Zipper-Type	Leucine Zipper	24	45	5.028
Pax4	MA0068.1	Helix-Turn-Helix	Homeo	1	1	4.894

RORA_2	MA0072.1	Zinc-coordinating	Hormone-nuclear Receptor	12	24	4.797
HNF1B	MA0153.1	Helix-Turn-Helix	Homeo	19	32	4.735
HNF1A	MA0046.1	Helix-Turn-Helix	Homeo	14	20	3.903
Ddit3::Cebpa	MA0019.1	Zipper-Type	Leucine Zipper	19	42	3.791
SPI1	MA0080.2	Winged Helix-Turn-Helix	Ets	46	373	3.542
Myb	MA0100.1	Helix-Turn-Helix	Myb	42	231	3.5
FEV	MA0156.1	Winged Helix-Turn-Helix	Ets	48	317	3.295
Pax6	MA0069.1	Helix-Turn-Helix	Homeo	6	7	3.278
Ar	MA0007.1	Zinc-coordinating	Hormone-nuclear Receptor	3	3	3.186
Nr2e3	MA0164.1	Zinc-coordinating	Hormone-nuclear Receptor	22	51	2.485
Tal1::Gata1	MA0140.1	Zipper-Type	Helix-Loop-Helix	17	44	2.176
MAX	MA0058.1	Zipper-Type	Helix-Loop-Helix	29	63	2.15
ELF5	MA0136.1	Winged Helix-Turn-Helix	Ets	45	441	2.077
Stat3	MA0144.1	Ig-fold	Stat	31	92	2.055
MIZF	MA0131.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	9	11	1.968
YY1	MA0095.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	43	619	1.925
Hand1::Tcfe2a	MA0092.1	Zipper-Type	Helix-Loop-Helix	39	207	1.78
Spz1	MA0111.1	Other	Other	21	49	1.679
T	MA0009.1	Beta-Hairpin-Ribbon	T	6	9	1.521
NR2F1	MA0017.1	Zinc-coordinating	Hormone-nuclear Receptor	14	20	1.33
Evi1	MA0029.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	6	10	1.295
znf143	MA0088.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	6	6	1.201

USF1	MA0093.1	Zipper-Type	Helix-Loop-Helix	33	84	1.183
EWSR1-FLI1	MA0149.1	Winged Helix-Turn-Helix	Ets	1	1	1.035
Sox2	MA0143.1	Other Alpha-Helix	High Mobility Group	10	15	1.031
RXR::RAR_DR5	MA0159.1	Zinc-coordinating	Hormone-nuclear Receptor	6	6	0.922
NR4A2	MA0160.1	Zinc-coordinating	Hormone-nuclear Receptor	38	223	0.911
NFKB1	MA0105.1	Ig-fold	Rel	15	24	0.548
REST	MA0138.2	Zinc-coordinating	BetaBetaAlpha-zinc finger	1	1	0.472
ZEB1	MA0103.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	49	626	0.388
ESR2	MA0258.1	Zinc-coordinating	Hormone-nuclear Receptor	8	10	0.364
HNF4A	MA0114.1	Zinc-coordinating	Hormone-nuclear Receptor	17	41	0.06
MZF1_5-13	MA0057.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	42	282	-0.055
RUNX1	MA0002.2	Ig-fold	Runt	41	211	-0.508
RELA	MA0107.1	Ig-fold	Rel	23	48	-0.813
Esrrb	MA0141.1	Zinc-coordinating	Hormone-nuclear Receptor	30	93	-0.997
SPIB	MA0081.1	Winged Helix-Turn-Helix	Ets	45	629	-1.01
RORA_1	MA0071.1	Zinc-coordinating	Hormone-nuclear Receptor	27	66	-1.044
STAT1	MA0137.2	Ig-fold	Stat	20	38	-1.216
RREB1	MA0073.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	6	6	-1.228
MYC::MAX	MA0059.1	Zipper-Type	Helix-Loop-Helix	13	15	-1.25
TP53	MA0106.1	Zinc-coordinating	Loop-Sheet-Helix	0	0	-1.418
CTCF	MA0139.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	11	13	-1.72
RXRA::VDR	MA0074.1	Zinc-coordinating	Hormone-nuclear Receptor	1	1	-2.24
NFATC2	MA0152.1	Ig-fold	Rel	39	339	-2.323
PPARG	MA0066.1	Zinc-coordinating	Hormone-nuclear Receptor	0	0	-2.389

Arnt::Ahr	MA0006.1	Zipper-Type	Helix-Loop-Helix	41	225	-2.43
NR1H2::RXRA	MA0115.1	Zinc-coordinating	Hormone-nuclear Receptor	0	0	-2.687
REL	MA0101.1	Ig-fold	Rel	30	97	-2.885
HIF1A::ARNT	MA0259.1	Zipper-Type	Helix-Loop-Helix	41	125	-2.902
SRF	MA0083.1	Other Alpha-Helix	MADS	2	2	-3.142
Myf	MA0055.1	Zipper-Type	Helix-Loop-Helix	30	90	-3.503
NF-kappaB	MA0061.1	Ig-fold	Rel	25	59	-3.683
Pax5	MA0014.1	Helix-Turn-Helix	Homeo	3	3	-3.984
Arnt	MA0004.1	Zipper-Type	Helix-Loop-Helix	23	39	-4.009
TLX1::NFIC	MA0119.1	Helix-Turn-Helix::Other	Homeo::Nuclear Factor I-CCAAT-binding	2	2	-4.114
ELK1	MA0028.1	Winged Helix-Turn-Helix	Ets	41	178	-4.754
ELK4	MA0076.1	Winged Helix-Turn-Helix	Ets	15	24	-5.16
Tcfcp2l1	MA0145.1	Other	CP2	26	102	-5.235
PPARG::RXRA	MA0065.2	Zinc-coordinating	Hormone-nuclear Receptor	16	31	-5.693
GABPA	MA0062.2	Winged Helix-Turn-Helix	Ets	30	61	-5.765
NFYA	MA0060.1	Other Alpha-Helix	NFY CCAAT-binding	17	26	-5.788
Mycn	MA0104.2	Zipper-Type	Helix-Loop-Helix	27	53	-6.058
Egr1	MA0162.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	21	32	-6.164
Zfp423	MA0116.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	15	27	-6.341
MZF1_1-4	MA0056.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	48	659	-6.607
Myc	MA0147.1	Zipper-Type	Helix-Loop-Helix	23	50	-6.657
E2F1	MA0024.1	Winged Helix-Turn-	E2F	24	43	-7.485

		Helix				
PLAG1	MA0163.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	2	3	-7.489
SP1	MA0079.2	Zinc-coordinating	BetaBetaAlpha-zinc finger	40	248	-7.904
NHLH1	MA0048.1	Zipper-Type	Helix-Loop-Helix	15	22	-8.069
EBF1	MA0154.1	Zipper-Type	Helix-Loop-Helix	35	111	-8.356
Zfx	MA0146.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	36	95	-8.857
ZNF354C	MA0130.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	47	645	-9.995
Klf4	MA0039.2	Zinc-coordinating	BetaBetaAlpha-zinc finger	44	264	-10.442
INSM1	MA0155.1	Zinc-coordinating	BetaBetaAlpha-zinc finger	21	40	-10.54

Class 5 representatives, Tcfcp2l1 is preferentially upregulated in embryonic stem cells as component of the LIF and BMP signaling pathways, self-renewal regulator and key reprogramming factor but has uncharacterized DNA-binding properties and function [Chen *et al.*, 2008]. RUNX1, the representative of class 6 is a critical regulator of CD41 expression in early embryos. It also controls the early stages of hematopoietic development and an essential regulator of blood cell specification [Tanaka *et al.*, 2012]. Class 7 representatives, TLX1::NFIC displayed cell type specific patterns and these footprinting patterns are highly correlated with gene expression differences. Also there are evidences where the homeoprotein TLX1 is known to interact with the CCAAT binding TF NFIC [Boyle *et al.*, 2011]. This suggests that the differential binding of the TLX1/NFIC complex in these cell types identified by the footprinting data is likely mediated by NFIC expression. Class 8 representative HOXA5 (Homeobox A5) is a member of the Homeobox family. During development it regulates organogenesis in lung, mammary, and tracheal tissues, and in adult tissues regulates mammary gland development and function. HOXA5 is also believed to function as a tumor suppressor (breast, colon, and lung) by transactivating p53 to promote p53-dependent and p53-independent apoptotic signaling. The expression of HOXA5 is regulated at least in part through epigenetic modifications of the HOXA5 gene in these tumors [Atabakhsh *et al.*, 2012]. The representative of class 9 named, TBP (TATA box

Binding Protein) is one of the most recognized DNA benders, and their interactions with TFs are majorly responsible for the spatiotemporal gene expression patterns [Hooghe *et al.*, 2012].

In class 10, T (T-box transcription family) is vital for morphogenetic movements in various processes of animal development [Yamada *et al.*, 2012]. All the important class representatives TFs for TFBS were provided in (**Figure 3.8.A** (classes 1-5) **and B** (classes 6-10)). It has been observed that conserved TFs are potent gene regulators for AD. We also explored this conservation of TFs towards their regulation mechanism for gene involved in AD. We found shared nature of TFs in AD which was associated with other diseases such as haematopoiesis, Epstein-Barr virus infection, viral carcinogenesis and other carcinomas, dominant optic atrophy, atrial fibrillation, coronary heart disease and sudden cardiac death. This association will provide important insight into regulatory processes common to AD, AG and other mentioned diseases. Major classes for secondary structure elements were dominated by helix-turn-helix for thefamilies' homeo, Arid, and Myb and its winged version for the families Forkhead, IRF, and Ets [Aravind *et al.*, 2005]. Another major class found was Zinc-coordinating which belongs to two families' hormone-nuclear receptor, and beta-beta-alpha-zinc finger (**Table 3.5**). These structural level constraints proposed plausible targets for neuronal tangles and plaques through normal and winged helix-turn-helix and beta-betaalpha structures. Linked TFBS for these physico-chemical elements could be manipulated to deal with involved complexities.

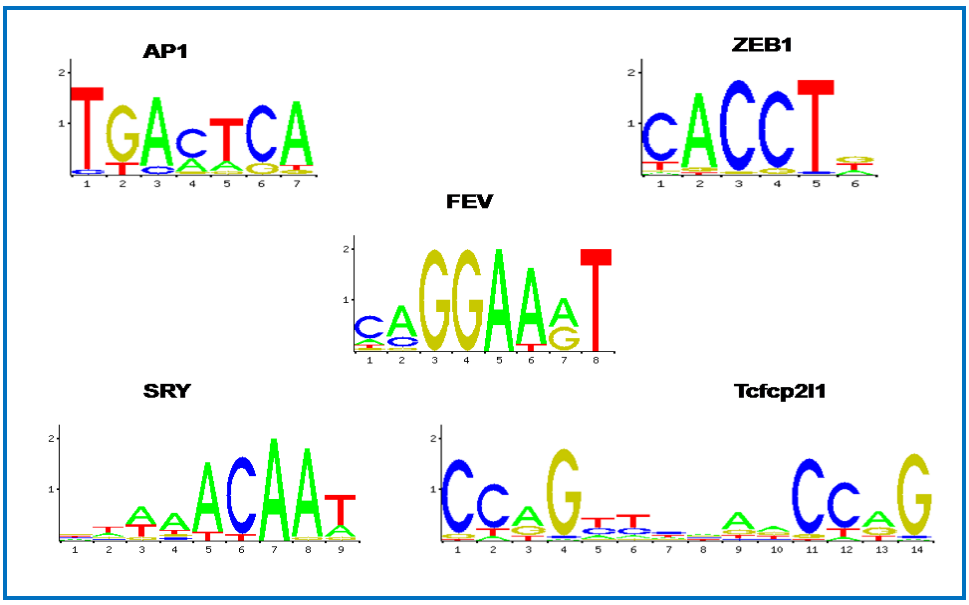


Figure 3.8.A.The sequence LOGO's for classes 1–5 representative TFs. Class representatives are AP1 (class 1), ZEB1 (class 2), FEV (class 3), SRY (class 4), Tcfcp2l1 (class 5)

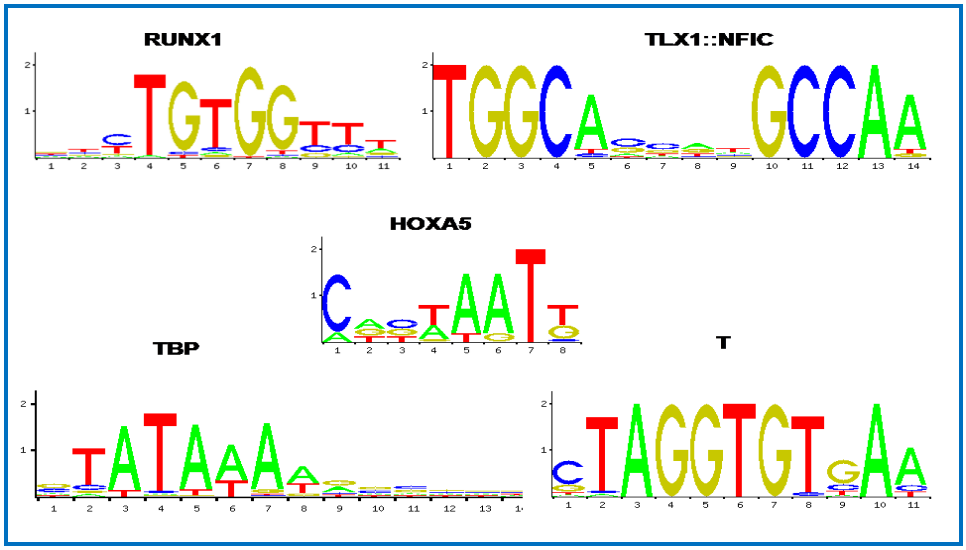


Figure 3.8.B.The sequence LOGO's from classes 6–10 representative TFs. Class representatives are RUNX1 (class 6), TLX1::NFIC (class 7), HOAX5 (class 8), TBP (class 9), T: T-box transcription family (class 10)

Another interesting observation that emerged from the analysis is association of miRNAs and their target genes. miRNAs regulate target genes at the posttranscriptional level and plays an important role in the development, and in other human diseases including heart

disease, schizophrenia and psoriasis. In this study, out of the 43 genes (significant hub genes identified through enrichment analyses), 13 genes were found to be miRNA targets (**Table 3.2**). There are some other genes such as LDB2 (LIM domain binding 2), DOPEY1 (dopey family member 1), DNM1L (dynamin 1-like) and EHD1 (EH-domain containing 1) are found to be common in both AD and AG, but are involved in different biological functions in AD. LDB2 is found in both TTGS and miRNA. DOPEY1 is found in miRNA and DNM1L is found to be common in KEGG whereas EHD1 is common in TTGS and KEGG. Differential patterns of interaction and association of these genes with diverse biological processes support the premise that all interactions are associations but not all associations are interactions.

3.3.4 Brain regions and their pathway mapping

In a recent study, microarray data of four different brain regions that are EC, HIP, PCC and MTG from AD affected and normal ones have been analyzed. Six sets of intersection genes were obtained from six comparisons- (1) EC and HIP; (2) EC and PCC; (3) EC andMTG; (4) HIP and PCC; (5) HIP and MTG; (6) PCC and MTG [for more details see Ray and Zhang, 2010]. Co-expressed networkswere built for each brain region using the intersection genes. When the genes of the present study were compared with the results of Ray and Zhang (2010), interestingly many common genes were found involved in various brain regions. At least one common gene is found in almost all regions. Between EC and HIP region, KCNAB2 and SPF3 are found, between EC and PCC region GPR22 is found, between HIP and PCC region KCNAB2 is found. There are four genes from this study named TBL1X, EFNB2, RND2 and CDH10 which were found to be involved between HIP and MTG region and TBL1X, EFNB2 genes between MTG, EC and HIP region [Ray and Zhang, 2010]. Concerned pathways such as Wnt, axon guidance, and Akt have also been found associated and the detailed descriptions and significance of these genes are described in**Table 3.2**. These genes and their associated pathways could be treated as hotspots while planning experimental procedures for association studies.

Wnt proteins are secreted morphogens, which are required for basic developmental processes, such as cell-fate specification, progenitor-cell proliferation and the control of asymmetric cell division, in many different species and organs. There are at least three different Wnt pathways: the canonical pathway, the planar cell polarity (PCP) pathway and

the Wnt/Ca2+ pathway. TBL1X is one of the core proteins of canonical Wnt signaling pathway. In this pathway, the major effect of Wnt ligand binding to its receptor is the stabilization of cytoplasmic beta-catenin through inhibition of the beta-catenin degradation complex. Beta-catenin is then free to enter the nucleus and activate Wnt-regulated genes through its interaction with TCF (T-cell factor) family TFs and concomitant recruitment of co-activators [Nalbantoglu *et al.*, 2012]. The hub of the canonical pathways are obtained as KC1AL (casein kinase I isoform alphalike), YWHAZ (protein kinase C inhibitor protein 1) and TBL1XR1 (F-box-like/WD repeat containing protein). TBL1XR1, also a core protein in the canonical Wnt signaling pathway, is involved in signal transduction and cytoskeletal assembly and plays an essential role in transcription activation mediated by nuclear receptors and has effects on cytotypic differentiation. Besides, low levels of TBL1XR1 gene expression cause breast cancer [Kadota *et al*., 2009].

TBL1X is a protein that plays an essential role in transcription activation mediated by nuclear receptors. It is a component of E3 ubiquitin ligase complex and could be a promising candidate for targeting APP regulatory inhibition. Axon guidance represents a key stage in the formation of neuronal network. Axons are guided by a variety of guidance factors, such as netrins, ephrins, slits, and semaphorins. These guidance cues are read by growth cone receptors, and signal transduction pathways downstream of these receptors converge onto the Rho GTPases to elicit changes in cytoskeletal organization that determine which way the growth cone will turn. The geneEFNB2 encodes an EFNB class ephrin (EPH) which binds to the EPHB4 and EPHA3 receptors [Hinck, 2004]. EFNB2 could be aputative marker for disease progression and analysis. There are several evidences where various inhibitors, biomarkers, and otherbiological entities associated with the progression of AD and other diseases have been proposed and could provide molecular insights for the cure of AD [Chou *et al*., 2000; Chou and Howe, 2002].

NOTCH2NL (Notch homolog 2 N-terminal-like), also known as N2N is identified as another putative candidate shared among ADand AG. NOTCH2NL is a 236 amino acid protein that has a nonspecific function in Notch signaling. The Notch signaling pathway controls cellular interactions important for the specification of a variety of fates in both invertebrates and vertebrates. The Notch genes are expressed in a variety of tissues in both the embryonic and adult organism, suggesting that the genes are involved in multiple

signalingpathways. The Notch proteins have been found to be over expressed or rearranged in human tumors. In addition, mutations inNotch genes may cause hyperplasia of the nervous system [Duan *et al.*, 2004]. Association of Notch signaling pathway and MAPKpathway is very well defined and identification of some marker genes involved (sharing biochemical signals) in AD and other diseasesuch as cancer; cardiovascular disease, diabetes etc. justify the purpose of such comparative analyses. The results obtained from nucleosome analysis are provided in "*http://www.sciencedirect.com/science/article/pii/S0022519313002853*". In this table output is divided into five columns. The first column of the table contain the name of genes, the second column contain the length of genes with respect to their DNA sequence in base pair (bp). The third column gives the information about the number of segments or subsequences (each of these sub-sequence is 150-bp long) in each gene sequence. The fourth column has the information about the presence or absence of nucleosomes in each segment of the genes. The fifth column gives the information about the presence or absence of linkers in the respective sequence. Range of linkers and their respective frequency is also mentioned in square brackets ([]) along with each entry in column 5. It is believed that generated information for nucleosomes, and linkers could proved to be biologically meaningful for future structure based studies associated with genes and proteins involved in AD.

3.3.5 Network motifs and their disease associated annotation

Despite many ground-breaking discoveries during the past century, we are far from having a complete understanding of theintricate network of molecular processes involved in diseases and still searching for the cures of most complex diseases. Diseases arecaused by the effect of several genes: for instance, comprehensive studies on mutations in complex diseases, such as breast cancer [Sjoblom *et al.*, 2006] or other types of cancer. In recent years many important properties of complex networks have been delineated.In particular, significant progress has been made in understanding the relationship between the structural properties of networks andthe nature of dynamics taking place on these networks [Sporns and Honey, 2006; Quackenbush, 2006]. The present study provides further support for the presence of "small-world" features in functional brain networks and demonstrates that AD is characterized byan association of small-world network distinctiveness. Interaction data was collected from resources such as KEGG, Reactome, and others for AD pathways. Network motif analysis was performed through FANMOD and crosschecked using MFinder, and MAVisto.

Identified network motifs were selected based upon the statistical criteria for the tools used (*Z*-score and *p*-value). We found the common network motifs which occur in transcriptional network such as SIM, Bifan, MIM, and complex networks (**Figure. 3.9a-n**). While annotating these motifs for their biological significance, we found specific ids matching with standard motif dictionary [Alon, 2006], which are mentioned asidentified motifs (**Figure. 3.10**).

Analyzed network motifs were inferred for their biological significance through association mapping overstandard biological network motifs. SIM [Shen-Orr *et al.*, 2002] are a family of structures with free parameters and are strong network motifs. In the SIM network motif, a master TF Xcontrols a group of target genes, *Y1, Y2,…., Yn*. Each of these target genes has only one input and is not being regulated by any other TF. Regulation sign is also same for all the genes in SIM and master TF X is usually autoragulatory. SIM has a dynamical function and can generate temporal programs of expression, where genes are activated one by one in a defined order. Temporal order associated with SIM exhibits "the earlier the protein functions in the pathway, the earlier the gene is activated" [Kalir and Alon, 2004]. The temporal order generated by SIM is maintained against mutations due to selective advantage affordedby just- when- needed production strategies. These are evidences of temporal order found in damage repair systems controlled bySIMs, where genes responsible for the mildest form of repair are turned off first, and for more severe damage are turned off later [Ronen *et al.*, 2002]. MIM arose as a generalization of SIM. Common four-mode motifs among neuronal networks are diamond, biparallel, and bifan. A bifan motif is built by two regulators and two regulated genes, with the two regulators jointly regulatingeach target gene (**Figure. 3.10, id282**).

To the best of our knowledge, this is first report on network motif analysis for AD associated biological networks. All the identified network motifs resembles the standard motifs and have been found associated with specific proteins for their activation or inhibition as an active role in these networks. Significant entries have been found for APP, APO-1, CAPN1, APAF1, APBB1 and others while performing annotations. As a multifactorial disorder, AD has been frequently linked to vascular risk factors like hypertension, obesity, diabetes, hyperlipidemia etc. in numerous prospective cohort studies of the general population [Qiu, 2012]. Previous studies have suggested that the molecular pathophysiology of AD significantly overlaps with

that of type-2 diabetes and the metabolic syndrome, most notably in insulin resistance [Craft et al., 2012]. Talbot et al. (2012) demonstrated that insulin resistance inAD occurs not only in peripheral tissues but also in the brain. The authors show that hippocampal brain slices in AD were lessresponsive to insulin than controls because of increased phosphorylation of IRS-1 that attenuated downstream Akt and ERK signaling. Brain insulin resistance in AD was not dependent on diabetes, or on the APOE4 genotype, which also affects the Akt pathway andis a major determinant of risk for non-Mendelian AD [Warren and Strittmatter, 2012]. The kinds of analyses propose common technical capabilities and treatment solutions not for AD but for other disease where biological markers are being shared. Generated sub-networks for all node sizes were annotated based on statistical criterion using mean-frequencies, standard deviation, *z*-score and *p*-values as demonstrated in **Table 3.6** and **Figure 3.10**.

