Identification of a novel family of presenilin homologues

Chris P. Ponting1,*, Mike Hutton2, Andrew Nyborg2, Matthew Baker2, Karen Jansen2 and Todd E Golde2

1MRC Functional Genetics Unit, University of Oxford, Department of Human Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK and 2Mayo Clinic Jacksonville, Department of Neuroscience, 4500 San Pablo Road, Jacksonville, FL 32224, USA

Received December 11, 2001; Revised and Accepted February 25, 2002

Presenilin 1 and presenilin 2 are polytopic membrane proteins, whose genes are mutated in some individuals with Alzheimer’s disease. Presenilins have been shown to influence limited proteolysis of amyloid β protein precursor (APP), Notch and ErbB4, and have been proposed to be γ-secretases that perform the terminal cleavage of APP. In this model, two conserved and apparently intramembranous aspartic acids participate in catalysis. Highly sequence-similar presenilin homologues are known in plants, invertebrates and vertebrates. In this work, we have used a combination of different sequence database search methods to identify a new family of proteins homologous to presenilins. Members of this family, which we term presenilin homologues (PSH), have significant sequence similarities to presenilins and also possess two conserved aspartic acid residues within adjacent predicted transmembrane segments. The PSH family is found throughout the eukaryotes, in fungi as well as plants and animals, and in archaea. Five PSHs are detectable in the human genome, of which three possess ‘protease-associated’ domains that are consistent with the proposed protease function of PSHs. Based on these findings, we propose that PSs and PSHs represent different sub-branches of a larger family of polytopic membrane-associated aspartyl proteases.

INTRODUCTION

Presenilin 1 (PS1) and presenilin 2 (PS2) are members of an evolutionarily conserved family of polytopic membrane proteins that are thought to span the membrane six to eight times (reviewed in 1). Presenilins were initially identified through genetic linkage analysis of families with autosomal dominant forms of Alzheimer’s disease (AD). To date, numerous mutations in the chromosome 14-encoded PS1 gene and several in the chromosome 1-encoded PS2 gene have been linked to rare forms of familial AD (see http://www.alzforum.org/members/resources/pres_mutations/index.html).

Studies of PSs and their role in the development of AD have led to an appreciation that they regulate a number of diverse functions in the cell. PSs have been shown to play a role in proteolytic processing of amyloid β protein precursor (APP) and APP family members (2,3), Notch and Notch family members (4,5), and ErbB4 (6). In all cases, PSs appear to be required to permit the release of the cytoplasmic tail of these proteins from a membrane-bound stub that is generated after ectodomain cleavage of the holoprotein. Significantly, inhibition of the release of the membrane-bound stub of these proteins alters signaling cascades that appear to be mediated by translocation of the cytoplasmic tail to the nucleus (6–8). PSs have also been shown to interact with a wide array of different proteins, and have been implicated in IP3-mediated release of endoplasmic reticulum (ER) calcium (9), capacitative calcium entry (10,11), β-catenin signaling (12–14) and protein trafficking (15).

Much work on PSs has focused on their role in processing APP (1). PSs have been shown to influence the final cleavage step in the release of the approximately 4 kDa amyloid β protein (Aβ) from APP, a proteolysis that is referred to as γ-secretase cleavage. This event generates Aβ from a membrane stub of the APP, which in turn is produced by a previous hydrolysis of the APP holoprotein by a membrane-bound aspartyl protease referred to as β-secretase. Significantly, γ-secretase generates Aβ peptides of varying lengths, with the two species of most interest being peptides of 40 amino acids (Aβ40) and 42 amino acids (Aβ42). In all cases studied to date, mutant PSs increase the relative level of Aβ42 production. As the longer Aβ42 aggregates more rapidly than other forms of Aβ and had previously been shown to be increased by AD-linked mutations in the APP, these studies not only implicated PSs as modulators of γ-secretase cleavage, but

*To whom correspondence should be addressed: Tel: +44 1865 272175; Fax: +44 1865 282651; Email: Chris.Ponting@anat.ox.ac.uk
also convinced many in the AD field that this longer form of Aβ was critical to AD pathogenesis (16).

