Research Article

Amino Acid Sequence and Structural Comparison of BACE1 and BACE2 Using Evolutionary Trace Method

Hoda Mirsafian,[1] Adiratna Mat Ripen,[2] Amir Feisal Merican,[1,3] and Saharuddin Bin Mohamad[1,3]

[1] *Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[2] *Allergy and Immunology Research Centre, Institute for Medical Research, Jalan Pahang, 50588 Kuala Lumpur, Malaysia*
[3] *Centre of Research for Computational Sciences and Informatics in Biology, Bioindustry, Environment,*
  Agriculture and Healthcare (CRYSTAL), University of Malaya, 50603 Kuala Lumpur, Malaysia

Correspondence should be addressed to Saharuddin Bin Mohamad; saharuddin@um.edu.my

Beta-amyloid precursor protein cleavage enzyme 1 (BACE1) and beta-amyloid precursor protein cleavage enzyme 2 (BACE2), members of aspartyl protease family, are close homologues and have high similarity in their protein crystal structures. However, their enzymatic properties differ leading to disparate clinical consequences. In order to identify the residues that are responsible for such differences, we used evolutionary trace (ET) method to compare the amino acid conservation patterns of BACE1 and BACE2 in several mammalian species. We found that, in BACE1 and BACE2 structures, most of the ligand binding sites are conserved which indicate their enzymatic property of aspartyl protease family members. The other conserved residues are more or less randomly localized in other parts of the structures. Four group-specific residues were identified at the ligand binding site of BACE1 and BACE2. We postulated that these residues would be essential for selectivity of BACE1 and BACE2 biological functions and could be sites of interest for the design of selective inhibitors targeting either BACE1 or BACE2.

1. Introduction

Members of aspartyl protease family are known to be associated with some pathological states such as breast cancer and pruritic inflammatory skin disease [1]. Furthermore, beta-amyloid precursor protein cleavage enzyme 1 (BACE1), a member of aspartyl protease family, is known to play an important role in the cellular pathways leading to Alzheimer's disease [2]. Beta-amyloid precursor protein cleavage enzyme 2 (BACE2) is a close homolog to BACE1 with 64% amino acid sequence similarity [3], and orthologs of both human BACE1 and BACE2 genes are present in marmoset, cattle, rabbit, guinea pig, rat, and mouse. Southan and Hancock reported the evolutionary history of BACE1 and BACE2, suggesting that the mammalian BACE1 and BACE2 are purifying selection compared to other classes analyzed [4].

BACE1 is a type 1 transmembrane protein which is highly expressed in brain and pancreas, and it can also be found in

other organs at much lower levels [5]. The current concept for Alzheimer's disease development involves the cleavage of amyloid precursor protein (APP) to amyloid beta (Aβ) peptide by BACE1 leading to the preliminary constituent of amyloid plaques in the brains of individuals with the disease [6]. BACE2 can be found in the majority of peripheral tissues with different expression levels, with kidney tissues showing the highest level of expression [7]. Though BACE2 is also expressed in the brain, some reports revealed the protective effect of the enzyme towards Alzheimer's disease, by reducing the pathogenic forms of Aβ being produced within the cells that expressed both BACE1 and BACE2 [8]. BACE1 competes with BACE2 over APP substrate, but their enzymatic activity generates different type of products in which the product of BACE1 but not BACE2 leads to Alzheimer's disease [9]. Furthermore, BACE1 is known as rate-limiting enzyme for Aβ formation leading to Alzheimer's disease, making it a good target for drug design to treat Alzheimer's disease. Since
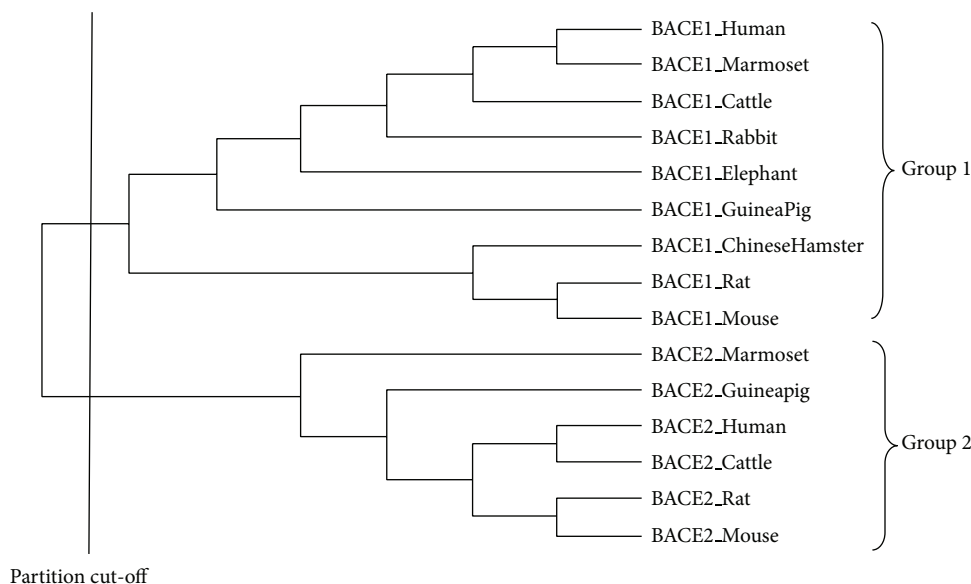
FIGURE 1: Evolutionary trace based cladogram of selected BACE1 and BACE2 protein sequences from UniProt database. ET partition cut-off divides the phylogenetic tree into Group 1 (BACE1) and Group 2 (BACE2).

