Graduate Theses, Dissertations, and Problem Reports

2008

Bioinformatic Study of gamma-Secretase and its Substrates

Fei Nan
West Virginia University

Bioinformatic Study of γ-Secretase and its Substrates

Fei Nan

Thesis Submitted to

School of Medicine

West Virginia University

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

Master of Science

In

Microbiology, Immunology, and Cell Biology

Dr. Bing-Hua Jiang, Chair

Dr. Donald Adjeroh, Co-Chair

Dr. Jia Luo

Department of Microbiology, Immunology, and Cell Biology`

Morgantown, West Virginia

2008

ABSTRACT

Bioinformatic Study of γ-Secretase and its Substrates

Fei Nan

γ-Secretase is an enzymatic complex critical to amyloid formation and other important pathophysiological processes. Non-specific γ-secretase inhibitors are not ideal for intervention of familial Alzheimer's disease. The four essential include Presenilin (PS1 or PS-2), APH-1, Pen-2, nicastrin (Nct). Also γ-secretase cleaves a variety of substrates (over 30). But different sites are cleaved even in a single substrate and no consensus sequence for cleavage by γ-secretase is identified. We hypothesize that consensus compositional sequence (or functional domain) in components of γ-secretase and/or its substrates should exist during γ-secretase regulation and specific γ-secretase inhibitor based on such consensus sequence (or functional domain) may act as a therapeutic intervention for familial Alzheimer's disease. We divide our research into two stages. At Stage I, we find the candidates by the methods of bioinformatics. The less, the better. By using the resources of genetic and protein databases and tools, selecting software to apply sequence analysis, modeling and simulation, and pattern recognition of γ-secretase regulation, we identify candidates of consensus sequences (or functional domains). At Stage II, we use the traditional biological methods to verify the critical role of the candidates including the protein purification, recognition, and functional domain interaction. This thesis work is particularly focused on Stage I.

Acknowledgements

I would like to take this great opportunity to thank my supervisor, Dr. Bing-Hua Jiang, for his endless and selfless support, guidance, insights and advice in the hard times, giving me constant encouragement. I would like to express my most profound gratitude to my co-advisor, Dr. Donald Adjeroh, for his greatness, scientific insights, inspiration and patience to keep me on the right track. I would like to thank Dr. Jia Luo for his time and constructive criticism in my thesis, which has helped develop my critical thinking. I would like to thank Dr. Xiaobo Zhou for his help and advice on experimental data as well as with presentation of my data. I would like to thank David Woo and Thomas Wessel for their open mind suggestion and discussion, which makes everyday an enjoyable day. I would also like to thank all people who helped me throughout my study.

Table of Contents

Abbreviations

Aß: Amyloid Beta

Aß40: Amyloid Beta 40

Aß42: Amyloid Beta 42

Aτ: Amyloid-τ

ACANA: Accurate Anchoring Alignment

ACH: Acethylcholine

ANOVA: One Way Analysis of Variance

APH-1: Anterior Pharynx-defective 1

APH-1a: Anterior Pharynx-defective 1 Homolog A (C. Elegans)

APH-1b: Anterior Pharynx-defective 1 Homolog B (C. Elegans)

APLP2: Amyloid Beta (A4) Precursor-Like Protein 2

APP: Amyloid Precursor Protein

BACE: ß-Secretase

BLAST: Basic Local Alignment Search Tool

BLOSUM: Blocks of Amino Acid Substitution Matrix

DNA: Deoxyribonucleic Acid

ELISA: Enzyme Linked Immunosorbent Assay

FAD: Familial Alzheimer's Disease

GAPDH: Dehydrogenase

GLUT1: Glucose Transporter 1

H_2O: Water

IgE: Immunoglobulin E

IgG: Immunoglobulin G

IL-: Interleukin

MEM: Maximal Exact Matches

MUM: Maximum Unique Match

MUSCLE: Multiple Sequence Comparison by Log-Expectation

NCT: Nicastrin

NK: Natural Killer

NOR1: Neuron-Derived Orphan Receptor 1

NSAID: Non-Steroidal Anti-Inflammatory Drugs

NUR77: Nerve Growth Factor IB

P3: Proximal Third

PAM: Point Accepted Mutation

PCA: Passive Cutaneous Assay

Pen-2: Presenilin Enhancer 2

PS1: Presenilin 1 (Particular on Human)

PS2: Presenilin 2 (Particular on Human)

PSen1: Presenilin 1

PSen2: Presenilin 2

RNA: Ribonucleic Acid

SCP: Sorted Common Prefix

T-COFFEE: Tree-Based Consistency Objective Function For Alignment Evaluation

TGF-β: Tumor Necrosis Factor Beta

Th1: T Helper 1

Th2: T Helper 2

TNF: Tumor Necrosis Factor

UV: Ultraviolet

VH: Vehicle

γ-Secretase: Gamma Secretase

α-secretase: Alpha Secretase

β-secretase: Beta Secretase

1. Introduction

1.1 Alzheimer's Disease and Amyloid-β

Alzheimer's disease is a neurodegenerative disease that, in its most common form, is found in people over age 65. Approximately 24 million people worldwide have dementia of which about 60% is due to Alzheimer's disease.
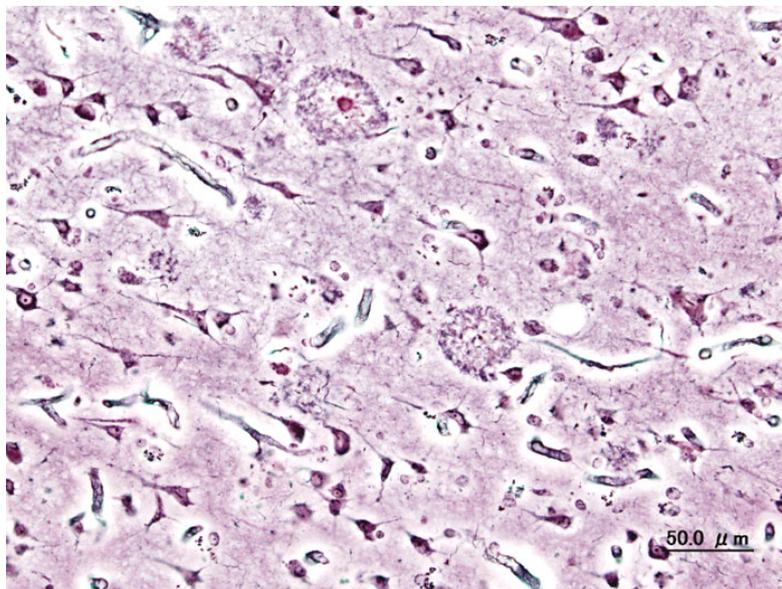
Figure 1. Histopathologic image of senile plaques seen in the cerebral cortex in a patient with Alzheimer's disease of presenile onset. Silver impregnation.

Plaques which contain misfolded peptides called amyloid-β (Aβ) are formed in the brain many years before the onset of Alzheimer's disease. Together, these plaques and neurofibrillary tangles form the pathological signs of the disease. These features can only be discovered at autopsy and help to confirm the clinical diagnosis. Medications can help

relieve the symptoms of the disease, but evidence proves that they canot change the course of the underlying pathology.

Experimental and genetic evidence has indicated that amyloid-β (Aβ) plays an essential role in the Alzheimer's disease pathogenesis. The malicious protein aggregates in the brain hippocampus and brain cortex of the patients of Alzheimer's disease as well as self-accumulation to form malicious oligomers which will cause neurodegeneration. Genetic defects identified include mutations in the Presenlin 1 gene on chromosome 14, the Presenilin 2 gene on chromosome 1, and the amyloid precursor protein (APP) gene on chromosome 21.

Alzheimer's disease has been identified as a protein misfolding disease due to the accumulation of abnormally folded amyloid-β (Aβ) in the brains of Alzheimer's disease patients (Hashimoto et al 2003). Amyloid-β (Aβ) is a short peptide that is a proteolytic byproduct of the transmembrane protein amyloid precursor protein (APP), whose function is unclear but thought to be involved in neuronal development (Rudolph et al 2000). Although Amyloid-β (Aβ) monomers are soluble and harmless, they undergo a dramatic conformational change at sufficiently high concentration to form a beta sheet-rich tertiary structure that aggregates to form amyloid fibrils (Ohnishi et al 2004) that deposit outside neurons in dense formations known as senile plaques or neuritic plaques, in less dense aggregates as diffuse plaques, and sometimes in the walls of small blood vessels in the brain in a process called amyloid angiopathy or congophilic angiopathy.

Recent research includes hypotheses centered on the effects of the misfolded and aggregated proteins, amyloid-β (Aβ) and amyloid-τ (Aτ). The τ hypothesis is supported by the long-standing observation that deposition of amyloid plaques do not correlate well with neuron loss (Mudher et al 2002). However, a majority of researchers support the alternative hypothesis that Aβ is the primary causative agent (Schmitz et al 2004).

1.2 Secretase

Secretases are enzymes that cut pieces off a longer protein that is embedded in the cell membrane. There are three types of components for secretase, namely α-secretase, β-secretase and γ-secretase.
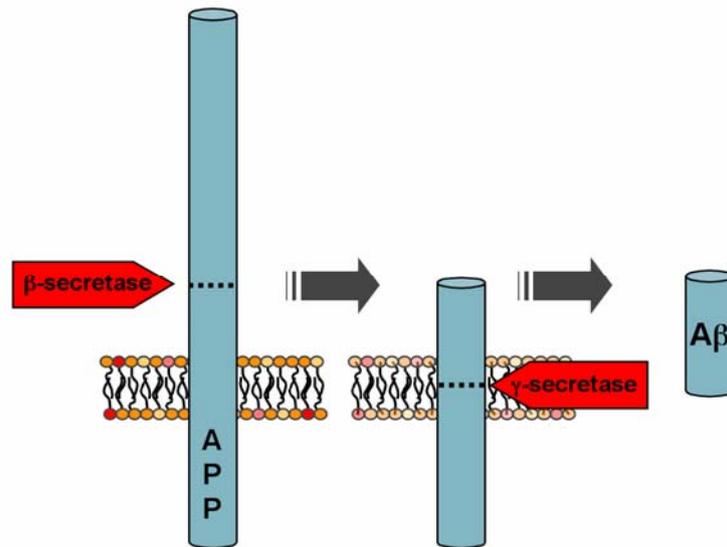
Figure 2. Amyloid-β is created from amyloid precursor protein, through two sequential proteolytic cleavages where secretases act on the amyloid precursor protein (APP) to cleave the protein into three fragments. Sequential cleavage by β-secretase (BACE) and γ-secretase produces the amyloid-β (Aβ) peptide fragment that aggregates into plaques in the brains of Alzheimer's disease patients.

Amyloid-β is heterogenous and formed from the precursor protein, called amyloid precursor protein, through two sequential proteolytic cleavages that involve β-secretase and γ-secretase. Specifically speaking, secretases act on the amyloid precursor protein (APP) to cleave the protein into three fragments. Sequential cleavage by β-secretase (BACE) and γ-secretase produces the amyloid-β (Aβ) peptide fragment that aggregates into clumps called "plaques" in the brains of Alzheimer's disease patients. If α-secretase acts on amyloid precursor protein (APP) first instead of β-secretase (BACE), no amyloid-

β (Aβ) is formed because α-secretase recognizes a target protein sequence closer to the cell surface than β-secretase (BACE). The non-pathogenic middle fragment formed by an α-secretase and γ-secretase cleavage sequence is referred to as P3.

However, together with β-secretase, γ-secretase will cleave the amyloid precursor protein by creating peptides of various lengths, namely Aβ40 and Aβ42. The longer one of Aβ40 and Aβ42 is more aggregating than the other one. The ratio of Aβ42 over Aβ40 is increased by familial Alzheimer's disease mutations occurring in the presenilin genes or in amyloid precursor protein, near the γ-secretase cleavage site. Thus, γ-secretase is able to determine the solubility of the created Aβ fragment and attract a wide popular attention in pharmaceutical drug industry.

The structure of the three secretases varies widely. The α-secretase gene has not been conclusively identified but is believed to be a metalloproteinase. β-secretase (BACE) is a transmembrane protein with an extracellular aspartic acid protease domain.

γ-secretase is an integral membrane protein, that cleaves single-pass transmembrane proteins at residues within the transmembrane domain. It is a protein complex containing presenilin, nicastrin, ACH-1, and PEN-2 and its biological mechanism remains under extensive research interests. Presenilin is believed to harbor the protease domain and represents an important example of a rare type of protease that cleaves targets within the cell membrane.

Several successful γ-secretase inhibitors have been discovered by screening drug libraries and designing aspartyl protease transition-state analogues based on the amyloid precursor protein (APP) substrate cleavage site. Most of these compounds are very general, not specific enough for γ-secretase cleavage site of amyloid precursor protein (APP), and equally inhibit the processing of other γ-secretase substrates, such as Notch and other cell-surface proteins and receptors involved in embryonic development, haematopoiesis, cell adhesion and cell/cell contacts.

Thus, it's more desirable for research scientists to focus on finding compounds that show selectivity for amyloid precursor protein (APP) cleavage and particularly inhibit the formation of the aggregating form of Aβ42. Compounds that target the substrate docking site instead of the enzyme active site could be another featured point worth of investigation as an alternative strategy. The finding that some NSAID analogues preferentially inhibit the formation of Aβ42 over Aβ40 and do not affect Notch processing has opened a new therapeutic window. The progress in design of selective inhibitors as well as recent results obtained in animal studies prove that γ-secretase remains among the best targets for the therapeutic control of amyloid build-up in Alzheimer's disease. The full understanding of γ-secretase regulation may yet uncover new therapeutic interests.

Besides the involvement in the pathogenesis of Alzheimer's disease, these proteins also have other functional roles in the cell. γ-secretase plays an essential role in developmental signalling by the transmembrane receptor Notch, freeing the cytoplasmic tail of Notch to travel to the cell nucleus to act as a transcription factor. Although β-secretase (BACE) cleaves the extracellular domains of several transmembrane proteins, its physiological function remains unknown.

1.3 Contributions

Many biochemists and biological scientists have tried to narrow down a way to determine the γ-secretase cleavage site mechanisms based on their primary, secondary, tertiary and quaternary structures. In this thesis, we are able to determine the consensus compositional sequences or functional domain of γ-secretase and its substrates as a possible therapeutic intervention. Then, we find the candidates of consensus sequence and functional domain by the methods of bioinformatics and collaborate to use traditional biological methods to verify the critical role of candidates including protein purification, recognition and domain functional interaction.

We also hope that we can apply the sequence comparison techniques and identify the candidate motifs and domains for all components of γ-secretase and its substrates. Meanwhile, we plan to biologically knockout and verify the functionality of those identified motifs and domains including protein purification, recognition and interaction. Finally we evaluate the efficacy of the bioinformatics tools and cast the direction of prospective improvements.

1.4 Organization

The thesis is organized as follows; Chapter 1 gives an overall introduction. Chapter 2 reviews the literature and related work. We present the typical genetic structures. Then, we go over the current methods in γ-secretase identification. Then, we describe the relation among amyloid-β (Aβ) and γ-secretase. Chapter 3 presents our primary methodology. We describe the bioinformatics approach to obtain the motifs and domains for amyloid-β (Aβ) and γ-secretase. We give evidence for the idenfitication. Chapter 4 presents the results of the application of our methodology. We conclude the thesis with a summary of the proposed work and future directions in Chapter 5.

2. Literature Review

2.1 Alzheimer's Disease hypothetic mechanisms

There are three major competing hypotheses on the possible cause of the Alzheimer's disease. The oldest, on which most currently available drug therapies are based, is known as the "cholinergic hypothesis" and suggests that Alzheimer's disease is due to reduced biosynthesis of the neurotransmitter acetylcholine. The medications that treat acetylcholine deficiency have served to only treat symptoms of the disease and have neither halted nor reversed it (Walker et al 2006). The cholinergic hypothesis has not maintained widespread support in the face of this evidence, although cholingeric effects have been proposed to initiate large-scale aggregation (Shen 2004) leading to generalized neuroinflammation (WenK 2006).

Recent research includes hypotheses centered on the effects of the misfolded and aggregated proteins, amyloid-β and amyloid-τ. The two positions differ, with one stating that the amyloid-τ protein abnormalities initiate the disease cascade, while the other states that amyloid-β (Aβ) deposits are the causative factor in the disease (Mudher et al 2002). The amyloid-τ hypothesis is supported by the long-standing observation that deposition of amyloid plaques do not correlate well with neuron loss (Schmitz et al 2004). However, a majority of researchers support the alternative hypothesis that amyloid-β (Aβ) is the primary causative agent (Mudher et al 2002).