Subsequent to these initial studies on AD-linked PS mutations, three lines of evidence have emerged to support a direct role of PS in γ-secretase cleavage of APP, as well as the γ-secretase cleavage that regulates the release of the cytoplasmic tails of Notch and ErbB4. First, at the genetic level, PSs appear to be essential for γ-secretase activity (17), probably owing to compensation by the more widely expressed PS1. However, a combined PS1/PS2 knockout completely abolishes γ-secretase activity (18,19). Second, biochemical evidence also supports the identification of PSs as γ-secretases. PSs co-fractionate with γ-secretase activity in a high-molecular-weight complex, and this in vitro activity can be recovered in the pellet after immunoprecipitation with antibodies to PS1 (20). In addition, mutation of two absolutely conserved aspartate residues (D257 and D385 in PS1, and D263 and D366 in PS2) results in PSs that act in a dominant-negative fashion – at least with respect to γ-secretase activity (21,22). This leads to the hypothesis that PSs represent a novel class of intramem-branous cleaving diapspartyl proteases, a hypothesis supported by inhibitor studies (23,24). Third, four groups have been able to show that inhibitors of γ-secretase activity bind with varying degrees of specificity to both PS1 and PS2 (25–28). Collectively, these experiments indicate that compounds, including some with properties of protease inhibitors, can bind to PSs. Moreover, given that two groups had specifically designed their inhibitors to be transition state analogues of aspartyl proteases, it certainly appears that, despite the lack of demonstrable homology to known aspartyl proteases, PSs may be just that.

More recently, the primary sequence encompassing one of the aspartate residues (D385 in PS1 and D366 in PS2) in PSs has been proposed to be similar to a region of type IV prepsin peptidases that contain a catalytic aspartate residue (29,30). This sequence motif similarity, and their common feature of spanning the membrane several times, suggests that PSs and the bacterial signal peptidases may be distinct members of a novel class of intramembranous proteases. Despite their local sequence similarities, PSs and type IV prepsin peptidases are not demonstrably homologous. In order to investigate whether PSs possess hitherto unappreciated divergent homologues, we undertook searches of protein and nucleotide sequence databases using PSI-BLAST (31) and a hidden Markov model package (32). These searches revealed a family of sequences with significant similarities to presenilins. This observation may assist in elucidating the molecular and cellular functions of PSs.

RESULTS

Human PS1 and PS2 sequences were compared with a non-redundant database (NR) using PSI-BLAST (31) and an E-value threshold of 2 × 10^{-3}. This search revealed that a mouse sequence of unknown function (GenInfo code 12849450) was marginally similar, albeit with non-significant statistics, to PS1 and PS2 (E = 1.8 and 2.3, respectively; 2 search rounds). When this mouse sequence was compared with NR, it was evident that it is but one of a large family of proteins with representatives among the fungi, plants, arthropods, nematode worms, vertebrates and archaea. This family showed three intriguing similarities with PSs. First, they are both families of transmembrane (TM) proteins whose sequences can be aligned almost throughout their entire lengths (Fig. 1). Second, they both contain conserved YD, GhGD and PAL motifs (h represents a hydrophobic residue) in the same colinear arrangement (Figs 1 and 2). Third, the YD and GhGD motifs, which in PSs include their candidate active site aspartic acid residues (21), are unusual in that they are present within predicted adjacent TM segments.

In order to investigate whether the newly identified protein family is homologous to PSs, we constructed its multiple sequence alignment and compared a global hidden Markov model (32), derived from the alignment, with a second non-redundant database, nrdb90. Results showed that presenilins represented all of the highest-scoring sequences, including a plant presenilin (Arabidopsis thaliana gene product At2g29900) whose sequence similarity to the family was significant (E = 5.3 × 10^{-3}). These significant sequence similarities demonstrate that the eukaryotic and archaeal proteins represent a novel family of presenilin homologues (PSHs).

Five distinct PSH genes were evident from the human genome sequence. PSH genes were found in fungi, such as Saccharomyces cerevisiae, and in archaea such as Archaeoglobus fulgidus. This is the first time that PS homologues have been detected in eukaryotes other than plants and animals, and in prokaryotes.

As predicted from their presence in the expressed sequence tag (EST) database, the human PSH RT-PCR experiments confirm that each of the PSHs is expressed in humans (Fig. 3). PCR products can be amplified both from a human brain cDNA library and from H4 cDNA. In these experiments, the H4 mRNA was treated with DNase I to remove trace DNA contamination prior to reverse transcription, and non-reverse-transcribed DNase I-treated RNA was used as control to show that the product originated from mRNA and not genomic DNA contamination. Thus, like other PSHs, the intronless PSH2 mRNA is expressed. Full-length cDNAs for each homologue can also be amplified from H4 cell cDNA (data not shown).