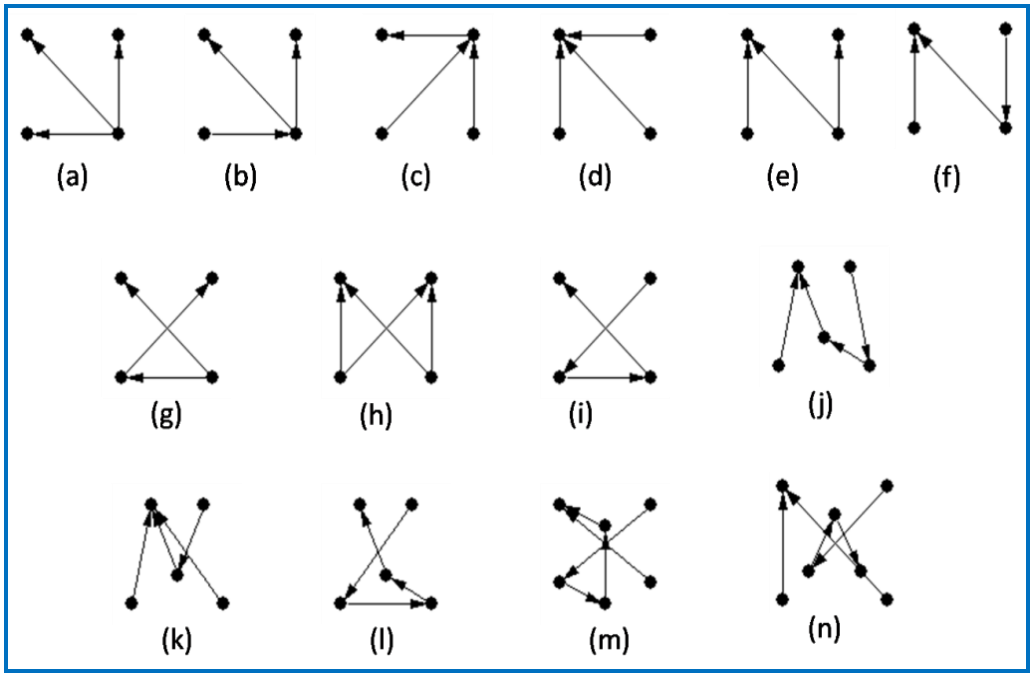


Figure 3.9. (a–n) Network motifs identified for AD associated pathways. Most frequent motifs belongs to four node class (9 types: a–i) followed by 5 node class (3 types: j–l), then 1 each for 6 and 7 node classes (1 type: m and n, respectively). Search was performed for 3–8 node motifs. No significant motif was found in 3 and 8 node classes

Table 3.6 Values of the statistical parameters for each recurrent motif in the AD pathway

Network Motif Image ID (Adjacency matrix)	Abreviation	Z-Score	*P*-value
000001100	3a	0.12669	0
000000110	3b	0.1163	0.017
000100100	3c	0.19624	0.017
0000001000011000	4a	2.7434	0.003
0000000110001000	4b	0.39431	0.322
0000000000001110	4c	4.5544	0.001
0000100010001000	4d	3.0859	0.001
0000100001000100	4e	3.1316	0.003
000000010000011000010000	5a	5.4823	0
000000000110000100010000	5b	2.7952	0.011
000000001000001001001000	5c	2.1103	0.037
000000000100000001001000100000100000	6a	4.5444	0
000000001000000001100000100000100000	6b	2.2233	0.006
000000000010000000001001000010000001000000010000000	7a	2.7986	0

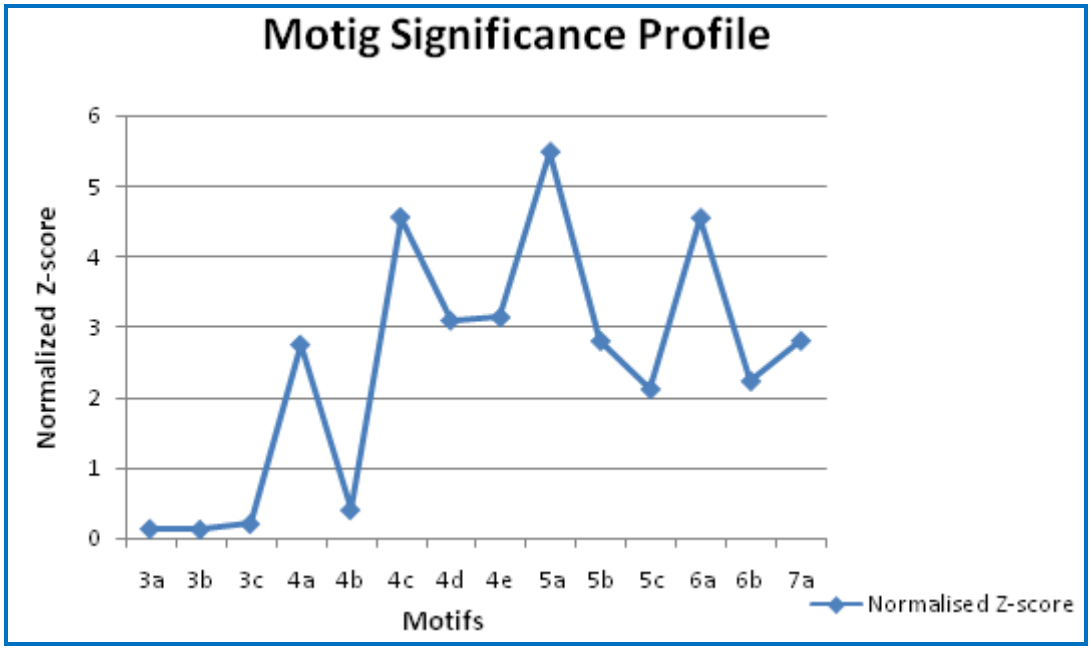


Figure 3.10.Significance profile for all randomly generated 3-7 node sub-graphs based on normalized *z*-score; the motif significance profile evidently exemplifies that when the complexity in AD pathway increases, the interactions among the nodes and intricacy in recognition of genes amplifies immensely. Network motifs of 4-6 nodes are more significant for this data set, while 3 node motifs are almost insignificant

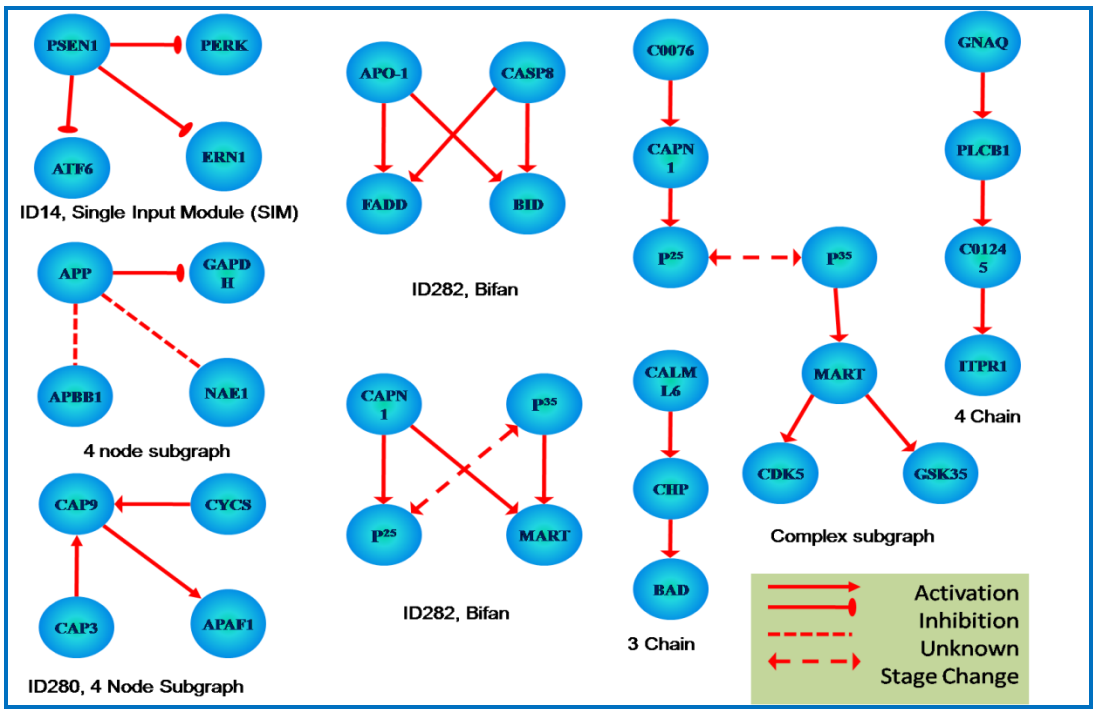


Figure 3.11.Biologically significant network motifs identified from AD pathways. Bifan, 4-node sub-graph, and SIM were common with activation, inhibition, and stage change properties. Respective gene/protein names are given for each node

This is a kind of study which not only performs computational enrichment analysis but also evaluate the performance of various tools, methods, applications available to analyze gene expression and network data. In this study similar kind of analyses were performed using various available popular, and best quality methods and tools which will definitely propose pros and cons of using freely available tools for academics and research purpose. This study revealed common sharing of important biological processes and genes among AD and supports previous studies and hypothesis for the same. Some novel genes and other variants for various biological processes have been reported associated with AD and could be implicated in biochemical events leads to AD form AG through pathways and interactions. Quantitative measurement and assessment of patho-physiological processes amongst AD could identify suitable gene candidates and provide more information about therapeutic targets. Modeling and simulation studies would provide system level insights while selecting these newly reported entities as workhorses after their experimental verifications.

3.4 CONCLUSION

In this study, an integrative systems biology approach is presented to cram a complex disease like AD and its associationwith AG and other diseases. Genome-wide expression profiling along with their interaction mapping studies allow researchers todiscover disease genes systematically. In this chepter, we studied several approaches for prioritizing genes by integrating geneexpression profiles. Unique study on network motifs and association of network motif entities with AD related markers revealed a new direction to its biological annotations. The results show the extensive links between AD and AG at a molecular level, identifying core biological processes and genes they share. Identified TFs and their respective TFBS information would be biologicallymeaningful for the associated cellular and molecular process for AD. Major classes for secondary structure elements were dominated by helix-turn-helix for the family homeo, and its winged version for the family forkhead. There is no doubt that ADand AG share common patho-physiological processes. It has been found that not only AG and AD share common biological processesbut also there is involvement of other important human disease with these biological processes. Interactions of TFBS, genes, andencoded proteins at molecular level for other disease such as diabetes, dominant optic atrophy, coronary heart disease, sudden cardiac arrest, gaucher disease, myriad carcinomas, and cirrhosis signifies putative association with AD and above mentioneddisease. Identified unique miRNA targets as a regulatory process for AD such as LDB2, and DOPEY1 could be verified to look fortheir active participation in the process of gene regulation and inhibitory activities. Given these results, a comprehensive analysisof both conditions in tandem, for example using the same tissues and microarray platforms across both age and AD progression, would be quite powerful. Such direct comparison would complement these analyses, permitting quantitative assessment of biologicaldifferences, as well as the similarities, between AG and AD.

CHAPTER-4

A genome wide association studyfor all key genes identified in our earlier research having direct or indirect impact on the metabolic pathways of alzheimer's disease: development of a web resource

ABSTRACT

The pathogenesis of AD is still a mystery however; genetic variants play a significant role in the pathogenesis of AD and several target genes contributing to its etiology. A number of mutations in the genes APP, PSEN1 and PSEN2 are described as the cause of up to 50% cases of the rare and early onset form of AD. ApoE *ε4* is the strongest risk gene for developing AD. Recently, ten more genes are identified which increase the risk of developing late onset AD. There are many other genes reported earlier to be involved in AD progression. All these prime genes point to potential new biomarkers for specific disease processes and possible new targets for future disease modifying treatment.Here we performed statistics based *in silico* analysis on genotype data of the prime genes for a specific hybrid population CEU i.e. CEPH (Utah residents with ancestry from northern and western Europe). Various markers were found and Linkage disequilibrium (LD) plots were also generated. We have predicted the possible impact of amino acid substitutions on the structure and function of proteins using biophysical and evolutionary comparisons and found mutations that might be damaging. Phosphorylation states were also predicted for all the genes under study. Generated information would be of utmost use to all the researchers working in the area of genomics and molecular genetics of the AD. We developed a unique web repository named '*ADDGAP*', where Genetic markers, Haplotype blocks, nsSNPs, tagSNPs, LD, and Phosphorylation states information are available in a distinctive manner and is available for academic and research use at *http://www.bioinfoindia.org/addgap/*.

4.1 INTRODUCTION

AD is the most recurrent cause of dementiawhich results in millions of death worldwide. It is a multifactorial neurodegenerative disorder with several target genes contributing to its etiology [Parihar and Hemnani, 2004]. Pathological, genetic, biochemical and modeling studies all point to a significant role of Aβ aggregation in AD [Manna and Mukhopadhyay, 2013]. The amyloid hypothesis determines that the production, aggregation and accumulation of Aβ in the brain give rise to a cascade of neurotoxic events that proceed to neuronal degeneration [Mattsson *et al.*, 2013]. Genetic factors are involved in 25 to 40% of AD patients and in some cases, AD segregates as an autosomal dominant trait in families [Loy *et al.*, 2014]. In these families, 3 genes are identified that, when mutated, cause AD which are; APP, PSEN1 and PSEN2. Together, these mutations are responsible for 30 to 50% of autosomal dominant AD cases [Cruts and Broeckhoven, 1998; Rongve *et al*., 2013]. They are important for pre-symptomatic diagnostics of patients of autosomal dominant AD families that segregate these mutations. Also, the identification of these genes and mutations has been enormously vital to the recent progress in understanding the biology of AD. In sporadic cases (majorly LOAD) the *ε4* allele of the APOE gene was recognized as a major risk factor contributing to the pathogenesis of AD in about 20% of the cases [Cruts and Broeckhoven, 1998].Along with APOE *ε4*, recently there are ten more genes being identified that increase the risk of developing LOAD [Karch *et al*., 2012;Rongve*et al.*, 2013]. However, other causative and risk genes are involved in AD and need to be identified to fully elucidate the etiology of AD [Maruszak *et al.*, 2009; Matsuda *et al.*, 2009; Herskowitz *et al.*, 2012; Natunen *et al*., 2013; Martiskainen *et al*., 2013; Panigrahi and Singh, 2013].

Further considering partial understanding of several potential genes which are suspected for their involvement in AD it is necessary to understand all the genetic aspects for various biological processes and functions. Here we performed statistics based *in silico* analysis on genotype data of the above mentioned prime genes for a specific population CEU i.e. CEPH (Utah residents with ancestry from northern and western Europe). Rationale of selecting this population was to look for a hybrid population with European ancestry and a long term USA residence. All other population data available in HapMap has non-hybrid or direct populations from Nigeria, China and Japan. Various markers were found and Linkage disequilibrium (LD) plots were also

generated [Thorisson *et al.*, 2005]. We have predicted the possible impact of amino acid substitutions on the structure and function of proteins using biophysical and evolutionary comparisons and found mutations that might be 'damaging' or being 'tolerated' by the native protein. There are other mutational databases available for AD like Alzheimer Research Forum (Alzforum, www.alzforum.org) and Alzheimer Disease & Frontotemporal Dementia Mutation Database (AD&FTDMDB, http://www.molgen.vib-ua.be/ADMutations) showing the mutation details of FAD [Kinoshita and Clark, 2007; Cruts *et al.*,2012]. By interrogating the genome-wide LD structures, the haplotypes they constitute, and selecting their tagSNPs, this chapter intends to offer comprehensive maps for complex disease genes association studies. The generated information for genetic markers (nsSNPs, tagSNPs), haplotype blocks, LD, and Phosphorylation states would be of utmost use to researchers working in the area of genomics and molecular genetics of AD and other associated disease. It provides an opportunity to experimental scientists to look for other hybrid populations for further research analysis.

Often AD is not diagnosed in its early symptomatic stages. One possible reason for the limited detection might be the lack of brief screening tests [Galvin *et al.*, 2010]. Much effort has been made to identify and verify diagnostic biomarkers for AD to improve antecedent detection and assist in the development of possible disease-modifying treatments at its early stage. To develop biomarkers to aid in disease diagnosis and prognosis, and assess disease risk are currently underway. Using biomarkers to identify affected individuals prior to the onset of clinical symptoms and associated synaptic or neuronal loss should enable novel clinical trial design and early mechanism based therapeutic intrusion [Fagan and Holtzman, 2010; Panigrahi and Singh, 2012].Biological markers for AD can serve as *in vivo* diagnostic indicators. Markers that change with disease progression may offer utility in assessing the rates of disease progression and the efficacy of potential therapeutic agents on AD pathology [Tarawneh andHoltzman, 2010; Sehgal and Singh, 2012]. Genetic association studies can be made more lucrative by exploiting LD patterns among close SNPs. HapMap currently proposes an opaque SNP map across the human genome in four population samples [Thorisson *et al.*, 2005].Haplotype analysis captures regional LD information of SNPs. Characterization of LD patterns across the human genome is at present an area of ongoing research. Computational identification and analysis of SNPs

could provide biologically significant information [Scacchi *et al*., 2007;Bettens *et al*., 2009;Fede *et al*., 2009; Miar *et al*., 2011; Singh *et al*., 2011]. HapMap was utilized in this study to gather genotype data with respect to their population genetics information. This will facilitate genome wide association analysis and the search for the genetic determinants of complex diseases. Here we performed *in silico* analysis to discover clues forbiomarkers latent in genotype data for the prime or candidate genes to resolve some unanswered or contradictory questions regarding AD.

4.2 MATERIALSANDMETHODS

4.2.1 Database Blueprint

Alzheimer's Disease Database for Genetic Association Studies and Phosphorylation States (*ADDGAP*) resource is developed by acomprehensive workflow.The complete workflow has been divided into components like: data collection, data analysis, interpretation of results, and presentation of results to the scientific community in the form of a database.

4.2.2 Gene and Protein Screening Process

Exploring several biological databases like NCBI and other published studies, we identified 41 genes which have direct involvement in the AD progression. A statistics based *in silico* and quantitative genetic studies have been performed, onthe above collectedcandidate or primegenes andtheir proteins.Alltheprimegenes were categorizedinto3majorclasses, onthebasis of their impact on different AD types (**Figure 4.1**).The first category is named as"EarlyonsetAD", in this category those genes are involved which have considerable impacton EOAD (majorly familial) progression. This is a rare form of AD that strikes people younger than age of 65 years (known to develop between the age 30 and 40 years) and less than 10% of all AD patients have this type. Since they experience premature aging, people with Down's syndrome are predominantly at risk for a form of early onset AD. Similarlythesecondcategoryistermedas "LateonsetAD"and genes havingsignificantroleintheLOAD(majorly sporadic)progression were collected inthisclass [Kowalska, 2004].This is the most common form of AD, accounting for about 90% of cases, and usually occurs after the age of 65 years. LOAD strikes almost half of all people over the age of 85

years and may or may not be hereditary. The causes of LOAD are not yet utterly understood, but they likely include an amalgamation of genetic, environmental, and lifestyle factors that persuade a person's risk for developing the disease. There is another category referred as "General AD"(third category) which was created for all other important genes which are suspected as risk genes for their involvement in AD progression, but their exact status is not being confirmed about their role in two previous categories (familial and sporadic) of AD. This category i.e. "General AD" also contained some genes which plays key role in the formation of Aβ plaques (prominent features found in AD brain) that are BACE1, FURIN, and γ-Secretase etc.

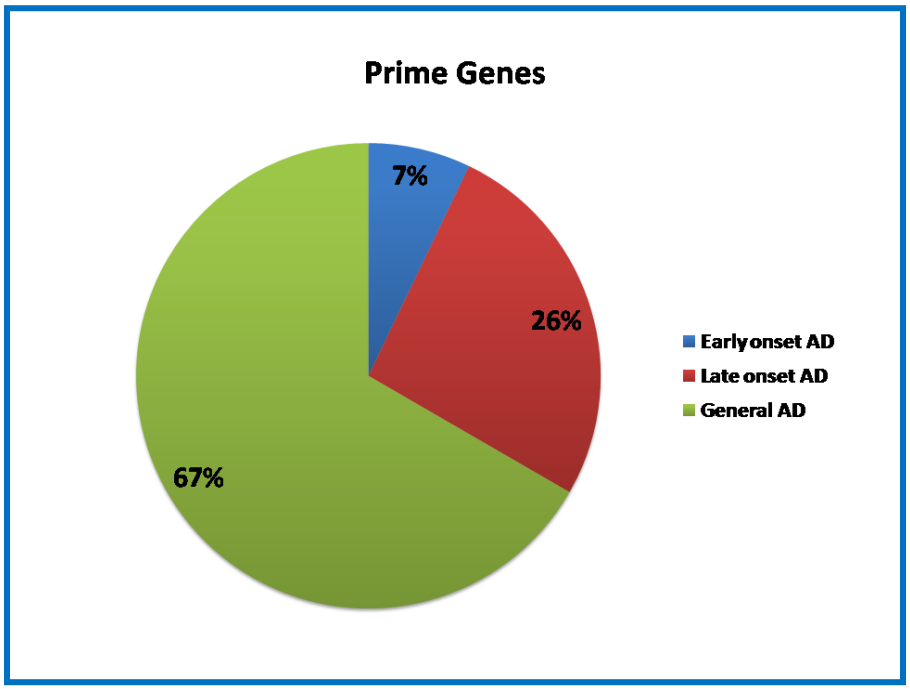


Figure 4.1.Catagorization of Prime genes according to their involvement in AD

4.2.3 Compilation of genotype data and their LD, and haplotype analysis

Parts of the human genome are inherited as blocks known as LD blocks; the polymorphisms located within these LD blocks are also inherited collectively with higher prospect. Generally, the genotypes of genetic variations in a LD block are also interrelated with each other [Savas *et al*., 2013]. Haplotypes are an anthology of many adjoining SNPs on a single chromosome. They

emerge to be more effective because in haplotype analysis polymorphic sites can be genetically distinct. Many genes that respond to drugs often have multiple polymorphisms. In haplotype analysis the combined effect of all of these polymorphisms can be measured. This may provide more information on the alliance of genes to disease [Shastry, 2003].The primary need for the analysis of LD and haplotype is the genotype data. Hence, in this study the genotype data compilation for all the prime genes for a specific population CEU has been compiled from HapMap. TagSNPs are an abridged set of SNPs that captures mainly the genetic variation in a region. They can be used in association studies to diminish the number of SNPs required to detect LD based association between a trait of interest and a region of the genome. To incorporate the selection of edifying markers ('tagSNPs'), this project provides genome-wide data in 269 individuals from four different population groups [Thorisson *et al.*, 2005], and supports the selection tagSNPs by exploiting redundancies among nearby polymorphisms due to LD [Johnson *et al*., 2001].

TagSNPs list for all the prime genes were generated using "Tagger Program" algorithm [Bakker *et al*., 2005]. The minimal coefficient of determination (r^2 value is ≥ 0.8) at which all alleles are to be captured. The composed genotype data for CEU population were further analyzed for several quantitative genetic parameters. For example haplotype frequencies and LD were calculated using Haploview software (Version 4.2) [Barrett *et al.*, 2005] developed at "The Broad Institute", which is based on the expectation maximization (EM) algorithm. The standardized disequilibrium coefficient (D' or D prime) and correlation coefficient (r^2) between these SNPs were also analyzed using the LD plot function of this software to find certain allelic combinations of SNPs. D' is the value of LD linking the two loci or blocks, which can be intended from the subsequent formula:

$$D' = \frac{D}{D_{max}}$$

$$Where,\ D = [(K11)(K22) - (K12)(K21)\ ]$$

and "D_{max}" depends upon the sign of D. If D is positive, then

$$Dmax = min\,[(b1a2)or\,(b2a1)]$$

While if, *D* is negative, then

$$Dmax = min\,[(b1a1)or(b2a2)]$$

Here, value of *D* in the vicinity of zero provides greater amount of hist orical recombination between the two blocks.Where, *b1* and *b2* are the frequencies of alleles at SNP1, *a1* and *a2* are the frequencies of alleles at SNP2 and *K11, K12, K21, K22* are the possible haplotype frequencies. One more essential parameter r^2 i.e. the correlation coefficient between these SNPs is also used here which is deliberated by:

$$r = \frac{D}{(b1b2a1a2)^{1/2}}$$

The squared coefficient of correlation (r^2) is frequently used to eliminate the arbitrary sign thus introduced in the correlation value.

4.3 RESULTS AND DISCUSSION

4.3.3 Findings

In this study, we have analyzed 41 genes (from literature survey and our previous studies [Fagan and Holtzman, 2010; Panigrahi and Singh, 2013]) having significant role in AD progression. Therefore the above genes are termed as prime genes and categorized on the basis of their functional role in AD. General informations for AD like overview, cause, precaution, diagnosis, treatment are hyperlinked in the 'Home page' of *ADDGAP* (In the 1st section present at the left side-bar). Other essential informations such as 'News and Events' and 'Notes' are also provided in the 2nd and 3rd sections just beneath the 1st section. User can access to other pages like 'Search page', 'Help page' 'Reference page' etc., by just clicking on the corresponding menu tab of the website (**Figure 4.2**). The original intent of our work was to provide the database *ADDGAP*, that includes an uncomplicated yet effectual way for the search and retrieval of essential genetic markers and associated information for AD prime genes. LD, haplotypes, other genetic markers, nsSNPs, tagSNPs, pathways, phosphorylation states and other related information were

collected for all the prime genes which have been categorized into 3 major classes as described in early section of this chapter.

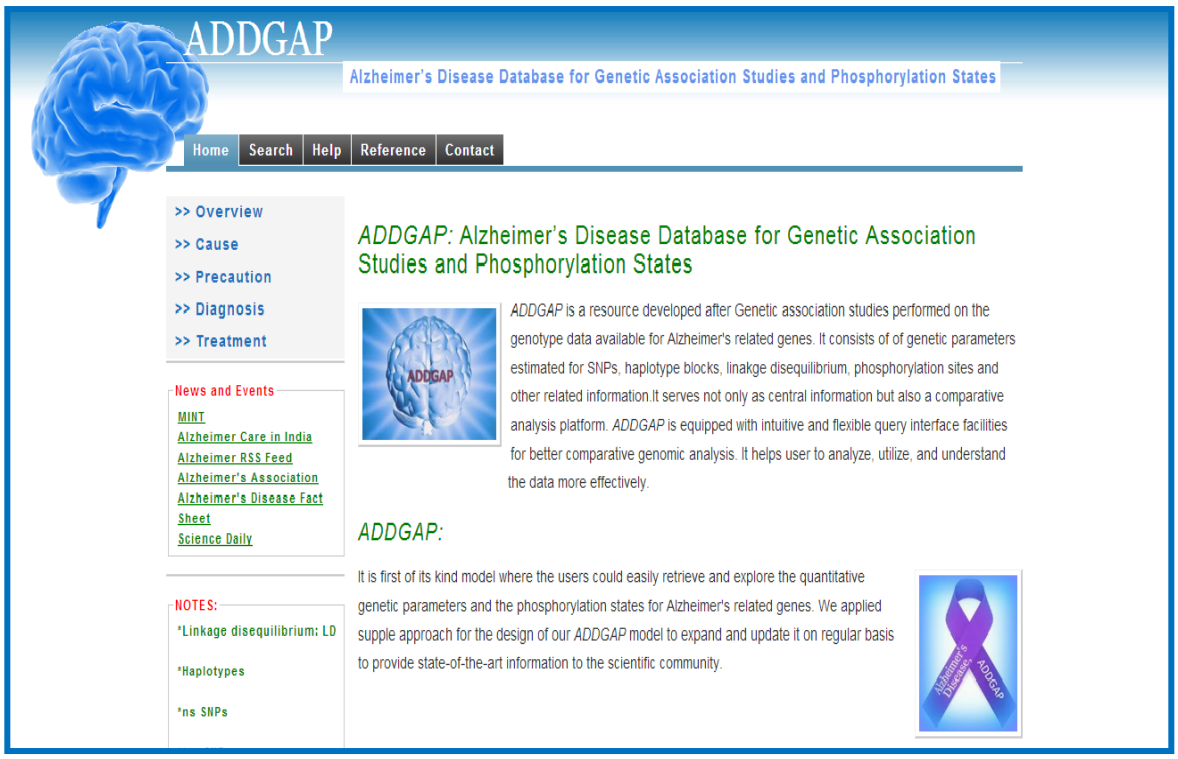


Figure 4.2.Home Page of *ADDGAP*

4.3.4 Web Outline

There are six different types of search options where user can browse our web repository named *ADDGAP*. Options are like search by category, nsSNPs, tag SNPs, haplotypes, LD, genetic markers and phosphorylation sites (**Figure 4.3.A and B**). There is one more search option i.e. advanced search for the user's expediency where the hybrid data is provided for all categories (**Figure 4.3.A).** The former search preference i.e., browse by category option, where user can search for any of the three categories and repossess information like all the genes implicated in that category, their Gene ID's, OMIM ID's, pathways involved, identical to the following **Figure 4.4**. The appropriate literature references corresponding to AD (for all the prime genes) which are linked to NCBI, OMIM, KEGG and PUBMED databases respectively, are also provided in the database. Additionally we provide general information about the overview, cause, precaution,

diagnosis, and treatment of AD being collected from various resources, literature and other databases. Options for all these searches are provided on left menu of the GUI in all the pages except search page.