The amyloid hypothesis is initially compelling because the gene for the amyloid precursor protein (APP) is located on chromosome 21, and patients with trisomy 21 (Down Syndrome) who thus have an extra gene copy almost universally exhibit Alzheimer's disease-like disorders by 40 years of age (Nistor et al 2006, Lott et al 2005). The traditional formulation of the amyloid hypothesis points to the cytotoxicity of mature aggregated amyloid fibrils, which are believed to be the toxic form of the protein responsible for disrupting the cell's calcium ion homeostasis and thus inducing apoptosis

(Yankner et al 1990). A more recent and widely supported hypothesis suggests that the cytotoxic species is an intermediate misfolded form of amyloid-β (Aβ), neither a soluble monomer nor a mature aggregated polymer but an oligomeric species (Blanchard et al 2000). Relevantly, much early development work on lead compounds has focused on the inhibition of fibrillization (Blanchard et al 2004, Porat et al 2006, Kanapathipillai et al 2005), but the toxic-oligomer theory would imply that prevention of oligomeric assembly is the more important process (Lee et al 2005) or that a better target lies upstream, for example in the inhibition of amyloid precursor protein (APP) processing to amyloid-β (Aβ) (Espeseth et al 2005).

It should be noted further that ApoE4, the major genetic risk factor for Alzheimer's disease, leads to excess amyloid build up in the brain cortex and hippocampus before Alzheimer's disease symptoms arise. Thus, amyloid-β (Aβ) deposition precedes clinical Alzheimer's disease (Polvikoski et al 1995). Another strong support for the amyloid-β (Aβ) hypothesis, which looks at amyloid-β (Aβ) as the common initiating factor for Alzheimer's disease, is that transgenic mice solely expressing a mutant human amyloid precursor protein (APP) gene develop first diffuse and then fibrillar amyloid plaques, associated with neuronal and microglial damage (Games et al 1995, Masliah et al 1996, Hsiao et al 1996).

The identification of the amino acid sequence for amyloid-β (Aβ) allowed cloning of its cognate protein, amyloid precursor protein (APP), and further confirmed it was a type I membrane protein resembling cell surface receptors. The majority of the current research is based on the "amyloid cascade hypothesis". It states that accumulation of amyloid-β (Aβ) peptide is the initiating event that triggers neuro-degeneration in both sporadic Alzheimer's disease cases and familial Alzheimer's disease cases. The accumulation is triggered by either overproduction, mutated processing, or mutated clearance of the amyloid peptide, and consequently leads to self association of amyloid-β (Aβ) as neuro-toxic fibrils. In recent years, genetic researches, which aim to link familial Alzheimer's disease mutation reported so far to the proteolytic events that generate amyloid-β (Aβ), further confirm the amyloid-β (Aβ) cascade hypothesis. There are plenty of evidences

which show that all the amyloid precursor protein (APP) pathogenic mutations are clustered near the β-secretase and γ-secretase cleavage sites (Evin et al 2002, St George-Hyslop 2000). Particularly, presenilin genes that encode essential components of the γ-secretase proteolytic complex, have the highest probability of familial Alzheimer's disease occurrence rates.

2.2 γ-Secretase

The γ-secretase proteolytic complex consists of four integral membrane proteins, namely presenilin, nicastrin, APH-1 and PEN-2. Presenilin includes PS1 and PS2 proteins. From a structural perspective, presenilin contains eight or nine transmembrane domains and two hypophlic aspartate residues in transmembrane domains which are suggested to form the catalytic site of the enzymic complex. Nicastrin consists of a single transmembrane domain and is highly glycosylated in the γ-secretase proteolytic complex. It has been reported that the DYIGS motif plays an essential role for the recognition of the γ-secretase substrates. APH-1 consists of seven transmembrane domains. The homologous proteins of APH-1a and APH-1b are the typical forms of APH-1 protein. PEN-2 is the presenilin enhancer, which consists of two transmembrane domains and is an integral component for the endoproteolysis of presenilin.

APH-1, PEN-2, presenilin, and nicastrin exist in at least three major states during initial assembly and subsequent maturation of the active γ-secretase proteolytic complex, the initial unassembled immature components, the immature complex and the mature and active complex. Biological studies suggest that APH-1 functions as a molecular scaffold for γ-secretase proteolytic complex. Transition of the immature complex to mature complex is associated with a presenilin endoproteolysis triggered conformational change. Motif mutations prevent APH-1 from performing its scaffolding role in initial assembly of the immature complex. APH-1, PEN-2, presenilin and nicastrin have multiple contacts with one another.

2.2.1 Presenilin

Initially found through linkage studies using mutations present in familial Alzheimer's cases in 1995 (Sherrington et al 1995), presenilins are a group of related multiple transmembrane proteins that function as a part of the γ-secretase proteolytic complex. Vertebrates species have two presenilin genes, namely PSEN1, located on chromosome 14 in humans, that encodes presenilin 1 (PS-1) and PSEN2, on chromosome 1 in humans, that codes for presenilin 2 (PS-2). Both genes contain conservation among species, with little difference between rat and human presenilins. The Caenorhabditis elegans has two genes that represent the presenilins and appear to be functionally similar, sel-12 and hop-1 (Smialowska et al 2006).
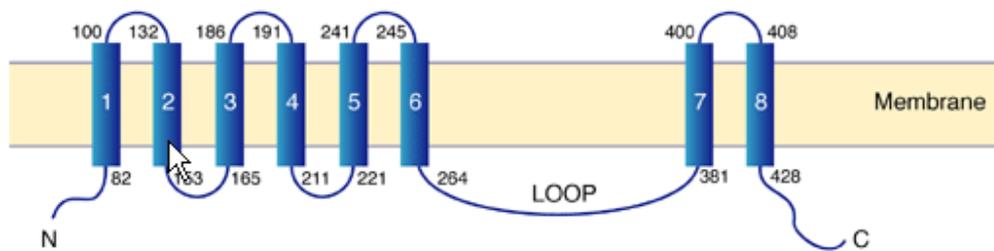
Figure 3. Presenilin-1 structural model. Presenilins will be cleaved in the α helical region of one of cytoplasmic loops between transmembrane domain 6 and transmembrane domain 7 to produce a larger N-terminal and a smaller C-terminal fragment which will create part of the functional protein.

Presenilins will be cleaved in the α helical region of one of cytoplasmic loops between transmembrane domain 6 and transmembrane domain 7 to produce a larger N-terminal and a smaller C-terminal fragment which will create part of the functional protein. Cleavage of presenilin 1 can be prevented by a mutation which causes the loss of exon 9,

and consequently results in loss of functionality. Mutations in the presenilin proteins are known to cause early onset Alzheimer's disease through mechanisms which are of great interests to scientists.

Recent research has produced a more widely accepted model with respect to the structure of presenilin-1. When initially investigated, the presenilin-1 gene was subjected to hydrophobicity analysis which predicted that the protein would contain ten trans-membrane domains. All previous models agreed that the first six putative transmembrane spanning regions cross the membrane. These regions correspond to the N-terminal fragment of presenilin-1 but the structure of the C-terminal fragment had been in investigation for an extended period of time. Due to the nature of presenilin-1with large numbers of hydrophobic regions, it is inconvenient for x-ray crystallography to determine a definitive proof of the three dimensional structure.

Generally, speaking, a significant percentage of the cases of Alzheimer's disease are not hereditary. However, there are a small percentage of cases that have an earlier age of onset and have a strong genetic influence. From patients of Alzheimer's disease, mutations in the presenilin proteins either presenilin-1 or presenilin-2 or the amyloid precursor protein (APP) can be tracked down. A significant percentage of these cases consists of mutant presenilin genes, either presenilin-1 or presenilin-2. An important part of the disease pathogenesis in Alzheimer's disease is the accumulation of Amyloid-β (Aβ) protein. To form Amyloid-β (Aβ) protein, amyloid precursor protein (APP) must be cleaved by two secretase enzymes, β-secretases and γ-secretase. Presenilin is the underlying substrate of γ-secretase that is responsible for the cleavage of amyloid precursor protein (APP) by γ-secretase.

γ-secretase can cleave amyloid precursor protein (APP) at several points within a small region of the protein which results in Amyloid-β (Aβ) protein of various lengths. The lengths associated with Alzheimer's disease are normally 40 and 42 amino acids long. Aβ-42 is more likely to aggregate to form plaques in the brain than Aβ-40. Presenilin mutations lead to an increase in the ratio of Aβ-42 produced compared to Aβ-40,

although the total quantity of Amyloid-β (Aβ) protein produced remains constantly unchanged (Citron et al 1997, Bentahir et al 2006).

2.2.2 Nicastrin
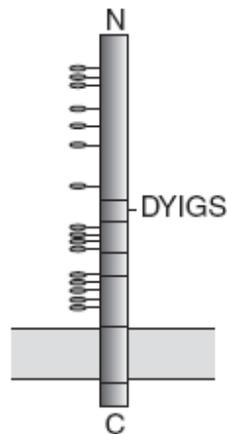
Figure 4. Nicastrin structural model. Nicastrin is an integral component of the γ-secretase proteolytic complex involved in processing amyloid precursor protein (APP) to the Alzheimer's disease associated peptide Amyloid-β (Aβ) peptides.

Nicastrin, or abbreviated as NCT, is a protein that is an integral component of the γ-secretase proteolytic complex, which is one of the proteases involved in processing amyloid precursor protein (APP) to the Alzheimer's disease associated peptide Amyloid-β (Aβ) peptides. The other proteins involved in the proteolytic complex are presenilin, which is the catalytically active component of the γ-secretase proteolytic complex, APH-1, and PEN-2 (Kaether et al 2006). Nicastrin by itself is not catalytically active, but has the capacity to promote the maturation and proper trafficking of the other proteins in the γ-secretase proteolytic complex. All of them will experience significant post-translational modification before becoming catalytically active in the cell (Zhang et al 2005). Besides

the role in amyloid precursor protein (APP) pathway, nicastrin has also been identified as a regulator of neprilysin, an enzyme involved in the degradation of Amyloid-β (Aβ) protein fragment (Pardossi-Piquard et al 2006).

2.2.3 APH-1
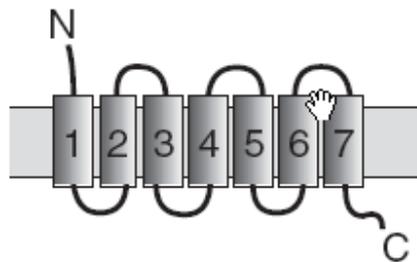
Figure 5. Anterior pharynx-defective 1 structural model. APH-1 is a protein gene product originally identified in the Notch signaling pathway in Caenorhabditis elegans as a regulator of the cell surface localization of nicastrin.

Anterior pharynx-defective 1, or abbreviated as APH-1, is a protein gene product originally identified in the Notch signaling pathway in Caenorhabditis elegans as a regulator of the cell surface localization of nicastrin (Goutte et al 2002). APH-1 homologs in other organisms, including humans, have since been identified as components of the γ-secretase proteolytic complex along with the catalytically active presenilin and the regulatory nicastrin and PEN-2. The γ-secretase proteolytic complex is a multimeric protease responsible for the intracellular proteolysis of transmembrane proteins such as the Notch protein and amyloid precursor protein (APP). γ-secretase cleavage of amyloid precursor protein (APP) is one of two proteolytic steps necessary to generate the peptide of Amyloid-β (Aβ). The Amyloid-β (Aβ) misfolded form is implicated in the pathogenesis of Alzheimer's disease (Kaether et al 2006). All of the

components of the γ-secretase proteolytic complex experience extensive post-translational modification, especially proteolytic activation. APH-1 and PEN-2 are considered as regulators of the maturation process of the catalytic component presenilin (Luo et al 2003). APH-1 consists of the α helical interaction motif glycine-X-X-glycine (GXXXG) that is essential to both assembly of the γ-secretase proteolytic complex and to the maturation of the components (Lee et al 2004).

2.2.4 PEN-2
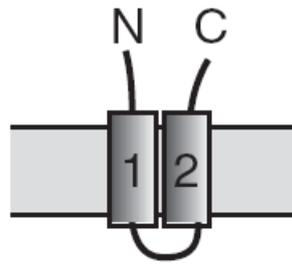
Figure 6. Presenilin enhancer 2 structural model. PEN-2 is a protein that is a regulatory component of the γ-secretase proteolytic complex responsible for proteolysis of amyloid precursor protein (APP).

Presenilin enhancer 2, abbreviated as PEN-2, is a protein that is a regulatory component of the γ-secretase proteolytic complex, the one responsible for proteolysis of transmembrane proteins such as the Notch protein and amyloid precursor protein (APP). The γ-secretase proteolytic complex consists of PEN-2, APH-1, nicastrin, and the catalytic subunit presenilin (Kaether et al 2006). PEN-2 is a 101 amino acid integral membrane protein with a topology structure such that both the N-terminus and the C-terminus face first the lumen of the endoplasmic reticulum (ER) and later the extracellular environment (Francis et al 2002). Numerous evidences have shown that a conserved sequence motif D-Y-L-S-F at the C-terminus, as well as the overall length of

the C-terminal tail, is required for the formation of an active gamma secretase complex (Hasegawa et al 2004).

2.3 Notch

Discovered in 1917 by Thomas Hunt Morgan when it was first noticed in a strain of the Drosophila melanogaster, its molecular analysis and sequencing has been widely studied in the 1980s. The Notch signaling pathway is a highly conserved cell signaling system in most multicellular organisms. Rather than crossing the membrane multiple passes between the extracellular and intracellular spaces, the notch receptor is a single-pass transmembrane receptor protein.

2.3.1 Mechanism

Besides the cleavage role of Amyloid-β (Aβ) peptides formation in γ-secretase proteolytic complex, presenilins are also implicated in the processing of notch, an important developmental protein. Mice that have the PS1 gene knocked out die early in development from developmental abnormalities similar to those found when notch is disrupted (Shen et al 1997). Together with nicastrin, APH-1 and PEN-2, these four proteolytic complex is essential for Notch signaling pathway in Caenorhabditis elegans embryos and is required for intramembrane proteolysis of Notch and β-amyloid precursor protein in mammalian cells.

The Notch protein sits like a trigger spanning the cell membrane, with part of it inside and part outside. Ligand proteins binding to the extracellular domain induce proteolytic cleavage and release of the intracellular domain, which enters the cell nucleus to alter gene expression.

2.3.2 Function

The Notch signaling pathway is essential for cell-cell communication. Notch signaling has also been involved in several processes that control multiple cell differentiation. Notch signaling is required for physiological angiogenesis and may promote tumor angiogenesis. A variety of strategies for blocking Notch signaling, in particular γ-secretase inhibition, are discussed as potential therapies for breast cancer and tumor angiogenesis. During the formation of the valve primordial and ventricular development and differentiation, notch regulates crucial cell communication events between endocardium and myocardium (Joaquín et al 2007). In the developing cardiac cushion, Notch signaling plays an essential role in the activity of the cardiac valve homeostasis and has implications in other human disorders involving the cardiovascular system. Meanwhile, notch signaling is often repressed in many cancers, and faulty Notch signaling is implicated in many diseases.

2.3.3 Notch Cascade

Because most ligands are also transmembrane proteins, the receptor is normally only triggered from direct cell to cell communication. Groups of cells can arrange themselves in such a way that if one cell expresses a given character, it can be switched off in neighbor cells by the intercellular Notch signal. This is how the majority of cells communicate one another to make large structures.

The Notch cascade consists of Notch and Notch ligands, as well as intracellular proteins transmitting the Notch signal to the cell's nucleus. The Notch receptor family is involved in the specification of cell fates during development in Drosophila and C. elegans.

2.4 A Bioinformatic Solution

In the research area of computational biology, sequence alignment is a method to derive the arrangement of sequences of DNA, RNA and protein sequence on the symbol-wise base-pair comparison basis so that the maximum similarity between sequences can be discovered. The biological implication is that we could now identify potential functional, systematic and phylogenetic relationships between the sequences. Under most circumstances, a matching equation with some data score value is used to measure how well two or more residue sequences can be aligned. The matching equation is related to the total occurrence of mismatch symbols or match symbols either on a percentage basis or an absolute score value with specific emphasis on special matching or mismatching. Gaps are defined as the blank spaces inserted between the nucleotide or amino acid residues to justify or align them with optimal matching score value. There are several existing approaches to solve either the global sequence alignment or the local sequence alignment. Theoretically speaking, the comparison between the test sequences and the sequence corpus database might uncover the potential motifs and domains within the test sequences. However, numerous evidences indicate there are more complicated mechanisms behind the wall. A straightforward alignment might not thoroughly reveal the background information.

Increasing numbers of observations and experiments indicate the inactivation of a certain gene has no major apparent phenotype and fitness changes of the species. In certain circumstance, the functionality of a gene can be seen as a replacement of another. The nature favors the redundancy of the genes. From an engineer's point of view, nature ensures that the major functionalities are substantially backed up in case of mechanical failure.