The proximity of PSH2 to the MAPT gene, which codes for the microtubule-associated protein tau, on chromosome 17 suggested that it might be a candidate gene for the variant of FTDP-17 that occurs in families that lack tau pathology and obvious mutations in tau (33–35). Sequence analysis was therefore performed in the probands from four families with frontotemporal dementia (FTD) with similar clinical and pathological phenotypes to these tau-negative families. However, in each case, the FTD family was not large enough to confirm linkage to chromosome 17. In addition, PSH2 was sequenced in control samples with H1/H1 and H2/H2 MAPT genotypes to identify single nucleotide polymorphisms (SNPs) in linkage disequilibrium with these haplotypes in the adjacent MAPT gene. This analysis demonstrated that
Figure 1. Multiple sequence alignment of presenilins (PSs; top) and PS homologues (PSHs; below) represented using CHROMA (http://www.lg.direc.co.uk/chroma/) and a 75% consensus. Alignment of the sequences is made using the 8-transmembrane (TM) model for PSs. PS1 and PS2 amino acids that are substituted in individuals with familial Alzheimer's disease (http://www.alzforum.org/members/resources/pres_mutations/) are shown in white-on-red. Amino acid changes due to non-synonymous single nucleotide polymorphisms in human PSH2 and PSH3 are shown in white-on-blue; the amino acid substitution (Ala-to-Pro) in PSH3 just prior to the putative active site residue-containing ‘LGLGD motif’ might be deleterious to function. Aspartic acids that are thought to be active site residues are shown in white-on-magenta and are indicated by ‘.’. The positions of seven of the predicted eight TM segments in PSs are shown; these correspond to the seven TM segments in PSHs. Numbers in parentheses represent numbers of amino acids excised from the alignment. ‘–’ represents a gap. The GenInfo numbers and amino acid limits of sequences are shown following the alignment. Proteins with asterisks (*) contain N-terminal protease-associated (PA) domains. Protein and/or gene names are shown followed by species abbreviations: Af, Archaeoglobus fulgidus; At, Arabidopsis thaliana; Ce, Caenorhabditis elegans; Dm, Drosophila melanogaster; Dr, Danio rerio (zebrafish); Hl, Helix lucorum (a mollusc); Hs, Homo sapiens; Hsp, Halobacterium sp. NRC-1; Sp, Schizosaccharomyces pombe; Ta, Thermoplasma acidophilum; Tv, Thermoplasma volcanium; Xl, Xenopus laevis.

**Human Molecular Genetics, 2002, Vol. 11, No. 9 1039**
PSH2 is highly polymorphic, with 12 SNPs being detected, of which nine alter the encoded amino acid sequence (Table 1). Nine of the PSH2 SNPs were apparently in complete linkage disequilibrium with the H1/H2 haplotypes in the adjacent MAPT gene. However, none of the SNPs were obviously pathogenic in the four FTD families, since all were present in control individuals or did not segregate with disease (R123Q and E209K).

Figure 1. (Continued).

Figure 2. Schematic representations of the predicted transmembrane topologies of presenilin 1 and human PSH2, which contains a protease-associated (PA) domain. The topology predictions are made on the basis of identifications of signal peptides and PA domains in PSHs and on the 8-TM model of PS topologies. The pink stars represent the putative active site aspartic acid residues in PSs and PSHs, and the red triangles represent conserved PAL(P) motifs.
of presenilins (36). Several animal and plant PSHs were found to contain N-terminal protease-associated (PA) domains that often co-occur with peptidase domains (37). This is consistent with a hypothesis that PSHs and, also by homology, PSs possess peptidase function.

Of the five human PSH genes, the PSH1 gene neighbors a locus on chromosome 12q24 with linkage to bipolar risk (38). PSH2 is the most proximal gene to the MAPT gene, encoding the tau protein, on chromosome 17q21.31. Hyperphosphorylated forms of tau are often found in the brains of patients with AD, and MAPT mutations are associated with frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17) (39,40). On one hand, the close proximity of a PS homologue to a gene whose dysfunction is intimately associated with neurodegenerative conditions might be coincidental. On the other hand, the close proximity (~50 kb) of these genes might be significant, since sequence analysis has revealed multiple (nine) SNPs in PSH2 that are in complete linkage disequilibrium with the extended H1/H2 haplotypes in the MAPT gene (41). The H1 haplotype is a genetic risk factor for the development of the sporadic tauopathies progressive supranuclear palsy (41) and cortical basal degeneration (42). In addition, the location of PSH2 might be relevant to families that have been described with FTDP-17 lacking tau pathology and without obvious mutations in the MAPT gene (33–35). Sequencing of PSH2 in multiple MAPT mutation-negative FTD families that have similar clinical and pathological phenotypes to these FTDP-17 families has thus far failed to reveal any evidence of pathogenic mutations. However, none of the sequenced tau-negative FTD families is large enough to demonstrate linkage to chromosome 17q, and thus we cannot exclude PSH2 as a candidate gene for ‘tau-negative’ FTDP-17. The remaining three PSH genes are on chromosomes 20q11.21, 19p13.3 and 15q21.2.