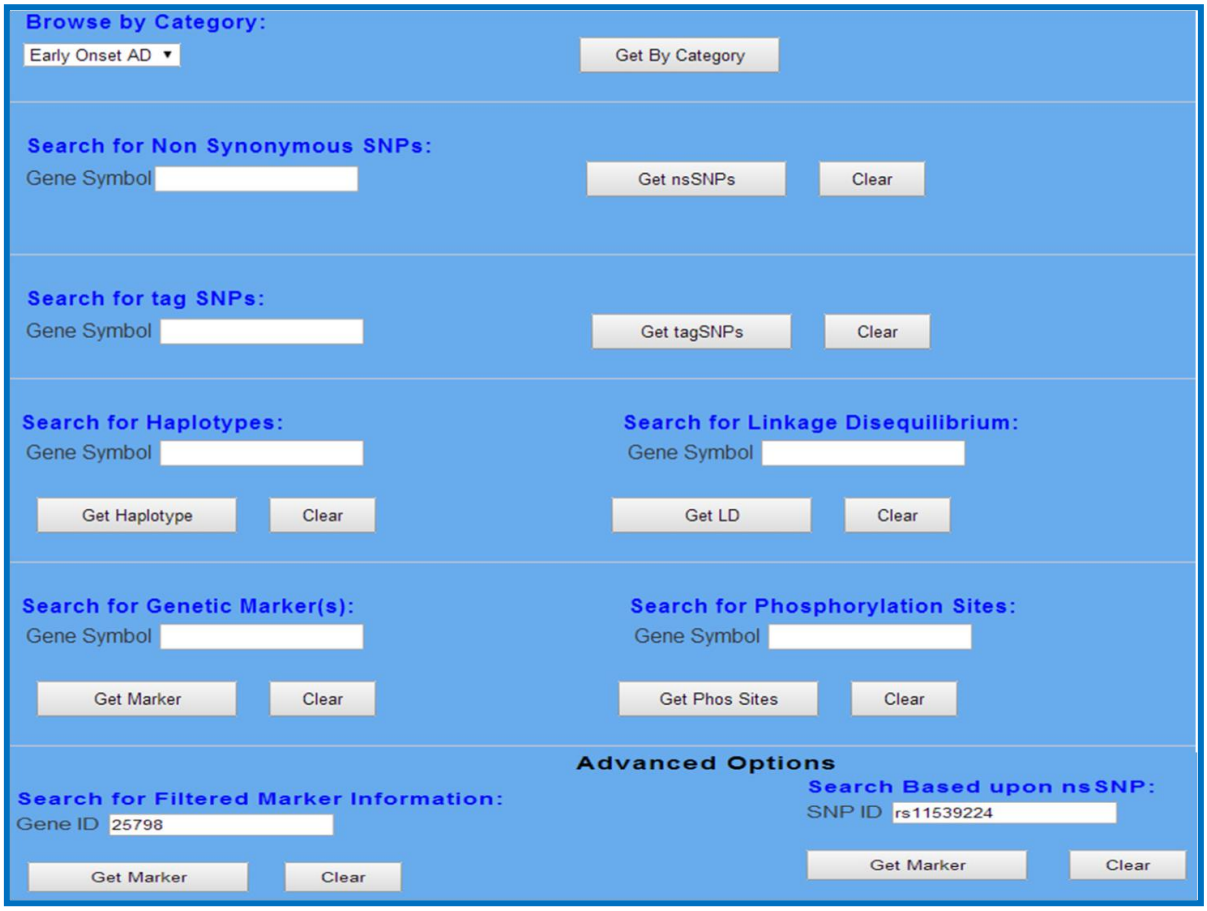


Figure 4.3.A.Search Page of *ADDGAP* with all possible search options

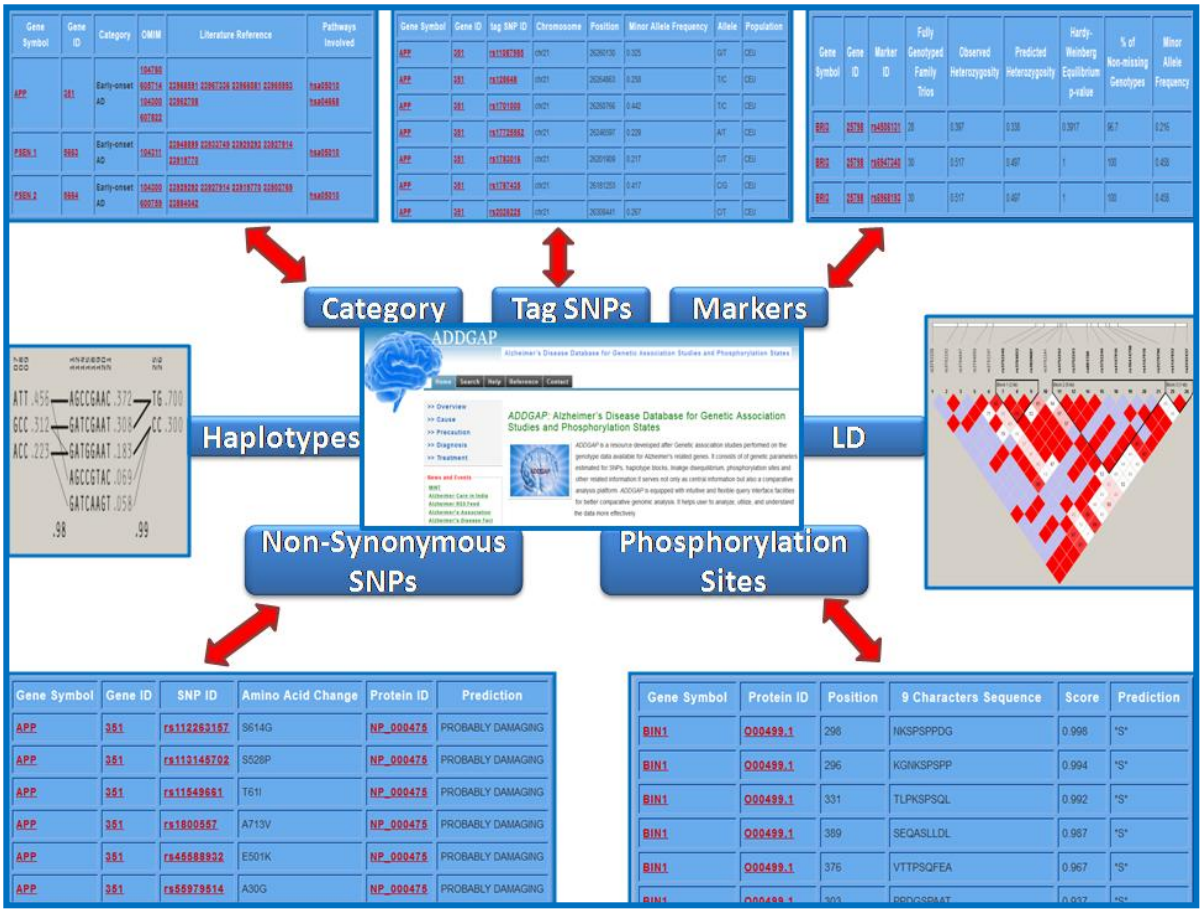
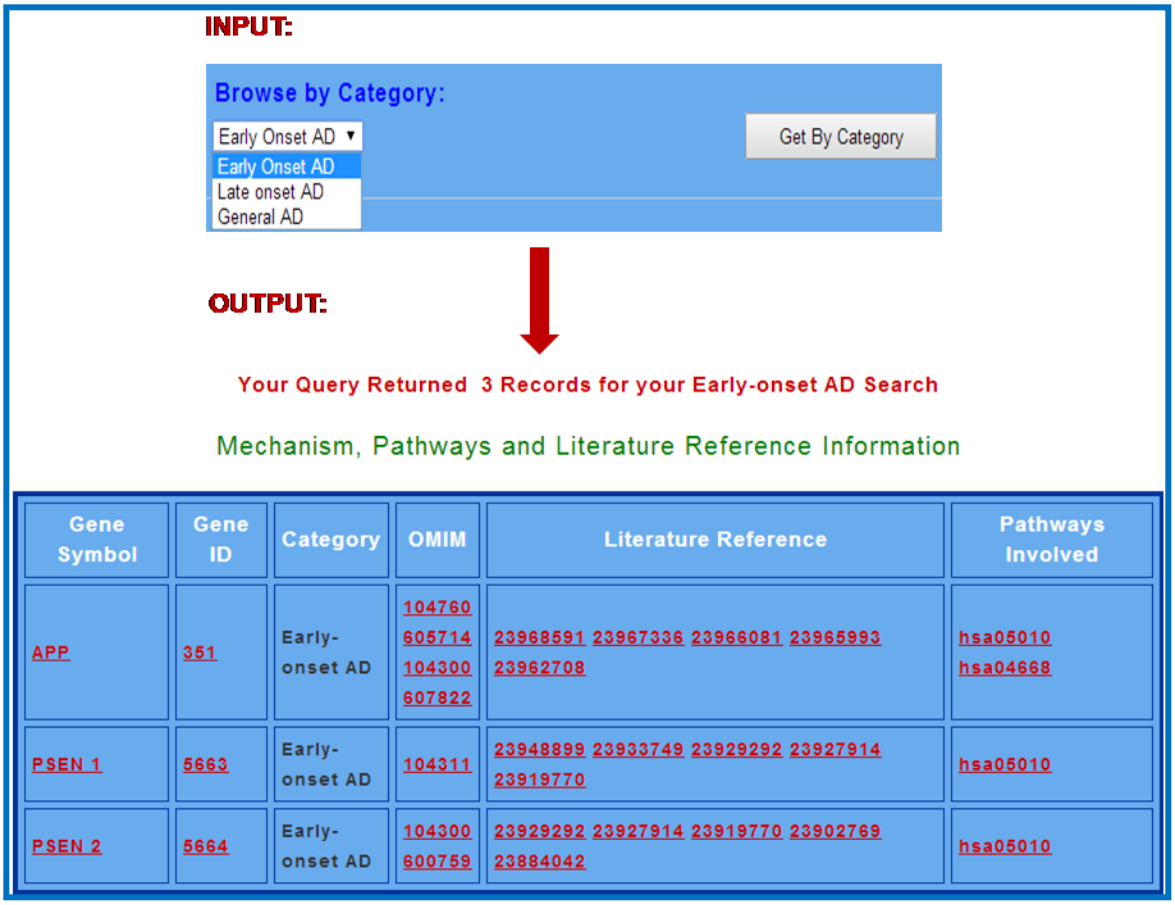


Figure 4.3.B.The web outline with essential facts and some pinpointing results generated and assembled using *ADDGAP*



Gene Symbol	Gene ID	Category	OMIM	Literature Reference	Pathways Involved
APP	351	Early-onset AD	104760 605714 104300 607822	23968591 23967336 23966081 23965993 23962708	hsa05010 hsa04668
PSEN 1	5663	Early-onset AD	104311	23948899 23933749 23929292 23927914 23919770	hsa05010
PSEN 2	5664	Early-onset AD	104300 600759	23929292 23927914 23919770 23902769 23884042	hsa05010

Figure 4.4.Input and Output for the search option, 'browse by category' (for Early-onset ADcategory) in *ADDGAP*

User can get the nsSNPs information of all the prime genes through non-synonymous SNPs search option in *ADDGAP*. By clicking on search for non-synonymous SNPs button, various nsSNPs for the selected gene (Gene Symbol) is displayed (**Figure 4.5**). SNP ID's are linked to dbSNP [Sherry *et al.*, 2001]; amino acid change column gives the position of the change in the sequence and the amino acid being replaced, while the prediction column shows whether the change is damaging or tolerant to the native protein [Schneider *et al.*, 1986; Cargill *et al.*, 1999; Brookes *et al.*, 2000; Palmer *et al.*, 2000; Ng and Henikoff, 2006; Kumar *et al.*, 2009; Adzhubei et al., 2013]. TagSNPs information for all the prime genes are available through tagSNPs search option in *ADDGAP* (**Figure 4.5**). By clicking on search for tagSNPs button, various tagSNPs for the selected genes (Gene Symbol) are displayed. In this search option

user can get the information about the tagSNP ID's whichare linked to dbSNP [Sherry *et al.*, 2001],with their chromosome number, position, minor allele frequency (MAF), allele information (amino acid change) and population from which the tagSNPs data were collected. All the genotype data, that has been collected for all the prime genes are from a particular group of hybrid population known as: CEPH or Utah residents with ancestry from northern and western Europe also called as CEU [Thorisson *et al.*, 2005].

Figure 4.5.Inputs and Outputs for the search options (by taking an example of gene APP), 'search for non-synonymous SNPs' and 'search for tagSNPs'

Information concerning the essential haplotypes can be acquired by clicking on the search for haplotypes option in *ADDGA*P. One can observe parameters like block which gives the current number of haplotype blocks in the particular query gene, number of markers column disports the quantity of markers present in a block. By clicking on the block option, a simple and deducible outlook of the haplotypes in that particular gene is generated and the marker numbers are depicted

on the top (**Figure 4.6**). Haplotype blocks are the clusters of haplotypes; allied SNPs which are conserved right through the genome in the form of patterns. These blocks or clusters correspond to the set of uninterrupted sites which either has petite or no signal of historical recombination. Population frequencies are shown for each haplotypes and common crossings from one cluster to the next are depicted by lines. Thicker lines represent more common crossings than the thinner ones. The multilocus *D'* is also being specifiedin the lower part of image.

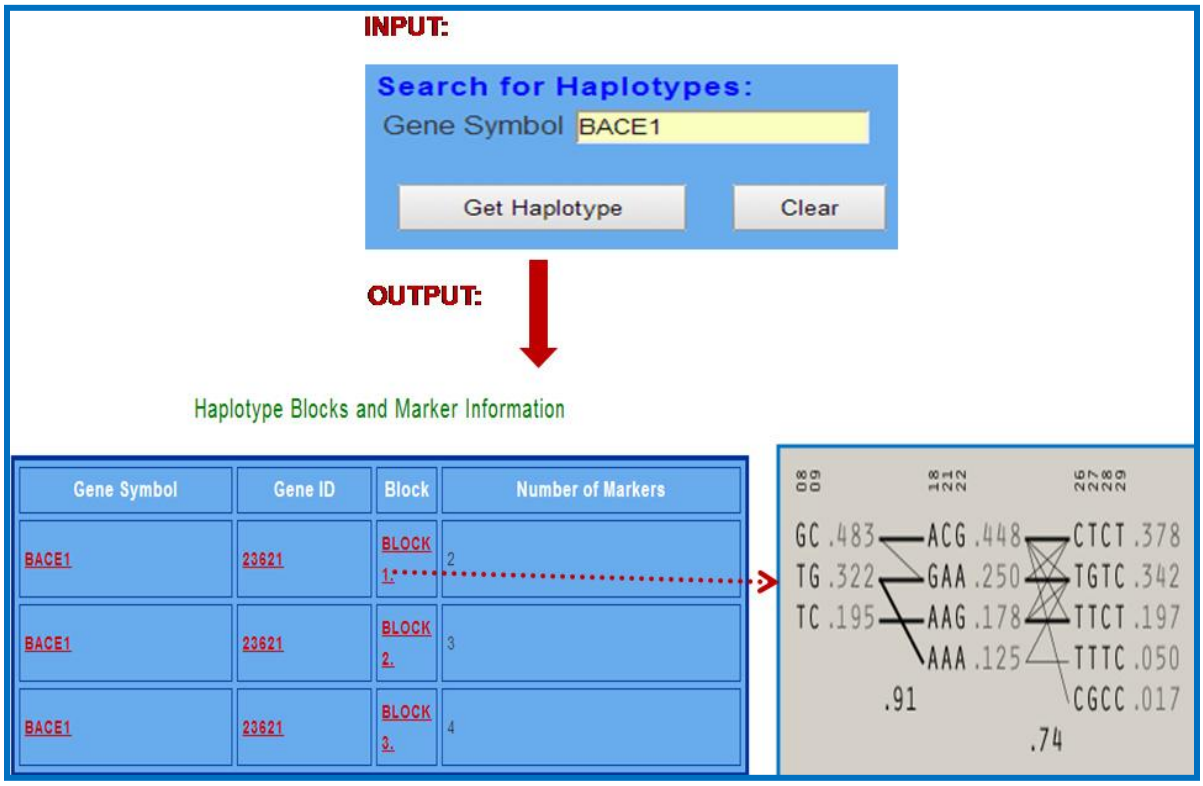


Figure 4.6.Input and Output for the search option 'Search for Haplotypes' in *ADDGAP*. There are 3 haplotype blocks found when searched for the gene 'BACE1' and by clicking on 'BLOCK 1', it shows the character-bar LD plot for block 1 in an another window showing 2 markers

ADDGAP database is also having a LD search option i.e. "Search for Linkage Disequilibrium" where, LD information of various prime genes are scrutinized and assembled (**Figure 4.7.A**). In this study, we have deliberated only the most extensive loci's, whose r^2 value is ≥ 0.6. It includes loci 1 and loci 2, which are the two loci under study, *D'* value between the two

loci, LOD which is the log of the likelihood odds ratio i.e. a measure of confidence in the value of *D'*, correlation coefficient value between loci 1 and loci 2, CI low and CI hi column represents 95% confidence lower bound and upper bound on *D'* respectively, distance (in bases) between the loci. LD image column gives an image of the generated LD plot (**Figure 4.7.B**). In LD plot; the number in each cell shows the LD parameter *D'* (×100). Each cell is color graduated correlating to the strength of LD among the two markers. Squares without numbers represent *D'* values of 1.0; all numbers represent the *D'* value articulated as a percentile. Red squares stand for pairs with LOD score for LD ≥ 2, blue squares symbolize *D'* = 1 but LOD<2, and white squares correspond to LOD<2 and D'<1.0.Characterizing LD is of vital significance for gene mapping studies and can provide insights into the biology of recombination, and human demographic history. It also provides information to map genes that are associated with quantitative characters and inherited diseases, and to understand the joint evolution of linked sets of genes.

There is another search option in *ADDGAP* i.e. search for markers option, where substantial markers recognized in this study have been accumulated and integrated. Here the user can access the marker specific parameters like Marker ID, fully genotyped family trios for the marker (0 for datasets with unrelated individuals); Hardy-Weinberg (H-W) equilibrium *p-value*, i.e. the probability that its deviation from H-W equilibrium could be explained by chance, the percentage of non-missing genotypes for the marker and MAF for the given marker (**Figure 4.8.A**). The marker's observed and predicted heterozygosity calculated by using the following formula:

$$[2 * \mathrm{MAF} * (1 - \mathrm{MAF})]$$

MAF of SNPs is one of the main factors affecting the resulting HapMap, being the factor upon which LD is calculated, haplotypes are constructed and tagSNPs are elected. The cutoff thresholds for the frequency of minor alleles used in the making of the map consequently have intense effects on the resolution of that map. A lower cutoff value is more appropriate for studies in which population-specific haplotypes are crucial. The most appropriate MAF cutoff values may differ between populations and a pilot study for selecting the most appropriate could be justified.

It seems that in general, if affordable, the 0.01 threshold is a good choice for capturing information [Johnson *et al.*, 2001; Barrett*et al.*, 2005; Thorisson *et al.*, 2005].

INPUT:

Search for Linkage Disequilibrium:

Gene Symbol BACE1

Get LD Clear

OUTPUT:

Your Query Returned 14 Records for your LD Search

Linkage Disequilibrium and Association Information

Gene Symbol	Gene ID	Loci 1	Loci 2	D Prime	Likelihood Odds Ratio (LOD)	Correlation b/w Loci1 and Loci2	Confidence Lower Bound(CI Lo)	Confidence Upper Bound(CI Hi)	Distance (bases) b/w Loci1 and Loci2	LD Image
BACE1	23621	rs535860	rs609332	1	17.07	1	0.87	1	15280	LD
BACE1	23621	rs548381	rs687740	1	29.78	0.966	0.93	1	6266	LD
BACE1	23621	rs7083	rs490460	1	28.49	0.905	0.92	1	7213	LD
BACE1	23621	rs593245	rs477036	0.964	24.73	0.866	0.87	1	264	LD

Figure 4.7.A.Input and Output for the search option 'Search for LD' in *ADDGAP*, for the gene BACE1

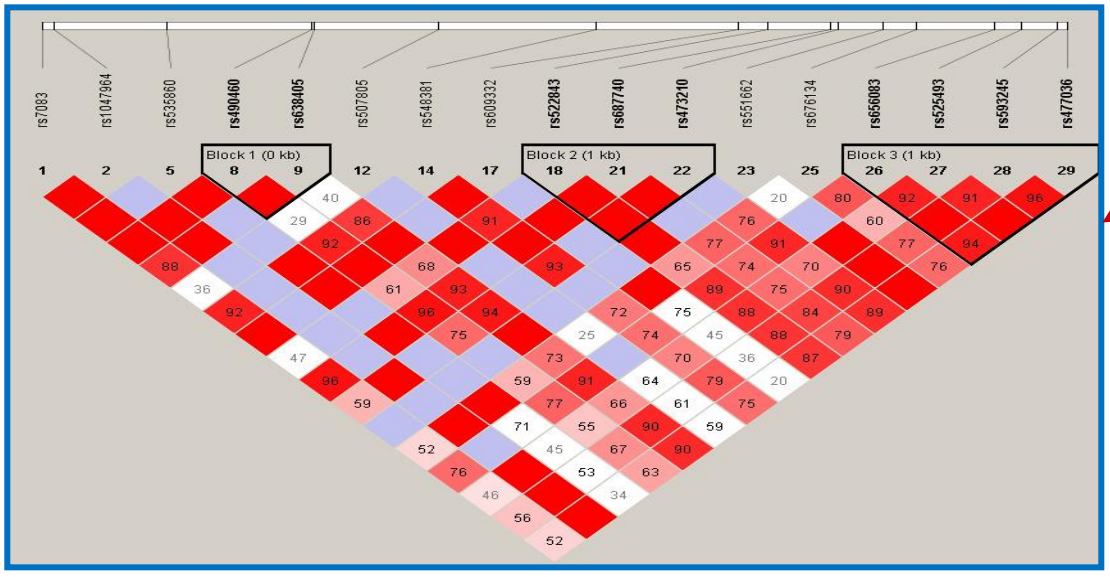
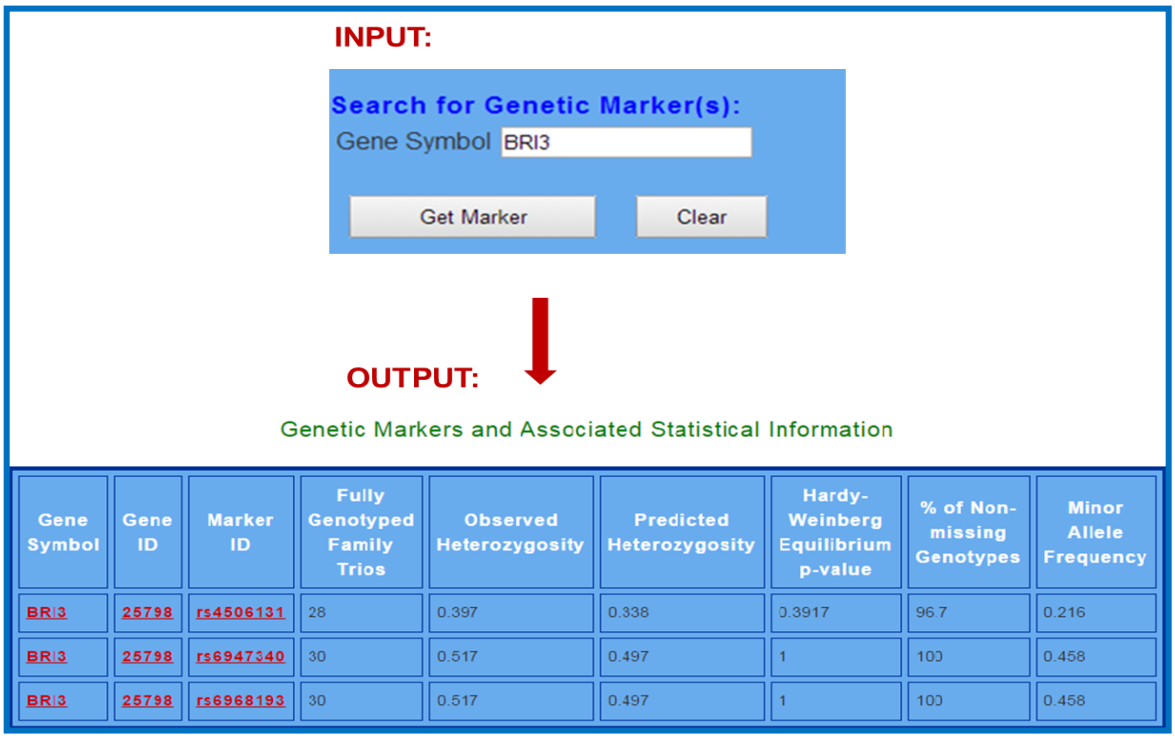


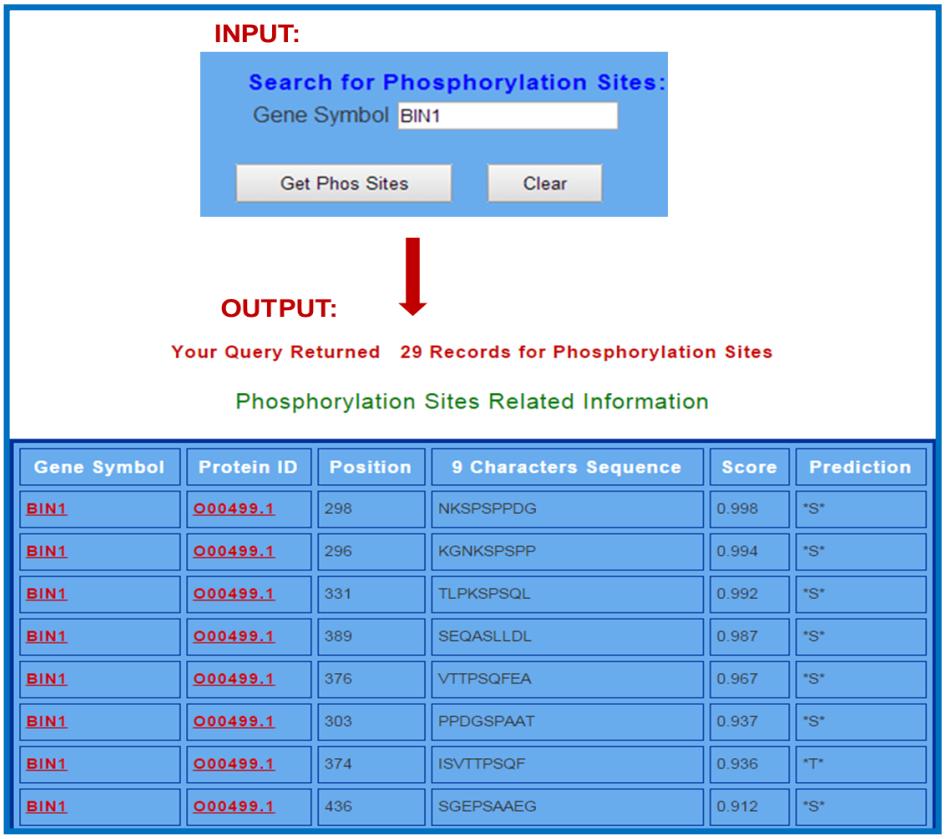
Figure 4.7.B. Clicking on the 'LD' present in the LD Image column (output table in Figure 4.7.A), it will show the image of the generated LD plot (for the gene BACE1) in a separate window

ADDGAP also provides the option for search of phosphorylation states, where user can get information for the predicted phosphorylation sites in the AD protein sequences (**Figure 4.8.B**). There are evidences where phosphorylation is being associatedwith NFTs in AD and suggest that such information might be crucial for the progression of the disease [Peel *et al.*, 2004]. Therefore we predicted and include phosphorylation information which will be crucial for the biological basis of disease progression. There is another study performed for similar kind of analysis and authors developed a database which provides state-of-the-art information on human DNA repair systems [Sehgal and Singh, 2014]. It is anticipated that *ADDGAP* will also be beneficial to the scientific community. The information includes; the protein ID, location of the phosphorylation site in the sequence, 9 character sequence depicting the phosphorylated residue at the correct center of the sequence, phosphorylation sites having predicted scores more than 0.5 has been selected for potential phosphorylation sites and the prediction column shows the putative phosphorylated residues (Serine (S) or Threonine (T) or Tyrosine (Y)) [Blom *et al.*, 1999; Kreegipuu et al., 1999; Savas and Ozcelik, 2005]. Overall phosphorylated residues (S, T, Y) for each group have been given in **Table 4.1**.



