CHAPTER 1: INTRODUCTION

1.1 Alzheimer disease

Alzheimer is the most common form of dementia, a general term for memory loss and other intellectual abilities serious enough to interfere with daily life (Alzheimer's Association, 2010). The most popular current hypothesis is that aggregated amyloid peptide (Aβ) is the major cause of Alzheimer's disease (AD) (Venugopal et al, 2008).

Aβ is a small fragment of a larger protein called "amyloid precursor protein" (APP). APP is an integral membrane protein can be presented in many membranous structures in the cell such as the endoplasmic reticulum and Golgi compartments. It also has the ability to be localized to the cell membrane. After reaching the cell membrane, APP is ultimately re-internalized and exposed to processing in lysosomes. Remarkably, APP was also identified in the membranes of synaptic preparations and was as well localized to postsynaptic densities, axons and dendrites (Schubert et al., 1991)

Three types of proteases, which are α, β and γ-secretases can cleave APP into smaller sections. β and γ-secretases cleave on the N and C-terminal ends respectively and causes to generating Aβ, which contains 42-43 amino acid residues Positioned partially within the ectodomain and partially within the transmembrane area of APP. Based on the amyloid hypothesis, the generation of Aβ is considered a major occurrence of AD. Nevertheless, cleavage of APP by α-secretases destroyed the Aβ fragment. The general belief is that amyloid formation is mitigated by α-secratase pathway (Nunan and Small, 2000).

BACE1 and BACE2 are two types of beta-secratase. Largest part of the Aβ in the human brain is most likely produced by BACE1. In AD cases, the BACE1 proteins and its activities were considerably amplified. In contrast to BACE1, BACE2 activity does not

contribute much to the fact. However, even minor alterations in BACE2 activity can have negative outcomes for human disease because both forms of BACE contend for the same substrate pool (Rachel et al., 2010).

In order to prevent the effects of Aβ, a number of strategiesare being explored by the scientists. A number of medicines aiming at curing Aβ are being tested during human clinical trials, however no clear and flaw less sign showing that Alzheimer symptoms can be improved or that they can protect the brain cells. At the present time, the aim of the experimental strategies which focus on Aβ is decreasing the production of Aβ, stopping it from getting aggregated, or increasing the process of Aβ removal from brain cells. Experimental medicine alters the behavior of secretases that cut APP into Aβ in order to reduce the production of Aβ (Alzheimer's Association 2008).

In this project I am going to study the conservation pattern of BACE1 and BACE2 throughout species and comparing their sequence and structure. As the protein structure and function seem to be closely related, understanding the relationship of BACE1 and BACE2 sequences and structure to function could help to design the selective inhibitors and prevent or reduce Aβ production.

1.2 Study Approach

The research of this project was done based on the Evolutionary Trace Method (ET). ET is a method which can be used to search for conserved residues within different branches of an evolutionary tree and identified the active site in a protein sequence. This method aims at extracting the residues which are functionally important. In homologous proteins, it extracts the residues from the sequence conservation patterns. The other purpose would be to create clusters which identify functional interfaces on their mapping onto the

protein surface, besides implying certain data regarding the relationships of protein sequence-structure-function (Lichtarge 1996, as sited by Spartt, 2000).

Multiple sequence alignment of a protein family and an evolutionary tree are required by this method that is structured on sequence identity and can be close to the functional categorization of the protein sequences (Figure 1 simplify the steps involved in ET method). This method divides the information which is inherent in a group of homologous proteins into high functional similarity within groups and functional variation between groups. This step involved the subdivision of the protein sequences into groups. A consensus sequence is formed for every single class and the sequences will later be compared in order to create the evolutionary trace sequence. The position of the residues which are invariant within these classes, but are various among themselves, are indicated by class-specific or trace residues.

At that point, in order to pinpoint a single, dominant spatial cluster of conserved and class-specific residues, the trace residues are shown in a three-dimensional structure of a family member. The residues which are important to the performance of all the proteins in the family will probably remain unchanged during the evolution and do not go through any mutation events. Meanwhile, the residues that contribute more to the functionary specificity of a particular subgroup of protein in the family will be conserved. The residues which are less vital to function will also be under evolutionary pressure to stay unchanged and can be subjected to a higher rate of mutation. In the end, clustering of low ranked residues is what we search for, where a functional site is represented by clusters of trace residues. A cluster formed by these residues can act as a binding site for the effectors' protein (Lichtarg & Sowa, 2002).

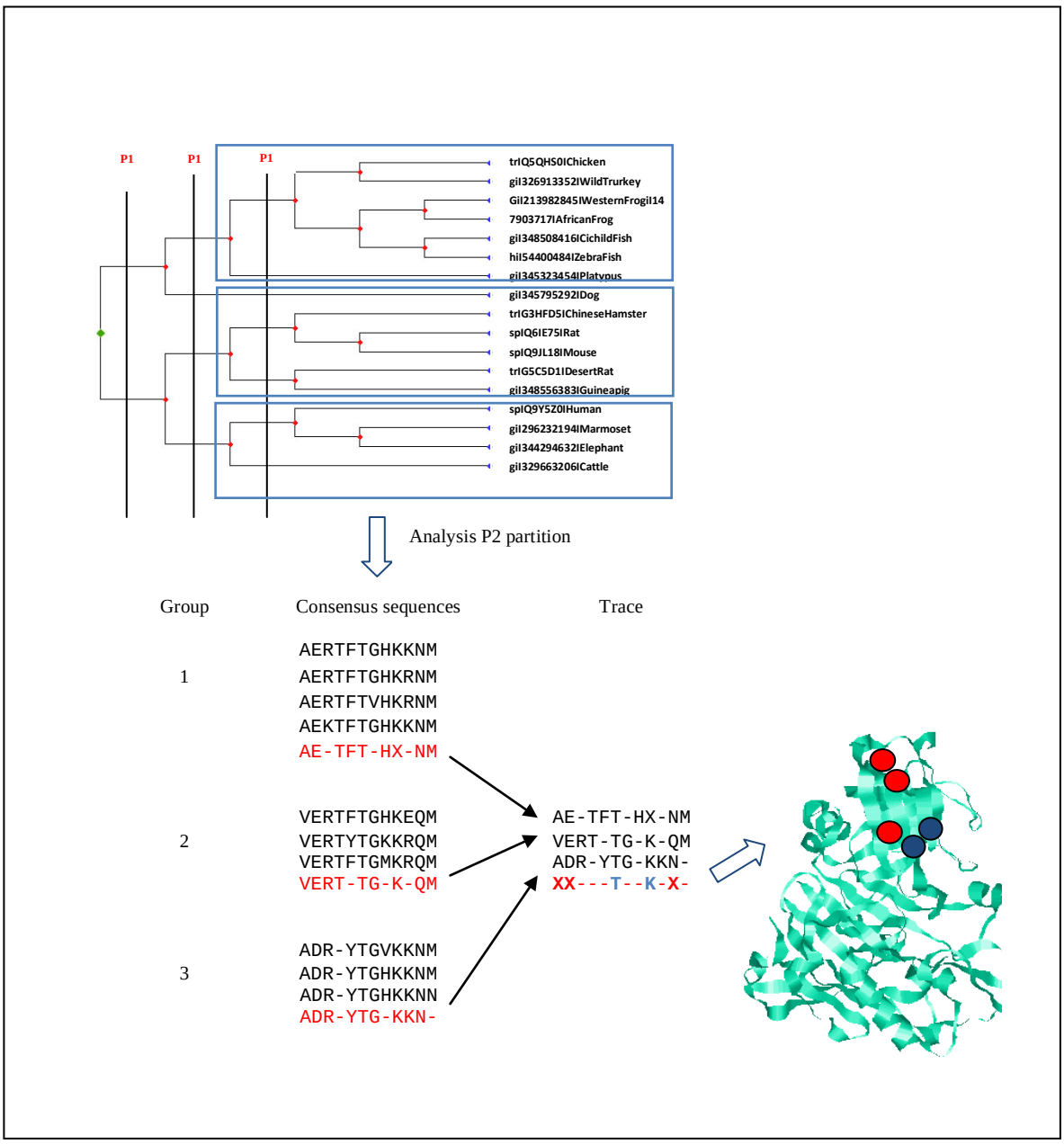

Figure1.1 Overview of the prediction of active site of protein by using Evolutionary Trace Method. A phylogenetic cut off into 3 partitions. Analysis at partition P2, derived the sequences of the phylogenetic tree in to 3 groups. Consensus sequences derived from the groups are colored red. Consensus sequences are compared to derive the tarce sequences. In trace sequences, X (red color) are residues that conserved within the groups but different in one group to another. Residues that are conserved within the all groups are labeled by amino acid short name (blue color).

1.3 Objective

The objective of this study is:

1. To identify the phylogenetic relationship of the BACE1 and BACE2 among different species by using evolutionary trace method.
2. To study the sequence and structure of BACE1 and BACE2 enzymes.
3. To compare the ligand binding site of BACE1 and BACE2.

CHAPTER 2: LITRETURE REVIWE

2.1 Alzheimer's disease

Alzheimer’s disease (AD) is a chronic advance neurodegenerative disorder, caused by three groups of symptoms. Cognitive dysfunction is the first group which is includes memory loss, language impairment, and loosing the ability for planning and understanding things and think intelligently. The second group is Psychiatric symptoms and behavioral disorder. Hallucinations, depression, delusions and commotion are the example of this group. The third group consists of difficulties with doing daily activities and quality of life for example some complex activities such as eating, dressing, shopping and driving. The symptoms of AD developed from calm symptoms of loosing memory to intense dementia (Burns & Iliff, 2009).

2.2 Molecular causes of AD

AD is characterized by the existance of proteinaceous deposits which exist in the brain and mostly contains extracellular amyloid plaques and congophilic amyloid angiopathy and intracellular neurofibrillary tangles. The extracellular amyloid deposits are primarily made of amyloid protein (Aβ). Aβ is derived by cleavage of amyloid protein precursor (APP) which is a type I transmembrane glycoprotein and contains 695-770 amino acids. The sequence of 42-43 amino acid residues encompassed by the Aβ region of APP is partly positioned in the ectodomain and the remaining parts are situated within the transmembrane domain of APP (Nunan & Small, 2000).

There are two ways for cleavage of APP, non-amyloidogenic and amyloidogenic pathways (Franketet al., 2007). α-secretase cleaved APP on the C-terminal direction of the Aβ

sequence at residues 16 and results in the generation of an 83-residue C-terminal fragment (C83) in the non-amyloidogenic pathway. Then gamma-secretase cleavage Subsequent and release a short peptide (p3) which contains the C-terminal region of the Aβ peptide. Cleavage by α-secretase preventing the generation and release of the Aβ peptide.

APP molecules that are not cleaved by the non-amyloidogenic pathway become a substrate for beta-secretase enzyme activity through the amyloidogenic pathway which leads to Aβ generation (Figure 2.1) (Nunan & Small, 2000).

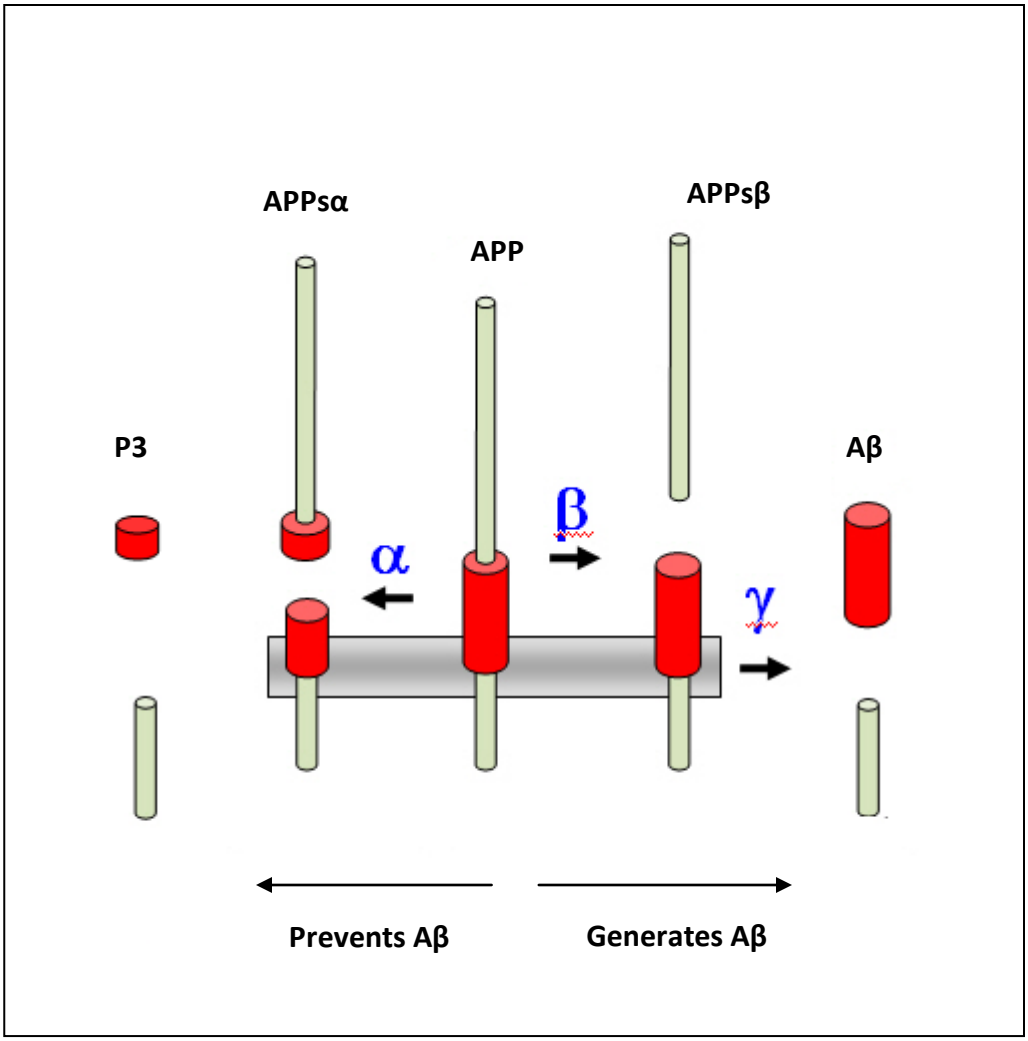

Figure 2.1 Non-amyloidogenic and amyloidogenic pathways of APP cleavages.

2.3 Beta Secretase

There are eight human enzymes in the pepsin-like family of aspartic proteases (A1): napsin A, pepsin A and C, cathepsin D and E, rennin, Beta amyloid cleaving enzymes 1 (BACE1) and 2 (BACE2).Two aspartic acids are located in the active site of these protein family members which used for catalysis the substrates (Ostermann et al., 2006).

BACE1 is a type1 transmembrane protein which is greatly expressed in brain and pancreas, it can also be found in other organs as well, at considerably smaller doses (Rachel et al., 2010). Based on the attained data, BACE1 only cleaved membrane-bound substrates efficiently and signified that the enzyme was likely to be membrane-bound or a close associate of a membrane protein. In addition, largest amount of BACE1 activity was noticed at acidic pH within the subcellular sections of the secretory pathway incorporates the trans-Golgi network (TGN) and endosomes (Cole & Vassar, 2007).

As a homologue of BACE1, BACE2 enjoys a 61.5% similarity and 45% identity at the amino acid level. In majority of peripheral tissues, BACE2 can be found at different levels, where kidney tissues have the top. It is believed that in brain BACE2 is mainly astrocytic, while BACE1 is basically neuronal. BACE2 is situated on chromosome 21 inside the critical region of obligate Down syndrome and might lead to amyloid pathology in these people (Rachel et al., 2010). It is also important to know that in breast cancers, BACE2 is over-expressed and with BACE1, they both participate in the processes of biology of human muscle, whether normal or abnormal (Ostermann et al., 2006).

2.4 APP Cleavage

APP is cleaved at the N terminus of Aβ (β-site) by BACE1 between amino acids Met-671 and Asp-672, producing a membrane-associated APP fragment of 99 aa (C99). Within the transmembrane domain of C99, there exists a second site which is named γ-site that can be cleavage by a γ-secretase to release Aβ (Figure 2.2). In addition, BACE1 can cleave APP between Tyr-10 and Glu-11 (β'-site) of the Aβ region with similar efficiency. Such Aβ fragments which produce cleavage at the said site was detected in senile plaques and it has been shown that the fragments of this size are further amyloidogenic and neurotoxic than a full-length Aβ (Farzan et al., 2000).

BACE2 and BACE1 compete over substrate and cleave APP at β site (Rachel et al., 2010). However, in Alzheimer's disease, unlike BACE1, BACE2 does not play a role in producing neuritic plaques (Ostermann et al., 2006).

BACE2 can cleave APP within the Aβ region, after Phe-19 and Phe-20 and precluding Aβ formation. Aβ production is stopped by the cleavage of APP created by BACE2 at this site. Moreover, BACE2 may be able to reduce the pathogenic forms of Aβ being produced (fragments beginning at Asp-1 or Glu-11) within the cells that express both BACE1and BACE2 (Figure 2.2).The Aβ fragments which are created by the internal cleavages possibly enjoy diverse clinical results, even though both enzymes cleave within Aβ. The maximum size of Aβ fragments which were produced by BACE1 can be found in senile plaques. These plaques contained the HHQK sulfate-binding region of Aβ which can be associated with sulfated proteoglycans. On the contrary, the internal fragments cleaved by BACE2 have not been seen in senile plaques and they also lack the HHQK domain.

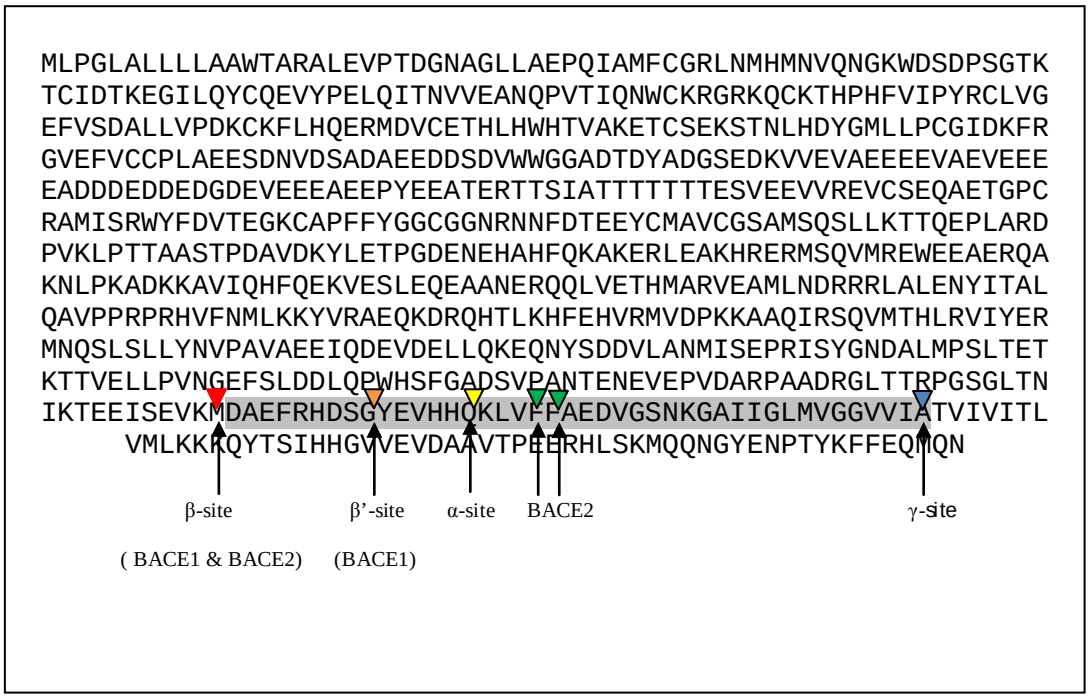

Figure 2.2 Human APP sequence, Aβ region (highlighted with grey color) and α , β, γ-secretase cleavages sites. The β-site is cleaved by BACE1 competing with BACE2 (red arrow). The β'-site is only cleaved by BACE1 (orange arrow). BACE2cleavages APP in two regions (green arrows).The α and γ-sites (yellow and blue arrows) are cleaved by α and γ-secratase.

2.5 BACE1 & BACE2 Structures

The structure of BACE1contains three different domains: an N-terminal ectodomain (Ser10-His145), interdomain (Val146-Ile193, Gly340-Ala358) and a cytosolic C-terminus (Arg194-Glu339, Cys359-Glu380). At the sites vulnerable against BACE1, the N-terminal domain is the protease domain and enjoys the right topological orientation for cleavage of APP. The protease domain has a conserved aspartic protease fold, with the substrate-binding cleft located between the N- and C-terminal lobes (Xu et al, 2011).

The overall structure of BACE2 follows the general fold of aspartic proteases of the A1 Family. It contains an N-terminal domain (Ser26-Asn161), a C-terminal domain (Lys207-Ala398) and a six-stranded anti-parallel β-sheet interdomain. The N-terminal and C-terminal domains are connected to each other by the interdomain and it consists of Leu16-Asp25 (forming the outermost strand), Ile162-Ile206 and Gly352-Pro370 (Figure 2.3). Compared to pepsin A or cathepsin D, the dimensions of the C-terminal domain of BACE2 is significantly bigger. The 28 residue C-terminal extension binds itself round the C-terminal domain, and is fastened to it stronger by creation of two disulfide bonds (between Cys371and Cys171, and between Cys395 and Cys230) (Ostermann et al., 2006). Two conserved catalytic Asp, Asp32 and Asp228 in BACE1 and Asp32 and Asp228 in BACE2 that are essential for the enzyme activities are situated at the interface of the N terminal and C terminal domain of their structure (Figure 2.3). In both structure the center of the active site is shielded from solvent by the so-called flap, residues Val67-Gly78 in BACE1 and residues Val83-Gly94 in BACE2. It was observed that, a flap is the most flexible region, which is able to cover the active site and adopting multiple conformational states in the various crystal structures.

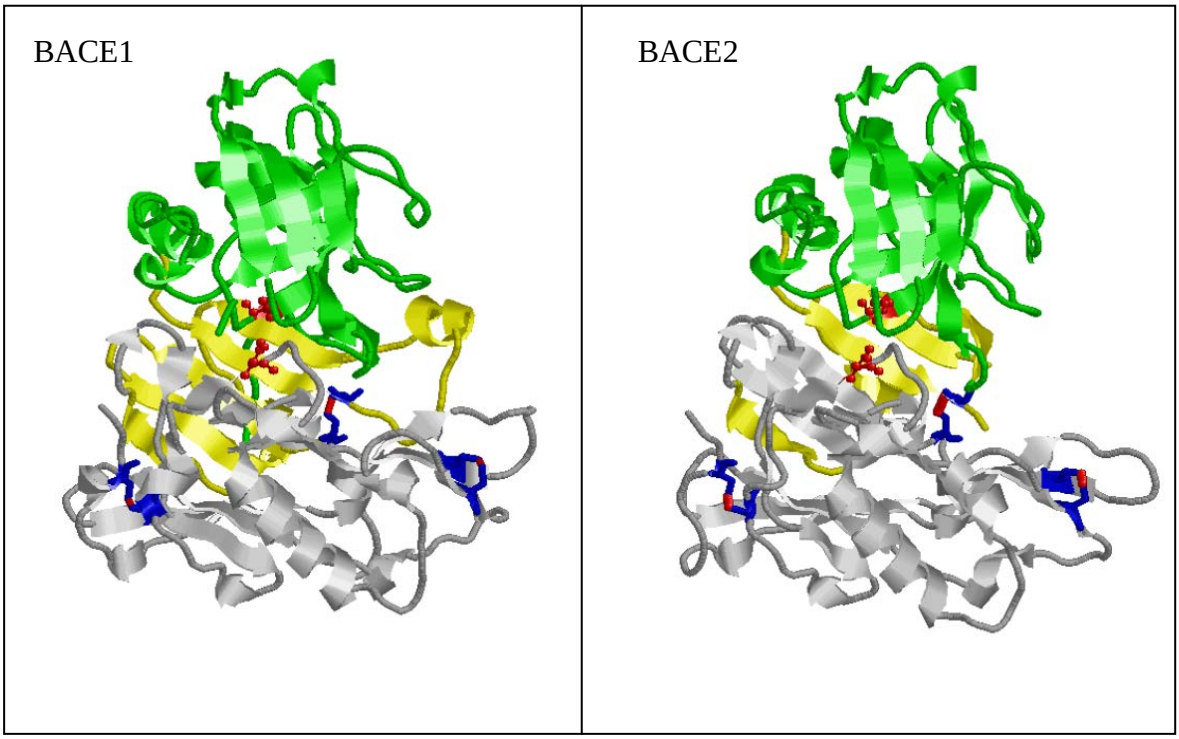

Figure 2.3 The overall structure of BACE1 and BACE2. The N-terminal, inter-domains and C-terminal are colored green, yellow and grey respectively. Two conserved catalytic Asp, Asp32 and Asp228 in BACE1 and Asp48 and Asp241 in BACE2 are displayed in form of “Ball and Stick” with red color. The Cys residues are presented in form of Sticks and blue color while, their disulfide bonds are colored red.