The mutation or deletion of a single gene often has little phenotypic effects due to the existence of the duplicate genes and the alternative metabolic pathways, regulatory networks. Particularly, multiple pathways and multiple genes can encode a specific enzyme in a pathway. In the case of a deletion or mutation of some gene, these

17

mechanisms could compensate the loss of the functionality via another copy of the same genotypic contents. For S. cerevisiae, at least a quarter of those gene deletions that have no phenotype are compensated by the duplicate genes. The necessity of the duplicate genes plays an imperative role towards the robustness of the genes against null mutations, a mutation with no functional change. In Drosophila, seven of the neurogenic phenotype deletions are extremely similar at the transcription and sequence level sharing the basic motif of transcriptional regulators. They are non-essentially individually, suggesting the functional redundancy of the encoded gene products.

Redundancy is wide spread in genomes of higher organisms. They are performing the same function and the inactivation of one of these genes has little or no effect on the biological phenotype. They present 4 cases that could explain why the genetic redundancy is common. A series of evidence show that redundancy itself appears to be a historical consequence of polyploidy in the vertebrate common ancestor.

2.5 Redundancy and Immunology

Just like redundancy is an imperative element in our routine daily life, it also performs a substantial role in our human mass body. Numerous evidences indicate that homologous genes and gene products often have redundant physiological functions. Complement is an integral part of innate immunity. It functions throughout the body, including the oral cavity. Complexity and redundancy are important to efficiently kill a wide range of pathogens, yet not cause destruction of host tissues.

Increasing numbers of observations and experiments indicates the inactivation of a certain gene has no major apparent phenotype and fitness of the species. In certain circumstance, the functionality of a gene can be seen as a replacement of another. The nature favors the redundancy of the genes.

The wide distributed redundancy of mast cells and macrophage plays an important role in the inflammation and tumor cytotoxicity. It helps human body to recover from the bacteria infections. For S. cerevisiae, at least a quarter of those gene deletions that have no phenotype are compensated by the duplicate genes (Gu et al 2003). The necessity of the duplicate genes plays an imperative role towards the robustness of the genes against null mutations, a mutation with no functional change.

Based on the comparative similarity of the induction kinetics of Nur77 and Nor-1 in T cells, the functional redundancy mechanism of the Nor-1 protein is used to explain the lack of a phenotype in the Nur77 mutant mice (Cheng et al 1997).

The phenotype of the knock-out mice has been studied by analyzing exhaustively all the possible genotypic double mutant combinations (Heber et al 2000) so as to link the relation between complementation and redundancy. They successfully identified the redundancy between APLP2 , a member of the Amyloid Precursor Protein (APP) involved in Alzheimer's disease, and both other family members and corroborate a key physiological role for APLP2. It had been confirmed that the proteins may serve primarily overlapping functions and complement to each other (Koch et al 1997, Steinbach et al 1998, White et al 1998).

The mutation or deletion of a single gene often has little phenotypic effects due to the existence of the duplicate genes and the alternative metabolic pathways, regulatory networks (Gu et al 2003). Particularly, it has been confirmed that, in Drosophila, seven of the neurogenic phenotype deletions are extremely similar at the transcription and sequence level sharing the basic motif of transcriptional regulators (Maier et al 1993). They are non-essentially individually, suggesting the functional redundancy of the encoded gene products. A series of evidence has proved that redundancy itself appears to be a historical consequence of polyploidy in the vertebrate common ancestor (Gibson et al 1998).

3. Methodology

3.1 Overview

Alignment of multiple biological sequences is one of the central problems in computational biology. Multiple sequence alignment has applications in various problems in biology, and can be applied respectively on DNA, RNA or protein sequences. Example applications can be found in whole-genome sequencing, identifying conserved regions and regulatory elements in related genomes, determining important regions for cite-directed mutatgenesis, construction of and analysis of phylogenic trees, phylogenic footprinting, etc. An alignment of the sequences is obtained by inserting gaps at chosen positions in each of the sequences, such that the resulting sequences all have the same length. The sequences are thus arranged into an array of symbols, whereby no column is allowed to contain only gaps. The goodness of an alignment is measured by the alignment value, which is determined using a scoring function. The problem is therefore to determine the optimal alignment. A multiple sequence alignment will result in a minimal alignment value. The result is that the sequences are now arranged such that the maximal similarity between them are now exposed, and can be exploited for various applications.

The problem of computing the optimal alignment for multiple sequences is known to be NP-complete, and thus various heuristics and approximations have been proposed. Most methods are typically based on either global alignments, or local alignment. Hybrid approaches aimed at integrating both local and global methods have also been reported. Example popular alignment programs include CLUSTAL W and its family, T-COFFEE, DIALIGN and its derivatives, MUSCLE, Align-m, GAME, etc. Gusfield and Mount provide a detailed discussion on the general problem of sequence alignment.

We propose a sort based method for constructing multiple sequence alignments. Our motivation is the recent result in linear time direct suffix sorting, whereby suffix arrays can be constructed in linear time and space in the worst case, without the need to first construct suffix trees. This has important implications for both the time required for computing the alignment, and the space needed for constructing the suffix arrays. We perform alignment by first identifying a set of anchor points, based on the suffix sorting output on the sequences. Final alignment and gap considerations can be made by a recursive application of the sort-based anchor point algorithm.

An important novelty in our approach is how biological mutation information is incorporated in each refinement step. Thus, the recursive alignment refinement is performed via sorting and alignment between anchor points using a biological mutated version of the original sequences. This sort-mutate-anchor paradigm separates our proposed sequence alignment algorithm from other previously proposed techniques.

In the next section, we briefly describe related work on sequence alignment, with emphasis on anchor-based methods. The following section presents a brief overview on linear time direct suffix sorting, the basis of our approach. Our algorithm is presented, along with an analysis of its results.

3.2 Methods

An approach that has been used to reduce the complexity of multiple sequence alignment is by the use of selected anchor points along the sequences. An anchor point is defined by the sequences involved in the anchor, the starting position of the anchor in the sequences, and the length of the anchor. In some situations, especially with user-defined anchors, the anchors are given priorities, which can help the algorithm in resolving consistency between the anchors. Anchor priorities can also be used to determine which anchor points should be chosen first, where there are multiple anchor points in the same region. Anchors help to restrict the search required in performing the alignment, by constraining the required search to only regions between two anchor points. Clearly, the performance of anchor-based algorithms depends critically on how the anchor points are chosen, and

to some extent, on how the alignment between anchor points is performed. The sequences involved in the alignment are represented as horizontal lines. The vertical lines indicate where an anchor has been established between sequences. Solid vertical lines represent strong anchors, while the dotted lines depict weak anchors.

Various anchor-based methods have been proposed. The methods differ in the way the anchor points are defined, in how they are computed, whether or not priorities are assigned, and in how the remaining sequences between anchors are aligned. The clusters observed in the sorted common prefix (SCP) data structure were proposed for determining regions for anchor point computations, using the notion of sphere of influence for the anchor points.

A maximum unique match (MUM) is a subsequence that occurs only once in each of the sequences involved in the alignment, and it is not contained in any other MUM. Here, the MUMs are used to form anchor points for the alignment. The use of MUMs follows the idea used in popular alignment programs such as BLAST and FASTA. Conceptually, if there is an MUM between the input sequences, it is most likely to be part of the global alignment. One problem with MUM-based alignment is that it does not work well in aligning sequences from relatively distant species. The basic idea has been modified, for instance, using criteria different from MUMs. Example, GAME used maximal exact matches (MEMs), rather than MUMs, while in LAGAN, Brudno et al used the idea of anchors, whereby anchor points are defined based on inexact matches.

In ACANA, anchors are determined by using the edit matrix produced during dynamic programming. By choosing near-optimal local alignments along the diagonals of the dynamic programming matrix, they recursively determine the best anchors for global alignment for each region, which are then fixed for the region. Final alignment is produced by connecting each region with the fixed anchor points.

Another way to view the anchors is in terms of constraints on the sequences to be aligned. Such constraints could be user-defined, or could be machine generated based on some

application dependent information. Myers et al proposed methods for progressive alignment incorporating sets of constraints, and studied the issue of consistency between constraints. Brown and Hudek proposed a tree-based method for anchor-based progressive alignment using spaced or seeded anchors. Spaced anchors are constraints in the alignment defined in terms of a binary pattern over the sequences, such that matches or mismatches can be allowed at a given position along the sequences, based on whether we have say a 1 or 0 at the given position. Sequence alignments using explicit user defined anchors were studied.

While user-defined constraints can be quite important in some applications, this makes the methods less flexible, especially in using the algorithms on applications different from their original purpose. Moreover, for very large sequences, for instance, whole genomes, or when many sequences are involved, defining these manual anchor points may become difficult. Automated anchor point determination will thus be needed in such situations. Such automated methods should be able to provide options for augmenting the automatically determined anchor points with user-specified anchors. Similar ideas can be used to incorporate further expertise, for instance, knowledge of locally conserved protein secondary structures, as constraints in the alignment.

3.2 The Problem and Main Contribution

The computational problem in multiple sequence alignment is often due to the model used for inducing gaps during the alignment. This is because the optimal alignment often requires an exhaustive evaluation of all the possibilities for gap insertion in the sequences being aligned. In fact, from one view point, the problem of sequence alignment can be viewed as that of appropriate gap induction within the sequences involved. For anchor-based alignment schemes, an additional problem is the selection of anchor points, and the development of appropriate anchor scoring functions to be used.

The major contribution of this work is a mechanism to cut down on the exhaustive search required for gap induction. The method we propose is anchor-based, whereby anchors are

determined based on the sorted suffixes and position-based analysis. In particular, we propose a sort-based alignment algorithm, using methods of linear-time worst-case suffix sorting as our basic building block. The sorts implicitly expose where alignments should be anchored, and which regions are not likely to form an alignment, thereby eliminating the need for exhaustive gap insertion and alignment score evaluation. We also develop a new anchor-point evaluation scheme based on the sorted suffixes, position-based sorts, and a proposed mutation mapping scheme.

After the initial sorting on the original sequences, we extend the sort-based approach using a novel mutation mapping scheme. Here, symbols are mutated based on a mutation map derived from standard mutation matrices, and the sorting procedures (suffix sorting and position-based sorts) on symbols between anchors are now performed on these mutated symbols. New anchor points are then determined within the previous anchors. Fixing the anchor points implies that subsequent steps can not change these anchors. Later alignment steps can only be performed on symbols within the same anchor region. Operations on symbols in two different anchor regions are forbidden. Sequence alignment is then performed as a recursive application of this sort-mutate-anchor paradigm on the symbols between each pair of anchor points.

3.3 Motif Determination

We propose a sort-based multiple sequence alignment algorithm. Based on the observation that a well matched alignment is also a rearrangement of highly repetitive segments and non-repetitive region across all the source sequences, the algorithm transforms the alignment for multiple sequences into a long range repetitive pattern identification problem where the advantage of the automated suffix sorting technique can be fully exploited. Since suffix sorting could reveal the repetition information across all sequences, we can home on the anchor points with the aid of suffix arrays. By constructing both strong and weak connections, the algorithm builds restraints to cross and align different sequences by their anchor points. By recursive repetition of the

procedure using information from biological substitution matrices, the global alignment for multiple sequences can be obtained.

3.3.1 Overview

We provide a generalized outline of the proposed algorithm. It includes subroutines of the source sequence concatenation, suffix sorting, anchor points identification and biological substitution matrices decomposition. The proposed algorithm takes advantage of suffix sorting by converting inter-sequences vertical alignment problem to one dimension horizontal repetitive pattern identification using suffix sorting. The general idea is to concatenate multiple source sequences into a simple sequence with special end-of-sequence symbols, then perform the suffix sorting. This sorting operation exposes the repetitive segments across the source sequences. We evaluate the suitability of these repetitive segments for anchor point's formation by computing their anchor scores. The segments with anchor scores above a threshold are marked as potential anchor regions. The starting position of each potential anchor region is a potential anchor point. The maximum valued anchor points are first selected and fixed onto the blank backbone sequences. Recursively, a less tightly correlated anchor point identification subroutine is performed between those previous identified anchor regions until all the anchor regions are exhausted.

3.3.2 Motif Selection

A one-to-one mapping relation has been noted from the sequence concatenation operation. The suffix sorting routine is then applied on the simple sequence. This will produce the suffix which contains all the suffixes, sorted according to their alphabetical orders. Essentially, input sequence will be partitioned, whereby each group contains suffixes that start with the identical symbol. This will create the re-sorted motif arrays. Given the one-to-one mapping, each suffix can be marked.

A sliding window is introduced to elaborate the anchor point selection process. The sliding window is defined as a covered region such that all suffixes fall into this region

can be factored to determine the anchor point. This region is marked which indicates the boundary of the potential anchor points. Anchor score for anchor point within the region Q is computed as shown in the downward sliding. The sliding window is then shifted to the next suffix by incrementation, namely sliding down the window by one. The new sliding window covers the region. A similar computation of anchor score can be obtained as well. The values are listed to the right of each sliding. The anchor score computation steps are repeated until the sliding window reaches the bottom of each. The algorithm will proceed to the next group after it finishes the designated group until all groups are processed. The suffix sorting routine is applied on sequence. Suffix is marked as being sorted, based on the alphabetical order. Suffix is marked with its corresponding positions.

3.4 Motif Determination with Mutation

One of the disadvantages of direct anchor point sorting is that it only identifies the strongest anchor points across the source sequences. The implicit structural and functional significance of biological sequence cannot be fully exploited. We introduce the biological mutation matrices of BLOSUM and PAM to further refine the multiple sequence alignment results between two consecutive strong anchor points. These matrices calculate a log-odds score for all possible substitutions of the 20 standard amino acids with typical applications in sequence alignment and multiple sequence alignment. Both BLOSUM62 and PAM250 are two widely referred biological mutation matrices. BLOSUM62 is the calculated matrix by using the observed substitutions between proteins which have *62* or more sequence identity. PAM250 means a *250* point mutations/100 amino acids which gives the different rates of mutations between pairs of amino acids.

After obtaining the preliminary results, the mutation-based suffix sort will continue to further refine the segmental alignment between two consecutive pre-determined anchor points. The biological mutation matrix will be introduced at this stage. The matrix denotes the different rates of mutations between pairs of amino acids. Notice the matrix is

a symmetric matrix. We can group the edges. Because the symmetric property of matrix, we define the mutation from alphabetically greater symbol to alphabetically smaller. We define group to contain all mutated vertices where there are parallel edges between each other. The grouping table is to be used in the mutation-based suffix sorting routine. This grouping table depends on the biological mutation table only and it is independent of the input source sequences. The table will be produced once during the alignment procedure.

Because the strong anchor points have been fixed in the first suffix sorting routine, the mutation based suffix sorting will be applied on the regions between two consecutive strong anchor points in a decreasing order. The suffix sorting routine is applied on the regions after each group mutation operation. The mutation information will be written to a supplemental table in order to recover the mutated sequences back to the origin during the finalizing steps.

The suffix sorting routine is performed after the mutation. This will anchor the group in the regions. A suffix sort after the mutation then produces the anchor points. Continuously performing the mutation, the algorithm will reach a stage when no more anchor sites can be found to achieve a positive profit gain. At this stage, for any regions between two adjacent anchor sites, the blank gap spaces are always at the end of the segment because whenever the algorithm tries to align a local anchor site, it always to shift the anchor sites to the left and seek fitness and adaptation. Following the mutation information, the algorithm performs the mutation once for each unique value and stores the mutation information into the complementary table. The complementary table will be used to recover the alignment to the original input source sequences after all mutations have been finished. The mutation is performed in the decreasing order such that the symbols with stronger affinity will be anchored first.

Each subsequence is recursively aligned within its individual anchor region following the mutation based suffix sorting routine. After each regional alignment is achieved, they are concatenated based on the strongest anchor points. The final results are obtained through concatenation of the regional alignment results from anchor points.

The traditional bottleneck in alignment is the search for gap positions. Notice that we do not perform direct search for best positions for gap insertion at each stage. With our approach to gap induction, it means that the traditional problem of exhaustive search for nest positions for gap insertion has now been transformed to the problem of searching for best anchor points. The problem of best positions for gap insertion is handled implicitly by the sorting stage and the mutation-mapping procedure.

As the sliding window moves along with the anchor sites, some anchor points might be found to be closer to some previously fixed anchor sites on the consensus sequence. The proposed algorithm will meet the creation of both anchor sites by insertion of gaps. The induction of gaps avoids the possibility of assigning two anchor points on the same location of the consensus sequence. Whenever two anchor points are identified on the same location of the consensus sequence, the alphabetically smaller anchor point is fixed at the current location and the alphabetically greater anchor point is fixed at the location next to the current one to the right. The algorithm fills the corresponding space with gaps across different sequences.