Gene Symbol	Gene ID	Marker ID	Fully Genotyped Family Trios	Observed Heterozygosity	Predicted Heterozygosity	Hardy-Weinberg Equilibrium p-value	% of Non-missing Genotypes	Minor Allele Frequency
BRI3	25798	rs4506131	28	0.397	0.338	0.3917	96.7	0.216
BRI3	25798	rs6947340	30	0.517	0.497	1	100	0.458
BRI3	25798	rs6968193	30	0.517	0.497	1	100	0.458

Figure 4.8.A.Input and Output for the search option 'Search for Genetic Marker(s)' in *ADDGAP*, for the gene BRI3



Gene Symbol	Protein ID	Position	9 Characters Sequence	Score	Prediction
BIN1	O00499.1	298	NKSPSPPDG	0.998	*S*
BIN1	O00499.1	296	KGNKSPSPP	0.994	*S*
BIN1	O00499.1	331	TLPKSPSQL	0.992	*S*
BIN1	O00499.1	389	SEQASLLDL	0.987	*S*
BIN1	O00499.1	376	VTTPSQFEA	0.967	*S*
BIN1	O00499.1	303	PPDGSPAAT	0.937	*S*
BIN1	O00499.1	374	ISVTTPSQF	0.936	*T*
BIN1	O00499.1	436	SGEPSAAEG	0.912	*S*

Figure 4.8.B. Inputs and Outputs for option 'Search for Phosphorylation Sites', for the gene BIN1

It also provides one more search category i.e. 'Advanced Option', as we have mentioned above, where user can search for the marker data using two advance search options that are "Search for Filtered Marker Information" and "Search Based upon ns SNP" (**Figure 4.9**). There are few resources available for AD but there is no such catalog for AD prime genes which provides all these essential quantitative genetic details which includes LD, haplotypes, nsSNPs, tagSNPs, disease related functional information, along with phosphorylation states on a common platform. In this regard *ADDGAP* is first of its kind model where the users could easily retrieve and explore the quantitative genetic parameters and the phosphorylation states for AD prime genes. We applied agile approach for the design of our *ADDGAP* model to expand and update it on regular basis to provide state-of-the-art information to the scientific community.

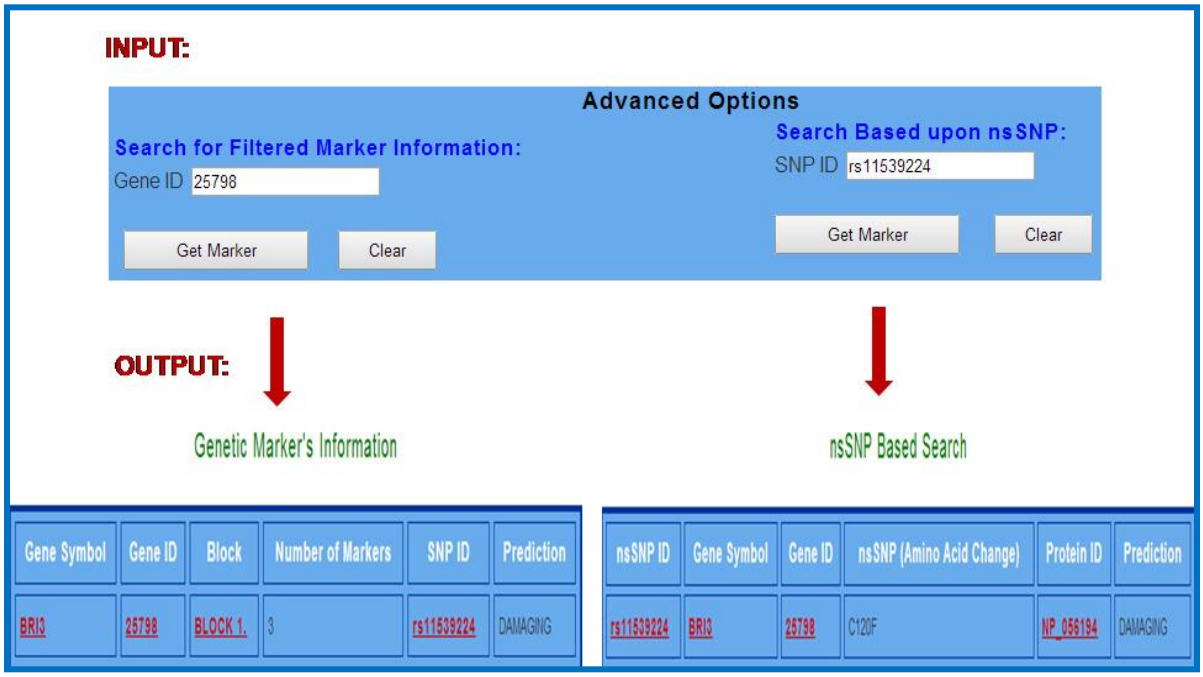


Figure 4.9.Inputs and Outputs for the search options (by taking an example of gene id 25798 and SNP id rs11539224), 'Search for Filtered Marker Information' and 'Search Based upon ns SNP' respectively

Table 4.1 No. of predicted nsSNPs that alter amino acid sequence and phosphorylation sites of the prime genes and proteins

Categories	**No. of Genes**	**Gene Symbols**	**No. of Predicted Damaging nsSNPs**	**Predicted Phosphorylated Residues**		
				S	**T**	**Y**
Early onset AD	3	*APP, PSEN1, PSEN2*	193	41	31	15

Late onset AD	11	*APOE ε4, CLU, PICALM, CR1, BIN1, MS4A, CD2AP, CD33, ABCA7, EPHA, TREM2*	59	133	68	28
General AD	27	*BACE1, NCSTN, FURIN, PLSCR1, LRP1, GGA1, APBB1, ATP1B1, ATP1B4, BRI3, CMPK1, CSNK1D, GAMMA SECRETASE, GGA2, GGA3,IGHG1, IGK@, ITM2B, ITM2C, KHSRP, MMP2, PDIA3, PSAP, RANBP9, RTN3, SELPLG, SORL1*	98	553	187	138

4.4 CONCLUSION

It is believed that *ADDGAP* will be a beneficial resource and convenient platform for researchers who are focusing on AD and its genetic studies. It serves not only as central information but also a comparative analysis platform. *ADDGAP* is equipped with intuitive and flexible query interface facilities for improved comparative genomic analysis. It helps user to analyze, utilize, and understand the data more effectively. We plan to update it on regular basis as per the availability of genotype data for other genetic entities for AD. Apart from several advantages of the project, *ADDGAP* have some limitations. The search criteria are limited and the database is population specific which we plan to resolve in its next version. We provide additional general information of AD besides all other quantative genetic parameters. We believe that the biological community would get benefit from having a new web repository having unique information of AD. It is anticipated that this web based comprehensive resource would serve as a useful accompaniment for analyzing AD and will also contribute scientific knowledge towards better understanding of AD and its regulatory processes.

CHAPTER-5

Study to test the binding of THC-δ9-tetrahydrocannabinol and derivatives on acetylcholine binding protein: a virtual screening and molecular docking study

ABSTRACT

Accumulation of Aβ and cholinergic deficiency are two prominent pathological descriptions for AD. The topographical characteristics of these different pathological processes in AD brain and how these relate to each other is still imprecise. There is a need to study these pathological descriptions and to prepare some putative molecules which could help in the cure of AD. AChE is an enzyme found in the synapse between nerve cells and muscle cells. It waits patiently and springs into action, soon after a signal is passed, breaking down the ACh into its two component parts that are acetic acid and choline. AChE is the target of many AD drugs which block the function of AChE and thus cause excessive ACh to accumulate in the synaptic cleft (**Figure 1.8, Chapter 1**). Indeed, it was shown that AChE inhibitors improve the cognitive abilities of AD patients at early stages of the disease development. Therefore, it may be implicated as a potential cure for neurological disorders such as AD and Myasthenia Gravis. Acetylcholine binding protein (AChBP) is a soluble protein found in the *snail lymnaea stagnalis*. It is produced and stored in glial cells and released in an ACh dependent manner into the synaptic cleft, where it modulates synaptic transmission. The study has been conducted to reveal the binding mode of the derivatives of delta-9-tetrahydrocannabinol (Δ9-THC) with AChBP. In this study we aimed to discover new molecules that can bind to the nicotinic ligand binding site. Here we provide a framework for the design of compounds with potential therapeutic applications and hint for the future design of new derivatives with higher potency and specificity. In the present study 3000 molecules were initially selected and screened using the Lamarckian Genetic algorithm (LGA) of AutoDock, from which top 20 lead molecules were chosen showing bond with TRP143 residue. Induced fit docking (IFD) protocol and TOPKAT were applied to these top 20 lead molecules to determine their binding potential and toxicity. QIKPROP was utilized for ADME properties. Molecular dynamics (MDs) study was done for the molecule which has shown maximum affinity during IFD. It is believed that the proposed set of putative or lead molecules ofΔ9-THC and its derivatives along with most potential candidate molecular agent (S1, i.e. molecule number 25) provide diagnostic tools to identify appropriate markers at earlier stages of the disease and will be helpful for AD drug development and for clinical trials.

5.1. INTRODUCTION

Tetrahydrocannabinol (THC), also known as delta-9-tetrahydrocannabinol (Δ9-THC), Δ1-THC or dronabinol, is the crucial psychoactive substance found in the cannabis plant. It was first isolated in 1964 [Gaoni and Mechoulam, 1964; Geller, 2007], in pure form. It is a glassy solid when cold and becomes viscous and sticky if warmed. An aromatic terpenoid, THC has a very low solubility in water, but has first-rate solubility in most organic solvents. Importantly, they also led on to the discovery that many of the effects produced by Δ9-THC and its synthetic cousins depend on the ability of these ligands to target a new family of receptors. Cannabinoid receptors 1 and 2 (CB1and CB2) receptors, both were G-protein coupled receptors (GPCR). The first endogenous cannabinoid receptor agonists (endocannabinoids) to be identified were N-arachidonoylethanolamine (anandamide) and 2-arachidonoylglycerol [Devane *et al.*,1992; Mechoulam*et al.*, 1995; Sugiura *et al.*, 1995], each of which can activate both CB1 and CB2 receptors and is synthesized on demand in response to elevations of intracellular calcium[Howlett *et al.*, 2002; Di Marzo*et al*., 2005].The pharmacological actions of THC result from its binding to the cannabinoid receptor CB1, located mainly in the central nervous system and the CB2 receptor mainly present in cells of the immune system. It acts as a partial agonist on both receptors, i.e., it activates them but not to their full extent [Iwamura *et al.*, 2001]. The psychoactive effects of THC are mediated by its activation of the CB1 receptor, which is the most abundant GPCR in the brain. The presence of these specialized receptors in the brain implied to researchers that endogenous cannabinoids are manufactured by the body, so the search began for a substance normally manufactured in the brain that binds to these receptors, the so called natural ligand or agonist, leading to the eventual discovery of anandamide, 2-arachidonoylglyceride, and other related compounds known as endocannabinoids [Hanus *et al.*, 2001].

This is similar to the story of the discovery of endogenous opiates (endorphins, enkephalins, and dynorphin), after the realization that morphine and other opiates bind to specific receptors in the brain. In addition, it has been shown that cannabinoids, through an unknown mechanism, activate endogenous opioid pathways involving the μ1 opioid receptor, precipitating a dopamine release in the nucleus accumbens. The effects of the drug can be suppressed by the CB1 cannabinoid receptor antagonist rimonabant (SR141716A) as well as opioid receptor antagonists

(opioid blockers) naloxone and naloxonazine [Lupica *et al.*, 2004]. THC's anticholinesterase action [Brown, 1972;Eubanks *et al.*, 2006], may implicate it as a potential treatment for neurological disorders. Nicotinic acetylcholine receptors (nAChRs) and the structurally allied GABAA, GABAC, 5HT3 serotonin, and glycine receptors are well studied, pharmacologically imperative ligands gated ion channels (LGICs) in the central and the peripheral nervous system[Le Novere and Changeux, 2001; Karlin*et al.*, 2002]. LGICs are implicated in important aspects of brain functioning and disease, mutations in these receptors lead to diseases such as congenital myasthenia gravis, epilepsy, alcohol abuse (nAChRs, GABAARs) or startle syndrome (glycine receptors) [Vafa and Schofield, 1998]. The super family of pentameric ligand gated ion channels (LGICs), including nAChR, 5-HT3, GABAA and GABAC, and glycine receptors, mediates chemical synaptic transmission. The nAChRs are extensively studied and can be divided into muscle and neuronal types [Corringer *et al.*, 2000]. The muscle type, with stoichiometry (α1)2β1γδ, is found at the neuromuscular junction and in the electric organs of fish such as the electric ray Torpedo californica [Sussman *et al.*, 1991]. It presents a common drug target for muscle relaxants. The neuronal nAChRs, which are hetero- or homo-meric (for example, (α4)2(β2)3 or (α7)5), are located on both pre- and post-synaptic nerve terminals. Clinical applications of Δ9-THC and nabilone are available for appetite stimulation and antiemesis and medicines containing Δ9-THC and cannabidiol (CBD) such as Sativex (GW Pharmaceuticals, Salisbury, Wiltshire, UK), are available for the symptomatic relief of neuropathic pain in patients with multiple sclerosis and of cancer pain.

Clinical evidences were obtained supporting the introduction of Δ9-THC or other cannabinoid receptor agonists into the clinical treatments, such as the management of disorders such as glaucoma and cancer, and for the relief of post-operative pain, spasms and spasticity caused by multiple sclerosis and painful spasticity triggered by spinal cord injury [Grotenhermen and Müller-Vahl, 2012]. It has been evidently found that the brain and associated cortex, hippocampus, cerebellum and basal ganglia are rich in CB1 receptor [Herkenham *et al.*, 1991]. CB1 receptors are behind the cognition, learning and memory effects of cannabinoids. Cannabinoids are reported as neuroprotective in Aβ induced neurological damage, decreasing neuroinflammation and regulating neuroregeneration in adult brain [Benito *et al.*, 2003; Iuvone *et*

al., 2004; Campbell and Gowran, 2007]. THC was found to be more effective inhibitor of the acetylcholinesterase induced accumulation of the Aβ protein than the approved drugs donepezil and tacrine. Two studies indicate that THC also has an anticholinesterase action [Brown, 1972; Eubanks *et al.*, 2006], which may implicate it as a potential treatment for AD and Myasthenia Gravis. They are important drug targets as they mediate nicotine addiction in smokers and the positive effects of nicotine on cognition, memory and attention in patients, for example AD and Parkinson's diseases [Paterson and Nordberg, 2000]. Therefore this study has been conducted, in order to reveal the binding mode of the derivatives of Δ9-THC with AChBP to discover new molecules that can bind to the ACh and nicotinic ligand binding site and finally leads to the cure of neurological disorders such as AD and Myasthenia Gravis. It is anticipated that this study will also provide a framework for the design of compounds with potential therapeutic applications and hint to the future design of new derivatives with higher potency and specificity.

5.2 MATERIALS & METHODS

All computational analyses were carried out on DELL PowerEdge T110II server and DELL Precision T3600 workstation with 8 GB RAM each and Intel Xeon processorE3-1230v2, 3.30 and E5-1603, 2.80 GHz. respectively.

5.2.1 Molecular Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second molecule when bound to each other to form a stable complex [Lengauer and Rarey, 1996]. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs [Kitchen *et al.*, 2004]. Molecular docking can be thought of as a problem of "lock-and-key". Here, the protein can be thought of as the "lock" and the ligand can be thought of as a "key". Molecular docking may be defined as an optimization problem, which would describe the "best-fit" orientation of a ligand that binds to a particular protein of interest. However, since both the ligand

and the protein are flexible, a "hand-in-glove" analogy is more appropriate than "lock-and-key" [Jorgensen, 1991]. During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational adjustments resulting in the overall binding is referred to as "induced-fit" [Wei *et al.*, 2004]. We applied similar approach for the docking of AChBP to the designed derivative molecules to take the advantage of "lock-and-key" mechanism.

Two approaches are particularly popular within the molecular docking community. One approach uses a matching technique that describes the protein and the ligand as complementary surfaces [Meng *et al.*, 1992; Morris *et al.*, 1998]. The second approach simulates the actual docking process in which the ligand-protein pairwise interaction energies are calculated [Feig *etal.*, 2004]. Docking was performed where energy evaluations were combined through grids of affinity potential employing various search algorithms to find the suitable binding position for a ligand on a given protein. While docking, polar hydrogens were added to the ligands. Docking of AChBP to these molecules was carried out using Lamarckian Genetic Algorithm (LGA) [Morris *et al.*, 1998], which was performed using AutoDock 4.2 package [Norgan *et al.*, 2011] with standard docking protocol on the basis of a population size of 150 randomly placed individuals; a maximum number of 2.5 *107 energy evaluations, a mutation rate of 0.02, a crossover rate of 0.80 and an elitism value of 1. Ten independent docking runs were carried out for each ligand and results were clustered according to the 1.0 Å RMSD criteria. The grid maps representing the proteins were calculated using auto grid and grid size was set to 60*60*60 points with grid spacing of 0.375 Å. The coordinate of the docked protein along with the ligand was visualized within 6.5 Å region to discover the H-bond interaction with active site residue of AChBP [Pettersen *et al.*, 2004]. In order to generate a comparative study for more comprehensible outcomes, CDOCKER was used in this study with standard parameters.

5.2.1.1 Substrate selection

Derivatives of Δ9-THC were chosen from different Chemical Databases like PUBCHEM and CHEMBANK [Bolton *et al.*, 2008; Seiler *et al.*, 2008] and the chemical structures were generated from SMILES (Simplified Molecular Input Line Entry Specification) notation by using the

ChemSketch Software [2007. ACD/ChemSketch Freeware, version 11.01, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com]. Derivatives were drawn and converted by software to comprehend the *in silico* work.

5.2.1.2 Rigid docking

Three thousand molecules were selected for the study and initially screened using the AutoDock 4.2. AChBP is a soluble protein found in the *snail lymnaea stagnalis*[Smit *et al.*, 2001]. It is produced and stored in glial cells and released in an ACh dependent manner into the synaptic cleft, where it modulates synaptic transmission. Mature AChBP is 210 residues long and forms a stable homopentamer. It aligns with the N-terminal domains of pentameric LGICs and lacks the transmembrane and intracellular domains present in the super family. Nearly all residues that are conserved within the nAChR family, present in AChBP, including those that are relevant for ligand binding. Moreover, AChBP binds known nAChR agonists and competitive antagonists such as ACh, nicotine, d-tubocurarine and a-bungarotoxin. Therefore, AChBP can be used as an example of the N-terminal domain of a α-subunit of nAChRs. We have taken the X-Ray structure of AChBP complexed with carbamylcholine from RCSB (Research Collaboratory for Structural Bioinformatics) Protein Data Bank (PDB) [Berman *et al.*, 2000] having PDB ID '1UV6', and resolution 2.50 Å [Celie *et al.*, 2004]. To study the binding mode of ligand, the active site of the protein has to be find out which was done by literature review as well as software. Combining a novel algorithm for rapid binding site identification and evaluation with easy to use property visualization tools, AutoDock provides an efficient means to find the characteristics of ligand binding sites and gives ways to better exploit them. In order to generate a comparative study for more comprehensible outcomes, CDOCKER protocol of Discovery Studio 3.5 (Accelrys Software Inc.) was used in this study.

5.2.1.3 Induced Fit Docking (IFD)

The starting structures for virtual study experiments of AChBP were retrieved from the RCSB-PDB with the PDB ID 1UV6. Bound ligands, waters beyond 5 Å, ions, molecules and hetero-atoms were removed from the complexes. Missing disulphide bonds were added. The H-bonds

were optimized using protassign at pH 7.0. Restrained minimization was done by impref with the convergences of the heavy atoms to RMSD 0.30 Å using OPLS-2005 force field. The ligands used were sketched using Maestro 9.3 and saved in sdf format. The ligands were prepared by using the Ligprep application of the Maestro 9.3 using the force field OPLS-2005. IFD procedure combines rigid-receptor docking with protein refinement [Sherman *et al.*, 2005; Sherman *et al.*, 2006]. IFD protocol integrates the principle of MPS (Multiple Protein Structure) technique by accounting protein flexibility during the docking of ligands [Bowman *et al.*, 2007]. Schrödinger recommended IFD procedures were followed. During the first phase of IFD, it performs an initial softened-potential docking of the ligands to the rigid receptor (PDB ID 1UV6), with van der Waal radii scaling of 0.5Å for the ligands and protein. Sampling of the protein was performed using Prime for each top 20 ligand poses which were ranked by Glide Score. Residues within the range of 5Å of ligand poses were refined which comprises of side-chain conformational search and optimization, tailed by full minimization of the residues and the ligand. The flexibility of protein was taken into consideration while utilizing this technique. Complexes which lie within 30.0 kcal/mol of the minimum energy structure were re docked into each low-energy. We have utilized induced-fit structure with default Glide settings having van der Waal radii scaling of 1.0 Å for protein and 0.8 Å for the ligand. The complexes obtained were ranked according to the IFD score which reflects both the docking as well as solvation energy.

5.2.2 Toxicity Prediction

Toxicity prediction by computer assisted technology (TOPKAT) models is based on a list of substructures and continuous-valued descriptors. The software is consisting of different modules for the prediction of multiple toxic endpoints. The models for this system are based on predefined lists of chemical descriptors, including 2D structural alerts (SAs) and continuous-valued descriptors. A number of systems effectively automate Quantitative structure-activity relationship (QSAR) predictions and TOPKAT is probably the best known and developed protocol. TOPKAT provides predictions for a variety of endpoints, based on QSAR analyses of large heterogeneous databases. However, the TOPKAT methodology does not allow for the development of models from a mechanistic standpoint. A unique feature of the software, is the optimum prediction space algorithm [Eriksson *et al.*, 2003] which allows for an estimate of the confidence of the prediction

to be assigned. TOPKAT study was performed for the molecules which were obtained upon the initial screening by AutoDock and CDOCKER showing bonding with TRP143 and other important residues (**Table 5.2**).

5.2.3 Prediction of Drug like Property

The ADME (Absorption, Distribution, Metabolism and Excretion) profile has a major impact on the likelihood of success of a lead molecule. Poor ADME characteristics are one of the reasons why drug candidates don't succeed in clinical trials. ADME properties of the top scoring docked compounds were predicted using QikProp module of Schrödinger 2012. The module predicts properties such as log Po/w , IC50 values for blockage of HERG K+channels, log BB, overall CNS activity, Caco-2 and MDCK cell permeability, logKhsa for human serum albumin binding and human oral bioavailability [Jorgensen and Duffy, 2002] (**Table 5.3**). Prediction of these characteristics proves the suitable candidates as potent markers and could be taken further for experimental verifications and clinical trials.