BACE1 activity is known to cause Alzheimer's disease, it is essential to design a selective drug that could specifically target BACE1 over BACE2 [2].

Bioinformatics tools have been widely used for predicting protein functions. Various computational methods for predicting protein structure and functions have been developed. Methods to predict protein functions can be divided into sequence-based method and structure-based method [10]. In this study, evolutionary trace (ET) method which relies on both amino acid sequence and structural information to analyze functional sites of proteins is used because of the advantage of having molecular and structural information in a single analysis [11]. It identifies the amino acid conservation pattern of a group of proteins and transfers the information onto known 3D protein structures. Since amino acid residues that are vital for function or structural stability of a protein tend to be conserved across species, any differences at the functional sites would indicate possible differences for biological activity. Here, we reported the application of the ET method on BACE1 and BACE2 to extract the conservative and distinctive features of both enzymes in terms of molecular sequences and 3D structures that would provide beneficial details for the design of selective drug targeting BACE1 and BACE2.

2. Materials and Methods

2.1. Dataset. Human BACE1 and BACE2 sequences with UniProt accession numbers P56817 and Q9Y5Z0, respectively, were used as query sequences for BLASTP [12] searches against the UniProt database [13]. Selected BACE1 and BACE2 mammalian sequences with more than 80% identity to their respective query sequence were selected form BLASTP search results.

2.2. Evolutionary Trace Analysis. Multiple sequence alignment was carried out on the selected sequences using ClustalW [14] and a phylogenetic tree was generated using Neighbor-joining algorithm method and visualized by Phylo-Draw [15]. After the phylogenetic tree had been constructed, a vertical cut-off line was generated at the node of the tree which divided the tree into two groups with one group containing all BACE1 sequences while the other was with all BACE2 sequences (Figure 1).

Then, the sequences were separated based on groups. Sequences within each group were separately aligned together and consensus sequences were obtained for a particular group. The consensus sequences were classified as neutral, group-specific, and conserved residues. Conserved residues are amino acid residues that are conserved in the multiple sequence alignment while neutral residues are amino acid residues that are not conserved. Group-specific residues are amino acid residues that are conserved within a particular group but differ from other groups. The consensus sequences were then compared and an ET sequence was derived. ET sequences were compared to the query sequences (P56817, Q9Y5Z0). To identify the location of conserved and group-specific amino acid residues, the ET sequence was aligned with the query sequences (Figure 2).