3.5 Conclusion

We proposed a suffix sorting-based multiple-sequence alignment algorithm with anchor points. We simplify the alignment score functions so as to better reflect the actual implication. We analyze the complexity requirements of the proposed algorithm. We correctly test the proposed algorithm on benchmark sequence data so far. Meanwhile, we apply the mutation based alignment scheme into the current algorithm. The current algorithm treats each individual mismatch uniformly. However, we believe that introducing the mutation mapping will produce a more biologically-relevant alignment.

4. Motifs of γ-Secretase and Its Substrates

In this study, we employ the designed bioinformatic software packages to study the target motif substrates of γ-secretase and those relevant domains. As known in most publication, γ-secretase contains four types of components of Presenilin, APH-1, Pen-2 and Nicastrin. We apply the sequence scanning technique to all of these four component sets. We collect the sequence data for the four γ-secretase substrates from NCBI official database. Mutiple species sequence data are collected to cross examine the target motif substrates and domain information. Without loss of generality, we take Nicastrin as an example. Figure 7 is the motif identification results for Nicastrin [Homo Sapiens].
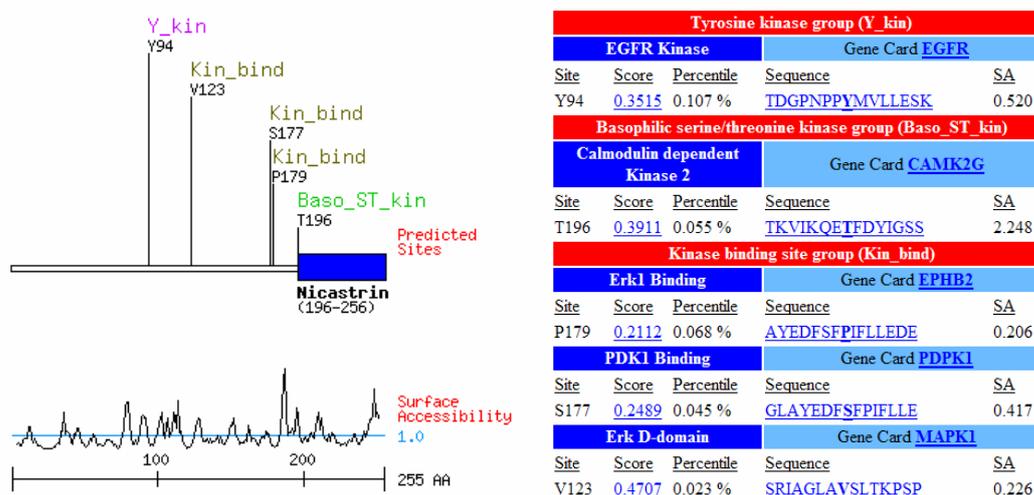
Figure 7. Speculative Motif sites of Nicastrin [Homo Sapiens].

We pass the Nicastrin [Homo Sapiens] to the motif finding program. There are 3 groups, totalling 5 motifs identified. Three groups include one motif of tyrosine kinase group, one motif of basophilic serine/threonine kinase group and three motifs of kinase binding site group. Their gene cards' ID codes are also maintained. Meanwhile, there is one Nicastrin(196-256) domain associates with the input sequence. Through the sequence scanning technique, we are able to speculate sets of Nicastrin target motif substrates for γ-secretase.

Meanwhile, we also apply the multiple sequence techniques to estimate the Nicastrin motifs identification. Through this method, we pass the Nicastrin [Homo Sapiens] domain sequences to the multiple sequence alignments and hidden Markov models covering many common protein domains and families. We are able to successfully identify the DYIGS groups, which are reported by the γ-secretase review paper. The DYIGS groups are marked in blue in the multuple sequence alignment table. Besides that, we also identify some potential motif sites, as marked in red in the following tables. We identify E will occur two spaces before DYIGS, forming E-DYIGS pair. E (Glutamic acid) is then speculated to be located two spaces before DYIGS. Glutamic acid might play a essential role for the regulatory reactions in γ-secretase target mitifs substrates.

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 260  Type: P  January 27, 2007 02:37  Check: 4069  ..

 Name: query_196-256        Len:    260  Check: 9654  Weight: 1.00
 Name: Q7KS07_DROME_236-469  Len:    260  Check: 6993  Weight: 1.00
 Name: Q7PY38_ANOGA_276-510  Len:    260  Check: 5861  Weight: 1.00
 Name: NICA_MOUSE_273-498   Len:    260  Check: 4890  Weight: 1.00
 Name: NICA_ARATH_251-460   Len:    260  Check: 9664  Weight: 1.00
 Name: Q6Z743_ORYSA_248-457  Len:    260  Check: 9567  Weight: 1.00
 Name: NICA_CAEEL_278-508   Len:    260  Check: 7440  Weight: 1.00

//

                     1                                                  50
query_196-256        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Q7KS07_DROME_236-469 KFILVTCRLD TTTMFDGVGL GAMDSLMGFA VFTHVAYLL. ...KQLLPPQ
Q7PY38_ANOGA_276-510 RIVLVSTRTD TTTMFDGVGL GAMDSVVPFA VLVAVAHFL. ...SQALP..
NICA_MOUSE_273-498   RVVVAATRLD SRSFFWNVAP GAESAVASFV TQLAAAEAL. ...HKA.PDV
NICA_ARATH_251-460   PVVLTVASMD TASFFRDKSF GADSPISGLV ALLGAVDAL. ...SRV.DGI
Q6Z743_ORYSA_248-457 PIIMVTASQD SASFFRDRSL GADSPISGLI ALLTAVDAL. ...SHL.HDI
NICA_CAEEL_278-508   KYMMVTARMD SFGMIPEISV GEVSVLTSII SVLAAARSMG TQIEKWQKAS

                     51                                                 100
query_196-256        ~~~~~~~~~~ ~~~~ETFDYI GSSRMVYDME KGKFP..... .......VQL
Q7KS07_DROME_236-469 SKDLHNVLFV TFNGESYDYI GSQRFVYDME KLQFP....T ESTGTPPIAF
Q7PY38_ANOGA_276-510 SND.RNVLFL FFNGESYDYI GSQRFVYDLQ TGAFP....S RGTQTKPISM
NICA_MOUSE_273-498   TTLSRNVMFV FFQGETFDYI GSSRMVYDME NGKFP..... .......VRL
NICA_ARATH_251-460   SNLKKQLVFL VLTGETWGYL GSRRFLHELD LHSDA..... ....VAGLSN
Q6Z743_ORYSA_248-457 SNLKKQLVFA VFNGEAWGYL GSRKFLQELD QGADS..... ....VNGISS
NICA_CAEEL_278-508   NTSNRNVFFA FFNGESLDYI GSGAAAYQME NGKFPQMIRS DRTHIHPIRP

                     101                                                150
query_196-256        ENVDSFVELG QVALRTSLE. ..LWMHTDPV SQKNESVRNQ ~~~~~~~~~~
Q7KS07_DROME_236-469 DNIDFMLDIG TLDDISN... ..IKLH.... ALNGTTLAQQ ILERLNNYAK
Q7PY38_ANOGA_276-510 DNIELMIDLG ALDNLTD... ....LHVY.. HAAPQPMATK VAELLRKINQ
NICA_MOUSE_273-498   ENIDSFVELG QVALRTSLD. ..LWMHTDPM SQKNESVKNQ VEDLLATLEK
NICA_ARATH_251-460   TSIETVLEIG SVGKGLSGGI NTFFAH.... KTRVSSVTNM TLDALKIAQD
Q6Z743_ORYSA_248-457 LLIDQVLEIG SVGKAISQGY PLFYAH.... AAGNSSISMK MVDALQSASE
NICA_CAEEL_278-508   NELDYILEVQ QIGVAKGRK. ..YYVHVD.. GERYQQNKTQ TDRVIDRIER

                     151                                                200
query_196-256        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

```
Q7KS07_DROME_236-469    SPRYGFNLNI .QSEMSAHLP PTSAQSFLRR DPNFNALILN ARPT...NKY
Q7PY38_ANOGA_276-510    QFGFGIQAGQ PVMTTN..LP PVSAQSFLRE NITFPAVIVA SRPA...NRF
NICA_MOUSE_273-498      SGAGVPEVVL RRLAQSQALP PSSLQRFLRA R.NISGVVLA DHSGSFHNRY
NICA_ARATH_251-460      SLASKNIKIL SADTANPGIP PSSLMAFMRK NPQTSAVVLE DFDTNFVNKF
Q6Z743_ORYSA_248-457    SLGSDNVKVK PAASSNPGVP PSSLMSFLGK NSSTPGLVLE DFDSQFSNRF
NICA_CAEEL_278-508      GLRSHAFDLE KPSGSGDRVP PASWHSFAKA DAHVQSVLLA PYGKEYEYQR

                        201                                            250
query_196-256           ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
Q7KS07_DROME_236-469    YHSIYDDADN VDFTYANTSK ..DFTQLTEV N..DFKSLN. PDSLQMKVRN
Q7PY38_ANOGA_276-510    YHSIYDDQEN LHYRYGNHSR KYDFTQLEDL DLGDRSDLYG KDSIQMRIRN
NICA_MOUSE_273-498      YQSIYDTAEN INVTYPEWQS .......... ...PEEDLNF VTDTAKALAN
NICA_ARATH_251-460      YHSHLDDLSN INS....... .......... .......... .....SSVVA
Q6Z743_ORYSA_248-457    YHSTLDGPAN VNS....... .......... .......... .....SSIAA
NICA_CAEEL_278-508      VNSILDKNE. .......... .......... ...WTEDERE KAIQEIEAVS

                        251
query_196-256           ~~~~~~~~~~
Q7KS07_DROME_236-469    VSSIVAMALY
Q7PY38_ANOGA_276-510    ASSLIGMSIY
NICA_MOUSE_273-498      VATVLARALY
NICA_ARATH_251-460      AASVVARTLY
Q6Z743_ORYSA_248-457    AAALIARSLY
NICA_CAEEL_278-508      TAILAAAADY
```

Table 1. Sequence alignment between Nicastrin [Homo Sapiens] and the other six sets of Nicastrin variation. DYIGS is reported from our measure as confirmed by the γ-secretase review paper. Some other similar motif patterns are also reported, which are marked in red above.

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 260  Type: P  January 28, 2007 01:42  Check: 9305  ..

 Name: query_274-499        Len:    260  Check: 4890  Weight: 1.00
 Name: Q7KS07_DROME_236-469 Len:    260  Check: 6993  Weight: 1.00
 Name: Q7PY38_ANOGA_276-510 Len:    260  Check: 5861  Weight: 1.00
 Name: NICA_MOUSE_273-498   Len:    260  Check: 4890  Weight: 1.00
 Name: NICA_ARATH_251-460   Len:    260  Check: 9664  Weight: 1.00
 Name: Q6Z743_ORYSA_248-457 Len:    260  Check: 9567  Weight: 1.00
 Name: NICA_CAEEL_278-508   Len:    260  Check: 7440  Weight: 1.00


//


                        1                                              50
query_274-499           RVVVAATRLD SRSFFWNVAP GAESAVASFV TQLAAAEAL. ...HKA.PDV
Q7KS07_DROME_236-469    KFILVTCRLD TTTMFDGVGL GAMDSLMGFA VFTHVAYLL. ...KQLLPPQ
Q7PY38_ANOGA_276-510    RIVLVSTRTD TTTMFDGVGL GAMDSVVPFA VLVAVAHFL. ...SQALP..
NICA_MOUSE_273-498      RVVVAATRLD SRSFFWNVAP GAESAVASFV TQLAAAEAL. ...HKA.PDV
NICA_ARATH_251-460      PVVLTVASMD TASFFRDKSF GADSPISGLV ALLGAVDAL. ...SRV.DGI
Q6Z743_ORYSA_248-457    PIIMVTASQD SASFFRDRSL GADSPISGLI ALLTAVDAL. ...SHL.HDI
NICA_CAEEL_278-508      KYMMVTARMD SFGMIPEISV GEVSVLTSII SVLAAARSMG TQIEKWQKAS

                        51                                             100
query_274-499           TTLSRNVMFV FFQGETFDYI GSSRMVYDME NGKFP..... .......VRL
Q7KS07_DROME_236-469    SKDLHNVLFV TFNGESYDYI GSQRFVYDME KLQFP....T ESTGTPPIAF
Q7PY38_ANOGA_276-510    SND.RNVLFL FFNGESYDYI GSQRFVYDLQ TGAFP....S RGTQTKPISM
NICA_MOUSE_273-498      TTLSRNVMFV FFQGETFDYI GSSRMVYDME NGKFP..... .......VRL
NICA_ARATH_251-460      SNLKKQLVFL VLTGETWGYL GSRRFLHELD LHSDA..... ....VAGLSN
Q6Z743_ORYSA_248-457    SNLKKQLVFA VFNGEAWGYL GSRKFLQELD QGADS..... ....VNGISS
NICA_CAEEL_278-508      NTSNRNVFFA FFNGESLDYI GSGAAAYQME NGKFPQMIRS DRTHIHPIRP

                        101                                            150
query_274-499           ENIDSFVELG QVALRTSLD. ..LWMHTDPM SQKNESVKNQ VEDLLATLEK
```

```
Q7KS07_DROME_236-469   DNIDFMLDIG TLDDISN...  ..IKLH.... ALNGTTLAQQ ILERLNNYAK
Q7PY38_ANOGA_276-510   DNIELMIDLG ALDNLTD...  ....LHVY.. HAAPQPMATK VAELLRKINQ
NICA_MOUSE_273-498     ENIDSFVELG QVALRTSLD.  ..LWMHTDPM SQKNESVKNQ VEDLLATLEK
NICA_ARATH_251-460     TSIETVLEIG SVGKGLSGGI NTFFAH.... KTRVSSVTNM TLDALKIAQD
Q6Z743_ORYSA_248-457   LLIDQVLEIG SVGKAISQGY PLFYAH.... AAGNSSISMK MVDALQSASE
NICA_CAEEL_278-508     NELDYILEVQ QIGVAKGRK. ..YYVHVD.. GERYQQNKTQ TDRVIDRIER

                       151                                            200
query_274-499          SGAGVPEVVL RRLAQSQALP PSSLQRFLRA R.NISGVVLA DHSGSFHNRY
Q7KS07_DROME_236-469   SPRYGFNLNI .QSEMSAHLP PTSAQSFLRR DPNFNALILN ARPT...NKY
Q7PY38_ANOGA_276-510   QFGFGIQAGQ PVMTTN..LP PVSAQSFLRE NITFPAVIVA SRPA...NRF
NICA_MOUSE_273-498     SGAGVPEVVL RRLAQSQALP PSSLQRFLRA R.NISGVVLA DHSGSFHNRY
NICA_ARATH_251-460     SLASKNIKIL SADTANPGIP PSSLMAFMRK NPQTSAVVLE DFDTNFVNKF
Q6Z743_ORYSA_248-457   SLGSDNVKVK PAASSNPGVP PSSLMSFLGK NSSTPGLVLE DFDSQFSNRF
NICA_CAEEL_278-508     GLRSHAFDLE KPSGSGDRVP PASWHSFAKA DAHVQSVLLA PYGKEYEYQR

                       201                                            250
query_274-499          YQSIYDTAEN INVTYPEWQS .......... ...PEEDLNF VTDTAKALAN
Q7KS07_DROME_236-469   YHSIYDDADN VDFTYANTSK ..DFTQLTEV N..DFKSLN. PDSLQMKVRN
Q7PY38_ANOGA_276-510   YHSIYDDQEN LHYRYGNHSR KYDFTQLEDL DLGDRSDLYG KDSIQMRIRN
NICA_MOUSE_273-498     YQSIYDTAEN INVTYPEWQS .......... ...PEEDLNF VTDTAKALAN
NICA_ARATH_251-460     YHSHLDDLSN INS....... .......... .......... .....SSVVA
Q6Z743_ORYSA_248-457   YHSTLDGPAN VNS....... .......... .......... .....SSIAA
NICA_CAEEL_278-508     VNSILDKNE. .......... .......... ...WTEDERE KAIQEIEAVS