5.2.4 Molecular Dynamics Simulation

The first step in order to obtain results from a molecular dynamics (MD) simulation is the geometric optimization, that is, a procedure to eliminate bad contact between atoms and incorrect angles and dihedrals in the structure. To do this a force field must be associated to the system, ions and counter ions must be added in the sense to neutralize the system and the solvent molecules must be added to build periodic boundary conditions. Besides, an ensemble must be considered in order to use the statistical thermodynamics analysis. The system considered in this study was composed by two monomers and one ligand (S1 molecule (number 25 in table 1 and 2)) between then, placed in the binding region. After the addition of counter ions (Na+) and solvation water molecules (TIP3P model) the all system contains almost 35 thousand atoms. The force field used was the AMBER ff03 [Hawkins *et al.*, 1995; Hawkins *et al.*, 1996] and the minimization procedure used 20 thousand steps divided into two parts: the first one was with 15 thousand steps using the steepest descending algorithm and the other 5 thousand steps were done using the gradient conjugated algorithm. The temperature was maintained at 310K by Nosé-Hoover

thermostat [Hoover, 1985] and the pressure was controlled by the Parrinelo-Rahmanbarostat [Parrinello and Rahman, 1981], constituting the constant-temperature, constant-pressure (NPT) ensemble. In sense to account the intermolecular interactions, a cut off radius of 8Å was used in the Particle Mesh Edwald method [Darden *et al.*, 1993]. The simulation time was 2ns with a time step of 8fs.

5.3 RESULTS AND DISCUSSION

5.3.1 Selection of potent lead compounds

An integrated multifaceted approach is applied to analyze this polygenic disease (**Figure 5.1**). Molecular docking approach using LGA was carried out to elucidate the extent of specificity of AChBP towards different classes of Δ9-THC. Combining an efficient algorithm for rapid binding site identification and evaluation with easy-to-use property visualization tools, the software has provided an efficient means to find and better exploit the characteristics of ligand binding sites. Selection of potent inhibitors was done on the basis of well-established standard parameters such as binding energies and Lipinski's Rule of 5. Total number of molecules was 3000 in number which were virtually screened from different databases on the basis of the structural similarity of Δ9-THC. Exact query "similar to C12 = CC(CCCCC) = CC(O[H]) = C1 [C@@] 3([H]) [C@@] ([H]) (C(C)(C)O2) CCC (C) = C3 using the Tanimoto metric with a distance of 0.5". The docking result of the study of 3000 molecules demonstrated that the binding energies were in the range of -14.34 kcal/mol to -3.26 kcal/mol, with the minimum binding energy of -14.34 kcal/mol. After getting the above results, we have used 2 filtration criteria for filtered out the most significant molecules. Those are, 1st the molecules having binding energies below '-9.00 kcal/mol' and 2ndthose molecules showing hydrogen bond with the active site residue 'TRP143'. 80 molecules were filtered out using the 1st filtrations and out of them 20 lead molecules were selected using the 2nd criteria.

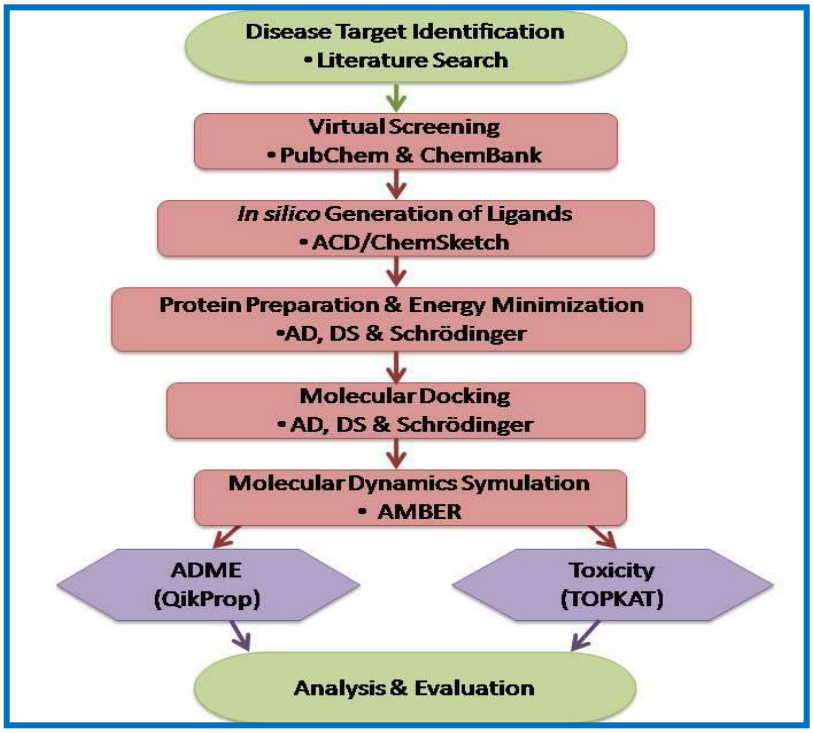
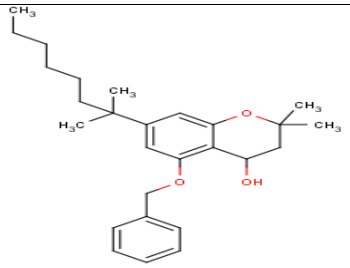


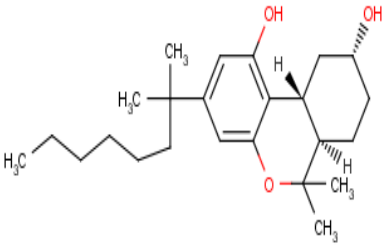
Figure 5.1.Pictographic representation of an integrated multifaceted approach applied in this study which describes the overall methodology (AD: AutoDock 4.2, DS: Discovery Studio 3.5)

5.3.2 Molecular Docking

Initial screening of 3000 molecules which was done by utilizing the LGA of AutoDock 4.2, helped us to screen out the top 20 lead molecules which were showing bond with TRP143 residue, which is essential for the biological activity. The top scoring 20 lead molecules have binding energy ranging from -14.34 kcal/mol to -9.88 kcal/mol. IFD protocol was applied to the top scoring 20 lead molecules, in order to determine the binding affinity and conformation of these top scoring molecules during the ligand induced receptor movement. This protocol accounts for both ligand and receptor flexibility and accurately predicts the conformation of protein-ligand binding complexes. The IFD score obtained was maximum for molecule 25 with an IFD score of -459.966 and minimum for molecule 43 with an IFD score of -454.942 (**Table 5.1**, Top 10 lead molecules are given). The top scoring lead molecule 25 was then subjected to MDs study and named as S1 molecule in this study.

Table 5.1The chemical name, XP-Glide dock score, IFD score and chemical structures of top 10 lead molecules according their IFD score

Molecule's No.	Molecule's Name	XP-Glide dock score	IFD score	Structure
25 (S1)	I 2,2-dimethyl-7-(2-methyloctan-2-yl)-5-phenylmethoxy-3,4-dihydrochromen-4-ol	-5.751	-459.966	
40	CID= 44537255	-6.435	-459.336	
33	9-(hydroxymethyl)-6,6-dimethyl-3-(2-methyloctan-2-yl)-6a,7,8,9,10,10a-hexahydrobenzo[c]chromen-1-ol	-5.121	-458.571	
11	(6aR)-6-(aminomethyl)-6,9-dimethyl-3-pentyl-6a,7,10,10a-tetrahydro-6H-benzo[c]chromen-1-ol	-6.300	-458.504	
50	3- (1,1- dimethylheptyl) -6,6,9- trimethyl- 6a,7,10,10a-tetrahydro- 6H-benzo[c]chromen- 1- ol	-4.432	-458.359	

61	6H-Dibenzo(b,d)pyran-1,9-diol, 6a,7,8,9,10,10a-hexahydro-6,6-dimethyl-3-(4-methoxy-1-methylbutyl)	-5.424	-458.231	
52	Methyloctylpyran	-4.159	-457.796	
46	Dimethyl-heptyl tetrahydrocannabinol	-5.356	-457.774	
49	3-(1,2-dimethylheptyl)-6,6,9-trimethyl-6a,7,10,10a-tetrahydro-6H-benzo[c]chromen-1-ol	-5.219	-457.684	
41	Canbisol [INN]6H-Dibenzo(b,d)pyran-1,9-diol, 6a,7,8,9,10,10a-hexahydro-6,6-dimethyl-3-(1,1-dimethylheptyl)-,(6a-alpha,9 alpha,10a-beta)-	-4.161	-455.583	

5.3.3 Molecular Dynamics Simulation Results

Using the conditions described in Material and Methods the lead molecule S1 was analyzed with respect to the energetic equilibrium, hydrogen bonding both as donor and acceptor as well and the important residues that it may interact with. It has been shown that the S1 lead molecule reaches the energetic equilibrium in the MD simulation (**Figure 5.2**). The **Figures 5.3**and **Figures 5.4**

show the hydrogen bonds between S1 molecule and the residues that lie at a 3.5Å of distance of it. The **Figure 5.3** shows the S1 molecule as acceptor of hydrogen bonds and there is almost one hydrogen bond in every frame of the simulation. Besides, the molecule S1 can accept one or two more hydrogen bond as the simulation runs. Furthermore, in the **Figure 5.4** the molecule S1 acts as a donor of hydrogen bonds. As in the case of acceptor, there is one hydrogen bond between S1 molecule and the residues in every frame of the simulation. However, there are few situations where the S1 molecule can donate more than two hydrogen bonds. These results suggest that the S1 molecule can form up to six hydrogen bonds, but four or five hydrogen bonds are frequently formed and these bonds contribute to the permanence of the S1 molecule in the binding site.

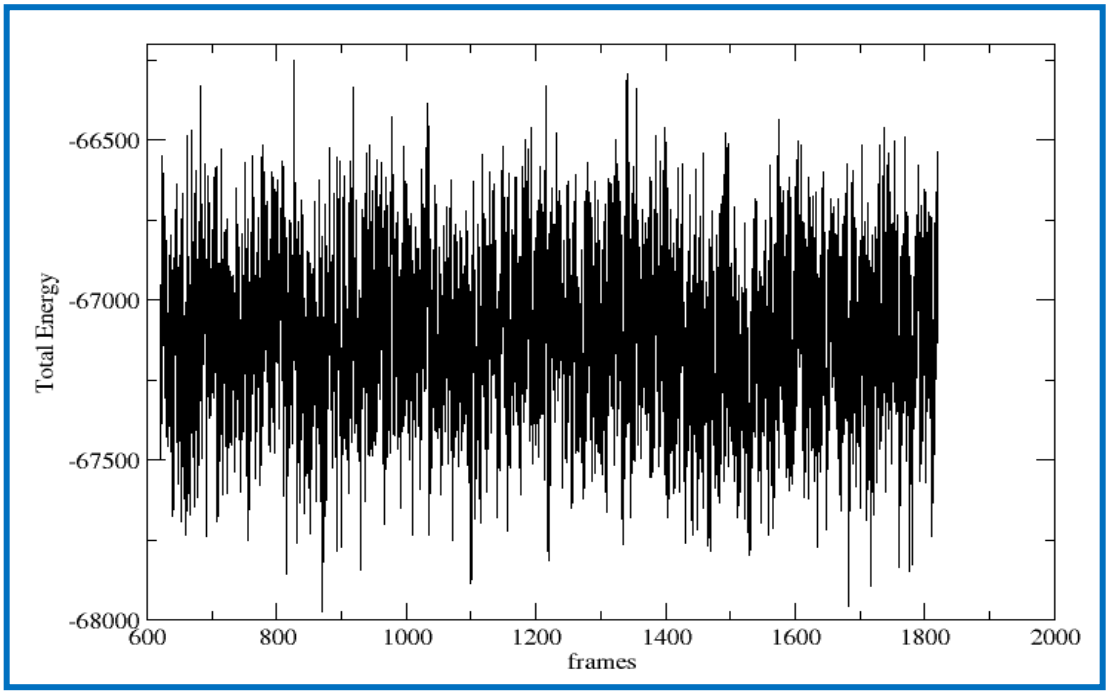


Figure 5.2.Simulation of the S1 molecule for its total energy and time frames. S1 reaches the energetic equilibrium in the molecular dynamics simulation

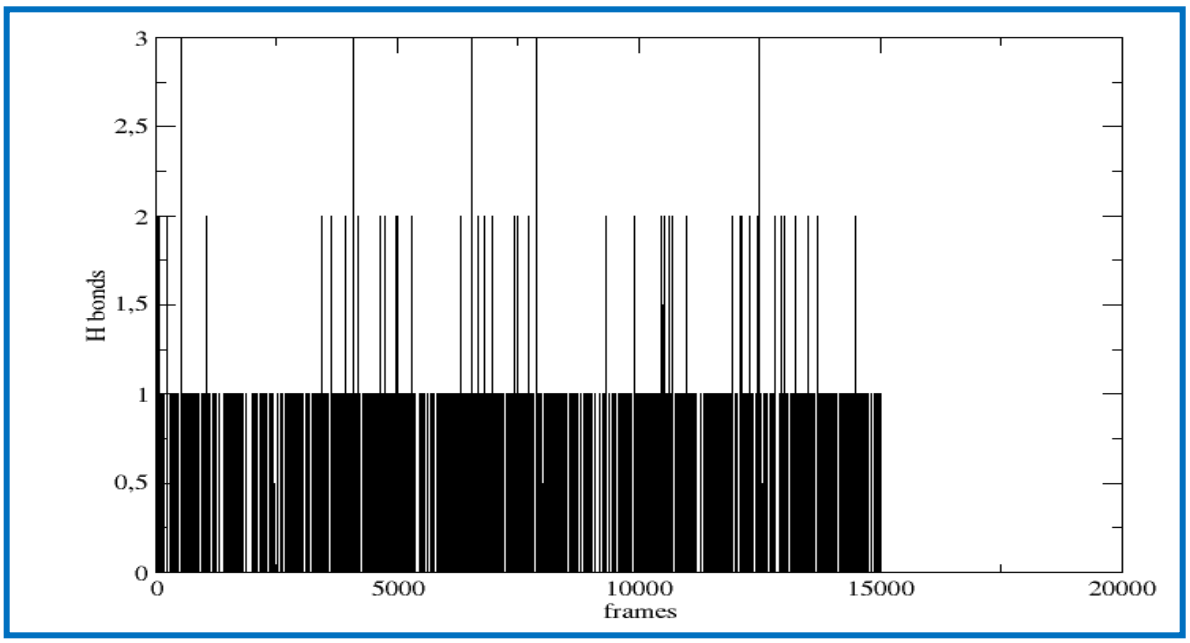


Figure 5.3.Representation of hydrogen bonds accepted by S1 molecule. Hydrogen bonds between S1 molecule and the residues that lie at a 3.5Å of distance of it were shown. There is almost one hydrogen bond in every frame of the simulation

Important interactions between S1 molecule and some residues in the interaction site of the protein were shown in **Figures 5.5** and **5.6**, and can be noted that there are important hydrophobic interactions that, with the hydrogen bonds formed, can explain the interaction of S1 molecule in the binding site. In the S1 interaction site, the Asp194 side chain moves systematically to form a hydrogen bond to the hydroxyl group of S1 molecule which is critical for ACh binding affinity (**Figure 5.5**). This hydrogen bonding is important as it may affect the ligand affinity by stabilizing the C loop conformation in its binding site. Similar association of S1 with Asp194 and Tyr185 and additionally their further interaction makes the process more proficient as the alteration of acidic to aromatic side chains or vice versa amongst the involved residues could be functionally significant. As expected, Tyr185 contributes aromatic contacts to the choline binding. Hydroxyl groups of Tyr89, and Ser142 makes a close and consistent contact with the ligand. Strong associations were found for Glu193, and Gly141, interestingly on the same aromatic ring of S1 molecule, where majority of vicinity residues are hydrophobic and polar (**Figure 5.5 and 5.6**).

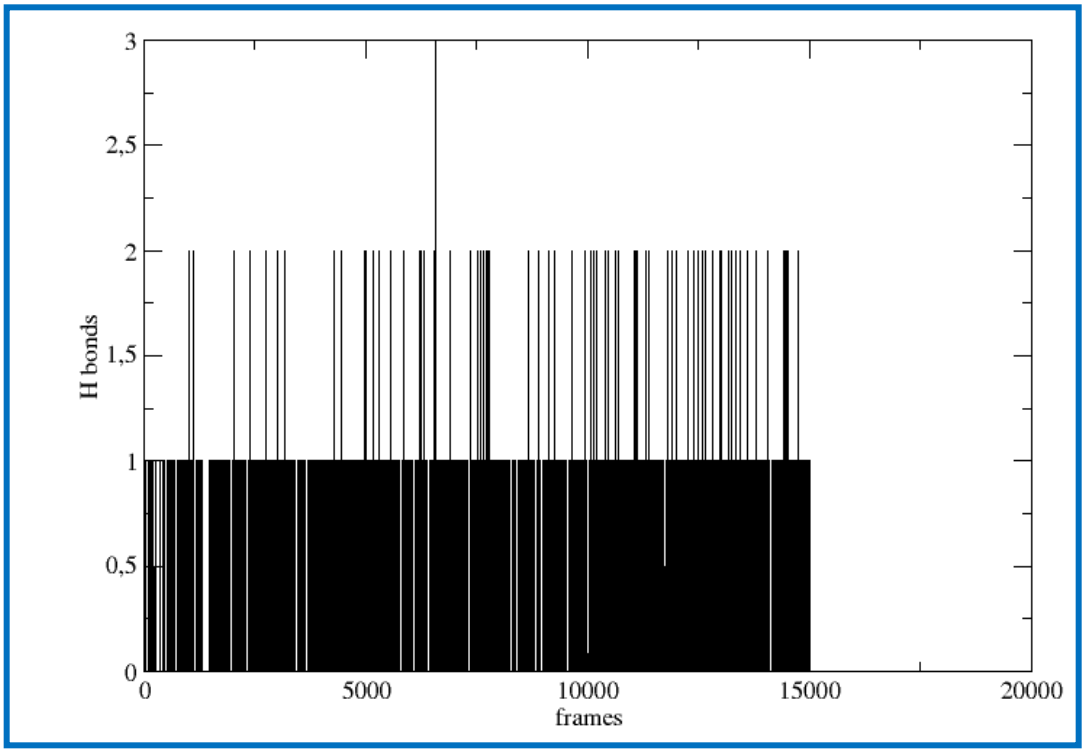


Figure 5.4.Simulation of the S1 molecule as donor of hydrogen bonds. Similar to the case of acceptor, there is one hydrogen bond between S1 molecule and the residues in every frame of the simulation

5.3.4 TOPKAT and QIKPROP Results

TOPKAT protocol predicts toxicity endpoints using the following *in silico* models: NTP Rodent Carcinogenicity, FDA Rodent Carcinogenicity, Weight of Evidence Carcinogenicity, Carcinogenic Potency TD50, Ames Mutagenicity, Developmental Toxicity Potential, Rat Oral LD50, Rat Maximum Tolerated Dose, Rat Inhalational LC50, Rat Chronic LOAEL, Skin Irritancy, Skin Sensitization, Ocular Irritancy, Aerobic Biodegradability, Fathead Minnow LC50, and Daphnia EC50. The objectives behind using the above model in our study was to optimize therapeutic ratios of lead compounds, prioritizing promising compounds for further development/investment, screening compounds generated via HTS systems, assessing pharmaceutical, setting dose-ranges for animal assays etc. The detailed results for each top listed lead molecules, by using TOPKAT module arenot shown (total 144 fields).**Table 5.2** shows the

major representative or important columns or sections out of the 144, and total 16 putative fields of interests were included.

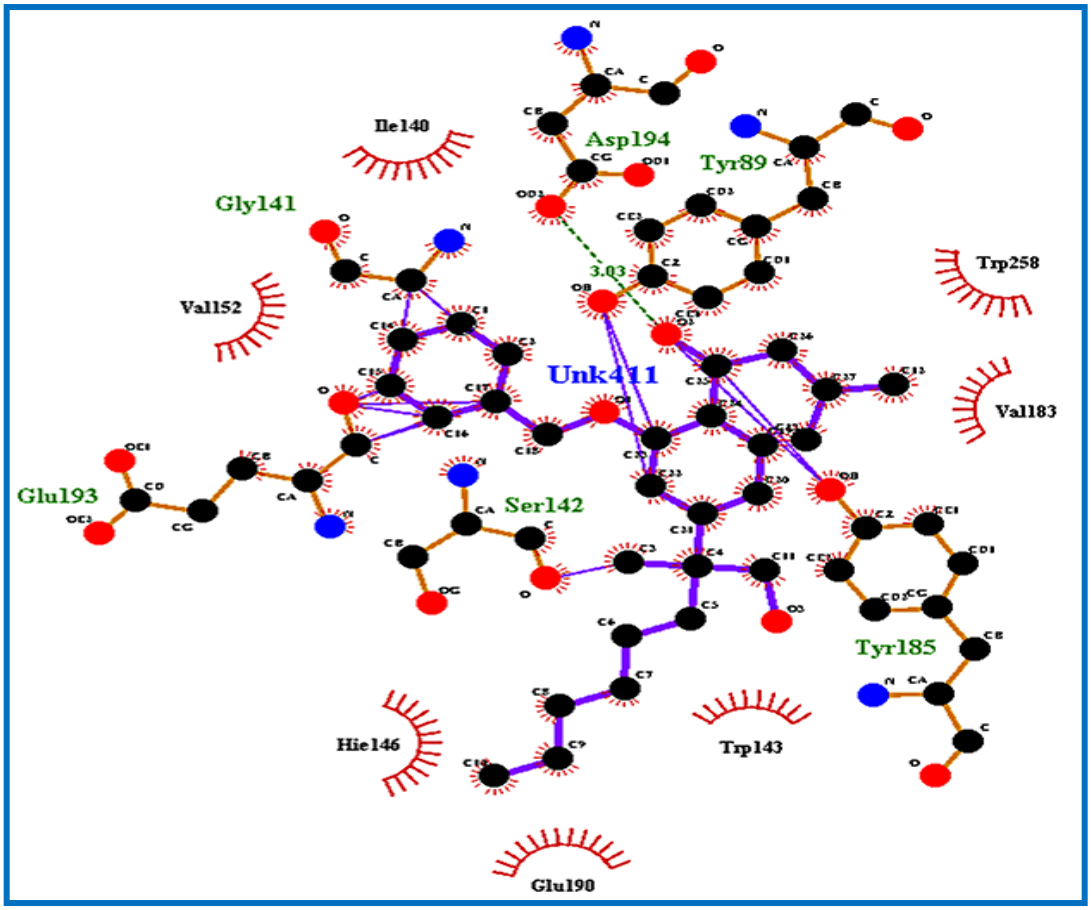


Figure 5.5.Ligplot for the interaction site of the S1 molecule with residues Ser142, Tyr89 and 185, Asp194, Gly141, and Glu193. Hydrogen bonding interaction with Asp194 is shown with dotted lines. Seven non-ligand residues such as Trp143, Glu190, Trp258 etc. shown as half sun, found involved in hrdrophobic interactions

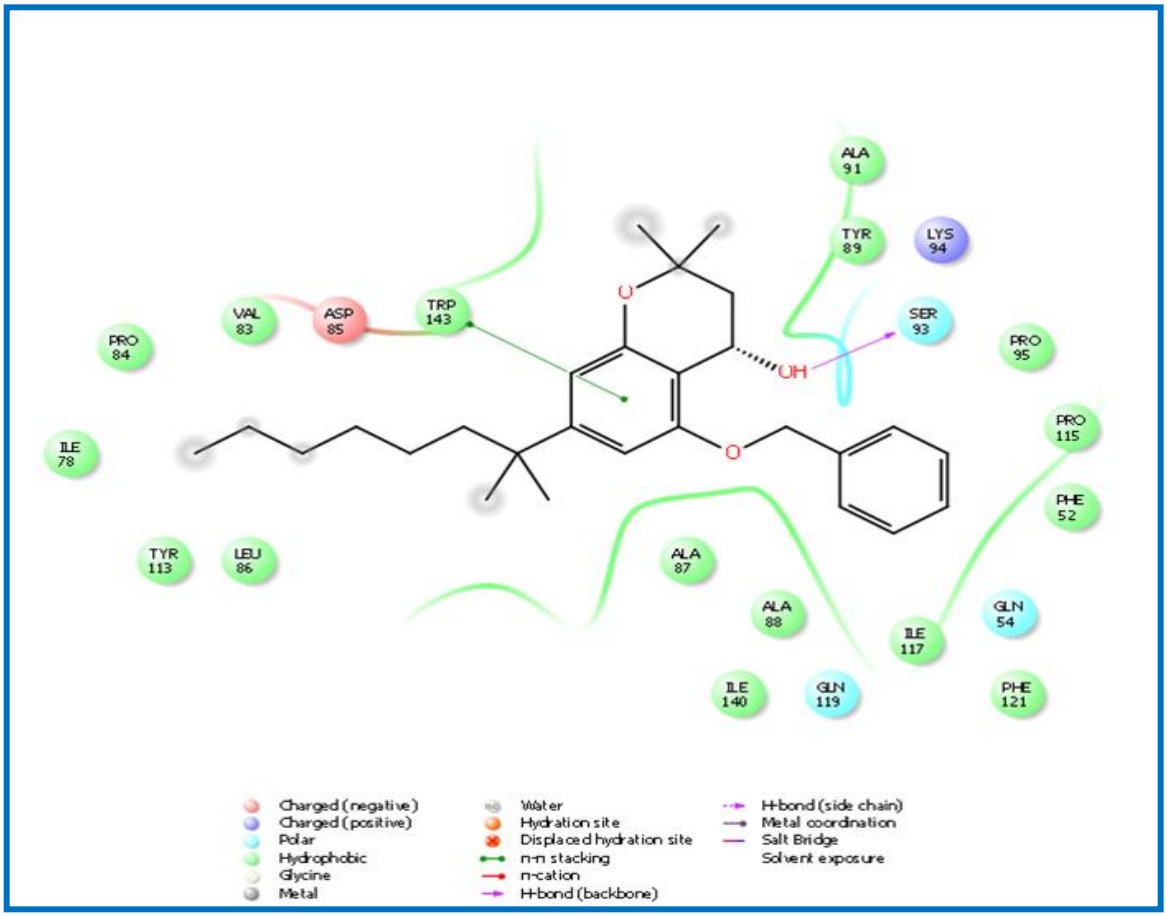


Figure 5.6.The 2-dimensional interaction site of the S1 molecule along with the active site residue Trp143 (π-π stacking), and Ser93 (hydrogen bonding: backbone)

The ADME properties were predicted and putative entries with following parameters: Predicted octanol/water partition co-efficient log p (QPlogPo/w); (range:-0.20 to 6.5), Predicted IC50 values for blockage of HERG K+ channels; (acceptable range: above-5.0), QPP Caco-Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells is a model for the gut blood barrier; (nm/sec)\25-poor[500-great], Q P log BB-Predicted brain/blood partition coefficient, QPP MDCK-Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the BBB; (nm/sec)\25-poor[500-great], Q P log KP-Predicted skin permeability; Q P log Khsa Prediction of binding to human serum albumin; (acceptable range:-1.5 to 1.5), and Percentage of human oral absorption; (\25 % is poor and [80 % is high), are shown in **Table 5.3**.