All of the PSHs appear to be widely expressed, since each is represented by numerous ESTs from diverse tissues and cell lines. Significantly, even though the intronless PSH2 might be predicted to be a non-translated pseudogene resulting from retrotransposition of an ancestral PSH mRNA, it does in fact appear to be expressed, at least at the mRNA level. In these initial non-quantitative RT–PCR experiments, expression of PSH2 appeared to be lower than that of other PSHs. More rigorous analysis of relative expression levels will be needed to determine the relative abundance of each of the PSH transcripts.

The identification of a novel family of PS homologues is likely to lead to insights into PS function and evolution. By inference from PSs, a diasparyl peptidase function for PSHs is plausible. However it is perhaps unlikely that PSHs cleave known PS substrates, since the PSH family, represented among the archaea and eukaryotes, appear to have arisen considerably before the appearance of PSs, Aβ, Notch and ErbB4. Moreover, double knockout of PS1 and PS2 appears sufficient to abolish γ-secretase cleavage of APP and Notch (18,19). In any case, additional experiments will need to be conducted to determine whether or not any of the PSHs can cleave known γ-secretase substrates. Based on their apparent expression in the brain, it is unlikely that known γ-secretase substrates are major substrates for the PSHs. However, as has been shown for other enzyme families, the identification of a wider repertoire of PS/PSH

---

**DISCUSSION**

The newly identified PSHs could be aligned over their entire lengths with PSs, encompassing PS TM segments 2–8 (Fig. 2). Despite their apparent lack of TM segment 1, they appear to have the same membrane topology as PSs as a consequence of their signal peptide sequences. Regions that are highly conserved between the two families include not only the putative active site aspartic acid residues, but also TM sequences and a ‘PALP’ motif that is reported to be obligatory for stabilization, complex formation and γ-secretase activities.
family members may now facilitate a greater understanding of the molecular and cellular functions of this biomedically important family. In particular, given the interest in developing γ-secretase inhibitors that target PSs as potential therapeutic agents in A D (1), information gathered by studying this protein family may offer important insights into the potential catalytic mechanisms and substrate-specificities of PS and PSH that will aid the development of such inhibitors.

**MATERIALS AND METHODS**

**Bioinformatic methods**

Known PS sequences were compared with a non-redundant protein sequence database (NR; ftp://ftp.ncbi.nih.gov/pub/blast/db/nr; 808 320 sequences) using PSI-BLAST (31) and an E-value threshold of 2 × 10⁻5. [An E-value (or expect value) for a given alignment score x represents the number of alignments with scores x or higher that are expected purely by chance in the database search.] Similar PSI-BLAST searches determined the extent of the protein family homologous to a mouse sequence (Geninfo code 12849450). A multiple sequence alignment of this family’s sequences was constructed using Clustal-W (43) and manually adjusted to minimize gap positions within secondary structures. A hidden Markov model of this alignment was constructed using HMMER (32) and compared with a second non-redundant protein sequence database (nrdb90; ftp://ftp.ebi.ac.uk/pub/databases/nrdb90/; 392 089 sequences).

Further members of the PSH sequence family were detected by searching a non-redundant nucleotide sequence (NT; ftp://ftp.ncbi.nih.gov/pub/blast/db/nt) or expressed sequence tag (ftp://ftp.ncbi.nih.gov/pub/blast/db/est) databases using BLAST. The positions of PSH genes in the human genome were determined using BLAT (http://genome.cse.ucsc.edu/cgi-bin/hgBlat?command=start). Signal peptides were detected using SignalP (http://www.cbs.dtu.dk/services/SignalP/), and co-occurring domains were identified using Pfam (http://www.sanger.ac.uk/Pfam/). Transmembrane topologies of PSs and PSHs were predicted using algorithms such as TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and TM PRED (http://www.ch.embnet.org/software/TMPRED/ form.html).