                       251
query_274-499          VATVLARALY
Q7KS07_DROME_236-469   VSSIVAMALY
Q7PY38_ANOGA_276-510   ASSLIGMSIY
NICA_MOUSE_273-498     VATVLARALY
NICA_ARATH_251-460     AASVVARTLY
Q6Z743_ORYSA_248-457   AAALIARSLY
NICA_CAEEL_278-508     TAILAAAADY
```

Table 2. Sequence alignment between Nicastrin [Mus musculus] and the other six sets of Nicastrin variation. DYIGS is reported from our measure as confirmed by the γ-secretase review paper. Some other similar motif patterns are also reported, which are marked in red above. "F - - F - G E - - D Y I G S" is speculated as a motif pattern, which might play an important functional role for "D Y I G S" group. "P P - S - - - F L R" and "N R - Y - S I - D - - - N" are other possible motif patterns from Nicastrin [Mus musculus].

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 260  Type: P  January 28, 2007 02:01  Check: 7558  ..

 Name: query_284-508        Len:   260  Check: 3143  Weight: 1.00
 Name: Q7KS07_DROME_236-469 Len:   260  Check: 6993  Weight: 1.00
 Name: Q7PY38_ANOGA_276-510 Len:   260  Check: 5861  Weight: 1.00
 Name: NICA_MOUSE_273-498   Len:   260  Check: 4890  Weight: 1.00
 Name: NICA_ARATH_251-460   Len:   260  Check: 9664  Weight: 1.00
 Name: Q6Z743_ORYSA_248-457 Len:   260  Check: 9567  Weight: 1.00
 Name: NICA_CAEEL_278-508   Len:   260  Check: 7440  Weight: 1.00

//

                       1                                             50
query_284-508          EFIMVATRID SHSFFWNIAP GAESAVSSFV THLAAAEAL. ...HKA.SDV
Q7KS07_DROME_236-469   KFILVTCRLD TTTMFDGVGL GAMDSLMGFA VFTHVAYLL. ...KQLLPPQ
Q7PY38_ANOGA_276-510   RIVLVSTRTD TTTMFDGVGL GAMDSVVPFA VLVAVAHFL. ...SQALP..
NICA_MOUSE_273-498     RVVVAATRLD SRSFFWNVAP GAESAVASFV TQLAAAEAL. ...HKA.PDV
```

```
NICA_ARATH_251-460      PVVLTVASMD TASFFRDKSF GADSPISGLV ALLGAVDAL. ...SRV.DGI
Q6Z743_ORYSA_248-457    PIIMVTASQD SASFFRDRSL GADSPISGLI ALLTAVDAL. ...SHL.HDI
NICA_CAEEL_278-508      KYMMVTARMD SFGMIPEISV GEVSVLTSII SVLAAARSMG TQIEKWQKAS

                        51                                                  100
query_284-508           HLLQRNIMFT FFQGETFDYI GSSRMVYDME KDKFP..... .......LRL
Q7KS07_DROME_236-469    SKDLHNVLFV TFNGESYDYI GSQRFVYDME KLQFP....T ESTGTPPIAF
Q7PY38_ANOGA_276-510    SND.RNVLFL FFNGESYDYI GSQRFVYDLQ TGAFP....S RGTQTKPISM
NICA_MOUSE_273-498      TTLSRNVMFV FFQGETFDYI GSSRMVYDME NGKFP..... .......VRL
NICA_ARATH_251-460      SNLKKQLVFL VLTGETWGYL GSRRFLHELD LHSDA..... ....VAGLSN
Q6Z743_ORYSA_248-457    SNLKKQLVFA VFNGEAWGYL GSRKFLQELD QGADS..... ....VNGISS
NICA_CAEEL_278-508      NTSNRNVFFA FFNGESLDYI GSGAAYQME NGKFPQMIRS DRTHIHPIRP

                        101                                                 150
query_284-508           ENIHSFVELN QVALRNG... SILWMHTDPV SRLNATVEPQ VKNLLDILSN
Q7KS07_DROME_236-469    DNIDFMLDIG TLDDISN... ..IKLH.... ALNGTTLAQQ ILERLNNYAK
Q7PY38_ANOGA_276-510    DNIELMIDLG ALDNLTD... ....LHVY.. HAAPQPMATK VAELLRKINQ
NICA_MOUSE_273-498      ENIDSFVELG QVALRTSLD. ..LWMHTDPM SQKNESVKNQ VEDLLATLEK
NICA_ARATH_251-460      TSIETVLEIG SVGKGLSGGI NTFFAH.... KTRVSSVTNM TLDALKIAQD
Q6Z743_ORYSA_248-457    LLIDQVLEIG SVGKAISQGY PLFYAH.... AAGNSSISMK MVDALQSASE
NICA_CAEEL_278-508      NELDYILEVQ QIGVAKGRK. ..YYVHVD.. GERYQQNKTQ TDRVIDRIER

                        151                                                 200
query_284-508           SSVGANITLQ .EVGFSQPLP PSSFQRFLRA R.HIPGVVLS DHQASFQNRY
Q7KS07_DROME_236-469    SPRYGFNLNI .QSEMSAHLP PTSAQSFLRR DPNFNALILN ARPT...NKY
Q7PY38_ANOGA_276-510    QFGFGIQAGQ PVMTTN..LP PVSAQSFLRE NITFPAVIVA SRPA...NRF
NICA_MOUSE_273-498      SGAGVPEVVL RRLAQSQALP PSSLQRFLRA R.NISGVVLA DHSGSFHNRY
NICA_ARATH_251-460      SLASKNIKIL SADTANPGIP PSSLMAFMRK NPQTSAVVLE DFDTNFVNKF
Q6Z743_ORYSA_248-457    SLGSDNVKVK PAASSNPGVP PSSLMSFLGK NSSTPGLVLE DFDSQFSNRF
NICA_CAEEL_278-508      GLRSHAFDLE KPSGSGDRVP PASWHSFAKA DAHVQSVLLA PYGKEYEYQR

                        201                                                 250
query_284-508           YQSMYDTPEN IQMQYPEGLS .......... ...PEETLEY VTDTAKSLAE
Q7KS07_DROME_236-469    YHSIYDDADN VDFTYANTSK ..DFTQLTEV N..DFKSLN. PDSLQMKVRN
Q7PY38_ANOGA_276-510    YHSIYDDQEN LHYRYGNHSR KYDFTQLEDL DLGDRSDLYG KDSIQMRIRN
NICA_MOUSE_273-498      YQSIYDTAEN INVTYPEWQS .......... ...PEEDLNF VTDTAKALAN
NICA_ARATH_251-460      YHSHLDDLSN INS....... .......... .......... .....SSVVA
Q6Z743_ORYSA_248-457    YHSTLDGPAN VNS....... .......... .......... .....SSIAA
NICA_CAEEL_278-508      VNSILDKNE. .......... .......... ...WTEDERE KAIQEIEAVS

                        251
query_284-508           VATVVARALY
Q7KS07_DROME_236-469    VSSIVAMALY
Q7PY38_ANOGA_276-510    ASSLIGMSIY
NICA_MOUSE_273-498      VATVLARALY
NICA_ARATH_251-460      AASVVARTLY
Q6Z743_ORYSA_248-457    AAALIARSLY
NICA_CAEEL_278-508      TAILAAAADY
```

Table 3. Sequence alignment between Nicastrin [Gallus gallus] and the other six sets of Nicastrin variation. DYIGS is reported from our measure as confirmed by the γ-secretase review paper. Some other similar motif patterns are also reported, which are marked in red above. "F - - F - G E - - D Y I G S" is speculated as a motif pattern, which might play an important functional role for "D Y I G S" group. "P P - S - - - F L R" and "N R - Y - S - - D - - - N" are other possible motif patterns from Nicastrin [Gallus gallus].

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 260  Type: P  January 28, 2007 02:11  Check: 1771  ..

 Name: query_267-500        Len:    260  Check: 7356  Weight: 1.00
 Name: Q7KS07_DROME_236-469 Len:    260  Check: 6993  Weight: 1.00
```

```
Name: Q7PY38_ANOGA_276-510  Len:    260  Check: 5861  Weight: 1.00
Name: NICA_MOUSE_273-498     Len:    260  Check: 4890  Weight: 1.00
Name: NICA_ARATH_251-460     Len:    260  Check: 9664  Weight: 1.00
Name: Q6Z743_ORYSA_248-457   Len:    260  Check: 9567  Weight: 1.00
Name: NICA_CAEEL_278-508     Len:    260  Check: 7440  Weight: 1.00

//


                         1                                                50
query_267-500            KFILVTCRLD TTTMFDGVGL GAMDSLMGFA VFTHVAYLL. ...KQLLPPQ
Q7KS07_DROME_236-469     KFILVTCRLD TTTMFDGVGL GAMDSLMGFA VFTHVAYLL. ...KQLLPPQ
Q7PY38_ANOGA_276-510     RIVLVSTRTD TTTMFDGVGL GAMDSVVPFA VLVAVAHFL. ...SQALP..
NICA_MOUSE_273-498       RVVVAATRLD SRSFFWNVAP GAESAVASFV TQLAAAEAL. ...HKA.PDV
NICA_ARATH_251-460       PVVLTVASMD TASFFRDKSF GADSPISGLV ALLGAVDAL. ...SRV.DGI
Q6Z743_ORYSA_248-457     PIIMVTASQD SASFFRDRSL GADSPISGLI ALLTAVDAL. ...SHL.HDI
NICA_CAEEL_278-508       KYMMVTARMD SFGMIPEISV GEVSVLTSII SVLAAARSMG TQIEKWQKAS


                         51                                               100
query_267-500            SKDLHNVLFV TFNGESYDYI GSQRFVYDME KLQFP....T ESTGTPPIAF
Q7KS07_DROME_236-469     SKDLHNVLFV TFNGESYDYI GSQRFVYDME KLQFP....T ESTGTPPIAF
Q7PY38_ANOGA_276-510     SND.RNVLFL FFNGESYDYI GSQRFVYDLQ TGAFP....S RGTQTKPISM
NICA_MOUSE_273-498       TTLSRNVMFV FFQGETFDYI GSSRMVYDME NGKFP..... .......VRL
NICA_ARATH_251-460       SNLKKQLVFL VLTGETWGYL GSRRFLHELD LHSDA..... ....VAGLSN
Q6Z743_ORYSA_248-457     SNLKKQLVFA VFNGEAWGYL GSRKFLQELD QGADS..... ....VNGISS
NICA_CAEEL_278-508       NTSNRNVFFA FFNGESLDYI GSGAAAYQME NGKFPQMIRS DRTHIHPIRP


                         101                                              150
query_267-500            DNIDFMLDIG TLDDISN... ..IKLH.... ALNGTTLAQQ ILERLNNYAK
Q7KS07_DROME_236-469     DNIDFMLDIG TLDDISN... ..IKLH.... ALNGTTLAQQ ILERLNNYAK
Q7PY38_ANOGA_276-510     DNIELMIDLG ALDNLTD... ....LHVY.. HAAPQPMATK VAELLRKINQ
NICA_MOUSE_273-498       ENIDSFVELG QVALRTSLD. ..LWMHTDPM SQKNESVKNQ VEDLLATLEK
NICA_ARATH_251-460       TSIETVLEIG SVGKGLSGGI NTFFAH.... KTRVSSVTNM TLDALKIAQD
Q6Z743_ORYSA_248-457     LLIDQVLEIG SVGKAISQGY PLFYAH.... AAGNSSISMK MVDALQSASE
NICA_CAEEL_278-508       NELDYILEVQ QIGVAKGRK. ..YYVHVD.. GERYQQNKTQ TDRVIDRIER


                         151                                              200
query_267-500            SPRYGFNLNI .QSEMSAHLP PTSAQSFLRR DPNFNALILN ARPT...NKY
Q7KS07_DROME_236-469     SPRYGFNLNI .QSEMSAHLP PTSAQSFLRR DPNFNALILN ARPT...NKY
Q7PY38_ANOGA_276-510     QFGFGIQAGQ PVMTTN..LP PVSAQSFLRE NITFPAVIVA SRPA...NRF
NICA_MOUSE_273-498       SGAGVPEVVL RRLAQSQALP PSSLQRFLRA R.NISGVVLA DHSGSFHNRY
NICA_ARATH_251-460       SLASKNIKIL SADTANPGIP PSSLMAFMRK NPQTSAVVLE DFDTNFVNKF
Q6Z743_ORYSA_248-457     SLGSDNVKVK PAASSNPGVP PSSLMSFLGK NSSTPGLVLE DFDSQFSNRF
NICA_CAEEL_278-508       GLRSHAFDLE KPSGSGDRVP PASWHSFAKA DAHVQSVLLA PYGKEYEYQR


                         201                                              250
query_267-500            YHSTYDDADN VDFTYANTSK ..DFTQLTEV N..DFKSLN. PDSLQMKVRN
Q7KS07_DROME_236-469     YHSIYDDADN VDFTYANTSK ..DFTQLTEV N..DFKSLN. PDSLQMKVRN
Q7PY38_ANOGA_276-510     YHSIYDDQEN LHYRYGNHSR KYDFTQLEDL DLGDRSDLYG KDSIQMRIRN
NICA_MOUSE_273-498       YQSIYDTAEN INVTYPEWQS .......... ...PEEDLNF VTDTAKALAN
NICA_ARATH_251-460       YHSHLDDLSN INS....... .......... .......... .....SSVVA
Q6Z743_ORYSA_248-457     YHSTLDGPAN VNS....... .......... .......... .....SSIAA
NICA_CAEEL_278-508       VNSILDKNE. .......... .......... ...WTEDERE KAIQEIEAVS