The dispositions of the candidate molecules within the organism were checked by ADMET PSA 2D (polar surface area) versus ADMET AlogP98 (the logarithm of the partition coefficient between n-octanol and water), and their suitability is represented in **Figure 5.7**. The ellipses define the regions of well-absorbed inhibitors are expected to be located. 8 lead compounds were found inside all 4 ellipses, while total 17 were found inside the territory of any of the ellipses (**Figure 5.7**), therefore satisfying the conditions for absorption by the intestines and the brain.

Table 5.2 Toxicity study of lead molecules through TOPKAT module

Ligand No.	TK Mouse Female NTP Predict.	TK Mouse Male NTP Predict.	TK Rat Female NTP Predict.	TK Rat Male NTP Predict.	TK Mouse FemaleFDA	TK MouseMale FDA	TK RAT FemaleFDA	TK RAT Male FDA	TK WOE Predict.	TK Ames Predict.	TK DTP Predict.	TKSkinIrritan.	TK Skin Sensitiz-ation	TK Ocular Irritan	TK Aerobic Biode. Predict.
25	NC	NC	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	Mild	Strong	Mild	D
40	C	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	None	Weak	None	D
33	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	Mild	None	Severe	D
11	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	Moderate	Weak	Severe	ND
50	NC	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	Weak	Severe	D
61	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	None	None	Severe	D
52	NC	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	Weak	None	D
46	C	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	Weak	None	D
49	C	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	Weak	None	D
41	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	Mild	Weak	Severe	D
3	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	None	Weak	Severe	ND
6	C	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	None	None	D

60	C	C	NC	NC	NC	NC	NC	NC	NC	NM	Toxic	Moderate	Strong	Severe	ND
53	C	C	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	Moderate	Weak	Severe	ND
43	NC	NC	NC	NC	NC	MC	NC	NC	NC	NM	Toxic	None	Strong	Mild	D

*** TK: TOPKAT, Predict.: Prediction, NC: Non-Carcinogen, MC: Multi-Carcinogen, NM: Non-Mutagen, D: Degradable, ND: Non-Degradable, Irritan.: Irritancy, Biodeg.:Biodegradability.**

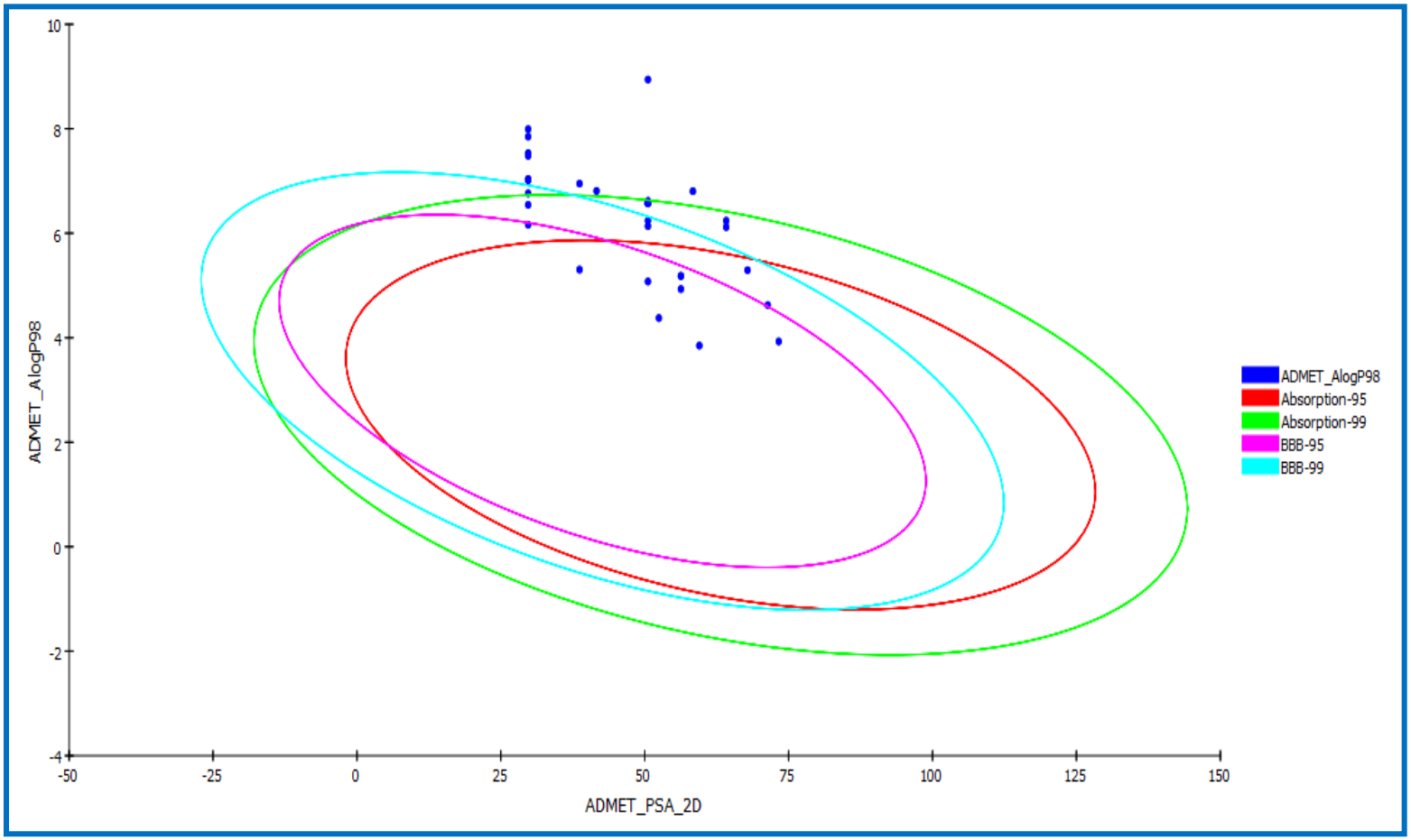


Figure 5.7.ADMET plot for ADMET PSA 2D (polar surface area) versus ADMET AlogP98 (the logarithm of the partition coefficient between n-octanol and water). 8 lead compounds fall in all the four ellipses

Table 5.3 Evaluation of ADME of molecules by QikProp. Lead molecule proposed (S1: 25) is being highlighted in yellow along with all its properties amongst 20 putative molecules

Mol.	mol_MW	dipole	Vol.	Donor HB	Accept HB	QPlogPo/w	QPlogS	QPlogHERG	QPPCaco	QPlogBB	QPPMDCK	QPlogKp	QPlogKhsa	Percent Human Oral Absorp.
3	362.508	2.817	1254.709	3	4.9	3.821	-5.344	-5.075	439.826	-1.523	203.601	-2.956	0.494	96.626
6	356.504	2.551	1275.489	1	3.2	5.834	-7.106	-5.505	4380.606	-0.259	2442.151	-1.123	1.175	100
11	329.481	2.988	1166.321	3	2.5	4.034	-4.409	-5.65	336.895	-0.367	168.852	-4.253	0.77	95.805
25	410.595	1.541	1479.434	1	3.2	7.325	-8.434	-6.407	5197.408	-0.375	2937.858	-0.299	1.617	100
26	388.589	2.864	1369.883	2	3.2	5.809	-7.034	-4.978	1591.967	-0.854	817.743	-2.06	1.25	100
33	388.589	2.127	1379.575	2	3.2	5.846	-7.15	-5.028	1473.038	-0.898	751.916	-2.125	1.276	100
35	388.589	2.864	1369.883	2	3.2	5.809	-7.034	-4.978	1591.968	-0.854	817.743	-2.06	1.25	100
37	388.589	2.127	1379.575	2	3.2	5.846	-7.15	-5.028	1473.038	-0.898	751.916	-2.125	1.276	100
40	388.589	2.772	1359.881	2	3.2	5.867	-6.709	-4.736	2279.765	-0.655	1205.555	-1.749	1.227	100
41	374.562	2.938	1328.677	2	3.2	5.535	-6.904	-4.977	1516.868	-0.814	776.128	-2.184	1.178	100
43	394.442	5.645	1209.06	0	5.25	4.429	-5.774	-5.339	2431.083	-0.183	2339.004	-1.525	0.554	100
46	370.574	1.413	1349.423	1	1.5	6.995	-8.115	-5.185	4387.296	-0.249	2446.183	-1.249	1.74	100
47	392.451	6.112	1211.81	1	6	3.968	-5.259	-5.361	1104.105	-0.701	550.604	-2.046	0.529	100
49	370.574	1.413	1347.053	1	1.5	6.977	-8.09	-5.166	4387.328	-0.248	2446.202	-1.258	1.733	100
50	370.574	2.372	1345.801	1	1.5	6.963	-8.134	-5.191	4396.041	-0.25	2451.453	-1.273	1.729	100
52	370.574	1.416	1379.151	1	1.5	7.177	-8.432	-5.328	4638.924	-0.297	2598.174	-1.186	1.781	100
53	316.483	1.32	1177.577	2	1.5	5.449	-6.133	-4.659	2520.663	-0.465	1343.816	-1.729	1.11	100
57	370.574	1.413	1347.053	1	1.5	6.977	-8.09	-5.166	4387.328	-0.248	2446.202	-1.258	1.733	100
60	314.467	2.637	1142.017	1	1.5	5.687	-6.575	-4.833	4482.991	-0.093	2503.904	-1.352	1.237	100
61	348.481	3.734	1191.849	2	4.9	4.016	-5.372	-4.771	1515.038	-0.719	775.116	-2.237	0.552	100

*** Mol.:Molecule, Vol.: Volume, Absorp.: Absorption.**

Q P log Po/w (-2.0 to 6.5) Predicted octanol/water partition co-efficient log p; (range:-0.20 to 6.5)

Predicted IC50 values for blockage of HERG K^+ channels; (acceptable range: above-5.0)

QPP Caco-Predicted apparent Caco-2 cell permeability in nm/sec. Caco-2 cells is a model for the gut blood barrier; (nm/sec)\25-poor [500-great

Q P log BB-Predicted brain/blood partition coefficient

QPP MDCK-Predicted apparent MDCK cell permeability in nm/sec. MDCK cells are considered to be a good mimic for the blood–brain barrier; (nm/sec)\25-poor [500-great

Q P log KP-Predicted skin permeability; Q P log Khsa Prediction of binding to human serum albumin; (acceptable range:-1.5 to 1.5)

Percentage of human oral absorption; (\25 % is poor and [80 % is high)

The IFD revealed the top scoring lead molecule25 was showing bonding with TRP143 and the toxicity profile obtained from TOPKAT, QIKPROP predicted that the molecule is non-carcinogenic, biodegradable, had high aqueous solubility and absolute oral bioavailability, making it a suitable drug candidate. Further the MD simulation study carried out utilizing the AMBER environment also confirmed its favorable interaction with the AChBP. It is believed that this report on the development of potential small molecule agents that binds toAChBP provides diagnostic tool to identify appropriate markers at earlier stages of the disease, which are essential for AD drug development and for clinical trials.

5.4 CONCLUSION

The molecular docking and simulation studies performed on 3000 molecules which have a structural similarity with Δ9-THC; there were 20 lead molecules which were showing good affinity with the AChBP and many of them displayed binding with TRP143, which is the key amino acid residue for triggering the concerned neurological pathway. The molecules showed no indication for mutagenicity, and tumorigenicity. Also, no indications for irritating and reproductive effects were found which was determined by the TOPKAT. The top scoring lead molecule S1 have shown score of -14.34 kcal/mol in the AutoDock simulation studies and IFD score of -459.966. S1 upon MD simulation study have demonstrated permanence in the receptor site by the formation of extensive hydrogen bonding. The results of molecular modeling show that the S1 molecule reaches at energetic equilibrium in the binding site and is stable there. Furthermore, the hydrogen bonds accepted and donated by this molecule, along with the hydrophobic interactions, can be used to suggest the binding stability of this molecule.Further *in vitro* and *in vivo* studies are required on these lead molecules as the binding mode and simulation study provided hints for the future design of new derivatives with higher potency and specificity. It is anticipated that the proposed lead molecule would serve as a potential candidate for experimental studies and will be prolific for neurological disorders such as AD and Myasthenia Gravis drug development.

OVERALL CONCLUSIONS AND FUTURE PROSPECTS

6.1 CONCLUSIONS

We have studied to analyze AD at molecular level. Theoverallstudy aimed for deciphering the biomolecules such as TFs, genes and proteins etc. and interactions among them at system level for better understanding of the bioprocesses involved in AD. The overall bottom up approach being utilized and their proposed outcomes to deal with AD as our system through various computational approaches for its system level understanding are given in **Figure 6.1**. Future prospects and the practical applications of our approach are also discussed briefly to provide new directions for the AD research.

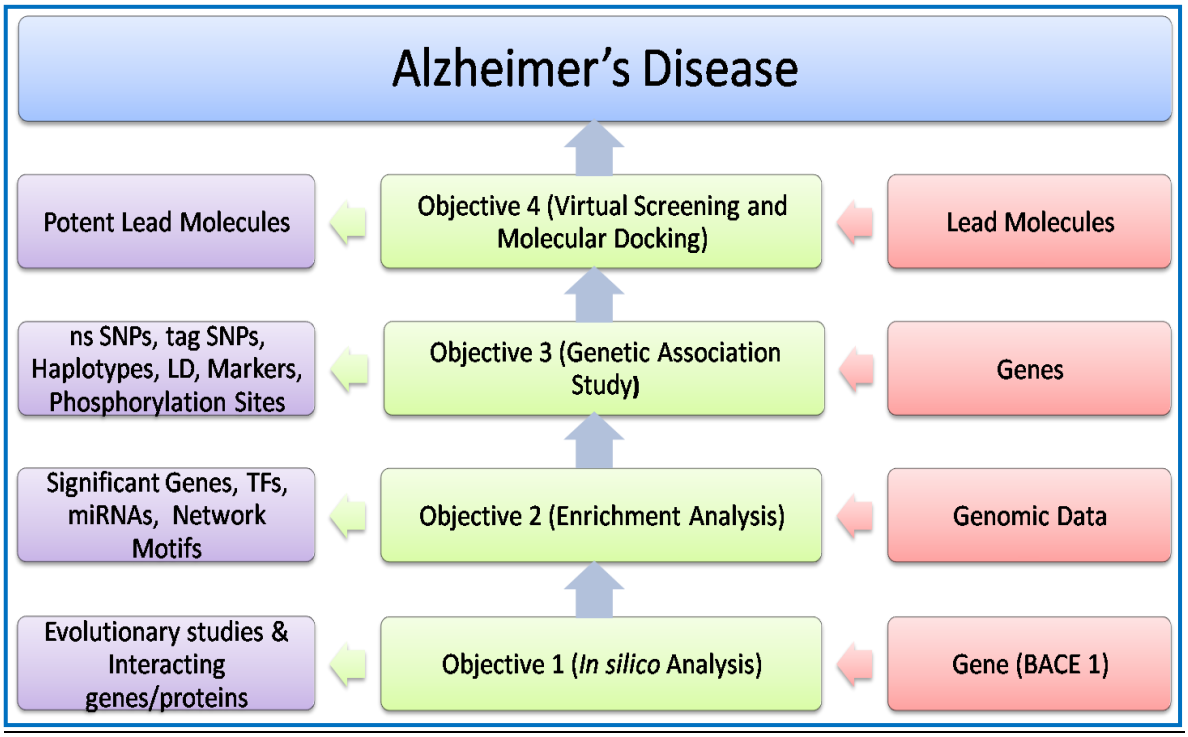


Figure 6.1.Bottom up approach being utilized with their corresponding outcomes to deal with AD as our system

Important findings of this thesis are being summarized here:-

- No such Phylogenetic tree is available based on BACE1 for species included in our tree.Phylogenetic results gave new direction to the position of lineages based on BACE1 evolutionary tree. Based upon combination of phylogenetic and functional motif analysis we suggest that guinea pig has an advantage over traditional rat and

mouse as model organism, because it is more close to mammalian class, even though all three belongs to class rodentia. Additionally, identification of Cys-Arg mutation at C-terminal will explore plausible secondary cause of variations among human BACE1 isoforms.

- From the PPI studies we have found five additional important proteins closely associated with BACE1 which can be promising candidates for AD research studies. It will also provide a platform for elucidation of regulation mechanism of BACE1 and will also helpful in designing new drug targets and inhibitors for the prevention and cure of AD.

- The Differential Gene Expression (After multiple comparisons through *t* and ANOVA *F*-statistics) results showed the extensive links between AD and AG at molecular level, identifying core biological processes and genes they share.

- It has been found through ranked list of genes and gene enrichment analysis that not only AG and AD share common patho-physiological processes but also there is involvement of other important human disease with these biological processes.Interactions of TFBS, genes, and encoded proteins at molecular level for other disease such as diabetes, dominant optic atrophy, coronary heart disease, sudden cardiac arrest, Gaucher disease, myriad carcinomas, and cirrhosis signifies putative association among AG, AD and all the above mentioned disease.

- It has been observed that conserved TFs are potent gene regulators for AD. We also explored this conservation of TFs towards their regulation mechanism for gene involved in AG, AD and other diseases. We found shared nature of TFs of AD which was associated with other disease such as haematopoiesis, Epstein-Barr virus infection, viral carcinogenesis and other carcinomas, dominant optic atrophy, atrial fibrillation, coronary heart disease and sudden cardiac death. Major classes for secondary structure elements were dominated by helix-turn-helix for the families' homeo, Arid, and Myb and its winged version for the families Forkhead, IRF, and Ets. Another major class found was Zinc-coordinating which belongs to two families' hormone-nuclear receptor, and beta-beta-alpha-zinc finger. These structural level

constraints proposed plausible targets for neuronal tangles and plaques through normal and winged helix-turn-helix and beta-beta alpha structures. Linked TFBS for these physico-chemical elements could be manipulated to deal with involved complexities.

- Genes like C1orf115, DPF3, PSMD4, USP25, KCNA5, LZTS1, CSPG5, and SLC25A6 are found to be novel candidates associated with AD, and could be useful for targeting either different brain regions (conditioned to their presence) or various biomolecular entities for designing treatment strategies for AD and other diseases.

- Some genes are found to be involved in multiple processes such as PSME3 and PSME4 are involved in PPI, TTGS, and KEGG pathways. These two genes are also found to be involved in normal AG process. While cross checking the output of GO enrichment analysis the above mentioned two genes were found in nucleoplasm with high level of significance (*p-value* 5.43E-04; GOrilla). It indicates how a particular method might not capture all the information latent in biological data and similar analysis with other tools, or methods could provide insightful annotations. Based upon this analysis we proposed that Bioinformatics/Computational analysis has to be done using multiple tools/approaches.

- Similarly genes such as TBL1X (transducin (beta)-like 1X-linked) and KIAA0528 are found to be involved in TTGS, miRNA targets, and also involved in both AD and AG related metabolic processes.

- When the genes of the present study (Chapter 3) were compared with the results of Ray and Zhang (2010), interestingly many common genes were found involved in various brain regions. At least one common gene is found in almost all regions. Between EC and HIP region, KCNAB2 and SPF3 are found, between EC and PCC region GPR22 is found, between HIP and PCC region KCNAB2 is found. There are four genes from this study named TBL1X, EFNB2, RND2 and CDH10 which were found to be involved between HIP and MTG region and TBL1X, EFNB2 genes between MTG, EC and HIP region [Ray and Zhang, 2010]. Concerned pathways such as Wnt, axon guidance, and Akt have also been found associated. These genes and

their associated pathways could be treated as hotspots while planning experimental procedures for association studies.

- Novel information for network motifs such as BiFan, MIM, and SIM and their close variants has also been discovered and this implicit information will help to improve research into AD.Biologically significant network motifs identified for AD pathways.

- There are few resources available for AD but there is no such catalog for AD prime genes which provides all these essential quantitative genetic details which includes LD, haplotypes, nsSNPs, tag SNPs, disease related functional information, along with phosphorylation states on a common platform. In this regard,*ADDGAP* is first of its kind model where the users could easily retrieve and explore the quantitative genetic parameters and the phosphorylation states for AD prime genes.

- The molecules showed no indication for mutagenicity, tumorigenicity and also, no indications for irritating and reproductive effects were found which was determined by the TOPKAT. The top scoring molecule S1 have shown score of -14.34 kcal/mol in the AutoDock simulation studies and IFD score of -459.966. The results of molecular modelling show that the S1 molecule reaches at energetic equilibrium in the binding site and is stable there. Furthermore, the H-bonds accepted and donated by this molecule, along with the hydrophobic interactions, can be used to suggest the binding stability of this molecule.

- Poor ADME characteristics are one of the reasons why drug candidates don't succeed in clinical trials. ADME properties of the top scoring docked compounds were predicted using QikProp module of Schrödinger 2012. The module predicts properties such as log $P_{o/w}$, IC50 values for blockage of HERG K^+ channels, log BB, overall CNS activity, Caco-2 and MDCK cell permeability, logKhsa for human serum albumin binding and human oral bioavailability.Prediction of these characteristics prove the suitable candidates as potent markers and could be taken further for experimental verifications and clinical trials.

6.1 FUTURE PROSPECTS

- Further *in vitro* and *in vivo* study is required on these lead molecules as the binding mode and simulation study provided hints for the future design of new derivatives with higher potency and specificity.

- The top scoring docked compound (lead molecule) showed no indication for mutagenicity, and tumorigenicity. Also, no indications for irritating and reproductive effects were found. It is anticipated that the proposed molecule would serve as a potential candidate for experimental studies and will be prolific for neurological disorders such as AD.

- Allthe projected bio-molecules (from all chapterss) would serve as potential candidates for experimental studies and will be productive for neurological disorders such as AD.

- It is believed that generated information for nucleosomes, and linkers could proved to be biologically meaningful for future structure based studies associated with genes and proteins involved in AD.

- We applied supple approach for the design of*ADDGAP* model to expand and update it on regular basis to provide state-of-the-art information to the scientific community.

- The biologically meaningful information generated through computational analyses would be of utmost use to scientific community after experimental verifications.

BIBLIOGRAPHY

"Dementia definition". *MD Guidelines*. Reed Group. Retrieved 2009-06-04.

2010 Alzheimer's Disease Facts and Figures. Rep. Vol. 6. Chicago: Alzheimer's Association, 2010. Print. Alzheimer's and Dementia.

Adzhubei I., *et al.* "Predicting functional effect of human missense mutations using PolyPhen-2," *Curr.Protoc. Hum. Genet.*, Chapter **7**, Unit7.20. doi:10.1002/0471142905.hg0720s76, 2013.

Alim M.A., *et al.*"Demonstration of a role for alpha-synuclein as a functional microtubule-associated protein," *J. Alzheimers Dis.*, vol. 6, pp. 435-42; discussion 443-9, 2004.

Alon U. Introduction to Systems Biology: Design Principles of Biological Circuits, London, UK:Chapman and Hall, 2006.

Altaf-Ul-Amin M., *et al.*"Systems biology in the context of big data and networks," *Biomed. Res. Int.*, vol. 2014, article id.428570, 2014.

Alzheimer's Association 2010.Web.01 Oct. 2010, (*http://alz.org*).

Alzheimer's Association Alzheimer's Disease Facts and Figures, 2014, (*http://www.alz.org/alzheimers_disease_facts_and_figures.asp*).

American Health Assistance Foundation (AHAF): Alzheimer's Disease, Macular Degeneration and Glaucoma. Web.15 Aug. 2010, (*http://www.ahaf.org*).

Ames B.N., *et al.*"Oxidants antioxidants and the degenerative diseases of ageing," *Proc. Natl. Acad. Sci. USA.*, vol. 90, pp. 7915-22, 1993.

Aravind L., *et al.*"The many faces of the helixturn- helix domain: transcription regulation and beyond," *FEMS Microbiol.Rev.*, vol. 2, pp. 231-62, 2005.

Artimo P., *et al.* "Expasy: SIB bioinformatics resource portal," *Nucleic Acids Res.*, vol. 40(W1), pp. W597-W603, 2012.

Ashburner M., *et al.* "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, pp. 25-9, 2000.

Ashkenazy H., *et al.* "ConSurf: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids," *Nucleic Acids Res.*, vol.38 (Web Server issue), pp. W529-W533, 2010.

Atabakhsh E., *et al.* "RanBPM expression regulates transcriptional pathways involved in development and tumorigenesis," *Am. J. Cancer Res.*, vol. 2, pp. 549-65, 2012.

Bachurin S.O. "Medicinal chemistry approaches for the treatment and prevention of Alzheimer's disease," *Med. Res. Rev.*, vol. 23, pp. 48-88, 2003.

Bakker P.D., *et al.*"Efficiency and power in genetic association studies," *Nat. Genet.*, vol. 37, pp. 1217-23, 2005.

Barrett J.C., *et al.* "Haploview: analysis and visualization of LD and haplotype maps," *Bioinformatics*, vol. 21, pp. 263-5, 2005.

Bartus R.T. "On neurodegenerative diseases, models, and treatment strategies: lessons learned and lessons forgotten a generation following the cholinergic hypothesis," *Exp. Neurol.*, vol. 163, pp. 495-529, 2000.

Battistin L. and Cagnin A. "Vascular cognitive disorder.A biological and clinical overview," *Neurochem.Res.*, vol. 35, pp. 1933-8, 2010.

Bell G., *et al.* "Computer science: beyond the data deluge," *Science*, vol. 323, pp. 1297-8, 2009.

Benito C., *et al.* "Cannabinoid CB2 receptors and fatty acid amide hydrolase are selectively overexpressed in neuritic plaque-associated glia in Alzheimer's disease brains," *J. Neurosci.*, vol. 23, pp. 11136-41, 2003.

Berbenetz N.M., *et al.*"Diversity of eukaryotic DNA replication origins revealed by genome-wide analysis of chromatin structure," *PLoS Genet.*, vol. 6, pp. e1001092, 2010.

Berchtold N.C. and Cotman C.W. "Evolution in the conceptualization of Dementia and Alzheimer's disease: Greco-Roman period to the 1960s," *Neurobiol. Aging*, vol. 19, pp. 173-89, 1998.