Crystal structures of human BACE1 and BACE2 with the PDB codes 1FKN and 2EWY, respectively, were retrieved from Protein Data Bank (PDB) [16] and visualized by Python based Molecular Viewer software [17]. In order to analyze the ligand binding site of BACE1 and BACE2, only chain A of both 3D structures was selected, and amino acid residues within 5 Å distance from the ligand were defined as ligand binding sites. Ligand binding residues were compared to the ET sequence to identify their trace status and later mapped onto the 3D structure.

```
BACE1_Human    --------MAQALPWLLLWMGAGVLPAHGTQHGIRLPLRS----------GLGGAPLGLR
BACE2_Human    MGALARALLLPLLAQWLLRAAPELAPAPFTLP-LRVAAATNRVVAPTPGPGTPAERHADG
ET             --------------XX-L------X----------XX-X-----------G------XXX


BACE1_Human          LPRETDEEPEEPGRRGSFVEMVDNLRGKSGQGYYVEMTVGSPPQTLNILVDTGSSNFAVG    41
BACE2_Human          LALALEPALASPAGAANFLAMVDNLQGDSGRGYYLEMLIGTPPQKLQILVDTGSSNFAVA    57
ET                   LXX-X-X--------X-FXXMVDNLXGXSGXGYYXEM--GXPPQX--ILVDTGSSNFAVX
                                                                      *  *  **

BACE1_Human    42    AAPHPFLHRYYQRQLSSTYRDLRKGVYVPYTQGKWEGELGTDLVSIPHGPNVTVRANIAA   101
BACE2_Human    58    GTPHSYIDTYFDTERSSTYRSKGFDVTVKYTQGSWTGFVGEDLVTIPKGFNTSFLVNIAT   117
ET                   X-PH-X-X-YXX-X-SSTY--XXX-VXVXYTQGXWXGXXGXD-VXIP--XN-XXL-N-AX
                                             ******

BACE1_Human    102   ITESDKFFINGSNWEGILGLAYAEIARPDDSLEPFFDSLVKQTHVPNLFSLQLCGAGFPL   161
BACE2_Human    118   IFESENFFLPGIKWNGILGLAYATLAKPSSSLETFFDSLVTQANIPNVFSMQMCGAGLPV   177
ET                   IXESXXFFXXGX-WXGILGLAYA-XAX------------------------XCGAGXPX
                          *    *  *         *

BACE1_Human    162   NQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYN   221
BACE2_Human    178   AGS---GTNGGSLVLGGIEPSLYKGDIWYTPIKEEWYYQIEILKLEIGGQSLNLDCREYN   234
ET                   ------XXXGGSXXXGGIXX-LY-G-XWYTPIXXEWYYXXXIXXXEIXGQ-LXXDCXEYN
                                                            *

BACE1_Human    222   YDKSIVDSGTTNLRLPKKVFEAAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFP   281