                         251
query_267-500            VSSIVAMALY
Q7KS07_DROME_236-469     VSSIVAMALY
Q7PY38_ANOGA_276-510     ASSLIGMSIY
NICA_MOUSE_273-498       VATVLARALY
NICA_ARATH_251-460       AASVVARTLY
Q6Z743_ORYSA_248-457     AAALIARSLY
NICA_CAEEL_278-508       TAILAAAADY
```

Table 4. Sequence alignment between Nicastrin [Drosophila melanogaster] and the other six sets of Nicastrin variation. DYIGS is reported from our measure as confirmed by the γ-secretase review paper. Some other similar motif patterns are also reported, which are marked in red above. "F - - F - G E - - D Y I G S" is speculated as a motif pattern, which

might play an important functional role for "D Y I G S" group. "P P - S - - S F L R" and "N - - Y H S - - D - - - N" are other possible motif patterns from Nicastrin [Drosophila melanogaster].

Similar speculation can also be reached following the protocols. The analysis on Presenilin-1 can have much similar results like Figure 8.
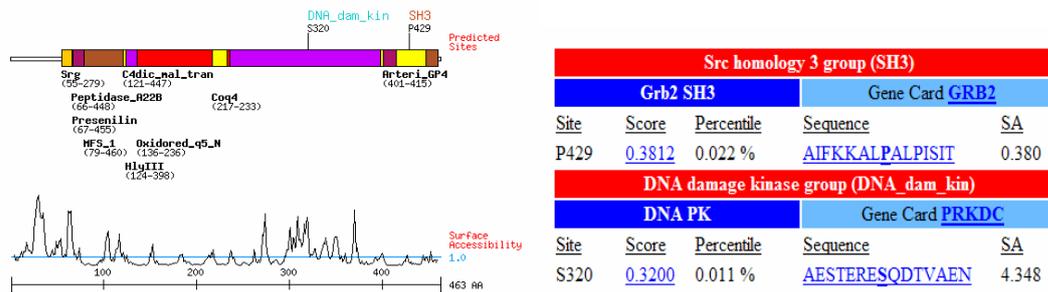
Figure 8. Speculative Motif sites of Presenilin-1 [Homo Sapiens].

The motif pattern identification will be achieved following an automated protocol specialized to process γ-secretase. Some unprocessed multiple sequence alignment data are indicated for Presenilin-1. It's straightforward to see that the sliding window method can help obtain the motif patterns for γ-secretase by measure the short-range repetitious subsequence segments.

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 480  Type: P  January 28, 2007 22:26  Check: 1804  ..

 Name: query_67-455     Len:     480  Check: 9257  Weight: 1.00
 Name: PSN_CAEEL_38-435  Len:     480  Check: 8470  Weight: 1.00
 Name: PSN2_RAT_76-439  Len:     480  Check: 7001  Weight: 1.00
 Name: PSN1_RAT_70-459  Len:     480  Check:  744  Weight: 1.00
 Name: PSN_DROME_92-532  Len:     480  Check: 1070  Weight: 1.00
 Name: HOP1_CAEEL_1-349  Len:     480  Check: 6457  Weight: 1.00
 Name: SPE4_CAEEL_5-465  Len:     480  Check: 8805  Weight: 1.00

//

               1                                                  50
query_67-455      DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRK DG.QLIYTPF
PSN_CAEEL_38-435  VEEAELKYGA SHVIHLFVPV SLCMALVVFT MNTITFYSQN NGRHLLYTPF
PSN2_RAT_76-439   EEELTLKYGA KHVIMLFVPV TLCMIVVVAT IKSVRFYTEK NG.QLIYTPF
PSN1_RAT_70-459   DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRK DG.QLIYTPF
PSN_DROME_92-532  EEEQGLKYGA QHVIKLFVPV SLCMLVVVAT INSISFYNST DV.YLLYTPF
HOP1_CAEEL_1-349  MPRTKRVYSG KTITGVLYPV AICMLFVAIN VKLSQPEQQE QS.KVVYGLF
```

```
SPE4_CAEEL_5-465    RSISSELVRS SQLRWTLFSV IANMSLTLSI WIGVYNMEVN SELSKTYFLD

                    51                                                   100
query_67-455        TEDTETVGQR ALHSILNAAI MISVIVVMTI LLVVLYKYRC YKVIHAWLII
PSN_CAEEL_38-435    VRETDSIVEK GLMSLGNALV MLCVVVLMTV LLIVFYKYKF YKLIHGWLIV
PSN2_RAT_76-439     TEDTPSVGQR LLNSVLNTLI MISVIVVMTI FLVVLYKYRC YKFIHGWLIM
PSN1_RAT_70-459     TEDTETVGQR ALHSILNAAI MISVIVVMTI LLVVLYKYRC YKVIHAWLIV
PSN_DROME_92-532    HEQSPEPSVK FWSALANSLI LMSVVVVMTF LLIVLYKKRC YRIIHGWLIL
HOP1_CAEEL_1-349    HS......YD TADSGTITLY LIGFLILTTS LGVFCYQMKF YKAIKVYVLA
SPE4_CAEEL_5-465    PSFEQTTGNL LLDGFINGVG TILVLGCVSF IMLAFVLFDF RRIVKAWLTL

                    101                                                  150
query_67-455        SSLLLLFFFS FIYLGEVFKT YN.......V AVDYITVALL IWNFGVVGMI
PSN_CAEEL_38-435    SSFLLLFLFT TIYVQEVLKS FD.......V SPSALLVLFG LGNYGVLGMM
PSN2_RAT_76-439     SSLMLLFLFT YIYLGEVFKT YN.......V AMDYPTLFLA VWNFGAVGMM
PSN1_RAT_70-459     SSLLLLFFFS FIYLGEVFKT YN.......V AVDYITVALL IWNFGVVGMI
PSN_DROME_92-532    SSFMLLFIFT YLYLEELLRA YN.......I PMDYPTALLI MWNFGVVGMM
HOP1_CAEEL_1-349    NSIGILLVYS VFHFQRIAEA QS.......I PVSVPTFFFL ILQFGGLGIT
SPE4_CAEEL_5-465    SCLLILFGVS AQTLHDMFSQ VFDQDDNNQY YMTIVLIVVP TVVYGFGGIY

                    151                                                  200
query_67-455        SIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWILLA VISVYDLVAV
PSN_CAEEL_38-435    CIHWKGPLRL QQFYLITMSA LMALVFIKYL PEWTVWFVLF VISVWDLVAV
PSN2_RAT_76-439     CIHWKGPLVL QQAYLIVISA LMALVFIKYL PEWSAWILG AISVYDLVAV
PSN1_RAT_70-459     AIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWILLA VISVYDLVAV
PSN_DROME_92-532    SIHWQGPLRL QQGYLIFVAA LMALVFIKYL PEWTAWAVLA AISIWDLIAV
HOP1_CAEEL_1-349    CLHWKSHRRL HQFYLIMLAG LTAIFILNIL PDWTVWMALT AISFWDIVAV
SPE4_CAEEL_5-465    AFFSNSSLIL HQIFVVTNCS LISVFYLRVF PSKTTWFVLW IVLFWDLFAV

                    201                                                  250
query_67-455        LCPKGPLRML VETAQERNET LFPALIYSST MVWLVNM... ...AEGDPEA
PSN_CAEEL_38-435    LTPKGPLRYL VETAQERNEP IFPALIYSSG VIYPYVLVTA VEN..TTDPR
PSN2_RAT_76-439     LCPKGPLRML VETAQERNEP IFPALIYSSA .MVWTVG... .....MAK..
PSN1_RAT_70-459     LCPKGPLRML VETAQERNEP IFPALIYSST MVWLVNM... ...AEGDPEA
PSN_DROME_92-532    LSPRGPLRIL VETAQERNEQ IFPALIYSST VVYALVNTVT PQQSQATASS
HOP1_CAEEL_1-349    LTPCGPLKML VETANRRGDD KFPAILYNSS SYVNEVD... ......SP..
SPE4_CAEEL_5-465    LAPMGPLKKV QEKASDYSKC VLNLIMFSAN EKRLTAGSNQ EETNEGEEST

                    251                                                  300
query_67-455        QRRVSK.N.. .......... .......... ..........S KYNAE.....
PSN_CAEEL_38-435    EPTSSDSN.. .......... .........T STAFPGEASC SSETP.....
PSN2_RAT_76-439     ....LD.... .......... .......... .........PS SQGAL.....
PSN1_RAT_70-459     QRRVPK.N.. .......... .......... .........P KYSTQ.....
PSN_DROME_92-532    SPSSSNST.. .TTTRATQ.. .........N SLASPEAAAA SGQRT.....
HOP1_CAEEL_1-349    ..DTTR.... .......... .......... .......... ..........
SPE4_CAEEL_5-465    IRRTVKQTIE YYTKREAQDD EFYQKIRQRR AAINPDSVPT EHSPLVEAEP

                    301                                                  350
query_67-455        .STERES.QD TVAENDDGGF SEEWE..... A......... ......QRDS
PSN_CAEEL_38-435    ...KRPKVKR IPQKVQIESN TTAST..... T......... ......QNSG
PSN2_RAT_76-439     ...QLPY... DPEMEEDSYD SFGEP..... S......... ..........
PSN1_RAT_70-459     .GTEREETQD TGTGSDDGGF SEEWE..... A......... ......QRDS
PSN_DROME_92-532    .....GN.SH PRQNQRDDGS VLATEGMPLV TFKSNLRGNA EAAGFTQEWS
HOP1_CAEEL_1-349    .....SN.ST PLTEFNNSSS SRLLE..... S......... ........DS
SPE4_CAEEL_5-465    SPIELKEKNS TEELSDDESD TSETS..... SGSSNLSSSD SSTTVSTSDI

                    351                                                  400
query_67-455        HLGPHRsTPE SRAAVQELSS ....SILAG. ...ED..... .....PEERG
PSN_CAEEL_38-435    VRVE....RE LAAERPTVQD ....ANFHR. ..HEE..... ......EEPG
PSN2_RAT_76-439     ..YP....EA FEAPQPGYP. .........G EEPEE..... .....EEERG
PSN1_RAT_70-459     HLGPHRsTPE SRAAVQELSG ....SILTS. ...ED..... .....PEERG
PSN_DROME_92-532    ANLS....ER VARRQIEVQS tqsgNAQRSN EYRTVTAPDQ NHPDGQEERG
HOP1_CAEEL_1-349    LLRP....PV IPRQIREVR. .......... .......... .....EVEGT
SPE4_CAEEL_5-465    STAE....EC DQKEWDDLVS ....NSLPNN DKRPATAADA L....NDGEV

                    401                                                  450
query_67-455        VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_CAEEL_38-435    VKLGLGDFIF YSVLLGKASS Y..FDWNTTI ACYVAILIGL CFTLVLLAVF
PSN2_RAT_76-439     VKLGLGDFIF YSVLVGKAAA TGNGDWSTTL ACFIAILIGL CLTLLLLAVF
PSN1_RAT_70-459     VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_DROME_92-532    IKLGLGDFIF YSVLVGKASS Y..GDWTTTI ACFVAILIGL CLTLLLLAIW
```

36

```
HOP1_CAEEL_1-349    IRLGMGDFVF YSLMLGNTVQ T..CPLPTVV ACFVSNLVGL TITLPIVTLS
SPE4_CAEEL_5-465    LRLGFGDFVF YSLLIGQAAA S..GCPFAVI SAALGILFGL VVTLTVFSTE


                    451                            480
query_67-455        KKALPALPIS ITFGLVFYFA TDYLVQPFMD
PSN_CAEEL_38-435    KRALPALPIS IFSGLIFYFC TRWIITPFVT
PSN2_RAT_76-439     KKALPALPIS ITFGLIFYFS TDNLVRPFMD
PSN1_RAT_70-459     KKALPALPIS ITFGLIFYFA TDYLVQPFMD
PSN_DROME_92-532    RKALPALPIS ITFGLIFCFA TSAVVKPFME
HOP1_CAEEL_1-349    QTALPALPFP LAIAAIFYFS SHIALTPFTD
SPE4_CAEEL_5-465    ESTTPALPLP VICGTFCYFS SMFFWEQLYG
```

Table 5. Unprocessed sequence alignment scheme for PSEN1 protein [Homo sapiens]

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 480  Type: P  January 28, 2007 22:29  Check: 956  ..

 Name: query_43-431     Len:     480  Check: 8409  Weight: 1.00
 Name: PSN_CAEEL_38-435 Len:     480  Check: 8470  Weight: 1.00
 Name: PSN2_RAT_76-439  Len:     480  Check: 7001  Weight: 1.00
 Name: PSN1_RAT_70-459  Len:     480  Check:  744  Weight: 1.00
 Name: PSN_DROME_92-532 Len:     480  Check: 1070  Weight: 1.00
 Name: HOP1_CAEEL_1-349 Len:     480  Check: 6457  Weight: 1.00
 Name: SPE4_CAEEL_5-465 Len:     480  Check: 8805  Weight: 1.00


//


                    1                                               50
query_43-431        DEELTLKYGA KHVIMLFIPV TLCMVVVVAT IKSVSFYTQK DGQQLIYTPF
PSN_CAEEL_38-435    VEEAELKYGA SHVIHLFVPV SLCMALVVFT MNTITFYSQN NGRHLLYTPF
PSN2_RAT_76-439     EEELTLKYGA KHVIMLFVPV TLCMIVVVAT IKSVRFYTEK NG.QLIYTPF
PSN1_RAT_70-459     DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRK DG.QLIYTPF
PSN_DROME_92-532    EEEQGLKYGA QHVIKLFVPV SLCMLVVVAT INSISFYNST DV.YLLYTPF
HOP1_CAEEL_1-349    MPRTKRVYSG KTITGVLYPV AICMLFVAIN VKLSQPEQQE QS.KVVYGLF
SPE4_CAEEL_5-465    RSISSELVRS SQLRWTLFSV IANMSLTLSI WIGVYNMEVN SELSKTYFLD


                    51                                              100
query_43-431        REDTETVGQR ALHSMLNAII MISVIVVMTL VLVVLYKYRC YKVIQAWLFF
PSN_CAEEL_38-435    VRETDSIVEK GLMSLGNALV MLCVVVLMTV LLIVFYKYKF YKLIHGWLIV
PSN2_RAT_76-439     TEDTPSVGQR LLNSVLNTLI MISVIVVMTI FLVVLYKYRC YKFIHGWLIM
PSN1_RAT_70-459     TEDTETVGQR ALHSILNAAI MISVIVVMTI LLVVLYKYRC YKVIHAWLIV
PSN_DROME_92-532    HEQSPEPSVK FWSALANSLI LMSVVVVMTF LLIVLYKKRC YRIIHGWLIL
HOP1_CAEEL_1-349    HS......YD TADSGTITLY LIGFLILTTS LGVFCYQMKF YKAIKVYVLA
SPE4_CAEEL_5-465    PSFEQTTGNL LLDGFINGVG TILVLGCVSF IMLAFVLFDF RRIVKAWLTL


                    101                                             150
query_43-431        SNLLLLFFFS LIYLGEVFKT YN.......V AMDYFTLALI IWNFGVVGMI
PSN_CAEEL_38-435    SSFLLLFLFT TIYVQEVLKS FD.......V SPSALLVLFG LGNYGVLGMM
PSN2_RAT_76-439     SSLMLLFLFT YIYLGEVFKT YN.......V AMDYPTLFLA VWNFGAVGMV
PSN1_RAT_70-459     SSLLLLFFFS FIYLGEVFKT YN.......V AVDYITVALL IWNFGVVGMI
PSN_DROME_92-532    SSFMLLFIFT YLYLEELLRA YN.......I PMDYPTALLI MWNFGVVGMM
HOP1_CAEEL_1-349    NSIGILLVYS VFHFQRIAEA QS.......I PVSVPTFFFL ILQFGGLGIT
SPE4_CAEEL_5-465    SCLLILFGVS AQTLHDMFSQ VFDQDDNNQY YMTIVLIVVP TVVYGFGGIY


                    151                                             200
query_43-431        CIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWLILA AISVYDLLAV
PSN_CAEEL_38-435    CIHWKGPLRL QQFYLITMSA LMALVFIKYL PEWTVWFVLF VISVWDLVAV
PSN2_RAT_76-439     CIHWKGPLVL QQAYLIVISA LMALVFIKYL PEWSAWVILG AISVYDLVAV
PSN1_RAT_70-459     AIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWLILA VISVYDLVAV
PSN_DROME_92-532    SIHWQGPLRL QQGYLIFVAA LMALVFIKYL PEWTAWAVLA AISIWDLIAV
HOP1_CAEEL_1-349    CLHWKSHRRL HQFYLIMLAG LTAIFILNIL PDWTVWMALT AISFWDIVAV
SPE4_CAEEL_5-465    AFFSNSSLIL HQIFVVTNCS LISVFYLRVF PSKTTWFVLW IVLFWDLFAV


                    201                                             250
query_43-431        LCPKGPLRIL VETAQERNEA IFPALIYSST MVWLFNM... ADSAETRN..
PSN_CAEEL_38-435    LTPKGPLRYL VETAQERNEP IFPALIYSSG VIYPYVLVTA VEN..TTDPR
```

```
PSN2_RAT_76-439     LCPKGPLRML VETAQERNEP IFPALIYSSA .MVWTVG... .....MAK..
PSN1_RAT_70-459     LCPKGPLRML VETAQERNET LFPALIYSST MVWLVNM... ...AEGDPEA
PSN_DROME_92-532    LSPRGPLRIL VETAQERNEQ IFPALIYSST VVYALVNTVT PQQSQATASS
HOP1_CAEEL_1-349    LTPCGPLKML VETANRRGDD KFPAILYNSS SYVNEVD... ......SP..
SPE4_CAEEL_5-465    LAPMGPLKKV QEKASDYSKC VLNLIMFSAN EKRLTAGSNQ EETNEGEEST

                    251                                            300
query_43-431        ..NSSH.... .......... .......... ........PV PQQEN.....
PSN_CAEEL_38-435    EPTSSDSN.. .......... .........T STAFPGEASC SSETP.....
PSN2_RAT_76-439     ....LD.... .......... .......... .......PS SQGAL.....
PSN1_RAT_70-459     QRRVPK.N.. .......... .......... .........P KYSTQ.....
PSN_DROME_92-532    SPSSSNST.. .TTTRATQ.. .........N SLASPEAAAA SGQRT.....
HOP1_CAEEL_1-349    ..DTTR.... .......... .......... .......... ..........
SPE4_CAEEL_5-465    IRRTVKQTIE YYTKREAQDD EFYQKIRQRR AAINPDSVPT EHSPLVEAEP

                    301                                            350
query_43-431        ...QVAM.AP TAQPEDDGGF TPAWV..... D......... ......HQQH
PSN_CAEEL_38-435    ...KRPKVKR IPQKVQIESN TTAST..... T......... ......QNSG
PSN2_RAT_76-439     ...QLPY... DPEMEEDSYD SFGEP..... S......... ..........
PSN1_RAT_70-459     .GTEREETQD TGTGSDDGGF SEEWE..... A......... ......QRDS
PSN_DROME_92-532    .....GN.SH PRQNQRDDGS VLATEGMPLV TFKSNLRGNA EAAGFTQEWS
HOP1_CAEEL_1-349    .....SN.ST PLTEFNNSSS SRLLE..... S......... ........DS
SPE4_CAEEL_5-465    SPIELKEKNS TEELSDDESD TSETS..... SGSSNLSSSD SSTTVSTSDI

                    351                                            400
query_43-431        QLGPMQsTEE SRRQIQEMPS ....ARPPP. ...PA..... ...DDDEERG
PSN_CAEEL_38-435    VRVE....RE LAAERPTVQD ....ANFHR. ..HEE..... ......EERG
PSN2_RAT_76-439     ..YP....EA FEAPQPGYP. .........G EEPEE..... ....EEERG
PSN1_RAT_70-459     HLGPHRsTPE SRAAVQELSG ....SILTS. ...ED..... .....PEERG
PSN_DROME_92-532    ANLS....ER VARRQIEVQS tqsgNAQRSN EYRTVTAPDQ NHPDGQEERG
HOP1_CAEEL_1-349    LLRP....PV IPRQIREVR. .......... .......... .....EVEGT
SPE4_CAEEL_5-465    STAE....EC DQKEWDDLVS ....NSLPNN DKRPATAADA L....NDGEV

                    401                                            450
query_43-431        VKLGLGDFIF YSMLVGKASA TASGDWNTTL ACFVAILIGL CLTLLLLAIF
PSN_CAEEL_38-435    VKLGLGDFIF YSVLLGKASS Y..FDWNTTI ACYVAILIGL CFTLVLLAVF
PSN2_RAT_76-439     VKLGLGDFIF YSVLVGKAAA TGNGDWSTTL ACFIAILIGL CLTLLLLAVF
PSN1_RAT_70-459     VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_DROME_92-532    IKLGLGDFIF YSVLVGKASS Y..GDWTTTI ACFVAILIGL CLTLLLLAIW
HOP1_CAEEL_1-349    IRLGMGDFVF YSLMLGNTVQ T..CPLPTVV ACFVSNLVGL TITLPIVTLS
SPE4_CAEEL_5-465    LRLGFGDFVF YSLLIGQAAA S..GCPFAVI SAALGILFGL VVTLTVFSTE

                    451                   480
query_43-431        KKALPALPIS ITFGLVFYFA TDNLVRPFMD
PSN_CAEEL_38-435    KRALPALPIS IFSGLIFYFC TRWIITPFVT
PSN2_RAT_76-439     KKALPALPIS ITFGLIFYFS TDNLVRPFMD
PSN1_RAT_70-459     KKALPALPIS ITFGLIFYFA TDYLVQPFMD
PSN_DROME_92-532    RKALPALPIS ITFGLIFCFA TSAVVKPFME
HOP1_CAEEL_1-349    QTALPALPFP LAIAAIFYFS SHIALTPFTD
SPE4_CAEEL_5-465    ESTTPALPLP VICGTFCYFS SMFFWEQLYG
```

Table 6. Unprocessed sequence alignment scheme for Psen1 protein [Danio rerio]

```
!!AA_MULTIPLE_ALIGNMENT 1.0
 squid.msf  MSF: 480  Type: P  January 28, 2007 22:35  Check: 1932  ..