In BACE1, both the presence or absence of an inhibitor within the active site and the crystal packing are shown to influence the flap's conformation. An open conformation of the flap is mostly observed in the apo structures, while direct hydrogen bonding interaction between main-chain atoms of the flap and the inhibitor is a prerequisite for the flap to adopt a closed conformation in the crystal structures of complexes. When BACE2 complexed with inhibitors the flap adopts a closed conformation similar to that found in BACE (Xu et al., 2011)

The findings suggest that, even though the overall structure of BACE2 is so alike to the one in BACE1, a conformational difference exists in the C-terminal extension and in three loops, residues Ala42-Arg50, Phe108-Trp115 and Leu316-Thr 324 in BACE1 and Gly58-Tyr67, Phe124-Trp131and Ile319-Tyr332 in BACE2.Theseloops are packed closely against and their conformations might be influenced by a symmetry-related molecule in the crystal environment (Ostermann et al., 2006).

CHAPTER 3: METHODOLOGY

3.1 Selection of a protein sequence

The first step of Evolutionary Trace method is selection of candidate or seed protein sequence from the protein databases for a database search of homologous sequences. There are lot of protein database such as NCBI, SwissProt, Uniprot, Protein Data Bank (PDB) and GenBank. In this project, the sequences of my candidate proteins which are human's BACE1 and BACE2 (Uniprot Accession Number P56817, Q9Y5Z0) derived in FASTA format from Uniprot and GenBank databases and the PDB structures of them are derived from Protein data bank (PDB).

To ensure that two structures file which are obtained from PDB are identical with the candidate proteins in Uniprot, a pairwise sequence alignment between the sequence from PDB and the sequence from Uniprot had been performed. The result of alignment indicates that these two sequences are exactly identical.

3.2 Database Similarity Searching

Blast (Basic Local Alignment Search Tool), is known to be one of the most prominent tools for searching the database. As a software package, BLAST is able to find the local alignments with high scoring between a query sequence of interest and some target database (Spartt, 2000). There are some types of BLAST program which can be used in different types of database searching. In my case I chose BLASTP program where it allows the searching of protein database against a protein query sequence. By using BLAST program the list of all similar sequences resulted in BLAST page. Sequences which have

high percentage similarity with candidate proteins, lower the E-value and higher the score value have been chosen.

3.3 Multiple Sequence Alignment

After Blast Multiple Sequence Alignment (MSA) has to be done. The MSA protein aims at assembling the amino acid sequences in alignments, in a way that reflect their biological relationship, whether evolutionary, functional, structural, or a combination of the three (Gharegozlo, 2009).

The program that I used for this task is ClustalX which is a recent windows interface designed for the commonly utilized progressive multiple sequence alignment program ClustalW. The advantage of this system is that, it is easy to use and performing multiple sequences of large numbers of sequences and profile alignments and analyzing the results by providing an integrated system (Thompson et al., 1997). Selected sequences were run in ClustalX to derive the MSA.

3.4 Phylogenetic Tree Construction

Once the ClustalX has made an alignment, it can draw phylogenetic trees. As a figure, phylogenetic tree portrays the proximity of relationship between different organisms. Phylogenetic tree is in fact a graph that is formed by branches and nodes (indicating where two species become one). Each of the said branches is connected to the very two neighboring nodes in a bifurcating tree. The interactions among the internal nodes related to their ancestry and decent is represented by the branches. The present taxonomic units or species are represented by the terminal (external) nodes. Today, numerous and

various methods are being used to construct the phylogenetic tree. Some of these methods include; maximum parsimony, maximum likelihood and neighbor joining (Spartt, 2000).

The method that I chose for constructing the phylogenetic trees in ClustalX is neighbor-joining (NJ), which is a distance-based method. This method enjoys a greedy algorithm that tries to reduce the amount of all branch-lengths which are constructed on phylogenetic trees. The beginning is with a star-formed tree on which each of the leaves corresponds to a species, and repeatedly chooses the two nodes neighboring the root and through placing in a new node between the root and the two selected nodes, joins them together. In order to join the nodes it selects the two nodes which decrease the length of the branches in the new tree which is created (Lakshmi & Appa Rao, 2001).

The results of MSA were used to generate the phylogenetic trees with NJ method within ClustalX program. After drawing the trees, since the ClustalX does not provide any tool for viewing phylogeny tree, the tree saved in PHYLIP file format and open with an appropriate viewing program such as draw tree, Phylip, or PhyloDraw. In This project I used PhyloDraw program, and rectangle cladogram style to view the tree .The tree is then used to analysis the molecular evolution of species group.

3.5 Molecular Evolution Analysis

The tree would be cut into several vertical branches (partition cut off, P) so as to examine the data . The calculation of an evolutionary trace is possible for every single cut which is made in the tree. Based on the protein in which we are interest, we can determine the ideal number of branches to cut. A consensus sequence is collected and grouped for every single branch of the tree which is cut at a particular cut point. Every sequence in each groups were examined and a consensus sequences were derived, containing information of conserved or

not-conserved residues within the groups. All the consensus sequences from every groups of the particular partition cut off were arranged together and ET sequences were obtained.

ET sequences have the information of the protein sequences conservation pattern derived from the phylogenetic tree. A particular site will conserved as consensus if all the amino acid residues in that particular site are the same (show as the single letter amino acid on the trace sequences).

A group specific site is a site that is conserved within the group, however it is different in the group of the particular partition cut off. The residues at the group specific site are written as "X" in the trace sequences. The neutral site is site with no conservation and contains a gap in the alignment which will be completely ignored by the trace.

3.6 Protein Mapping

Protein mapping is used as the final step to find out the protein ligand binding site. For this task the 3D structure of BACE1 (1FKN) and BACE2 (2EWY) are derived from PDB databases. Then I used Rasmol program, a molecular graphics program which is used to create visualizations of proteins, nucleic acids and small molecules.

Through Rasmol program, for both structure the mapping is restricted to 5.0 angstrom within the ligand binding site of the protein. Those residues that are present within the restriction area had been observed and identified. The result of protein mapping will be then compared with the consensus sequences to identify the active site. If a residue that had been identified in protein mapping is also a conserved residue in consensus sequence, then it is considered present at the active site (Spartt, 2000).

CHAPTER4: RESULT

4.1 Candidate Sequences Selection

My Candidate proteins sequences which are human's BACE1 and BACE2 (P56817, Q9Y5Z0) were derived from UniProtKB database in FASTA format (Supplement Figure1).

4.2 Similarity Sequence searching

The human's BACE1 and BACE2 sequences were used as seed sequences and put in the Blast program within UniProt databases to search for related sequences between other species. When list of similar sequences were given as the result in BLAST page, for selecting the species sequences, the result was filtered based on the high percentage similarity and lower E-value. Sequences with 50% and above similarity percentage were selected. To ensure no BACE1 and BACE2 were thoroughly inspected, I searched within GenBank database, and found some BACE1 and BACE2 precursor and predicted proteins sequences from various species, and add them to my sequences' list. To find the similarity between those sequences which were derived from GenBank with human's BACE1 and BACE2 sequences, the pairwise sequence alignment were done.

The lists of the selected sequences from the databases are shown in Tables 4.1 for BACE1 and Table 4.2 for BACE2. In both tables, the "*" symbol was put instead of the E value of those sequences were not from BALST search and individually selected from GenBank database.

Table4.1 BACE1's sequences from different species with their accession ID, definition, species usual name, number of amino acid, similarity, score and E-value.

Accession	Definition	Usual name	No. amino acid	Similarity %	Score, bits	E-Value
P56817	Beta-secretase 1 (*Homo sapiens*)	Human	501	100.0%	2,665	0.0
XP_002708419.1	Beta-site APP-cleaving enzyme (Predicted) (*Oryctolagus cuniculus*)	European rabbit	501	99.0%	2611.0	*
XP_002807322.1	Beta-secretase 1-like (Predicted) (*Callithrix jacchus*)	Monkey	501	99.4%	2641.0	*
XP_003418287.1	Beta-secretase 1-like (*Loxodonta Africana*)	African bush elephant	502	98.0%	2568.0	*
G5BHK2	Beta-secretase 1 (*Heterocephalus glaber*)	Desert rat	426	98.0%	2,229	0.0
Q2HJ40	Beta-secretase 1 (*Bos taurus*)	Cattle	501	98.0%	2,603	0.0
P56818	Beta-secretase 1 (*Mus musculus*)	Mouse	501	96.0%	2,564	0.0
P56819	Beta-secretase 1 (*Rattus norvegicus*)	Rat	501	96.0%	2,566	0.0
G3IAK4	Beta-secretase 1 (*Cricetulus griseus*)	Chinese hamster	501	96.0%	2,547	0.0
Q1KLR6	Beta-secretase 1 (*Cavia porcellus*)	Guinea pig	473	91.0%	2,390	0.0
Q5QHS1	Beta-site APP cleaving enzyme (*Gallus gallus*)	Chicken	334	85.0%	1,458	1.0×10-159
NP_991267.1	Beta-secretase 1 precursor (*Danio rerio*)	Zebra fish	505	82.4%	1961.5	*
Q0P4T5	Beta-siteAPP-cleaving enzyme1(*Xenopus tropicalis*)	Western clawed frog	502	80.0%	2,069	0.0
XP_546508.3	Beta- secretase 1(Predicted) (*Canis lupus familiaris*)	Dog	392	76.8%	2011.0	*

Table 4.2 BACE2's sequences from different species with their accession ID, definition, Species usual name, number of amino acid, similarity, score and E-value.

Accession	Definition	Usual name	No. amino acid	Similarity %	Score, bits	E-Value
Q9Y5Z0	Beta-secretase 2 (*Homo sapiens*)	Human	518	100.0%	2,687	0.0
NP_001192991.1	Beta-secratase2 precursor (Bos Taurus)	Cattle	514	94.2%	2475.5	*
XP_003464002.1	Beta-secretase 2-like (Predicted) (*Cavia porcellus*)	Guinea pig	513	92.1%	2355.0	*
G5C5D1	Beta-secretase 2 (*Heterocephalus glaber*)	Desert rat	415	89.0%	1,970	0.0
Q9JL18	Beta-secretase 2 (*Mus musculus*)	Mouse	514	88.0%	2,392	0.0
Q6IE75	Beta-secretase 2 (*Rattus norvegicus*)	Rat	514	88.0%	2,405	0.0
G3HFD5	Beta-secretase 2 (*Cricetulusgriseus*)	Chinese hamster	378	82.0%	1,777	0.0
XP_002761485.1	Beta-site APP-cleaving enzyme 2 (predicted) (Callithrix jacchus)	Monkey	452	80.7%	2109.0	*
XP_535595.3	Beta-secratase 2(Predicted) (Canis lupus familiaris)	Dog	459	79.6%	2043.0	*
Q5QHS0	Beta-site APP cleaving enzyme 2 (*Gallus gallus*)	Chicken	392	78.0%	1,649	0.0
NP_001080615.1	Beta-site APP-cleaving enzyme 2 precursor (*Xenopus laevis*)	African clawed frog	500	77.2%	1694.0	*
NP_001135590.1	Beta-site APP-cleaving enzyme 2 precursor (*Xenopus Tropicalis*)	Western clawed frog	499	77.0%	1723.5	*
XP_001511090.2	Beta-secretase 2 (Predicted) (*Ornithorhynchus anatinus*)	Platypus	427	72.2%	1840.5	*

Table 4.2 continued.

Accession	Definition	Usual name	No. amino acid	Similarity %	Score, bits	E-Value
XP_003203003.1	Beta-secretase 2-like(Predicted) (*Meleagris gallopavo*|)	Wild Turkey	420	70.2%	1676.5	*
Xp_003441750.1	Beta-secratase2-like (Predicted (Oreochromis niloticus)	Cichild Fish	509	68.6%	1521.0	*
XP_003419020.1	Beta-secretase 2-like(Predicted) (*Loxodonta Africana*)	African Bush Elephant	323	60.4%	1559.5	*
NP_001005991	Beta-secretase2 precursor (*Danio rerio*)	Zebra fish	372	51.8%	1087.0	*

4.3 Multiple Sequence Alignments

After selecting species proteins, the amino acid sequences of the selected proteins in FASTA format were derived from the database and save in the text file for doing multiple sequence alignment in ClustalX and constructing phylogenetic tree to perform similarity comparison among them. Two separate text file were provided for BACE1 and BACE2's sequences. Then for comparing these two enzymes sequences with each other among the species by multiple sequence alignment, their all taken sequences were saved to gather in one text file. To make the BACE1 and BACE2 sequences easier to identify in subsequent output, I called the ones with BACE1 sequences number one and the ones with BACE2 sequences number two.

In this step, 3 multiple sequence alignment analyses were done for the group of sequences. Firstly, the groups of BACE1 and BACE2 were analyzed separately. Then all the proteins sequences containing BACE1 and BACE2 were grouped together for multiple sequence analyses.

The result of multiple sequence alignment analyses of BACE1 group with 14 selected sequences and BACE2 with 17 selected sequences from various species are displayed in Supplement Figure 4 and Supplement Figure 5. Also the result of multiple sequence alignment of BACE1 and BACE2 were grouped together is displayed in Supplement Figure 6.

4.4 Phylogenetic Tree Construction

By using the alignment results in ClustalX, 3 phylogenetic trees with neighbor-joining method were constructed for BACE1, BACE2, and BACE1 & BACE2 for further

analysis. The trees were saved as PHYLIP file (.ph) and run on PhyloDraw program. In PhyloDraw the rectangle cladogram diagram was used for viewing the trees.

Figure 4.1 displays the BACE1's phylogenetic tree. In this tree the first node was divided into two groups. The first group is consists of 11 species which are all mammalians except zebra fish, western frog and chicken. Human with other two mammalians are located in the second group. The BACE2's phylogenetic tree is displayed as Figure 4.2. As same as BACE1's phylogenetic tree, the first node was divided in to two groups. There are eight species in group1, contains, dog, birds, Fishes and amphibian, however Group 2 involves human and other eight mammalians. Figure 4.3 shows the Phylogenetic tree of BACE1& BACE2's sequences together. In this tree we have two major groups. The first group consisted of all BACE1's sequences while the BACE2's sequences are located in the second group.

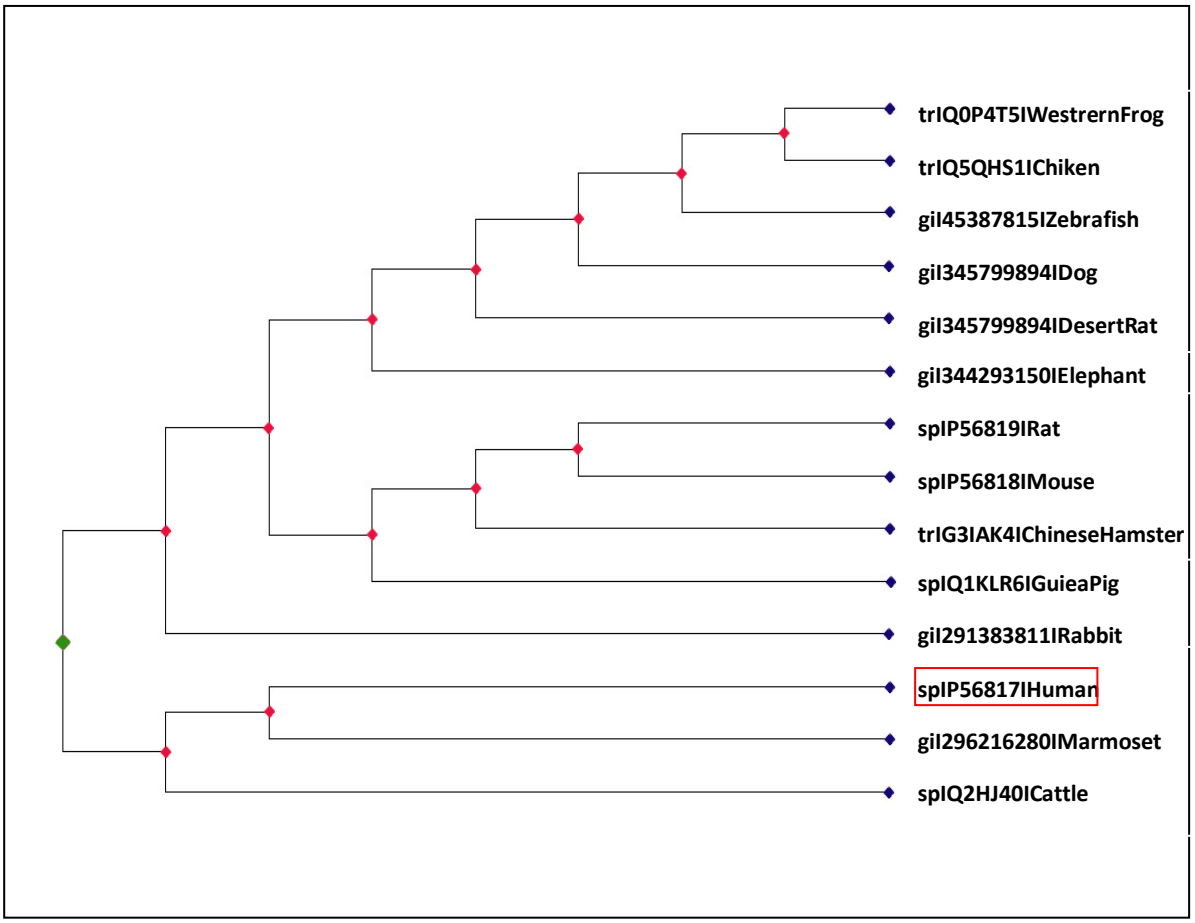

Figure 4.1 Phylogenetic tree of BACE1's sequences, the red box indicate the query sequence.

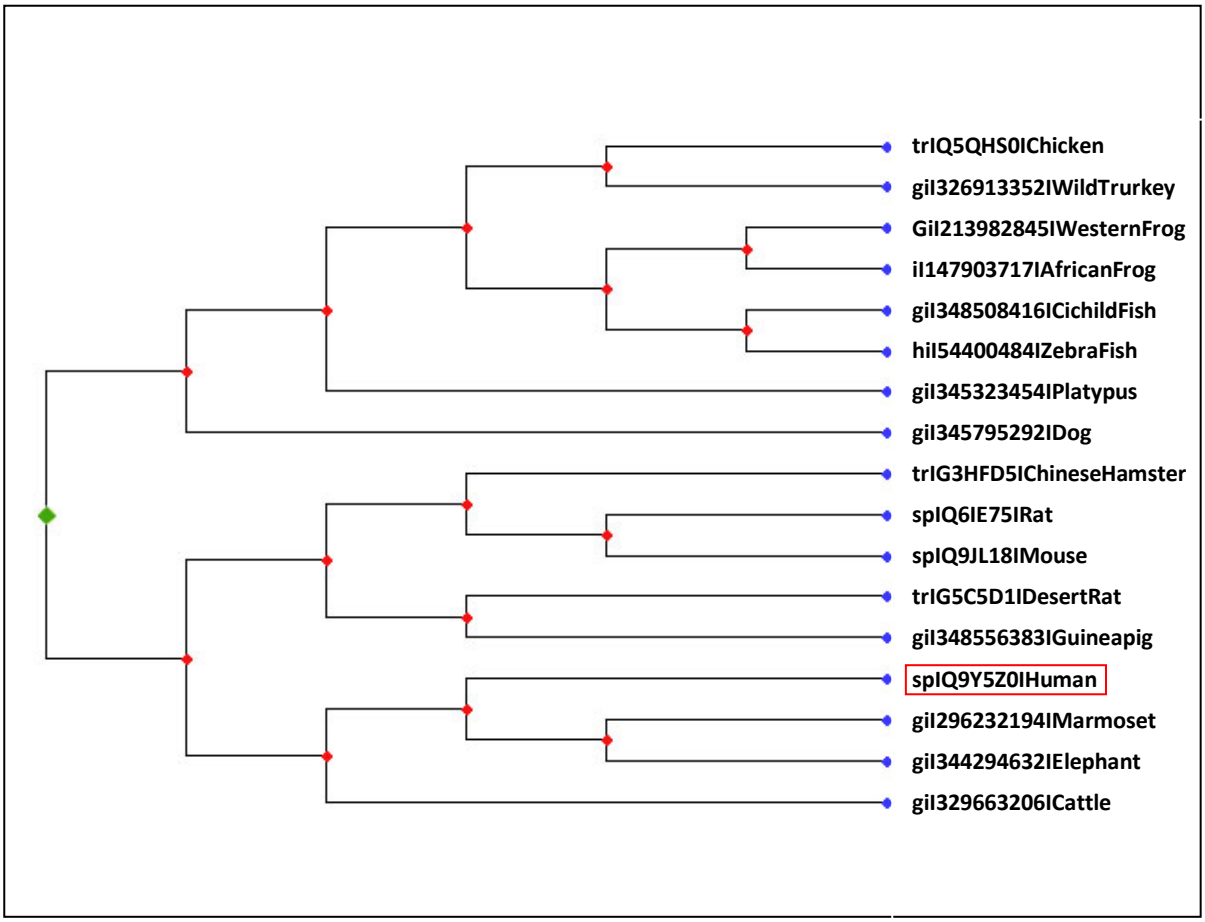

Figure 4.2 Phylogenetic tree of BACE2's sequences, the red box indicate the query sequence.