Berman H.M., *et al.* "The protein data bank," *N. A. R.*, vol. 28, pp. 235-42, 2000.

Bertram L. "Alzheimer's disease genetics current status and future perspectives," *Int. Rev. Neurobiol.*,vol. 84, pp. 167-84, 2009.

Bertram L., *et al.* "Systematic meta-analyses of Alzheimer disease genetic association studies: The AlzGene database," *Nat. Genet.*, vol. 39, pp. 17-23, 2007.

Bettens K., *et al.* "APP and BACE1 miRNA genetic variability has no major role in risk for Alzheimer disease," *Human Mutati.*, vol. 30, pp. 1207-13, 2009.

Blalock E., *et al.* "Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 2173-8, 2004.

Blennow K., *et al.*"Alzheimer's disease," *Lancet*, vol. 368, pp. 387-403, 2006.

Blennow K., *et al*. "Longitudinal stability of CSF biomarkers in Alzheimer's disease," *Neurosci.Lett.*, vol. 419, pp. 18-22, 2007.

Blom N., *et al.* "Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites," *J. Mol. Biol.*, vol. 294, pp. 1351-62, 1999.

Bolton E.E., *et al.* "PubChem: integrated platform of small molecules and biological activities," *Annu. Rep. in comput.Chem.*, vol. 4, pp. 217-41, 2008.

Bowman A.L., *et al.* "Small molecule inhibitors of the MDM2-p53 interaction discovered by ensemble-based receptor models," *J. Am. Chem. Soc.*, vol. 129, pp. 12809-14, 2007.

Boyle A.P., *et al.*"High-resolution genome-wide *in vivo* foot printing of diverse transcription factors in human cells," *Genome Res.*, vol. 21, pp. 456-64, 2011.

Breitner J.C. "NSAIDs Alzheimer's disease: how far to generalize from trials?," *Lancet Neurol.*, vol. 2, pp. 527, 2003.

Brookes A.J., *et al*. "HGBASE: a database of SNPs and other variations in and around human genes," *N. A. R.*, vol. 28, pp. 356-60, 2000.

Brown H. "Possible anticholinesterase-like effects of trans (−) δ8 and-δ tetrahydrocannabinol as observed in the general motor activity of mice," *Psychopharmacol.*, vol. 27, pp. 111-6, 1972.

Brown M. "Gene therapy success for Alzheimer's," *Drug Discov.Today*, vol. 8, pp. 474-5, 2003.

Bryson H.M. and Benfield P. "Donepezil Drugs," *Aging*, vol. 10, pp. 234-9 [discussion 231-40], 1997.

Bush A.I. "The metallobiology of Alzheimer's disease," *Trends. Neurosci.*, vol. 26, pp. 207-14, 2003.

Cadieux M.A., *et al.* "Needs of people with dementia in long-term care: A systematic review," *Am. J. of Alzheimer's Dis. Other Demen.*, vol. 28, pp. 723-33, 2013.

Callebaut W. "Scientific perspectivism: a philosopher of science's response to the challenge of big data biology," *Stud. Hist. Philos. Biol. Biomed.Sci.*, vol. 43, pp. 69-80, 2012.

Campbell V.A. and. Gowran A. "Alzheimer's disease; taking the edge off with cannabinoids?," *Br. J. Pharmacol.*, vol. 152, pp. 655-62, 2007.

Cargill M., *et al*. "Characterization of single nucleotide polymorphisms in coding regions ofhuman genes," *Nat. Genet.*, vol. 22, pp. 231-8, 1999.

Carter D.B. and Chou K.C. "A model for structure dependent binding of Congo Red to Alzeheimer beta-amyloid fibrils," *Neurobiol.Aging*, vol. 19, pp. 37-40, 1998.

Celie P.H *et al.* "Nicotine and carbamylcholine binding to nicotinic acetylcholine receptors as studied in AChBP crystal structures," *Neuron*, vol. 41, pp. 907-14, 2004.

Chang K.A. and Suh Y.H. "Pathophysiological roles of amyloidogenic carboxy-terminal fragments of the beta amyloid precursor protein in Alzheimer's disease," *J. Pharmacol.Sci.*, vol. 97, pp. 461-71, 2005.

Chen H.Y., *et al.*"Effect of the beta secretase-1 inhibitor on the amyloid C-terminal fragment of amyloid precursor protein processing in a hyperphosphorylated tau rat model," *Genet.Mol.Res.*, vol. 13, pp. 6213-27, 2014.

Chen W., *et al.* "iNuc-PhysChem: a sequence-based predictor for identifying nucleosomes via physicochemical properties," *PLoS. One*, vol. 7, pp. e47843, 2012.

Chen X., *et al.* "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," *Cell*, vol. 133, pp. 1106-17, 2008.

Choi S.H., *et al.*"Memory impairment and cholinergic dysfunction by centrally administered Aβ and carboxyl-terminal fragment of Alzheimer's APP in mice," *FASEB J.*, vol. 15, pp. 1816-8, 2001.

Chou K.C. "Insights from modelling the tertiary structure of BACE2," *J. Proteome Res.*, vol. 3, pp. 1069-72, 2004.

Chou K.C. "Modeling the tertiary structure of human cathepsin-E," *Biochem.Biophys. Res. Commun.*, vol. 331, pp. 56-60, 2005.

Chou K.C. "Review: structural bioinformatics and its impact to biomedical science," *Curr. Med. Chem.*, vol. 11, pp. 2105-34, 2004.

Chou K.C. and Howe W.J. "Prediction of the tertiary structure of the betasecretase zymogen," *Biochem.Biophys. Res. Commun.*, vol. 292, pp. 702-8, 2002.

Chou K.C., *et al.*"Prediction of the tertiary structure of a caspase-9/inhibitor complex," *FEBS Lett.*, vol. 470, pp. 249-56, 2000.

Clarke B., *et al.* "BACE-1 hydroxyethylamine inhibitors using novel edge-to-face interaction with Arg-296," *Bioorg. Med. Chem. Lett.*, vol. 20, pp. 4639-44, 2010.

Cleveland W.S. and Devlin S.J. "Locally weighted regression: an approach to regression analysis by local fitting," *J. Am. Stat. Assoc.*, vol. 83, pp. 596-610, 1988.

Cocabelos R., *et al.* "Molecular strategies for the first generation of antidementia drugs (I). Tacrine and related compounds," *Drugs Today*, vol. 30, pp. 259-337, 1994.

Collin Y., *et al*. "Understanding the molecular basis of Alzheimer's disease using a *Caenorhabditis elegans* model system," *Brain Stru. Funct.*,vol. 214, pp. 263-83, 2009.

Corder E.H., *et al*. "Gene dose of Apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families," *Science*, vol. 261, pp. 921-3, 1993.

Corringer P.J., *et al.* "Nicotinic receptors at the amino acid level," *Annu. Rev. Pharmacol. Toxicol.*, vol. 40, pp. 431-58, 2000.

Craft S., *et al.* "Intranasal insulin therapy for Alzheimer disease and amnestic mild cognitive impairment," *Arch. Neurol.*, vol. 69, pp. 29-38, 2012.

Croft D., *et al.* "Reactome: a database of reactions, pathways and biological processes," *Nucleic Acids Res.*, vol. 39, pp. D691-D697, 2010.

Cruts M., *et al.* "Locus-specific mutation databases for neurodegenerative brain diseases," *Human Mutat.*, vol. 33, pp. 1340-4, 2012.

Cruts M. and Broeckhoven C.V. "Molecular genetics of Alzheimer's disease,"*Annals. Medicine*,vol.30, pp. 560-5, 1998.

Csiszar A., *et al.* "Synergistic effects of hypertension and aging on cognitive function and hippocampal expression of genes involved inβ-amyloid generation and Alzheimer's disease," *Am. J. Physiol. Heart. Circ. Physiol.*, vol. 305, pp. H1120-H1130, 2013.

Cui Q., *et al.* "Molecular dynamics-solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex," *J. Mol. Biol.*, vol. 379, pp. 787-802, 2008.

Cummings J.L. and Benson D.F. "Subcortical dementia.Review of an emerging concept," *Arch. Neurol.*,vol. 41, pp. 874-9, 1984.

Darden T., "Particle mesh Ewald - an n.Log(n) method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089-92, 1993.

Derouesne C. "Alzheimer and Alzheimer's disease: the present enlighted by the past. An historical approach," *Psychol. Neuropsychiatr.Vieil.*, vol. 6, pp. 115-28, 2008.

Devane W.A., *et al.* "Isolation and structure of a brain constituent that binds to the cannabinoid receptor," *Science*, vol. 258, pp. 1946-9, 1992.

Deza E. and Deza M.M. Encyclopedia of Distances, Second Ed., Springer Heidelberg New York Dordrecht London, pp. 94, 2009, DOI: 10.1007/978-3-642-30958-8.

Di Marzo V., *et al.* "The biosynthesis, fate and pharmacological properties of endocannabinoids, in: Cannabinoids," *Handb. Exp. Pharmacol.*, vol.168, pp. 147-85, 2005.

Doerfler P., *et al.* "Presenilin-dependent gamma-secretase activity modulates thymocyte development," *Proc. Natl. Acad. Sci. USA.*, vol.98, pp. 9312-7, 2001.

Dominguez D.I. and De S.B. "Novel therapeutic strategies provide the real test for the amyloid hypothesis of Alzheimer's disease," *Trends. Pharmacol.Sci.*, vol. 23, pp. 324-30, 2002.

Duan Z., *et al.* "A novel notch protein, N2N, targeted by neutrophil elastase and implicated in hereditary neutropenia," *Mol. Cell. Biol.*, vol. 24, pp. 58-70, 2004.

Duncan D.T., *et al.*"WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit," *BMC.Bioinform.*, vol. 11, pp. 10, 2010.

Dziuda D.M., "Basic analysis of gene expression microarray data," Data Mining for Genomics and Proteomics: analysis of Gene and Protein Expression Data, John Wiley and Sons (Publ.), Hoboken, pp. 17-93, 2010.

Eden E., *et al.* "Discovering motifs in ranked lists of DNA sequences," *PLoS. Comput.Biol.*, vol. 3, pp. e39, 2007.

Eden E., *et al.* "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists," *BMC. Bioinform.*, vol. 10, pp. 48, 2009.

Edgar R.C. "MUSCLE: Multiple sequence alignment with high accuracy and high throughput," *Nucleic Acids Res.*, vol. 32, pp. 1792-7, 2004.

Emilien G.R., *et al*. Alzheimer Disease: Neuropsychology and Pharmacology, Berlin: Birkhauser Publ. Ltd., Basel, 2004.

Eriksson L., *et al.* "Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs," *Environ. Health Perspect.*, vol. 111, pp. 1361, 2003.

Ertekin-Taner N. "Genetics of Alzheimer disease in the pre- and post-GWAS era," *Alzheimers Res. Ther.*, vol. 2, pp. 3, 2010.

Eubanks L.M., *et al.* "A molecular link between the active component of marijuana and Alzheimer's disease pathology," *Mol. Pharmaceutics.*, vol. 3, pp. 773-7, 2006.

Fagan A.M. and Holtzman D.M. "Cerebrospinal fluid biomarkers of Alzheimer's disease," *Biomark in Medicine*, vol. 4, pp. 51-63, 2010.

Fede G.D., *et al.* "A recessive mutation in the APP gene with dominant negative effect on amyloidogenesis," *Science*, vol. 323, pp. 1473-7, 2009.

Feig M., *et al.* "Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures," *J. Comput. Chem.*, vol. 25, pp. 265-84, 2004.

Forstl H. and Kurz A. "Clinical features of Alzheimer's disease," *Eur. Arch. of Psychiatry and Clinical Neurosci.*, vol. 249, pp. 288-90, 1999.

Frank M.,*et al.* "Intracellular amyloid-β in Alzheimer's disease," *Nat. Reviews Neurosci.*, vol. 8, 499-509, 2007.

Galvin J.E.,*et al.* "Relationship of dementia screening tests with biomarkers of Alzheimer's disease," *Brain*, vol. 133, pp. 3290-300, 2010.

Gaoni Y. and Mechoulam R. "Isolation, Structure, and Partial Synthesis of an Active Constituent of Hashish," *J. Am. Chem. Soc.*, vol. 86, pp. 1646-7, 1964.

Gatz M., *et al.* "Role of genes and environments for explaining Alzheimer disease," *Arch.Gen.Psychiatry*, vol. 63, pp. 168-74, 2006.

Geller T. "Cannabinoids: A secret history," *Chemical Heritage News magazine*. vol.v25 (2), 2007.

Gotz J., *et al.* "Common features between diabetes mellitus and Alzheimer's disease," *Cell. Mol. Life Sci.*, vol. 66, pp. 1321-5, 2009.

Gotz J., *et al.* "Formation of neurofibrillary tangles in P301L tau transgenic mice induced by Aβ42 fibrils," *Science*, vol. 293, pp. 1491-5, 2001.

Grotenhermen F. and Muller-Vahl K. "he therapeutic potential of cannabis and cannabinoids," *Dtsch Arztebl Int.*, vol 109, pp. 495-501, 2012.

Gu R.X., *et al.* "Possible drug candidates for Alzheimer's disease deduced from studying their binding interactions with alpha 7 nicotinic acetylcholine receptor," *Med. Chem.*, vol. 5, pp. 250-62, 2009.

Guenette S., *et al.* "Essential roles for the FE65 amyloid precursor protein-interacting proteins in brain development," *EMBO J.*, vol. 25, pp.420-31, 2006.

Hake A.M. "Use of cholinesterase inhibitors for treatment of Alzheimer's disease," *Cleveland Clin. J. Med.*,vol.68, pp. 608-16, 2001.

Haniu M., *et al.* "Characterization of Alzheimer's beta -secretase protein BACE. A pepsin family member with unusual properties," *J. Biol.Chem.*, vol.275, pp.21099-106, 2000.

Hanus L., et al. "2-arachidonyl glyceryl ether, an endogenous agonist of the cannabinoid CB1 receptor," *Proc. Natl. Acad. Sci.*, vol. 98, pp. 3662-5, 2001.

Hardy J. and Selkoe D.J. "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics," *Science*, vol. 297, pp. 353-6, 2002.

Hattori C., *et al*. "BACE1 interacts with nicastrin," *Biochem. Biophys. Res. Commun*., vol. 293, pp.1228-32, 2002.

Hawkins G.D., *et al*. "Pairwise Solute Descreening of Solute Charges from a Dielectric Medium," *Chem. Phys. Lett.*, vol. 246, pp. 122-9, 1995.

Hawkins G. D., *et al.* "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium," *J. Phys. Chem.*, vol. 100, pp. 19824-39, 1996.

He W., *et al.* "Reticulon family members modulate BACE1 activity and amyloid-beta peptide generation," *Nat. Med.*, vol. 10, pp. 959-965, 2004.

He X., *et al.* "Biochemical and Structural Characterization of the Interaction of Memapsin 2 (β-Secretase) Cytosolic Domain with the VHS Domain of GGA Proteins," *Biochemistry*, vol.42, pp.12174-80, 2003.

He X., *et al*. "GGA Proteins Mediate the Recycling Pathway ofMemapsin 2 (BACE)," *J. Biol. Chem.*, vol. 280, pp.11696-703, 2005.

Hebert L.E., *et al*. "Alzheimer disease in the United States (2010-2050) estimated using the 2010 Census," *Neurology*, vol. 80, pp. 1778-83, 2013.

Hebert S.S., *et al*. "Presenilin-1 interacts directly with the beta-site amyloid protein precursor cleaving enzyme (BACE1)," *Neurobiol. Dis*.,vol. 13, pp. 238-45, 2003.

Helmuth L. "New therapies.New Alzheimer's treatments that may ease the mind," *Science*, pp. 297, vol. 1260-2, 2002.

Herkenham M., *et al.* "Characterization and localization of cannabinoid receptors in rat brain: a quantitative in vitro auto-radiographic study," *J. Neurosci.*, vol. 11, pp. 563-83, 1991.

Herskowitz J.H., *et al.* "GGA1-mediated endocytic traffic of LR11/SorLA alters APP intracellular distribution and amyloid-β production," *Mol. Biol. Cell.*, vol. 23, pp. 2645-57, 2012.

Hinck L. "The Versatile Roles of Axon Guidance," *Cues Tissue Morphog. Dev. Cell*, vol. 7, pp. 783-793, 2004.

Ho S.S.J., *et al.*"oPOSSUM: integrated tools for analysis of regulatory motif over-representation," *Nucleic Acids Res.*, vol. 35, pp. W245-W252, 2007.

Hommet C., *et al.* "Review of cerebral microangiopathy and Alzheimer's disease: relation between white matter hyperintensities and microbleeds," *Dementia Geriatric Cognitive Dis.*, vol. 32, pp. 367-78, 2011.

Hong L., *et al.* "Structural features of human memapsin 2 (beta-secretase) and their biological and pathological implications," *Acta.Biochim. Biophys.Sin. (Shanghai*), vol.36, pp.787-92, 2005.

Hong L., *et al*."Structural Features of Human Memapsin 2 (β-Secretase) and Their Biological and Pathological Implications," *Acta.Biochim. Biophys. Sin*. (*Shanghai*), vol. 36, pp. 787-92, 2004.

Hooghe B., et al. "A flexible integrative approach based on random forest improves prediction of transcription factor binding sites," *Nucleic Acids Res.*, vol. 40, pp. e106, 2012.

Hoover W. "Canonical dynamics - equilibrium phase-space distributions," *Phys. Rev. A.*, vol. 31, pp. 1695-7, 1985.

Hotelling H. "Analysis of a complex of statistical variables into principle components," *J. Educ. Psychol.*, vol. 24, pp. 417-41, 1933.

Howlett A., *et al.* "International Union of Pharmacology. XXVII. Classification of cannabinoid receptors," *Pharmacol. Rev.*, vol. 54, pp. 161-202, 2002.

Hrncic R., *et al.* "Antibody-mediated resolution of light chain-associated amyloid deposits," *Am. J. Pathol.*, vol. 157, pp. 1239-46, 2000.

Huang X., *et al.* "Cu (II) potentiation of alzheimer abeta neurotoxicity. Correlation with cell-free hydrogen peroxide production and metal reduction," *J. Biol. Chem.*, vol. 274, pp. 37111-6, 1999.

Huang Y., *et al.* "An image based system biology approach for Alzheimer's disease pathway analysis," *IEEE NIH Life Sci. Syst. Appl. Workshop*, pp. 128-32, 2009.

Huber S.J., *et al.*"Cortical vs Subcortical dementia neuropsychological differences," *Arch. Neurol*., vol. 43, pp. 392-4, 1986.

Hunt C.E. and Turner A.J. "Cell biology, regulation and inhibition of beta-secretase (BACE-1)," *FEBS J.*, vol. 276, pp. 1845-59, 2009.

Imbimbo B.P., *et al.*"Efficacy and safety of eptastigmine for the treatment of patients with Alzheimer's disease," *Neurology*, vol. 52, pp. 700-8, 1999.

Iuvone T., *et al.* "Neuroprotective effect of cannabidiol, a non-psychoactive component from Cannabis sativa, on beta-amyloid-induced toxicity in PC12 cells," *J. Neurochem.*, vol. 89, pp. 134-41, 2004.

Iwamura H., *et al.*"In vitro and in vivo pharmacological characterization of JTE-907, a novel selective ligand for cannabinoid CB2 receptor,"*J. Pharmacol. Exp. Ther.*, vol. 296, pp. 420-5, 2001.

Johnson G.C., *et al*. "Haplotype tagging for the identification of common disease genes,"*Nat. Genet.*, vol. 29, pp. 233-7, 2001.

Johnson S.C. "Hierarchical clustering schemes," Psychometrika., vol. 2, pp. 241-254, 1967.

Johnston J.A., *et al.* "Expression and activity of beta-site amyloid precursor protein cleaving enzyme in Alzheimer's disease," *Biochem. Soc. Trans.*, vol.33, pp.1096-100, 2006.

Jones R.W., *et al.*"The Atorvastatin/Donepezil in Alzheimer's Disease Study (LEADe): design and baseline characteristics," *Alzheimers Dement.*, vol. 4, pp. 145-53, 2008.

Jorgensen W.L. "Rusting of the lock and key model for protein-ligand binding," *Science*, vol. 254, pp. 954-5, 1991.

Jorgensen W.L. and Duffy E.M. "Prediction of drug solubility from structure," *Adv. Drug Deliv.Rev.*, vol. 54, pp. 355-66, 2002.

Kadota M., *et al.* "Identification of novel gene amplifications in breast cancer and coexistence of gene amplification with an activating mutation of PIK3CA," *Cancer Res.*, vol. 69, pp. 7357-65, 2009.

Kalir S. and Alon U. "Using a quantitative blueprint to reprogram the dynamics of the flagella gene network," *Cell*, vol. 117, pp. 713-20, 2004.

Kametaka S., *et al.* "Identification of phospholipid scramblase 1 as a novel interacting molecule with beta-secretase (beta -site amyloid precursor protein (APP) cleaving enzyme (BACE))," *J. Biol. Chem.*, vol. 278, pp. 15239-45, 2003.

Kandasamy K., *et al.* "NetPath: a public resource of curated signal transduction pathways," *Genome Biol.*, vol. 11, pp. R3, 2010.

Kanehisa M. and Goto S. "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, pp. 27-30, 2000.

Kann M.G. "Protein interactions and disease: computational approaches to uncover the etiology of diseases," *Briefings Bioinform.*, vol. 8, pp. 333-46, 2007.

Karch C.M., *et al.* "Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains," *Plos One*,vol. **7**, pp. e509, 2012.

Karlin S., *et al.* "Amino acid runs in eukaryotic proteomes and disease associations," *Proc. Natl. Acad. Sci.*, vol. 99, pp. 333-8, 2002.

Kashtan N., *et al.* "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs," *Bioinformatics*, vol. 20, pp. 1746-58, 2004.

Katzung B.G. "Basic and clinical pharmacology:Introduction to autonomic pharmacology", Ed: 8, *The McGraw Hill Companies*, pp. 75-91, ISBN 978-0-07-160405-5, 2001.

Katoh K., *et al.* "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform," *Nucleic Acids Res.*, vol. 30, pp. 3059-66, 2002.

Kinoshita J. and Clark T. "Alzforum," *Methods Mol. Biol.*, vol. 401, pp. 365-81, 2007.

Kitchen D.B., *et al.* "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nat. Rev. Drug discov.*,vol. 3, pp. 935-49, 2004.

Kohonen T."Analysis of a simple self-organizing process," *Biol. Cybern.*, vol. 44, pp. 135-140, 1982b.

Kohonen T."Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, pp. 59-69, 1982a.

Kong W., *et al.* "Exploring matrix factorization techniques for Significant genes identification of Alzheimer's disease microarray gene expression data," *BMC.Bioinform.*, vol. 12, pp. S7, 2011.

Kordower J.H., *et al.*"Cellular delivery of trophic factors for the treatment of Huntington's disease: is neuroprotection possible," *Exp. Neurol.*, vol. 159, pp. 4-20, 1999.

Kowalska A., "Counseling and genetic testing for families with Alzheimer's disease," *Neurol. and Neurosur.Pol.*, vol. 38, pp. 495-501, 2004.

Krauthammer M., *et al*. "Molecular triangulation: bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 15148-53, 2004.

Kreegipuu A., *et al.*"PhosphoBase, a database of phosphorylation sites: release 2.0," *N. A. R.*, vol. 27, pp. 237-9, 1999.

Kriegebaum B.C., *et al.*"The expression of the transcription factor FEV in adult human brain and its association with affective disorders," *J. Neural Trans.*, vol. 117, pp. 831-836, 2010.

Kubis A.M. and Janusz M. "Alzheimer's disease: new prospects in therapy and applied experimental models," *Postepy Hig. Med. Dosw.(Online)*, vol. 62, pp. 372-92, 2008.

Kumar P., *et al.* "*In Silico* Theoretical Molecular Modeling for Alzheimer'sDisease: The Nicotine-Curcumin Paradigm in Neuroprotection and Neurotherapy,"*Int. J. Mol. Sci.*, vol.12, pp.694-724, 2011.

Kumar P., *et al*. "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nature Protocols*, vol. 4, pp. 1073-81, 2009.

Kuo W., *et al*. "Analysis of matched mRNA measurements from two different microarray technologies," *Bioinformatics*, vol. 18, pp. 405-412, 2002.

Kwak B., *et al.* "Statins as a newly recognized type of immunemodulator," *Nat. Med.*, vol. 6, pp. 1399-402, 2000.

Le Novere N. and Changeux J.P. "LGICdb: the ligand-gated ion channel database," *Nucleic Acids Res.*, vol. 29, pp. 294-5, 2001.

Lee H.S., *et al.* "Rational drug repositioning guided by an integrated pharmacological network of protein, disease and drug," *BMC. Syst. Biol.*, vol. 6, article id. 80, 2012.

Lee J. H., *et al.*"The neuronal sortilin related receptor gene SORL1 and late-onset Alzheimer's disease," *Curr. Neurol. Neurosci. Rep.*, vol. 8, pp. 384-91, 2008.

Lee K.W., *et al.*"Progressive neuronal loss and behavioral impairments of transgenic C57BL/6 inbred mice expressing the carboxy terminus of amyloid precursor protein," *Neurobiol.Dis.*, vol. 22, pp. 10-24, 2006.

Lengauer T.and Rarey M. "Computational methods for biomolecular docking," *Curr.Opin.Struct.Biol.*, vol. 6, pp. 402-6, 1996.

Liang W.S., *et al.*"Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set," *Physiol. Genomic*, vol. 33, pp. 240-56, 2008.

Licata L., *et al.* "MINT, the molecular interaction database: 2012 update," *Nucleic Acids Res.*, vol. 40 (Database issue), pp. D857-61, 2012.

Lin J.S., *et al.* "Screening for cognitive impairment in older adults: A systematic review for the U.S. preventive services task force," *Annals. internal. Medicine*, vol. 159, pp. 601-12, 2013.

Llado A., *et al*. "A novel PSEN1 mutation (K239N) associated with Alzheimer's disease with wide range age of onset and slow progression," *Eur. J. Neurol.*,vol. 17, pp. 994-6, 2010.