BACE2_Human    235   ADKAIVDSGTTLLRLPQKVFDAVVEAVARASLIPEFSDGFWTGSQLACWTNSETPWSYFP   294
ET                   XDKXIVDSGTTXLRLPXKVFXAVV-XXXX-SXXXXFX-GFWXG-QLACW-XXXTPW-XFP
                      * * * **** *

BACE1_Human    282   VISLYLMGEVTNQSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGF   341
BACE2_Human    295   KISIYLRDENSSRSFRITILPQLYIQPMMGAGLN-YECYRFGISPSTNALVIGATVMEGF   353
ET                   XISXYLX-EX--XXSFR-TILPQXYXQPXMXXX---YXCYXFX-S-SXXXXVXGATXMEGF
                                                          *  *

BACE1_Human    342   YVVFDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVM   401
BACE2_Human    354   YVIFDRAQKRVGFAASPCAEIAGAAVSEISGPFSTEDVASNCVPAQSLSEPILWIVSYAL   413
ET                   YV-FDRA-KRXGFA-SXC---XX---X---G-----------X----------------

BACE1_Human    402   AAICALFMLPLCLMVCQWCCLRCLRQQHDDFADDISLLK----   440
BACE2_Human    414   MSVCGAILLVLIVLLLLPFRCQRRPRDPEVVNDESSLVRHRWK   456
ET                   ---C--------------------X-X---D----------
```
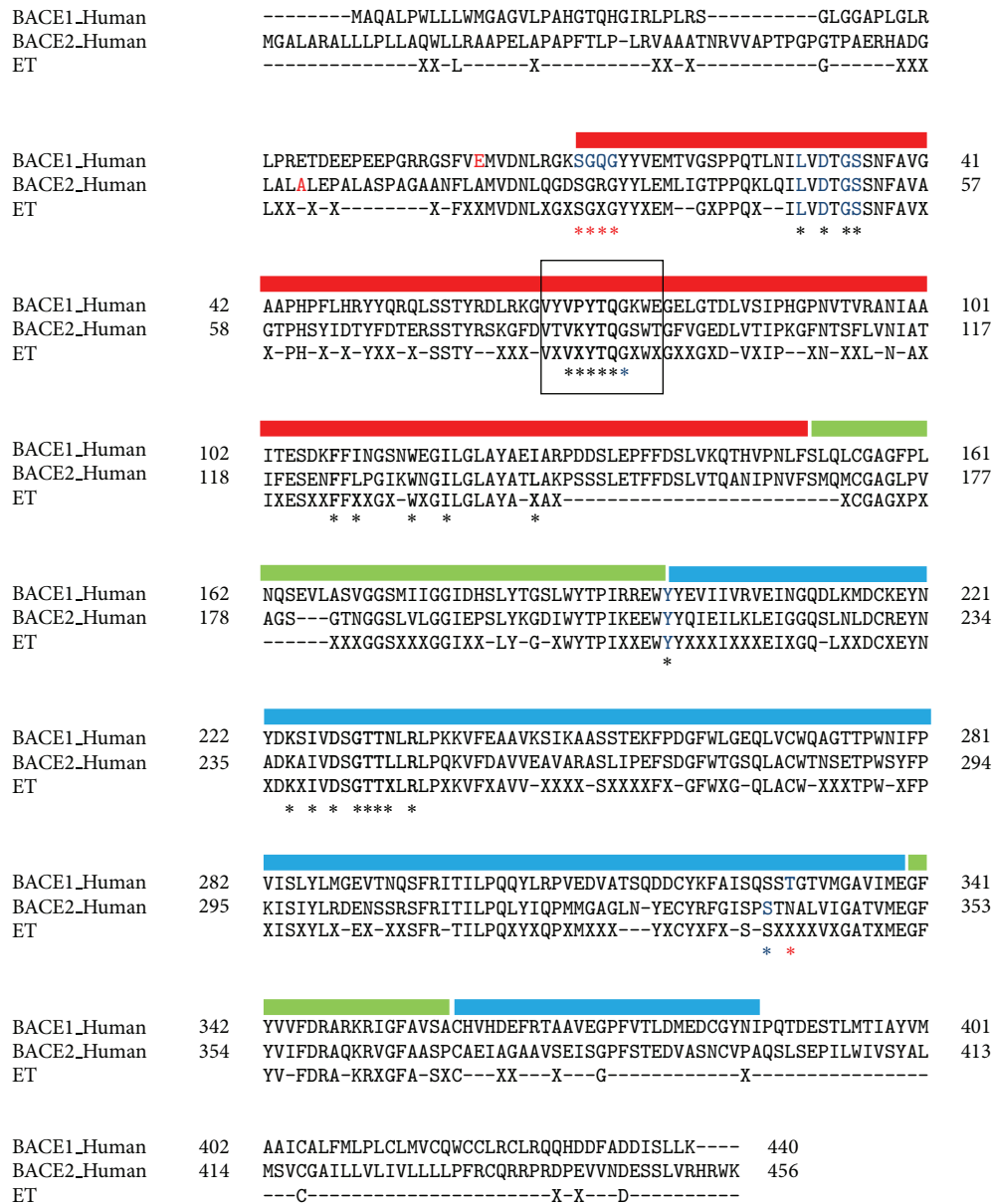
FIGURE 2: ET sequence compared to the seed sequences of BACE1 (P56817) and BACE2 (Q9Y5Z0). The domains within BACE1 and BACE2 sequences are marked with the coloured thick lines. Red and blue lines are for the N-terminal and C-terminal domains, respectively, whereas the green lines marked the interdomain regions within the structures. Amino acid residues for the flap region are marked within the box. Conserved residues are shown as their specific amino acid in the ET sequences, while group-specific residues are shown as alphabet "X." Neutral residues are marked as "-" in the ET sequence. The amino acid numbering of BACE1 and BACE2 sequences is according to their 3D structures obtained from PDB file, 1FKN, and 2EWY, respectively. The starting amino acid residues are highlighted in red. Amino acid residues that are involved in the ligand binding sites of BACE1 and BACE2 are marked with asterisk. Red coloured asterisks are binding sites that are only for BACE1, while blue coloured asterisks are the ligand binding sites that are only for BACE2.