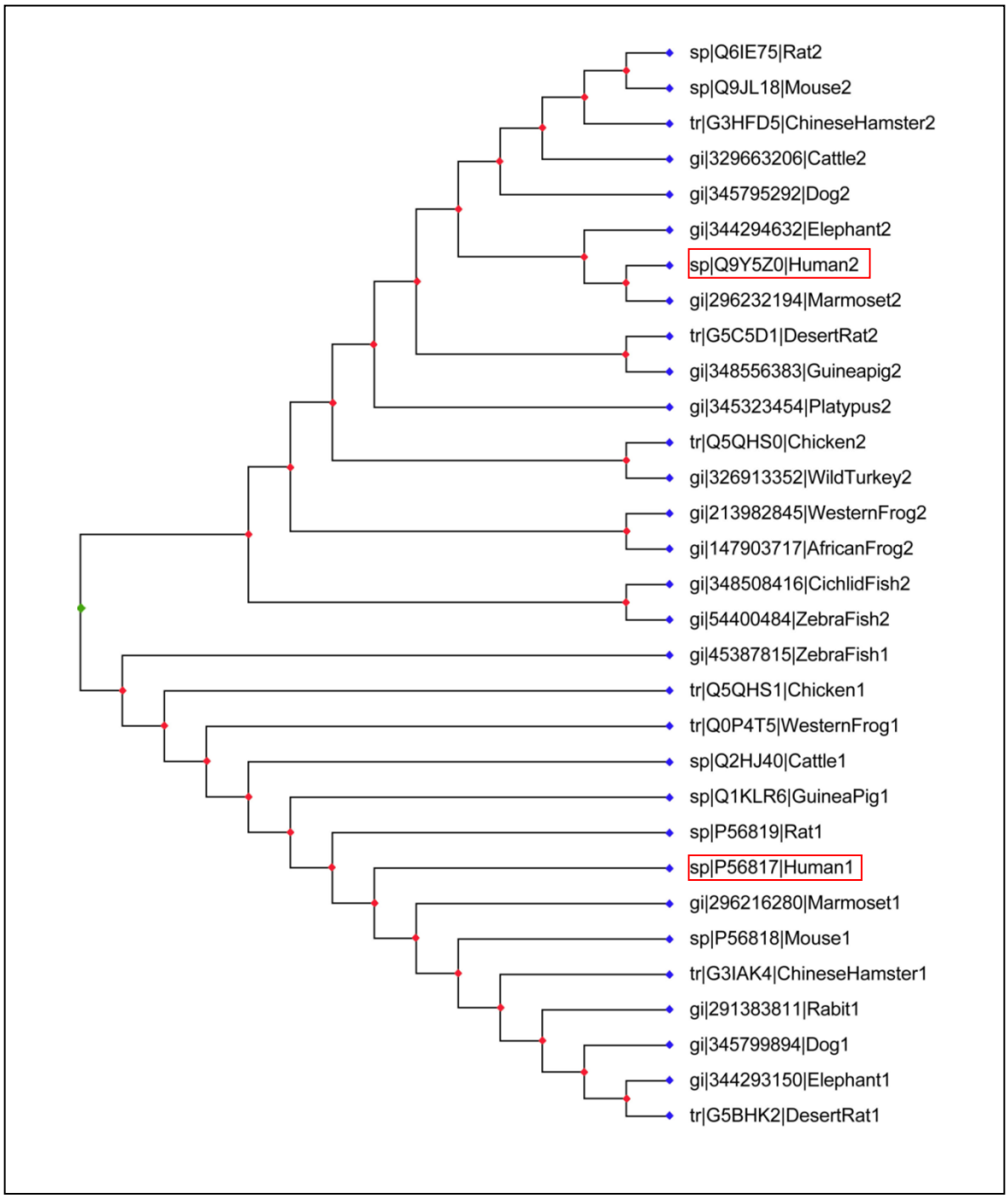

Figure 4.3 Phylogenetic tree of BACE1 & BACE2's sequences, the red box indicate the query sequences.

4.5 Molecular Evolution Analysis

In this step I compared the alignment's result of BACE1, BACE2, and BACE1 & BACE2 together based on their phylogenetic trees to get the consensus sequence. After the phylogenetic tree had been constructed, the tree had been divided to different groups by vertically cut. Then in the alignment file the sequences are separated based on groups. We compared the sequences with each other within each groups to find the conservation residues. The conserved residues within the sequences are labeled by amino acid short name while the conservation residues presented by X in each group, and the "-" represents for any conserved residue, where it may only present as a non-functional residue or a structural site to support the shape of the protein.

4.5.1 BACE1 sequence analysis

The Partitions P1 and P2 were performed to separate the groups of BACE1 species into subgroups for further analysis (Figure 4.4).

4.5.1.1 Partition P1

In the first partition, P_1, the tree was cleaved into two major groups, Group1 (G1) and Group2 (G2) (Figure 4.5). Group1 contains 11 sequences while Group2 only has 3 members. Based on the p1 partition in the alignment file the sequences were separated in to two groups to perform the ET analysis (Supplement Figure 7). All the p1 partition's trace sequences of BACE1 are displayed as Figure 4.6.

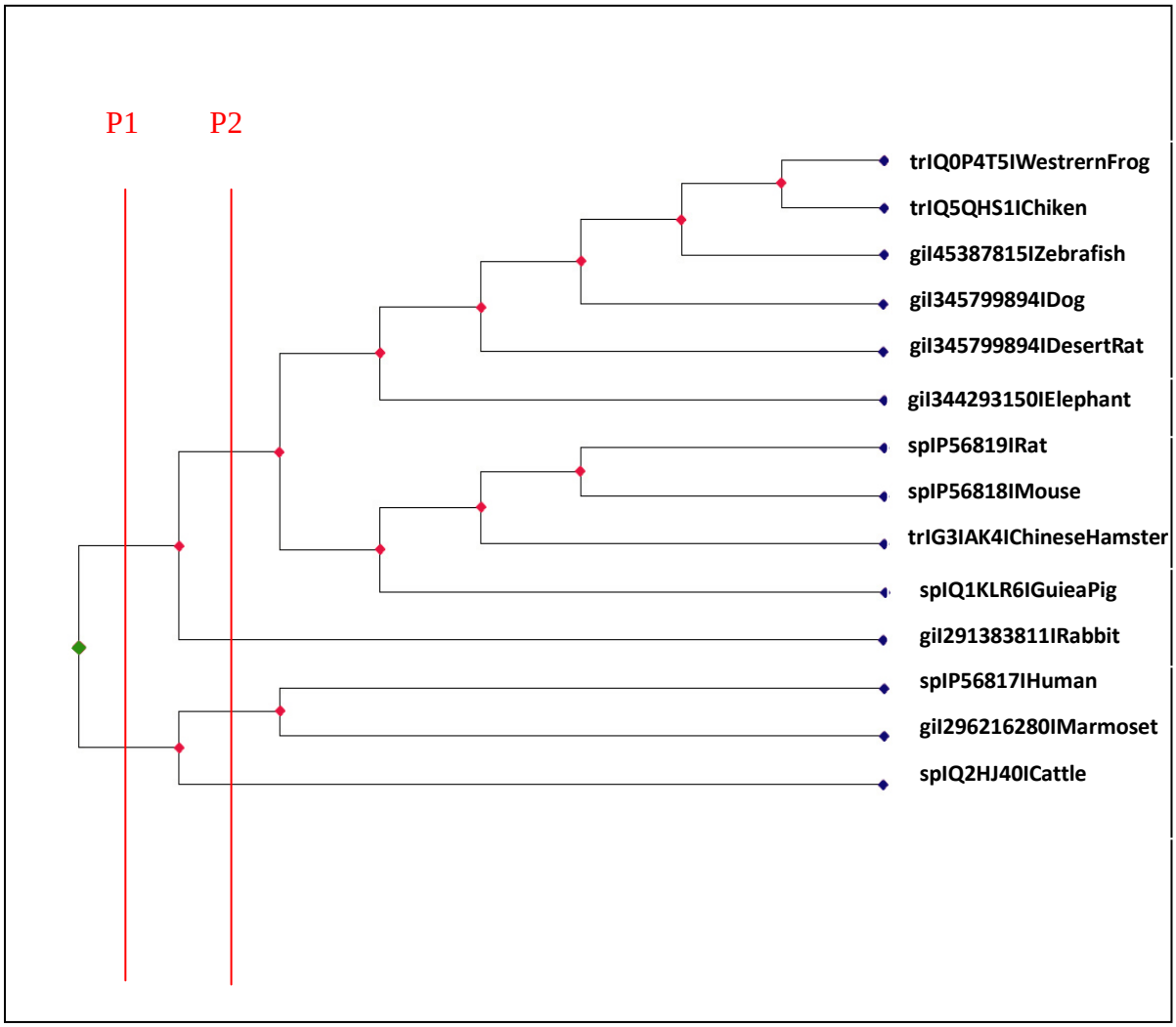

Figure 4.4 P1 and P2 partition of BACE1 phylogenetic tree.

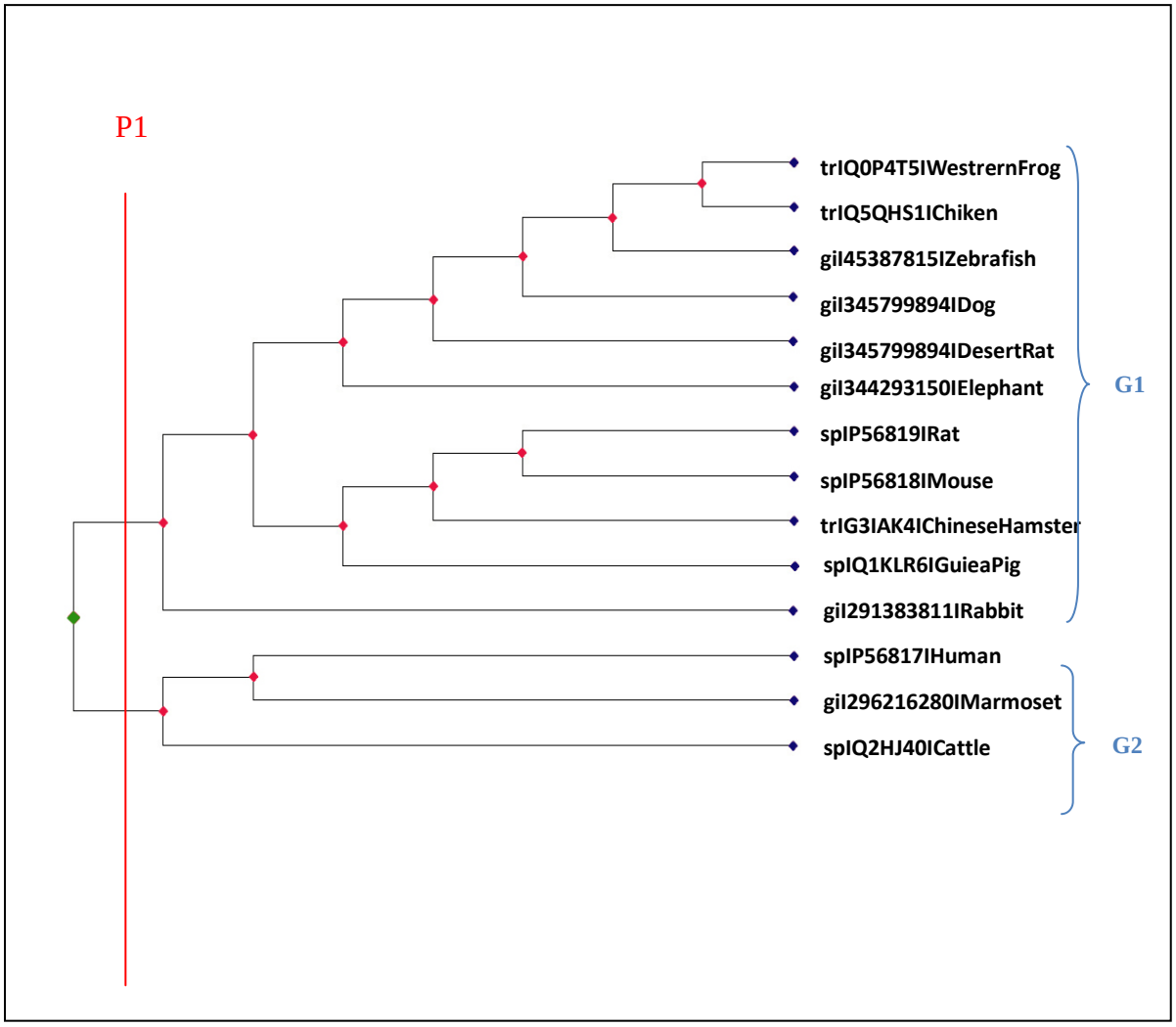

Figure 4.5 P1 partition of BACE1 phylogenetic tree.

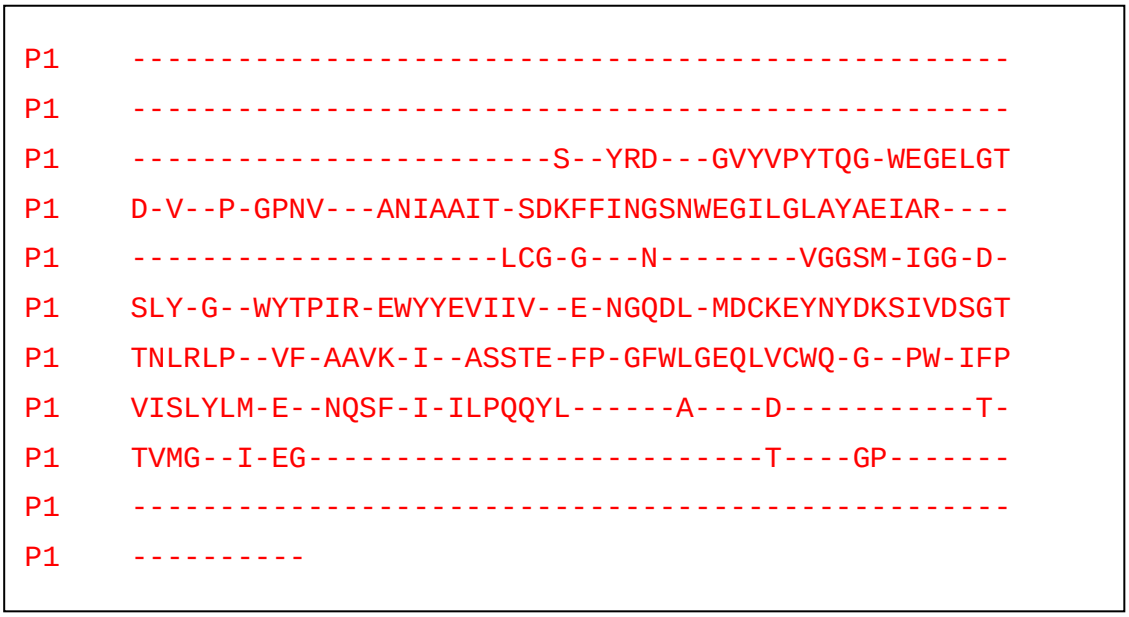

Figure 4.6 P1 partition' trace residues of BACE1 phylogenetic tree

4.5.1.2 Partition P2

After the first partition was done, the same step of grouping will continually be carried on in the second partition P_2 to extracting the trace residues (Figure 4.7).

The Second partition, P2 involves two groups .Group1 contains 10 spices and Group 2 involves human and marmoset. Two spices sequences which are rabbit and cattle were ignored in grouping and extracting. The reason is that, there are no more branches diverge for them. The result of sequences separating in the alignment file based on these two groups is displayed as Supplement Figure 8. All the p2 partition's trace sequences of BACE1 are displayed as Figure 4.8

4.5.1.3 Trace Residues Analysis of BACE1

The trace residues that derived from P1and P2 partition were compared with each other as Figure 4.9**.** As all the trace residues which were derived from P1 and P2 partition are the same with each other, I used the P1 trace sequences to compare with the protein mapping to identify conserved residues and active site. After choosing P1 partition, the position of each trace residues from P1 partition highlighted in human BACE1 a sequence which was derived from BACE1 alignment result and PDB structure sequences (Figure 4.10).

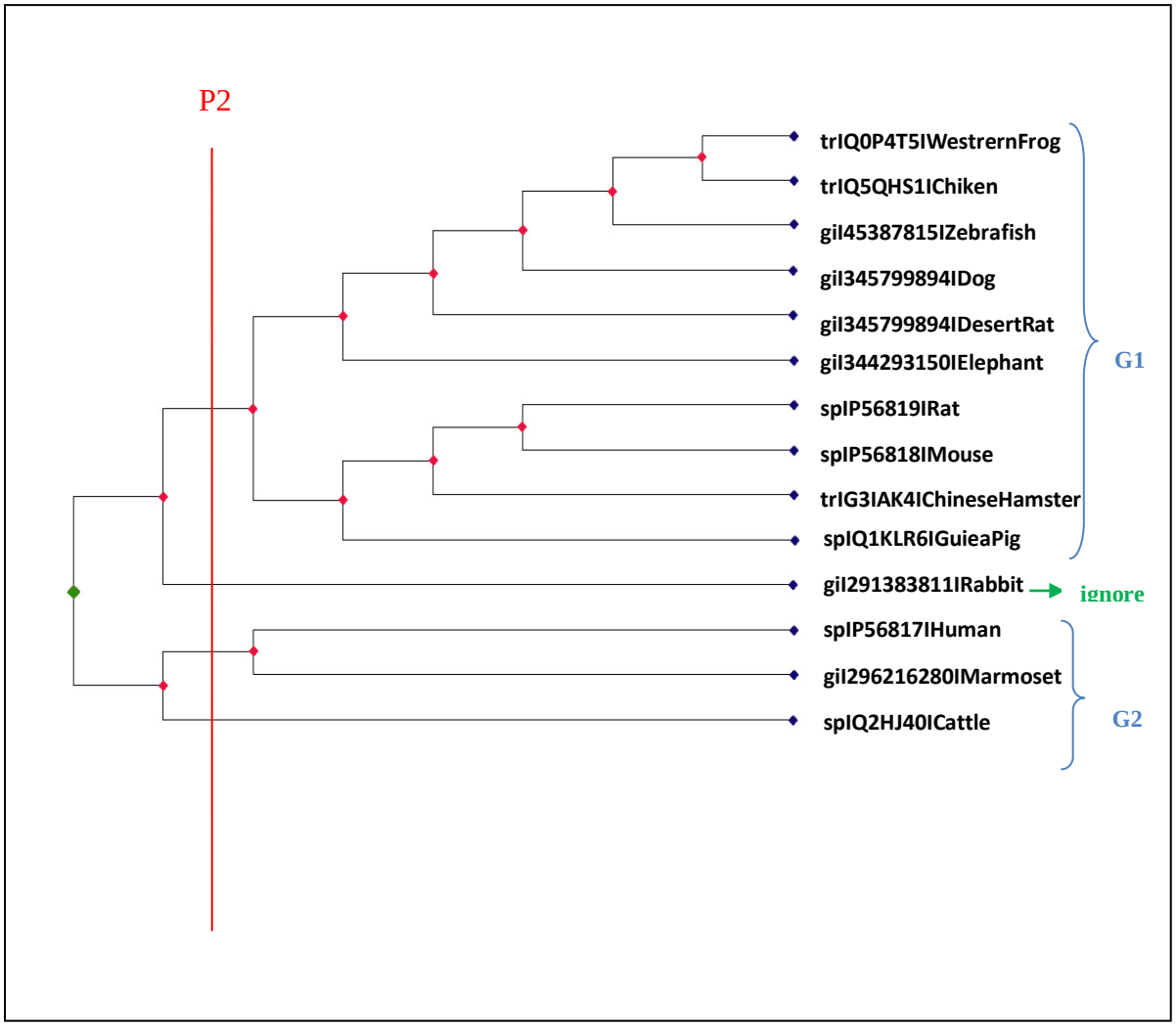

Figure 4.7 P2 partition of BACE1 phylogenetic tree.

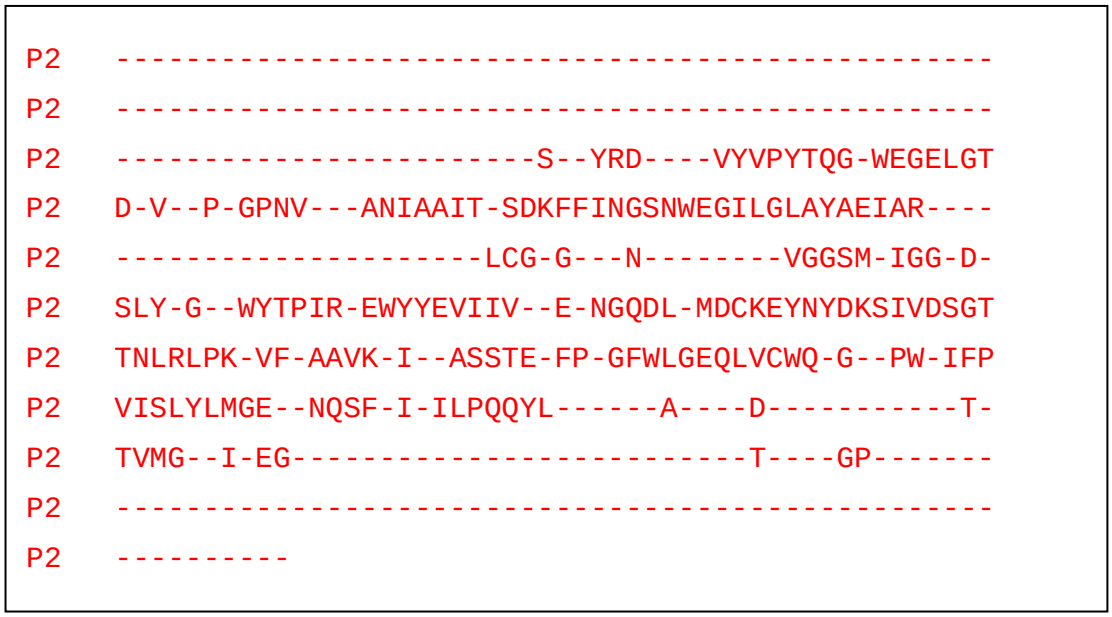

Figure 4.8 P1 partition' trace residues of BACE1 phylogenetic tree

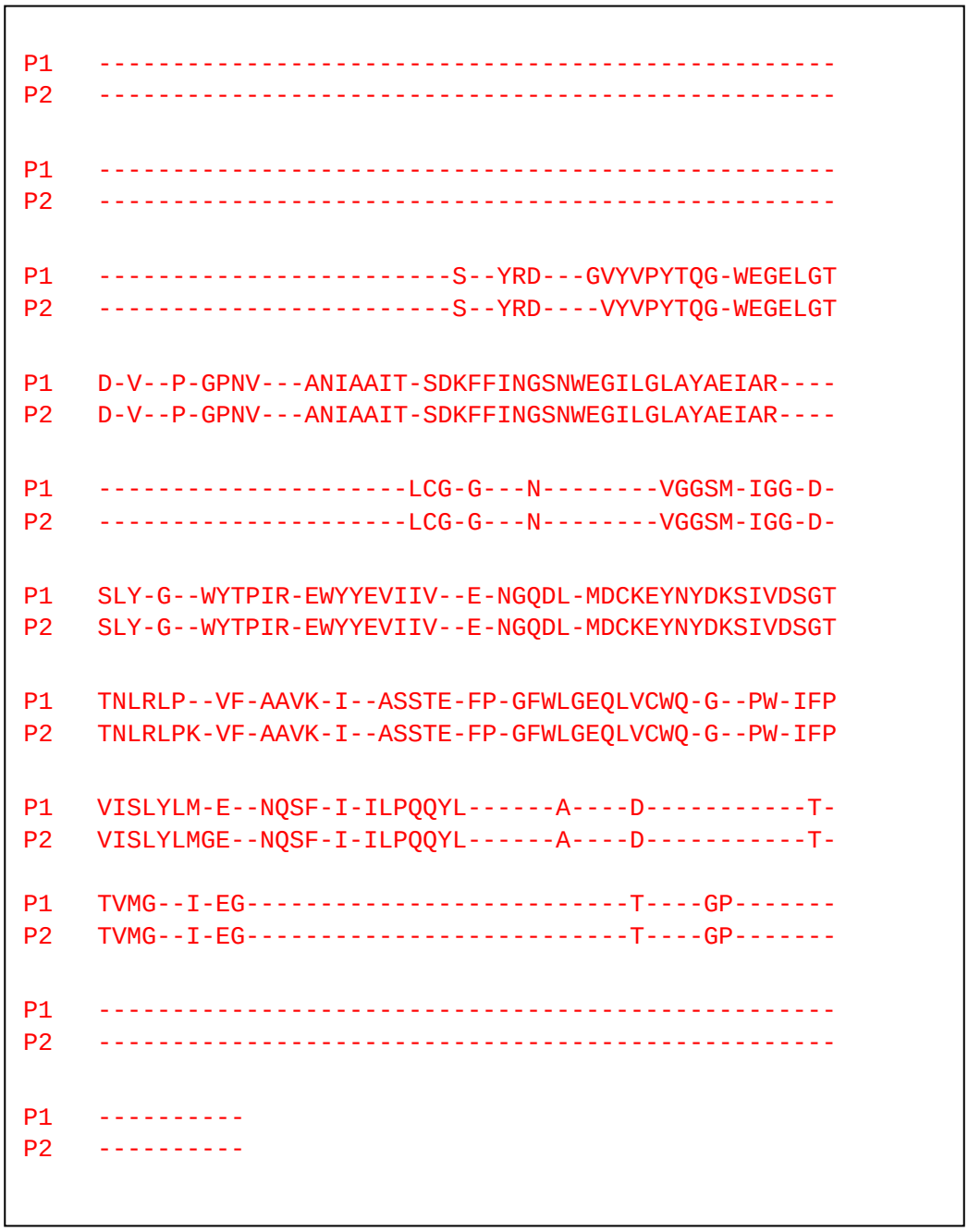

Figure 4.9 P1 and P2 partition**'** trace residues of BACE1 phylogenetic tree

Figure 4.10 Position of each trace residues from P1 partition in human BACE1 sequence and PDB structure.