 Name: query_71-459      Len:    480  Check: 9385  Weight: 1.00
 Name: PSN_CAEEL_38-435  Len:    480  Check: 8470  Weight: 1.00
 Name: PSN2_RAT_76-439   Len:    480  Check: 7001  Weight: 1.00
 Name: PSN1_RAT_70-459   Len:    480  Check:  744  Weight: 1.00
 Name: PSN_DROME_92-532  Len:    480  Check: 1070  Weight: 1.00
 Name: HOP1_CAEEL_1-349  Len:    480  Check: 6457  Weight: 1.00
 Name: SPE4_CAEEL_5-465  Len:    480  Check: 8805  Weight: 1.00


 //
```

```
                       1                                                  50
query_71-459           DEELTLKYGA KHVIMLFVPV TLCMVVVAT IKSVSFYTRK DG.QLIYTPF
PSN_CAEEL_38-435       VEEAELKYGA SHVIHLFVPV SLCMALVVFT MNTITFYSQN NGRHLLYTPF
PSN2_RAT_76-439        EEELTLKYGA KHVIMLFVPV TLCMIVVVAT IKSVRFYTEK NG.QLIYTPF
PSN1_RAT_70-459        DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRK DG.QLIYTPF
PSN_DROME_92-532       EEEQGLKYGA QHVIKLFVPV SLCMLVVVAT INSISFYNST DV.YLLYTPF
HOP1_CAEEL_1-349       MPRTKRVYSG KTITGVLYPV AICMLFVAIN VKLSQPEQQE QS.KVVYGLF
SPE4_CAEEL_5-465       RSISSELVRS SQLRWTLFSV IANMSLTLSI WIGVYNMEVN SELSKTYFLD


                       51                                                 100
query_71-459           TEDTETVGQR ALHSILNAAI MISVIVIMTI LLVVLYKYRC YKVIHAWLII
PSN_CAEEL_38-435       VRETDSIVEK GLMSLGNALV MLCVVVLMTV LLIVFYKYKF YKLIHGWLIV
PSN2_RAT_76-439        TEDTPSVGQR LLNSVLNTLI MISVIVVMTI FLVVLYKYRC YKFIHGWLIM
PSN1_RAT_70-459        TEDTETVGQR ALHSILNAAI MISVIVVMTI LLVVLYKYRC YKVIHAWLIV
PSN_DROME_92-532       HEQSPEPSVK FWSALANSLI LMSVVVVMTF LLIVLYKKRC YRIIHGWLIL
HOP1_CAEEL_1-349       HS......YD TADSGTITLY LIGFLILTTS LGVFCYQMKF YKAIKVYVLA
SPE4_CAEEL_5-465       PSFEQTTGNL LLDGFINGVG TILVLGCVSF IMLAFVLFDF RRIVKAWLTL


                       101                                                150
query_71-459           SSLLLLFFFS FIYLGEVFKT YN.......V AVDYVTVALL IWNFGVVGMI
PSN_CAEEL_38-435       SSFLLLFLFT TIYVQEVLKS FD.......V SPSALLVLFG LGNYGVLGMM
PSN2_RAT_76-439        SSLMLLFLFT YIYLGEVFKT YN.......V AMDYPTLFLA VWNFGAVGMV
PSN1_RAT_70-459        SSLLLLFFFS FIYLGEVFKT YN.......V AVDYITVALL IWNFGVVGMI
PSN_DROME_92-532       SSFMLLFIFT YLYLEELLRA YN.......I PMDYPTALLI MWNFGVVGMM
HOP1_CAEEL_1-349       NSIGILLVYS VFHFQRIAEA QS.......I PVSVPTFFFL ILQFGGLGIT
SPE4_CAEEL_5-465       SCLLILFGVS AQTLHDMFSQ VFDQDDNNQY YMTIVLIVVP TVVYGFGGIY


                       151                                                200
query_71-459           AIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWLILA VISVYDLVAV
PSN_CAEEL_38-435       CIHWKGPLRL QQFYLITMSA LMALVFIKYL PEWTVWFVLF VISVWDLVAV
PSN2_RAT_76-439        CIHWKGPLVL QQAYLIVISA LMALVFIKYL PEWSAWVILG AISVYDLVAV
PSN1_RAT_70-459        AIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWLILA VISVYDLVAV
PSN_DROME_92-532       SIHWQGPLRL QQGYLIFVAA LMALVFIKYL PEWTAWAVLA AISIWDLIAV
HOP1_CAEEL_1-349       CLHWKSHRRL HQFYLIMLAG LTAIFILNIL PDWTVWMALT AISFWDIVAV
SPE4_CAEEL_5-465       AFFSNSSLIL HQIFVVTNCS LISVFYLRVF PSKTTWFVLW IVLFWDLFAV


                       201                                                250
query_71-459           LCPKGPLRML VETAQERNET LFPALIYSST MVWLVNM... ...AEGDPEA
PSN_CAEEL_38-435       LTPKGPLRYL VETAQERNEP IFPALIYSSG VIYPYVLVTA VEN..TTDPR
PSN2_RAT_76-439        LCPKGPLRML VETAQERNEP IFPALIYSSA .MVWTVG... ...MAK..
PSN1_RAT_70-459        LCPKGPLRML VETAQERNET LFPALIYSST MVWLVNM... ...AEGDPEA
PSN_DROME_92-532       LSPRGPLRIL VETAQERNEQ IFPALIYSST VVYALVNTVT PQQSQATASS
HOP1_CAEEL_1-349       LTPCGPLKML VETANRRGDD KFPAILYNSS SYVNEVD... ......SP..
SPE4_CAEEL_5-465       LAPMGPLKKV QEKASDYSKC VLNLIMFSAN EKRLTAGSNQ EETNEGEEST


                       251                                                300
query_71-459           QRRVPK.N.. .......... .......... .........P KYNTQ.....
PSN_CAEEL_38-435       EPTSSDSN.. .......... .........T STAFPGEASC SSETP.....
PSN2_RAT_76-439        ....LD.... .......... .......... .......PS SQGAL.....
PSN1_RAT_70-459        QRRVPK.N.. .......... .......... .........P KYSTQ.....
PSN_DROME_92-532       SPSSSNST.. .TTTRATQ.. .........N SLASPEAAAA SGQRT.....
HOP1_CAEEL_1-349       ..DTTR.... .......... .......... .......... ..........
SPE4_CAEEL_5-465       IRRTVKQTIE YYTKREAQDD EFYQKIRQRR AAINPDSVPT EHSPLVEAEP


                       301                                                350
query_71-459           .RAERET.QD SGSGNDDGGF SEEWE..... A......... ......QRDS
PSN_CAEEL_38-435       ...KRPKVKR IPQKVQIESN TTAST..... T......... ......QNSG
PSN2_RAT_76-439        ...QLPY... DPEMEEDSYD SFGEP..... S......... ..........
PSN1_RAT_70-459        .GTEREETQD TGTGSDDGGF SEEWE..... A......... ......QRDS
PSN_DROME_92-532       .....GN.SH PRQNQRDDGS VLATEGMPLV TFKSNLRGNA EAAGFTQEWS
HOP1_CAEEL_1-349       .....SN.ST PLTEFNNSSS SRLLE..... S......... ........DS
SPE4_CAEEL_5-465       SPIELKEKNS TEELSDDESD TSETS..... SGSSNLSSSD SSTTVSTSDI


                       351                                                400
query_71-459           HLGPHRsTPE SRAAVQELSG ....SILTS. ...ED..... .....PEERG
PSN_CAEEL_38-435       VRVE....RE LAAERPTVQD ....ANFHR. ..HEE..... ....EERG
PSN2_RAT_76-439        ..YP....EA FEAPQPGYP. .........G EEPEE..... ....EEERG
PSN1_RAT_70-459        HLGPHRsTPE SRAAVQELSG ....SILTS. ...ED..... .....PEERG
PSN_DROME_92-532       ANLS....ER VARRQIEVQS tqsgNAQRSN EYRTVTAPDQ NHPDGQEERG
HOP1_CAEEL_1-349       LLRP....PV IPRQIREVR. .......... .......... .....EVEGT
SPE4_CAEEL_5-465       STAE....EC DQKEWDDLVS ....NSLPNN DKRPATAADA L....NDGEV
```

39

```
                401                                                   450
query_71-459    VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_CAEEL_38-435 VKLGLGDFIF YSVLLGKASS Y..FDWNTTI ACYVAILIGL CFTLVLLAVF
PSN2_RAT_76-439 VKLGLGDFIF YSVLVGKAAA TGNGDWSTTL ACFIAILIGL CLTLLLLAVF
PSN1_RAT_70-459 VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_DROME_92-532 IKLGLGDFIF YSVLVGKASS Y..GDWTTTI ACFVAILIGL CLTLLLLAIW
HOP1_CAEEL_1-349 IRLGMGDFVF YSLMLGNTVQ T..CPLPTVV ACFVSNLVGL TITLPIVTLS
SPE4_CAEEL_5-465 LRLGFGDFVF YSLLIGQAAA S..GCPFAVI SAALGILFGL VVTLTVFSTE


                451                   480
query_71-459    KKALPALPIS ITFGLVFYFA TDYLVQPFMD
PSN_CAEEL_38-435 KRALPALPIS IFSGLIFYFC TRWIITPFVT
PSN2_RAT_76-439 KKALPALPIS ITFGLIFYFS TDNLVRPFMD
PSN1_RAT_70-459 KKALPALPIS ITFGLIFYFA TDYLVQPFMD
PSN_DROME_92-532 RKALPALPIS ITFGLIFCFA TSAVVKPFME
HOP1_CAEEL_1-349 QTALPALPFP LAIAAIFYFS SHIALTPFTD
SPE4_CAEEL_5-465 ESTTPALPLP VICGTFCYFS SMFFWEQLYG
```

Table 7. Unprocessed sequence alignment scheme for Psen1 protein [Mus musculus]

```
        !!AA_MULTIPLE_ALIGNMENT 1.0
         squid.msf  MSF: 480  Type: P  January 28, 2007 22:40  Check: 4201  ..

         Name: query_65-453     Len:      480  Check: 1654  Weight: 1.00
         Name: PSN_CAEEL_38-435 Len:      480  Check: 8470  Weight: 1.00
         Name: PSN2_RAT_76-439  Len:      480  Check: 7001  Weight: 1.00
         Name: PSN1_RAT_70-459  Len:      480  Check:  744  Weight: 1.00
         Name: PSN_DROME_92-532 Len:      480  Check: 1070  Weight: 1.00
         Name: HOP1_CAEEL_1-349 Len:      480  Check: 6457  Weight: 1.00
         Name: SPE4_CAEEL_5-465 Len:      480  Check: 8805  Weight: 1.00


        //


                1                                                     50
query_65-453    DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRY DG.QLIYTPF
PSN_CAEEL_38-435 VEEAELKYGA SHVIHLFVPV SLCMALVVFT MNTITFYSQN NGRHLLYTPF
PSN2_RAT_76-439 EEELTLKYGA KHVIMLFVPV TLCMIVVVAT IKSVRFYTEK NG.QLIYTPF
PSN1_RAT_70-459 DEELTLKYGA KHVIMLFVPV TLCMVVVVAT IKSVSFYTRY DG.QLIYTPF
PSN_DROME_92-532 EEEQGLKYGA QHVIKLFVPV SLCMLVVVAT INSISFYNST DV.YLLYTPF
HOP1_CAEEL_1-349 MPRTKRVYSG KTITGVLYPV AICMLFVAIN VKLSQPEQQE QS.KVVYGLF
SPE4_CAEEL_5-465 RSISSELVRS SQLRWTLFSV IANMSLTLSI WIGVYNMEVN SELSKTYFLD


                51                                                    100
query_65-453    TEDTDSVGQR ALNSILNTAI MISVIIVMTI LLVVLYKYRC YKVIHGWLII
PSN_CAEEL_38-435 VRETDSIVEK GLMSLGNALV MLCVVVLMTV LLIVFYKYKF YKLIHGWLIV
PSN2_RAT_76-439 TEDTPSVGQR LLNSVLNTLI MISVIVVMTI FLVVLYKYRC YKFIHGWLIM
PSN1_RAT_70-459 TEDTETVGQR ALHSILNAAI MISVIVVMTI LLVVLYKYRC YKVIHAWLIV
PSN_DROME_92-532 HEQSPEPSVK FWSALANSLI LMSVVVVMTF LLIVLYKKRC YRIIHGWLIL
HOP1_CAEEL_1-349 HS......YD TADSGTITLY LIGFLILTTS LGVFCYQMKF YKAIKVYVLA
SPE4_CAEEL_5-465 PSFEQTTGNL LLDGFINGVG TILVLGCVSF IMLAFVLFDF RRIVKAWLTL


                101                                                   150
query_65-453    SSLLLLFFFS YIYLGEVFKT YN.......V AMDYITLALM IWNFGVVGMI
PSN_CAEEL_38-435 SSFLLLFLFT TIYVQEVLKS FD.......V SPSALLVLFG LGNYGVLGMM
PSN2_RAT_76-439 SSLMLLFLFT YIYLGEVFKT YN.......V AMDYPTLFLA VWNFGAVGMV
PSN1_RAT_70-459 SSLLLLFFFS FIYLGEVFKT YN.......V AVDYITVALL IWNFGVVGMI
PSN_DROME_92-532 SSFMLLFIFT YLYLEELLRA VN.......I PMDYPTALLI MWNFGVVGMM
HOP1_CAEEL_1-349 NSIGILLVYS VFHFQRIAEA QS.......I PVSVPTFFFL ILQFGGLGIT
SPE4_CAEEL_5-465 SCLLILFGVS AQTLHDMFSQ VFDQDDNNQY YMTIVLIVVP TVVYGFGGIY


                151                                                   200
query_65-453    CIHWKGPLLL QQAYLIMISA LMALVFIKYL PEWTTWLILA VISVYDLIAV
PSN_CAEEL_38-435 CIHWKGPLRL QQFYLITMSA LMALVFIKYL PEWTVWFVLF VISVWDLVAV
PSN2_RAT_76-439 CIHWKGPLVL QQAYLIVISA LMALVFIKYL PEWSAWVILG AISVYDLVAV
PSN1_RAT_70-459 AIHWKGPLRL QQAYLIMISA LMALVFIKYL PEWTAWLILA VISVYDLVAV
```

```
PSN_DROME_92-532   SIHWQGPLRL QQGYLIFVAA LMALVFIKYL PEWTAWAVLA AISIWDLIAV
HOP1_CAEEL_1-349   CLHWKSHRRL HQFYLIMLAG LTAIFILNIL PDWTVWMALT AISFWDIVAV
SPE4_CAEEL_5-465   AFFSNSSLIL HQIFVVTNCS LISVFYLRVF PSKTTWFVLW IVLFWDLFAV


                   201                                                 250
query_65-453       LSPKGPLRML VETAQERNET LFPALIYSST MIWLVKM... ...ADGDPGF
PSN_CAEEL_38-435   LTPKGPLRYL VETAQERNEP IFPALIYSSG VIYPYVLVTA VEN..TTDPR
PSN2_RAT_76-439    LCPKGPLRML VETAQERNEP IFPALIYSSA .MVWTVG... .....MAK..
PSN1_RAT_70-459    LCPKGPLRML VETAQERNEP LFPALIYSST MVWLVNM... ...AEGDPEA
PSN_DROME_92-532   LSPRGPLRIL VETAQERNEQ IFPALIYSST VVYALVNTVT PQQSQATASS
HOP1_CAEEL_1-349   LTPCGPLKML VETANRRGDD KFPAILYNSS SYVNEVD... ......SP..
SPE4_CAEEL_5-465   LAPMGPLKKV QEKASDYSKC VLNLIMFSAN EKRLTAGSNQ EETNEGEEST