3. Results and Discussion

We analyzed BACE1 and BACE2 protein sequences from several mammalian species using ET method to identify the functional sites that lead to different functional properties of BACE1 and BACE2. Developed by Lichtarge et al. in 1996, ET method enables the identification of conservation pattern within homologous proteins by comparing both the amino acid sequence and protein crystal structure information [11]. ET method applications were also reported in the identification ligand binding site [18] and the design of a high hydrophobicity laccase [19], an enzyme with industrial importance.

Our ET analysis revealed that 189 amino acid residues were conserved which is 37.7% of human BACE1 and 36.4% of human BACE2. Furthermore, 123 group-specific residues

(a)   (b)
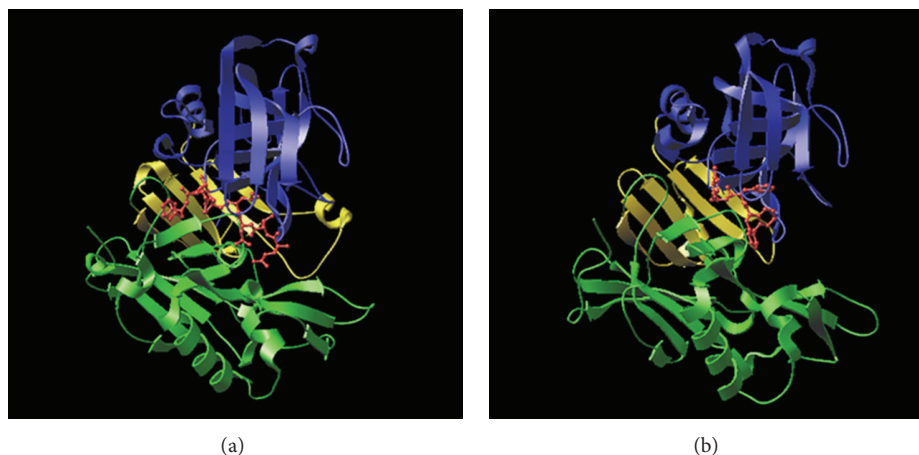
FIGURE 3: Chain A of crystal structures of BACE1 (a) and BACE2 (b). In both structures, the N-terminal, inter-, and C-terminal domains are coloured blue, yellow, and green, respectively. The ligands are displayed with "Ball and Stick" and red colour.
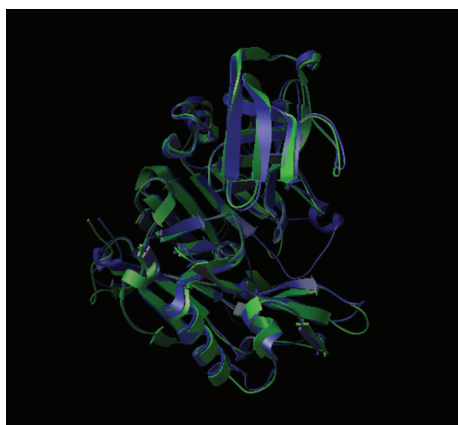

FIGURE 4: Superimposition of chain A of crystal structures of BACE1 (1FKN) and BACE2 (2EWY), where BACE1 is coloured blue and BACE2 is coloured green.

were identified, which comprise 24.5% of human BACE1 and 23.7% of human BACE2.

BACE1 and BACE2 are close homologs that not only share sequence similarity, but also have very similar 3D structure. Overall structures of BACE1 and BACE follow the general fold of aspartyl protease family comprising an N-terminal domain, a C-terminal domain, and an interdomain that connects the N-terminal and C-terminal domains (Figure 3) [9].