4.5.1.4 Protein structure of BACE1

As mentioned before, the BACE1 was used in this project is 1FKN which was derived from PDB databases. This structure has two chains, A and B. As both chains are the same, for analyzing the structure easily I focused on chain A.

To display only chain A through Rasmol program, the structure data derived from PDB file had to be filtered before. The part that contains atoms and ligand (written as HETAM in PDB file) only within the chain A had been cut out, and then input into Rasmol program for viewing. The 3-dimensional structure of chain A and its ligand is displayed in Figure 4.11. In this figure, the left side shows the structure of chain A. Chain A consists of α helix and β sheet which are colored by red and yellow in this Figure. Also there are different loops with grey color which made the connection between those α helix and β sheet. The right sight of the figure presents the ligand which is the polypeptide and consists of 7 residues. It is displayed in form of "sticks" and grey color.

By using Rasmol program, chain A and its ligand are examined different views or displays by rotation. The structure had been rotated by 0^{o}, 90^{o}, 180^{o}, and 270^{o} about the y-axis. The protein had been displayed in form of "spacefill" and grey color while the ligand had been displayed in form of " sticks" with red color (Figure 4.12).

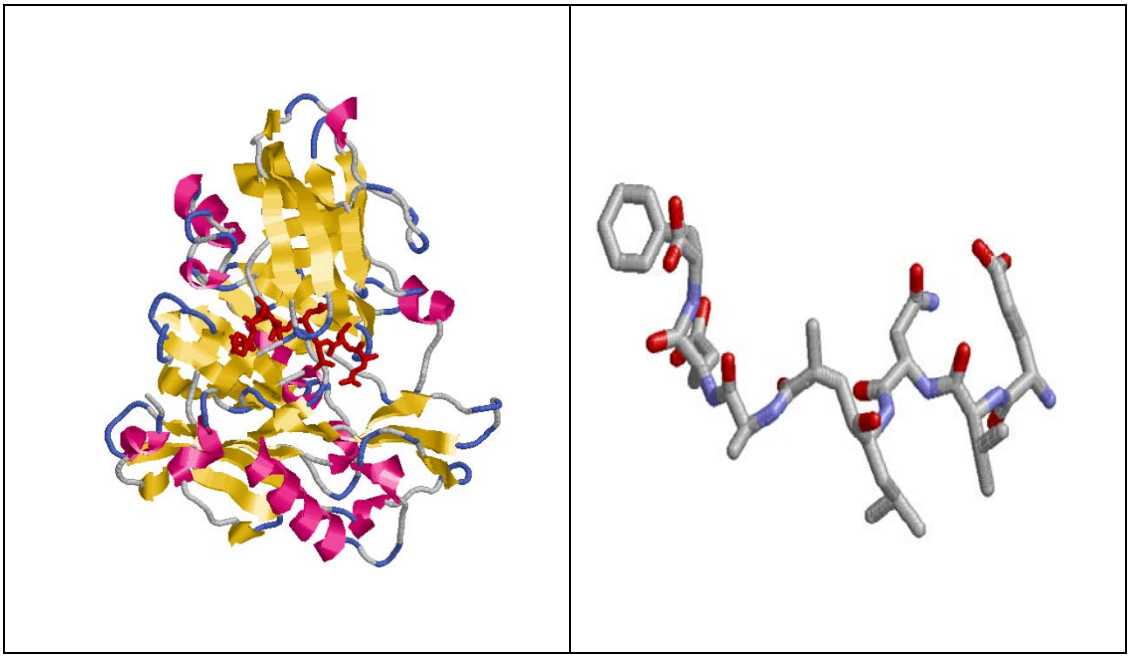

Figure 4.11 The 3-dimensional structure of BACE1's chain A and its ligand.

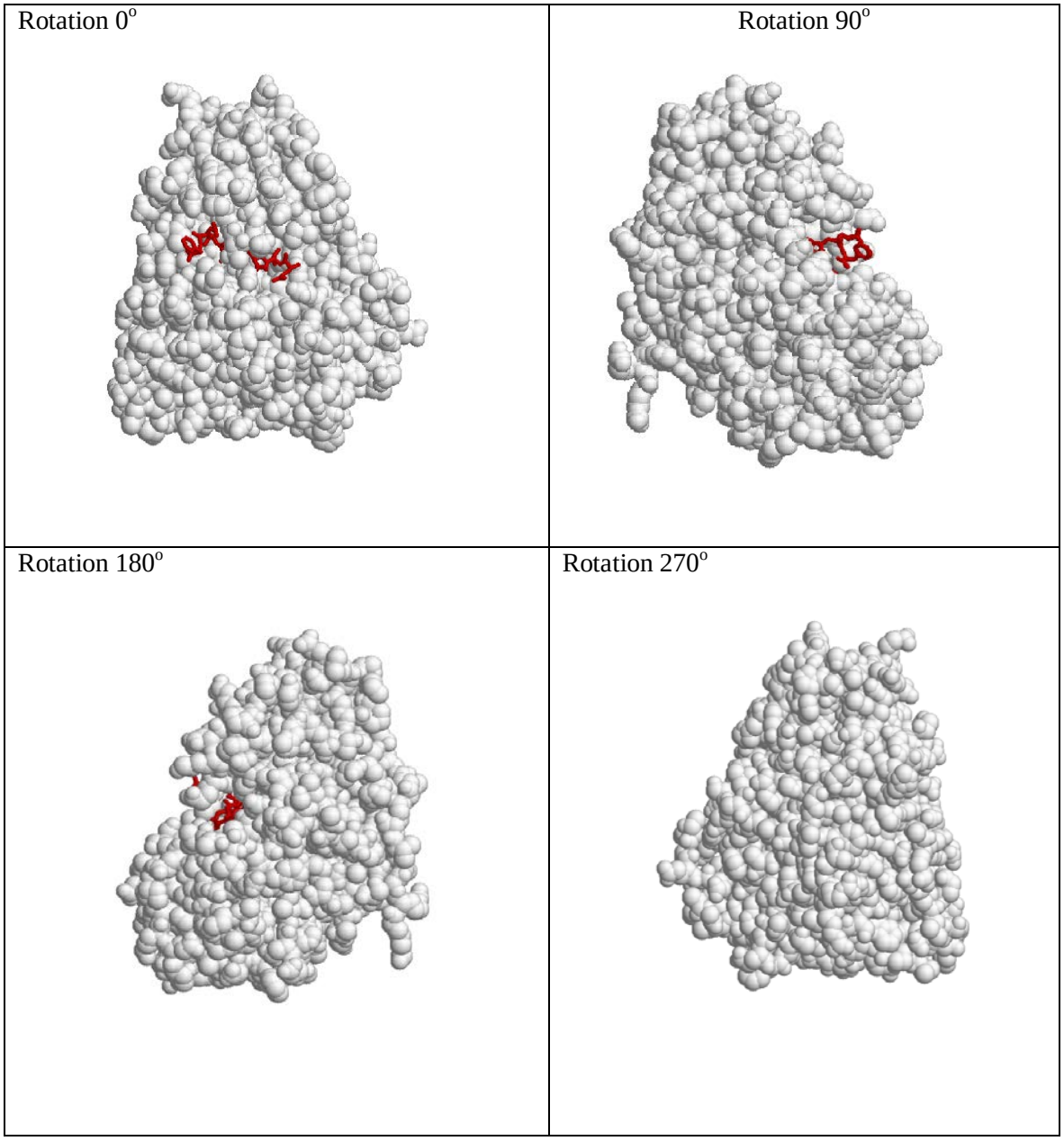

Figure 4.12 The display of the BACE1's chain A with its ligand from different view direction.

4.5.1.5 Trace residues mapping on BACE1 Protein structure

Protein mapping is the next step after obtained the trace residues from the analysis in molecular evolutionary, to predict the active site. First of all, the P1 trace residues were combined with chain A of the BACE1 structure (Figure 4.13). Then, the conserved residues of P1 and their position on the chain A were identified and listed in Table 4.3.

4.5.1.6 Ligand binding site of BACE1 protein structure

The 1FKN is binding to a ligand which is an inhibitor polypeptide. The ligand binding site was selected within the chain A of BACE1 structure by the proximity distance of the amino acid residues within 5.0 Armstrong to the ligand as the ligand binding site (Figure 4.14).

All the atoms of the amino acid residues within 5.0 Armstrong of the ligand were presented in Figure 4.15. By comparing the trace residues with the ligand binding site, conserved ligand binding residues were identified (Table 4.4). There are 8 residues in the active site of BACE1 which are not conserved in 3 species based on ET analysis. The reason is that their sequences length is short and the first part of their sequences only containing the gaps in multiple alignment result. Those residues are present in the table by '*' symbol.

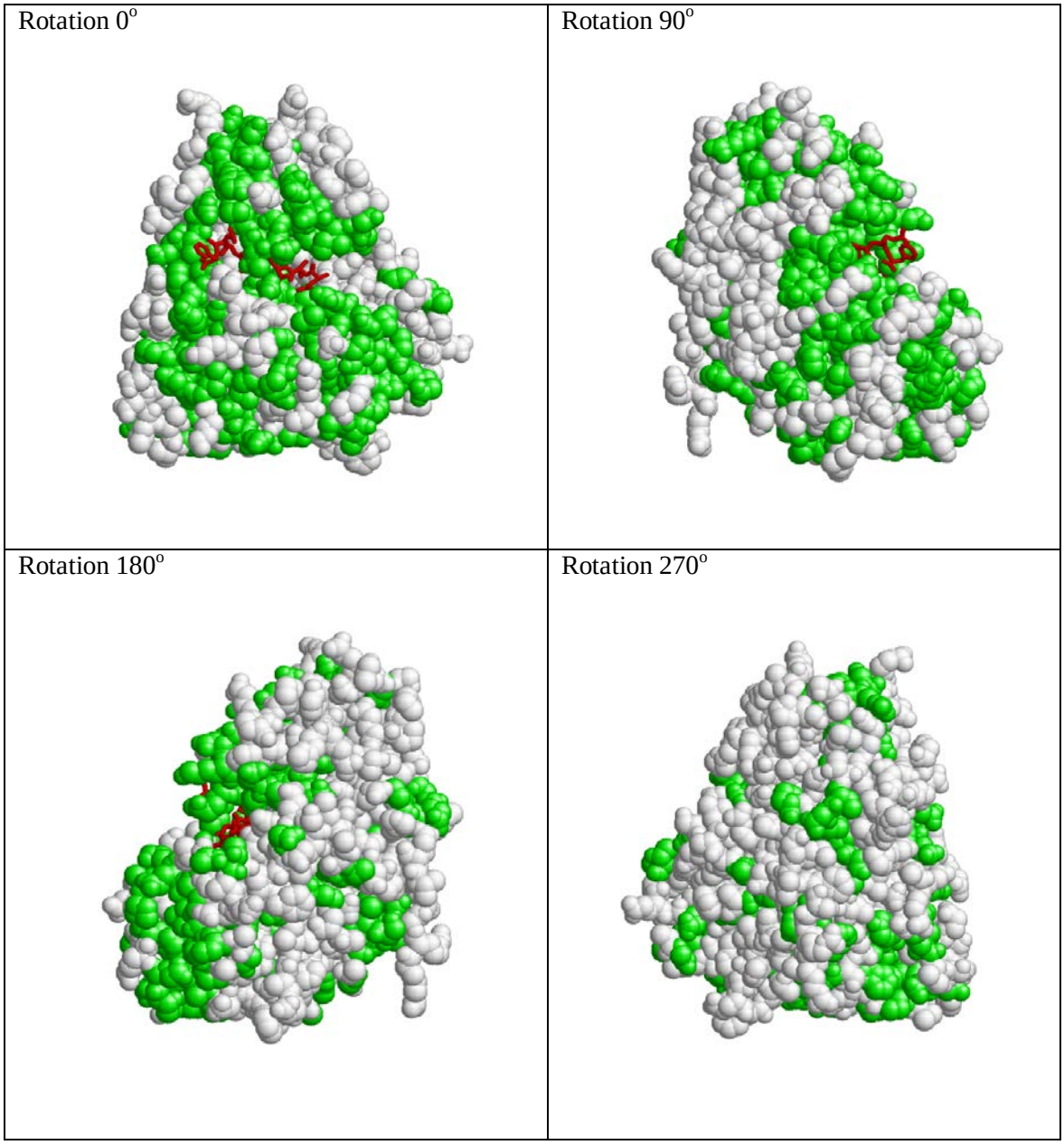

Figure 4.13 The 3-dimensional structure of BACE1's chain A with conserved residues from different view direction.

Table 4.3 The conserved residues and position in the structure of 1FKN protein.

Ser57	Tyr60	Arg61	Asp62	Gly66
Val67	Tyr68	Val69	Pro70	Tyr71
Thr72	Gln73	Gly74	Trp76	Glu77
Gly78	Glu79	Leu80	Gly81	Thr82
Asp83	Val85	Pro88	Gly90	Pro91
Asn92	Val93	Ala97	Asn98	Ile99
Ala100	Ala101	Ile102	Thr103	Ser105
Asp106	Lys107	Phe108	Phe109	Ile110
Asn111	Gly112	Ser113	Asn114	Trp115
Glu116	Gly117	Ile118	Leu119	Gly120
Ile121	Ala122	Tyr123	Ala124	Glu125
Ile126	Ala127	Arg128	Leu154	Cys155
Gly156	Gly158	Asn162	Val170	Gly171
Gly172	Ser173	Met174	Ile176	Gly177
Gly178	Asp180	His181	Ser182	Lue183
Tyr184	Gly186	Trp189	Tyr190	Thr191
Pro192	Ile193	Arg194	Glu196	Trp197
Tyr198	Tyr199	Glu200	Val201	Ile202
Ile203	Val204	Glu207	Asn209	Gly210

Table 4.3 The conserved residues and position in the structure of 1FKN protein, continued.

Gln211	Asp212	Tyr220	Met215	Asp216
Cys217	Lys218	Glu219	Tyr220	Asn221
Tyr222	Asp233	Lys224	Ser225	Ile226
Val227	Asp228	Ser229	Gly230	Tyr231
Tyr232	Asn233	Leu234	Arg235	Leu236
Pro237	Val240	Phe241	Ala243	Ala244
Val245	Lys246	Ile248	Ala251	Ser252
Ser253	Thr254	Glu255	Phe257	Pro258
Gly260	Phe261	Trp262	Leu263	Gly264
Glu265	Gln266	Leu267	Val268	Cys269
Trp270	Gln271	Gly273	Pro276	Trp279
Phe280	Pro281	Val282	Ile283	Ser284
Leu285	Tyr286	Leu287	Met288	Glu290
Asn293	Gln294	Ser295	Phe296	Ile298
Ile300	Leu301	Pro302	Gln303	Gln304
Tyr305	Leu306	Ala313	Asp318	Thr330
Thr332	Val333	Met334	Gly335	Thr338
Glu340	Gly341	Thr368	Gly373	Pro374

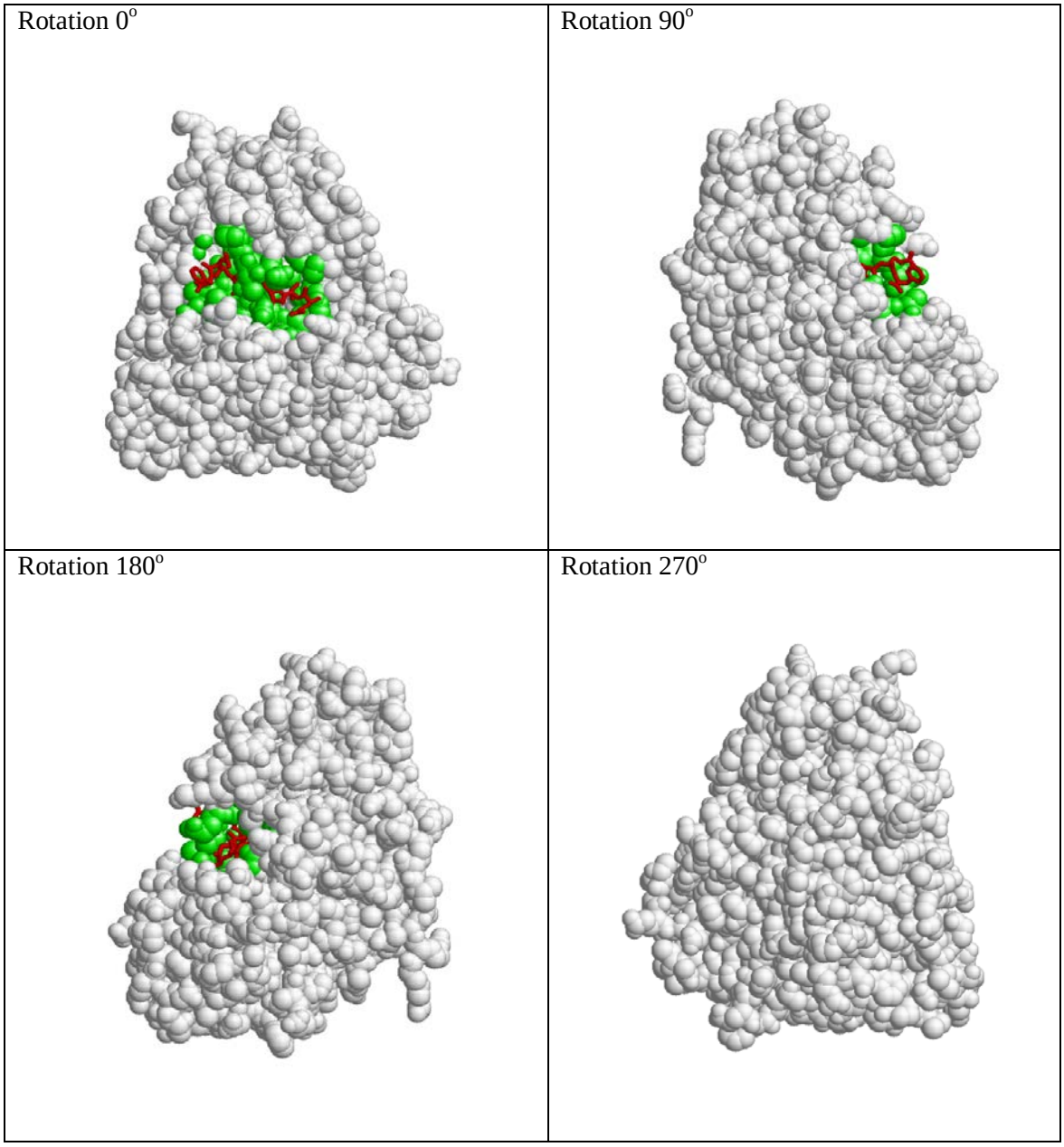

Figure 4.14 The 3-dimensional structure of ligand and trace residues within chain A of BACE1 which are selected within 5.0 Armstrong from different view direction.

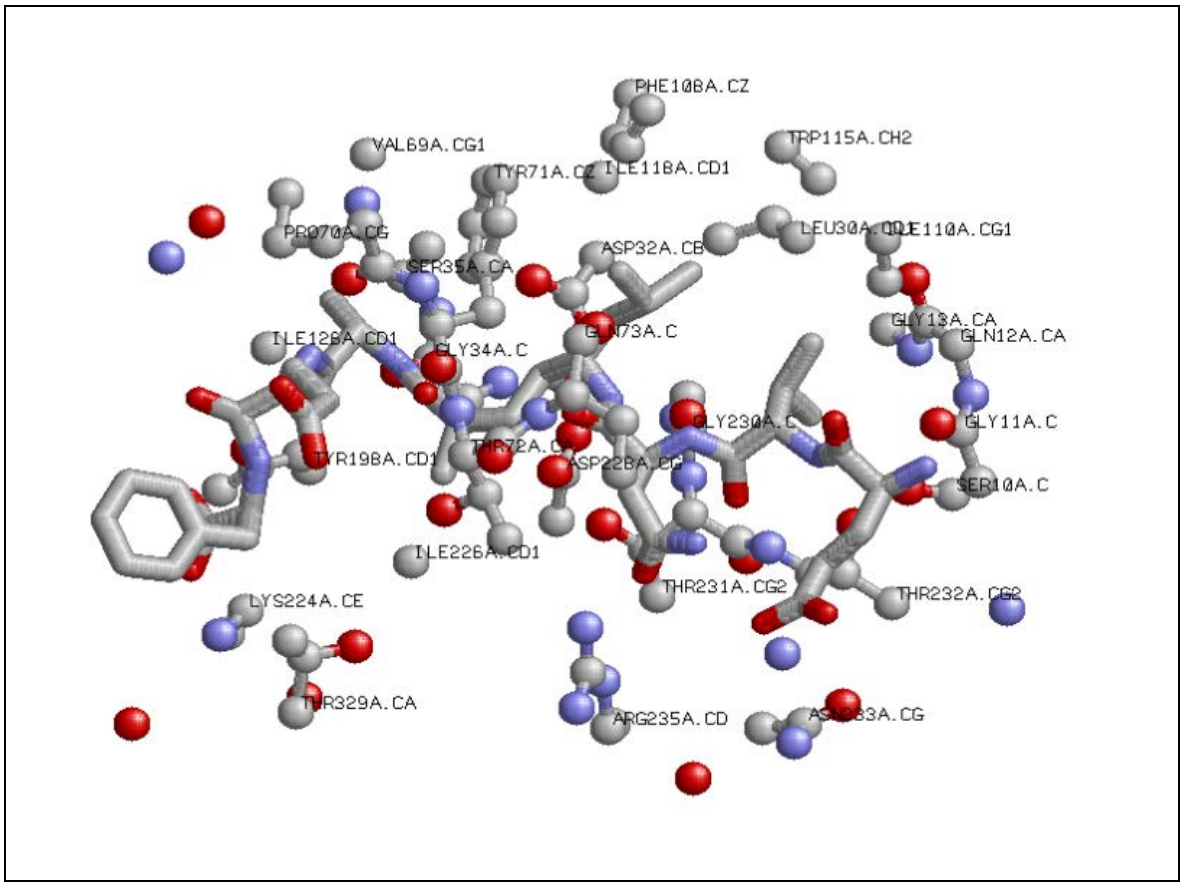

Figure 4.15 The display of ligand and trace residues restricted within 5.0 Armstrong of BACE1's chain A. The ligand is displayed in form of "sticks" while the conserved residues are displayed in form of "Ball & Stick".