**RT-PCR**

RNA was extracted from H4 (human neuroglioma) cells using the RNeasy RNA extraction kit (Qiagen, Chatsworth, CA) according to the manufacturer's instructions. RNA was treated with DNase using the DNA-free kit (Ambion, Austin TX), and cDNA was synthesized from 5 μg of RNA using Superscript Reverse Transcriptase (Invitrogen, Carlsbad, CA), also according to the manufacturer’s instruction. A whole-brain cDNA library was purchased from Invitrogen. PCR was performed with Taq DNA polymerase (Invitrogen) on a Hybaid MBS 0.2G thermocycler using a touchdown program with the following cycling parameters: 10 cycles annealing at 60°C, followed by 25 cycles annealing at 55°C. Primers for the PCR reactions are shown in Table 2. PCR products were analyzed on a 2% NuSieve gel (BioWhittaker, Rockland, ME).

**PCR and sequence analysis**

Primers were designed to generate three separate products: Product #1 was amplified using primers 1F and 3R, giving a 1126 bp product, then sequenced with 1F, 2R, 3F, 3R, and 1R. Product #2 was generated using primers 3F and 5R, giving a 1088 bp product, then sequenced with 3F, 4F, 4R, 5R, 1R, and 7R. The primer sequences are shown in Table 3. Reactions contained 50 ng of genomic DNA in a 50 μl mixture containing 20 pmol of each primer, 0.2 mM dNTPs, 1 unit of Taq buffer and Q-solution (Qiagen, Califonia, UK). Amplifications were performed oil-free in Hybaid Touchdown thermal cyclers (Hybaid, Cambridge, UK). Conditions were 35 cycles of 94°C for 30 s, 60°C to 50°C touchdown annealing for 30 s, and 72°C for 45 s, with a final extension of 72°C for 10 min. All products were

**Table 2. PCR primers for amplification of PSH**

<table>
<thead>
<tr>
<th>mRNA</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Product (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSH1</td>
<td>5’GGCTTGCTCTCTCTATGC3’</td>
<td>5’CCAAATAGGAAGGGCGGG3’</td>
<td>218</td>
</tr>
<tr>
<td>PSH2</td>
<td>5’GAGCTGCCCTTATCCCC3’</td>
<td>5’CAAGGGAGCTGTTGACAT3’</td>
<td>226</td>
</tr>
<tr>
<td>PSH3</td>
<td>5’GCTTTGAGCTCACCTTCG3’</td>
<td>5’TTCCTTTTTTTTCCAGGCC3’</td>
<td>309</td>
</tr>
<tr>
<td>PSH4</td>
<td>5’CTGCACCATCGCCTATGGC3’</td>
<td>5’GGACCTTGCAAAAGCCGCT3’</td>
<td>179</td>
</tr>
<tr>
<td>PSH5</td>
<td>5’GGTGCAGTGGAAAAGGAGG3’C</td>
<td>5’TGTCTGCAAGTCGTTCACC3’</td>
<td>206</td>
</tr>
<tr>
<td>Actin</td>
<td>5’GGCAACCGCTCATTGCC3’</td>
<td>5’ACCACACTTGCCACCTCA3’</td>
<td>290</td>
</tr>
</tbody>
</table>

**Table 3. PCR and internal sequencing primers**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F</td>
<td>GACTGCTGTGTGACAAAAAGG</td>
</tr>
<tr>
<td>1R</td>
<td>GTGAAAGCTGAGTTAACCC</td>
</tr>
<tr>
<td>2F</td>
<td>ATGAGGCACCAGCGGAC</td>
</tr>
<tr>
<td>3F</td>
<td>TTCTCTGAGGAGTAATCC</td>
</tr>
<tr>
<td>3R</td>
<td>CAGGAGCTGTCCTATGGGACG</td>
</tr>
<tr>
<td>4F</td>
<td>CTGCTCTGCGGCAATAC</td>
</tr>
<tr>
<td>4R</td>
<td>CACATACGAAGCGGCAACAG</td>
</tr>
<tr>
<td>5F</td>
<td>CTGCAGCTGTGCCGCTGCA</td>
</tr>
<tr>
<td>5R</td>
<td>ATCTGCTGTGCCCCCTCTGCTT</td>
</tr>
<tr>
<td>6F</td>
<td>AGCTTACGGCTCTTCTGGAG</td>
</tr>
<tr>
<td>6R</td>
<td>TGAAGGCTGACAGCAAGATG</td>
</tr>
<tr>
<td>7F</td>
<td>GACCCGAGCTGATACCTCC</td>
</tr>
<tr>
<td>7R</td>
<td>ATGCCCCAGATTCTGTGTG</td>
</tr>
</tbody>
</table>
References