Superimposition of the crystal structures of BACE1 (1FKN) and BACE2 (2EWY) gave a root-mean-square deviation (RMSD) of 1.46 over 373 C-alpha atoms indicating that the structures of BACE1 and BACE2 are very similar to each other (Figure 4). Although BACE1 and BACE2 have the same biological activity against the APP substrate, their cleavage sites on APP are different, producing different cleavage products which are amyloidogenic for BACE1 and nonamyloidogenic for BACE2.

We defined the ligand binding site as all amino acid residues within 5 Å from the ligand. Twenty-eight of amino acid residues in BACE1 and 24 of amino acid residues in BACE2 were identified as ligand binding site. The ligand binding site residues of BACE1 and BACE2 together with their evolutionary trace status are summarized in Table 1. The entire ligand binding site residues were identified as conserved or group-specific residues and no neutral residues were found in the ligand binding site. This conservation pattern indicates that the active site for both BACE1 and BACE2 is well preserved for the aspartic protease activity. Two catalytic aspartic acid residues which are located in ligand binding site and essential for the enzyme activity are conserved in our analysis (BACE1: Asp32 and Asp228; BACE2: Asp48 and Asp241). However, we identified 4 residues that are group-specific in the ligand binding sites (Table 1). These group-specific residues are Pro70, Ile110, Ile126, and Asn233 of BACE1 substituting Lys86, Leu126, Leu142, and Leu246 of BACE2, respectively. Since we generated a partition cut-off on the phylogenetic tree that differentiates BACE1 and BACE2 into 2 separate groups, we believe that these 4 residues would play a pivotal role in determining the selectivity of enzymatic activity.

We mapped the ligand binding sites and their ET status onto the chain A of crystal structures of BACE1 (1FKN) and BACE2 (2EWY) (Figure 5). Two out of the 4 group-specific residues identified at the ligand binding site comprise Ile110 and Ile126 for BACE1 substituted with Leu126 and Leu144 for BACE2, respectively. Although these amino acid residues have similar physicochemical properties which are small and hydrophobic, the difference in the orientation of the alkyl group of Ile and Leu residues within the 3D structure of the protein may affect the binding selectivity of the ligand to BACE1 and BACE2.

Members of aspartyl protease family are known to have a loop structure that covers the active site of enzyme upon binding with the ligand. This structure is known as the flap region, and it is flexible and can adopt different structural conformations upon binding with the substrate. It is assumed that this structure will shield the active site from the solvent [20]. The flap region in BACE1 and BACE2 involves residues

Table 1: Amino acid residues at the ligand binding sites of BACE1 and BACE2. The amino acid residue numbering is according to their respective 3D structure (1FNK and 2EWY).

BACE1		BACE2	
Amino acid residues	Trace status	Amino acid residues	Trace status
Ser10	Conserved	**	—
Gly11	Conserved	**	—
Gln12	Conserved	**	—
Gly13	Conserved	**	—
Leu 30	Conserved	Leu46	Conserved
Asp32	Conserved	Asp48	Conserved
GLy34	Conserved	Gly50	Conserved
Ser35	Conserved	Ser51	Conserved
Val69	Conserved	Val85	Conserved
Pro70	Group-specific	Lys86	Group-specific
Tyr71	Conserved	Tyr87	Conserved
Thr72	Conserved	Thr88	Conserved
Gln73	Conserved	Gln89	Conserved
*	—	Gln90	Conserved
Phe108	Conserved	Phe124	Conserved
Ile110	Group-specific	Leu126	Group-specific
Trp115	Conserved	Trp131	Conserved
Ile118	Conserved	Ile134	Conserved
Ile126	Group-specific	Leu142	Group-specific
Tyr198	Conserved	Tyr211	Conserved
Lys224	Conserved	**	—
Ile226	Conserved	Ile239	Conserved
ASP228	Conserved	Asp241	Conserved
Gly230	Conserved	Gly243	Conserved
Thr231	Conserved	Thr244	Conserved
Thr232	Conserved	Thr245	Conserved
Asn233	Group-specific	Leu246	Group-specific
Arg235	Conserved	Arg248	Conserved
Thr329	Conserved	**	—
*	—	Ser337	Conserved

* Ligand binding site not identified in BACE1.
** Ligand binding site not identified in BACE2.