Table 4.4 The conserved residues positions in the structure which are predicted present in active site.

Ser10*	Gly11*	Gln12*	Gly13*	Leu30*	Asp32*	Gly34*
Ser35*	Val69	Pro70	Tyr71	Thr72	Gln73	Phe108
Ile110	Ile115	Ile118	Ile126	Tyr198	Lys224	Ile226
Asp228	Gly230	Thr231	Thr232	Asn233	Arg235	Thr329

4.5.2 BACE2 sequence analysis

As same as BACE1, the same step of grouping BACE2's sequences were carried on. Three Partitions, P1, P2 and P3 were performed to separate the groups of BACE2 species into subgroups for further analysis (Figure 4.16).

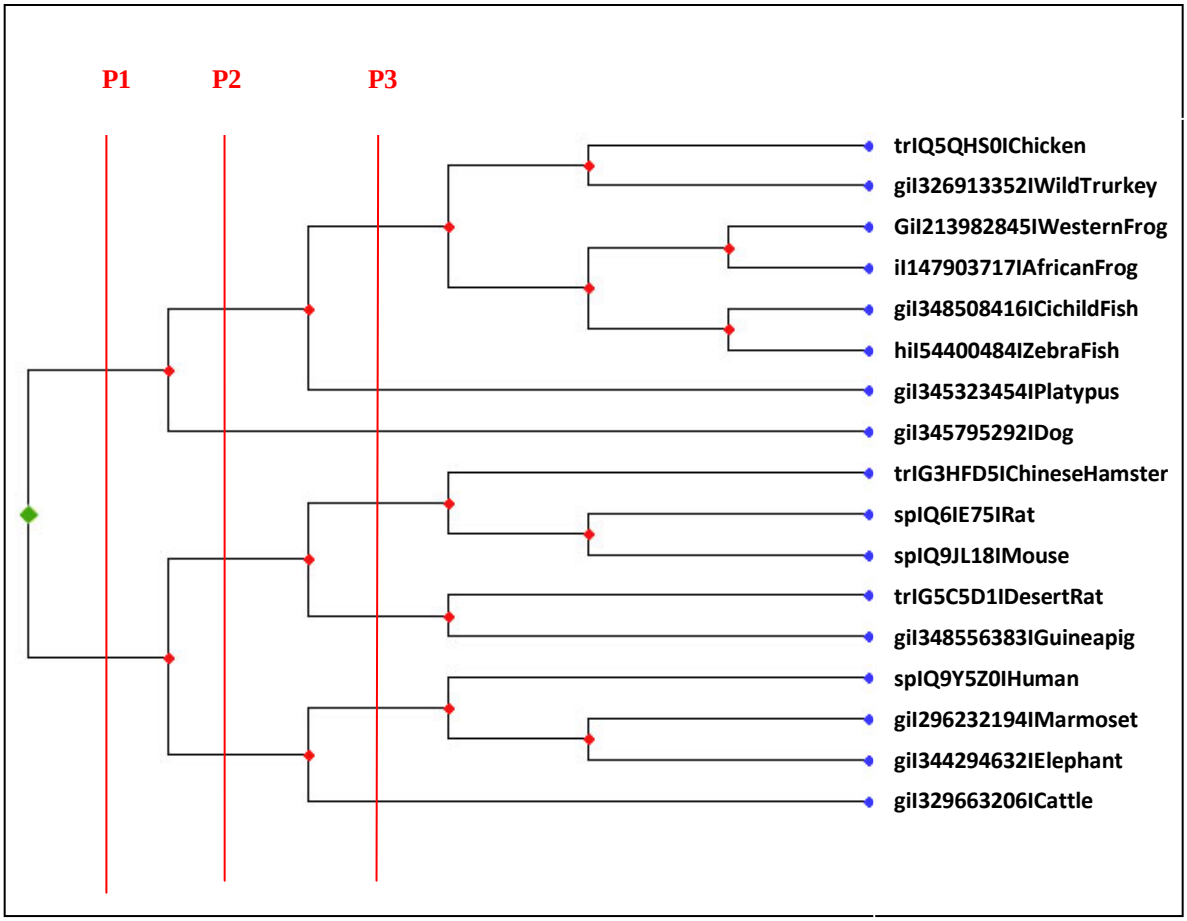

Figure 4.16 P1, P2 and P3 Partitions of BACE2's phylogenetic tree.

4.5.2.1 Partition P1

In the first stage, P1, the tree were vertically cut and the sequences were cleaved into two major groups, Group1 (G1) and Group2 (G2).There are 8 species in group1 and 9 species in group2 (Figure 4.17). Based on the P1 partition in the alignment file the sequences were separated in to two groups (Supplement Figure 9). All the P1 partition's trace residues of BACE2 are displayed as Figure 4.18.

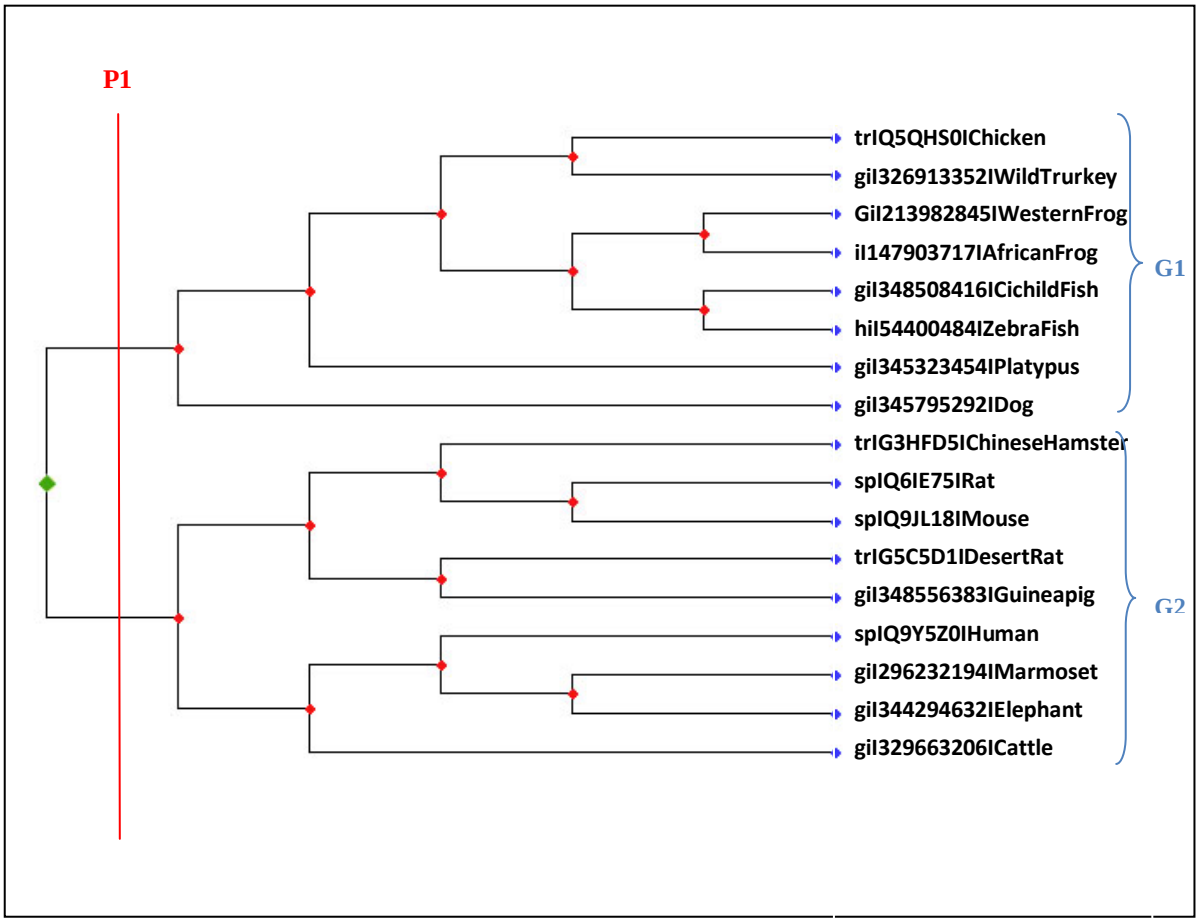

Figure 4.17 P1 partition of BACE2's phylogenetic tree.

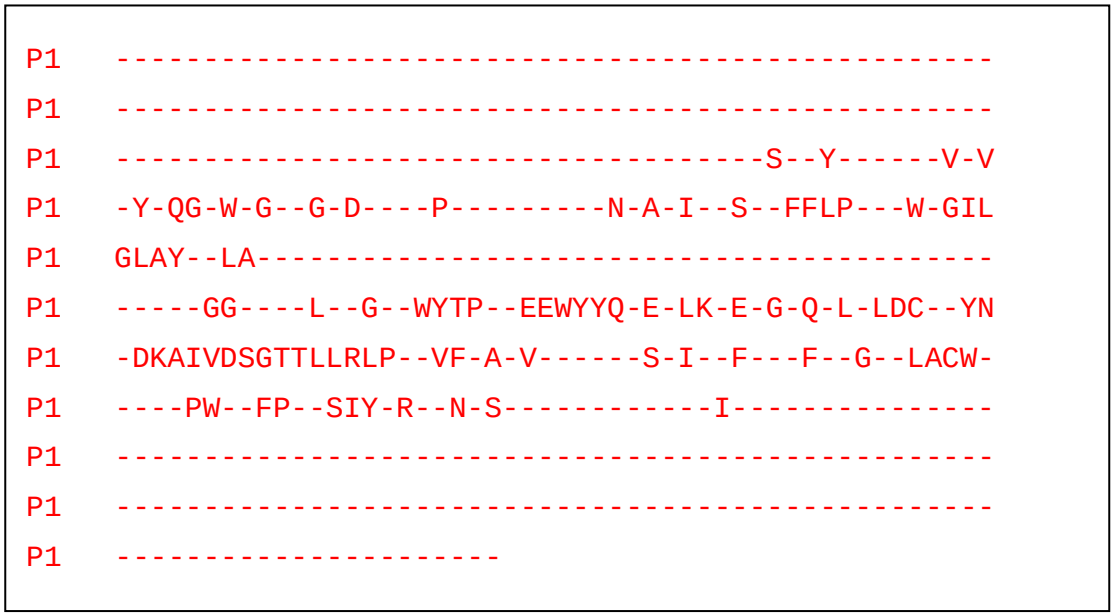

Figure 4.18 P1 partition's trace residues of BACE2 phylogenetic tree.

4.5.2.2 P2 partition

The Second grouping, P2 involves three groups and one species' sequence which is dag was ignored in grouping and extracting (Figure 4.19). As same as privies step in the alignment file, the sequences were separated based on these tree groups to perform the ET analysis (Supplement Figure 10). All the P2 partition's trace residues of BACE2 are displayed as Figure 4.20.

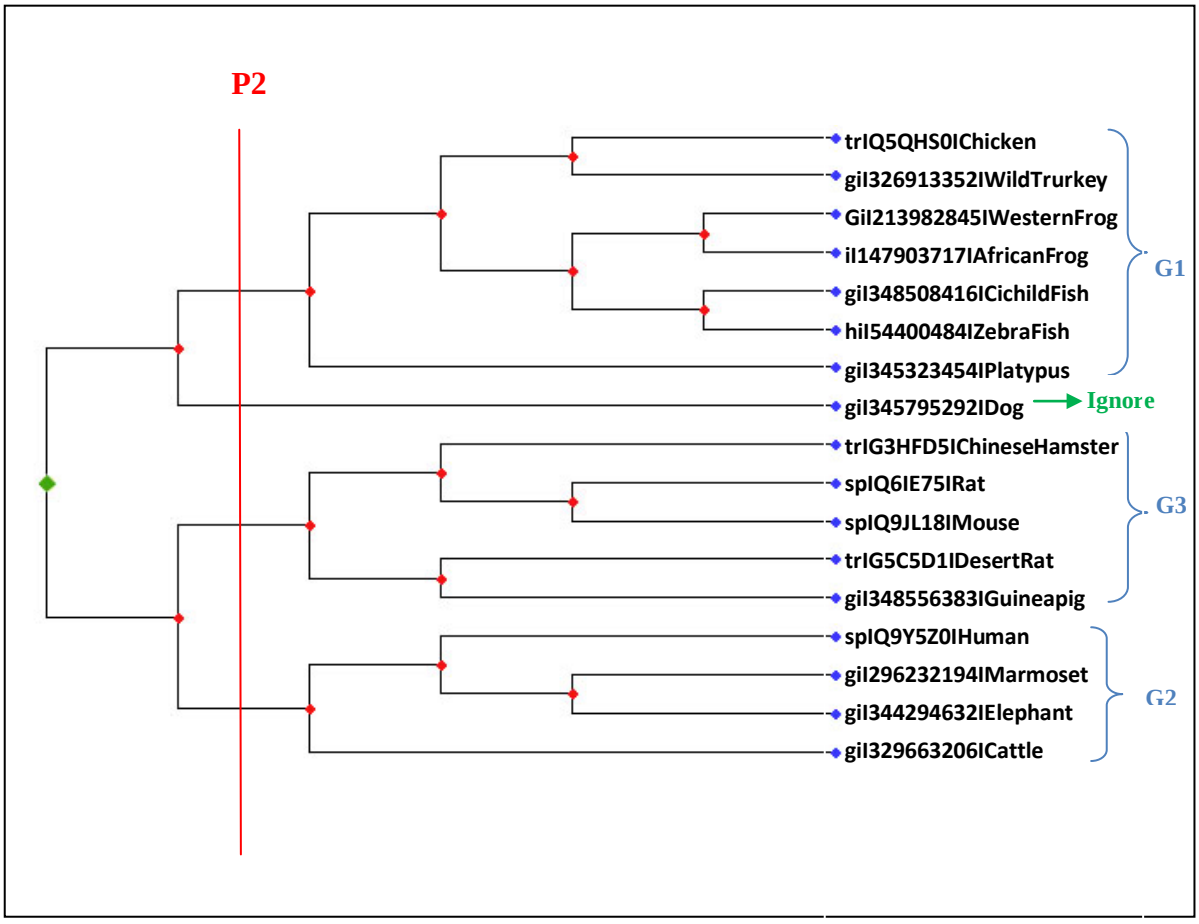

Figure 4.19 P2 partition of BACE2's phylogenetic tree.

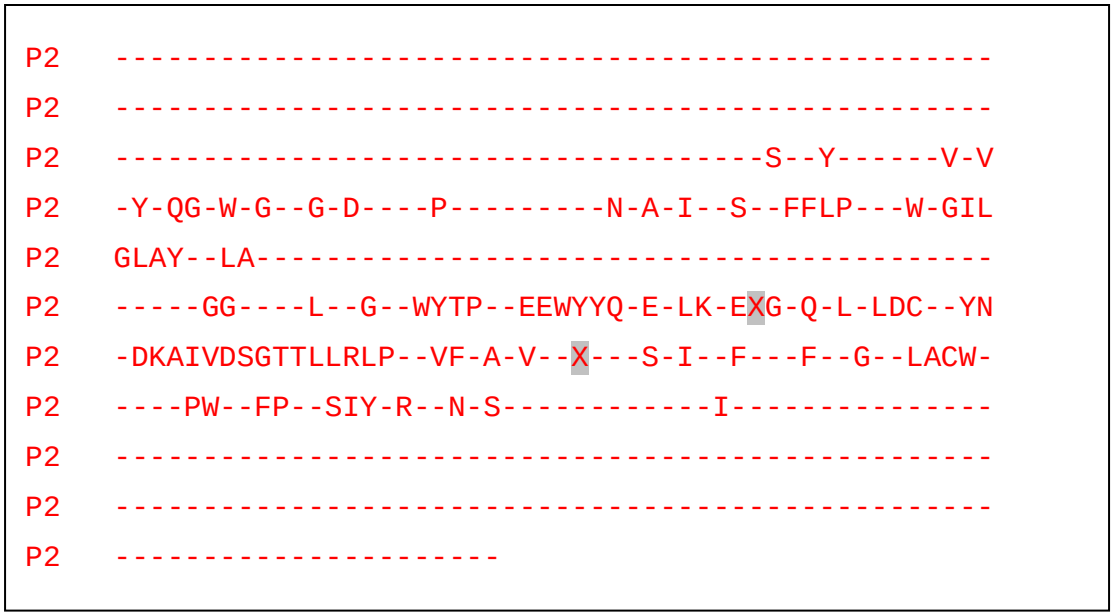

Figure 4.20 P2 partition's trace residues of BACE2 phylogenetic tree.

4.5.2.3 Partition P3

The third grouping, P3 involves four groups and tree species sequences which are platypus, dog, and cattle were ignored in this grouping (Figure 4.21).The results of separating sequences based on this partition and ET analysis were displayed as Supplement Figure 11. All the P3 partition's trace residues of BACE2 are displayed as Figure 4.22.

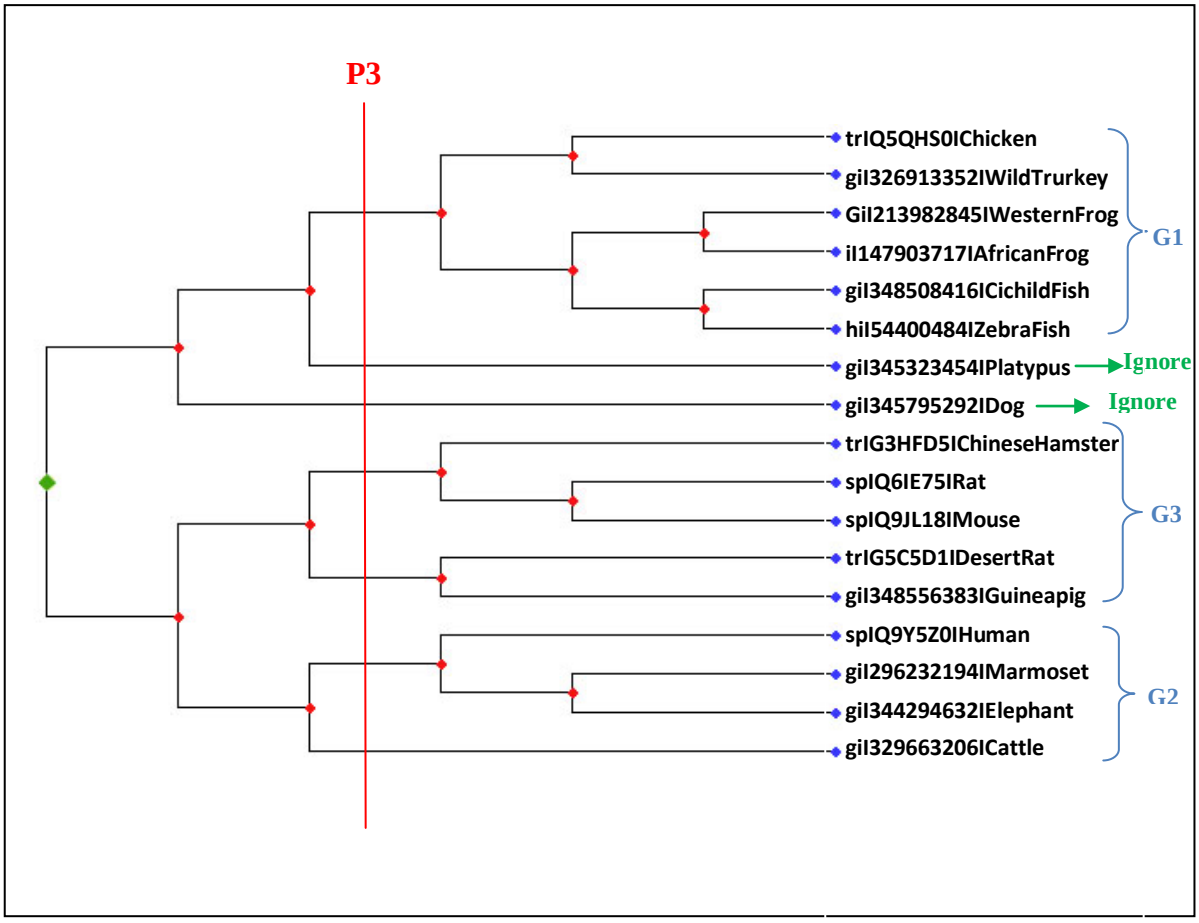

Figure 4.21 P3 partition of BACE2 phylogenetic tree.

```
P3   ------------------------------------------------
P3   ------------------------------------------------
P3   -----------------------------------S--Y------V-V
P3   -Y-QG-W-G-XG-D----P-X--X---XN-A-I--S--FFLP---W-GIL
P3   GLAY--LA----------------------------------------
P3   -----GG--P-L--G--WYTP—-EEWYYQXE-LK-EXG-Q-L-LDC--YN
P3   -DKAIVDSGTTLLRLP--VF-A-V--X---SLI--F---F--G--LACW-
P3   ----PW--FP--SIY-R--N-S------------I---------------
P3   ------------------------------------------------
P3   ------------------------------------------------
P3   ----------------------
```

Figure 4.22 P3 partition**'s** trace residues of BACE2 phylogenetic tree.

4.5.2.4 Trace Residues Analysis of BACE2

The trace residues that derived from P1, P2 and P3 partitions of BACE2 were compared with each other as Figure 4.23. I chose the P3 partition's trace residues to compare with the protein mapping and identify conserved residues and active site. The reason why I chose P3 is that the sequences within the group are more closely related each other and more conserved sites obtained. Also the number of group specific residues in this position is more than other partition.

After choosing the P3 partition, in the next step the position of each trace residues from P3 partition, highlighted in human BACE2 a sequence which was derived from BACE2 alignment result (Figure 4.24).