                   251                                                 300
query_65-453       KQSASKKT.. .......... .......... .......... .YNAQ.....
PSN_CAEEL_38-435   EPTSSDSN.. .......... ........T STAFPGEASC SSETP.....
PSN2_RAT_76-439    ....LD.... .......... .......... ........PS SQGAL.....
PSN1_RAT_70-459    QRRVPK.N.. .......... .......... .........P KYSTQ.....
PSN_DROME_92-532   SPSSSNST.. .TTTRATQ.. .........N SLASPEAAAA SGQRT.....
HOP1_CAEEL_1-349   ..DTTR.... .......... .......... .......... ..........
SPE4_CAEEL_5-465   IRRTVKQTIE YYTKREAQDD EFYQKIRQRR AAINPDSVPT EHSPLVEAEP


                   301                                                 350
query_65-453       APVAQPR.SD SAASDDNGGF DTAWE..... E......... ......QRNE
PSN_CAEEL_38-435   ...KRPKVKR IPQKVQIESN TTAST..... T......... ......QNSG
PSN2_RAT_76-439    ...QLPY... DPEMEEDSYD SFGEP..... S......... ..........
PSN1_RAT_70-459    .GTEREETQD TGTGSDDGGF SEEWE..... A......... ......QRDS
PSN_DROME_92-532   .....GN.SH PRQNQRDDGS VLATEGMPLV TFKSNLRGNA EAAGFTQEWS
HOP1_CAEEL_1-349   .....SN.ST PLTEFNNSSS SRLLE..... S......... .........DS
SPE4_CAEEL_5-465   SPIELKEKNS TEELSDDESD TSETS..... SGSSNLSSSD SSTTVSTSDI


                   351                                                 400
query_65-453       QIGQINsTPE SRAAVQALPS ....NNPPS. ...ED..... .....PEERG
PSN_CAEEL_38-435   VRVE....RE LAAERPTVQD ....ANFHR. ..HEE..... ......EERG
PSN2_RAT_76-439    ..YP....EA FEAPQPGYP. .........G EEPEE..... .....EERG
PSN1_RAT_70-459    HLGPHRsTPE SRAAVQELSG ....SILTS. ...ED..... .....PEERG
PSN_DROME_92-532   ANLS....ER VARRQIEVQS tqsgNAQRSN EYRTVTAPDQ NHPDGQEERG
HOP1_CAEEL_1-349   LLRP....PV IPRQIREVR. .......... .......... .....EVEGT
SPE4_CAEEL_5-465   STAE....EC DQKEWDDLVS ....NSLPNN DKRPATAADA L....NDGEV


                   401                                                 450
query_65-453       VKLGLGDFIF YSVLVGKASA TASGDWNTTL ACFVAILIGL CLTLLLLAIF
PSN_CAEEL_38-435   VKLGLGDFIF YSVLLGKASS Y..FDWNTTI ACYVAILIGL CFTLVLLAVF
PSN2_RAT_76-439    VKLGLGDFIF YSVLVGKAAA TGNGDWSTTL ACFIAILIGL CLTLLLLAVF
PSN1_RAT_70-459    VKLGLGDFIF YSVLVGKASA TASGDWNTTI ACFVAILIGL CLTLLLLAIF
PSN_DROME_92-532   IKLGLGDFIF YSVLVGKASS Y..GDWTTTI ACFVAILIGL CLTLLLLAIW
HOP1_CAEEL_1-349   IRLGMGDFVF YSLMLGNTVQ T..CPLPTVV ACFVSNLVGL TITLPIVTLS
SPE4_CAEEL_5-465   LRLGFGDFVF YSLLIGQAAA S..GCPFAVI SAALGILFGL VVTLTVFSTE


                   451                            480
query_65-453       KKALPALPIS ITFGLVFYFA TDYLVQPFMD
PSN_CAEEL_38-435   KRALPALPIS IFSGLIFYFC TRWIITPFVT
PSN2_RAT_76-439    KKALPALPIS ITFGLIFYFS TDNLVRPFMD
PSN1_RAT_70-459    KKALPALPIS ITFGLIFYFA TDYLVQPFMD
PSN_DROME_92-532   RKALPALPIS ITFGLIFCFA TSAVVKPFME
HOP1_CAEEL_1-349   QTALPALPFP LAIAAIFYFS SHIALTPFTD
SPE4_CAEEL_5-465   ESTTPALPLP VICGTFCYFS SMFFWEQLYG
```

Table 8. Unprocessed sequence alignment scheme for Psen1 [Xenopus tropicalis]

5. Discussion and Conclusion

γ-Secretase is an enzymatic complex which is critical to amyloid formation and other important pathophysiological processes. Non-specific γ-secretase inhibitors are not ideal for the intervention of familial Alzheimer's disease. The four essential inhibitors include Presenilin (PS1 or PS-2), APH-1, Pen-2, nicastrin (Nct). Also γ-secretase cleaves a variety of substrates (over 30). But different sites are cleaved even in a single substrate and no consensus sequence for cleavage by γ-secretase is identified. We predict consensus structural sequence in components of γ-secretase and its substrates should exist during γ-secretase regulation and specific γ-secretase inhibitor based on such consensus may act as a therapeutic intervention for familial Alzheimer's disease. We accomplish Stage I and successfully verify Stage II of our initial proposal. We find the candidates by the comparative methods of bioinformatics. By using the resources of motif identification and domain prediction databases and tools, selecting software to apply sequence analysis, modeling and simulation, and pattern recognition of γ-secretase regulation, we find out the candidates of consensus sequences. We collaborate to use the traditional biological methods and verify the critical role of the candidates including the protein purification, recognition, and functional domain interaction.

Reference

Bentahir M. et al. Presenilin clinical mutations can affect γ-secretase activity by different mechanisms, Journal of Neurochemistry (2006) 96, 732-742

Blanchard BJ, Hiniker AE, Lu CC, Margolin Y, Yu AS, Ingram VM (2000). "Elimination of Amyloid beta Neurotoxicity.". J Alzheimers Dis 2 (2): 137-149.

Blanchard B, Chen A, Rozeboom L, Stafford K, Weigele P, Ingram V (2004). "Efficient reversal of Alzheimer's disease fibril formation and elimination of neurotoxicity by a small molecule.". Proc Natl Acad Sci U S A 101 (40): 14326-32.

Laurence Eng-Chee Cheng, Francis Ka-Ming Chan, Dragana Cado and Astar Winoto1, "Functional redundancy of the Nur77 and Nor-1 orphan steroid receptors in T-cell apoptosis" The EMBO Journal Vol.16 No.8 pp.1865–1875, 1997

Citron M. et al. Mutant presenilins of Alzheimer's disease increase production of 42-residue amyloid β-protein in both transfected cells and transgenic mice, Nature medicine, vol 3, number 1, January 1997

Espeseth A, Xu M, Huang Q, Coburn C, Jones K, Ferrer M, Zuck P, Strulovici B, Price E, Wu G, Wolfe A, Lineberger J, Sardana M, Tugusheva K, Pietrak B, Crouthamel M, Lai M, Dodson E, Bazzo R, Shi X, Simon A, Li Y, Hazuda D (2005). "Compounds that bind APP and inhibit Abeta processing in vitro suggest a novel approach to Alzheimer disease therapeutics.". J Biol Chem 280 (18): 17792-7.

Evin G, Weidemann A. (2002) Biogenesis and metabolism of Alzheimer's disease Aβ amyloid peptides. Peptides; 23 (7): 1285-97

Francis R, McGrath G, Zhang J, Ruddy DA, Sym M, Apfeld J, Nicoll M, Maxwell M, Hai B, Ellis MC, Parks AL, Xu W, Li J, Gurney M, Myers RL, Himes CS, Hiebsch R, Ruble C, Nye JS, Curtis D. (2002). aph-1 and pen-2 are required for

Notch pathway signaling, gamma-secretase cleavage of betaAPP, and presenilin protein accumulation. Dev Cell 3(1):85-97.

Games D, Adams D, Alessandrini R, Barbour R, Berthelette P, Blackwell C, Carr T, Clemens J, Donaldson T, Gillespie F (1995). "Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein.". Nature 373 (6514): 523-7.

Gibson, T. J. and Spring, J. "Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins." Trends Genet. 14:46–49.1998

Goutte C, Tsunozaki M, Hale VA, Priess JR. (2002). APH-1 is a multipass membrane protein essential for the Notch signaling pathway in Caenorhabditis elegans embryos. Proc Natl Acad Sci USA 99(2):775-9.

Gu Z, Steinmetz L, Gu X, Scharle, C, Dacis, R and Li W. Role of Duplicate Genes in Genetic Robustness against Null Mutations. Nature 421:63-66 2003

Hasegawa H, Sanjo N, Chen F, Gu YJ, Shier C, Petit A, Kawarai T, Katayama T, Schmidt S, Mathews P, Schmitt-Ulms G, Fraser PE, St George-Hyslop P. (2004). Both the Sequence and Length of the C Terminus of PEN-2 Are Critical for Intermolecular Interactions and Function of Presenilin Complexes. J Biol Chem 279(45): 46455-46463

Hashimoto M, Rockenstein E, Crews L, Masliah E (2003). "Role of protein aggregation in mitochondrial dysfunction and neurodegeneration in Alzheimer's and Parkinson's diseases.". Neuromolecular Med 4 (1-2): 21-36.

Heber Sabine, Jochen Herms, Vladan Gajic, Johannes Hainfellner, Adriano Aguzzi,3 Thomas Rulicke, Hans Kretzschmar, Cornelia von Koch, Sangram Sisodia,

Phillippe Tremml, Hans-Peter Lipp. "Mice with Combined Gene Knock-Outs Reveal Essential and Partially Redundant Functions of Amyloid Precursor Protein Family Members." The Journal of Neuroscience, November 1, 2000, 20(21):7951–7963

Hsiao K, Chapman P, Nilsen S, Eckman C, Harigaya Y, Younkin S, Yang F, Cole G (1996). "Correlative memory deficits, Abeta elevation, and amyloid plaques in transgenic mice.". Science 274 (5284): 99-102.

Jie Shen et al. Skeletal and CNS defects in presenilin-1 deficient mice, Cell, Vol. 89, 629–639, May 16, 1997

Joaquín Grego-Bessa et al (Mar 2007). "Notch Signaling Is Essential for Ventricular Chamber Development". Developmental Cell 12 (3): 415-429.

Kanapathipillai M, Lentzen G, Sierks M, Park C (2005). "Ectoine and hydroxyectoine inhibit aggregation and neurotoxicity of Alzheimer's beta-amyloid.". FEBS Lett 579 (21): 4775-80.

Kaether C, Haass C, Steiner H. (2006). Assembly, trafficking and function of gamma-secretase. Neurodegener Dis 3(4-5):275-83.

Von Koch CS, Zheng H, Chen H, Trumbauer M, Thinakaran G, van der Ploeg LH, Price DL, Sisodia SS (1997) Generation of APLP2 KO mice and early postnatal lethality in APLP2/APP double KO mice. Neurobiol Aging 18:661– 669.

Lee K, Shin B, Shin K, Kim D, Yu J (2005). "A hybrid molecule that prohibits amyloid fibrils and alleviates neuronal toxicity induced by beta-amyloid (1-42).". Biochem Biophys Res Commun 328 (4): 816-23.

Lee SF, Shah S, Yu C, Wigley WC, Li H, Lim M, Pedersen K, Han W, Thomas P, Lundkvist J, Hao YH, Yu G. (2004). A conserved GXXXG motif in APH-1 is critical for assembly and activity of the gamma-secretase complex. *J Biol Chem* 279(6):4144-52.

Lott I, Head E (2005). "Alzheimer disease and Down syndrome: factors in pathogenesis.". Neurobiol Aging 26 (3): 383-9.

Luo WJ, Wang H, Li H, Kim BS, Shah S, Lee HJ, Thinakaran G, Kim TW, Yu G, Xu H. (2003). PEN-2 and APH-1 coordinately regulate proteolytic processing of presenilin 1. J Biol Chem 278(10):7850-4.

Maier D, Marte BM, Schafer W, Yu Y and Preiss A, "D Maier, BM Marte, W Schafer, Y Yu and A Preiss", Proceedings of the National Academy of Sciences, 90:5464-5468.1993

Masliah E, Sisk A, Mallory M, Mucke L, Schenk D, Games D (1996). "Comparison of neurodegenerative pathology in transgenic mice overexpressing V717F beta-amyloid precursor protein and Alzheimer's disease.". J Neurosci 16 (18): 5795-811.

Mudher A, Lovestone S (2002). "Alzheimer's disease-do tauists and baptists finally shake hands?". Trends Neurosci 25 (1): 22-6.

Nistor M, Don M, Parekh M, Sarsoza F, Goodus M, Lopez GE, Kawas C, Leverenz J, Doran E, Lott IT, Hill M, Head E (2006). "Alpha- and beta-secretase activity as a function of age and Aβ in Down syndrome and normal brain.". Neurobiol Aging (epub). PMID 16904243.

Ohnishi S, Takano K (2004). "Amyloid fibrils from the viewpoint of protein folding.". Cell Mol Life Sci 61 (5): 511-24.

Pardossi-Piquard R, Dunys J, Yu G, St George-Hyslop P, Alves da Costa C, Checler F. (2006). Neprilysin activity and expression are controlled by nicastrin. J Neurochem 97(4):1052-6.

Polvikoski T, Sulkava R, Haltia M, Kainulainen K, Vuorio A, Verkkoniemi A, Niinist? L, Halonen P, Kontula K (1995). "Apolipoprotein E, dementia, and cortical deposition of beta-amyloid protein.". N Engl J Med 333 (19): 1242-7.

Porat Y, Abramowitz A, Gazit E (2006). "Inhibition of amyloid fibril formation by polyphenols: structural similarity and aromatic interactions as a common inhibition mechanism.". Chem Biol Drug Des 67 (1): 27-37.

Rudolph Tanzi and Ann Parson, (2000). "Decoding Darkness: The Search for the Genetics Causes of Alzheimer's Disease", Perseus Press.

Schmitz C, Rutten B, Pielen A, Schafer S, Wirths O, Tremp G, Czech C, Blanchard V, Multhaup G, Rezaie P, Korr H, Steinbusch H, Pradier L, Bayer T (2004). "Hippocampal neuron loss exceeds amyloid plaque load in a transgenic mouse model of Alzheimer's disease." Am J Pathol 164 (4): 1495-502.

Shen Z (2004). "Brain cholinesterases: II. The molecular and cellular basis of Alzheimer's disease." Med Hypotheses 63 (2): 308-21.

Sherrington R. et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease, Nature, Vol 375, 29 June 1995

Smialowska, A.,Baumeister, R., Presenilin function in Caenorhabditis elegans, Neurodegenerative Diseases. Vol. 3(4-5)(pp 227-232), 2006.

Spasic D. et al. Presenilin-1 maintains a nine transmembrane topology throughout the secretary pathway, J. Biol. Chem., Vol. 281, Issue 36, 26569-26577, September 8, 2006

Steinbach JP, Muller U, Leist M, Li Z-W, Nicotera P, Aguzzi A (1998) Hypersensitivity to Seizures in b-amyloid precursor protein deficient mice. Cell Death Diff 5:858–866.

St George-Hyslop PH. (2000) Genetic factors in the genesis of Alzheimer's disease. Ann N Y Acad Sci; 924: 1-7

Walker LC, Rosen RF (2006). "Alzheimer therapeutics: What after the cholinesterase inhibitors?". Age Ageing 35: 332-335.

Wenk G. (2006) "Neuropathologic changes in Alzheimer's disease.". J Clin Psychiatry 64 Suppl 9: 7-10.

White AR, Zheng H, Galatis D, Maher F, Hesse L, Multhaup G, 7962 J. Neurosci., November 1, 2000, 20(21):7951–7963 Heber et al. Mice with Combined Gene Deficiencies of APP Family Members Beyreuther K, Masters CL, Cappai R (1998) Survival of culturedneurons from amyloid precursor protein knock-out mice against Alzheimer's amyloid-beta toxicity and oxidative stress. J Neurosci 18:6207– 6217.

Yankner B, Duffy L, Kirschner D (1990). "Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides.". Science 250 (4978): 279-82.

Zhang YW, Luo WJ, Wang H, Lin P, Vetrivel KS, Liao F, Li F, Wong PC, Farquhar MG, Thinakaran G, Xu H. (2005). Nicastrin is critical for stability and trafficking but

not association of other presenilin/gamma-secretase components. J Biol Chem 280(17):17020-6.