Val67 to Glu77 and Val83 to Thr93, respectively (Figure 2). In BACE1, Tyr71 of the flap region is known to play an important role in the conformational flexibility of flap region upon binding to ligand (reviewed in [21]). The amino acid sequences of flap in BACE1 and BACE2 are highly conserved. As shown in Figure 2, five residues of the flap were identified at the ligand binding site. Our analysis based on the ET method using a broader range of mammalian BACE1 and BACE2 sequences revealed that, in this region, Pro70 of BACE1 is substituted with Lys86 of BACE2. Since the cyclic property of proline residue may affect the flexibility of the flap region, this group-specific residue could be a crucial site that would lead to selectivity in the enzymatic activity of BACE1 and BACE2.
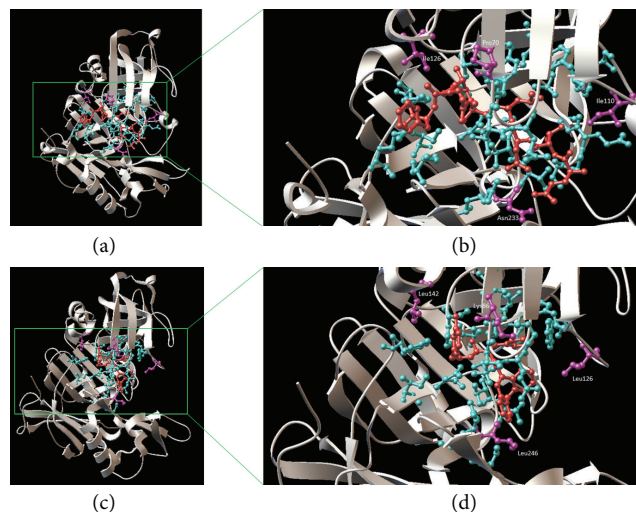

(a) (b) (c) (d)

Figure 5: The display of ligand, conserved, and group-specific residues within the ligand binding sits of BACE1 (a, b) and BACE2 (c, d). The ligands are coloured red while conserved and group-specific residues are coloured cyan and purple.

The last group-specific residue at the ligand binding site is Asn233 of BACE1 substituted with Leu246 of BACE2. The observation indicated that, in BACE2, the binding site at this region is more hydrophobic than the one in BACE1. This difference can be exploited for the design of a selective drug targeting either BACE1 or BACE2.

Designing a drug that could selectively inhibit BACE1 could avoid the unwanted side-effects of also inhibiting other aspartic protease family members including BACE2. BACE1 and BACE2 are close homologs and they are competing against each other for the same substrate (Aβ). As a result, the design of selective BACE1 or BACE2 inhibitors is rather challenging. Analyzing the sequences of a group of selected BACE1 and BACE2 proteins and mapping the details on known 3D structure reveal the differences between these two groups that would be beneficial for the selective drug design.

4. Conclusion

ET analysis on amino acid sequences and protein structures of BACE1 and BACE2 from several mammalian species enabled us to identify the distinctive features of BACE1 and BACE2 amino acid sequences. Mapping the ET analysis onto a known 3D structure of BACE1 and BACE2 revealed that their active sites are well conserved. Four group-specific residues were identified in the ligand binding sites of BACE1 and BACE2. The residues are Pro70, Ile110, Ile126, and Asn233 of BACE1 substituting Lys86, Leu126, Leu142, and Leu246 of BACE2, respectively. These group-specific residues would be the reason for cleavage site selectivity in BACE1 and BACE2 biological function and would be the potential residues for the design of selective and specific inhibitors targeting either BACE1 or BACE2.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] K. C. Chou, "Modeling the tertiary structure of human cathepsin-E," *Biochemical and Biophysical Research Communications*, vol. 331, no. 1, pp. 56–60, 2005.

[2] S. L. Cole and R. Vassar, "The Alzheimer's disease β-secretase enzyme, BACE1," *Molecular Neurodegeneration*, vol. 2, no. 1, article 22, 2007.