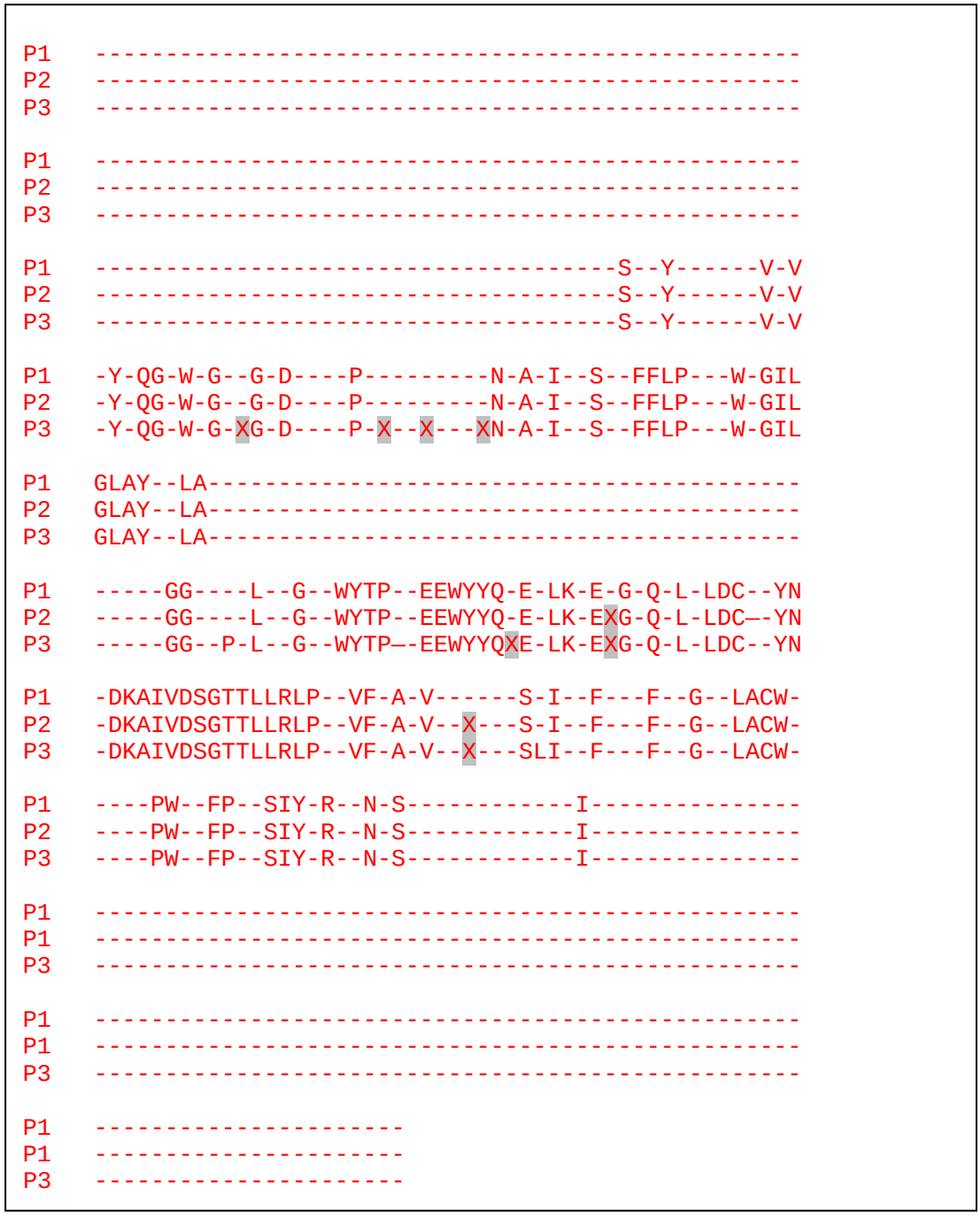

Figure 4.23 P1, P2 and P3 partition's trace residues of BACE2 phylogenetic tree

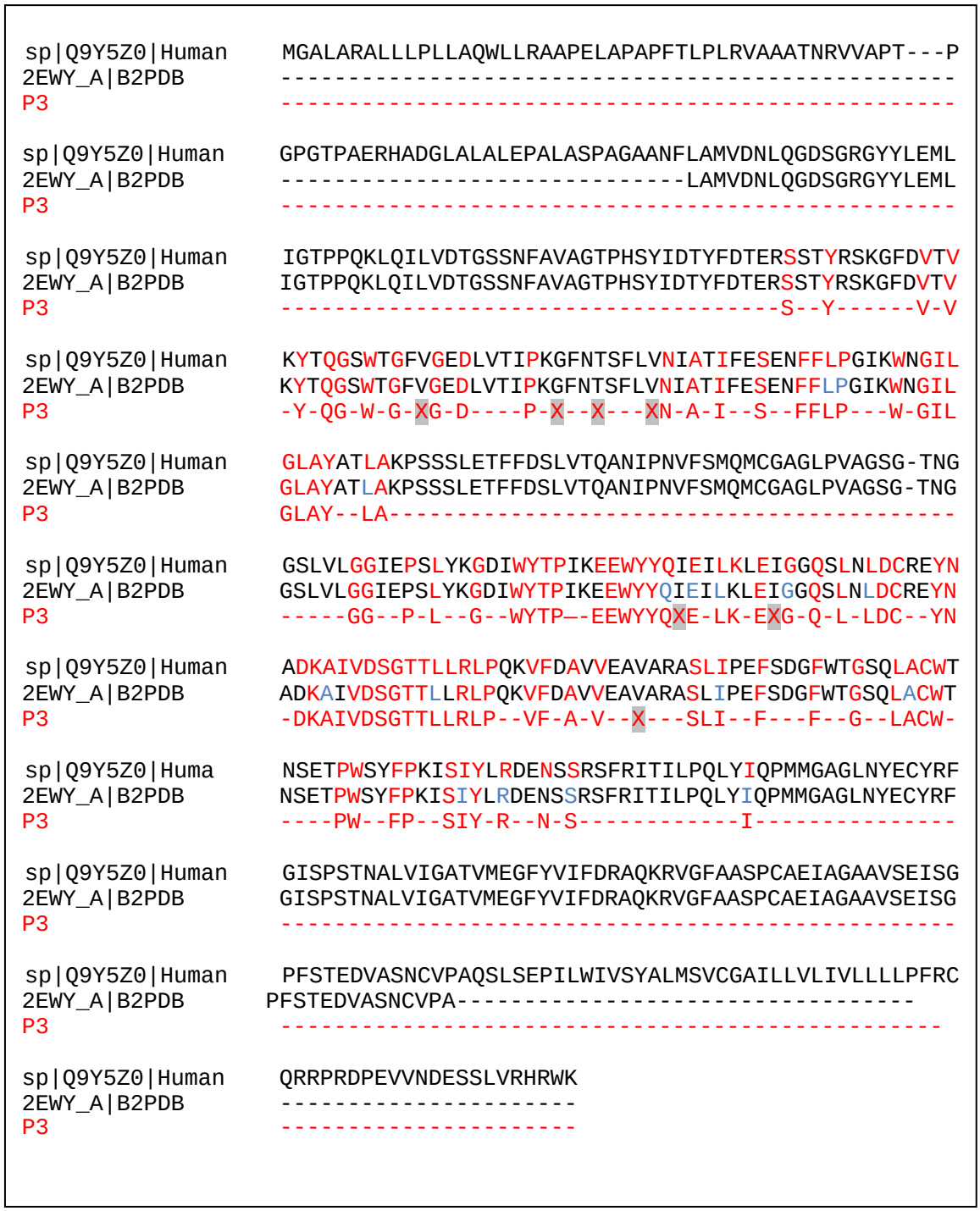

Figure 4.24 Position of each consensus sequences from P3 partition in human BACE2 sequences and PDB structure sequences.

4.5.2.4 Protein structure of BACE2

The BACE2's structure that was used in this project is 2EWY which was derived from PDB databases. This structure has four chains, A, B, C and D which are similar, and consists of same number of atoms. As all chains are the same, I only focused on chain A. To display the chain A only through Rasmol program, the same step as 1FKN were done and chain A was cut out, and then input into Rasmol program for viewing. The 3-dimensional structure of 2EWY's chain A with ligand display as Figure 4.25. Chain A consists of α-helix and β-sheet which were colored by red and yellow in this Figure. Also there are different loops with grey color which made the connection between those α-helix and β-sheet .The ligand displays in the left side of the figure in form of "sticks" and grey color.

By using Rasmol program, the structure of chain A it's ligand were rotated by 0^{o}, 90^{o}, 180^{o}, and 270^{o} about the y-axis. The protein was displayed in form of "spacefill" and grey color while the ligand was displayed in form of sticks with red color (Figure 4.26).

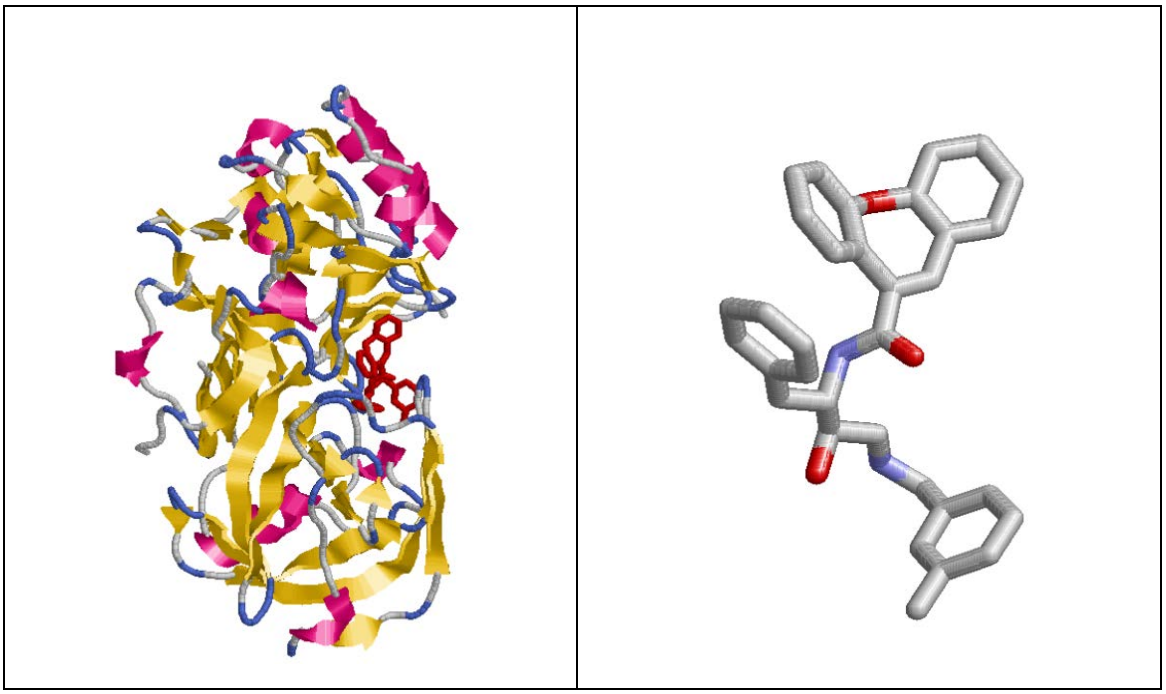

Figure 4.25 The 3-dimensional structure of BACE2's chain A and its ligand.

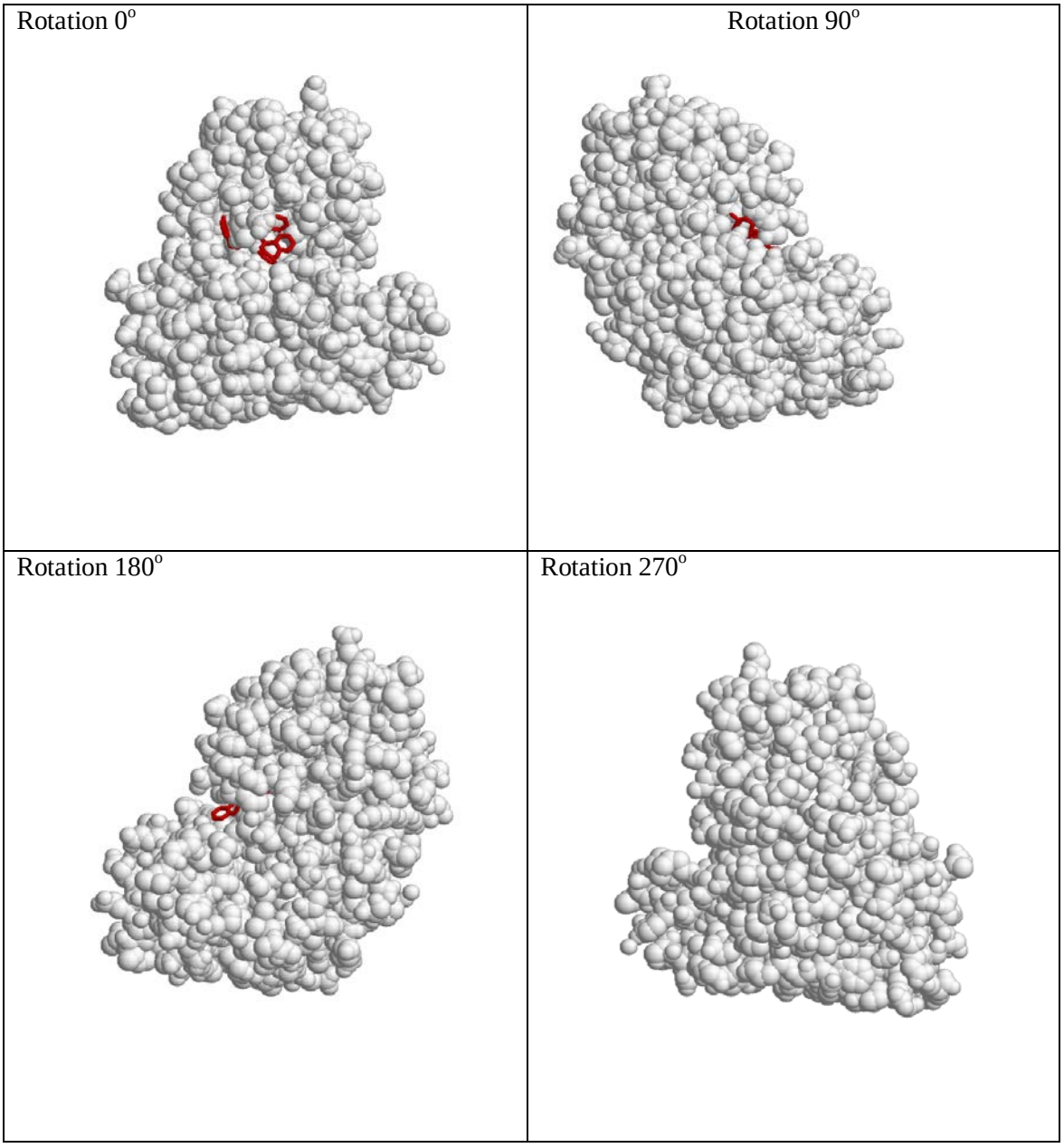

Figure 4.26 The display of the BACE2's chain A with its ligand from different view direction.

4.5.2.6 Trace residues mapping on BACE2 Protein structure

As same as BACE1's Protein mapping steps, at first, the P3 trace residues were combined with chain A of the BACE2 structure (Figure 4.27). Then, the conserved residues of P3 and their position on the chain A were identified and listed in Table 4.5.

4.5.1.6 Ligand binding site of BACE1 protein structure

The 2EWY is binding to a ligand which is an inhibitor polypeptide. The ligand binding site was selected within the chain A of BACE1 structure by the proximity distance of the amino acid residues within 5.0 Armstrong to the ligand as the ligand binding site (Figure 4.28).

All the atoms of the amino acid residues within 5.0 Armstrong of the ligand were presented in Figure 4.29. By comparing the trace residues with the ligand binding site, conserved ligand binding residues were identified (Table 4.4). There are five residues in the ligand binding site which are not-conserved in two species based on ET analysis. The reason is that their sequences length is short and the first part of their sequences only containing the gaps in multiple alignment result. Those residues are present in the table with '*' symbol.

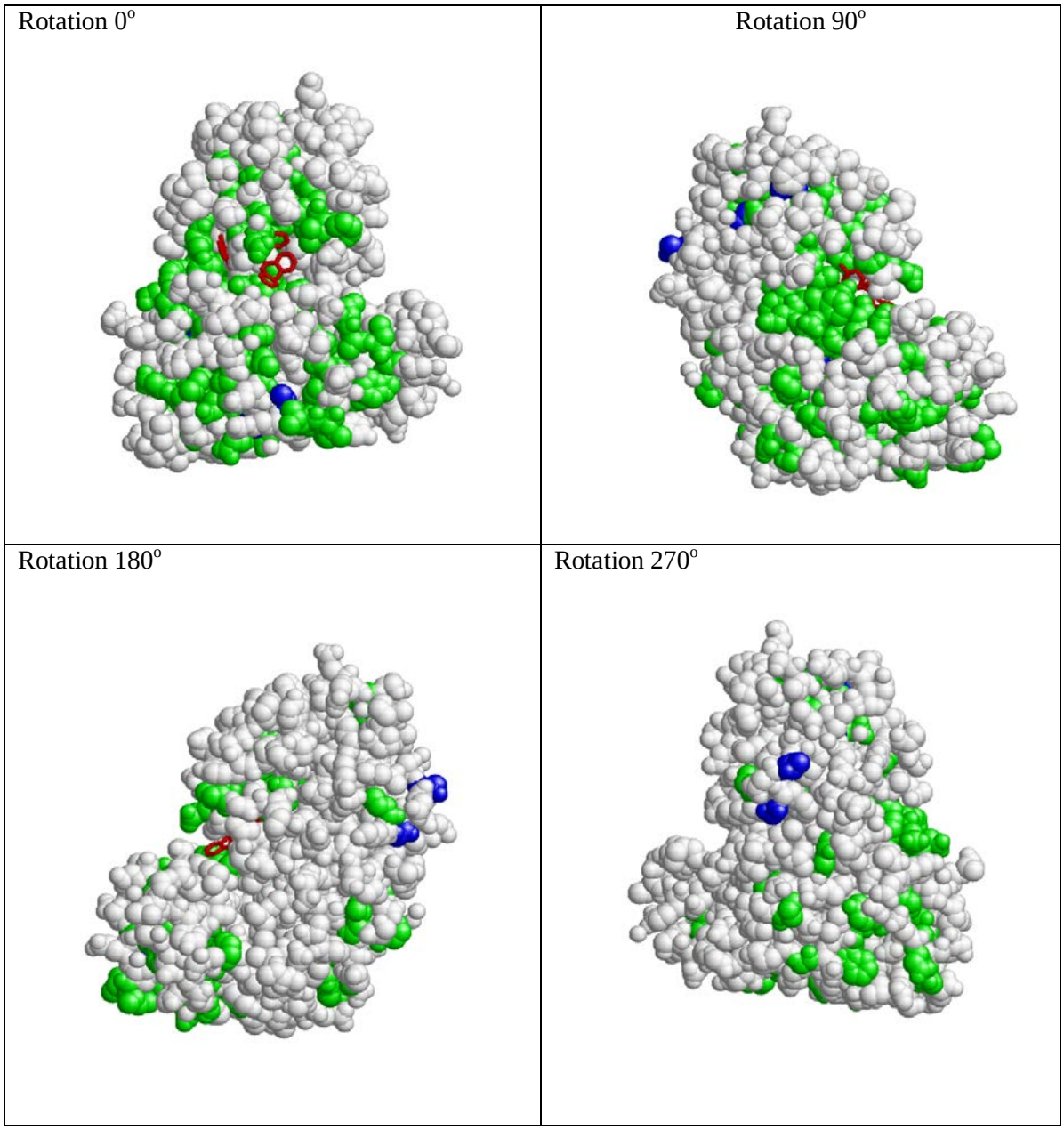

Figure 4.27 The 3-dimensional structure of BACE2's chain A with conserved residues from different view direction.

Table 4.5 The conserved residues and position in the structure of 2EWY protein.

Ser73	Tyr76	Val83	Val85	Tyr87	Gln89
Gly90	Trp92	Gly94	Gly97	Asp99	Pro104
Asn114	Ala116	Ile118	Ser121	Phe124	Phe125
Leu126	Phe127	Trp131	Gly133	Ile134	Leu135
Gly136	Leu137	Ala138	Tyr139	Leu142	Ala143
Gly190	Gly191	Phe194	Leu196	Gly199	Trp202
Tyr203	Thr204	Pro205	Glu208	Glu209	Trp210
Tyr211	Tyr212	Gln213	Glu215	Leu217	Lys218
Glu220	Gly222	Gly224	Leu226	Leu228	Asp229
Cys230	Tyr233	Asn234	Asp236	Lys237	Ala238
Ile239	Val240	Asp241	Ser242	Gly243	Thr244
Thr245	Leu246	Leu247	Arg248	Leu249	Pro250
Val253	Phe254	Ala256	Val258	Ser265	Leu266
Ile267	Phe270	Phe274	Gly277	Leu280	Ala281
Cys282	Trp283	Pr0289	Trp290	Phe293	Pro294
Ser297	Ile298	Tyr299	Arg301	Asn304	Ser306
Ile319					

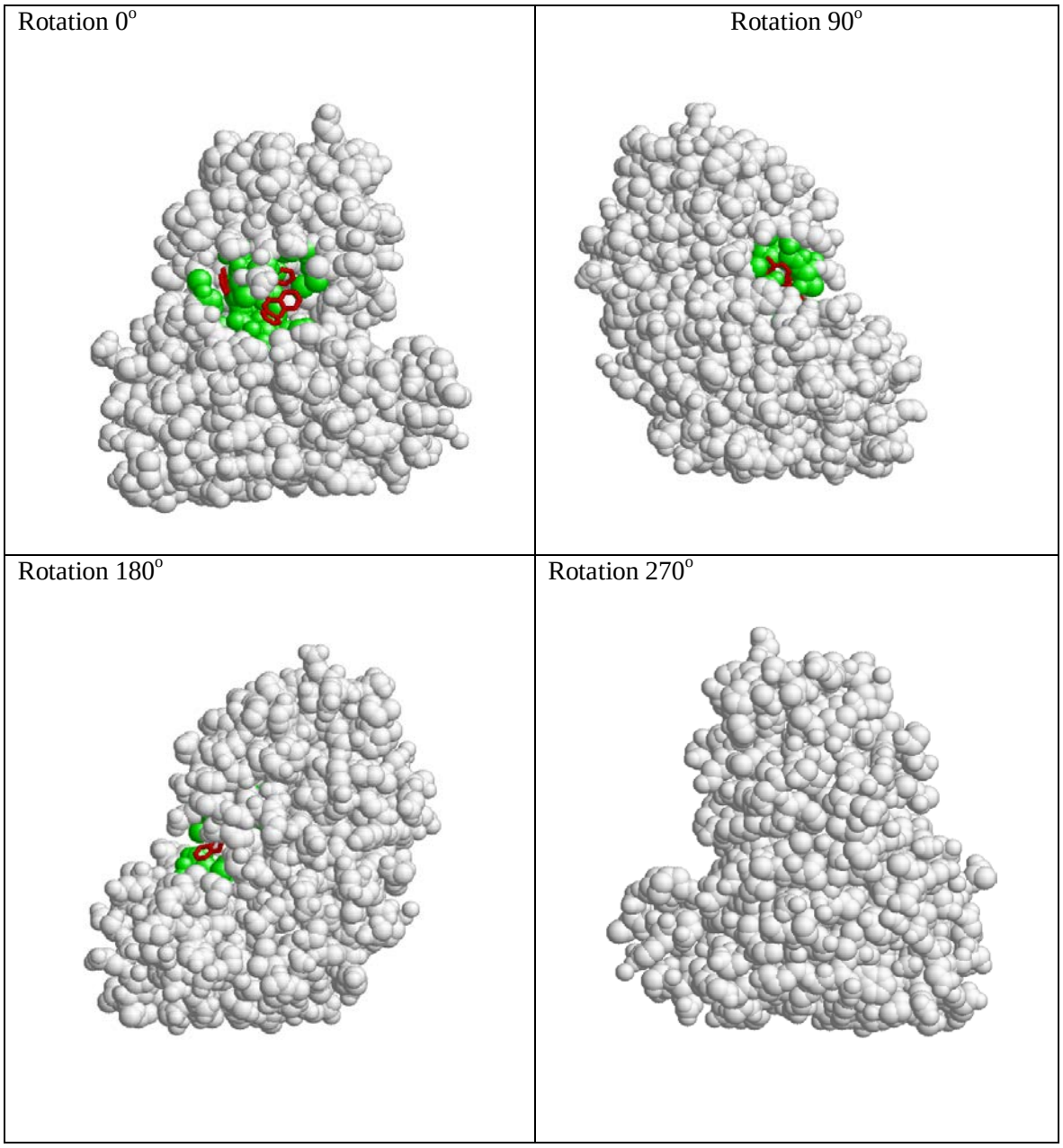

Figure 4.28 The 3-dimensional structure of ligand and trace residues within BACE2's chainA which are selected within 5.0 Armstrong from different view direction.