Loy C.T., *et al*. "Genetics of Dementia," *Lancet.*, vol. 383, pp. 828-40, 2014.

Lu T., *et al.*, "Gene regulation and DNA damage in the ageing human brain," *Nature*, vol. 429, pp. 883-91, 2004.

Lu X., *et al.* "Hubs in biological interaction networks exhibit low changes in expression in experimental asthma," *Mol. Syst. Biol.*, vol. 3, pp. 105, 2007.

Lupica C.R., *at al.* "Marijuana and cannabinoid regulation of brain reward circuits,"J. Pharmacol., vol. 143, pp. 227-234, 2004.

Lynn B.C., *et al.*"Quantitative Changes in the Mitochondrial Proteome fromSubjects with Mild Cognitive Impairment, Early Stage and Late Stage Alzheimer'sdisease," *J. Alzheimers Dis.*, vol.19, pp.325-39, 2010.

Macian F., *et al.* "Partners in transcription: NFAT and AP-1," *Oncogene*, vol. 20, pp. 2476-89, 2001.

Maes O.C., *et al.*"Transcriptional profiling of Alzheimer blood mononuclear cells by microarray," *Neurobiol.Aging*, vol. 28, pp. 1795-1809, 2007.

Mandelkow E. "Alzheimer's disease, the tangled tale of tau," *Nature*, vol. 402, pp. 588-9, 1999.

Manna M. and Mukhopadhyay C. "Binding, Conformational Transition and Dimerization of Amyloid-β Peptide on GM1-Containing Ternary Membrane: Insights from Molecular Dynamics Simulation," *Plos One*, vol. 8, pp. e71308, 2013.

Martiskainen H., *et al.* "Targeting ApoE4/ApoE receptor LRP1 in Alzheimer's disease," *Expert Opin.Ther.Targets.*, vol. 17, pp. 781-94, 2013.

Maruszak A., *et al.* "PIN1 gene variants in Alzheimer's disease," *BMC. Medical Genetics*, vol.10, pp. 1-11, 2009.

Matsuda S., *et al.* "BRI3 inhibits amyloid precursor protein processing in a mechanistically distinct manner from its homologue dementia gene BRI2," *J. Biol. Chem.*, vol. 284, pp. 15815-25, 2009.

Mattson M. P. "Pathways towards and away from Alzheimer's disease," *Nature*, vol. 430, pp. 631-9, 2004.

Mattsson N., *et al.* "CSF protein biomarkers predicting longitudinal reduction of CSF β-amyloid 42 in cognitively healthy elders,"*Translational Psychiatry*, vol. 3, pp. e293, 2013.

Mayo Clinic Medical Information and Tools for Healthy Living.2010 Mayo Foundation for Medical Education and Research.Web.21 Nov. 2010, (*http://www.mayoclinic.com/*).

McGeer E.G. and McGeer P.L. "Inflammatory processes in Alzheimer's disease," *Prog.Neuropsychopharmacol. Biol. Psychiatry*, vol. 27, pp. 741-9, 2003.

Mcgeer P.L. and Mcgeer E.G. "NSAIDs and Alzheimer disease: epidemiological, animal model and clinical studies," *Neurobiol. Aging*, vol. 28, pp. 639-47, 2007.

Mechoulam R., *et al.* "Identification of an endogenous 2-monoglyceride, present in canine gut, that binds to cannabinoid receptors,"*Biochem. Pharmacol.*, vol. 50, pp. 83-90, 1995.

Meng E.C., *et al.* "Automated docking with grid-based energy evaluation," *J. Comput. Chem.*, vol. 13, pp. 505-24, 1992.

Meng Y., *et al.* "Association between SORL1 and Alzheimer's disease in a genome-wide study," *NeuroReport.*, vol. 18, pp. 1761-4, 2007.

Miar A., *et al.* "Amyloid Precursor Protein Gene (APP) Variation in Late-Onset Alzheimer's Disease," *J. Mole.Neurosci.*, vol. 45, pp. 5-9, 2011.

Miller J.A., *et al.*"A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging," *J. Neurosc.*, vol. 28, pp. 1410-20, 2008.

Milo R., et al. "Network motifs: simple building blocks of complex networks," *Science*, vol. 298, pp.824-7, 2002.

Mintseris J. and Weng Z. "Structure,function and evolution of transient andobligate protein-protein interactions," *Proc. Natl. Acad. Sci. U. S. A.*,vol.102, pp.10930-5, 2005.

Morris G.M., *et al.* "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function," *J. Comput. Chem.*, vol.19, pp. 1639-62, 1998.

Mudher A. and Lovestone S. "Alzheimer's disease- do tauists and baptists finally shake hands," *Trends Neurosci.*, vol. 25, pp. 22-6, 2002.

Murayama K. S., *et al.*"Extracellular release of BACE1 holoproteins from human neuronal cells," Biochem.*Biophys. Res. Commun.*, vol. 338, pp. 800-7, 2005.

Naj A.C., *et al.*"Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease," *Nat. Genet.*, vol. 43, pp. 436-41, 2011.

Nakamura S. "Senile dementia and pre-senile dementia," *Tohoku.J. Exp. Med.*, vol. 161, pp. 49-60, 1990.

Nalbantoglu B., *et al.* Wnt signaling network in *Homo sapiens*, In: Paula Bubulya (Ed.), Cell Metabolism - Cell Homeostasis Stress Response, 2012.

Natunen T., *et al*. "Elucidation of the BACE1 Regulating Factor GGA3 in Alzheimer's Disease," *J. Alzheimer's Dis.*, vol. 37, pp. 217-32, 2013.

Naylor S. and Chen J.Y. "Unraveling human complexity and disease with systems biology and personalized medicine," *Per. Med.*, vol. 7, pp. 275-89, 2010.

Nei M. and Kumar S. "Molecular Evolution and Phylogenetics," Oxford University Press, New York (333pp), 2000.

Newman J.C. and Weiner A.M. "L2L: a simple tool for discovering the hidden significance in microarray expression data," *Genome Biol.*, vol. 6, pp. r81, 2005.

Ng P.C. and Henikoff S. "Predicting the effects of amino acid substitutions onprotein function," *Annu.Rev. Genomics Hum. Genet.*, vol. **7**, pp. 61-80, 2006.

Norgan A.P., *et al.* "Multilevel parallelization of AutoDock 4.2," *J. cheminfo.*, vol. 3, pp. 1-9, 2011.

Nunomura A., *et al.*"Oxidative damage is the earliest event in Alzheimer's disease," *J. Neuropathol. Exp. Neurol.*, vol. 60, pp. 759-67, 2001.

Oddo S., *et al*. "Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer's disease," *Neurobiol. Aging*, vol. 24, pp. 1063-70, 2003.

Osborn G.G. and Saunders A.V. "Current treatments for patients with Alzheimer Disease,"*J. Am. Osteopath Assoc.*, vol. 110 (Suppl. 8), pp. S16-26, 2010.

Ottolenghi C., *et al.* "Determination and stability of sex," *BioEssays: News Rev. Mol. Cell. Dev. Biol.*, vol. 29, pp. 15-25, 2007.

Palmer J.S., *et al.* "Melanocortin-1receptor polymorphisms and risk of melanoma: Is the association explainedsolely by pigmentation phenotype?," *Am. J. Hum. Genet.*, vol. 66, pp. 176-86, 2000.

Panigrahi P.P. and Singh T.R. "Computational analysis for functional and evolutionary aspects of BACE-1 and associated Alzheimer's related proteins," *IJCI Stud.*, vol. 1, pp. 322-32, 2012.

Panigrahi P.P. and Singh T.R. "Computational studies on Alzheimer's disease associated pathways and regulatory patterns using microarray gene expression and network data: Revealed association with aging and other diseases,"*J. Theor. Biol.*, vol. 334, pp. 109-21, 2013.

Parihar M.S. and Hemnani T. "Alzheimer's disease pathogenesis and therapeutic interventions,"*J. Clinical Neurosci.*, vol. 11, pp. 456-67, 2004.

Parrinello M. and Rahman A. "Polymorphic transitions in single-crystals- a new molecular-dynamics method," *J. Applied Phys., vol. 52*, pp. 7182-90, 1981.

Paterson D. and Nordberg A. "Neuronal nicotinic receptors in the human brain," *Prog.Neurobiol.*, vol. 61, pp. 75-111, 2000.

Peel A.L., *et al.*"Tau phosphorylation in Alzheimer's disease: potential involvement of an APP-MAP kinase complex," *Neuromolecular Med.*, vol. 5, pp. 205-18, 2004.

Pettersen E.F., *et al.* "UCSF Chimera-a visualization system for exploratory research and analysis," *J. Comput.Chem.*, vol. 25, pp. 1605-12, 2004.

Qiu C. "Preventing Alzheimer's disease by targeting vascular risk factors: hope and gap," *J. Alzheimers Dis.*, vol.32, pp. 721-31, 2012.

Quackenbush J. "Microarray analysis and tumor classification," *N Engl J. Med.*, vol. 354, pp. 2463-72, 2006.

Ray M. and Zhang W. "Analysis of Alzheimer's disease severity across brain regions by topological analysis of gene co-expression networks," *BMC. Syst. Biol.*, vol. 4, pp. 136, 2010.

Ray M., *et al.* "Variations in the transcriptome of Alzheimer's disease reveal molecular networks involved in cardiovascular diseases," *Genome Biol.*, vol. 9, pp. r148, 2008.

Reitz C., *et al.* "SORCS1 alters amyloid precursor protein processing and variants may increase Alzheimer's disease risk," *Ann. Neurol.*, vol. 69, no. 1, pp. 47-64, 2011.

Ripich D.N. and Horner J. "The Neurodegenerative Dementias: Diagnoses and Interventions," *The ASHA Leader*, vol. 9, pp. 4-5, 2004.

Rivals I., *et al.* "Enrichment or depletion of a GO category within a class of genes: which test?," *Bioinformatics*, vol. 23, pp. 401-7, 2007.

Ronen M., *et al.* "Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, pp. 10555-60, 2002.

Rongve A., *et al.*"Alzheimer's disease and genetics," *Tidsskr. Nor. Laegeforen.*, vol. 133, pp. 1449-52, 2013.

Ruan J. and Zhang W. "Identification and evaluation of functional modules in gene co-expression networks," In: Proceedings of RECOMB Satellite Conferences on Systems Biology and Computational Proteomics, San Diego, CA, pp. 57-76, 2006.

Ruan J., *et al.* "A general co-expression network-based approach to gene expression analysis: comparison and applications," *BMC. Syst. Biol.*, vol. 4, pp. 8, 2010.

Saeed A.I., *et al.*"TM4: a free, open-source system for microarray data management and analysis," *Biotechniques*, vol. 34, pp. 374-8, 2003.

Said M.R., *et al.*"Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 18006-11, 2004.

Sandelin A., *et al.*"JASPAR: an open-access database for eukaryotic transcription factor binding profiles," *Nucleic Acids Res.*, vol. 32, pp. D91–D94, 2004.

Savas S. and Ozcelik H. "Phosphorylation states of cell cycle and DNA repair proteins can be altered by the nsSNPs,"*BMC. Cancer*, vol. 5, pp. 107, 2005.

Savas S., *et al.* "Special considerations in prognostic research incancer involving genetic polymorphisms,"*BMC. Medicine*,vol.11, pp. 149, 2013.

Scacchi R., *et al*. "A mutation screening by DHPLC of PSEN1 and APP genes reveals no significant variation associated with the sporadic late-onset form of Alzheimer's disease," *Neurosci. Lett.*, vol. 418, pp. 282-5, 2007.

Scarpini E., *et al.* "Treatment of Alzheimer's disease: current status and new perspectives," *Lancet Neurol.*, vol. 2, pp. 539-47, 2003.

Schmalhofer O., *et al.* "E-cadherin, β-catenin, and ZEB1 in malignant progression of cancer," *Cancer Metastasis Rev.*, vol. 28, pp. 151-66, 2009.

Schneider T.D., *et al*. "Information contentof binding sites on nucleotide sequences," *J. Mol. Biol.*, vol. 188, pp. 415-31, 1986.

Schreiber F. and Schwöbbermeyer H. "Mavisto: a tool for the exploration of network motifs. Bioinformatics," vol. 21, pp. 3572-4, 2005.

Scodellaro C. and Pin S. "The ambiguous relationships between aging and Alzheimer's disease: A critical literature review," *Dementia (London)*,vol. 12, pp. 137-51, 2013.

Scott L.J. and Goa K.L. "Galantamine: a review of its use in Alzheimer's disease," *Drugs*, vol. 60, pp. 1095-122, 2000.

Sealfon R., *et al.* "GOLEM: an interactive graphbased gene-ontology navigation and analysis tool," *BMC. Bioinformatics*, vol. 7, pp. 443, 2006.

Sehgal M. and Singh T.R. "DRGAS: A database of functional genetic variants and their phosphorylation states in human DNA repair systems," *DNA Repair*, vol. 16, pp. 97-103, 2014.

Sehgal M. and Singh T.R. "Identification and analysis of biomarkers for repair proteins: a bioinformatic approach," *J. Nat. Sci. Biol. Med.*, vol. 2, pp. 139-46, 2012.

Seiler K.P., *et al.* "ChemBank: a small-molecule screening and cheminformatics resource database," *N. A. R.*,vol. 36 (suppl 1), pp. D351-D359, 2008.

Shachar R., *et al.* "A systems-level approach to mapping the telomere-length maintenance gene circuitry," *Mol. Syst. Biol.*, vol. 4, pp. 172, 2008.

Shastry B.S., "SNPs and haplotypes: Genetic markers for disease and drug response (Review)," *Int. J. Mole. Medicine*,vol.11, pp. 379-82, 2003.

Shen-Orr S.S., *et al.*"Network motifs in the transcriptional regulation network of Escherichia coli," *Nat. Genet.*, vol. 31, pp. 64-8, 2002.

Sherry S.T., *et al*. "dbSNP: the NCBI database of genetic variation," *N. A. R.*, vol. 29, pp. 308-11, 2001.

Sherman W., *et al.* "Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects," *J. Med. Chem.*, vol. 49, pp. 534-53, 2005.

Sherman W., *et al.* "Use of an Induced Fit Receptor Structure in Virtual Screening," *Chem. Biol. Drug Des.*, vol. 67, pp. 83-84, 2006.

Sherwin B.B. "Estrogen and cognitive aging in women," *Trends.Pharmacol. Sci.*, vol. 23, pp. 527-34, 2002.

Shobab L.A., *et al*. "Cholesterol in Alzheimer's disease," *Lancet.Neurol.*, vol. 4, pp. 841-52, 2005.

Singh T.R., *et al.* "Computational identification and analysis of single-nucleotide polymorphisms and insertions/deletions in expressed sequence tag data of Eucalyptus," *J. Genetics*, vol. 90, pp. e34-e38, 2011.

Sjoblom T., *et al.* "The consensus coding sequences of human breast and colorectal cancers," *Science*, vol. 314, pp. 268-74, 2006.

Smit A.B., *et al.* "A glia-derived acetylcholine-binding protein that modulates synaptic transmission," *Nature*, vol. 411, pp. 261-8, 2001.

Smith M.A., *et al.*"Amyloid-Beta and Tau serve antioxidant functions in the aging and Alzheimer brain," *Free Radic. Biol. Med.*, vol. 33, pp. 1194-9, 2002.

Sonkusare S.K., *et al.*"Dementia of Alzheimer's disease and other neurodegenerative disorders-memantine, a new hope," *Pharmacological Res.*, vol. 51, pp. 1-17, 2005.

Sporns O. and Honey C.J. "Small world inside big brains," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, pp.19219-20, 2006.

Stark C., *et al.* "BioGRID: a general repository for interaction datasets," *Nucleic Acids Res.*, vol. 34, pp. D535-D539, 2006.

Stekel D. "Bioinformatics, Microarray," Analysis of Differentially Expressed Genes, Cambridge University Press, New York, pp. 110-138, 2003.

Su J.H., *et al*. "Plaque biogenesis in brain aging and Alzheimer's disease. I. Progressive changes in phosphorylation states of paired helical filaments and neurofilaments," *Brain Res.*, vol. 739, pp. 79-87, 1996.

Sugiura T., *et al.* "2-Arachidonoylglycerol: a possible endogenous cannabinoid receptor ligand in brain,"*Biochem.Biophys. Res. Commun.*, vol. 215, pp. 89-97, 1995.

Suh Y.H. and Checler F. "Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease," *Pharmacol Rev.*, vol. 54, pp. 469-525, 2002.

Sussman J.L., *et al.* "Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein," *Science*, vol. 253, pp. 872-9, 1991.

Suyama M., *et al.* "PAL2NAL: robust conversion of protein sequencealignments into the corresponding codon alignments," *Nucleic Acids Res.*,vol.34, (Web Server issue), pp. W609-W612, 2006.

Szklarczyk D., *et al.* "The STRING database in: functional interactionnetworks of proteins, globally integrated and scored," *Nucleic Acids Res.*, vol.39, (Database issue), pp. D561-8, 2011.

Talbot K., *et al.* "Demonstrated brain insulin resistance in Alzheimer's disease patients is associated with IGF-1 resistance, IRS-1 dysregulation, and cognitive decline," *J. Clin. Invest.*, vol. 122, pp. 1316-38, 2012.

Tamura K., et al, "MEGA5: Molecular Evolutionary Genetics Analysis usingmaximum likelihood, evolutionary distance, and maximum parsimony methods,"*Mol. Biol. Evol.*, vol. 28, pp. 2731-9, 2011.

Tanaka Y., *et al.* "The transcriptional programme controlled by Runx1 during early embryonic blood development," *Dev. Biol.*, vol. 366, pp. 404-419, 2012.

Tarawneh R. and Holtzman D.M. "Biomarkers in translational research of Alzheimer's disease," *Neuropharmacology*, vol. 59, pp. 310-22, 2010.

Tatsuno G., *et al.* "Functional Analysis of β-secretase Using Mutagenesis and Structural Homology Modeling," in Alzheimer's Disease: Advances in Etiology, Pathogenesis and Therapeutics. Iqbal K., Sisodia S. S. and Winblad B. (Eds), John Weily & Sons, Ltd. Macromolecular Structural Group, vol. 15, pp. 800-807, 2001.

Terry A.V. and Buccafusco J.J. "The cholinergic hypothesis of age and Alzheimer's disease-related cognitive deficits: recent challenges and their implications for novel drug development," *J. Pharmacol. Exp. Ther.*, vol. 306, pp. 821-7, 2003.

Thorisson G.A., *et al*. "The International HapMap Project Web site," *Genome Research*, vol. 15, pp. 1592-3, 2005.

Tilgner H., *et al.* "Nucleosome positioning as a determinant of exon recognition," *Nat. Struct. Mol. Biol.*, vol. 16, pp. 996-1001, 2009.

Toda N., *et al.* "A conformational restriction approach to the development of dual inhibitors of acetylcholinesterase and serotonin transporter as potential agents for Alzheimer's disease," *Bioorg.Med.Chem.*, vol. 11, pp. 4389-415, 2003.

Tusher V.G., *et al.*"Significance analysis of microarrays applied to the ionizing radiation response," *Proc.Natl. Acad. Sci.*, vol. 98, pp. 5116-21, 2001.

Vafa B. and Schofield P.R. "Heritable mutations in the glycine, GABAA, and nicotinic acetylcholine receptors provide new insights into the ligand-gated ion channel receptor superfamily," *Int. Rev. Neurobiol.*, vol. 42, pp. 285-332, 1998.

Van Dam D. and De Deyn P.P. "Drug discovery in dementia: the role of rodent models," *Nat. Rev. Drug Discov.*, vol. 5, pp. 956-70, 2006.

Van Gool W.A., *et al.*"Effect of hydroxychloroquine on progression of dementia in early Alzheimer's disease: an 18- month randomised, double-blind, placebo-controlled study," *Lancet*, vol. 358, pp. 455-60, 2001.

Vassar R., *et al.* "β-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE,"*Science*, vol. 286, pp.735-41, 1999.

Venugopal C., *et al.* "Beta-secretase: Structure, Function, and Evolution," *CNS Neurol. Disord. Drug Targets*, vol. 7, pp.278-94, 2008.

Von Arnim C.A., *et al.*"The low density lipoprotein receptor-related protein (LRP) is a novel beta-secretase (BACE1) substrate," *J. Biol. Chem.*, vol. 280, pp. 17777-85, 2005.

Von Eichborn J.,*et al.* "Structural features and evolution of protein-protein interactions," *Genome Inform.*, vol. 22, pp. 1-10, 2010.

Wang B., *et al.* "Isoform-specific knockout of FE65 leads to impaired learning and memory," *J. Neurosci. Res.*, vol. 75, pp. 12-24, 2004.

Warren J. and Strittmatter M.D."Alzheimer's disease: the new promise,"*J. Clin.Invest.*, vol. 122, pp. 1191, 2012.

Wei B.Q., *et al.* "Testing a flexible-receptor docking algorithm in a model binding site," *J. Mole.Biol.*, vol. 337, pp. 1161-82, 2004.

Wei D.Q., *et al.* "Theoretical studies of Alzheimer's disease drug candidate [(2,4-dimethoxy) benzylidene]-anabaseine dihydrochloride (GTS-21) and its derivatives," *Biochem. Biophys. Res. Commun.*, vol. 338, pp. 1059-64, 2005.

Wenk G.L. "Neuropathologic changes in Alzheimer's disease," *J. Clin.Psychiatry.*, vol 64 (Suppl. 9), pp. 7-10, 2003.

Wenk G.L., *et al.*"The effects of mitochondrial failure upon cholinergic toxicity in the nucleus basalis," *Neuroreport*, vol. 7, pp. 1453-6, 1996.

Wernicke S. and Rasche F. "FANMOD: a tool for fast network motif detection," *Bioinformatics*,vol.22, pp. 1152-3, 2006.

Whitehouse P.J. "The concept of subcortical and cortical dementia: another look," *Ann. Neurol*., vol. 19, pp. 1-6, 1986.

Wilkins R.H. and Brody I.A. "Alzheimer's Disease," *Arch. Neurol*.,vol.21, pp. 109-10, 1969.

Witt A., *et al.* "Memantine hydrochloride," *Nat. Rev. Drug. Discov.*, vol. 3, pp. 109-10, 2004.

Yamada A., *et al.*"In silico and in vivo identification of the intermediate filament vimentin that is downregulated downstream of Brachyury during Xenopus embryogenesis," *Gene*, vol. 491, pp. 232-6, 2012.

Yang Y.H., *et al.*"Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation," *Nucleic Acids Res.*, vol. 30, pp. e15, 2002.

Yasuda T., *et al.* "Nucleosomal structure of undamaged DNA regions suppresses the non-specific DNA binding of the XPC complex," *DNA Repair (Amst.)*, vol. 4, pp. 389-95, 2005.

Zempel H., *et al.* "Aβ oligomers cause localized Ca2+ elevation, missorting of endogenous Tau into dendrites, Tau phosphorylation, and destruction of microtubules and spines," *J. Neurosci.*, vol. 30, pp. 11938-50, 2010.

Zhang B., *et al.* "WebGestalt: an integrated system for exploring gene sets in various biological contexts," *Nucleic Acids Res.*, vol. 33, pp. W741-W748, 2005.

Zhang Y.W., *et al.*"APP processing in Alzheimer's disease," *Mole.Brain*, vol.4, 3, 2011.

PUBLICATIONS AND PRESENTATIONS

PAPERS IN INTERNATIONAL REFEREED JOURNALS:

Priya P. Panigrahi and Tiratha Raj Singh. Computational analysis for functional and evolutionary aspects of BACE-1 and associated Alzheimer's related proteins. *International Journal of Computational Intelligence Studies*, vol. 1, pp. 322-32, 2012.

Priya P. Panigrahi and Tiratha Raj Singh. Computational analysis for Alzheimer's disease associated pathways and regulatory patterns using Microarray gene expression and network data reveal association with other diseases. *Journal of Theoretical Biology*, vol. 334, pp. 109-21, 2013.

Priya P. Panigrahi, Manika Sehgal and Tiratha Raj Singh, *ADDGAP*: A comprehensive repository on Alzheimer's disease for genetic association studies and phosphorylation states, 2014. [Under-Revision]

Priya P. Panigrahi, Ramit Singla, Moacyr Comar JuniorVikas Jaitak, and Tiratha Raj Singh,*In silico* Screening and Molecular Interaction Studies to Test the Binding Affinity of Δ9-Tetrahydrocannabinol and its Derivatives on Acetylcholine Binding Protein. 2014. [Under-Review]

PRESENTATIONS IN NATIONAL AND INTERNATIONAL CONFERENCES:

Priya P. Panigrahi and Tiratha Raj Singh "Functional annotation of alzheimer's related proteins through protein-protein interactions and other *in silico* methods" in 2nd international bioinformatics conference under the aegis of IFIP–TC 5 and CSI held during 22-25 September, 2011 at S.S. Dempo College, Panji, Goa, India.

Digvijay Singh Chauhan, Priya P. Panigrahi, **andTiratha Raj Singh.** Development of early phase diagnostics for Alzheimer's disease: use of diabetes mellitus as surrogate biomarker. International Conference on Bioinformatics (INCOB), Oct. 3-5, 2012, Bangkok, Thailand.

Priya P. Panigrahi, Ankit Gupta, Gautam Mehta, Ahmed Moussa, and Tiratha Raj Singh, Web Repository for Mitochondrial and Neurological Associated Human Disease, ISCB Africa ASCBC Conference on Bioinformatics, 14th March, 2013, Rabat, Morocco.

Priya P. Panigrahi and Tiratha Raj Singh "Enrichment analysis for Alzhiemer's disease associated pathways and regulatory patterns", in 3rd International bioinformatics conference under the aegis of IFIP–TC 5 and CSI held during 23-26 September, 2013 at MANIT, Bhopal, India.

ACADEMIC ACHIVEMENTS:

Resource Person in Computer Aided Drug Design module for one month Summer Training held during 1st to 30th June, 2012 organized by Department of Biotechnology and Bioinformatics, JUIT, Waknaghat, Solan-173234, H.P., India.

Resource Person in the workshop on Computer Application in Genomics and Proteomics (Sponsored by the Dept. of Biotechnology, Govt. of India) held during 2nd to 3rd March, 2012 At Department of Agricultural Biotechnology and Bioinformatics Center, CSK Himachal Pradesh Krishi Vishvavidyalaya, Palampur, H.P., India.