[3] R. Vassar and P. C. Kandalepas, "The β-secretase enzyme BACE1 as a therapeutic target for Alzheimer's disease," *Alzheimer's Research and Therapy*, vol. 3, no. 3, article 20, 2011.

[4] C. Southan and J. M. Hancock, "A tale of two drug targets: the evolutionary history of BACE1 and BACE2," *Frontiers in Genetics*, vol. 4, article 293, 2013.

[5] R. Vassar, B. D. Bennett, S. Babu-Khan et al., "β-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE," *Science*, vol. 286, no. 5440, pp. 735–741, 1999.

[6] M. Farzan, C. E. Schnitzler, N. Vasilieva, D. Leung, and H. Choe, "BACE2, a β-secretase homolog, cleaves at the β site and within the amyloid-β region of the amyloid-β precursor protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 17, pp. 9712–9717, 2000.

[7] B. D. Bennett, S. Babu-Khan, R. Loeloff et al., "Expression analysis of BACE2 in brain and peripheral tissues," *The Journal of Biological Chemistry*, vol. 275, no. 27, pp. 20647–20651, 2000.

[8] N. Ostermann, J. Eder, U. Eidhoff et al., "Crystal structure of human BACE2 in complex with a hydroxyethylamine transition-state inhibitor," *Journal of Molecular Biology*, vol. 355, no. 2, pp. 249–261, 2006.

[9] R. R. Ahmed, C. J. Holler, R. L. Webb, F. Li, T. L. Beckett, and M. P. Murphy, "BACE1 and BACE2 enzymatic activities in Alzheimer's disease," *Journal of Neurochemistry*, vol. 112, no. 4, pp. 1045–1053, 2010.

[10] J. D. Watson, R. A. Laskowski, and J. M. Thornton, "Predicting protein function from sequence and structural data," *Current Opinion in Structural Biology*, vol. 15, no. 3, pp. 275–284, 2005.

[11] O. Lichtarge, H. R. Bourne, and F. E. Cohen, "An evolutionary trace method defines binding surfaces common to protein families," *Journal of Molecular Biology*, vol. 257, no. 2, pp. 342–358, 1996.

[12] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *Journal of Molecular Biology*, vol. 215, no. 3, pp. 403–410, 1990.

[13] A. Bairoch, R. Apweiler, C. H. Wu et al., "The universal orotein resource (UniProt)," *Nucleic Acids Research*, vol. 33, pp. D154–D159, 2005.

[14] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, vol. 22, no. 22, pp. 4673–4680, 1994.

[15] J.-H. Choi, H.-Y. Jung, H.-S. Kim, and H.-G. Cho, "PhyloDraw: a phylogenetic tree drawing system," *Bioinformatics*, vol. 16, no. 11, pp. 1056–1058, 2000.

[16] H. M. Berman, J. Westbrook, Z. Feng et al., "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[17] F. S. Michel, "Python: a programming language for software integration and development," *Journal of Molecular Graphics and Modelling*, vol. 17, no. 1, pp. 57–61, 1999.

[18] S. B. Mohamad, A. L. Ong, and A. M. Ripen, "Evolutionary trace analysis at the ligand binding site of laccase," *Bioinformation*, vol. 2, no. 9, pp. 369–372, 2008.

[19] S. B. Mohamad, A. L. Ong, R. F. Khairuddin, and A. M. Ripen, "A combination of evolutionary trace method, molecular surface accessibility and hydrophobicity analysis to design a high hydrophobicity laccase," *In Silico Biology*, vol. 10, no. 3-4, pp. 145–153, 2010.

[20] Y. Xu, M. Li, H. Greenblatt et al., "Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations," *Acta Crystallographica D: Biological Crystallography*, vol. 68, no. 1, pp. 13–25, 2012.

[21] J. Yuan, S. Venkatraman, Y. Zheng, B. M. McKeever, L. W. Dillard, and S. B. Singh, "Structure-based design of β-site APP cleaving enzyme 1 (BACE1) inhibitors for the treatment of Alzheimer's disease," *Journal of Medicinal Chemistry*, vol. 56, no. 11, pp. 4156–4180, 2013.