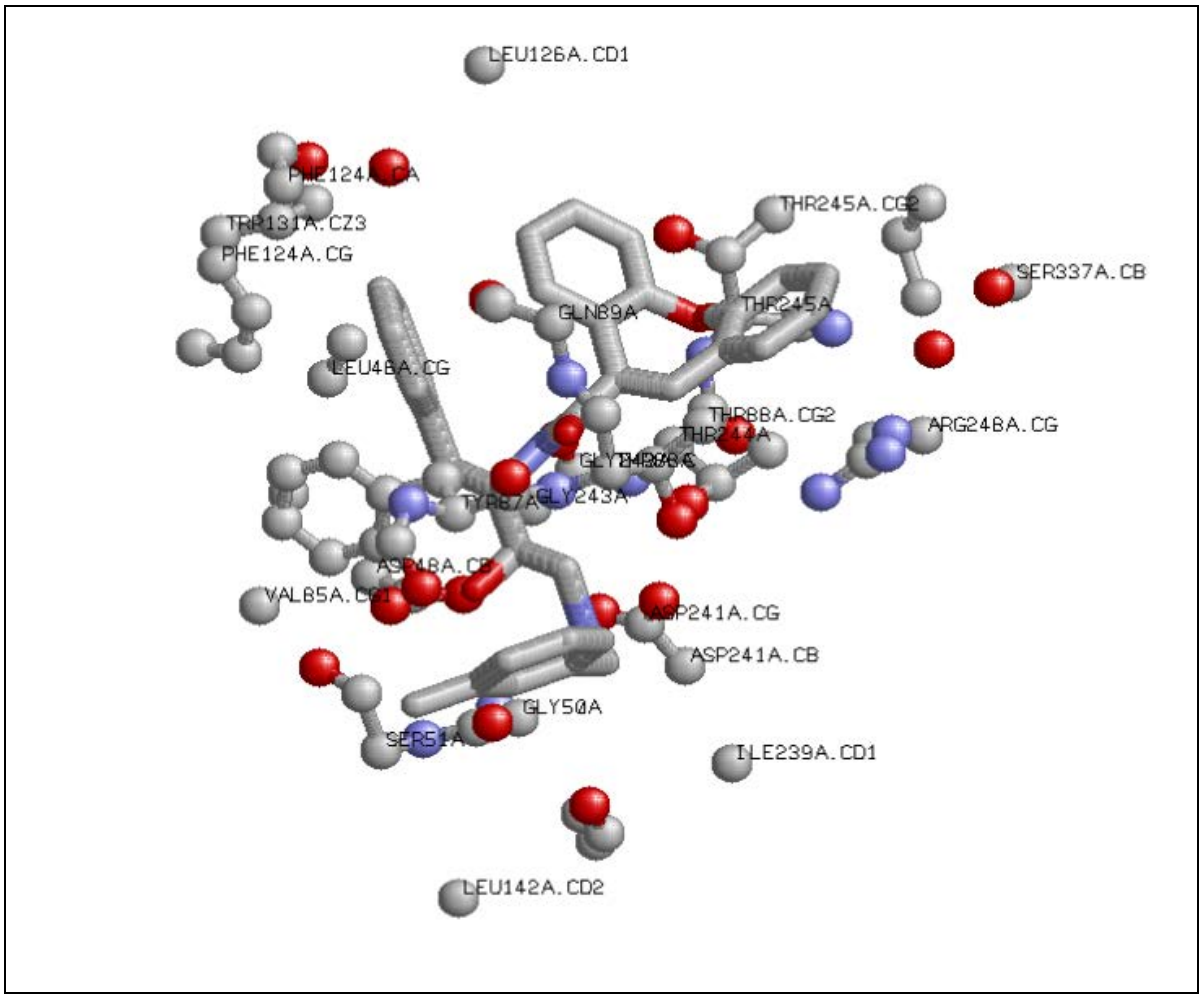

Figure 4.29 The display of ligand and conserved residues within the ligand binding site of BACE2. The ligand is displayed in form of "sticks" while the conserved residues are displayed in form of "Ball & Stick".

Table 4.6 The conserved residues positions in the BACE2 structure which are predicted present in active site.

Leu46*	Asp48*	Gly50*	Ser51*	Val85	Lys86*	Tyr87
Thr88*	Gln89	Gly90	Phe124	Leu126	Trp131	Ile134
Leu142	Tyr211	Ile239	Asp241	Gly243	Thr244	Thr245
Leu246	Arg248	Ser337				

4.5.3 BACE1 & BACE2 sequences analysis

In this step, to compare BACE1 and BACE2 protein sequences with each other and determine the conserved residues which are the same between them, the alignment result of BACE1& BACE2 sequences together based on their phylogenetic trees were analyzed.

4.5.3.1 Partition P_1

In P1 partition, the tree was vertically cut and the sequences were cleaved into two major groups, Group1 (G1),which is involve BACE1 sequences and Group2 (G2) which is contained the sequences of BACE2 (Figure 4.30). Based on the P1 partition in the alignment file the sequences were separated into two groups (Supplement Figure 12). All the P1 partition's trace residues of BACE1& BACE2 are displayed as Figure 4.31.

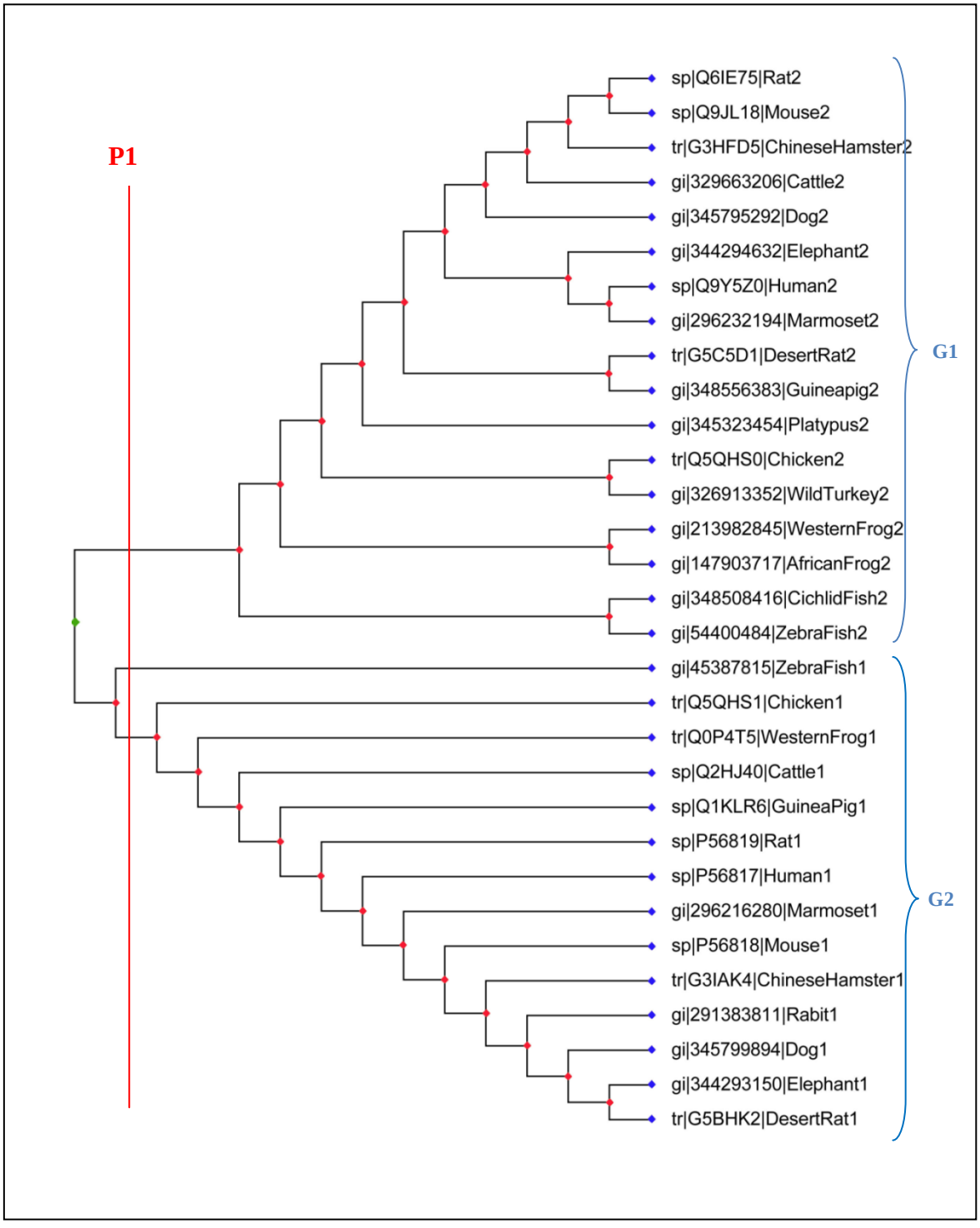

Figure 4.30 P1 partition of BACE1 & BACE2 phylogenetic tree.

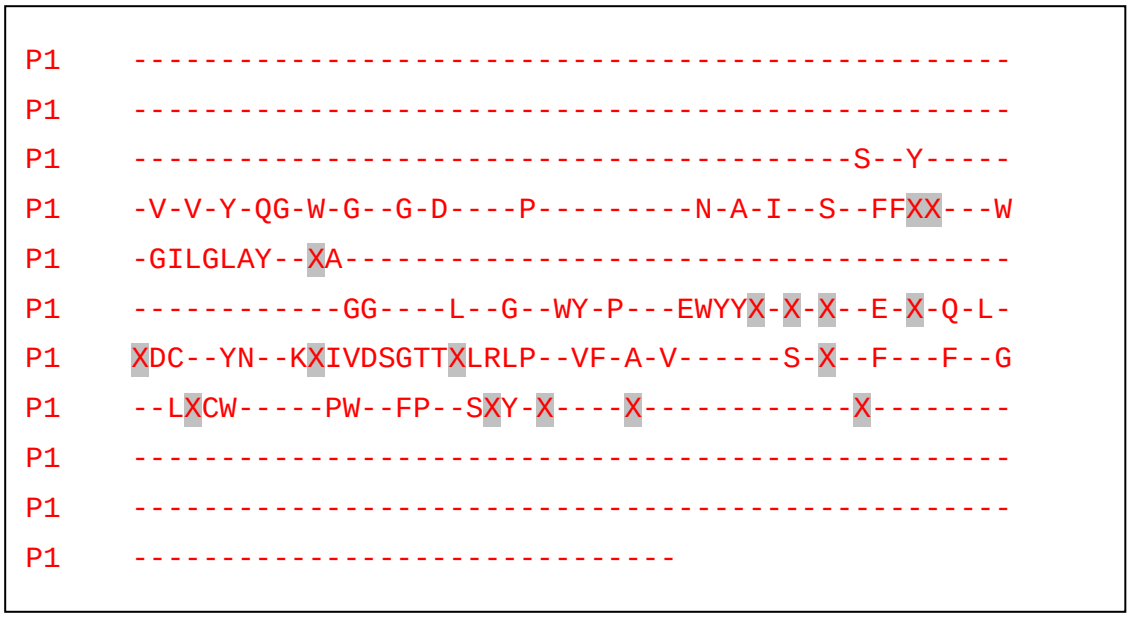

Figure 4.31 P1 partition' trace residues of BACE1 & BACE2 phylogenetic tree.

4.5.3.2 Trace Residues Analysis of BACE1 and BACE2

After determine the trace residues of P1 partition, in the next step, the position of each trace residues was highlighted in human BACE1 and BACE2 sequences and their PDB structure sequences (Figure 4.32).

4.5.3.3 Trace residues mapping on BACE2 Protein structure

In the final step, all those conserved residues within the 3D structure of BACE1 and BACE2 were selected and colored by green, while group specific residues were colored by blue (Figures 4.33, 4.34).

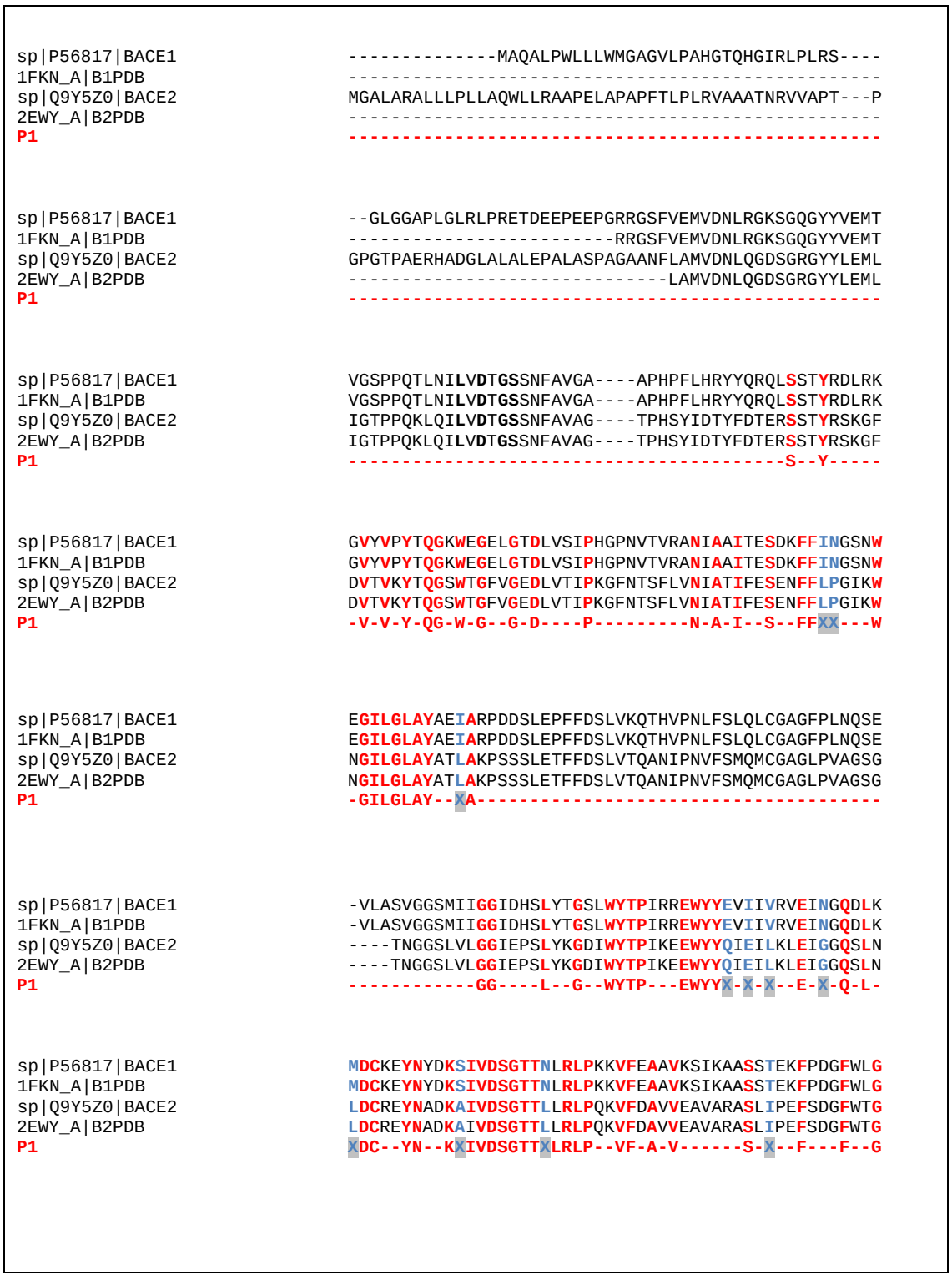

Figure 4.32 Position of each consensus sequences from P1 partition in human BACE1 & BACE2 sequences.

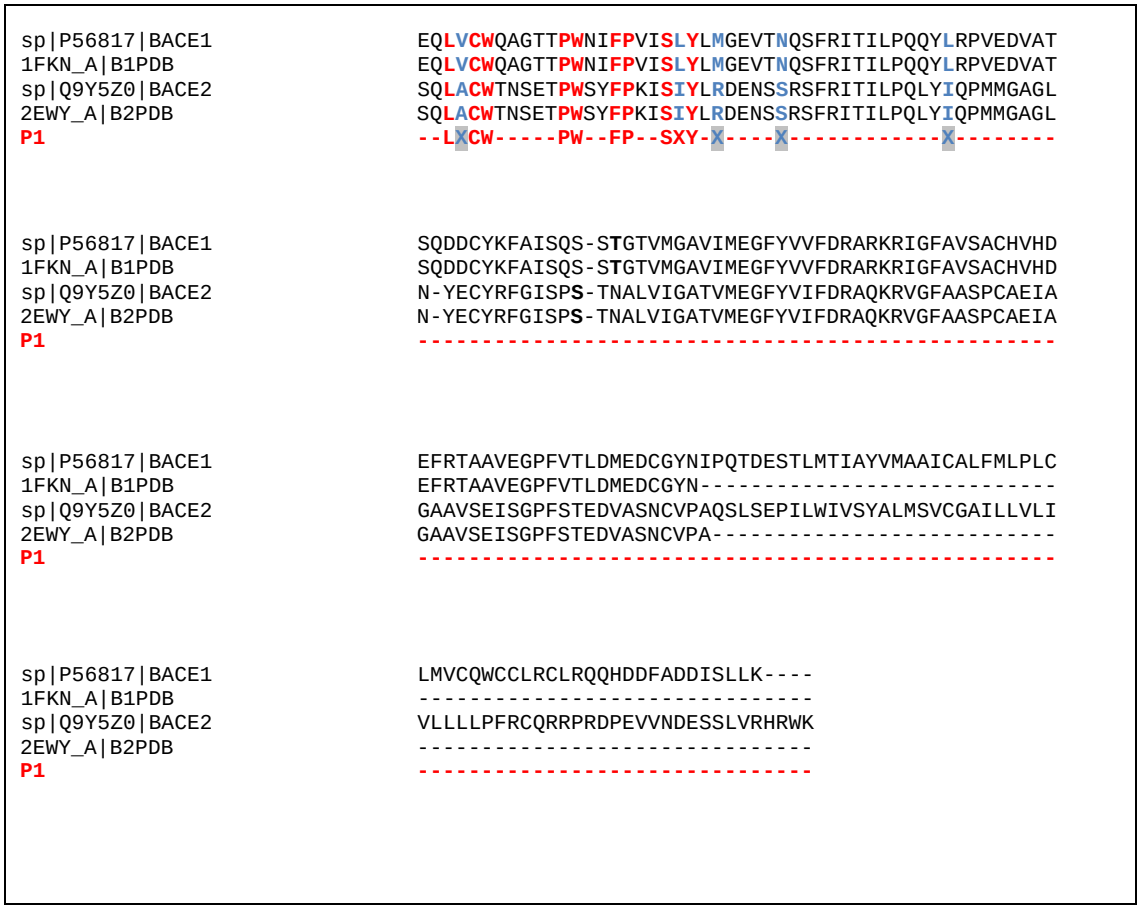

Figure 4.32 Position of each consensus sequences from P1 partition in human BACE1 & BACE2 sequences, continued.

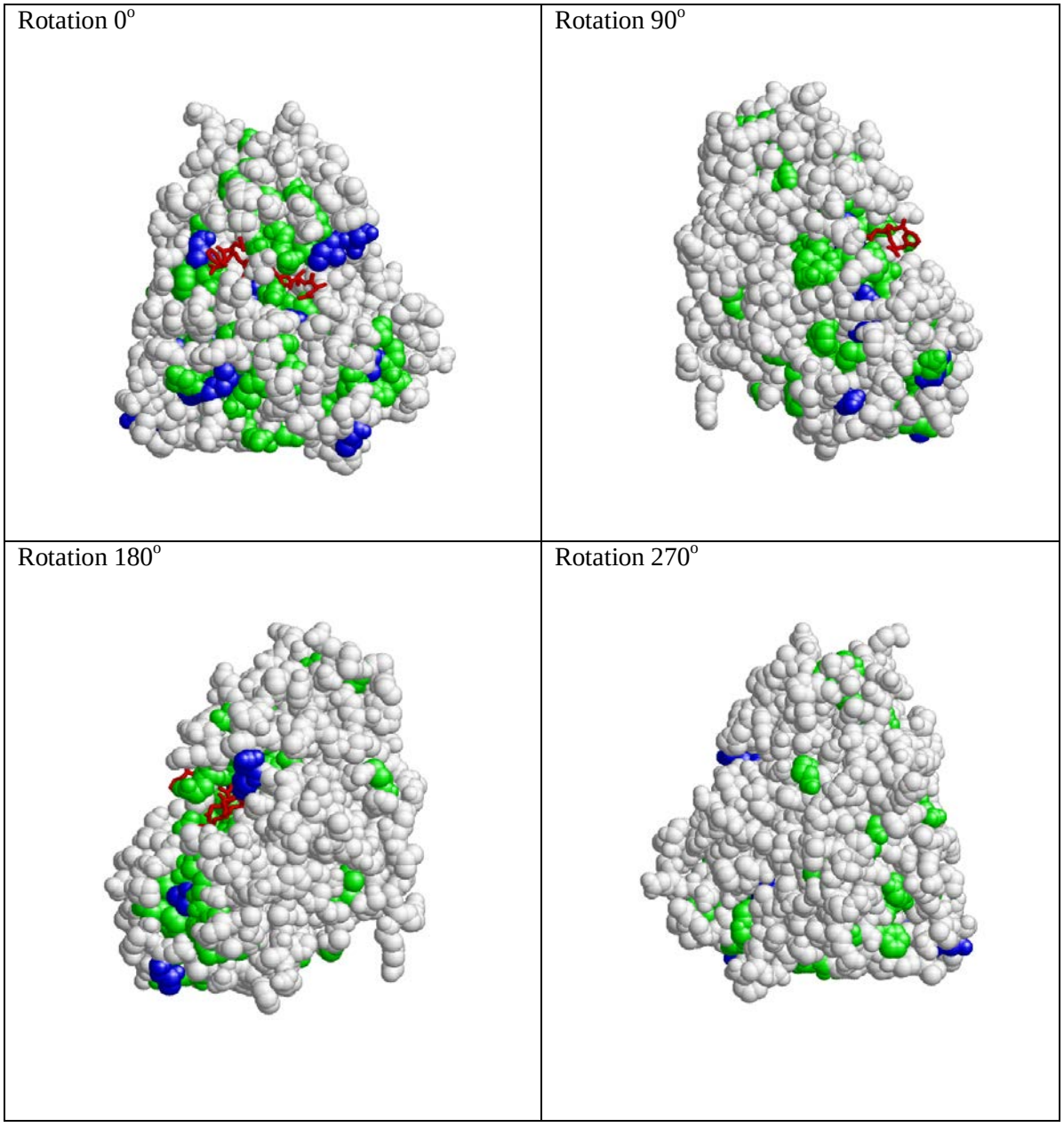

Figure 4.33 The 3-dimensional structure of BACE1'schain A with conserved residues and group specific residues from different view direction.

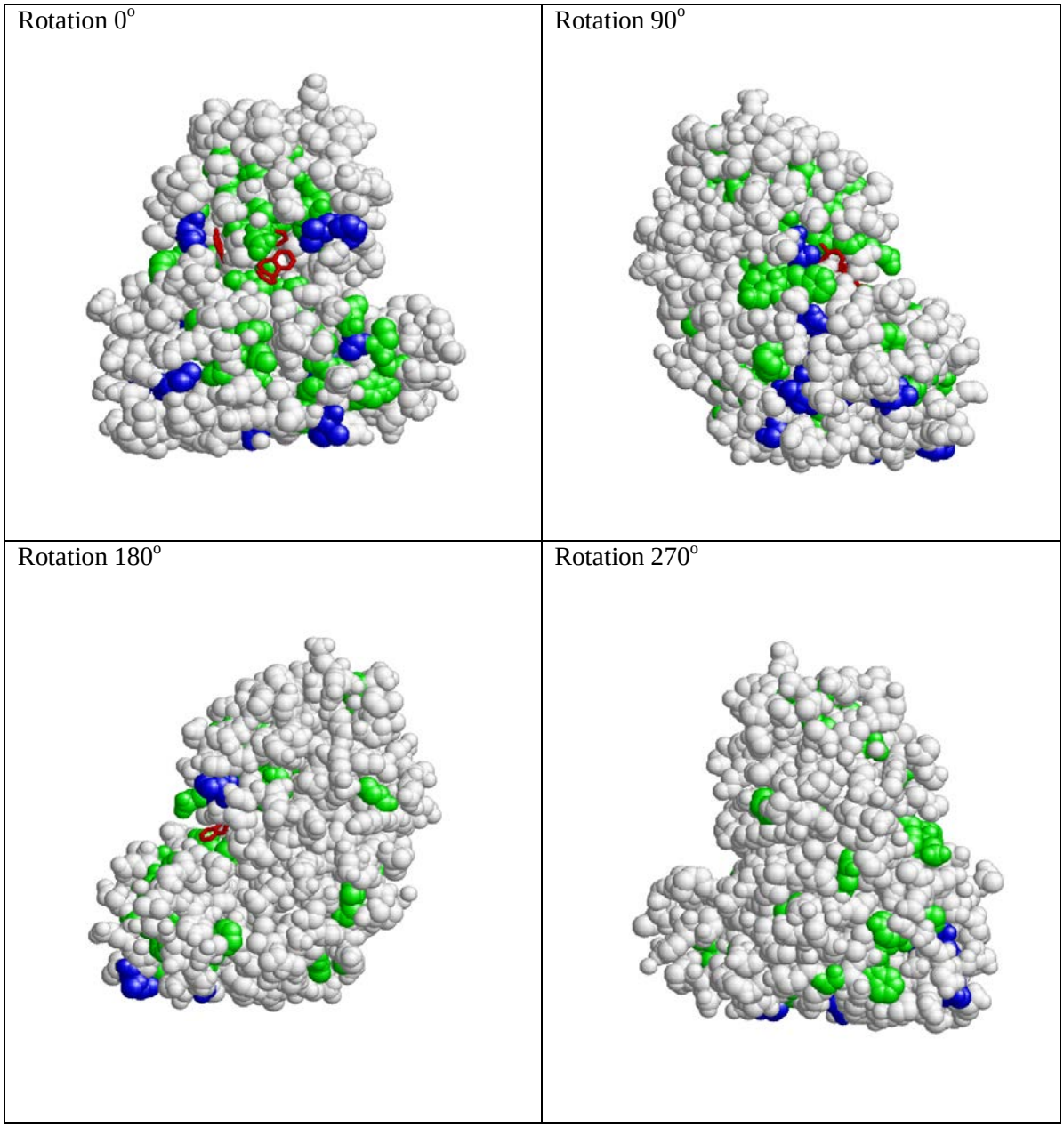

Figure 4.34 The 3-dimensional structure of BACE2's chain A with conserved residues and group specific residues from different view direction.

CHAPTER 5: DISCUSSION

5.1 Evolutionary Trace Analysis Result

Amyloid beta (Aβ) is the preliminary and main constituent of amyloid plaques seen in the brains of every single individual suffering from Alzheimer's disease (AD) and is caused by APP cleaves by β-secretase. BACE1 and BACE2 are two major forms of the β-secretase enzyme. The fragments of Aβ which are generated by the said internal cleavages might enjoy varied clinical consequences, even though both of the enzymes cleave APP. Aβ fragments generated by BACE1 is amyloidogenic and neurotoxic and cause to AD. On the contrary, in the senile plaques in AD, no BACE2-cleaved internal fragments were observed and they appear to be less amyloidogenic and neurotoxicas compared to the BACE1cleaved fragment neurotoxic and cause to AD (Farzan et al, 2000).

In this study, I compared the conservation pattern of BACE1 and BACE2 throughout species by using evolutionary trace method. The evolutionary trace is a method to identify active sites in a protein sequence by looking for conserved residues in the branches of an evolutionary tree. By using this method we can rank the functional significance of amino acids in the protein structure and also imply some data about the protein sequence-structure-function relationship.

I retrieved my candidates sequences which are human's BACE1 and BACE2 form Uniprot database. Then By using BLASTP program, the wide range of species' sequences were chosen according to their similarity with the human's BACE1 and BACE2.All the selected BACE1 sequences have high percentage similarity with the human BACE1 sequence from 99% to76%. On average, the length of those similar sequences is 480 except chicken which is 334 and dog which is 392 (Table 4.1). Also there is a high percentage similarity between human BACE2 sequences with other selected species from 94 to 51.8%

and somewhat similar sequences length, except chicken, Chinese hamster, elephant and zebra fish which have the short sequences in comparison with human (Table 4.2).

In the next step, to find out the evolutionary relationship between the sequences, multiple sequence alignment was performed to align those sequences based on their similarity, and then were used to construct a phylogenetic tree for evolutionary trace (ET) analysis and identified the conserved residues of BACE1 and BACE2 sequences.

According to my ET analysis, in human's BACE1, the number of conserved residues is 185 out of 501 which is 36% of the whole sequence. However there is no group specific residue observed for BACE1. The longest region of conserved residues is located in the middle of the sequences while the first and end of the sequences does not have any conserved residues (Figure 4.10).

The number of conserved residues in human's BACE2 sequence is 96 out of 518 which is 18% of the whole sequence. Also some group specific residues are identified in P2 and P3 partition of BACE2 phylogenetic tree (Figure 4.24). I found that, as same as BACE1 the longest region of conserved residues is located in the middle of the sequences while, the first and end of the sequences does not have any conserved residues.

After obtaining the conservation residues, I used Rasmol program to visualize the structure of BACE1 and BACE2 enzymes through protein mapping and prediction of the key functional residues and active site of the enzyme The active site of an enzyme is also known as ligand binding site. Therefore, each residue that is surrounding the ligand is considered as key functional residue. But, a problem arises in protein mapping is that we do not know whether that residue which is surrounding the ligand is conserved or not. To solve this problem, the result of protein mapping has to be comparing with the consensus sequence to identify the key functional residues. Those residues that are predicted as key

functional residues in protein mapping are identified as a part of active site if they are conserved also in the consensus sequence.

According to my analysis, there are 20 residues of BACE1 and 19 residues of BACE2 present in the active site of their structures and conserved between species (Tables 4.4, 4.6).

5.2 BACE1 & BACE2 structural comparison

The structure of BACE1and BACE2 is very similar to each other with an RMSD of 0.82over 335 Ca atoms. Nevertheless, it should be noted that the BACE1 and BACE2 sequences and the structure that created their function are different in terms of conformation. Referring to resent studies, the important difference between BACE1 and BACE2 structures is in tree loops. The first loop is in the N-terminal domain, consists of amino acid residues Ala42-Arg50 in BACE1 and Gly58-Tyr67 in BACE2.The second loop is also located in the N-terminal domain and involves amino acid residues Phe108-Trp115 in BACE1 and Phe124-Trp131 in BACE2.The third loop is in the C-terminal and contains amino acid residues Leu316-Thr 324 in BACE1 and Ile319-Tyr332 in BACE2.

The result of this study showed that, the first and third loops residues of BACE1 and BACE2 are not conserved between species. In the second loop all the residues are conserved in BACE1 and BACE2, except Gly128, Ile129, and Lys 130 in BACE2 which are not conserved between species based on my ET analysis.

5.3 BACE1 & BACE2 Active site comparison

The residues which are presented in the active site of BACE1 and BACE2 structure are located between N terminal and C terminal domain. The most important key functional residues in the active site are two aspartic acid residues, Asp32 and Asp228 in BACE1 and Asp48 and Asp241 in BACE2, which are supposed to be conserved between species. However according to my ET analysis, the Asp32 of BACE1 is not conserved in dog, desert rat, and zebra fish, also Asp48 of BACE2 is not conserved in chicken and elephant because their sequence in that region is truncated.

It has been observed that, there is a flap in the center of the active site of BACE1 and BACE2 structure which is contained residues Val67-Gly78 in BACE1 and Val83-Gly94 in BACE2 (Ostermann et al, 2006). The flap can adopt different conformations which are assumed to be able to control substrate access to the active site, to place the substrate in the right geometry for catalysis to happen and also to eliminate the hydrolysis products (Shimizu et al., 2008).

According to my ET analysis, in BACE1, residues Val69, Pro70, Tyr71, Thr72, Gln73 of the flap are presented in the active site, which are all conserved between species. Also in BACE2, residues Val85, Lys86, Tyr87, Thr88, Gln89, and Gly90 are identified in the active site which are located in the flap and conserved except Lys86 and Thr88. The reason is that, Lys 86 in Western frog, African frog and cichlid fish replaced by Arg and Thr88 in chicken and wild turkey, replaced by Ser. The important different of BACE1 and BACE2's flaps is Pro70 of BACE1 which is replaced by Lys86 in BACE2. These residues may have the effect on the flexibility of the flap.

The other prominent region in the active site of BACE1 and BACE2 is the second loop. Tree conserved active site residues of BACE1 and BACE2 are located in this position

which are involved Phe108, Ile110, Trp115 in BACE1 and Phe124, Leu126, Trp131 in BACE2.The important different between BACE1 and BACE2 in second loop is, Ile110 in BACE1, which is occupied by Leu126 in BACE2. These residues are may be of interest for the design of selective BACE1 or BACE2 inhibitors.

The two other active site residues, Ile126 and Asn 233 in BACE1 are replaced by, Leu142 and Leu246 in BACE2. All those residues are conserved and may play role in the different behavior of BACE1 with BACE2.

The position of all consensus sequences and group specific residues in human's BACE1 and BACE2 sequences and structure sequences were displayed as Figure 4.35. In this figure the residues which are presented in the active site of the BACE1 are highlighted by grey color while in the BACE2, those residues are highlighted by yellow color. The N-terminal, interdomain and C-terminal are present by blue, black and green arrows. The flap and tree loops regions that adopt different conformations in BACE1 compared to BACE2 are selected within the black and green color box.

The overall structure of BACE1 compared with BACE2 is presented as Figure 4,36. The N-terminal, inter-domains and C-terminal are colored green, grey and yellow respectively. The flap is displayed by orange color. The tree loops regions that adopt different conformations in BACE1 compared to BACE2 are presented by blue color. Also 4 residues which are presented in the active site of BACE1 and replayed by other residues in BACE2 are displayed in form "ball and stick" with red color.

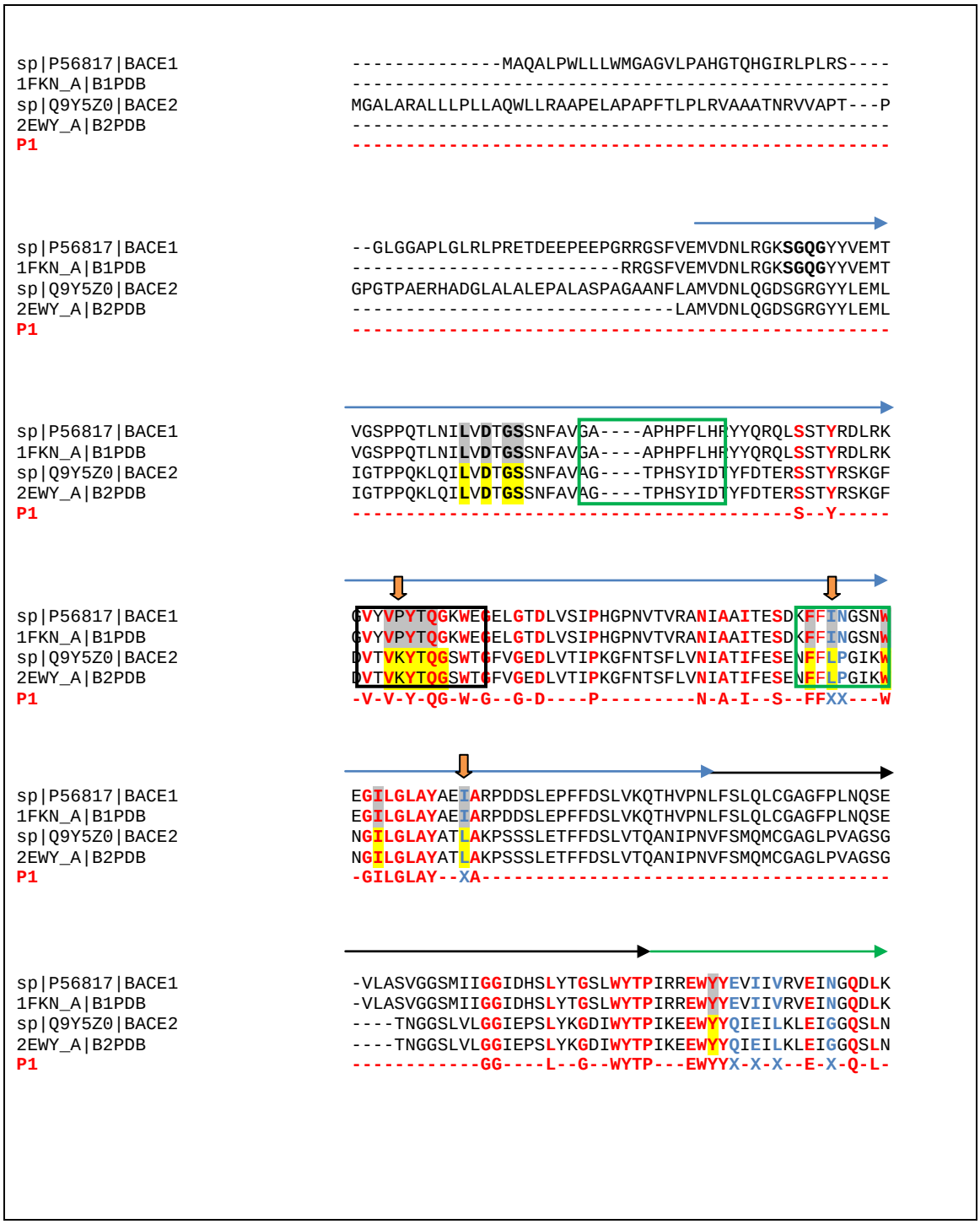

Figure 4.35 The position of all consensus sequences and group specific residues in human's BACE1 and BACE2 sequences.

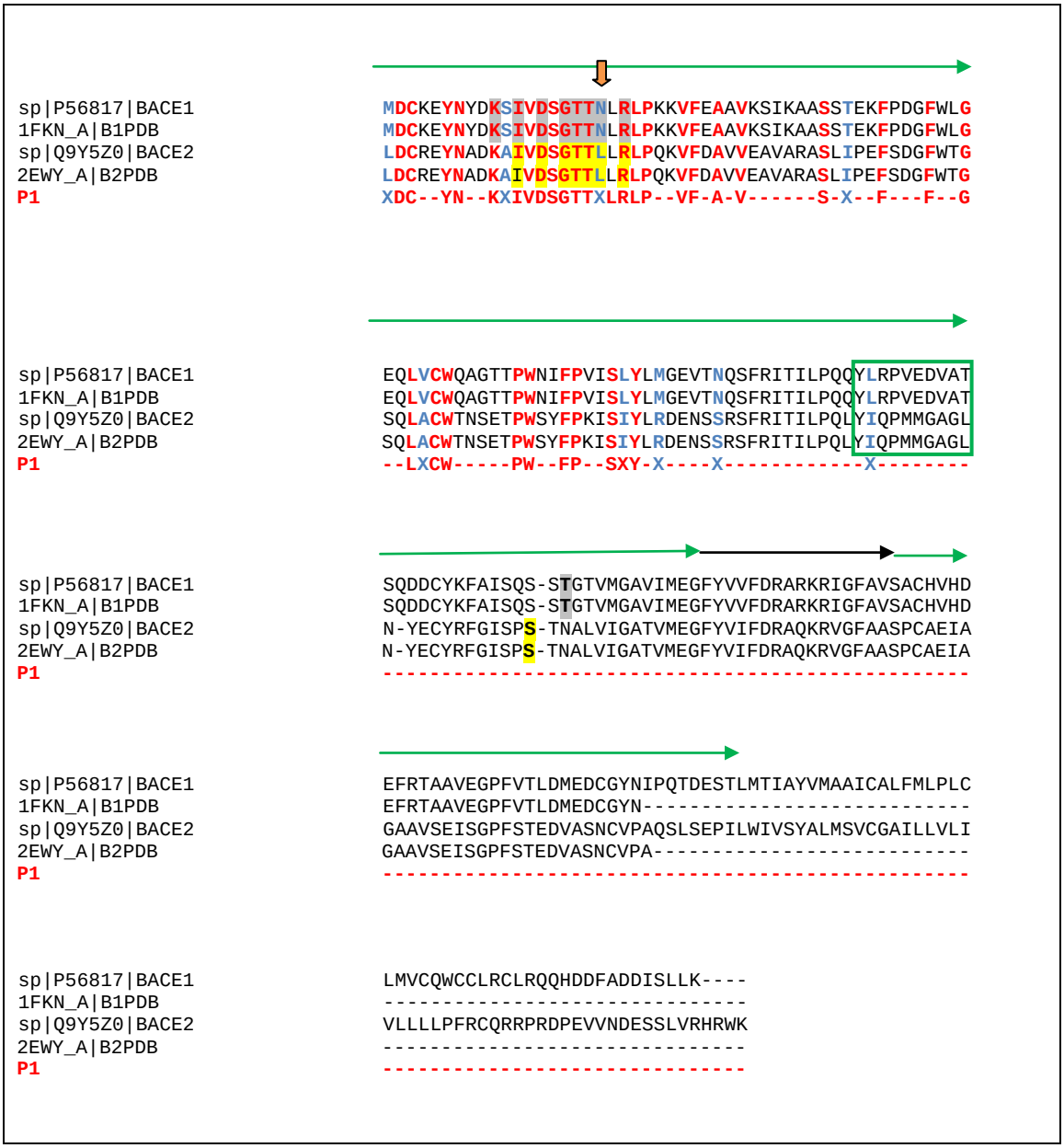

Figure 4.35 The position of all consensus sequences and group specific residues in human's BACE1 and BACE2 sequences, continued.

BACE1 **BACE2**

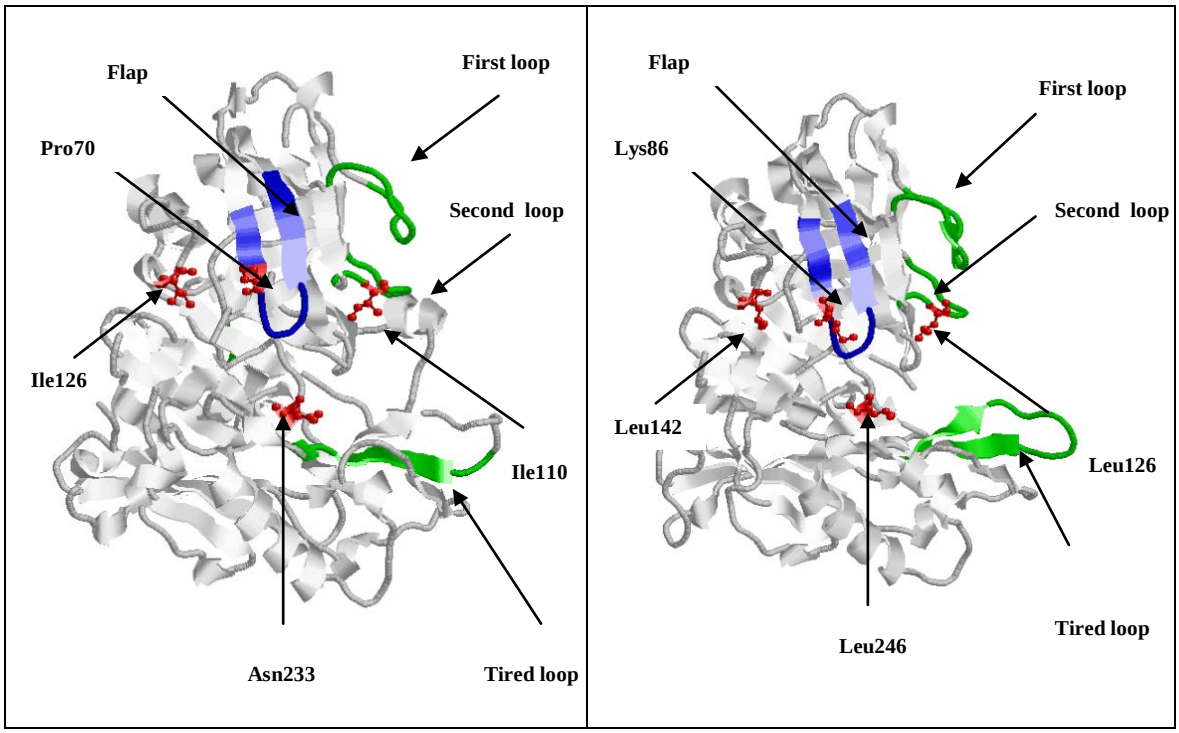

Figure 4,36 The overall structure of BACE1 compared with BACE2.

CHAPTER6: CONCLUSION

Beta-secretase is the rate limiting enzymatic activity in the production of the amyloid-β peptide (Aβ) and is thought to be involved in Alzheimer's disease (AD) pathogenesis. BACE1 and BACE2 are two types of Beta-secretase with 61.5% similarity and 45% identity at the amino acid level. Aβ fragments generated by BACE1 is cause to AD. In contrast, BACE2-cleaved internal fragments have not been observed in senile plaques in AD. In this study, I compared the relationship of the BACE1 and BACE2 among different species by using Evolutionary Trace method and identified the amino acid conservation pattern at the functional site of their structure.

The results of this study indicated that, the number of conserved residues in human's BACE1 sequence is 185 out of 501 which is 36% of whole sequence. Also in human's BACE2 the number of conserved residues is 96 out of 518 which is 18% of whole sequence. In both BACE1 and BACE2 structures, a cluster of conservation amino acid residues were identified at the proximity of the ligand which being 10% of whole conserved residues in BACE1 and 18% of whole conserve residues in BACE2. These conserved residues play the important functional role for enzymes activity. There are also some conserved residues scattered around the 3structures of both enzymes which are may be important on the structure stabilization or for attachment of protein surface with other protein.

Although the structure of BACE1 and BACE2 is very similar to each other, there are conformational differences between BACE1 and BACE2 sequences and structure which made their function different with each other. One of the important different between BACE1 and BACE2, is in the flap region which is able to cover the active site, adopting multiple conformational states in the various crystal structures. In this region, Pro70 of

BACE1 which is a conserved residue replaced by Lys86 in BACE2 which is not-conserved in some species. These residues may have the effect on the flexibility of the flap for binding of substrates or inhibitors to the enzymes.

The other different between BACE1 and BACE2's active site is residues Ile110 in BACE1, located in the second loop of the enzyme which occupied by Leu126 in BACE2. These residues are conserved and may be of interest for the design of selective BACE1 or BACE2 inhibitors.

The two other important residues, Ile126 and Asn 233 in BACE1 are replaced by, Leu142 and Leu246 in BACE2. All those residues are conserved and may play role in the different behavior of BACE1 with BACE2.